Volume 3 Issue 2

ISSN: 2321-1857

International Journal of

Business Analytics & Intelligence

October 2015

Introducing

'Analytically Yours'

A Column on Analytics

By

Prof. Arnab Laha

IIM Ahmedabad

A Publication of Publishing India Group

International Journal of Business Analytics & Intelligence

Editorial Message

Dear Readers,

It gives me immense pleasure to announce that Professor Arnab Laha of IIM Ahmedabad has started a new column on analytics at IJBAI. His column, Analytically Yours, will now be an integral component of IJBAI. In the present issue, he reflects on "Interval-Valued Data Analysis" in the Big Data context with a couple of real life scenarios.

In consonance with our reputation of publishing scholastic research papers, the current issue features research papers on a wide range of applications like healthcare, supply chain, retail, finance, information technology and port. The first paper on Multi Attribute Group Decision Making (MAGDM) problem explores the application of TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) and Borda's function for supplier selection. The second paper applies machine learning techniques like Random Forest (RF) along with Gradient Boosting Machine (GBM) and Support Vector Regression (SVR) on health care data to forecast Zoonotic Disease. The third paper applies regression to identify the impact of the ratios of DuPont model (profit margin, capital turnover, financial cost ratio, finance structure ratio and tax effect ratio) on Return on Equity for IT sector in India. The fourth paper on applying analytics at ports introduces system dynamics to our readers and figures out the impact of Inefficiency Level (IL) on Average Turn Around Time (ATRT). The fifth paper applies structural equation modelling to understand the green supply chain in retail domain. The sixth paper is on financial analytics where the authors explore the effectiveness of investment in intellectual capital of Indian knowledge companies.

Thus, we are exposing our readers to diverse applications of analytics techniques to solve business problems. I am sure that our readers will enjoy and learn a lot from the present issue. Do let us know your thoughts and views.

Merry Christmas and Happy New Year!

Sincerely yours,
Tuhin Chattopadhyay
Editor-in-Chief
International Journal of Business Analytics and Intelligence (IJBAI)

International Journal of Business Analytics and Intelligence

Volume 3 Issue 2 October 2015

ISSN: 2321-1857

Column – Analytically Yours

Research Papers

Journal is available online at www.publishingindia.com

Article can be accessed online at http://www.publishingindia.com

Analytically Yours

Interval-Valued Data Analysis

Arnab Kumar Laha*

Interval-valued data arise in real-life in many different ways. Weather data published daily in newspapers contains only the maximum and minimum temperature readings during a day for a city, stock market data contains the highest and lowest traded price of a stock or a stock-index on a day etc. In the Big Data context often data are aggregated for certain features of interest giving rise to interval-valued data. For example, a credit card company need not store the entire history of credit card transactions of its customers but instead may store only the minimum and maximum amounts transacted by them. Thus for each customer the bank stores an interval which is the range of the amounts transacted by the customer.

You must now be wondering about the kind of decision problems that can be solved by using such interval data. Here are two examples:

1) Imagine yourself to be an investor in the stock market. You may like to know the usual range of variation of returns on a certain stock during the course of a day to decide whether to invest in that stock or not.

2) Imagine yourself to be a functionary of a credit card company in-charge of preventing credit card frauds. You would be interested in the usual range of amounts transacted by customers of different categories. If the transaction amount for a transaction falls outside the range specified for customers of that category, then you may take necessary action to verify the credentials of the person making the transaction before allowing the same.

Interval-valued data is a special type of Symbolic data. There can be many other forms of Symbolic data as discussed in Billard, 2011. In this short note we discuss analysis of interval-valued data with some applications.

Let $[a_i, b_i]$, i=1,...,n be a random sample of n intervals. We are interested in providing a summary of the data by providing a representative "mean interval". A naive approach is to use $[\tilde{a},\tilde{b}]$ as the "mean interval" where $\tilde{a} = \frac{a_1 + ... + a_n}{n}$ and $\tilde{b} = \frac{b_1 + ... + b_n}{n}$. While this method is expected to perform well for datasets in which there are no outliers, it may perform quite badly in case the dataset has outliers. A robust alternative can be to use the interval $[a',b']$ where $a' = \text{median}\{a_1,...,a_n\}$ and $b' = \text{median}\{b_1,...,b_n\}$. A third approach described by Le-Rademacher and Billard (2011) is described below.

Assuming that the variable of interest (say, temperature) of the i-th day is uniformly distributed in the interval $[a_i, b_i]$ we can compute the mean $\Theta_{1i} = (a_i+b_i)/2$ and the variance $\Theta_{2i} = (b_i - a_i)^2/12$. The random variables $\mathbf{\Theta}_i = (\Theta_{1i}, \Theta_{2i})$ are referred to as the internal parameters of the i-th interval $[a_i, b_i]$. Le-Rademacher and Billard (2011) assumes that the random variables Θ_{1i} and Θ_{2i} are independent with Θ_{1i} distributed as $N(\mu, \sigma^2)$ and Θ_{2i} distributed as $Exp(\beta)$ (where $\beta = E(\Theta_{2i})$) for all i =1,..., n. They obtain the MLEs of μ, σ and β based on the observed interval-valued data. Then the "mean interval" $[\hat{a},\hat{b}]$ is computed by solving $\hat{\mu} = \frac{\hat{a}+\hat{b}}{2}$ and $\hat{\beta} = \frac{(\hat{b}-\hat{a})^2}{12}$. This yields $\hat{a} = \hat{\mu} - \sqrt{3\hat{\beta}}$ and $\hat{b} = \hat{\mu} + \sqrt{3\hat{\beta}}$. The standard errors of $\hat{a}$ and $\hat{b}$ can be easily computed and they are $\text{se}(\hat{a}) = \text{se}(\hat{b}) = \sqrt{\{\text{se}(\hat{\mu})\}^2 + 3\left\{\text{se}\left(\sqrt{\hat{\beta}}\right)\right\}^2}$. Note that $\text{Cov}(\hat{\mu},\sqrt{\hat{\beta}}) = 0$ since the Θ_{1i} and Θ_{2i} are assumed to be

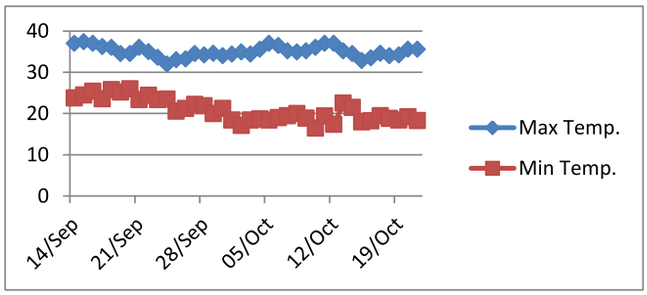
* Professor, Indian Institute of Management Ahmedabad, Gujarat, India. Email: arnab@iimahd.ernet.in

independent for all i=1,...,n. It is not difficult to see that the results can be easily extended to those cases where the internal parameters may have distributions other than those postulated in Le-Rademacher and Billard (2011).

A natural question to ask would be that which of these three "mean intervals" gives us the best representation for a given interval-valued dataset. For that purpose we define for any proposed interval [x,y] a measure of inaccuracy as $L([x,y]) = \frac{1}{n}\sum_{i=1}^{n}(|x-a_i|+|y-b_i|)$. Note that L([x,y]) = 0 is the best case which is attained only when $a_i = x$ and $b_i = y$ for all i = 1,..., n. Further observe that, an interval which yields a lower value of L gives a better representation of the data than another which yields a higher value of L. We compute the value of this inaccuracy measure L for the above three intervals and the one which gives the least value of L is chosen as the mean interval for representing the given interval dataset.

We illustrate the above methodology with two real life datasets. The first dataset correspond to the maximum and minimum temperatures at Ahmedabad during the period 14th September 2015 to 21st October 2015. The data is described graphically in Figure 1 below.

Fig 1: Maximum and Minimum Temperatures at Ahmedabad for the Period 14th September, 2015 to 21st October, 2015



Simple computations give $[\tilde{a},\tilde{b}]$=[20.8, 35.0] and $[a',b']$= [20, 34.9]. To compute $[\hat{a},\hat{b}]$ first the internal parameters (Θ_{1i}, Θ_{2i}) are computed for i=1,...n. The distribution of the internal parameters did not satisfy the assumptions made in the paper by Le-Rademacher and Billard (2011). However that does not cause any hindrance to the computation of the mean interval. We use the sample mean as an estimate of the population mean and compute $\hat{a} = \frac{1}{n}\sum_{i=1}^{n}\Theta_{1i} - \sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}\Theta_{2i}\right)}$ and $\hat{b} = \frac{1}{n}\sum_{i=1}^{n}\Theta_{1i} + \sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}\Theta_{2i}\right)}$. Simple computations now yield $[\hat{a},\hat{b}]$= [20.7, 35.1]. Thus we find that the three methods give very similar estimates of the mean interval. Now to choose one amongst the three we use the measure of inaccuracy L. Again, simple computations yield L([20.8, 35.0])= 3.426, L([20, 34.9])= 3.368 and L([20.7, 35.1])= 3.426. Hence, the interval [20, 34.9] gives the best representation of the variation of temperature during the given period.

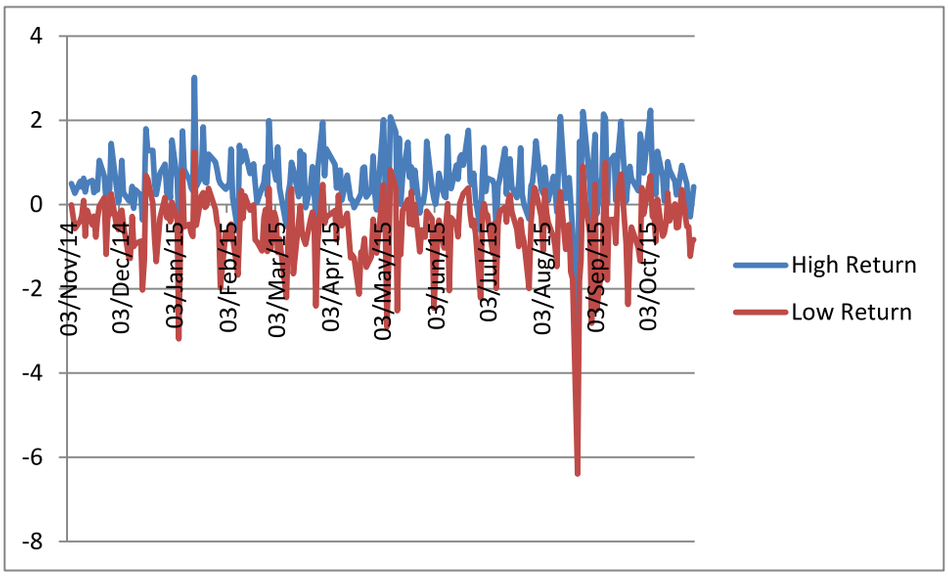
As a second example we consider CNX NIFTY data for the period 3rd November, 2014 to 30th October, 2015. Let H(t), L(t) and C(t) denote the High, Low and Closing values of NIFTY for day t. These are used to compute the variables High and Low Returns of day t (which are denoted as HR(t) and LR(t) respectively) as $HR(t) = \frac{H(t)-C(t-1)}{C(t-1)} \times 100$ and $LR(t) = \frac{L(t)-C(t-1)}{C(t-1)} \times 100$. The data is described graphically in Figure 2 below:

As in the earlier case we compute the mean interval by all the three methods and also the measure of inaccuracy (L). The results are given in Table 1 below:

Table 1: The Mean Intervals Obtained by the Three Methods and Their Measure of Inaccuracy

	$[\tilde{a},\tilde{b}]$	$[a',b']$	$[\hat{a},\hat{b}]$
Obtained Interval	[-0.59, 0.62]	[-0.47, 0.51]	[-0.65, 0.68]
Measure of Inaccuracy (L)	1.058	1.036	1.082

Fig 2: High and Low Returns of CNX NIFTY during the period 3rd November, 2014 to 30th October, 2015



We observe that there is substantial difference among the three intervals possibly due to the presence of outliers in the dataset. Since the interval [-0.47, 0.51] yields the least value of the inaccuracy measure (L) we use this interval for representing the variation of returns on CNX NIFTY within a day. An investor investing in CNX NIFTY would be advised to expect a variation in return between -0.47% to +0.51% during the course of a normal day.

References

Billard, L. (2011). Brief overview of symbolic data and analytic issues. *Statistical Analysis and Data Mining,* 4(2), 149-156.

Le-Rademacher, J., & Billard, L. (2011). Likelihood functions and some maximum likelihood estimators for symbolic data. *Journal of Statistical Planning and Inference,* 141, 1593-1602.

Article can be accessed online at http://www.publishingindia.com

A Multiple Attribute Group Decision-making Model for Selecting the Best Supplier

Mohammad Azadfallah*

Abstract

In the current literature, supplier selection is an important Multi Attribute Group Decision Making (MAGDM) problem which heavily contributes to the overall supply chain performance. Several solutions for the above problem are proposed. In this paper, TOPSIS and Borda's function approach, which is one of these methods, is discussed. So that, in the present model, first TOPSIS is used to find the individual preference ordering. Then, Borda's function is used to find the collective preference orderings. Finally, a simple example is provided in order to demonstrate its applicability and effectiveness of the proposed method.

Keywords: MAGDM, Topsis, Borda's Function, Supplier Selection Problem

Introduction

Multi Attribute Decision Making (MADM) is an important component of modern decision science. The theory and methods of MADM have been extensively applied to the fields of engineering project, economy, management and military affairs, such as investment decision making, venture capital project evaluation, facility location, bidding, maintenance services, military system efficiency evaluation, development ranking of industrial sectors, comprehensive evaluation of economic performance, etc. Essentially, MADM is to select the most desirable alternative(s) from a given finite set of alternatives according to a collection of attributes by using a proper means. It mainly consists of two stages: 1. collect decision information. The decision information generally includes the attribute weights and the attribute values, especially, how to determine the attribute weights is an important research topic in MADM. 2. Aggregate the decision information through some proper approaches (Xu, 2015).

The increasing complexity of the socio-economic environment makes it less and less possible for a single expert or decision maker to consider all relevant aspects of a problem. Therefore, complex decision problems usually have to be conducted by integrating a group of expert's knowledge and experiences. Generally, the practice of Multiple Attribute Group Decision Making (MAGDM) is to invite internal experts, external experts, or their combination of related fields to evaluate each attribute of every alternative individually. In the MAGDM, the experts usually have diverging opinions, because they often come from different specialty fields and each expert has his/ her unique characteristics with regard to knowledge, skills, experience and personality. Thus, how to obtain the maximum degree of consensus or agreement from these experts for the given alternatives is an interesting and important research topic (Li & Sun, 2012). On the other side of this coin, supplier selection is a fundamental issue of supply chain area that heavily contributes to the overall supply chain performance, and also, it is a hard problem since supplier selection is typically a Multi Criteria Group Decision problem (Izadikhah, 2012). In this paper to solve this problem, TOPSIS and Borda's function approach is proposed.

TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) proposed by Yoon and Hwang is one

* Researcher, Business Studies and Development Office, Saipa yadak (Saipa after sales services), Islamic Republic of Iran.
Email: m.azadfallah@yahoo.com

of the widely used techniques in Multi Attribute Decision Making. TOPSIS can rank a finite number of feasible alternatives in order of preference according to the features of each attribute of every alternative and select a suitable alternative that conforms to the decision maker's ideal. The basic concept of TOPSIS technique is that the selected alternative will have the shortest Euclidean distance from the ideal solution and the farthest Euclidean distance from the anti-ideal solution (Lin, Lee, Chang & Ting, 2008). In addition, the Borda rule is an appropriate procedure in multi-person decision making when several alternatives are considered (Lapresta *et al*., 2008). It is based on a majority rule relation (Hwang & Yoon, 1981). In other words, the Borda technique assigns ranks to options based on the rational that the higher position of an option on the voters list, the higher the rank assigned. The voting position of an option is determined by adding the ranks for each option from every voter using the Borda vote aggregation function. The winner is an option that receives the highest score calculated such that all options are assigned a score starting with 0 for the last favourable solution, 1 for the second worst, 2 for the third worst, and so on. All scores are weighted by the number of voters, resulting in the Borda score for each option (Mysiak, 2010). Nevertheless, the TOPSIS and Borda's function approaches are the ordinal and cardinal approach. The ordinal (ranking) and cardinal (rating) approach allows committee members to individually evaluate each candidate and to find the collective preference ordering so that, TOPSIS can be used to find the individual preference ordering. Then Borda's function can be used to find the collective preference orderings (Hwang & Lin, 1987).

This paper aims to use a numerical example to illustrate the process of the proposed MAGDM method (TOPSIS and Borda's function approach) in supplier selection context.

The paper is organised as follow. In the second section, the literature and in the third section, the proposed approach is discussed. Numerical example is provided in the next section. The paper is concluded in the fifth and the last section.

Literature Review

In recent years, TOPSIS has been successfully adopted in various fields, e.g. location analysis, construction processes, human resource management, transportation, product design, manufacturing, water management, and quality control, and has demonstrated satisfactory results (Lin *et al*., 2008). TOPSIS has also been employed in Multi Attribute Group Decision Making environment. For instance, Chen (2000), Wanga and Leeb (2007), and Krohling and Campanharo (2011) extended the TOPSIS method for fuzzy environment. Saghafian and Hejazi (2005) and Lin *et al*., (2008), proposed a modified fuzzy TOPSIS method for group decision problem. Boran, Genç, Kurt and Akay (2009) combined TOPSIS method with intuitionistic fuzzy set to select appropriate supplier in-group decision-making environment. Ghaemi Nasab and Malkhalifeh (2010) extended TOPSIS for group decision making based on the type-2 fuzzy positive and negative ideal solutions. Yue (2011) developed TOPSIS method for determining weights of decision makers with interval numbers. Huang and Li (2012) proposed a new aggregation rule for the use of TOPSIS in-group decision context. Izadikhah (2012) used an extended TOPSIS method for group decision making with Atanassovs interval-valued intuitionistic fuzzy numbers to solve the supplier selection problem under incomplete and uncertain information environment. And Zhang (2014) presented an MAGDM model to the assessing project risks. In addition, in the literature, Borda approach and TOPSIS and Borda's function have been applied in various fields. For instance, Lapresta *et al*. (2008) used linguistic labels as inputs in the Borda count. Shahbandarzadeh and Haghighat (2010) used LINMAP for target market selection so that LINMAP technique is used separately for each level (strengths, weaknesses, opportunities and threats), and at last unified the results of each level by using of Borda method. Anisseh and Yusuff (2011) suggested a new method under the linguistic framework for heterogeneous group decision making. To accomplish this, an integrated fuzzy group decision-making method based on Borda count is proposed. Kim and Ye (n.d.) proposed a general fuzzy grey decision making method, which takes into consideration of the grey degree of the weight and the attribute value at the same time, for the MADM where attributes have the generalised super mixed-type values given by real number, interval value, linguistic value, and uncertain linguistic value. Then, four ranks are obtained by four methods of plan evaluation such as the evaluation by the relative approach degree of grey TOPSIS. Finally, the weighted Borda method is used to obtain final rank by using the results of four methods. Fu (2008) developed three extended TOPSIS models,

the pre-model, post-model, and inter-model, associated with three approaches to aggregating group preferences (the pre-operation, post-operation, and inter-operation), which depend on Dempster's rule or its modifications, some social choice functions (i.e. Borda's function and some mean approaches). Gharakhlou *et al.* (2010) used TOPSIS, fuzzy and SAW techniques for making changes in the route network to facilitate rescue operations in urban disasters. Next, the results were combined by means of Borda method. Kim *et al*. (n.d.), proposed an interval weight determining method and three methods of grey interval relation decision making: the evaluation of plans by the relative approach degree of grey TOPSIS method, the evaluation by the relative approach degree of grey incidence and the evaluation by the relative approach degree of grey incidence using maximum entropy estimation. The final rank of plans has been obtained by weighted Borda method considering the above three ranking results.

It should be emphasized at this point that, according to Hojjati and Anvari (2014), decision makers are not limited just to one method of MADM in critical issues, because it is possible that various MADM methods attain different results. In order to get over this problem, aggregate method (Borda method) has been introduced. In addition, in the literature, many studies exist on using Borda's function. However, most of them are limited to combine the results of different MADM methods (i.e., Hojjati & Anvari, 2014; Hashemi & Zamani, 2015 etc.). However, but in this paper we used Borda's function for to unify the results of each DM (expert) where same method is used (TOPSIS).

This paper focused on the application of MAGDM models, particularly TOPSIS and Borda's function approach for solving the supplier selection problem. In the next section, the proposed method will be considered.

Proposed Approach

Topsis Method

TOPSIS assumes that we have m alternatives (options) and n attributes/criteria and we have the score of each option with respect to each criterion. Let x_{ij} score of option i with respect to criterion j. we have a matrix $X = (x_{ij})_{m.n}$ matrix. Let J be the set of benefit attributes or criteria (more is better). Let J' be the set of negative attributes or criteria (less is better). The idea of Topsis can be expressed in a series of steps (Tayeb, Ahcene, Omar & Mouloud, 2007):

Step 1: Obtain performance data for n alternatives over k criteria. Raw measurements are usually standardised; converting raw measures x_{ij} into standardised measures s_{ij}. Construct normalised decision matrix. This step transforms various attribute dimensions into non-dimensional attributes, which allows comparisons across criteria. Normalize scores or data as follows:

$$r_{ij} = X_{ij} / \sqrt{\sum X_{ij}^2} \text{ for } i = 1, \ldots, m; j = 1, \ldots, n.$$

Step 2: Develop a set of importance weights wk, for each of the criteria. The basis for these weights can be anything, but usually, is ad hoc reflective of relative importance. Scale is not an issue if standardizing was accomplished in step 1. Construct the weighted normalised decision matrix. Assume we have a set of weights for each criteria wj for j = 1, …, n. multiplies each column of the normalised decision matrix by its associated weight. An element of the new matrix is:

$$V_{ij} = w_j r_{ij}$$

Step 3: Determine the ideal and negative ideal solutions.

Ideal solutions:

$$A^* = \{v_1^*, \ldots, v_n^*\}, \text{ where}$$

$$V_j^* = \{\max_i (v_{ij}) \text{ if } j \varepsilon J; \min_i (v_{ij}) \text{ if } j \varepsilon J'\}$$

Negative ideal solutions:

$$A' = \{v_1', \ldots, v_n'\}, \text{ where}$$

$$V_j^* = \{\min_i (v_{ij}) \text{ if } j \varepsilon J; \max_i (v_{ij}) \text{ if } j \varepsilon J'\}$$

Step 4: Calculate the separation measures for each alternative. The separation from the ideal alternative is:

$$S_i^* = [\sum_j (v_j^* - v_{ij})^2]^{\wedge 1/2} \quad i = 1, \ldots, m.$$

Similarly, the separation from the negative ideal alternative is:

$$S_i' = [\sum_j (v_j' - v_{ij})^2]^{\wedge 1/2} \quad i = 1, \ldots, m.$$

Step 5: calculate the relative closeness to the ideal solution C_i^*:

$C_i^* = S_i' / S_i^* + S_i'$ $0<C_i^*<1$

Step 6: Rank order alternatives by maximizing the ratio in step 5. Select the option with C_i^* closest to 1.

Borda's Function

The method Borda function is the rank-order method. With m candidates in A, assign marks of m-1, m-2, …, 1, 0 to the first ranked, second ranked, … last ranked candidate for each individual, then determine the Borda score for each candidate as the sum of the individual marks for that candidate. Then the candidate with the highest Borda score is declared as the winner. The Borda score of a candidate x is equivalent to the sum of the number of individuals that have x preferred to y for all yϵA /{X}.

Borda's function:

Let

$F_B (X) = \sum_{y \epsilon A} \# (i: X P_i Y)$

And the candidates are ranked in the order of the value of F_B (Hwang & Lin, 1987).

TOPSIS and Borda's function Approach

In this approach, TOPSIS can be used to find the individual preference ordering. Then Borda's function can be used to find the collective (social) preference orderings. With m candidates in A, scores of m-1, m-2, …, 1, 0 can be assigned to the first ranked, second ranked, …, last ranked candidate by each committee member. The Borda score (the sum of the committee members scores) can be determined for each candidate. Finally, the candidates are ranked according to their Borda scores (Hwang & Lin, 1987).

Numerical Example

In this section, a numerical example is used to illustrate the application of the proposed method. Assume that there are four committee members, who are experts (Experts; E1, E2, E3, and E4), seven alternatives (Suppliers; S_1, S_2, …, and S_7), and five criteria (C_1=quality, C_2=on-time delivery, C_3=service, C_4=responsiveness, and C_5=price and cost). As you see, the performance values for each expert are shown in table 1-4. In addition, several researchers have argued that the equal weight rule is often a highly accurate simplification of the decision making process (Birnbaum, 1998). Thus, W_j= (0.2, 0.2, 0.2, 0.2, and 0.2).

In the process of group decision making, each committee member (expert) provides the performance values, the alternatives with respect to each attribute (having common criteria for committee members). For expert 1:

Table 1: Performance Value for E1*

Criteria / Alternative	C_1	C_2	C_3	C_4	C_5^-
S_1	7	7	1	1	103
S_2	9	7	5	5	130
S_3	5	9	1	5	121
S_4	3	1	3	3	109
S_5	9	5	9	3	115
S_6	1	3	7	7	117
S_7	7	1	5	7	125

*note, that all attributes (criteria) except C5 are the benefit criteria (in the all tables).

For expert 2:

Table 2: Performance Value for E2*

Criteria / Alternative	C_1	C_2	C_3	C_4	C_5^-
S_1	1	7	7	7	101
S_2	1	1	5	5	132
S_3	9	5	9	3	123
S_4	5	3	5	5	108
S_5	3	3	5	9	115
S_6	3	9	3	1	119
S_7	3	9	5	5	121

For expert 3:

Table 3: Performance Value for E3*

Criteria / Alternative	C_1	C_2	C_3	C_4	C_5^-
S_1	9	1	7	9	105
S_2	7	7	1	5	130
S_3	1	7	5	1	122
S_4	9	3	9	1	111
S_5	3	9	1	3	115

Criteria / Alternative	C_1	C_2	C_3	C_4	C_5^-
S_6	1	5	9	1	116
S_7	3	1	3	7	123

For expert 4:

Table 4: Performance value for E4*

Criteria / Alternative	C_1	C_2	C_3	C_4	C_5^-
S_1	1	9	1	3	106
S_2	3	5	7	9	131
S_3	3	5	9	3	123
S_4	3	7	5	1	107
S_5	5	3	1	5	118
S_6	3	9	1	1	117
S_7	1	3	5	9	119

In this section, we have used TOPSIS to find the individual preference ordering. i.e., for E1:

1. Calculate the normalised decision matrix

For instance, for R_{11}:

$R_{11}= 7/ \sqrt{7^2+9^2+5^2+3^2+9^2+1^2+7^2}=0.408$

$$R_{ij} = \begin{vmatrix} .408 & .477 & .072 & .077 & .331 \\ .524 & .477 & .362 & .387 & .418 \\ .291 & .614 & .072 & .387 & .389 \\ .175 & .068 & .217 & .232 & .351 \\ .524 & .341 & .651 & .232 & .370 \\ .058 & .205 & .507 & .542 & .376 \\ .408 & .068 & .362 & .542 & .402 \end{vmatrix}$$

2. Calculated the weighted decision matrix (as noted earlier, W_j= 0.2, 0.2, 0.2, 0.2, and 0.2). For instance, for V_{11}:

$V_{11}=0.408*0.2 = 0.082$

$$V_{ij} = \begin{vmatrix} .082 & .095 & .014 & .015 & .066 \\ .105 & .095 & .072 & .077 & .084 \\ .058 & .123 & .014 & .077 & .078 \\ .035 & .014 & .043 & .046 & .070 \\ .105 & .068 & .130 & .046 & .074 \\ .012 & .041 & .101 & .108 & .075 \\ .082 & .014 & .072 & .108 & .080 \end{vmatrix}$$

3. Determine the ideal and negative ideal solutions

A* = {v1*… vn*}; where, Vj* = {max (vij) if j ε J; min (vij) if j ε J'};

For instance, for v1*:

= max (0.082, 0.105, 0.058, 0.035, 0.105, 0.012, 0.082) =0.105

A* = (.105 .123 .130 .108 .066)

A^- = {$v1^-$ … vn^-}; where, Vj^- = {min (vij) if j ε J; max (vij) if j ε J^-}

For instance, for $v1^-$:

= min (0.082, 0.105, 0.058, 0.035, 0.105, 0.012, 0.082) =0.012

A^- = (.012 .014 .014 .015 .084)

4. Calculated the separation measures

$Si^*= [\Sigma_{j=1}^{7} (vj^* – vij) 2]^{\wedge 1/2}$, i = 1, 2, 3, 4, 5.

For instance, for S_1:

$S_1= \sqrt{(0.082 – 0.105)^2+ (0.095-0.123)^2 + … + (0.066-0.066)^2} = 0.153$

S_1 =.153, S_2=.073, S_3=.129, S_4=.168, S_5=.083, S_6=.128, and S_7=.126

$Si^-= [\Sigma_{j=1}^{7} (vj^- - vij) 2]^{\wedge 1/2}$, i = 1… m.

For instance, for S_1:

$S_1= \sqrt{(0.082-0.012)^2 + (0.095-0.014)^2+ … + (0.066-0.084)^2} = 0.109$

S_1 =.109, S_2 =.150, S_3=.134, S_4=.050, S_5=.162, S_6=.130, and S_7=.130

5. Calculated the relative closeness to the ideal solution

$C_1^* = S1^-/ S1^* + S1^- = 0.109 / (0.109+0.153) =0.417$

$C_2^* = 0.672$
$C_3^* = 0.509$
$C_4^* = 0.230$
$C_5^* = 0.661$
$C_6^* = 0.505$
$C_7^* = 0.507$

6. Rank the preference order. According to the descending Order of Ci*, the preference order is (for E_1):

S2 > S5 > S3 ≈ S7 > S6 > S1 > S4

Similarly, for E_2:

$C_1^* = 0.461$, $C_2^* = 0.244$, $C_3^* = 0.634$, $C_4^* = 0.431$, $C_5^* = 0.466$, $C_6^* = 0.386$, $C_7^* = 0.488$

For E_3:

$C_1^* = 0.617$, $C_2^* = 0.507$, $C_3^* = 0.358$, $C_4^* = 0.502$, $C_5^* = 0.424$, $C_6^* = 0.403$, $C_7^* = 0.387$

For E_4:

$C_1^* = 0.309$, $C_2^* = 0.664$, $C_3^* = 0.551$, $C_4^* = 0.402$, $C_5^* = 0.436$, $C_6^* = 0.341$, $C_7^* = 0.479$

A comparison of the test results is given in Table 5.

Table 5: Comparison Results (for Four Experts)

Rank / *Expert*	*Results*
E_1	S2 > S5 > S3 ≈ S7 > S6 > S1 > S4
E_2	S3 > S7 > S5 > S1 > S4 > S6 > S2
E_3	S1 > S2 > S4 > S5 > S6 > S7 > S3
E_4	S2 > S3 > S7 > S5 > S4 > S6 > S1

Now, to aggregate the preference ordering into a consensus ordering, the Borda method is used. So, can more assurance to the results by applying a mathematical model. Therefore, 6, 5, …, 1, 0 scores are assigned to the first rank, second rank, …, and last rank. i.e., for S_1:

S1=1+3+6+0=10

Similarly, S2=17

$S_3 = 14.5$
$S_4 = 8$
$S_5 = 15$
$S_6 = 6$
$S_7 = 13.5$

From the above results, it can be easily derived that, the implied ranking is as follow:

S2 > S5 > S3 > S7 > S1 > S4 > S6

Therefore, the best alternative is S_2, since it is superior to all the other alternatives. Meanwhile, S_6 have very bad performance.

Concluding Remarks

In the current literature, supplier selection is typically a Multi Attribute Group Decision problem (MAGDM). Since, how to obtain the maximum degree of consensus or agreement from these experts for the given alternatives is an interesting and important topic, in this paper to solve this problem, TOPSIS and Borda's function approach is proposed, and can provide more assurance to the results by applying a mathematical model. Moreover, the result implicitly shows the MAGDM (i.e., TOPSIS and Borda's function approach) is an effective approach in dealing with this kind of decision problem. Hence, to increase your chance of finding an appropriate supplier for your companies, we suggest using the proposed models in this paper. In addition, further research can apply this proposed approach to other managerial issue or compares with another MAGDM method.

Reference

Anisseh, M., & Yusuff, R. B. M. (2011), A fuzzy group decision making model for multiple criteria based on Borda count. *International Journal of the physical sciences,* 6(3), 425-433.

Birnbaum, M. H. (1998). *Measurement - judgment and decision-making.* Academic Press.

Boran, E., Genç, S., Kurt, M., & Akay, D. (2009). A multi criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. *Expert Systems with Applications,* 36(2009), 11363-11368.

Chen, C. T. (2000). Extensions of the TOPSIS for group decision making under fuzzy environment. *Fuzzy Sets and Systems,* 114(2000), 1-9.

Fu, C. (2008). Extended TOPSIS for belief group decision making, *Journal of Service Science & Management,* 1, 11-20.

Gharakhlou, M., Faraji Sabokbar, H. A., & Givehchi, C. (2010). access enhancement by making changes in the route network to facilitate rescue operations in urban disasters. *International Journal Environ. Res.,* 4(2), 183-192.

Ghaemi, N. F., & Malkhalifeh, M. R. (2010). Extension of TOPSIS for group decision making based on the type-2 fuzzy positive and negative ideal solutions. *International Journal industrial mathematics,* 2(3), 199-213.

Hashemi, Z. M., & Zamani F. M. (2015). Proposing a multi attribute decision making model on human force productivity, *GMP Review,* 18(1), 187-199.

Hashemi, S. M. H., & Anvari, A. (2014). An integrated MCDM method in ranking the major lean practices based on four attributes. *MAGNT Research Report* (ISSN: 1444-8939), 2(7), 285-298.

Huang, Y. Sh., & Li, W. H. (2012). A study on aggregation of TOPSIS ideal solutions for group decision making. *Group Decis Negot*, 21, 461-473.

Hwang, C. L., & Yoon, K. (1981). Multiple attribute decision making: Methods and applications. Springer-Verlag.

Hwang, C. L., & Lin, M. J. (1987). *Group decision-making under multiple criteria: Methods and applications.* Springer-Verlag.

Izadikhah, M. (2012). Group decision-making process for supplier selection with TOPSIS method under internal valued intuitionistic fuzzy number, Advances in fuzzy systems, doi: 10.1155/2012/407942, 1-14.

Kim, G., & Ye, F. (n.d.). super-mixed multiple attribute group decision-making method based on hybrid fuzzy grey relation approach degree. Retrieved from http://arxiv.org/ftp/arxiv/papers/1207/1207.1501.pdf, 1-10.

Kim, et al. (n.d.). hybrid grey interval relation decision making in artistic talent evaluation of player. Retrieved from http://arxiv.org/ftp/arxiv/papers/1207/1207.3855.pdf), 1-8.

Krohling, R. A., & Campanharo, V. C. (2011). Fuzzy TOPSIS for group decision making: a case study for accidents with oil spill in the sea. *Expert Systems with Applications*, 38(2011), 4190-4197.

Lapresta, *et al*., (2008), defining the Borda count in a linguistic decision making context, preprint submitted to Elsevier Science, 15 October 2008, (available in: http://www.researchgate.net/publication/228359030_Defining_the_Borda_count_in_a_linguistic_decision_making_context).

Lin, Y. H., Lee, P. C., Chang, T-P., & Ting, H-l. (2008). Multi attribute group decision making model under the condition of uncertain information. *Automation in Construction,* 17, 792-797.

Li, F., & Sun M. (2012). Approach to reaching consensus in triangular fuzzy multiple attributes group decision-making. *Journal of Information and Computational Science,* 9(12).

Mysiak, J. (2010). MDSS, decision methods, the guide for the mDSS version 5. Retrieved from http://www.hetsymodoeu/mdss/, 1-36.

Shahbandarzadeh, H., & Haghighat, F. (2010). Evaluation of the strategies of target market selection on the basis of IFE and EFE matrixes using LINMAP technique. *Iranian Journal of Management Studies (IJMS),* 3(3), 41-58.

Saghafian, S., & Hejazi, S. R. (2005). Multi criteria group decision making using a modified fuzzy TOPSIS procedure, Proceeding of the 2005 international conference on computational intelligence for modeling, control and automation, and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'05).

Tayeb, S., Ahcene, B., Omar, P. J. S., & Mouloud, B. K. (2007). Equipment selection by numerical resolution of the Hessian matrix and TOPSI algorithm. *Asian Journal of Information Technology,* 6(1), 81-88.

Wanga, Y. J., & Leeb, H. S. (2007). Generalizing TOPSIS for fuzzy multiple criteria group decision-making. *Computers and Mathematics with applications,* 53, 1762-1772.

Xu, Z. (2015). *Uncertain multi attributes decision-making; methods and applications.* Springer-Verlag, Berlin Heidelberg.

Yue, Z. (2011). An extended TOPSIS for determining weights of decision makers with interval numbers. *Knowledge based Systems,* 24(2011), 146-153.

Zhang, Z. (2014). An approach to multi attribute group decision-making and its application to project risk assessment. *Journal of software,* 9(2).

Article can be accessed online at http://www.publishingindia.com

Ensemble Approach for Zoonotic Disease Forecasting using Machine Learning Techniques

Vikash Chandra Sharma*, David Frankenfield, Anupam Gupta***, Rama Krishna Singh ****

Abstract

More than two-third of emerging infectious diseases in recent decades are zoonotic in origin. Timely prediction of these diseases which migrate from animals to humans and preventive measures to stop the loss in terms of morbidity and mortality is the requirement of healthcare industry. Avian Influenza is one of the zoonotic diseases that have created havoc in recent past especially in Asian subcontinent. In past, attempts have been made to predict influenza using traditional time-series techniques (AR, MA, ARMA, ARIMA etc.) as well as machine learning techniques to capture the cyclicity and seasonality of these virus strains. In current research an effort has been made to utilize the Empirical Mode Decomposition (EMD) to extract the Intrinsic Mode function (IMF) and then apply state of art Machine Learning (ML) techniques to predict the series. Several machine learning techniques like Random Forest (RF) along with Gradient Boosting Machine (GBM) and Support Vector Regression (SVR)have been applied on the decomposed series. Exogenous models showed variables like temperature, humidity and precipitation have been incorporated to improve upon the forecast. An ensemble approach of ML models showed significant improvement over the traditional models in terms of long term forecast accuracy.

Keywords: Random Forest, Gradient Boosting Machine, Support Vector Regression, Machine Learning, Avian Influenza

Introduction

In healthcare industry, application of time-series modeling and prediction of future outbreak of certain infectious diseases and disease events which occur in a cyclic or rhythmic pattern are very crucial. The forecasting of disease helps to predict the course of disease, warn healthcare experts and adopt control measures to prevent disease outbreaks. The US Agency for International Development launched its Emerging Pandemic Threats Programme in late 2009 to build an early warning system to detect and reduce the impacts of zoonotic diseases.

Zoonotic diseases are a group of infectious diseases naturally transmitted from animals to humans. Avian influenza (AI) under consideration is one of those zoonotic diseases which posea major threat to mankind in recent years. It refers to the disease caused by infection with avian (bird) influenza (flu) Type A viruses. Humans are affected by AI virus subtypes H5N1 and H9N2 and swine influenza virus subtypes H1N1 and H3N2. The AI (H5N1) virus subtype, a highly pathogenic AI virus, first infected humans in 1997 during a poultry outbreak in Hong Kong and China. Since its widespread re-emergence in 2003 and 2004, this avian virus has spread from Asia to Europe and Africa and has become entrenched in poultry in some countries, resulting in millions of poultry infections and many human deaths. The mortality and morbidity associated with this disease have devastated communities in some countries and led to global changes in public health. Countries not only suffered huge economic loss but in some instances closed down - global travel and trade networks. These vulnerabilities emphasize the need for a systematic, pre-emptive, advanced and improved predictive modeling approach to predict the emergence of such pandemics that could impact the health risk to susceptible human population.

AI also has some cyclic or repeating pattern to capture which, traditional time-series predictions are performed

* Senior Consultant at UnitedHealth Group, Noida, Uttar Pradesh, India. Email:vikash_c_sharma@uhc.com
** Director at United HealthCare, Greater Denver Area, United States. Email: david_l_frankenfield@uhc.com
*** Director at UnitedHealth Group, New Delhi, India. Email: anupam_gupta@uhc.com
**** Associate Director at UnitedHealth Group, Noida, Uttar Pradesh, India.Email: rama_k_singh@uhc.com

using the autoregressive integrated moving average (ARIMA) (Box & Jenkins, 1970 technique. ARIMA model attempts to filter out high frequency noise in the data to detect local trends based on linear dependence in observations in the series. ARIMA models even though widely applied incorporate lot many assumptions. First, it assumes linear relationship between independent and dependent variables. Real-world relationships are often non-linear and therefore more complex than the assumptions build into ARIMA model. As a result this model does not perform well when data structure is complex. Also, these models assume a constant standard deviation of errors over time. This assumption can be removed when ARIMA is used in conjunction with a Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) (Engle, 1995) model. GARCH technique attempts to characterize model's non-constant standard deviations in a time-series but it comes with its own challenges and optimizing the parameters for GARCH is always a challenge.

Another challenge in time series prediction is the prediction horizon. When the prediction horizon increases, the uncertainty of future trends also increases, rendering a more tough prediction problem. Researchers have always wanted to extract the maximum knowledge from the past values to better utilize them for long-term time series prediction. More recently, new classes of regression models along with machine learning techniques have been developed to address the challenges associated with classical methods. Literature suggests the usage of Random Forest (Kane, Price, Scotch & Rabinowitz, 2014) technique for superior prediction accuracy when compared to the classical models. This paper makes use of an ensemble of Random Forest with Gradient Boosting Machine and Support Vector Regression models to come up with a better forecast for longer duration which can then be used for future planning and taking preventive steps to contain the disease from spreading and transforming into epidemic.

Data

Data for avian influenza virus were collected from the online web-based application (EMPRES-i) (http://empres-i.fao.org/eipws3g/)which has been designed to support veterinary services by facilitating the organization and access to regional and global disease information. This platform is a global animal disease information system including emergent zoonoses and other high impact animal diseases. In this research, data were considered for a period of Jan'06 to Jul'14 for some Asian countries (China, India, Nepal, Bangladesh, Vietnam, and South Korea) because these countries historically seemed to witness majority of outbreak cases of AI. Fig. 1 demonstrates the number of outbreak cases in Asian countries considered in the research work.

Fig. 2 shows the monthly time plot for avian influenza outbreak cases in nations under consideration.

Avian influenza was forecasted taking into consideration few exogenous meteorological variables like daily temperature (minimum and maximum temperature), relative humidity and precipitation. The data were obtained from http://globalweather.tamu.edu/ for the Asian countries in the study for a period of Jan'06 to Jul'14. The data from different sources were merged into a single data set and was then brought to monthly level taking the number of outbreak cases, average minimum temperature, average maximum temperature, average humidity, and average precipitation.

Methodology

Preliminary Analysis

The model was fitted considering data for a period of Jun'06 to Dec'12 and the forecast was validated on a period of Jan'13 to Jul'14. This period was selected considering the availability of data and to maintain the consistency of training and validation period across all modeling methods. Univariate time series analysis was performed on the data using few of the traditional forecasting techniques like Holt Winter's and ARIMA models to get a benchmark estimate. Improvement was observed with application of models like AR-GARCH model over traditional forecasting models. Actual vs fitted records are highlighted to show the performance of traditional models in Fig. 3. A detailed R syntax is shared in Appendix A for reference.

However, real world forecasting processes involve complex nonlinear series having large number of components. In the study for zoonotic diseases, occurrences are driven by lot many factors like weather, humidity, cleanliness, income group, etc. It is difficult to analyse such disease as

Fig. 1: Avian Influenza Outbreaks for a Period of Jan'06 to Jul'14

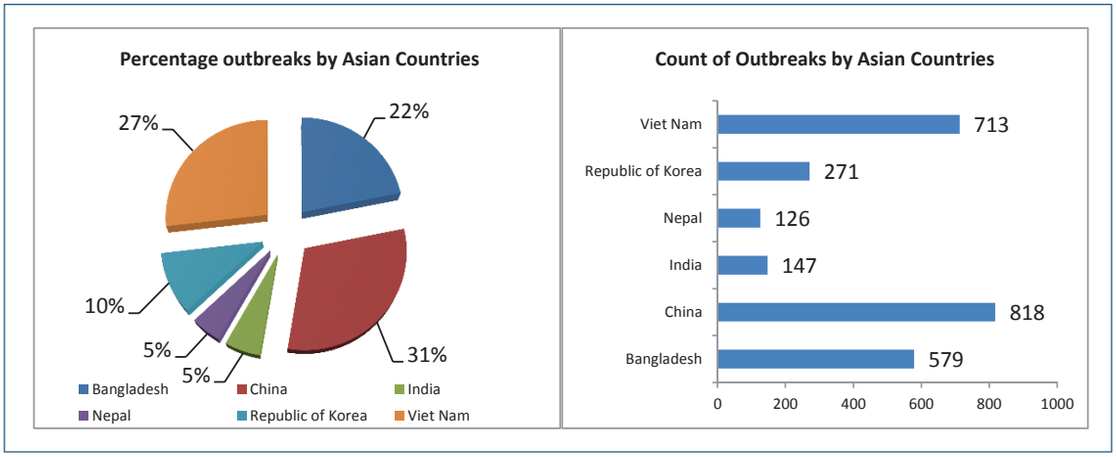


Fig. 2: Time Plot for Avian Influenza Outbreaks in Asia

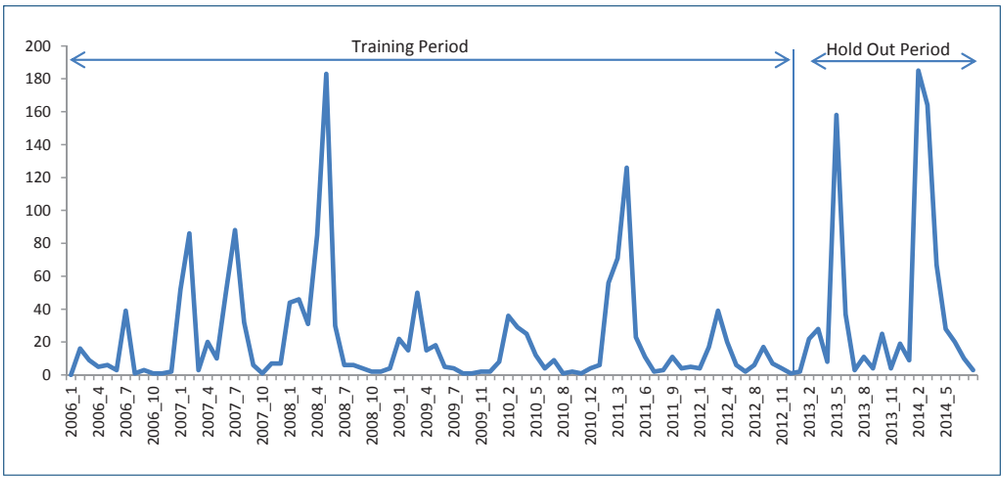
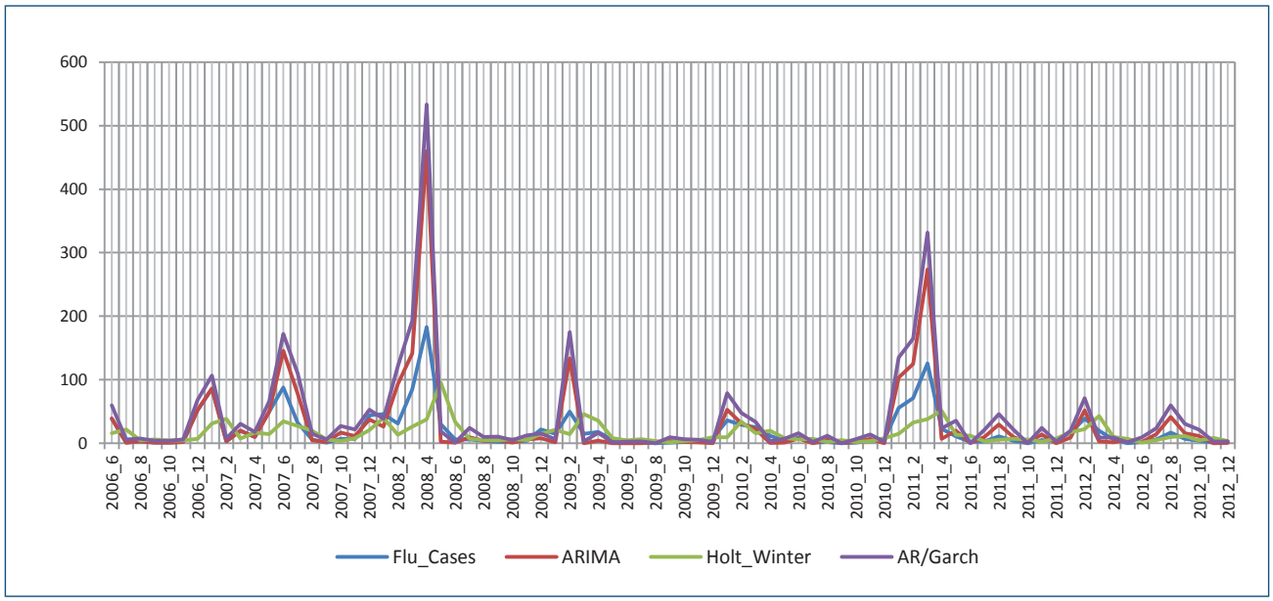


Fig. 3: Actual vs Fitted for Traditional Models



its components, when interacting with each other, mask and distort the regularities which need to be identified. This gives rise to the requirement to break down the process under consideration into individual components and analyse each and every component separately. Analysis of individual component and consideration of contribution they make into the process at hand helps us understand the process better as well as increases forecast reliability.

In the research work, decomposing the monthly influenza outbreaks using Empirical Mode Decomposition (EMD) (Wu & Hu, 2006) technique was done. Exogenous variables like temperature, humidity, and precipitation were used and a set of machine learning techniques like Gradient Boosting Machine, Support Vector Regression, and Random Forest were applied on EMD decomposed data to come up with the forecasted value. Models were compared on basis of mean absolute percentage error (MAPE) on the hold out validation period. For all modeling exercise, R-Studio programming environment was used and various packages were considered from CRAN (http://CRAN.R-project.org).

Empirical Mode Decomposition

Empirical Mode Decomposition is a decomposition technique which was proposed as the fundamental part of the Hilbert-Huang transform (HHT) (Huang, Shen, Long, Wu, Shin, Zheng, Yen, Tung & Liu, 1998). In contrast to other decomposition techniques, the EMD decomposes any given data into intrinsic mode functions (IMF) that are not set analytically and are determined by analysed sequence only. The basic functions are determined directly from the input data. An IMF resulting from the EMD shall satisfy the following requirements:

1. The number of IMF extrema (the sum of the maxima and minima) and the number of zero-crossings must either be equal or differ at most by one;
2. At any point of an IMF, the mean value of the envelopes defined by local maxima and local minima shall be zero.

Decomposition contains a family of frequency ordered IMF components. Each successive IMF contains lower frequency oscillations than the preceding one.

Fig. 4 depicts the analysed sequence of the thin blue line which is the actual series under consideration. The envelopes are shown in green. Mean is calculated based on the two envelopes and then subtracted from the initial sequence. To obtain the final IMF, new maxima and minima shall be identified and all the above steps repeated until stoppage criteria are met. This recursive processof subtracting the mean of envelopes from the initial sequence is called sifting. Fig. 5 shows the sifting process applied on the monthly Avian Flu cases recorded for Asian countries. The sifting process continues until the mean value of the minima and maxima envelopes becomes zero and that is the first IMF extract.

Monthly avian influenza outbreak data were first tested for various transformations and then square root transformation was considered to reduce on the variance

Fig. 4: Plotting the Envelopes and their Mean

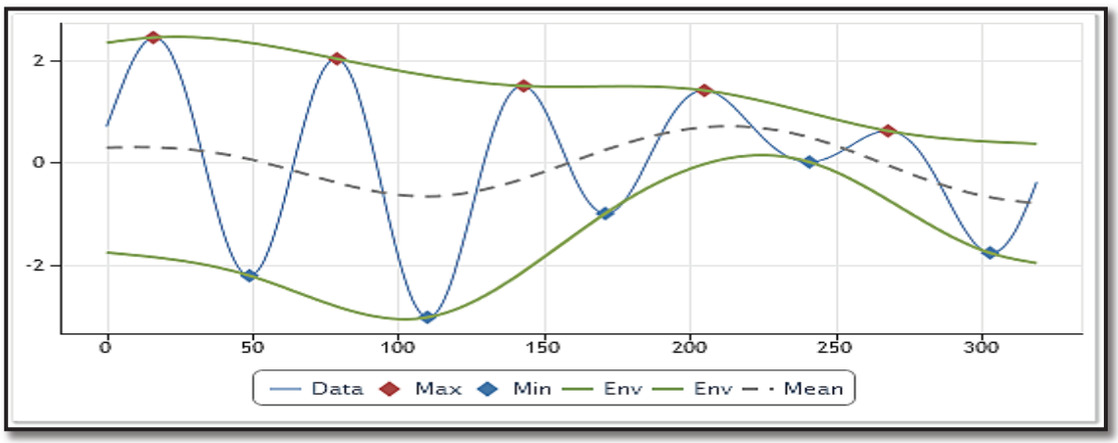


Fig. 5: Sifting Simulations on Monthly Avian Flu cases in Asia

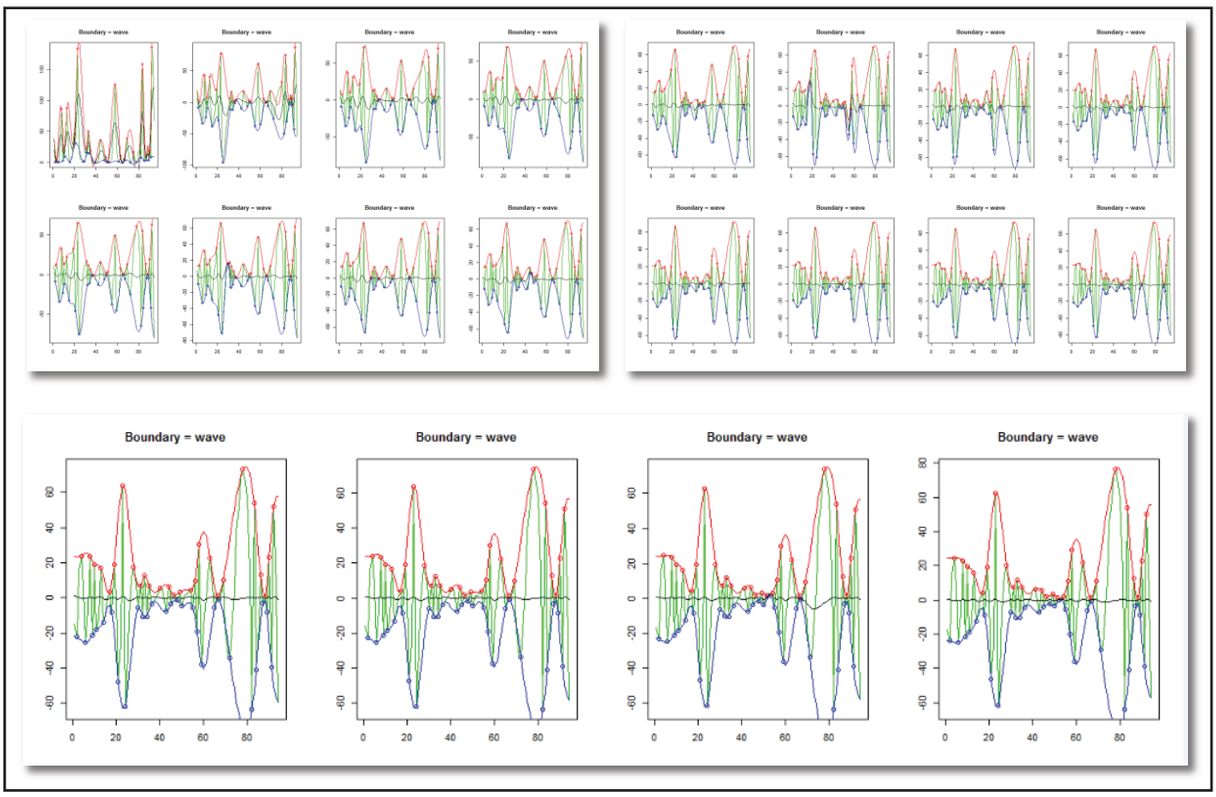
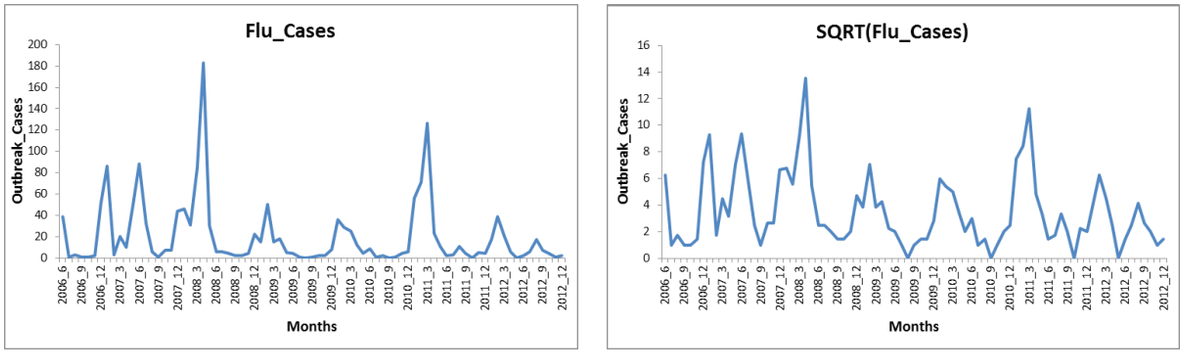


Fig. 6: Modified Series Using Square Root Transformation



along with handling months which had no outbreak cases as shown in Fig. 6.

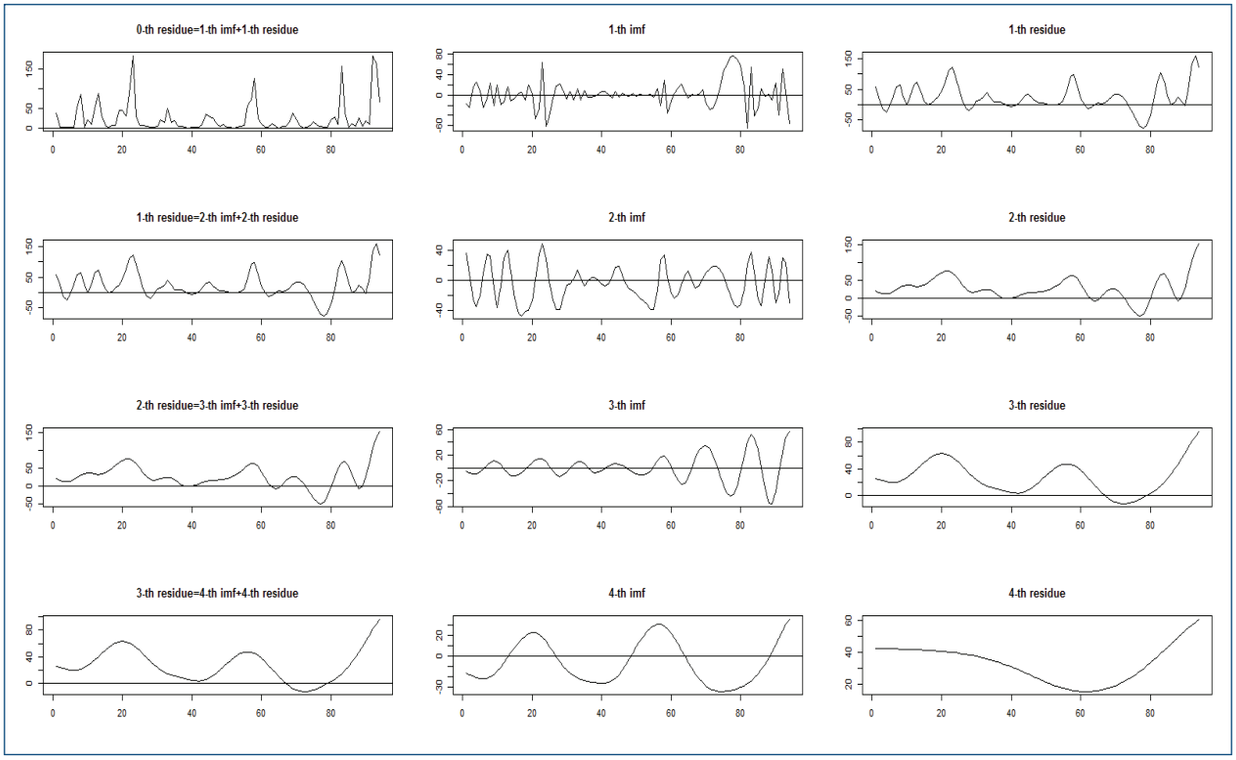
This transformed series was then decomposed into various IMFs using the EMD algorithm. Fig. 7 demonstrates the decomposition of the monthly series into varying frequency IMF series.

A maximum of 4 IMF series was extracted using the EMD algorithm for the given flu data. Along with the above 4 IMF series, the final residue series was also to be forecasted for a longer horizon.These extracted series were then forecasted using various machine learning techniques. All of these individual forecasts were summed up to obtain the final forecasted series.

Application of Machine Learning

Comparative study was performed on the given dataset with a combination of machine learning models random forest, support vector machine, and gradient boosting

Fig. 7: EMD Decomposition of Monthly Avian Flu Cases



machine.

Random forest (RF) is a typical machine learning technique which starts by creating decision trees in a recursive fashion. It selects a subset of available features and recursively partitions the data in the regression space until the amount of variation in the subspace is small. Random forest as a technique is greedy and as a result, does not necessarily converge to the global optimal solution. In order to avoid such indecisive convergence, a collection or ensemble of locally optimal trees is done which is termed as bagging. The ensemble of such trees is known as forest. Variables considered were the lagged values of temperature, humidity, precipitation, and seasonality indicators. All of these variables were scaled and centered. The model used 1000 trees with a grid search approach to sample the efficient number of features to be selected to build the final model with least root mean square error.

Another machine learning technique Support Vector Regression (SVR) (Scholkopf, 1997)is applied for forecasting in regression framework by introducing an alternative loss function. The loss function is modified to include a distance measure. It employs a rich class of non-linear modeling functions via kernels. For the current research, svmPoly kernel was used to decipher the support vectors.This kernel takes in three parameters namely degree, scale, and cost. A grid search was performed to choose these parameters automatically. Root mean square error was the metric considered to select the efficient parameters for every model.

Finally, a class of machine learning models Gradient Boosting Machine (GBM) (Friedman, 2001) which is again a tree-based model involving a recursive addition to the initial learning from the residuals was applied. It fits a tree-based model on the residuals using the specified list of variables at hand and explains the variance in the residuals. Total number of trees specified for model building was 500 with interaction depth as 5 and learning weight ofiteration was 0.1.

Fig. 8 shows the fit for the class of machine learning models discussed. Detailed syntax used for developing these machine learning models along with IMF extraction is appended in Appendix B for reference.

Fig. 8: Fit using Random Forest, Support Vector and Gradient Boosting

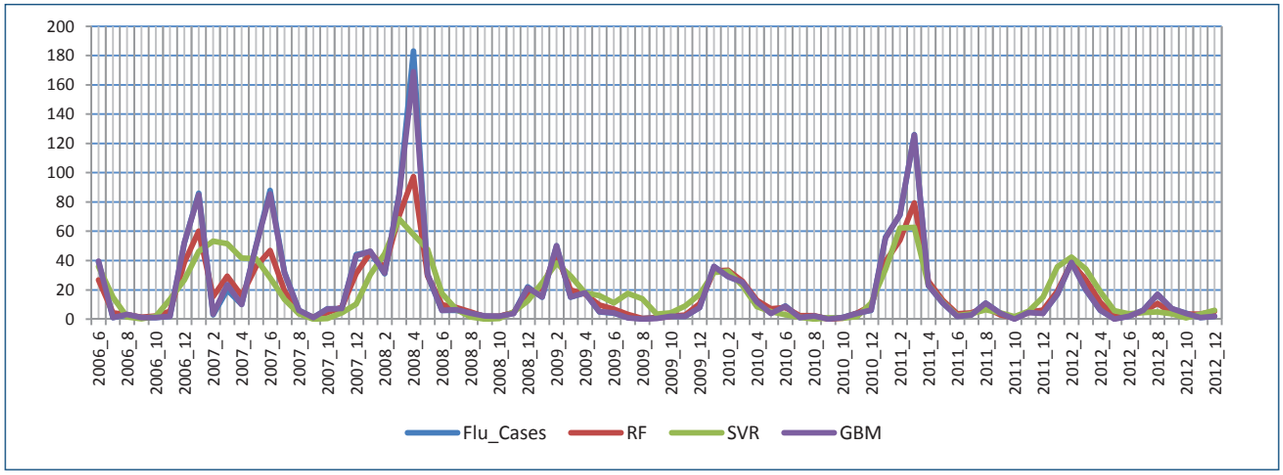
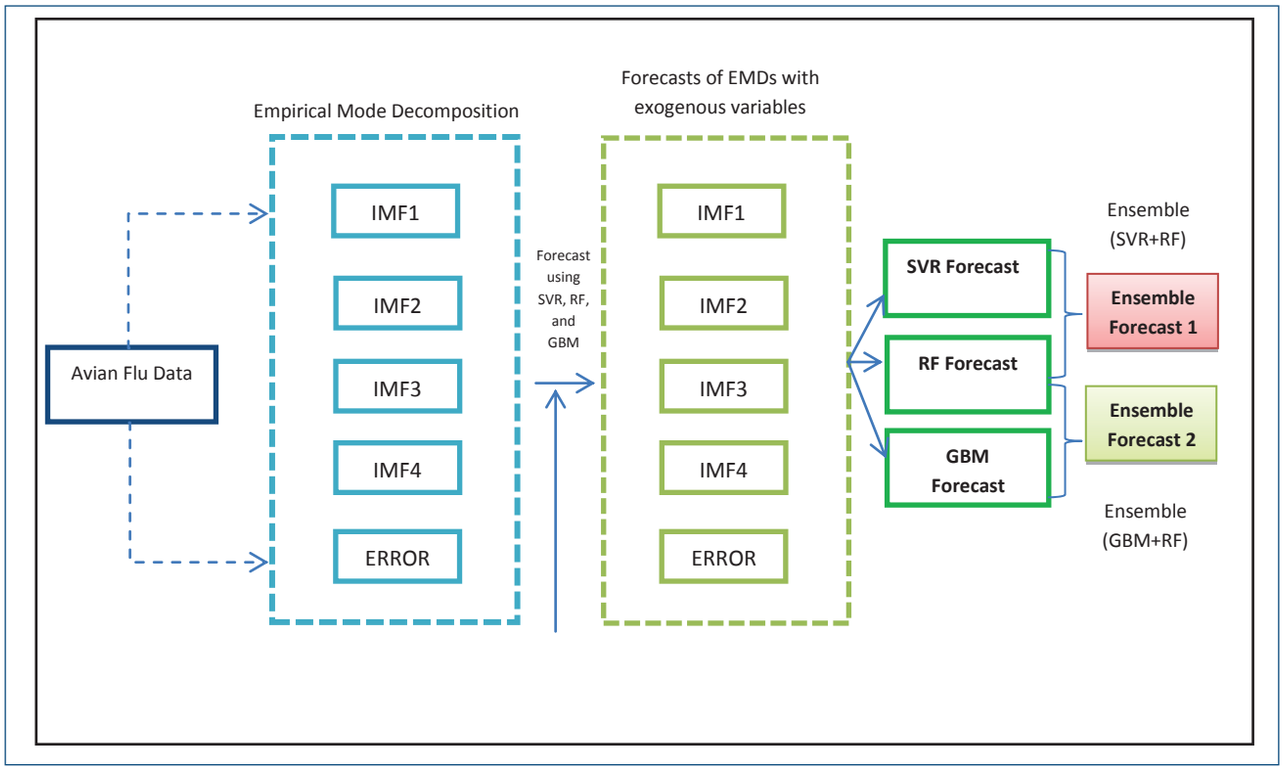


Fig. 9: Demonstration of Ensemble of various Machine Learning Models



Ensemble Approach

A state of art ensemble approach for harnessing the power of all the machine learning models was applied for coming up with the efficient solution. A flowchart explaining the ensemble strategy is shown in Fig. 9.

Ensemble was done in order to come up with a stable forecast for longer horizon. Each of these machine learning techniques had a better fit when compared to any of those classical benchmark techniques applied during the preliminary analysis. A lot of numerical iterative methods along with some intuitive combination of models were performed to come up with ensemble coefficients.

Table 1: Comparison of Mean Absolute Percentage Errors

Forecasting Methods	*MAPE(DEVJun'06 – Dec'12)*	*MAPE(VAL Jan'13 – Jul'14)*
Traditional Univariate Forecasting Methods		
Holt's Winters	152.0%	108.5%
ARIMA	90.7%	89.5%
AR-GARCH	102.3%	88.9%
Machine Learning Techniques		
Support Vector Regression	75.4%	69%
Random Forest	65%	60%
Gradient Boosting Machine	24.3%	49.2%

As shown in Fig. 9, Ensemble Forecast 1 is a combined result of first half of a year using SVR and second half of a year using RF model. Similarly, Ensemble Forecast 2 is a combined result of first half of a year using GBM and second half of a year using RF model.

Results

A summary shown in Table 1 compares results of individual machine learning models and shows improvement overtraditional univariate models. The model performance is compared in the hold out period (Jan'13 to Jul'14) basis the mean absolute percentage error. There was a significant improvement observed in the forecasts that were obtained using machine learning methods as compared to the traditional methods.

The results from the ensemble approach seemed to reduce on the error when compared to individual machine learning approaches and the traditional approaches. Ensemble 2, which constituted of intuitive ensemble of GBM and RF, considered the output of Gradient Boosting Machine for the first half of a year and output of Random Forest for the next half of the year to come up with the complete year forecast. This approach of ensemble significantly is reduced on the MAPE when compared to the best benchmark set by traditional models. For the hold out period Ensemble 2 model is reduced on the MAPE to 44.6% in comparison to best performing traditional model at 88.9% MAPE. Table 2 is a tabular view to the MAPE obtained in the development and validation periods for the two ensemble approaches used.

A visual display to show the model performance in the hold out period for different ensembled model is shown in Fig. 10. The green line which is the ensemble of GBM and RF is seen to trace the peaks to some extent and follow similar pattern as the actuals.

Mean absolute percentage error obtained from all the machine learning models along with ensemble models is further analysed and a comparison is done for

Table 2: Mean Absolute Percentage Errors for the Ensembled Models

Forecasting Methods	*MAPE (DEVJun'06 – Dec'12)*	*MAPE (VAL Jan'13 – Jul'14)*
Traditional Univariate Forecasting Methods		
Holt's Winters	152.0%	108.5%
ARIMA	90.7%	89.5%
AR-GARCH	102.3%	88.9%
Machine Learning Techniques		
Support Vector Regression	75.4%	69%
Random Forest	65%	60%
Gradient Boosting Machine	24.3%	49.2%
Ensemble of Machine Learning Methods		
Ensemble 1 (SVR+RF)	70.6%	67%
Ensemble 2 (RF+GBM)	28.7%	44.6%

Fig. 10: Graph Showing the Actual vs Forecast for Ensemble 1 and Ensemble 2 Techniques

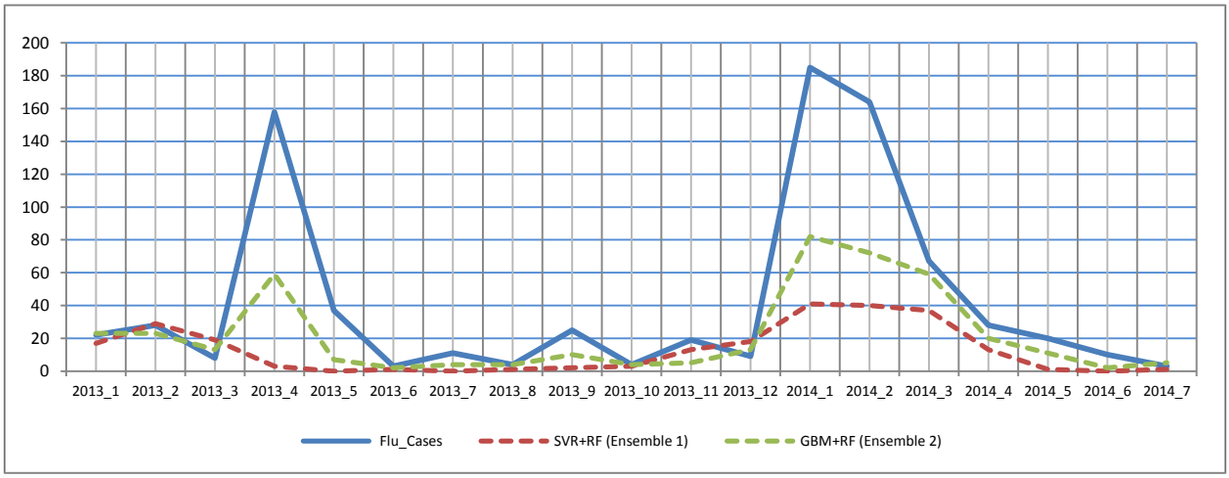
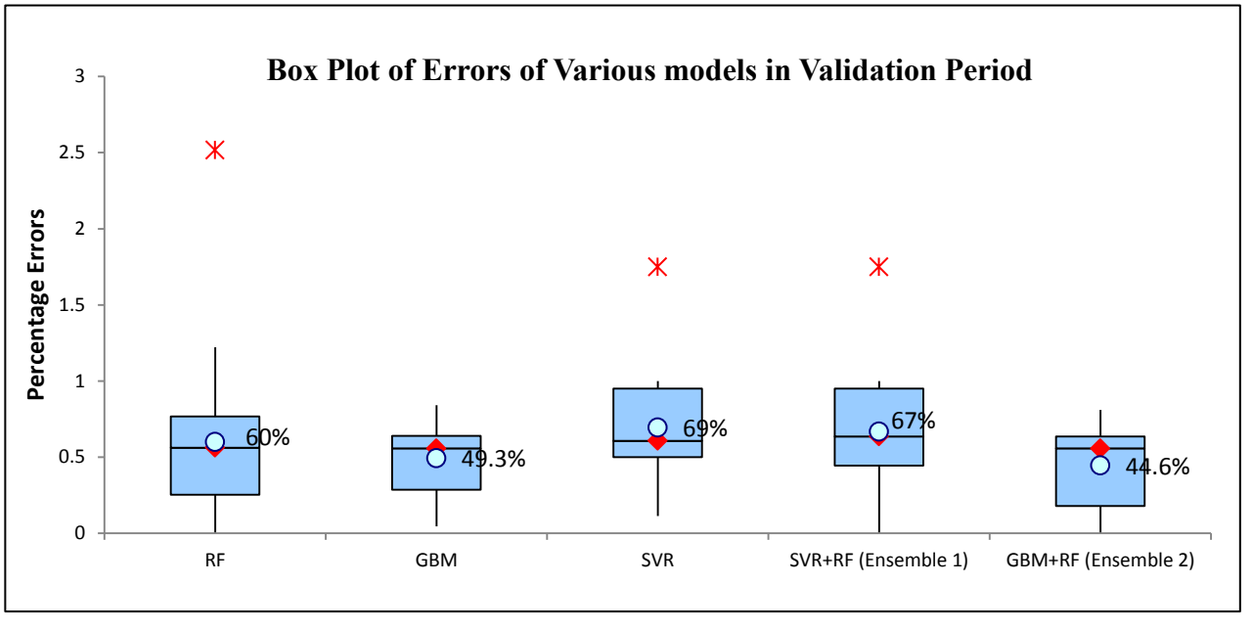


Fig. 11: Boxplot Analysis of Errors of Various Machine Learning Models and their Ensembles



the validation period (Jan'13 to Jul'14) using the boxplot technique.

Fig. 11. shows the distribution of errors for various machine learning models (RF, SVR and GBM) and their ensemble models (Ensemble 1 and Ensemble 2). It is seen that there are outliers in Random Forest, Support Vector Regression, and Ensemble 1 approaches due to which, these models do not yield a better MAPE in the validation period. In other words, these models fail to predict few data points closely. In Ensemble 2, it can be seen that the power of two models (GBM and RF) have reduced the MAPE to 44.6%, denoted by a small white dot. Also, there are no outliers for this ensemble demonstrating uniformity in error distribution and stability in the forecast.

Conclusion

Each method of forecasting has its own strength and weaknesses and hence an ensemble of these non-linear techniques tries to minimise on their shortcomings. In the present research work, ensemble of machine learning

techniques random forest and gradient boosting machine models provide an enhanced predictive ability over existing time series models (ARIMA) for the prediction of Avian Influenza outbreaks in Asian regions. This ensemble model takes advantage of each of its components random forest and gradient boosting, and recursive learning component, to generate good prediction efficiency. Also, the proposed approach is capable of handling time series prediction over a longer horizon. As next steps for its improvement additional factors could be incorporated into the predictive model. Also, more granular study for any specific country/location and incorporating the Geographic Information System to track the outbreak would improve the findings of research work further.

Acknowledgement

We would sincerely like to thank the leadership of Advanced Research and Analytics (ARA) for supporting our research work. We would also like to pay special gratitude tothe head of Instructional Design team (ARA) for helping us in figuring out relevant data for the analysis. Last but not the least, we also acknowledge the support and critical feedback of our lead time-series modeler.

References

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day

Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives,* 15(4), 157-168.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics,*29, 1189-1232.

Husin N. A., Salim N., & Ahmad, A. R. (2008). Modeling of dengue outbreak in Malaysia: A comparison of Neural Network and Nonlinear Regression Model. *Proc. International Symposium in Information Technology*, 1-4.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu H. H. (1998).The empirical mode decomposition method and the Hilbert spectrum for non-stationary Time Series Analysis. *Proc Roy Soc London A,* 454, 903-995

Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014).Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *Kane et al. BMC Bioinformatics,* 15, 276

Liu, Y. P., Wang, Y., & Wang, Z. (2013). RBF prediction model based on EMD for forecasting GPS precipitable water vapor and annual precipitation. *Advanced Materials Research*, 765, 2830-2834.

Labate, D., Foresta, F. L., Occhiuto, G., Morabito, F. C., Lay-Ekuakille, A., & Vergallo, P. (2013). Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG. *A comparison. Sensors Journal, IEEE,* 13(7), 2666-2674.

Wu, M. C., & Hu, C. K. (2006). *Empirical mode decomposition and synchrogram approach to cardiorespiratory synchronization*. Phys. Rev. E **73**, 051917

R Core Team: R: (2014). *A language and environment for statistical computing.*Vienna: R Foundation for Statistical Computing. Retrieved from http://www.Rproject.org/ISBN 3-900051-07-0

Scholkopf, B. (1997). *Support Vector Learning*. R. Oldenbourg Verlag, Munich

The United Nations Food and Agricultural Organization: EMPRES-iGlobal Animal Disease Information System. Retrieved from http://empres-i.fao.org/eipws3g/

the National Centers for Environmental Prediction: Global Weather Data for SWAT. Retrieved from http://globalweather.tamu.edu/

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York

Appendix A

Code for forecasting monthly Avian Influenzausing Traditional techniques like Holt's Winter, ARIMA, AR-GARCH

```
#########Tile: Flu Forecasting using Traditional techniques like Holt's Winter, ARIMA, AR-GARCH

# Removes the junk variables from the memory

rm(list=ls())

#########loading required packages for Analysis

library(TSA);library(tseries);library(forecast);library(fGarch);library(PerformanceAnalytics);library(rmgarch)

#########Setting working directory

setwd("/Users/desktop/DATASET/Research")

#########Reading Dataset

flu<- read.csv("flutrain.csv")

flu<- flu[,4]

hist(flu)

hist(exp(flu))

length(flu)

# Converting Data to time series object

t_data<- ts(flu$Flu_Cases[1:79], frequency=12, start=c(2006,6))

#Holt Winter

holt<- HoltWinters(t_data)

holt_forecast<- forecast.HoltWinters(holt, h = 19)

holt_forecast

plot(holt_forecast)

holt<- as.numeric(holt_forecast$mean)

write.csv(round(holt), file="Holt.csv")

#########creating reg variable to have intercept

t <- seq(1,79,1)

nt<- seq(80,98,1)

str(flu)

#transforming time series.

flue<- ts(sqrt(flu),frequency=12)

class(flu)

# creating Arima using Auto.Arima

flu.auto<- auto.arima(flu)

adf.test(flu)

adf.test(flu,k=12)

adf.test(diff(flu,lag=12))

adf.test(diff(diff(flu,lag=12)))

acf(flu,lag.max=200)

pacf(flu,lag.max=200)

acf(diff(diff(flu,lag=12)))

eacf(diff(diff(flu,lag=12)))

#craetingArima model manually

flu.arima<- arima(flu,order=c(0,0,1),

seasonal=list(order=c(0,1,1),period=12),xreg=t)

fluarpreA<- predict(flu.auto,n.ahead=19)

fluarpre<- predict(flu.arima,n.ahead=19,newxreg=nt)

#writing csv file

write.csv(fluarpre$pred,file="ffores.csv")

write.csv(fluarpreA$pred,file="fforea.csv")

plot(residuals(flu.arima),type="p")

res<- residuals(flu.arima)

resa<- residuals(flu.auto)

Box.test(resa)

resasq<- resa*resa

Box.test(res)

Box.test(resasq)
```

```
resq<- res*res

Box.test(resq)

plot(resq,type="l")

acf(resq)

pacf(resq)

eacf(resq)

eacf(resasq)

#Aarch Modeling for Residual

#spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = #c(1,1)), mean.model = list(armaOrder = c(0, 0), include.mean = FALSE), # distribution.model = "norm")

eacf(resasq)

#creating Garch model from residual from autoarima

resag<- garchFit(~garch(1,0),data=resa,cond.dist="norm",include.mean=FALSE)

resagp<- predict(resag, n.ahead=19)

eacf(resq)

#creating Garch model for residual of Manual Arima Model

resgp<-garchFit(~garch(2,0),data=res,cond.dist="norm",include.mean=FALSE)

resqg<- predict(resgp, n.ahead=19)

write.csv(resqg$standardDeviation,file="sd1.csv")

write.csv(resagp$standardDeviation, file="sd2.csv")
```

Appendix B

Code for forecasting monthly Avian Influenzausing Machine Learnings methods (Random Forest, Gradient Boosting Machine and Support Vector Regression)

```
rm(list=ls())

root<- "C:\\Users\\vshar50\\Documents\\Research_n_Development\\Research Papers\\DATA"

setwd(root)

library(data.table)

data<- as.data.frame(fread("Weather_Data.csv"))

t_data<- ts(sqrt(data$Flu_Cases)[1:79], frequency =12, start= c(2006,6))

library(EMD)

### Extracting the first IMF by sifting process

par(mfrow=c(2,3))

tryimf<- extractimf(t_data, check=TRUE, boundary="wave")

### Empirical Mode Decomposition

par(mfrow=c(4,3))

try<- emd(t_data, boundary="wave", plot.imf=TRUE)

# Collecting IMF

# These are the series to be forecasted

imf1= try[[1]][1:98]

imf2= try[[1]][99:196]

imf3= try[[1]][197:294]

imf4= try[[1]][295:392]

error= try[[2]]

training<- cbind(data[1:79,5:ncol(data)], error[1:79])

names(training)<- c(names(data)[5:ncol(data)], "error")

testing<- cbind(data[80:98,5:ncol(data)], error[80:98])

names(testing)<- c(names(data)[5:ncol(data)], "error")

library(caret)

# setup learning method

require(randomForest)

library(doParallel)

# try the random forest fit

# using parallel computation if available

set.seed(9)
```

```
rfGrid = expand.grid(mtry = c(3,5,7,9,11,15,20))
cluster<- makeCluster(detectCores())
registerDoParallel(cluster)
# applies for each classification or regression fit
fitControl<- trainControl(
method = "repeatedcv",
number = 5,
repeats = 5,
classProbs = FALSE,
verboseIter = TRUE,
preProcOptions=list(thresh=0.95,na.remove=TRUE,verbose=TRUE),
seeds = NA,
allowParallel = TRUE
)
paste(names(training),collapse="+")
#################################################
# Forecasting IMFs 1st , 2nd, 3rd, 4th and error using Random Forest
cluster<- makeCluster(detectCores())
registerDoParallel(cluster)
#Random Forest Code
fit.raf<- train(error~<List of Variables>,
data=training,
method="rf",
preProcess=c("center","scale"),
tunelength=10,
tuneGrid = rfGrid,
trControl=fitControl,
ntree = 1000,
importance=TRUE,
metric="RMSE")
stopCluster(cluster)
predicted.raf<- predict(fit.raf,newdata=testing)
fitted.raf<- predict(fit.raf,newdata=training)
#################################################
#Support Vector Machine
library(e1071)
cluster<- makeCluster(detectCores())
registerDoParallel(cluster)
svm<- train(error~<List of Variables>,
data=training,
method = "svmPoly",
trControl = fitControl,
preProc = c("center", "scale"),
tuneLength = 10,
metric = "RMSE")
stopCluster(cluster)
predicted.svm<- predict(svm,newdata=testing)
fitted.svm<- predict(svm,newdata=training)
#################################################
#Gradient Boosting Machine
set.seed(9999)
cluster<- makeCluster(detectCores())
registerDoParallel(cluster)
gbmFit<- train(error~<List of Variables>,
data = training,
method = "gbm",
```

```
trControl = fitControl,
verbose = FALSE,
## Only a single model can be passed to the
## function when no resampling is used:
tuneGrid = data.frame(interaction.depth = 5,
n.trees = 500,
shrinkage = .1),
metric = "RMSE")
stopCluster(cluster)
predicted.gbm<- predict(gbmFit,newdata=testing)
fitted.gbm<- predict(gbmFit,newdata=training)
#Predicted Residue
par(mfrow=c(1,1))
pres<- error[80:98]- (predicted.raf)
plot(predicted.svm, type="l")
plot(error[80:98], type="l")
plot(error, type="l")
#Predicted and Fitted
raf<- c(fitted.raf,predicted.raf)
svm<- c(fitted.svm,predicted.svm)
gbm<- c(fitted.gbm,predicted.gbm)
final<- cbind(data$Key,error, raf,svm,gbm)
write.csv(final,"LogError.csv", row.names=FALSE)
#Storing the Predicted IMF's
imf1_pred<- predicted.raf
imf2_pred<- predicted.svm
imf3_pred<- predicted.svm
imf4_pred<- predicted.gbm
error_pred<- predicted.raf
final_series=imf4_pred+imf3_pred+imf2_pred+imf1_
pred+error_pred
final_series<- final_series^2
final_series1=ifelse(final_series<0,0,round(final_series))
flu=imf1+imf2+imf3+imf4+error
flu<- data$Flu_Cases[80:98]
abs_error= abs(final_series1-flu)
MAPE= mean(abs_error/flu*100)
```

Article can be accessed online at http://www.publishingindia.com

Use of Five point Du Pont model and Regression tools to study Information Technology Sector of India

Kartik Sachdev*, Arushi Jamaiyar **

Abstract

This paper specifically concentrates on 5-point Du Pont analysis of the Indian IT software industry. Step by step development of 5-point model from the initial 2-point model is discussed. Trend of the industry over the past eight years is analysed and the effect of recession is studied using the time series analysis. Cross sectional analysis has been performed to discuss the significant focus points of different companies and these have been benchmarked with industry standards. Also, regression analysis has been performed to comprehend the relative impact of the five ratios on return on equity. Finally, using regression analysis, Indian IT industry is compared with IT industry of the USA and the differentiating factors are examined.

Keywords: DuPont, Regression, IT, Recession, Financial Ratio

Introduction

Acceptance of a decision to an organisation's financial management depends on the answer to question "Will this decision increase the value of the owner's equity?" Fundamentally, every decision that affects productivity, product costs, tax limits, etc. directly impacts the overall profit generated for investors. Likewise, any decision which impacts the type or amount of debt and equity used affects the financial structure.

The financial statements, namely, Balance Sheet and Statement of Profit and Loss, of an organisation accommodate significant and extremely important information about its financial health. For any firm, there are numerous models to determine how well IT is operating. Among these the DuPont model has become one of the significant tools of financial analysis for assessment of the profitability. IT can thus be applied to industries as well, to comprehensively analyze the important factors and examine its overall growth pattern.

In this paper we have used the 5-point Du Pont model comprising following ratios – profit margin, capital turnover, finance structure ratio, financial cost ratio, and tax effect ratio to analyze the Indian IT industry as this sector is proved to be very crucial in the growth and development of India. The Indian software industry has been a story of exceptional success. IT has helped wipe out the image of India as a poverty stricken backward nation to the one with highly intelligent techno-savvy brains, carving out a niche for India in the software world.

A time series analysis has been carried out to identify how the various Du Pont ratios have supplemented the trend of change in the overall return on equity over the years. In the study apart from the internal factors such as financial structure, sales, etc., the effect of environment over the years has also been discussed. The period of recession has been critically examined and its impact on the Indian IT companies brought out.

Regression tools have also been employed to compute the weighted dependence of ROE on its constituent components and the physical analogies are drawn. The same has been compared with that of the USA to highlight the distinguishing attributes.

Literature Review

The DuPont Model is an important financial analysis tool in imparting an overview as well as focus points for an

* Overall Coordinator at Training and Placement Cell, IIT Delhi, India. Email: kartik2094@gmail.com
** Financial Analyst at Deutsche Bank, Mumbai, Maharashtra, India. Email: arushi.jamaiyar@gmail.com

industry or organisation. IT helps to identify the significant areas of strength and weakness evident in the financial statements. Hence, IT can be used as a compass in the process by directing the organisation toward important areas.

The Du Pont model was created by F. Donaldson Brown who brought about the model while he was allocated to clear up the finances in General Motors. The elegance of return on asset (ROA) being impacted by an efficiency measure and a profitability measure led to the DuPont model becoming a widely-used tool of financial analysis (Liesz, 2002). In 1970s, attention in financial analysis shifted to return on equity (ROE) from return on asset (ROA), and the DuPont model was enhanced to incorporate the ratio of total assets to equity. Later, further modification and addition of two more factors let to development of five-step DuPont model.

Progress of the DuPont model parallels the advancements in the financial analysis itself. Over the years, three distinct versions of DuPont model have been developed and utilised to help explain the fundamental drivers of profitability and return.

Two-step Du Pont Model

The actual DuPont model of financial ratio analysis was created in1918 by an engineer named Brown at DuPont who was allocated with analysing the finances of a firm that DuPont was acquiring. Brown identified that a mathematical relationship prevailed between two frequently calculated ratios, namely net asset turnover (an efficiency measure) and net profit margin (a profitability measure), and return on asset(ROA). The product of the net asset turnover and the net profit margin is equal to ROA, as shown in Equation 1 below.

Eq. 1: (net income / sales) x (sales / total assets) = (net income / total assets) i.e. ROA

At this phase of time, the usual corporate aim was to maximize ROA and the understanding that ROA was affected by both efficiency and profitability led to the development of a system of control and planning for all operating decisions within a business organisation. Until 1970s, this was the commanding form of financial analysis (Blumenthal, 1998).

Three-step Du Pont Model

In 1970s the normally assumed aim of financial management became "boosting the invested value of the firm's owners" (Gitman, 1998) and the attention shifted from return on asset ROA to return on equity ROE. This led to the first significant moderation of the original Du Pont model. The method that company uses to finance its operations, i.e. its use of leverage or debt became a third subject of recognition for financial analysts in addition to profitability and efficiency. The new ratio is known as the leverage multiplier, which is computed by the equation (total assets/equity). The three-step DuPont model is illustrated in Equations 2 and 3 below.

Eq. 2: ROA x (total assets / equity) = ROE

Eq. 3: (net income / sales) x (sales / total assets) x (total assets / equity) = ROE

The three-step DuPont model (also called the "Du Pont identity") was a significant tool to demonstrate the interconnectedness of a company's balance sheet and its income statement, and to evolve straightforward policies for raising the company's ROE. The three-step DuPont model gives an outstanding method to get a fast snapshot view of the all-round performance of a company in three critical areas of ratio analysis (Isberg, 1998). However, even the modified DuPont model had its critics. The DuPont model does not sufficiently differentiate between "unfavourable" debt and "favourable" debt, on the basis of preferred stock in a company's financial structure (Boyd, 1989).

Five-step DuPont Model

Recently yet another modification was offered by Hawawini and Viallet (1999) to the DuPont model. This modification resulted in disintegration of ROE in five different ratios. In this moderation they acknowledge, that the statements of finance prepared by firms for their annual reports (that are significant to tax collectors and creditors) may not be at all times of prime importance to managers taking financial and operating decisions (Brigham and Houston). Hence, on this basis Hawawini and Viallet redefined the conventional balance sheet into a "managerial balance sheet" that is "a more suitable tool for analyzing the impact of operating decisions to the company's financial performance" (Hawawini & Viallet,

2000, p. 68).

This redefined managerial balance sheet takes into account the notion of "invested capital" instead of total assets, and the notion of "capital employed" instead of total liabilities and owner's equity present in the conventional balance sheet. The basic difference is in the consideration of the short-term "working capital" balances. As a part of invested capital, this managerial balance sheet takes into account a net value called "working capital requirement" (calculated as: [inventories + accounts receivable + prepaid expenses] – [accrued expenses + accounts payable]).

The 5-step DuPont model is illustrated below in Equation 4.

Eq. 4: (EBIT / sales) x (sales / invested capital) x (EBT / EBIT) x (invested capital / equity) x (EAT / EBT) = ROE

where invested capital = net fixed assets + working capital requirement + cash

The decisions that incorporate the management of firm's operating liabilities (accruals and accounts payable) and operating assets (mostly accounts receivable and inventories) and the disposal and procurement of fixed assets are referred as company's operating decisions. These are reflected in the first two ratios-

1. Operating profit margin: (Earnings before Interest Taxes or EBIT / sales)
2. Capital turnover: (sales / invested capital)

The decisions that regulate the mix of equity and debt utilised to fund the company's operating decisions are referred as firm's financing decisions. These are reflected in the third and fourth ratios of the five-step DuPont model. These are:

3. Financial cost ratio: (Earnings before Taxes or EBT / EBIT)
4. Financial structure ratio: (invested capital / equity)

The last factor of a company's ROE takes into consideration the effect of business taxation. The return on equity ROE is inversely proportional to the tax rate on company's earnings before taxes (EBT). This is reflected in the fifth ratio of the five-step DuPont model.

5. Tax effect ratio: (Earnings After Taxes or EAT / EBT)

Analysis of Information Technology Industry

Overview

The Indian IT industry was born in 1974 when a group of Mumbai based conglomerates supplied programmers to the overseas IT firms for the first time. This opened the doors for a potential global market opportunity for India as well as the world. Advances in technology and communication, coupled with very low operating costs encouraged many western software companies to outsource their products to India. In the 1990s, the government relaxed the restrictions and liberalised the economy leading to a further growth.

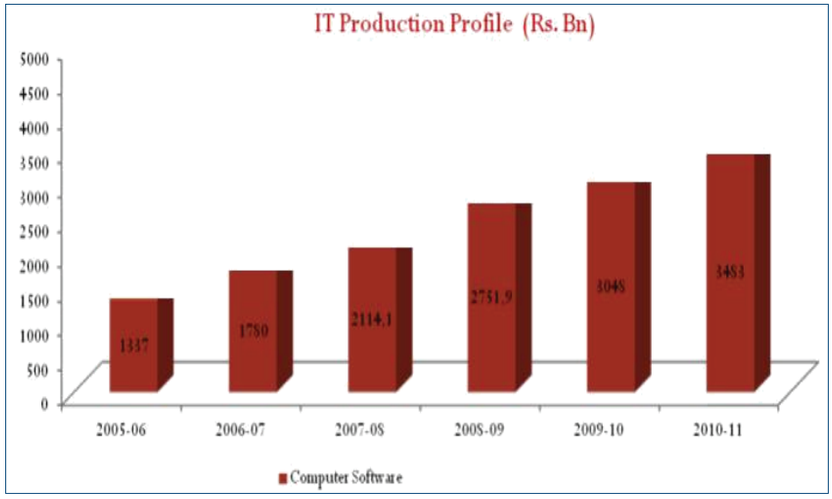
In 2007- 2008, the world witnessed one of the worst economic crises in its history. Large financial institutions crumbled, economies shrunk drastically and stock markets crashed across the globe. India was also hard hit by the economic slump with the software companies clocking a very low growth rate. The rising inflation and sense of insecurity caused demands to dip dramatically with about 40% of the Western companies cutting down their IT expenses.

However, the industry was able to cope with the slowdown by introducing methods of cost saving, thereby recording a good growth and helping India in surviving the blow. In 2005, the software exports contributed 15-20% of the total exports from India. During 2004-2009, the major market for the IT products and services was primarily the USA, accounting for about 60% of the total exports. The industry recorded a turnover of $60 billion in 2009 with a 78% contribution from exports and a CAGR of over 30%.

Today, the software industry contributes to 8% of the GDP and has its base in the Silicon Valley of India, Bangalore. IT ranks fourth in India's total FDI share and accounts for approximately 37% of total private equity and venture investments in the country. The industry stepped up with a 13.1% CAGR in the financial years 2008-2013 despite the global economic constraints.

The bar graph shown in Fig. 1 depicts the growth of the software industry over the years.

Fig. 1: Growth of Software Industry Over the Years



Source: Department of IT, Ministy of Communication and IT

Industry Structure

India's IT industry can be divided into five main components, viz. software products, IT services, engineering and R&D services, ITES/BPO (IT-enabled services/Business Process Outsourcing) and hardware.

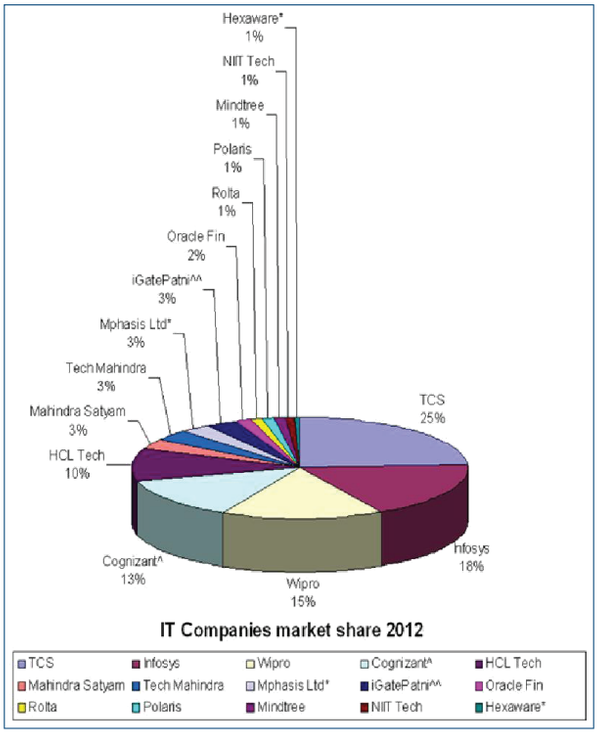
The IT sector is highly competitive sector, as more and more MNCs and startups try to invest in the technology hubs of India and replicate of the models of offshoring. Most of the industries are clustered at Bangalore, Pune, Delhi, and Chennai. However, the major chunk of the industry (81%) is captured by the five major players, namely TCS, Infosys, Wipro, Cognizant, and HCL technologies, thus making IT an oligopoly.

Table 1 shows the contribution of exports in the incomes of the companies. IT can thus be inferred that exports

Table 1: Contribution of Exports in the Income of Companies

Company	Export	Domestic
Tata Consultancy Services (TCS)	93	7
Wipro Technologies	76	24
Infosys Technologies	99	1
HP India	18	82
IBM	58	42
Cognizant Technology Solutions	100	
Ingram Micro		100
HCL Technologies	93	7
HCL Infosystems		100
Redington India		100
Cisco India	92	8
Oracle India		
Intel India	90	10
Accenture	93	7
SAP India	79	21
Dell India		100
Tech Mahindra	98	2
Microsoft India	90	10
MphasiS	100	
Patni Computer Systems	99	1

Fig. 2: IT Companies Market Share



continue to be major part of the revenue. During the financial year 2013, exports accounted for 68% of the industry's revenue.

Data

- We have used the following sources for data collection-
- capitaline.com for collecting balance sheet and income statement for Indian companies
- finance.yahoo.com for collecting balance sheet and income statement of US companies
- google.com for other data collection like market share, tax rate, interest rate, etc.

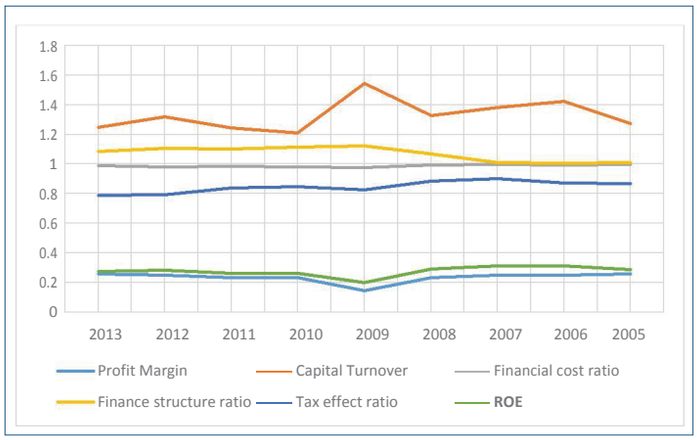
For each company the five ratios considered in DuPont model were calculated using the data taken from balance sheets and income statements.

The companies used in this study constitute a market share of more than 90 percent. The regression analysis have been performed using the regression tool in Excel.

Time Series Analysis (2005-2013)

In Fig. 3, the complete sector comprising large, medium and small software companies has been covered.

Fig. 3: Analysis of Large, Medium and Small Software Companies



In all the five ratios, there has been a significant change in the years 2008-2010, for instance the ROE, profit margin showing a sharp dip. The capital turnover ratio also reached a peak at the same time along with an increase in the finance structure ratio curve. The major reason for these changes can be explained by the economic slump that the world witnessed in the years 2008-2010.

ROE and Profit Margin

2010-2013 - Impact of the Western Economic Crisis And Falling Rupee

The ongoing global economic slowdown has impacted the Indian IT software sector in many ways. The Indian IT software industry has witnessed an overall gain since corporate in the Western countries are focusing on cutting costs and hence the outsourcing has been increased by them. This has caused more business opportunities to come in the hands of the Indian IT software sector. An advantage has been the fact that the depreciating rupee versus the dollar has caused the Indian outsourcing businesses to reap the effect in terms of appreciated rupee revenues. This has also made the competition among these companies fierce since their services cost less in dollar terms.

This can be observed from the graph given below, as the ROE and profit margin curves have been showing a steady rise from 2010-2013. Thus, both these components have a close positive correlation.

2008-2010 - Impact of the recession on the Indian IT sector

Lag phase: The severe economic depression that resulted from the subprime mortgage crisis of 2007, engulfed the entire world economy. The western world thus experienced a profound economic crunch, low growth and squeezed budgets during the years 2008-2010. India being one of the fastest growing tech markets and a major exporter of IT goods and services, was also hit hard by the global economic crisis of this period.

- About 43% of the Westerners, including USA, cut down their IT spends which lead to a drastic decline in the overall profit of the Indian software companies.
- The slowing US economy witnessed 70% of the firms negotiating at lower rates with suppliers and nearly 60% cutting back on contractors. With budgets squeezed, just

over 40% of companies planned to increase their use of offshore vendors, thereby adversely affecting the income of the Indian IT sector.

This can be reflected from the sudden dip in the curve of ROE and profit margin during the financial year 2009.

Growth phase: Despite the slowdown the industry displayed resilience in countering the negative effects. The IT services and outsourcing markets underwent a structural transformation, wherein more work was moved to lower cost offsite locations, which increased the cost savings of the industry. This led to a dramatic increase

Fig. 4: Profit Margin and ROE of IT Industry

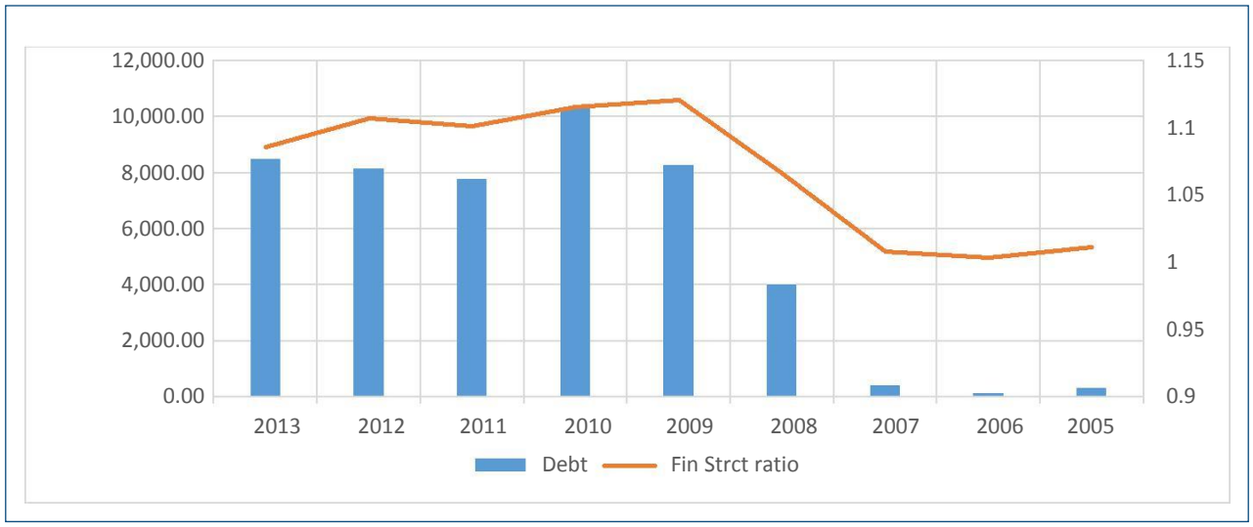


Fig. 5: Financial Structure Ratio Over the Years

in the ROE and profit margin from 2009-2010, as can be seen from the curve in Fig. 4.

Financial Structure Ratio

The financial structure ratio also saw a sharp increase in the years 2007-2009. This can be explained by the fact that due to the economic depression, the companies increased their debt significantly from Rs 400 crores in 2007 to Rs 8273 crores in 2009. From 2010 onwards, there has been a decline in the trend of borrowings, thereby reducing the financial structure ratio.

Financial Cost Ratio

Tax Effect Ratio

This ratio shows almost a similar trend as ROE as the tax rates have been fairly constant over the years. However, during 2010-2013, IT has shown an increase due to the lowering of corporate tax rates by the Indian government, as indicated in Fig. 6.

Fig. 6: Tax Effect Ratio Over the Years

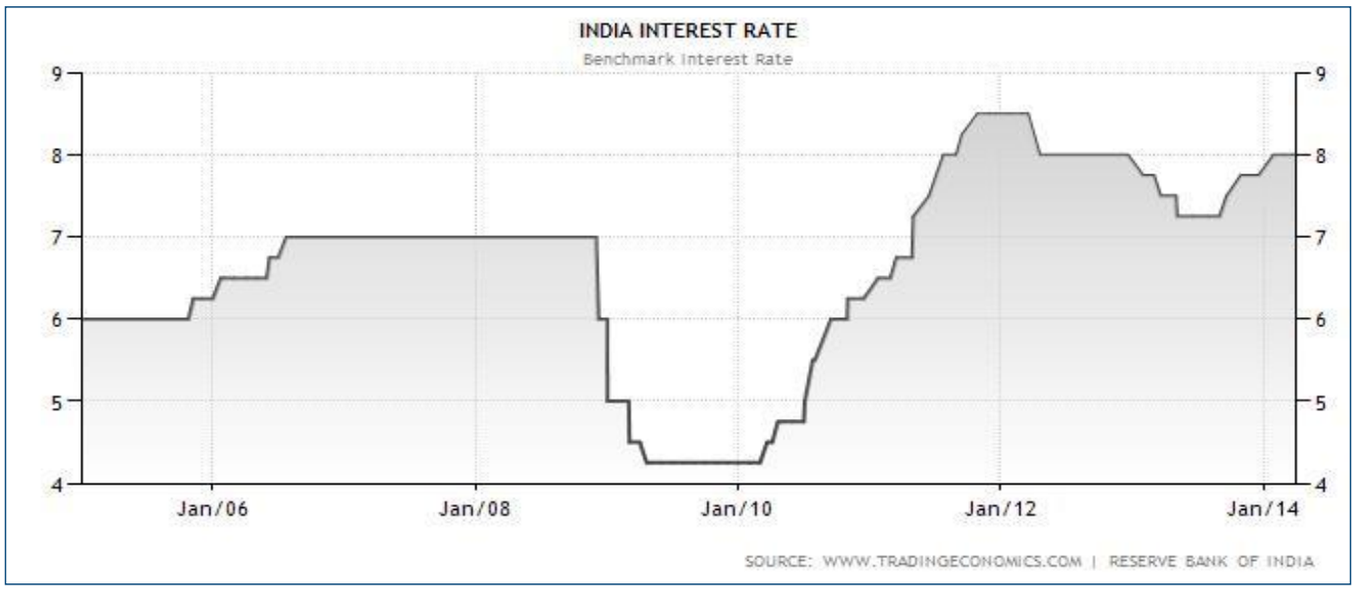
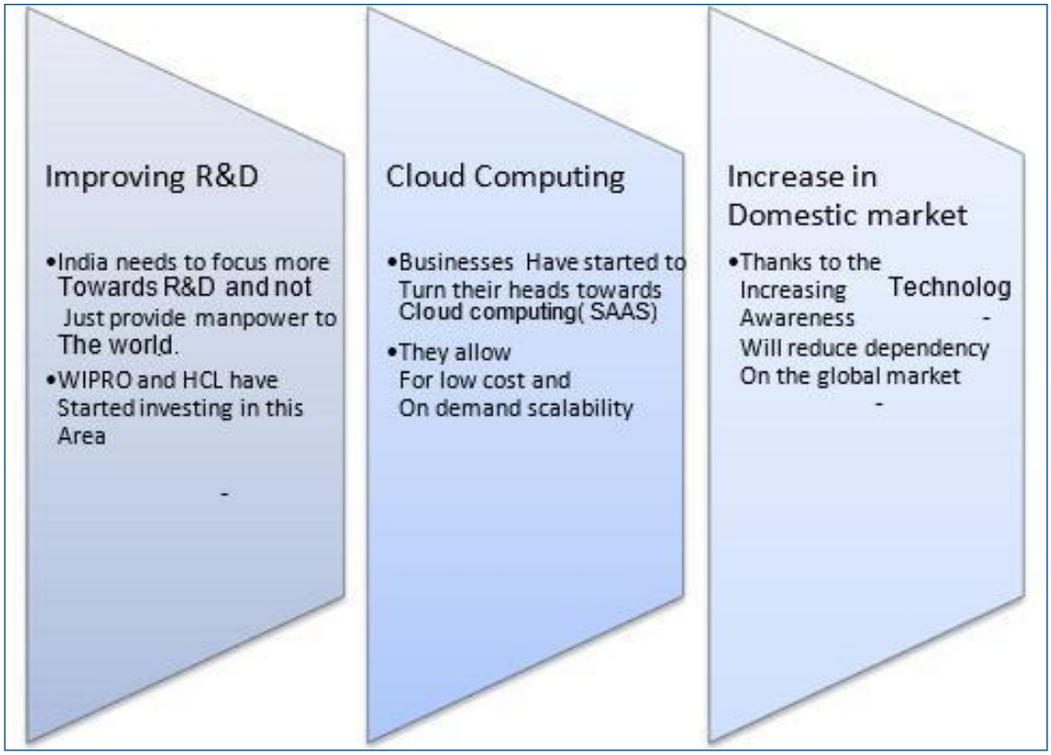


Fig. 7: Legend for Cross Sectional Analysis Results



Cross-sectional Analysis of the Indian IT sector

The various ratios of the companies are compared with the industry reference. On the basis of comparisons, analysis have been made to show what are the strong points of a particular company and other are the areas that is can work upon.

The legend used for all the five parts is specified as shown in Fig. 7.

Profit Margin Analysis

The profit margin of the following companies is significantly lower than the industry average-Capgemini, HP, IBM, Polaris, and Tech Mahindra. These companies need to focus on improving the profit incurred on the sales.

Table 2: Cross Sectional Analysis (Profit margin)

	Profit Margin								
Company	2013	2012	2011	2010	2009	2008	2007	2006	2005
IT Industry	0.259307	0.247892	0.232013	0.234253	0.143599	0.233357	0.249146	0.250461	0.257826
Capgemini	0.125092	0.132854	0.113711	0.158718	0.136575	NA	NA	NA	NA
Cognizant	0.291472	0.210089	0.249279	0.224975	0.407705	0.26689	0.305816	0.350605	0.344752
HCL Techno	0.35596	0.267692	0.196779	0.240443	0.242604	0.18687	0.282949	0.213901	0.225179
HP	0.113912	0.119466	0.108079	0.087457	0.178704	0.105792	0.084603	0.11085	0.15285
IBM	0.081379	0.118455	0.103813	0.073027	0.063146	NA	NA	NA	NA
Igate	0.25814	0.341694	0.326214	0.254434	0.329632	0.301048	0.27233	0.356269	0.348609
Infosys	0.29857	0.316412	0.34687	0.332504	0.340254	0.317692	0.313392	0.306948	0.295732
Mphasis	0.213781	0.214455	0.256256	0.281782	0.254007	0.182944	0.131574	0.128411	0.215409
Oracle	0.445501	0.475765	0.413706	0.323417	0.307569	0.234321	0.239735	0.240684	0.265159
Polaris	0.10906	0.132552	0.150503	0.129717	0.105473	0.067432	0.103856	0.030365	0.088365
TCS	0.310598	0.328095	0.292934	0.274752	0.22679	0.263824	0.275451	0.272514	0.260557
Tech Mahindra	0.156805	0.128081	0.180142	0.226069	0.249066	0.109283	0.04816	0.195889	0.091835
Wipro	0.218614	0.198516	0.216249	0.242204	0.169895	0.199902	0.226535	0.225189	0.240152

Table 3: Cross Sectional Analysis (Capital Turnover)

Capital Turnover									
Company	2013	2012	2011	2010	2009	2008	2007	2006	2005
IT Industry	1.249876	1.31986	1.24297	1.21169	1.54443	1.32819	1.38061	1.42383	1.27304
Capgemini	2.0124168	2.19292	2.19106	1.56004	2.55739	NA	NA	NA	NA
Cognizant	0.9665366	1.03797	1.00925	0.93496	1.04061	1.17457	1.45725	1.20849	0.97958
HCL Techno	1.1321288	1.12538	0.99833	0.82361	1.25636	1.47702	1.21412	1.20354	0.51748
HP	3.4606543	2.73663	2.34298	1.18927	1.32904	1.57177	1.63787	1.46584	1.33096
IBM	2.7615457	2.97507	3.57975	3.31716	2.85536	NA	NA	NA	NA
Igate	0.6793622	0.70976	0.56296	0.65394	0.52436	0.47272	0.44585	0.5349	0.65167
Infosys	1.1050735	1.07971	1.1304	1.08179	1.00322	1.18704	1.2106	1.21215	1.34189
Mphasis	0.9193988	0.91112	0.97025	1.33413	1.61029	1.24439	1.74229	1.33174	0.74004
Oracle	0.4594202	0.52302	0.48134	0.55326	0.67844	0.65476	0.66268	0.87061	0.82824
Polaris	1.5699907	1.57218	1.54942	1.488	1.73093	1.59299	1.5918	1.29505	1.29422
TCS	1.523249	1.61138	1.50147	1.53248	1.68273	1.7218	1.86859	2.00146	2.36689
Tech Mahindra	0.9871647	1.02696	0.99192	0.91465	2.33301	2.79757	2.99256	2.05475	1.92532
Wipro	1.1310248	1.09792	1.01537	1.02917	1.25743	1.16252	1.47018	1.60785	1.48066

Oracle, TCS, and Infosys have outperformed the others owing to their vast export market. For example- Oracle, Infosys, and TCS generated 96 percent, 94 percent and 91 percent respectively of the revenue through exports.

Capital Turnover Analysis

Capgemini, HP, IBM, Polaris, and TCS have performed better than the industry average in terms on capital turnover ratio. Better utilisation of the capital employed has allowed them to outperform other companies. Table 3 also shows the effect of recessionon Tech Mahindra and Wipro. The companies have not been able to recover from the dipping of sales and increase in capital invested during the time of initial recession. HCL, Cognizant, Infosys and Oracle enjoy a major market share i.e. have high amount of sales but need to work on reducing the capital employed.

Financial Cost Ratio

HP, Cognizant, Infosys, Oracle, and TCS have marginal or no interest payments owing to leverage free capital. Tech Mahindra and Wipro need to significantly work upon in reducing their debts as they are paying our 12% and 5% respectively of their operating income as interest, which is pretty high than the industry average i.e. 1.2%.

Table 4: Cross Sectional Analysis (Financial Cost Ratio)

Financial Cost Ratio									
Company	2013	2012	2011	2010	2009	2008	2007	2006	2005
IT Industry	0.9876409	0.979865	0.986631	0.981301	0.974888	0.991676	0.995691	0.994921	0.996656
Capgemini	0.9645779	0.963224	1	0.999975	0.999702	NA	NA	NA	NA
Cognizant	1	1	1	1	1	1	1	1	1
HCL Techno	0.9834545	0.960427	0.927124	0.919182	0.976971	0.978677	0.989107	0.980887	0.983688
HP	1	1	1	1	1	1	0.998754	0.998104	0.994125
IBM	0.9862465	0.985679	0.960315	0.871219	0.813472	NA	NA	NA	NA
Igate	0.9950599	0.993971	0.988548	0.984563	0.984432	0.971789	0.98357	0.999415	0.999326
Infosys	0.9995717	0.999757	0.999829	0.999887	0.999734	0.999702	0.999805	0.999759	0.999635
Mphasis	0.9817012	0.984396	0.99756	0.998501	0.989636	0.996402	0.997288	0.989861	0.939398
Oracle	0.9999003	0.999879	0.999797	0.999692	1	1	1	1	1
Polaris	0.9882142	0.992487	0.99843	0.996879	0.995661	0.990724	0.993963	0.973747	0.986455
TCS	0.9980539	0.998775	0.997705	0.998505	0.998555	0.999317	0.999178	0.998542	0.9951
Tech Mahindra	0.8823148	0.849309	0.878666	0.845291	0.997713	0.975284	0.948353	1	1
Wipro	0.9533708	0.907162	0.976718	0.982759	0.947446	0.967433	0.997738	0.998664	0.996838

Table 5: Cross Sectional Analysis (Financial Ratio)

Financial Structure Ratio									
Company	2013	2012	2011	2010	2009	2008	2007	2006	2005
IT Industry	1.0858516	1.107453591	1.101716915	1.115982432	1.121024406	1.066509784	1.008211319	1.003819233	1.011427047
Capgemini	1.1121057	1.149437141	1	1	1.003450331	NA	NA	NA	NA
Cognizant	1.0739717	1.09909624	1.000255724	1	1	1	1	1	1
HCL Techno	1.1200929	1.235012669	1.208721402	1.283109732	1.147274843	1.00787911	1.011938616	1.005194362	1.033807929
HP	1.0641841	1	1	1.000036656	1	1	1.000036986	1.000448368	1.008996235
IBM	1.0930521	1.115450357	1.115024654	1.4440995	1.996192538	NA	NA	NA	NA
Igate	1.0003549	1.000330267	1.000293407	1.000694114	1.000930345	1.00138538	1.001557236	1.002130195	1.002904197
Infosys	1.0086477	1.003327879	1.000705716	1.001020367	1	1	1	1	1
Mphasis	1.0083971	1.078185743	1.078466527	1.000295679	1.062480862	1.012922482	1.004853623	1.001579779	1.002950526
Oracle	1.0082895	1.00745313	1.00876447	1	1	1	1	1	1
Polaris	0.8199264	0.797183575	0.865158664	0.870353098	0.880922199	0.821473159	0.848055408	0.651960784	0.894675151
TCS	1.0212924	1.018368138	1.012625437	1.00236428	1.003052141	1.001658358	1.006296074	1.006328742	1.036355972
Tech Mahindra	1.4305011	1.501887779	1.517021277	1.744723899	1	1.077336373	1.055808656	1	1
Wipro	1.261487	1.229204394	1.247826005	1.312578424	1.400634444	1.32921357	1.025535385	1.007812537	1.012691009

Financial Structure Ratio

Most of the big market players like Infosys, TCS, Oracle, and HP have net capital invested to equity ratio less than the industry average. As discussed earlier, debt to equity ratio of Tech Mahindra and Wipro is comparatively has been comparatively large.

Tax Effect Ratio

The tax effect ratio is impacted by the tax policies of the government, which are constant for all companies. However, allowance of no tax on debt impacts the tax effect ratio as companies based on debt have to pay lesser tax. Tax management methods to decrease the tax burden are adopted, to one degree or another, by most of the companies.

Return on Equity

TCS having the highest market share also has very high return on equity. Other players like HCL, HP, Infosys, and Cognizant have return on equity more than industry benchmark and also significant market share. It can also be seen that post-recession, return on equity of Tech Mahindra and Wipro have tipped over the industrial benchmark which were higher during pre-recession period. The two companies have not been able to recover from the losses of lag phase of recession period.

Table 6: Cross Sectional Analysis (ROE)

Return on equity									
Company	2013	2012	2011	2010	2009	2008	2007	2006	2005
IT Industry	0.27364065	0.28093037	0.26265405	0.26341728	0.19981886	0.28954336	0.31051987	0.31028923	0.28668858
Capgemini	0.19515449	0.21987986	0.21287726	0.2145625	0.33328106	NA	NA	NA	NA
Cognizant	0.25350835	0.21181878	0.24312533	0.19892584	0.4061425	0.29973682	0.41803827	0.4008477	0.30625445
HCL Techno	0.36186893	0.29522385	0.20451431	0.21406199	0.28590636	0.24282777	0.32169739	0.24783083	0.11513018
HP	0.41951346	0.32693489	0.19989543	0.11470987	0.22129529	0.14895904	0.12010553	0.14026902	0.17968024
IBM	0.18455898	0.28255054	0.33432431	0.24898267	0.29895049	NA	NA	NA	NA
Igate	0.14782717	0.22075287	0.16940513	0.1543511	0.1514821	0.0931555	0.09520685	0.1717116	0.19603329
Infosys	0.24218379	0.2528079	0.28463891	0.26296886	0.2633418	0.3267449	0.33135656	0.33891776	0.35102218
Mphasis	0.14515639	0.16823508	0.22946639	0.34274005	0.41331404	0.22621613	0.22606211	0.15861703	0.14510879
Oracle	0.14114262	0.17435938	0.18794487	0.15817112	0.19824073	0.14607932	0.14787282	0.17670412	0.17565033
Polaris	0.15359862	0.1832044	0.2039341	0.16747	0.16021152	0.0875422	0.13966833	0.02503435	0.1009332
TCS	0.39267372	0.44157152	0.38662797	0.37167693	0.34925667	0.40970626	0.46622343	0.4843484	0.55145812
Tech Mahindra	0.15599598	0.13377091	0.20588061	0.25911327	0.52450824	0.26514165	0.07425968	0.36817984	0.14723609
Wipro	0.23319507	0.19238682	0.2271883	0.27684516	0.23762076	0.26383422	0.30493326	0.31469445	0.30553671

Table 7: Regression Data -IT Sector

Regression Data - IT Sector (2013)

Company	ROE	Profit Margin	Capital Turnover	Financial cost rat	Finance structure	Tax effect ratio
Ancient Techno Holding	0.063116557	0.137326268	0.936030011	0.986216285	1.081515696	0.460357174
Capgemini	0.195154488	0.125091868	2.012416806	0.964577889	1.112105651	0.72268357
Cognizant	0.253508348	0.291471681	0.966536607	1	1.073971726	0.837885634
Geometric Ltd	0.138896735	0.120781047	1.492110891	0.993954322	1	0.775399865
HCL Techno	0.361868928	0.355959659	1.132128798	0.983454514	1.120092911	0.815166544
Hexaware	0.375829939	0.389786095	1.162870549	0.998733406	1.019277081	0.814501378
HP	0.419513461	0.113912373	3.460654329	1	1.064184139	1
IBM	0.184558979	0.081379241	2.761545712	0.986246454	1.09	0.761803514
Igate	0.147827167	0.258139909	0.67936216	0.995059938	1.000354941	0.846826392
Infosys	0.242183788	0.298569815	1.105073488	0.999571673	1.008647724	0.728038852
KPIT	0.116144084	0.196986453	0.614566431	0.942385965	1.32685735	0.767369127
Mastek	0.078950809	0.075752303	1.043430895	0.995405819	1.080232224	0.928923077
Mindtree	0.257973662	0.17715287	1.782272457	0.997644842	1.023673594	0.800047214
Mphasis	0.145156387	0.213780555	0.91939882	0.981701192	1.008397121	0.746024503
NIIT	0.218612041	0.200692254	1.423617175	0.9887351	1.043925562	0.741311548
Oracle	0.141142624	0.445500619	0.459420166	0.999900326	1.00828952	0.684002765
Polaris	0.153598617	0.109059941	1.569990699	0.988214182	1.107131036	0.81992638
Syntel	0.265569168	0.334161453	0.946599963	1	1.040623217	0.806791489
TCS	0.392673725	0.310597666	1.523249031	0.998053871	1.021292448	0.814251636
Tech Mahindra	0.155995984	0.156804984	0.987164703	0.882314835	1.4305011	0.79845815
Wipro	0.233195072	0.218614406	1.131024826	0.953370824	1.26148703	0.784194529

Regression Data - IT Sector (2009)

Company	ROE	Profit Margin	Capital Turnover	Financial cost rati	Finance structure	Tax effect ratio
Ancient Techno	0.005231623	0.024290174	0.435223241	0.852564103	1.010473517	0.57443609
Capgemini	0.333281056	0.13657513	2.557385931	0.99970168	1.003450331	0.95121005
Cognizant	0.406142503	1.040606764	0.407704779	1	1	0.95729545
Geometric Ltd	0.229313286	0.130089374	1.764543907	0.976881134	1.064939219	0.960258986
HCL Techno	0.28590636	0.242603786	1.256358743	0.976971446	1.147274843	0.836880087
Hexaware	0.162750776	0.249548407	0.681902166	0.997997844	1	0.958333333
HP	0.221295287	0.178703778	1.329041933	1	1	0.931750742
IBM	0.298950487	0.063145841	2.855359687	0.813471503	1.996192538	1.021054494
Igate	0.151482103	0.329631699	0.524355905	0.984432193	1.000930345	0.889439739
Infosys	0.263341804	0.340254218	1.003222	0.999734113	1	0.771675532
KPIT	0.365792541	0.12008374	2.239562864	0.902576593	1.690384615	0.891492686
Mastek	0.27352016	0.158608806	1.708520308	0.998417888	1.000486131	1.010458483
Mindtree	0.056584207	0.049524743	1.523748376	0.644044321	1.26278377	0.921966206
Mphasis	0.413314039	0.254006697	1.61028964	0.989635539	1.062480862	0.961024793
NIIT	0.294950842	0.182302615	1.802273029	0.991807424	1.002899517	0.902508668
Oracle	0.198240735	0.307568891	0.678439742	1	1	0.950034139
Polaris	0.160211521	0.10547291	1.73092543	0.995661434	1.000518717	0.880922199
Syntel	0.241871863	0.276371896	1.040992681	1	1	0.84070532
TCS	0.349256673	0.226789774	1.682733586	0.998554534	1.003052141	0.91371464
Tech Mahindra	0.52450824	0.249065719	2.333014354	0.997712717	1	0.904722604
Wipro	0.237620756	0.169894698	1.257433481	0.947445723	1.400634444	0.838185969

Regression Analysis

Using multiple regression model, by taking the five ratios of DuPont model (profit margin, capital turnover, financial cost ratio, finance structure ratio, and tax effect ratio) as independent variables, we wish to analyse their relationship with one dependent variable (return on equity). The model used in this report is

$$ROE = \alpha + \beta 1\ (PM) + \beta 2\ (CT) + \beta 3\ (FC) + \beta 4\ (FS) + \beta 5\ (TE) + \varepsilon$$

where, β 1, β 2, β 3, β 4 & β 5 are standardised coefficients, α is constant and ε is estimate error.

The regression analysis have been computed using the data of 21 Indian IT software companies for the years 2013 and 2009 to illustrate the change in effect of the above mentioned variables during the period of recession. The data used for computation of the ratios has been taken from balance sheet and profit and loss statement available on the capitaline website.

Table 8: Regression Analysis– IT Sector

Regression Analysis - IT Sector (2013)		
Regression Statistics		
R Square	0.819887255	
Observations	21	
ANOVA		
	Significance F	
Regression	3.91607E-05	
	Coefficients	*t Stat*
Intercept	-0.653714975	-0.47895
Profit Margin	0.852837485	6.30735
Capital Turnover	0.11912004	5.59518
Financial cost ratio	0.177836994	0.15768
Finance structure ratio	0.151493167	0.55223
Tax effect ratio	0.233706622	1.88415

Regression Analysis - IT Sector (2009)		
Regression Statistics		
R Square	0.822619118	
Observations	21	
ANOVA		
	Significance F	
Regression	3.51E-05	
	Coefficients	*t Stat*
Intercept	-0.414312883	-1.658499406
Profit Margin	0.417933146	4.83491417
Capital Turnover	0.157590514	5.322510562
Financial cost ratio	0.495161644	2.53662445
Finance structure ra	0.026401455	0.376730422
Tax effect ratio	0.169705846	0.983366706

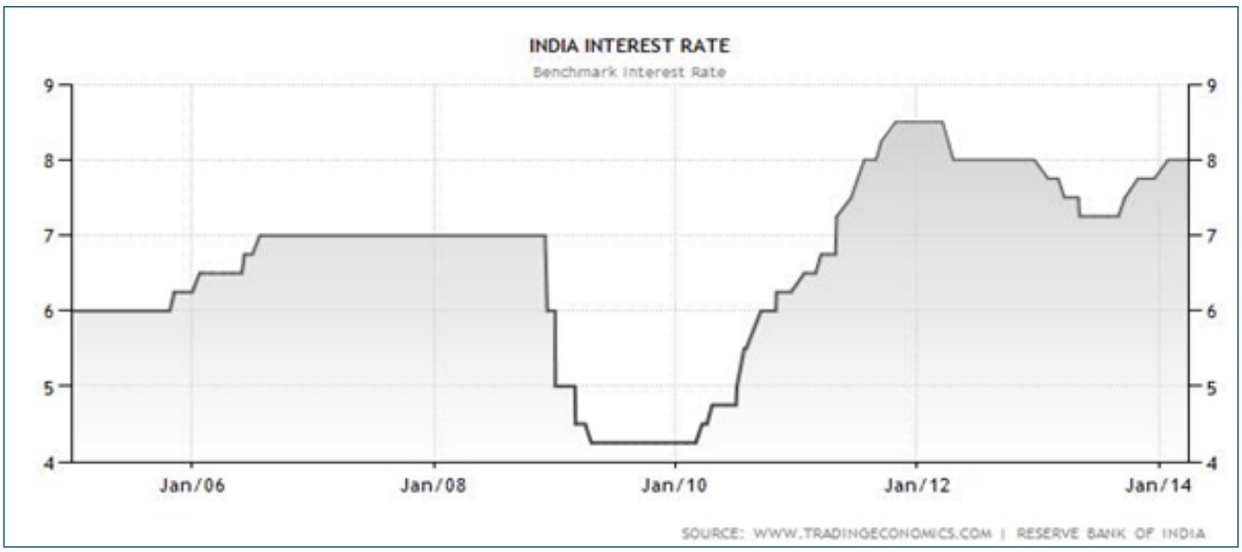
Results and Estimations

The value of R-square is 81% and 82% for the year 2013 and 2009 respectively. This shows that this model is accurate. The result shows that there is positive relationship between all variables taken into consideration. As described by the theory return on equity shows positive relationship between profit margin, capital turnover and tax effect ratio. Increase in sales, profit or decrease in tax percentage directly leads to increase in ROE.

Net Financial Leverage Factor

The net effect of finance structure ratio and finance cost ratioon ROE is a function of relative change of the two components. The net relative effect of using financial leverage is defined by net leverage multiplier, the product of two terms. As per the results of regression, there has been an increase in ROE owing to the fact that the relative reduction in income is less than the relative reduction in equity. As per the theoretical analysis, this happens when

Fig. 8: Interest Rates in India Over the Years



the interest rate on debt is less than the basic earning power. This holds true for the IT industry, hence these terms show a positive relationship with ROE.

Comparison of Years 2013 and 2009

Financial Cost Ratio Coefficient

IT had a significantly high contribution in 2009 which became normal during 2013. It can be explained as, for every 1 unit increase of financial cost in 2013, ROE increased by 0.177 whereas during 2009 the increase in ROE was 0.49 for equivalent increase in financial cost. This can be justified by the falling interest rates during 2009-10. Low interest rates during 2009 explain the coefficient being higher in 2009 as compared to 2013.

Profit Margin Coefficient

Profit margin had a very significant contribution during 2013, much higher than IT had during recession phase of 2009. As explained in the time series analysis of profit margin, opening up of opportunities and rapid growth of IT industry after the lag phase of recession has been the major factor for the change in contribution of profit margin to ROE.

Financial Structure Ratio Coefficient

The ratio had significantly less contribution to ROE during 2009. However, rise of debt has been witnessed post 2009 leading to an increase in ratio coefficient in 2013.

Capital turnover ratio and tax effect ratio coefficient have more or less remained contact over the considered two years.

Comparison of Indian IT Sector with US IT Sector using Regression Analysis

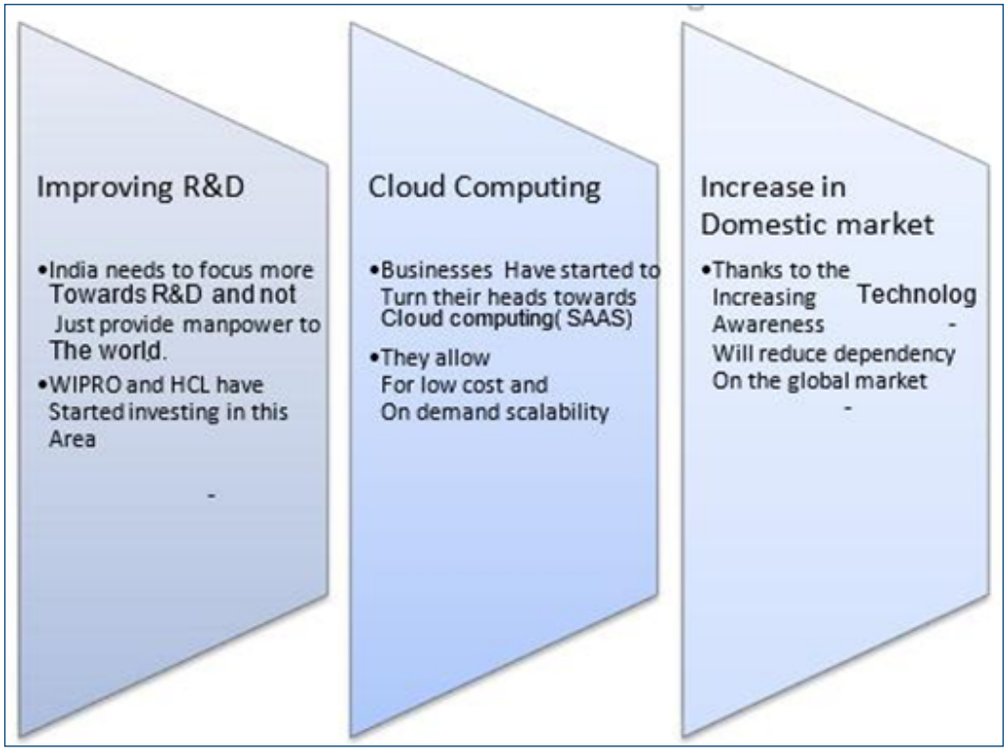
Capital turnover ratio has a significant impact to ROE of USA companies. For every 1 unit change in capital turnover ratio, return on equity changes by 0.38 whereas the change incurred in Indian companies is 0.119.

Also, financial structure ratio coefficient of USA companies is less than that of companies of India, demonstrating IT plays a less role in ROE. This can be justified by the debt to equity structure differences between the companies of two areas.

Table 9: Regression Analysis Comparison between India and USA

Regression Analysis - USA Companies(2013)		Indian Companies(2013)
	Regression Statistics	
R Square	0.76522364	0.819887255
Observations	13	21
ANOVA		
	Significance F	*Significance F*
Regression	1.96E-04	3.92E-05
	Coefficients	*Coefficients*
Intercept	-0.735602741	-0.653714975
Profit Margin	0.806216697	0.852837485
Capital Turnove	0.3838689	0.11912004
Financial cost r	0.146266094	0.177836994
Finance structur	0.081109216	0.151493167
Tax effect Ratio	0.284557551	0.233706662

Fig. 9: Scope of IT industry in India



Future Scope of IT Industry

India enjoys a market share of 52% in the global software industry and has a scope of further profit. The year 2012 was turbulent for the Indian software companies. Nevertheless, the software giants reclaimed their territories in 2013 and have started tracing a trajectory of higher growth rate of 12% as compared to 10% last year. This happened with an increase in the global technology spending, thus creating opportunities for the sector to grow.

Conclusion

The DuPont model has come a long way from being a simple two step ratio analysis to the one that can extensively explain the ups and downs of an industry in terms of its turnover, financial leverage, and tax effects. This has been made possible by the breaking up of the original components into five meaningful factors that can help to determine the overall return on equity of the firm and frame its strategies.

In the paper, major companies of the software industry were covered and their corresponding analysis done extensively. The IT sector was picked up as IT is a large contributor to the GDP of India and its study can help in the formulation of various policies and strategies.

The Time series analysis of the Indian IT sector threw light on the impact of various factors on the DuPont ratios and their corresponding impact on the overall financial performance of the software companies. A significant observation was how the global economic depression of 2007-2010 guided the profit margin, return on equity, and financial structure ratios of the industry greatly and the way companies formulated their strategies to overcome it.

The cross sectional analysis was useful in comparing the DuPont ratios of the companies with the industry average and understand why a certain firm was making loss and while the others were fairly well and vice versa.

Through the regression tools, it was derived that the return on equity (ROE) has a positive dependence on the five DuPont ratios. Here clearly the dependent variable taken was ROE and the independent ones were its five components. It can be observed that the profit margin has a larger impact on the ROE recently (coefficient being 0.85 in 2013) than it was during the recession period (coefficient being 0.4 during 2008-2009). During this

period, the financial structure ratio had a greater because the companies had started becoming more leveraged than and thus the net ROE depended greatly on the debts, i.e the financial structure of the firms.

Furthermore, the comparisons drawn between the Indian and US firms using regression helped us understand how the five parameters impacted the ROE differently across the two nations.

Thus it can be clearly observed that the five point Du Pont method was significantly useful for examining the trends of the Information Technology sector in India. It is a relatively new area of study and can be extensively exploited in studying financial structures or business models and making recommendations for further improvements.

References

Annual Report 2012-13, NASSCON

Bhatnagar, S. (2006). India's Software Industry, Technology, Adaption and Exports: How Some Developing Countries Got IT Right, World Bank.

Boyd, G. M. (1989). Some suggestions for a "new and improved" Du Pont model. *Journal of Financial Education.*

Gitman, L. J. (2000). *Principles of financial management* (8th ed.). Addison Wesley Publishers

Hawawini, G., & Viallet, C. (2000). *Finance for Executives.* South-Western College Publishing.

Hawawini, G., & Viallet, C. (1999). *Finance for executives*, South-Western College Publishing.

Human Resource and Skill Requirements in the IT and ITES Industry Sector, National Skill Development Corporation.

Isberg, S. C. (1998). Financial analysis with the Du Pont ratio: A useful compass. *Credit & Financial Management Review.*

Liesz, T. J. (2004). Really modified DuPont analysis: Five ways to improve return on equity. *The Economics review*, 81(3), 231-243.

Malik, P. (2011). An Overview of the Indian Software Industry Profile, R&D features and Future.*Global ICT Industry: Changing Landscape?- The future of Europeon ICT R&D.*

Swapna, N., & Sujata, N. Trends of IT Industry in Indian Economy – An Analysis. *International Journal of Computer Science & Informatics* (IJCSI), ISSN (PRINT), 2231–5292, 2(1), 196-198

Article can be accessed online at http://www.publishingindia.com

Analytics for Decision Making at Ports

Mrinal Kumar Dasgupta*

Abstract

Ports serve as an important link in global supply chain. Worldwide more than 75 percent of cargo move by sea. Over the years, the Indian Union has endeavoured to invest on major ports of the country to meet up to the global standards. Yet the share of major ports under the government of India has decrease from 90 to 70 percentage of total sea borne cargo in the country. The major ports lost its share to the minor ports under the state governments. Two reasons could be hypothesized for the said problem. One, the investments are not made in the right direction and other that the efficiency needs to be improved in functioning of the ports. In this paper an attempt has been made to identify the dimensions of port performance and the causality between the dimensions. It chooses to take average turn round time (ATRT) as an indicator of port performance. The paper proposes an analytical framework to identify the causality that would aid the decision makers. The causal approach has been based on identifying the dimensions (factors) using multi-variate data analysis, establishing the linear causal association between the ATRT and the factors, analyzing the relationship so obtained to propose an System Dynamics model for policy simulation by the decision makers.

Keywords: Port Performance Indicators, Average Turn Round Time, Causality, Analytical Framework, Multi-Variate Data Analysis, System Dynamics

Introduction

Ports serve as an important link in global supply chain. Logistics and supply chain are concerned with physical and information flows and storage from raw materials through to the final distribution of the finished products. Logistics concerns the efficient transfer of goods from the source of supply through the place of manufacture to the point of consumption in a cost-effective way whilst providing an acceptable service to the customer. The key areas representing the major components of distribution and logistics include transport, warehousing, inventory, packaging and information. Logistics is an important activity making extensive use of the human and material resources that affect a national economy. Amstrong and Associates (2007) found that for the main European and North American economies, logistics represents between about 8 percent and 11 percent of the gross domestic product of each country. For developing countries this range is higher at around 12 percent to 21 percent – with India at about 17 percent and China at 21 percent. The substantial costs involved in logistics signifies the importance of understanding the nature of logistics costs and identifying means of keeping these costs to a minimum.

The breakdown of the costs of the different elements within logistics has also been addressed in various surveys. One survey of US logistics costs undertaken by Establish/ Herbert Davis (2008) indicated that transport was the most important element at 50 percent, followed by inventory carrying costs (20 percent), storage/ warehousing (20 percent), customer service/ order entry (7 percent) and administration (3 percent). The survey also produced a pan-European cost breakdown which placed transport at about 40 percent, warehousing at about 32 percent, inventory carrying cost at about 18 percent, customer service/ order entry at about 5 percent and administration at about 5 percent. In both studies the transport cost element of distribution was the major constituent part.

In a global context, more products are moved far greater distance because of the concentration of production facilities in low-cost manufacturing locations and because companies have developed concepts such as focus factories, some with a single global manufacturing point for certain products. Long-distance modes of transport have thus become much more important to the development of efficient logistics operations that have global perspective. The broad approach for selecting the suitable mode of transport is split into four stages covering operational

* Research Officer, Indian Maritime University – Kolkata Campus, Email : mrinalsa@rediffmail.com

factors, transport mode characteristics, consignment factors and cost and service requirements. A modal choice matrix for international logistics as observed by Rushton et al (2010) is given below.

Table 1: Modal Choice Matrix

Size of order/ load	*Short*	*Medium*	*Long*	*Very Long*
100T	Road	Road/Rail	Rail/Sea	Sea
20T	Road	Road	Road/Rail	Rail/Sea
Pallet	Road	Road	Road/Rail	Air/Sea
Parcel	Post/ Road	Post/Road/ Air	Post/road/ air	Post/Air

Source : The Handbook of Logistics & Distribution Management (2010)

For goods traded globally maritime transport plays an important role as seventy percent of goods traded globally by volume are transported using sea routes. Access to a global network of reliable, efficient and cost-effective maritime transport services thus often becomes a necessary condition in today's highly competitive global scenario. A well connected maritime transport sector ensures an efficient and cost effective supply chain and implies vast positive spillovers in terms of increased trade.

Ports are essentially points at which sea-borne cargo is transferred from one mode of transport to another. Ports play a vital part in inter-modal transport networks (sea, road, and rail). In order to make whole transport operation smooth, both the inward transport to the port and outward transport from the port must be speedy and efficient. In addition, the handling of cargo i.e., discharge from ship to wharf, movement from wharf to stack yard and from stack yard to lorry or railway wagon, in case of imports, and discharge from lorry/railway wagon, movement to stack yard and loading to the ship, in case of exports, must be efficient and cost effective.

In this paper literature survey has been carried out to identify different port performance indicators. The selected indicators have been subjected to factor analysis for identifying the dimensions of port performance. The relationship between the port performance measured through dimensions so obtained (from the factor analysis) and the average turn round time (ATRT) of vessels , the dependant variable, has been established through multiple regression. The analysis has been done using SPSS. A causal model has been developed based on System-Dynamics approach, to study the impact of change in port performances on number of vessels and total cargo throughput.

Literature Review

In the 1976's United Nations Conference on Trade and Development (UNCTAD) published a document about the port performance indicators and since then it is seen by the researchers in this area as a reference (UNCTAD, 1976). In this document there are several types of indicators to evaluate the operational and financial performance.

Kek Choo Chung (1993) indicates that the operational performance of a port is generally measured in terms of the speed with which a vessel is dispatched, the rate at which cargo is handled and the duration that cargo stays in port prior to shipment or post discharge. According to Chung a more progressive approach would also like to know how extensively and intensively the port's assets are being utilized as well as how well the operations perform financially. Indicators to measure these performances are determined generally in relation to the tonnage of ships calling at the port and of the volume of cargo handled since port services in the main are rendered to ships and cargo.

The evolution of the concept of logistics, in which the operators are classified according to its level of intervention in the supply chains and designated as Transport Service Providers (TSP) allows us to understand that the measurement of the efficiency level of these entities is not confined to quantitative aspects and proves that qualitative indicators are necessary (Antão et al., 2005).

Thai (2007), opined that if security measures and initiatives are not carefully designed and effectively implemented, they can negatively impact the whole maritime transport chain. Security improvements resulting from maritime security requirements may also bring about some benefits to the business performance for the organization. Bryan et al. (2006) concluded that port infrastructure plays an important role in supporting other welsh businesses. Farrel (2009) describes how container terminal efficiency declines as the terminal becomes more congested.

Yan and Liu (2010) indicated that the number of berths and the capital deployed are the most sensitive measures impacting performance of most container ports. The

analysis also reveals that container ports located in different continents behave differently. The results show that vessel turnaround time is highly correlated with crane allocation as well as the number of containers loaded and discharged. The benefits of such model include giving port operators opportunity to determine optimum crane allocation to achieve the desired turnaround time given the quantity of containers to be processed (Mokhtar & Shah , 2006).

Most of the recent studies used the methodology of data envelopment analysis (DEA) to measure port efficiencies. Martinez-Budria et al. (1999) analyzed 26 Spanish ports using DEA-BCC model and concluded that larger ports produced higher efficiencies. Tongzon (2001) analysed the efficiency of four Australian and twelve other ports using the DEA – Additive and DEA – CCR model and argued that container handling operation is the most important component of the service offered by port authorities. Tongzon (2008) pointed out that operational efficiency does not solely depend on a port's size and function. The study by Wang and Cullinane (2006) included European container terminals with annual throughput of over 10,000 TEUs from 29 countries. They concluded that most of the terminals under study showed inefficiency and that large-scale production tended to be associated with higher efficiency. Yongrok Choi (2011) used DEA and its variant models for 13 major sea ports in North East Asia including the seven largest container ports and concluded that investment in infrastructure does not improve efficiency, rather self created logistics demand and strategic alliances do improve the efficiency. Chudasama (2009) identifies the efficient and inefficient major ports of India and discovers the sources of inefficiency for the inefficient ports on the basis of DEA. Lee et al. proposed a new procedure based on DEA (Data Envelopment Analysis) called RDEA (Recursive Data Envelopment Analysis) and applied it to rank 16 international container ports in Asia Pacific region in terms of operational efficiency.

Bhatt and Gaur (2011) , concluded that after privatisation of the container terminals the performance of the terminals was relatively closely matched. The competition of securing the cargo had led to matching efficiencies on quay side where ships turnaround times and client satisfaction are closely related. However, they found that yard side efficiencies in evacuation of cargo were suffering major differences.

Blonigen and Wilson (2006) developed and applied a straightforward approach to estimate port efficiency by using detailed data on U.S. imports and associated import costs, yielding estimates across ports, products, and time. These measures are then incorporated into a gravity trade model where they estimated that improved port efficiency significantly increases trade volumes. The study provides new measures of ocean port efficiencies through simple statistical tools using U.S. data on import flows from 1991 through 2003.

Stochastic frontier model was used by Coto, Banos and Rodriguez in 2000 to measure efficiency of Spanish Ports. Their analysis resulted in a conclusion that efficiency and size are not related and that autonomous ports are less efficient than the rest.

A similar study and methodology used by Notteboom, Coeck and Van den Broeck in 2000 to measure efficiency of 36 European container terminals which concluded that hub ports are more efficient than feeder ports and that efficiency and size relationship is a function of type of port. They observed no relationship between type of ownership of port or terminal and the efficiency level.

This observation on relationship of ownership and efficiency is further contradicted by Jose Tongzon and Wu heng (NUS, Singapore) in 2005 when they conclude that private participation in ports is useful to improving efficiency however complete privatization is not the answer to improve efficiency of a port and that the relationship is an inverted bell shape curve.

The literature survey reveals that different discrete analytical tools have been used by the researchers to draw conclusion on factors and dimensions affecting port performance and their relationships. It is felt that a well-defined analytical framework is required to define the causality amongst the factors and their variables. The causal model will enable the port managers to take the right decision.

In this paper an analytical framework has been described to arrive at a causal model for decision making by port managers.

Factor Analysis

Factor Analysis is a multi-variate technique which is used to uncover the latent structure (dimension) of a set of variables (Gorsuch, 1983, Rummel, 1970). It reduces

attribute space from a larger number of variables to a smaller number of factors and as such is a 'non-dependent' procedure. Factor analysis could be used for any of the following purposes (Hair et al,1998) :

- To reduce a large number of variables to a smaller number of factors for modeling purposes, where the large number of variables precludes modeling all the measures individually.
- To select a subset of variables from a larger set based on which original variables have the highest correlations with the principal component factors.
- To create a set of factors to be treated as uncorrelated variables as one approach to handling multicollinearity in such procedures as multiple regression.
- To validate a scale or index by demonstrating that its constituent items load on the same factor, and to drop proposed scale items which cross-load on more than one factor.
- To establish that multiple tests measure the same factor, thereby giving justification for administering fewer tests.
- To identify clusters of cases and/or outliers.
- To determine network groups by determining which sets of people cluster together (using Q-mode factor analysis).

Factor Analysis is an interdependence technique, whose primary purpose is to define the underlying structure among variables in the analysis (Hair et al 2007). While using multivariate analysis, the number of variables increases. Univariate techniques are limited to a single variable, but multivariate techniques can have tens, hundreds or even thousands of variables. If we have few variables, they may all be distinct and different. When we add more and more variables, more and more overlaps i.e., correlation is likely among the variables. As the variables become correlated, the problem becomes how to manage these variables – grouping highly correlated variables together, labeling or naming the groups, and perhaps even creating a new composite measure that can represent each group of variables. Factor analysis provides the tools for analyzing the structure of the interrelationships among a large number of variables by defining sets of variables that are highly interrelated, known as factors. These groups of variables (factors), that are by definition highly inter-correlated, are assumed to represent dimensions within data. If one is only concerned with reducing the number of variables, then dimensions can guide in creating new composite measures. On the other hand, if one has a conceptual basis for understanding the relationships among variables, then the dimensions may actually have meaning for what they collectively represent. In the latter case, these dimensions may correspond to concepts that cannot be adequately described by a single measure. Factor analysis presents several ways of representing these groups of variables for use in other multivariate techniques.

Factor analysis can be either exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). The purpose of exploratory factor analysis is to identify the factor structure or model for a set of variables. This often involves determining how many factors exist, as well as the pattern of the factor loadings. EFA is generally considered to be more of theory generating than a theory testing procedure. In contrast, CFA is based on strong theoretical and/or empirical foundation that allows the researchers to specify an exact factor model in advance. This model usually specifies which variables will load on which factors as well as such things as which factors are correlated. It is more a theory testing procedure than EFA. The factors are the latent (unobserved), hypothetical, underlying concepts deduced from the correlations between the measured variables of the instrument or test.

Systems Dynamics

System Dynamics, developed by Forrester (1961, 1968) identifies cause-effect relationships and structures them in a feedback control framework to understand the dynamic behavior of the systems. The approach professes causality doctrine associated with determinism. System Dynamics is a methodology that has ability to capture and model dynamic complexity of complex systems. Dynamic complexity refers to state where cause and effect are subtle and where effects over time interventions are not obvious (Senge, 1990).

Causal thinking is the key to organizing ideas in a system dynamics study. Instead of 'cause', 'affect' or 'influence' can be used to describe the related components in the system. Some influences can be logically deduced such as food intake increases weight or say if there is smoke there is fire or say use of seatbelts reduces highway fatalities. However unidirectional linear causality is not enough to describe the dynamics of a system. It requires identifying feedback loops that govern the dynamics of a system. A feedback can be said "that an initial cause ripples through a chain of causation ultimately to re-affect itself". Thus

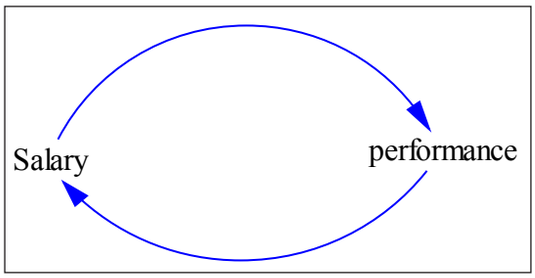
one key element of System Dynamics is to identify closed, causal feedback loops. The most important causal influences will be exactly those that are enclosed within feedback loop.

Causal loop diagrams represent the feedback structure of systems. It captures the causes of dynamics. For example, we know the better salary leads better performance while better performance also results in higher salary. That is, in "Salary vs Performance" dynamics it can be represented as:

- Salary → Performance
- Performance → Salary

The causal loop diagram can be shown as given in figure 1a below.

Figure 1a: Salary – Performance Loop

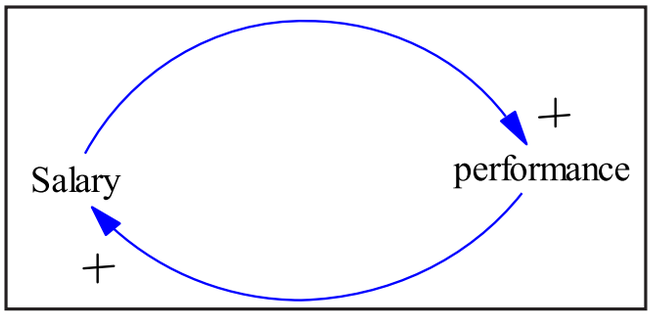


Adding a '+' or a '–' sign at each arrowhead conveys more information. For example;

- A '+' sign is used *if the cause increase, the effect increases* and *if the cause decrease, the effect decreases*
- A '-' sign is used *if the cause increases, the effect decreases* and *if the cause decreases, the effect increases*

Since it is established that salary leads to better performance and better performance leads to higher salary, the above diagram can be shown as (Figure 1b).

Figure 1b: Salary – Performance Loop with signs



The signs establish the polarity of the loop. The polarity results in positive or negative feedback loop. Positive feedback loops have the following characteristics:

- Have an even number of '–' signs
- Some quantity increase, a "snowball" effect takes over, and that quantity continues to increase
- The "snowball" effect can also work in reverse
- Generate behaviors of growth, amplify, deviation, and reinforce

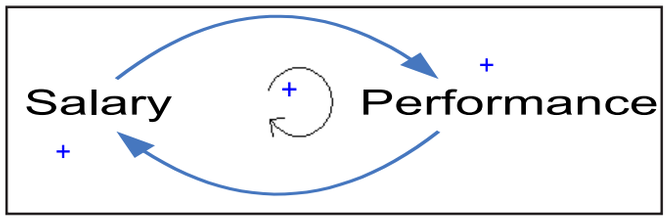
The notation used is to place ⊕ symbol in the center of the loop

Negative feedback loops have the following characteristics:

- Have an odd number of "–" signs
- Tend to produce "stable", "balance", "equilibrium" and "goal-seeking" behavior over time
- Notation: place ⊖ symbol in the center of the loop

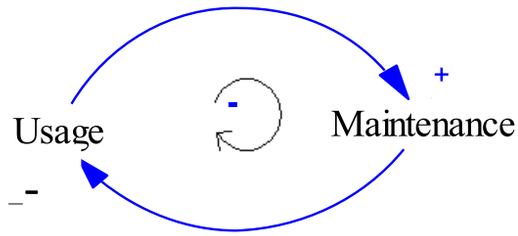
Thus the causal loop diagram for salary – performance dynamics would be as shown in figure 1c below:

Figure 1c: Salary – Performance Positive Loop



There are systems which have more than one feedback loop within them. A particular loop in a system of more than one loop is most responsible for the overall behavior of that system. The dominating loop might shift over time. When a feedback loop is within another, one loop must dominate. A stable condition will exist when negative loops dominate positive loops. An example of negative loop is the example of "usage versus maintenance". More we use a machine more is the maintenance required for that machine. More we maintain the machine less it is available for usage. Figure 1d shows the negative loop arising out of the nature of dynamics between usage of a machine and its maintenance.

Figure 1d: Usage – Maintenance Negative Loop



The Analytical Framework

i. Identify the factors and variables associated with each factor through literature review
ii. Use Factor Analysis on data related to port performance
iii. Identify the causality between factors using multiple regressions
iv. Define the Causal Model using Systems Dynamics Approach.

Factors and Variables Affecting Port Performance – Findings from Literature Review

The review of literature reveal that factors affecting Port efficiency emerging through these studies are Size, Competition – Intra and Inter port, Technology adopted and Management/Institutional structure. These factors are again interdependent and region specific. The important findings may be drawn on comparable terminals and indicators. These are summarized below.

i. The operational s performance of a port is generally measured in terms of the speed with which a vessel is dispatched, the rate at which cargo is handled and the duration that cargo stays in port prior to shipment or post discharge (Kek Choo Chung, 1993).

   The performance parameter that indicates this aspect is the turn round time (TRT).

ii. Container terminal efficiency declines as the terminal becomes more congested (Farrel, **2009)**.

   The performance parameter that manifests congestion is the pre berthing delay (PBD).

iii. The number of berths and the capital deployed are the most sensitive measures impacting performance of most container ports (Yan and Liu, 2010).

iv. Vessel turnaround time is highly correlated with crane allocation as well as the number of containers loaded and discharged (Yan and Liu, 2010).

   The average output per ship berth day (AOPSBD) is the parameter that reflects the impact of crane and moves per crane on vessel turnaround time.

v. Variations in port efficiency are linked to excessive regulation, the prevalence of organized crime, and the general condition of the country's infrastructure (Clark et al, 2004). They found that besides distance and containerization, the efficiency of ports is also important in determining maritime transport costs.

   The pre berthing delay and post operation time prior to departure or the non-working time reflect the delay owing to regulations and other factors. These time durations are reflected in TRT.

vi. Larger ports produced higher efficiencies (Martinez-Budria et al. 1999).

   Port's size is reflected in terms of number of berths and/or cargo throughput per annum of the port.

vii. Large-scale production tended to be associated with higher efficiency (Wang and Cullinane 2006).

   Cargo throughput and vessels handled reflect scale of production for ports.

Dynamics of a Port System

The dynamics of a port system as a part of the logistics chain, arising out of interaction of different variables that describes the system, may be explained as follows.

Ship's costs at ports constitute a significant part of the maritime transport costs and thus can significantly influence the logistics costs and hence the final price of a product. The total costs incurred in port are found by adding together (1) actual port costs and (2) the cost of ship's time in port.

$$T_c = f(S_c) \quad (1)$$

$$S_c = P_c + S_{tc} \quad (2)$$

$$P_c = P_d + P_p + P_b \quad (3)$$

$$P_b = g(I_f, O_e)$$

Where,

T_c = Maritime Transport Cost

S_c = Ship's cost at port

P_c = Actual Port Cost

S_{tc} = Cost of ship's time at port

P_d = Port dues

P_p = Pilotage charges

P_b = Berth hire charges

I_f = Infrastructural facilities

O_e = Operational Efficiency level)

Infrastructural facilities (I_f) constitutes loading, unloading and shore clearing equipment. The stay at berth (that results in berth hire charges) is dependent on number of right equipment and its operational efficiency level (O_e). Operational efficiency level (O_e) determines the output per ship per day. Berth hire charges also depends on the parcel load (total cargo carried by the ship) of the ship. As the tonnage increases, stay at berth increases.

Cost of ship's time at port (S_{tc}) includes opportunity cost (P_{oc}) due to non-working time at port. A ship may have to wait due to non-availability of berths or any other resources such as tugs, or may be owing to stoppage of work, delay in clearance of documents or any other managerial issues. In whole the "Cost of ship's time at port" is function of "Turn Round Time". The Turn Round Time (TRT) is defined as the time duration from the time ship reports to the reporting point of the port and till she leaves this point. Thus, Ship's cost at port (S_c) can be defined as sum total of one-time cost payable to the port per voyage (P_{ot}) and the variable component (P_{vt}) that is proportional to ship's time at port (TRT). P_{ot} includes Port dues (P_d) and Pilotage charges (P_p). Hence we redefine equation (2) as:

$$S_c = P_{ot} + P_{vt} \quad (4)$$

Where,

$$P_{vt} = P_b + P_{oc} \quad (5)$$

Thus, we can conclude that the "Turn Round Time" is a function of parcel load, infrastructural facilities, and Operational efficiency level)

Average charges calculated on the basis of shipping rates provided by the Maersk Sealand for the year 2006 for import of a container vessel in India reveals that almost 25% of the total freight charges are collected by the terminal or port operators (of which around 13% at destination i.e. at Indian ports) for various port related activities. Thus, an increase in the efficiency level of the ports may have a significant effect on the logistics costs. Besides, an increase in the efficiency level of the ports may also increase the overall efficiency level of the supply chain.

Case of Port Sector in India

The study has been done based on the data collected on containerized cargo handled by 12 major ports of India, a brief description of which is given below.

Port Sector in India

India accounts for 7517 km of coastal line spread over 13 states and Union Territories. There are 12 major ports and about 200 non-major ports. Among the non-major ports only 60 per cent are functioning actively. The thirteen major ports are Kolkata Port Trust, Paradeep Port Trust, Visakhapatnam Port Trust, Ennore, Chennai Port Trust, Tutucorin Port Trust, Cochin Port Trust, New Mangalore Port Trust, Mormugao Port Trust, Jawaharlal Nehru Port Trust, Mumbai Port Trust, and Kandla Port Trust. In India, about 95 % of cargo by volume and 70 % in terms of value are transported by sea. Cargo handled by Indian Ports increased from 21.30 million tonnes in the year 1950-51 to 849.88 million tonnes in 2009-10 at a compound annual growth rate of 6.45% . During the last 9 years (2000-01 to 2009-10) it has registered a compound annual growth rate of 9.75%. Table 2 shows below the trend of cargo traffic handled by the Indian ports with a breakup of its share among Major ports of India (controlled by Central Government) and non-major ports comprising ports controlled by State Maritime Boards which also include private ports. Table 3 shows the compound annual growth rates of the Major and non-major ports for various time periods. It is evident from the Tables that both the growth rate and the share of the non-major ports increased drastically after 1990-91 i.e., after the liberalization process started.

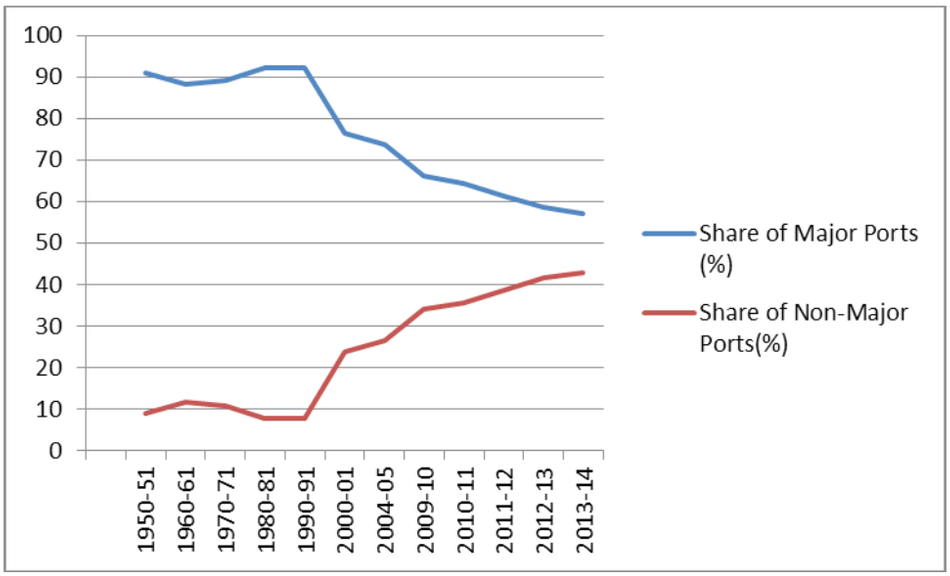
Figure 2 below shows the drop in share of major ports from around ninety percent in the year 1950-51 to less than sixty percent in the year 2013-14. This drop is

Table 2: Cargo handled by Major and Non-Major Ports in India

Year	*Major Ports (MT)*	*Non-Major Ports (MT)*	*Total (MT)*	*Share of Major Ports (%)*	*Share of Non-Major Ports(%)*
1950-51	19.38	1.92	21.30	90.99	9.01
1960-61	33.12	4.41	37.53	88.25	11.75
1970-71	55.58	6.69	62.27	89.26	10.74
1980-81	80.27	6.73	87.00	92.26	7.74
1990-91	151.67	12.78	164.45	92.23	7.77
2000-01	281.13	87.25	368.38	76.32	23.68
2004-05	383.62	137.83	521.45	73.57	26.43
2009-10	561.09	288.79	849.88	66.02	33.98
2010-11	570.03	314.85	884.88	64.42	35.58
2011-12	560.14	353.19	913.33	61.33	38.67
2012-13	545.79	387.87	933.66	58.46	41.54
2013-14	555.49	417.12	972.61	57.11	42.89

Source : Various issues of Major Port of India : A Profile, IPA

Figure 2: Share of Major Ports Vis-A-Vis non-Major Ports of India



inspite the fact that the liberisation process started in early nineties in India.

Figure 3 shows the rise in cargo handling by ports of India. It shows an increasing trend and India also poised to be one of the significant economies is the world.

The earning from international trade contributing to the GDP (gross domestic product) of the country is greatly dependent on the port performance.

Table 3 shows the comparison of compounded annual growth rate of cargo handled by major Indian ports pre and post liberalization period. The growth has not been significant in the post liberalization with respect to the pre-liberalization period.

Although the traffic through Indian Ports has registered a significant growth, that cannot be said for the efficiency level of most of the Indian ports. The reform process adopted by the port sector of India started more than a decade ago, without having much effect on the status of the Indian ports in the world map of ports. There is no sign of Indian ports being closer to the regional ports of Singapore or Colombo or Hong Kong or Port of Shanghai, in terms of cargo handling and efficiency. This is evident

Figure 3: Trend of Cargo handled by the Indian Ports since 1950-51

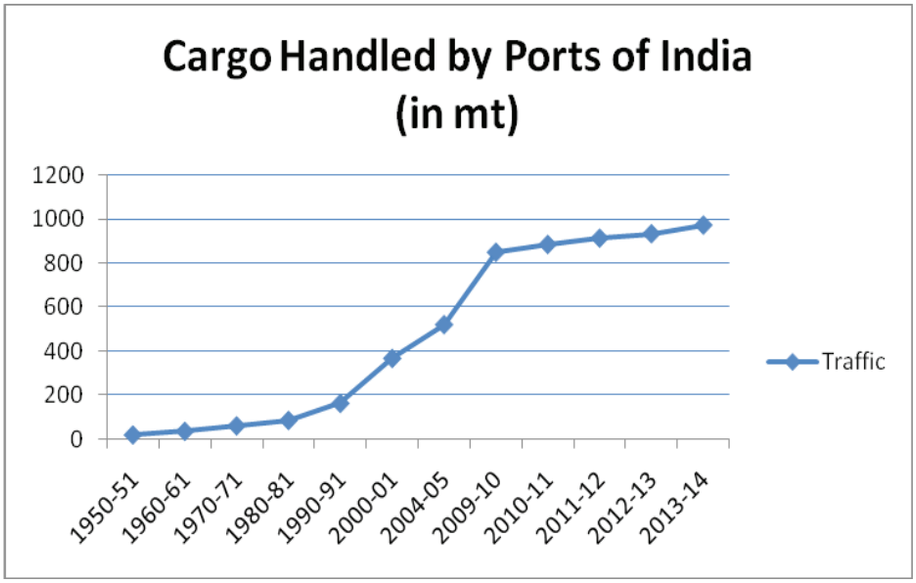


Table 3 : Compound Annual Growth Rate of Cargo handled by Indian Ports

Year	*Major Ports*	*Non Major Ports*	*All Ports*
Pre liberalization Era			
1950-51 to 1960-61	5.51	8.67	5.83
1960-61 to 1970-71	5.31	4.26	5.19
1970-71 to 1980-81	3.74	0.06	3.40
1980-81 to 1990-91	6.57	6.62	6.57
Post Liberalization Era			
1990-91 to 2000-01	6.35	21.82	8.40
2000-01 to 2010-11	7.32	13.69	9.16
20001-02 to 2013-14	5.38	12.79	7.75
Overall			
1950-51 to 2013-14	5.47	8.92	6.25

Source : Various issues of Major Port of India : A Profile, IPA

from Table 4 and Table 5. Table 4 shows port-wise container traffic in various major ports of India for the year 2012-13 whereas Table 5 shows container handled by top 10 container handling ports of the world in the year 2012. As it is seen, the total container handled by all major ports of India (7.704 million tonnes) is far less than the container handled by the port positioned at number 10 in the world. As a consequence, most of the Indian ports are still being visited mostly by the feeder vessels. This involves a longer time for the entire supply chain and in turn has its effect on overall transportation costs and trade cost for the shippers.

Table 4 : Major Ports of India - Port Wise Container Traffic (2012-13)

Port	*Container Traffic (In ,000 TEUs)*
Kolkata Dock System	463
Haldia Dock Complex	137
Paradip Port Trust	13
Visakhapatnam Port Trust	247
Chennai Port Trust	1540
Tuticorin Port Trust	476
Cochin Port Trust	335
New Mangalore Port Trust	48
Mormagao Port Trust	20

Port	Container Traffic (In ,000 TEUs)
Mumbai Port Trust	48
Jawaharlal Nehru Port Trust	4259
Kandla Port Trust	118
TOTAL	7704

Source : Major Ports of India – A Profile: 2012-2013, IPA

Table 5 : Top 10 Container Handling Ports of the World 2012

(In million TEUs)

Rank	Port	2011	2012
1	Shanghai	31.74	32.53
2	Singapore	29.94	31.65
3	Hong Kong	24.38	23.10
4	Shenzhen	22.57	22.94
5	Busan	16.18	17.04
6	Ningbo-Zhoushan	14.72	16.83
7	Guangzhou	14.42	14.74
8	Qingdao	13.02	14.50
9	Dubai	13.00	13.30
10	Tianjin	11.59	12.30

Source : Containerization International, 2012

Identifying Dimensions of Port Performance using Factor Analysis

Secondary Data for the study was collected, collated and compiled for the purpose of identifying the factors and the association, relationship and causation, if any, among the variables. At this stage data analysis was carried out by taking data from all individual major ports for the period 1990-91 to 2009-10. The data were collected and compiled from the publication of port data (titled Major Ports of India : A Profile) by Indian Ports Association, an apex body of all major ports of India; and the data by the major ports in their Annual Administration Reports.

In this study the analysis has been done in two parts. First all the operational indicators of performance of container handling (ship side) activities (except ATRT) of Major Ports of India are taken for factor analysis. Then a multiple regression has been done taking Average Turn Round Time of Vessels as the dependant variable and the Factors obtained from the factor analysis as the independent variables. All the analysis has been carried out using SPSS. At the next stage the system dynamics models have been built up.

Variables considered are listed below.

1. Container Traffic measured in '000 TEUs Terminal wise container traffic for the year 2012-13, unit, 000 TEUs. (Where available private terminal data were separately treated).
2. VTRAFFIC : Container Vessel Traffic in number
3. ATRT: Average Turn Round Time (ATRT) unit in days. It includes total time needed by a ship for entry pilotage, Pre berthing waiting time, stay at berth and exit time.
4. APBT :Average Pre Berthing Time in hours
5. APS : Average Parcel Size in tonnes
6. AOPSBD : Average Output per Ship Berth Day in tones
7. PNWTSP : Percentage of Non Working Time to Total time at Port
8. Draft: Refers to the available depth of water in the port. It determines the size of the ship that can visit the port. Measuring unit metre.
9. CRANES : Number of cranes handling containers (including both Quay and Yard cranes)
10. BERTHS : Number of Berths handling containers
11. CAPACITY :Container handling capacity of the terminal

RESULTS

7.1.1 The data set Factor Analysis carried out through Rotated factor matrix (varimax rotation) indicates 2 (two) factors with eigenvalue more than one (table 6). Factor 1 is loaded with Traffic, Vessel Traffic, Average Parcel Size (APS), Average Output per Ship Berthday (AOPSBD) and No. of Cranes , No. of Berths and Capacity whereas Factor 2 is loaded with Average Pre Berthing Time (APBT), Draft(negative), and Percentage of non-working time to total time of ship at Port (PNWTSP). As discussed with the experts the first factor is taken as a measure of the Capacity Dimension (CD) whereas the second factor is taken as a measure of the operational Inefficiency Level (IL) of the port/ terminal. KMO and Bartlett's Test of Sphericity are measures of sampling adequacy. The KMO obtained in this analysis is .814, i.e. over the acceptable value of 0.6. Bartlett's Test of Sphericity relates to the significance of the study and indicating the validity and suitability of the data collected to the problem being addressed through the study. The value of Bartlett's Test

Table 6: Rotated Component Matrix[a]

	Component	
	1	2
TRAFFIC	.967	.028
CAPACITY	.933	-.063
CRANE	.903	-.066
VTRAFFIC	.839	.139
AOPSBD	.822	-.343
APS	.645	-.110
BERTH	.600	.483
APBT	.223	.684
DRAFT	.231	-.571
PNWTSP	-.470	.492

Author's calculation

Extraction Method : Principal Component Analysis

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Table 7: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.814
Bartlett's Test of Sphericity	Approx. Chi-Square	1691.547
	Df	45
	Sig.	.000

of Sphericity is 0, that is, less than acceptable value of 0.05. The table 7 shows the results of the analysis.

Linear causality between ATRT and Factors

Multiple Regression has been done taking ATRT (Average Turn Round Time) as the dependent variable and CD (Capacity Dimension) and IL (Inefficiency Level) as independent variables.

The result (Table 8) shows that there is no significant relationship between the Capacity Dimension and ATRT.

However, there is a strong positive relationship between the Inefficiency Level and ATRT is given by

$$ATRT = 2.626 + 1.311\ IL \quad (1)$$

The IL dimension constitutes variables such as pre-berthing time (or in other words waiting time) and non-working time at berth. The increase in these values will lead to increase in ATRT of the vessel. In addition to this

Table 8: Regression Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.626	.067		39.294	.000
	REGR factor score 1 (CD or Capacity Dimension) for analysis 7	.120	.067	.073	1.791	.075
	REGR factor score 2 (IL or Inefficiency Level) for analysis 7	1.311	.067	.796	19.574	.000

a. Dependent Variable: ATRT

the navigable draft is negatively loaded with IL dimension meaning reduction in draft will lead to increase in ATRT. Ports with lower draft have lower scale of operation as the parcel load per vessel tends to be smaller and hence tend to be relatively inefficient compared to hub or larger ports. This corroborates with findings vi and vii enumerated in section 4 above.

The above relation confirms the basic premise that the efficiency level of the logistics chain is dependent on the efficiency level of the port. The increase in efficiency is expected to result in reduction in stay time of ships at the port, that is, decrease in ATRT.

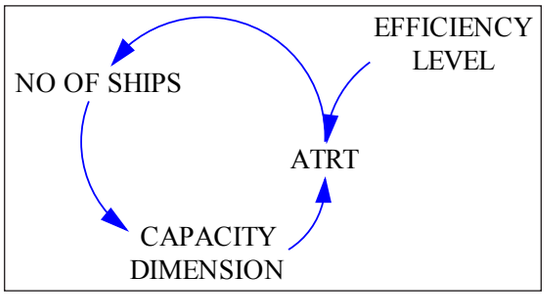
The above relationship is confined to the data set of the Indian major ports. For a different data set the Capacity Dimension and the Inefficiency Level may both or either effect ATRT. Hence, the decision maker would know the dimensions that needs attention improve the performance of the port.

Multi-dimensional Causality

The results of multi-variate analysis show a linear relationship with no circular relation. For example ATRT increases with ineffciency level or in other word it decreases with Efficiency Level. Decrease in ATRT in turn may result in increasing the attractiveness of the port causing more ships to call at the port. This sort of causility can be achieve by defining the causal model using System Dynamics as described above.

Figure 4 shows the causality between the two dimensions of port performance, namely, capacity and efficiency level with ATRT (average turn round time) and NO OF SHIPS (number of ships) calling at port

Figure 4: Loop Diagram Showing Causality between the Two Dimensions with ATRT and No. of Ships Calling at Port

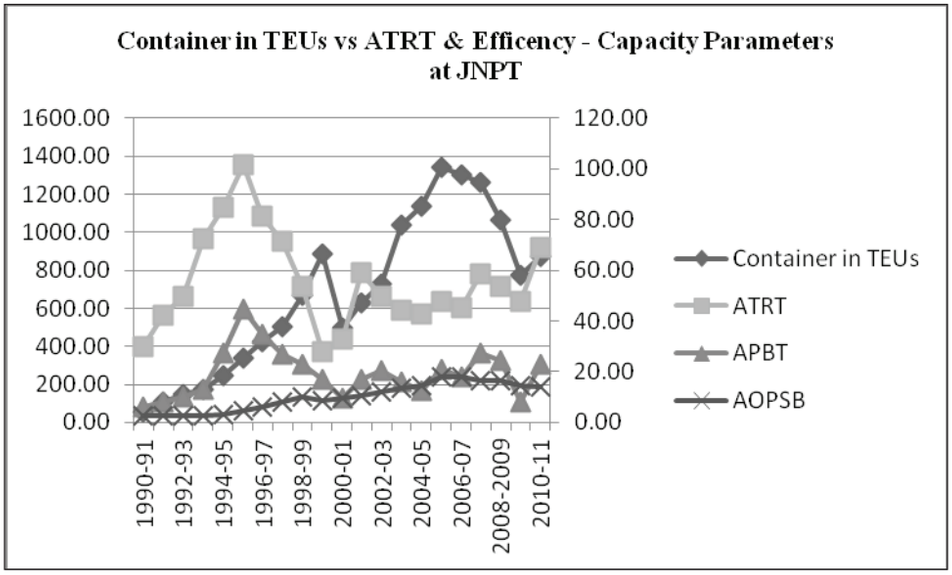


It demonstrates that influence of capacity and efficiency dimensions on ATRT in turn affects the number of ships calling at port. The behaviour of the system is governed by the negative loop. This implies that when the ATRT increases reflecting poor performance of the port, the carriers are expected to avoid that particular port. As such number of ships calling at port decreases while the reverse is expected to be observed when ATRT decreases. At the same time the number of ships calling at port would impact the capacity dimension, as its increase will tend to demand for more berths and or equipment, meaning that the number of ships calling at port causes reduction in capacity of the port, in turn affecting ATRT.

The policy implication of above model is that the decision maker has two different options to enhance its performance. One way would be to increase its capacity through increase in average ship day output (AOPSBD) (may be through modernizing or replacing the equipment,and/or training of manpower), and/or management restructuring (may be through privatisation). The other option would be enhance efficiency, may be through enhancing "soft measures" relating to cargo handling processes viz, documentation, ICT and statutary inspections, and/or increase in draft (may be through dredging, better disposal), and / or improved supervison resulting in reduction in non-working time of ships at the port.

The data analysis reveal that the inefficiency dimension has more weightage than capacity dimension, meaning that the Indian major ports need to streamline their business processes and take other measures such as efficient pilotage system to achieve better performance. These efforts should bring down the pre-berthing time and non-working time at ports, that is, reduction in values of the variables that constitutes the inefficiency level (IL) for the major Indian ports. Inaddition ports under study should endeavour create deep drafted operational points. However, the impact of effeciency level due to actions taken to improve the ATRT is not static, as improvement of ATRT is likely to increase number of ships causing capacity to fall, meaning a low performance measured through ATRT. Hence, the impact of the changed decisions has to be simulated and results monitored to achieve the corporate goal. The above model can, therefore, be described as a dynamic model based on principles of causality.

Figure 5: Container in TEUs vs ATRT & Efficiency - Capacity Parameters at JNPT



An improvement in ATRT is likely to attract more ships which cause the capacity to act as constraint, resulting in increase in ATRT unless the efficiency level increases to improve the productivity and hence stretch the capacity further. However, there is limit to enhancement of efficiency level, and subsequently may call for increase in capacity through investments.

Conclusion

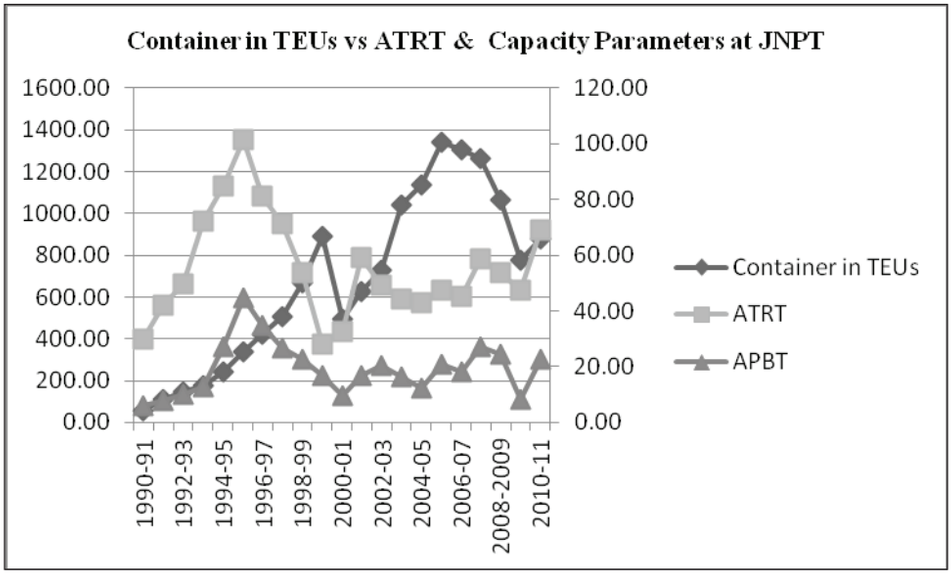
This study lays down the analytical framework for decision makers at ports enabling them to identify the dimensions, their associations, extent of influence on performance of ports and causality between them. The use of factor analysis identifies the dimensions while the beta coefficients of the dimensions, obtained from regressional analysis indicate their association with port performance indicator. The System dynamics model establishes the causality aiding in identification of effect of policies by the port decision makers.

The application of above study on major Indian ports reveal that Average Turn Round Time (ATRT) of a container vessel is dependent on one specific dimension, namely, inefficiency level (IL) of the Port. A decrease in the Inefficiency Level can significantly reduce the ATRT of a vessel and thus can help to reduce the overall transport costs in the Supply Chain. The reduction in pre-berthing time and non-working time can be effected through business process re-engineering and other soft measures. However, as the impact on efficiency caused by any action taken to reduce the ATRT is not static, a dynamic model based on causality must be used before any decision to achieve corporate goal.

The causal model identifies a stabilizing feedback loop that governs the dynamics of ships flow to a port assuming that there exists adequate demand for import and/or export cargo. The effect of efficiency parameters such as Average Output per Ship Berth-Day (AOPSB) and Average Pre-Berthing Time (ABPT) on ATRT which in turn effects the cargo flow through the ports can be observed from the figure 4 drawn on data for the period 1990-91 to 2010-2011 for container traffic (in TEUs) at JNPT (Jawaharlal Nehru Port Trust).

The ATRT in the above case decreases with increase in AOPSB and increases with increase in APBT (or the waiting time). Thus the model not only explains the causality but also identifies the limits to growth. An improvement in ATRT is likely to attract more ships which cause the capacity to act as constraint, resulting in increase in ATRT unless the efficiency level increases to improve the productivity and hence stretch the capacity further. However, there is limit to enhancement of efficiency level, and subsequently may call for increase in capacity through investments. Figure 5 shows that as container traffic grow the APBT increases meaning that increase in traffic causes the capacity to reach its upper limits resulting in waiting of ships.

Figure 6: Container in TEUs vs. ATRT & Efficiency - Capacity Parameters at JNPT



Further Scope of Research

The model proposed in the paper can be simulated using system dynamics software such as STELLA or VENSIM to see the impact of change on capacity and efficiency on ATRT affecting the number of ships calling at the port. The model can also be extended by including variables affecting the dimensions. This will aid the decision maker to carry out policy experimentation on varying internal variables and observe the behavior of the port system on being exposed to shock through variation of the exogenous variables.

References

Alderton, P. M. (1994). *Sea transport operations and economics.* Thomas Reed Publications.

Antão, P, Soares, C., & Gerretsen, A. (2005). Benchmarking analysis of European sea ports

and terminals. *Maritime Transportation and Exploitation of Ocean and Coastal Resources* (IMAM 2005), C. Guedes Soares, Y. Garbatov e N. Fonseca (Eds.), Vol. 2, Lisboa, pp. 1301-1310.

Blonigen, B. A., &. Wilson, W. W. (2006). New Measures of Port Efficiency using International Trade Data," NBER working paper 12052

Bryan J., et al., (2006). *Assessing the economic significance of port activity: Evidence from ABP operations in Industrial South Wales.*

Chung, K. C. (1993). Transportation, water and urban development department the world bank. *Port Performance Indicators*, 1(1).

Choi1, Y., & Zhang, N. (2011). Does proactive green logistics management improve business performance? A case of Chinese logistics enterprises. *African Journal of Business Management*, 5(17), 7564-7574.

Clark, X., Dollar, D., & Micco, A. (March 2004). Port Efficiency, Maritime Transport Costs And Bilateral Trade" Working Paper 10353, NBER Working Paper Series, National Bureau Of Economic Research, Cambridge

Clark, X., Dollar, D., & Micco, A. (2003). Maritime Transport Costs and Port Efficiency", LACEA Conference in Uruguay.

Coto, P., Baños, J., & Rodríguez, A. (2000). Economic efficiency in spanish ports: some empirical evidence. *Maritime Policy and Management: An International Journal of Shipping and Port Research,* 27(2), 169-174.

Cullinane, K. P. B., & Wang, T. F. (2006). The efficiency of european container Ports: A cross-sectional data enve lopment analysis. *International Journal of Logistics Research and Applications*, **9**, 19-31. Retrieved from http://dx.doi.org/10.1080/13675560500322417

De, P., & Rout, B. (September 2008). *Transportation cost and trade competitiveness: Empirical evidence from India*. RIS Discussion Paper.

Frankel, E. G. (1990). *Port planning and Development.* John Willey and Sons.

Farrel, S. (2009). *Factors influencing port efficiency : A case study of dar es salaam.*

Gorsuch, R. L. (1983). *Factor analysis (2nd ed.)*. Hillside, NJ: Lawrence Erlbaum Associates.

Hair, J.F., Jr.l Anderson, R. E.m Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis with readings (5th ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

IPA (Indian Port Association), (2001). *Guidelines for Capacity Computation,* New Delhi

IPA (Indian Port Association), (2004). *Major Ports of India - A profile: 2009 – 2010*, New Delhi

Kent, P. E., & Hochstein, A. (1998). Port reform and privatization in conditions of limited competition: The experience in Colombia, Costa Rica and Nicaragua. *Maritime Policy and Management*, 25(4), 313-333.

Lee, H. S., Chou, M. -T., & Kuo, S.-G. (2005). Evaluating port efficiency in asia pacific region with recursive data envelopment analysis. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 544 - 559,

Martinez-Budria, E., Diaz-Armas, R., Navarro-Ibanez, M., & Ravelo-Mesa, T. (1999). A study of the efficiency of Spanish port authorities using data envelopment analysis. *Int. J. Trans. Eco.,* 26(2), 237-53.

Mokhtar, K., & Shah, M. Z. (2006). A regression model for vessel turnaround time. *Tokyo Academic, Industry & Cultural Integration Tour,* 10-19

Notteboom, T., Coeck, C., & Broeck, J. V. D. (2000). Measuring and explaining the relative efficiency of container terminals by means of bayesian stochastic frontier models. *Maritime Economics and Logistics*, 2(2), 83-106

NCAER (National Council of Applied Economic Research). (2001). Questionnaire on Maritime Transport, New Delhi, 2, May .

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Rushton, A., Croucher, P., & Baker, P. (2010). *The handbook of logistics & distribution management* (5th ed.).

Ryoo, D. K., & Thanopolou, H. A. (1999). Liner alliances in the globalization era: A strategic tool for Asian container carriers. *Maritime Policy and Management*, 26(4), 349-367.

Senge, P. M. (1990). *The fifth discipline: The art & practice of the learning organization*. Newyork: Doubleday/Currency.

Sinha, D., (2005a). The dimensions of Indian port industry", paper accepted for *Indian Maritime Award, 2005, Ministry of Shipping, Roads and Highways, Department of Shipping*, Government of India.

Song, D. K., Cullinane, K., & Roe, M. S. (2001). *The Productive Efficiency of Container Terminals: An Application to Korea and the UK.* Aldershot: Ashgate.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modelling for a complex world,* Irwin Mcgraw Hill.

Sukyens, F., & Van De Voorde, E. (1998). A quarter of a century of port management in Europe : Objective and Tool. *Maritime Policy and Management*, 25(3), pp. 251-261.

TRW (Transport Research Wing), *Basic Port Statistics*, 2002-03, GOI, New Delhi.

UNCTAD. (1976). Port performance indicators', *TD/B/C.4/131/supp.1/Rev.1,United Nations Conference on Trade and Development, New York, US.*

Yan, J. W. H., & Liu, J. (2010). DEA models for identifying sensitive performance measures in container port evaluation. *Maritime Economics & Logistics,* 12, 215-236.

Zan, Y. (1999). Analysis of container port policy by the reaction of an equilibrium shipping market. *Maritime Policy & Management*, 26(4), 369-381.

Article can be accessed online at http://www.publishingindia.com

Applying Structural Equation Modeling for Green Supply Chain in Retail Domain

Aashima Sharma*, Arun Rajan, Richard Jose***, Purba Rao******

Abstract

The purpose of the research was to analyse the impact of Green Supply Chain Management (GSCM) initiatives, currently present in the retail domain in South India, on the economic performance and competitiveness of the retail companies. The GSCM concepts can be implemented in inbound, internal, and outbound stages of the retail supply chain. The research examines each of these functions separately and determines its impact on the overall performance. For this purpose, a conceptual model was developed from literature sources and data collected using a structured questionnaire circulated among mid-size retail firms in Southern India. Thereafter, a confirmatory model was tested using structural equation modeling to validate the conceptual hypothesis. The analysis identified that greening the different phases of the supply chain leads to an integrated green supply chain, which ultimately leads to competitiveness and economic performance. The research findings suggest that if the retail firms green their supply chains not only would they achieve substantial cost savings, but also enhance sales, market share, and exploit new market opportunities which lead to greater profit margins, all of which contribute to the economic performance of the firm.

Keywords: Structural Equation Modelling, Economic Performance, Competitiveness

Introduction to Green Supply Chain

Green Supply Chain Management (GSCM) has grown in popularity in the last few years. The concept is about giving more concern to the environment and its sustainability while considering the supply chain. GSCM is applying Green principles to the various processes of the supply chain such as usage of environment friendly materials, recycling products, where ever possible, and reducing the usage of certain chemicals which have adverse effects on the environment. Few of the benefits of GSCM are cost reduction, increased efficiency, resource sustainability, public relations benefits, adapting to regulation and risk reduction, and improved quality of products. All the same, the primary objective of GSCM is greening the environment through making all operations of suppliers, distributors, waste handlers, transporters and all other business partners, as environment friendly as possible.

The economic success of some of the biggest firms has been attributed to their efficient supply chain management. Retail chains like Wal-Mart, P&G have been able to drive their profits up significantly by managing their supply chains. Firms have moved from a traditional supply chain model which involves all parties working independently, to a highly integrated system wherein all the involved parties including suppliers and customers work closely with the purchasing, production, and distribution departments.

The supply chain systems which exist in the organised retail sector in India should ideally be like a well woven partnership between the retailer (customer) and the manufacturer (supplier). If this is achieved, it will create supply chains that are efficient with minimum losses. The challenges found in organised retail supply chain in India comprises problems in product sourcing, transparency issues, lack of specialised skills, improper man power management, existence of inefficient kirana stores,

* PGPM Graduate, Great Lakes Institute of Management, Chennai, Tamil Nadu, India. Email: aashima.pgpm14c@greatlakes.edu.in
** PGPM Graduate, Great Lakes Institute of Management, Chennai Tamil Nadu, India. Email: arun.rajan88@gmail.com
*** PGPM Graduate, Great Lakes Institute of Management, Chennai Tamil Nadu, India. Email: richard.pgpm14c@greatlakes.edu.in
**** Visiting faculty at IIM Ahmedabad, Gujarat, India. Email: purbarao@gmail.com

multiple taxes, inadequate infrastructure, real estate cost, quick response, customer loyalty, high connectivity, operational cost, forecasting etc. (Rajwinder, Bhirmaya, 2011). These may be classified as strategic challenges, environmental challenges, customer challenges, and supply chain (SC) challenges (Rajwinder *et al.*, 2011).The fragmented and un-integrated nature of supply chains in India affect the availability of various essential products needed by consumers. For instance in the organised retail sector in India the availability of fresh produce (vegetables and fruits) is very small. This is so for the nature of supply chain is very unorganised and fragmented. This shows the important role supply chains play in the organised retail sector in India. In the organised retail market in India, the role of supply chain is also very important because the Indian customer demands at affordable prices a variety of product mix. The supply chain ensures to the customer that all the various product offerings are available at reasonable price, in good service condition and in the promptness in responding to ever changing tastes of the customer.

Recently, this integration of the supply chain has taken yet another dimension and the focus of supply chain management is gradually shifting from supplying at minimum cost to supplying materials sustainably. Thus it is important for the integrated supply chain not only to deliver at best prices, best quality and at best possible variety but also in a manner that the products offered are environmentally sustainable. In addition the integrated supply chain which aims to be sustainable must also make sure that in all phases of the supply chain the materials are environment friendly and the production processes are clean and green (Corbett & Kleindorfer, 2003) and operations and sustainability (Kleindorfer *et al*., 2005). In India, environmentally and socially responsive supply chains are in the early adoption stages. A number of companies are realising the need to 'go green.' Customers are the key drivers, especially if they are large well known multinational companies and they want to know how the products are made, what impact future environmental legislations will have on the products they buy. Practitioners with this knowledge and a vision suggest that there is a need for the business to be environmentally sustainable and to communicate this effort to their customers, partners and the public.

However, the question remains whether greening the supply chain will be economically profitable for the business and how much will it affect competitiveness in the market. Literature reviews have shown that adopting techniques to green the supply chain leads to improved efficiency and performance. By re-evaluating a company's supply chain, from purchasing, planning, managing the use of materials to shipping and distributing final products, savings are often identified as a benefit of implementing green policies.

Introduction of GSCM in companies is often expected to increase its performance with several benefits. These could be lowered costs, efficient production system, less wastage, sustainability of resources, positive impact on financial performance, product differentiation & competitive advantages, risk reduction and conforming to local, and international regulation (Cervera & Flores, 2012). This research was targeted at studying the impact of greening the supply chain on a company's economic performance and competitiveness.

Components of Green Supply Chain

Greening the Inbound Function

The inbound phase of supply chain in general comprises various activities or actions required before the production or processing of goods and services begin. These include sourcing and handling of raw materials or merchandise, storage & inventory control, and all activities related to suppliers and procurement of materials. As far as the retail industry is considered, inbound activities are related to the procurement of merchandise goods from the manufacturers and producers. The activities related to implementing green methods in the inbound activities of a supply chain are considered as greening the inbound function. The relationship between the greening of inbound function, along with greening the internal/production and greening the outbound phases is researched through this paper with respect to their impact on competitiveness and economic performance.

The literature reviews shows that green sourcing and encouraging green suppliers leads to internal green product, process and managerial innovations, which in turn enhance competitive advantage (Chiou, Chan, Lettice, & Chung, 2011). The study suggests that the firms should work closely with suppliers both upstream and downstream for a better environmental goals and sustainability. Also it mentions that the companies, in their

effort to green the inbound function, should try to provide adequate technical support, assistance and guidance to suppliers so that they can implement environment friendly systems and processes and to organise environmental awareness seminars or training sessions which will lead to environmental performance which can give the firms a competitive advantage in the global market and hence greater economic performance.

According to Wang, Tianand Hu (2005) suppliers serve to be one of the most important players in the supply chain in maintaining the Green Supply Chain management to remain competitive. The study conducted an empirical analysis and has empirically verified the positive relationship between the competitive advantage of the manufacturer in winning orders and his criteria in selecting suppliers. Even if one of the suppliers fails in conforming to the Green methods, GSCM will not yield required success in competitiveness and economic performance. Su-Lee (2008) study shows that environmental requirements and support of buyers were positively correlated to their suppliers' willingness to participate in GSCM initiatives.

Greening the Internal / Production Function

The internal function, which also refers to production function, relates to the greening activities done as part as the firms initiatives towards the environmental sustainability. It basically includes cleaner production, efficient waste management techniques, cost reduction, eco efficiency etc. which are directly related to the firm. If this function is made environment friendly and sustainable, it is then referred to Green operations as well. When the retail industry is concerned, the internal activities can be green design, internal product delivery system, greening in internal transports, internal packaging (using less packaging materials) & ecolabelling, store management etc.

Green design is defined by Fiksel (1996) as the systematic consideration of product and process design issues associated with environmental safety and health over the full product life cycle. This would encompass the processes during new production and process development. It includes many disciplines, including environmental risk management, occupational health and safety, product safety, pollution prevention, resource conservation, and waste management all of which can ultimately lead to environmental and economic performance.

Green operations therefore, include activities related to product manufacture/remanufacture, usage, handling, logistics, and waste management. These activities are carried out once the design has been finalised (Lund, 1984). There exists research that shows consumers in certain segments are prepared to pay more for environment friendly goods (Thompson, 2007), though many other researchers disagree. Thompson says that by implementing relatively simple steps it is possible to save between 10 and 25 percent of energy costs. This is encouraging because this would be affordable for small retailers. He stresses that retailers should focus on the energy savings by switching over to low energy devices and usage of green energy. He also suggests easy techniques which can be implemented by retailers such as green home delivery by usage of electric cars, online sales which will reduce carbon print and green travel plans etc.

There are several other techniques which can be used in retail sector like green lease as described by Sinreich (2009). He describes green lease as one that provides for the sustainable construction, operation and renovation of a property and allocates the costs, benefits and responsibilities for sustainability in a manner that facilitates achievement of the desired green results.

The paper explores if (a) inbound function directly and positively impacts internal function

Greening the Outbound Function

Outbound logistics consists of activities involving distribution of finished product, material handling after production, order fulfilment, packaging, transportation, warehousing, and retailing of finished products to the ultimate consumer. (Davison, 2008)

Currently, most products in the market come in a form of packaging that prevents the product from damage and makes the product easy to handle. The use of packaging whether, it is made of glass, metal, paper or plastic, contributes heavily to the solid waste stream (Rao & Holt, 2005). Few of the areas where greening can be introduced in outbound logistics are environment friendly packaging, taking back packaging, wherever possible, environment friendly transportation, and environment friendly waste management.

The benefits of Green Supply Chain approach to industrial packaging are significant. In addition to environmental

benefits, these also include savings in packaging, waste disposal as well as other efficiencies (Verghese & Lewis, 2007). Eco-friendly packaging offers a vast potential for the Indian manufacturing companies in terms of cost reduction and resource conservation. With a growing consumer demand, limited resource availability and high wastage rate (40%), manufacturing companies will have to focus on sustainable as well as lean & green packaging alternatives. While some progressive steps have been taken by companies in this direction, the sector is still at a very nascent stage in India. The packaging companies in India have the potential to innovate and develop sustainable practices (Dharmadhikari, 2012).

The design of a logistics network and its planning are two of the more strategic issues facing logistics managers in this function. Some of the design and management criteria that support environmental planning in this area include fewer shipments, less handling, shorter movements, more direct routes, and better space utilisation. But, each of these issues includes trade-offs among delivery time, responsiveness, quality and cost, as well as environmental performance. Warehousing and delivery packaging design are two important issues in outbound (and inbound) logistics and distribution (Wu & Dunn, 1995). Standardised reusable containers, good warehouse layouts, easy information access all cut storage and retrieval movements and save on operating costs and are environmentally sounder (Toke, Gupta & Dandekar, 2012)

The paper explores if (b) inbound function directly and positively impacts outbound function.

(c) Internal function directly and positively impacts outbound function.

Competitive Advantage

Competitive advantage is a phenomenon under which companies occupy some niche positions where their competitors cannot imitate their business strategies and they can gain significant market benefits (Porter, 1980; Porter & van der Linde, 1995). In the context of green supply chain it is expected that companies that have high environmental ethics and standards can not only avoid the troubles that come with environmental protection protests, but also improve their corporate images leading to competitive advantage (Chen *et al*., 2006). Greening of supply chain brings about significant cost savings and resource conservation. These lead to competitive advantage in terms of cost. Also companies can acquire a competitive edge with respect to competition with the support of green initiatives which will change the perception of customers towards the firm and its products. Environmental sustainability image of the company can be treated as an intangible asset which will improve the goodwill of the company. The empirical results demonstrate that corporate environmental ethics positively affects corporate competitive advantage (Chang, 2011). Also competitiveness is expected to be directly related to the economic performance of the company in the future.

Greening effects of the inbound, internal, and outbound segments of the retail supply chain can be evaluated through the increase in the competitive advantage given to the company. The correlation between the competitive advantage and the economic performance can also be revealed as part of this study.

The paper also explores if

(d) Greening inbound function directly and positively impacts competitive advantage.

(e) Greening internal/production function directly and positively impacts competitive advantage.

(f) Greening outbound function directly and positively impacts competitive advantage.

Economic Performance

For an organisation the industry perception of its economic performance and benchmarking is very essential for positioning and also identifying the 'best practices' for an organisation in an industry. Thus it becomes critical to measure the various processes in Green Supply Chain in terms of economic performance.

Measuring performance and utilising metrics have an important role to play in setting objectives, evaluating performance, and determining future courses of actions. All the same performance measurement and metrics pertaining to supply chain practices have not received adequate attention from researchers or practitioners (Gunasekaran, Patel & McGaughey, 2004). Of the few existing research pertaining to the topic one may mention the work on environmental indicators across a green supply chain which have been found to lead to

environmental performance (Rao, 2014).

To meet efficiency and market demand objectives, the output of the processes enabled by the supply chain must be measured and compared with a set of standards. In order to be properly executable, the process indicator values need to be kept within a pre-assigned limit which remains relatively constant. This will allow comparison of planned and actual indicator values, and once done, the parameter values can be influenced through certain reactive measures in order to improve the performance or re-align the monitored value to the specified value (Gunasekaran *et al.*, 2004).

Research and studies have shown that greening the various phases of supply chain contributes to improved economic and environmental performance. But greening all the phases need not necessarily impact financial performance. The extents to which the various functions in the supply chain affect the economic and financial performance need to be measured (Rao & Holt, 2005).

The paper explores whether

(g) Inbound function directly and positively impacts economic performance.

(h) Internal function directly and positively impacts economic performance.

(i) Outbound function directly and positively impacts economic performance.

(j) Competitive advantage directly and positively impacts economic performance.

The Conceptual Model and Research Question

Based on literature review as given above the conceptual framework used for the research may be depicted as follows.

Research Methodology

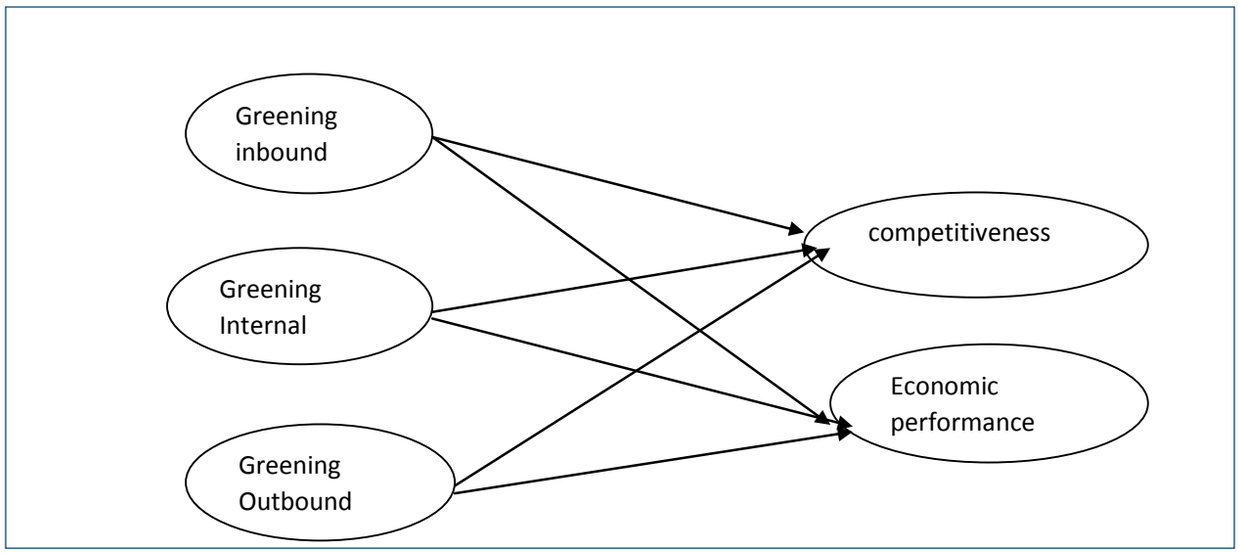
Data Sources

The data for this study are collected through primary sources. Since we don't have much empirical evidence on measuring the impact of incorporating green initiatives in supply chain, on a firm's competitiveness and economic performance in the retail sector, our reliance on secondary data is limited. Although similar studies have been conducted before, though not particularly focussing on the retail sector, we referred to those sources (research papers from EBSCO etc.) and got a fair idea of the relationship between greening and economic performance.

Instrument for Data Collection

The research instrument was a survey questionnaire to collect data for study. The questionnaire comprised

Fig 1: Conceptual Framework/Research Question for Green Supply Chain, Competitiveness and Economic Performance (Adapted from Rao & Holt, 2005)



multiple constructs, each construct having multiple questions, indicator variables within it that will measure the latent construct. The various constructs that exist in the questionnaire include inbound, internal, outbound functions of a supply chain and also questions directly measuring the economic performance and competitiveness.

Sample

The research target segment is retail outlets across South India. Excluding hypermarkets and super chains, mid-segment retail outlets were focussed. To bring diversity in the research the data collection focussed on different kinds of retail outlets like apparels, foot wears, bags, baby's products, branded stores etc.

Methodology

For data collection the researchers visited stores in Chennai and questioned people managing the supply chain i.e., the supervisors & managers of supply chain, store managers, logistics head etc. to get a perspective of the techniques being adopted by them. Once this was done, we got the questionnaires filled and used the data to perform further analysis.

The final sample size was 104, margin of error 9 %.

Structural Equation Modeling to be used for Validating Research question.

In Structural Equation Modeling approach one considers latent constructs which are looked upon as unobserved variables. These unobserved variables are constituted and measured by observed variables or indicator variables. The model comprising the linkages between unobserved variables and indicator variables associated with it is called measurement model. The model comprising linkages between unobserved variables is called structural model.

To carry out structural equation modelling one uses a conceptual model proposed by the research. This is validated or not validated by checking if the data collected supports the model or not. In the current research the proposed conceptual model was considered and evaluated by using AMOS Graphics for Windows Version 3.6 (Arbuckle , 1997) estimating the regression weight of each link (arrow)and the associated significance. This significance was evaluated with the statistic called "critical ratio" associated with the regression weight which helped to test the null hypothesis.

Ho: regression coefficient , beta= 0

If critical ratio > 1.96, one rejected the null hypothesis and considered the link to be significant If critical ratio < 1.96 , it implied that there was no statistical evidence to say that there was a cause and effect relationship between the two latent constructs concerned.

Upon running the algorithm provided a Chi square value, the degrees of freedom and the associated probability level, the p-value. The model was considered acceptable at 5% level of significance if the p-value >.05.

In addition to p - value, chi square/degrees of freedom, which is needed to be < 2.

GFI, AGFI were additional indicators to evaluate the validity of the model.

The latent constructs which were used in the model and the associated indicator variables were the same as given in the conceptual framework section. Several sets of analyses were conducted with the input being the descriptive statistics of the indicator variables and the correlation matrix for all of them. Also several sets of structural equation models were run to test variations of the model with alternate paths deleted to assess the importance of model aspects. The proposed model which would be validated have five latent constructs, each being measured by the indicator variables as given below:

Latent Construct: Greening the Inbound Function

Indicator Variables

Conducting awareness seminars for suppliers and contractors, bringing together suppliers in the same industry to share their know-how and problems, choosing suppliers based on environmental criteria, guiding suppliers to set up their own environmental programmes, urging/pressuring suppliers to take environmental actions, informing suppliers about the benefits of cleaner production and technologies etc.

Latent Construct: Greening the Internal Function

Indicator Variables

Optimisation of process to reduce solid waste and emissions, internal recycling of materials within the production phase, substitution of environmentally questionable materials, use of environment friendly raw materials, use of cleaner technology processes to save energy, water, and waste, etc.

Latent Construct: Greening the Outbound Function

Indicator Variables

Environmental improvement of packaging, taking back packaging, environment friendly waste management, use of environment friendly transportation, recovery of company's end-of-life products, etc.

Latent Construct: Economic Performance

Indicator Variables

Improved profit margin, sales enhancement, improvement in market share, new market opportunities, product price increase, etc.

Latent Construct: Competitiveness

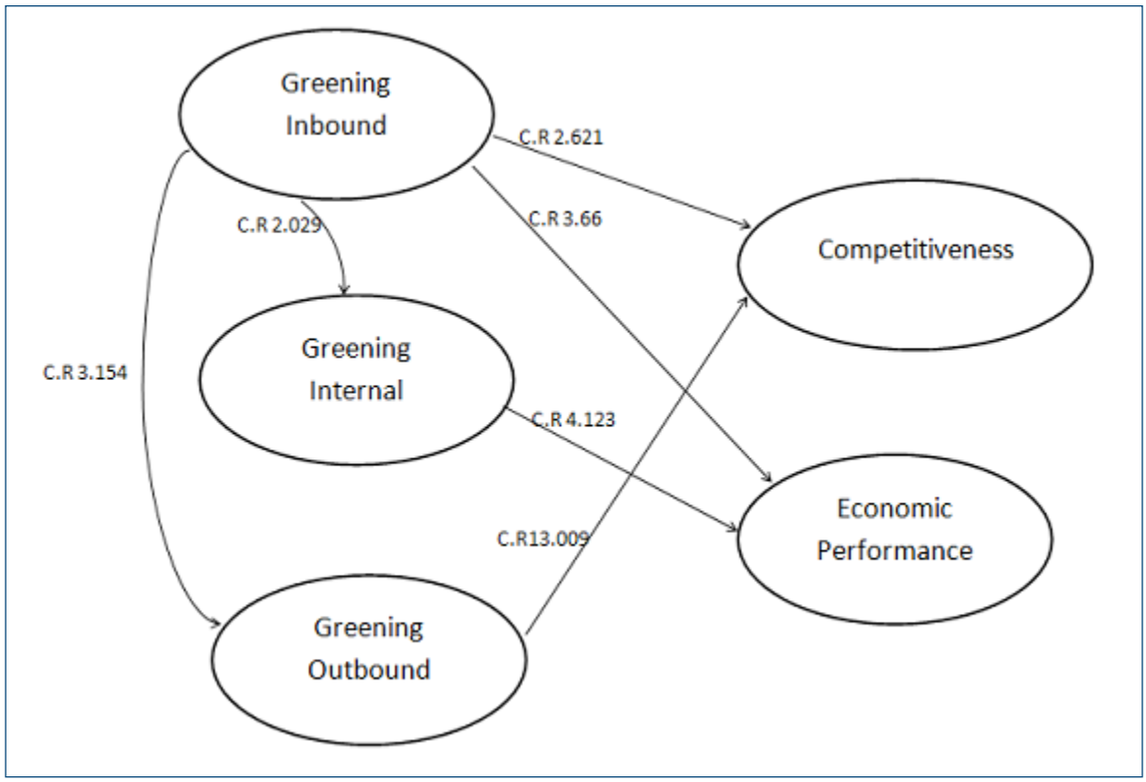
Indicator Variables

Cost savings, improved efficiency, productivity improvement, quality improvement, etc.

Results and Data Analysis

The Structural Equation Model was run with AMOS graphics software. Several runs were carried out until the overall model p-value became > .05 in fact much higher than.05. The other model indicators, GFI, AGFI, CFI etc. were all not very good but they indicated reasonably acceptable fit.

Fig. 2: Finalised Model with the Associated Statistics



Thus the overall convergence of the SEM model was considered significant, as indicated by the statistics detailed in Table 1.

Table 1: Convergence of the SEM Model

Model	*GFI*	*AGFI*	*CFI*	*P*	*CMIN/DF*
Default model	0.869	0.775	0.985	0.193	1.135
Saturated model	1		1		
Independence model	0.324	0.234	0	0	6.777

The chi-square values and associated p-value are highly acceptable indicating a good fit for the model. However, GFI and AGFI, both of which are measures that represent overall degree of fit (squared residuals from prediction compared to the actual data) are on the low side. For both of these, higher values would indicate better fit but no absolute threshold levels have been established for GFI (Rao, 2013, Hair *et al*., 1992).

This might be expected given the sample size, which makes estimation of maximum likelihood parameters not significant. Conversely, the measure of chi-square/degrees of freedom is 1.135; which falls within the recommended levels of 1.0-2.0, and indicates a significant model. One should also consider the value of CFI equal to 0.985, which is acceptable. Thus, given the limitation of a small sample size, we accept the validity of the model.

Fig. 2 gives the finalised model with the associated statistics as presented in Table 2. From the statistics as shown in Table 2, the evidence suggests that greening the different phases of the supply chain does lead to an integrated green supply chain where the green inbound function leads to a green outbound function of supply chain.

From the values of critical ratios and p values given in the table, we can conclude that greening the inbound and outbound functions leads to increased competitiveness as the critical ratios are above 1.96 (95% confidence) and p values are less than 0.05 indicating significance. However, nothing much can be said about the impact of internal function on competitiveness.

Similarly, economic performance can be improved by focusing on greening the internal and inbound functions of supply chain. We can see from the table, they have critical ratio of as high as 3.66 and 4.12 and almost negligible p values. This signifies a very strong linkage between them.

The environmental initiatives form the indicator variables and hence in order to improve the economic performance and competitiveness, the firms should focus on having integrated green supply chain functions.

- Inbound function green initiatives such as having a sustainable purchasing policy in place, choosing the right suppliers, taking initiatives to educate them on various environmental programs and have them implement systems like EMS, etc. will lead to a definite positive impact on our dependent variables and on the outbound function of supply chain.
- Internal function green initiatives such as having an environment friendly retail store design, using alternative sources of energy, clean fuel, organic packaging materials etc. can improve the economic performance of the retailer.
- Outbound function green initiatives such as using eco-friendly transportation, making reverse logistics efficient etc. aid in improving firm's competitiveness in the industry.

Table 2: Statistics of the Finalised Model

Regression Weights			*Estimate*	*S.E.*	*C.R.*	*P*
InternalFunction	<---	InboundFunction	0.376	0.185	2.029	0.042
Outbound	<---	InboundFunction	0.512	0.162	3.154	0.002
Competitiveness	<---	Outbound	0.919	0.071	13.009	***
Competitiveness	<---	InboundFunction	0.166	0.063	2.621	0.009
Competitiveness	<---	InternalFunction	-0.061	0.035	-1.751	0.08
EconomicPerformance	<---	InboundFunction	0.632	0.173	3.66	***
EconomicPerformance	<---	InternalFunction	0.422	0.102	4.123	***
Economic Performance	<---	Outbound	0.224	0.633	0.354	0.723

Discussion

This research concludes that greening the supply chain also has the same potential to lead to competitiveness and economic performance. The first contribution of the study is that it considers the green supply chain not in sections but in its entirety. Here we consider the distinct environmental initiatives in each phase as different indicator variables. The second contribution is that it establishes with empirical analysis that the green supply chain does lead to increased competitiveness and better economic performance. The third contribution is that this research has a theoretical basis and an empirical analysis where the model converged statistically with acceptable chi-square and p-values using Structural Equation Modelling.

From previous research by Klassen and McLaughlin (1996) environmental management is identified as a potential factor in the enhancement of financial performance and competitiveness of the firm.

The results obtained from the research demonstrates that the greening of inbound function and production significantly lead to greening outbound, which results in better competitiveness and economic performance of the firm.

From an industry perspective, firms are continuously striving to achieve competitiveness in their business activities in both the domestic and the global arena. These research findings suggest that if they green their supply chains not only would firms achieve substantial cost savings, but they would also enhance sales, market share, and exploit new market opportunities to lead to greater profit margins, all of which contribute to the economic performance of the firm.

The main limitation of this research is that the study was confined to the retail supply chain in South India, but the lack of empirical research in this region is also one of the main strengths of the paper. This paper will contribute to various retail firms which are planning to establish their footprint in South Indian region in terms of supply chain especially in the greening function. Future research should empirically test the relationships suggested in this paper to different parts of the country to enable comparative studies.

References

Arbuckle, J. (1997). *Amos Users' Guide Version 3.6*. Chicago IL: Smallwaters Corporation

Amemba, C. S. (n.d.). Green supply chain best practices in hospitality industry in Kenya. Retrieved from http://www.gifre.org/admin/papers/gjcmp/GREEN%20SUPPLY.pdf

Cervera, C. M., & Flores, J. L. M. (2012). *A conceptual Model for Green Supply Chain Strategy.*

Corbett, C. J., & Kleindorfer, P. R. (2003). Environmental management and operations management: introduction to the third special issue. *Production and Operations Management,* 12(3), 287-289

Chang , C.H. (2011).The influence of corporate environmental ethics on competitive advantage: The mediation role of green innovation. *Journal of Business Ethics*, 104, 361-370.

Chen, Y., Lai, S., & Wen, C. (2006). The influence of green innovation performance on corporate advantage in Taiwan. *Journal of Business Ethics,* 67, 331-339, 2006

Chiou, T.-Y., Chan, H. K., Lettice, F., & Chung, S. H. (2011). *The Influence of greening the suppliers and green innovation on environmental performance and competitive advantage in Taiwan,* 6, 822-836. Retrieved from http://www.academia.edu/788754/The_Influence_of_Greening_the_Suppliers_and_Green_Innovation_on_Environmental_Performance_and_Competitive_Advantage_in_Taiwan

Davison, E. D. (2008). Logistics Management-Research Starters Business (Online Ed.).

Dharmadhikari, S. (2012). Eco-friendly packaging in supply chain. *IUP Journal of Supply Chain Management*, 9(2), 7-18.

Fiksel, J. (1996). Achieving eco-efficiency through design for environment. *Enviromental Quality Management,* 5(4), 47-54. Retrieved fromhttp://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=ff40e4e0-83df-48e5-b40d-ff016bfe1fec%40sessionmgr12&vid=2&hid=2

Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, 87, 333-347.

Klieindorfer, P. R., Singhal, K., & Wassenhoven, L. N. V. (2005). Sustainable operation management. *Production and Operation Management,* 14(4), 482-492.

Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm Performance. *Management Science,* 42(8), 1199-1214.

Lee, S. (2008). Drivers for the participation of small and medium-sized suppliers in green supply chain initiatives. *Supply Chain Management: An International Journal*, 13(3), 185-198.

Lund, R. T. (1984). Remanufacturing. *Technology Review,* 87, 18-23.

Porter, M. E., & van der Linde, C. (1995). Green and competitive: Ending the stalemate. *Harvard Business Review,* 73, 120-134, 1995.

Rao, P. (2013). *Business Analytics, an Application Focus. PHI. New Delhi* Rao, P. (2014). Measuring environmental performance across a green supply chain, a managerial overview of environmental indicators. *Vikalpa,* 39(1).

Rao, P., & Holt, D. (2005). Do green supply chains lead to competitiveness and economic performance?*International Journal of Operations and Production Management,* 25(9), 898-916.

Rehman, M. A. A. & Shrivastava, R. L. "*An innovative approach to evaluate greensupply chain management (gscm) driversby using interpretive structural modeling (ism)*

Singh, R., Singh, H., Sandhu, B. A., & Metri, R. K. (2011). *Organizational Performance and Retail Challenges: A Structural Equation Approach.*

Sinreich, E. (2009). Greening the Retail Lease. *ICSC Research Review*, 16(2), 61-64.

Retrieved fromhttp://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=76bb4b2f-69e9-41a9-a481-bb0bcabaa5c7%40sessionmgr12&vid=2&hid=26

Singh, R., Sandhu, R. S., Metri, B. A., & Kaur, R. (2011). Organizational performance and Retail challenges: A structural eqation approach. *iBusiness,* 3, 159-168. Doi:10.4236/ib.2011.32022. Retrieved from http://www.scirp.org/journal/ib

Thompson, B. (2007). Green retail: Retailer strategies for surviving the sustainability storm. *Journal of Retail & Leisure Property,* 6(4), 281-286.

Toke, L. K., Gupta, R. C., & Dandekar, M. (2012).An empirical study of green supply chain management in Indian perspective. *International Journal of Applied Sciences and Engineering Researc*h, 1(2), 372-383.

Verghese, K., & Lewis, H. (2007). *Environmental innovation in industrial packaging: A Supply Chain Approach.*

Wang, D., Tian, Y., & Hu, Y. (2005). Empirical Study Of Supplier Selection Practices in Supply Chain Management in Manufacturing Companies. *International Journal of Innovation and Technology Management,* 2(4), 391-409.

http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=b4a2f4d1-831b-4117-bd88-51b9143ccac0%40sessionmgr10&vid=2&hid=26

Wu, H. J., & Dunn, S. C. (1995). *Environmentally responsible logistics systems. International Journal of Physical Distribution and Logistic Management,* 25(2), 20-38.

Article can be accessed online at http://www.publishingindia.com

A Study on Effectiveness of Investment in Intellectual Capital of Indian Knowledge Companies

Amitava Mondal,* Santanu Kumar Ghosh**

Abstract

Intellectual capital (IC) is non-monetary assets or resources without physical substances which are underlying factors of a firm's value creation process. Knowledge based companies mainly depend upon these type of assets for their value creation and competitive advantage. The present study makes an attempt to examine efficiency and effectiveness of investment in intellectual capital of 100 Indian knowledge companies during the period 2002 to 2011. In other words, this study examines the efficiency of intellectual capital management with regard to target level of Indian knowledge companies.

For the purpose of the study, 100 Indian knowledge-intensive companies comprising 32 software companies, 32 pharmaceuticals companies, and 36 banking and finance companies are selected on the basis of highest market capitalisation. For measuring the efficiency of intellectual capital Pulic's VAICTM (value added intellectual coefficient) is applied. This study examines, by applying partial adjustment (PAM) model, how fast the sample companies are improving the respective level of intellectual capital efficiency with respect to a target efficiency level.

The study results also indicate that the speed of achieving that target level of efficiency of sample companies is moderate. From the beta values of regression results it is also observed that IT companies are more efficient in intellectual capital management with regard to target level as compared to banks and pharmaceutical companies.

This is the first study in the IC literature that applies partial adjustment model to examine the speed of achieving target efficiency level by an individual knowledge company. However, this study confined to knowledge companies only.

Keywords: Intellectual Capital, Knowledge Company, VAICTM, Partial Adjustment Model

Introduction

Intellectual capital (IC) is non-monetary assets or resources without physical substance e.g., innovation, knowledge, research and development, employee training or customer satisfaction, which are underlying factors of a firm's value creation process (MERITUM, 2002; Lev & Zambon, 2003). The importance of IC resources in firm's value creation process has continuously increased due to the change from manufacturing-based economies to knowledge-based economies (Barth & Clinch, 1998; Kallapur & Kwan, 2004). Today, in the knowledge driven - economy IC is a key issue in strengthening a firm's competitive position and in achieving its objectives (Guthrie & Petty, 2000).

Another opinion is that the growth of knowledge economy has increased the importance of intellectual capital (Cabrita & Vaz, 2006) whereas traditional performance methods only consider attributes of physical and financial capital and lacking intellectual capital (Kujansivu, 2005). Research of different countries actively engaged in finding out methods for measuring intellectual capital. The main motivation comes from the failure of the traditional

* Assistant Professor, Department of Commerce, Kalna College, West Bengal, India. Email: a_mkolkata@yahoo.co.in
** Professor, Department of Commerce, The University of Burdwan, West Bengal, India. Email: shantanu.kaizen@gmail.com

performance measures to include the IC contributions which directly or indirectly affect the business performance. This has resulted in the development of various methods for measuring IC contribution. However, most of the methods are non-financial and non-comparable among companies in the industry. There are few non-financial measures (namely, calculated intangible value, value added intellectual coefficient (VAIC)) that produce IC value or efficiency which provides concrete basis for comparing the IC performance of different companies.

In the present study attempt has taken to measure the intellectual capital efficiency of 100 knowledge based companies and to examine the effectiveness of investment in intellectual capital of the above sample companies.

Current study is divided into following subsections. Second section provides an overview of VAIC framework. Third section discusses the advantages of VAIC framework and fourth section presents some results of empirical studies by applying VAIC framework. In fifth section the effectiveness of intellectual capital management in achieving the target efficiency level is examined. Lastly, sixth section concludes the study.

Measurement of Intellectual Capital Efficiency

Till date there is no unique, well-defined and widely accepted intellectual capital measurement model. Sveiby (2010) finds 42 different measurement models available in the literature. He classifies these models into four approaches, these are as follows;

- Market capitalisation approach,
- Direct intellectual capital measurement approach,
- Return on assets approach, and
- Scorecard approach.

Under market capitalisation approach the value of intellectual capital is determined by subtracting book value of assets from market value of the company. However, several problems associated with this approach like market value of a company varies from day to day and subject to the speculative bubbles of the stock market. Under second approach, IC values are determined by indentifying individual components of intellectual capital and assigning monetary value to them. This measurement approach suffers from subjectivity (Chan, 2009a) in indentifying intellectual components since there is no well accepted definition of intellectual capital and it varies from firm to firm. ROA approach calculates IC as the capitalisation of excess earnings of a firm over and above the industry average earnings. If a newly established knowledge firm fails to earn over and above industry above earnings then it may not be true that the firm has no intellectual capital. Methods under scorecard approach are popularly used to report intangible assets in management accounting. According to Chan (2009a) Scorecards are used to generate indicators and indices, and may not require the assignment of a financial value to the IC components.

The above four approaches do not include the popular method developed by Pulic (2000) known as value added intellectual capital coefficient (VAIC). K.H. Chan (2009a) suggested that VAIC may be viewed as fifth IC measurement approach. The main logic behind selecting it as a separate approach is that VAIC shows the value creation efficiency of company's intellectual ability by considering the all types (intellectual and physical) of resources. VAIC has widely been used in numerous empirical researches conducted worldwide. Relatively simple and quantitative approach that usage accounting information and producing efficiency indicators which are comparable among companies within the industry makes the approach popular. The procedure for calculating VAIC starts from determining the company's ability to create value added (VA). VA is the difference between sales (OUT) and inputs (IN) and can be represented in the following equations;

$$VA = OUT - IN$$

where output (OUT) represents the sales revenue that the company earned by selling all the products and services in the market in a particular time period. Inputs (IN), on the other hand, comprise all the expenses incurred in earning the above revenue except employee costs. In developing this model Pulic considered that in the knowledge economy the employee costs are not input costs rather those are investment by the company. Value added can also be calculated from the companies audited financial accounts, as follows;

$$VA = I + DP + D + T + M + R + W$$

where,

I = Interest expenses, DP= Depreciation, D= Dividends, T = Taxes Paid, Minority shareholders' interest, R= Retained earnings, W = Wages and Salaries.

VAIC™ is the sum of two indicators. These are: (i) Capital Employed Efficiency (CEE) – the indicator of VA efficiency of capital employed; and (ii) Intellectual Capital Efficiency (ICE) – the indicator of VA efficiency of company's Intellectual Capital base. Intellectual Capital Efficiency, on the other hand, is composed of (a) Human Capital Efficiency (HCE) – the indicator of VA efficiency of human capital; and (b) Structural Capital Efficiency (SCE) – the indicator of VA efficiency of structural capital.

These efficiency indicators are calculated as follows:

The relation between VA and HC (human capital) is termed as 'human capital coefficient' which shows how much VA is created by a rupee spent on employees. It actually indicates the ability of employees (HC) to create value in a company. Consistent with the views of other leading IC researchers (for example, Edvinsson, 1997; Sveiby, 2001, 2007), Pulic (1998) argues total salary and wage costs are an indicator of a firm's human capital (HC). HCE, therefore, is calculated as the ratio of total VA divided by the total salary and wages spent by the firm on it employees. Following equation shows this relationship algebraically:

HCE = VA/HC

where:

HCE = human capital efficiency coefficient;

VA = Value added and

HC = total salary and wages cost.

The second relation is 'structural capital coefficient' which indicates the contribution of structural capital in the value creation process. In Pulic's model structural capital is calculated by differentiating HC from VA. The higher the contribution of HC in the value creation, the lesser will be the structural capital (SC). Being an inverse relationship between HC and SC, the structural capital efficiency indicator is calculated as follows

SCE = SC / VA

where

SCE = structural capital efficiency coefficient;

SC = Structural capital; and

VA = Value Added.

SCE indicates the amount of SC needed to generate a rupee of VA and shows how successful in value creation process.

Since intellectual capital cannot generate value per se, physical and tangible capital must be combined with intellectual resources to create value. The third coefficient of VAIC model indicates the VA by one unit of physical capital, which is calculated as follows

CEE = VA/CE

where

CEE = capital employed efficiency coefficient; VA = Value added; and

CE = book value of the net assets.

Advantages of VAIC™

Recently, VAIC ™ method has gained popularity among the researchers to measure intellectual ability of companies. K.H. Chan (2009) supports the adoption of this technique as an effective method of measuring intellectual capital efficiency because. VAIC™ has a number of advantages listed by many researchers (Schneider, 1999; Goh, 2005; Tseng & Goo, 2005; Chan, 2009) which are summarised as below:

- It produces quantifiable, objective and quantitative measurements without requirement of any subjective grading and awarding of scores or scales.
- It generates indicators that are relevant, useful and informative to all stakeholders.
- It uses financially oriented measures therefore, calculated indicators, relations or ratios computed may be used for comparison along with traditional financial indicators commonly found in business.
- It uses very simple and straight forward procedures in the computation of the indexes and coefficients.
- It makes use of public or published financial data therefore, minimizes the question of reliability of IC efficiency.
- It can be employed to any size organisation ranging from small to large.

- It enhances the utility of traditional financial statements by incorporating measures of IC performance.
- The results are easily understood by those who have very basic understanding of accounting information.
- The commonly accepted definition of IC makes this method as the most appropriate method to measure IC performance of any organisation.

Application of VAIC™ Model

Several attempts have been taken by various authors to measure intellectual capital efficiency of business organisations and its impact on business performance are also examined. Most of them have applied Pulic's VAIC framework to measure the value creation efficiency of corporate intellectual capital and assess the linkage between intellectual capital and financial performance of companies. However, in few cases VAIC framework have also been used to examine the nexus between intellectual capital efficiency and IC disclosure and stock level of companies.

The VAIC framework is invented and applied by Pulic (2000) in Europe to examine the relation between market value added (MVA) and intellectual capital efficiency of 30 'FTSE-250' companies. He again used the model to benchmark banks operating in Croatian region (Pulic, 2002). In a similar study Bornemann (1999) examined the Croation national economy by analyzing 400 companies through VAIC. In a separate study Williams (2001) applied VAIC model to investigate intellectual capital efficiency and IC disclosure practices of 30UK companies.

In South Africa, Firer and Stainbank (2003), Firer and Williams (2003) use VAIC model to examine the association between intellectual capital efficiency and corporate financial performance of South African public companies.

Above mentioned studies are widely followed and referenced in subsequent studies by numerous authors of different countries. In India Kamath (2007, 2008), Ghosh and Mondal (2009), Pal and Soriya (2012) apply VAIC model to measure the intellectual capital efficiency of Indian companies. In Australia Clarke, Seng and Whiting (2010), Joshi, Cahill and Sidhu (2010) analysed the intellectual capital efficiency of Australian Companies by following VAIC rules. Ting and Lean (2009), Maheran and Muhammad (2009) of Malaysia used VAIC framework to examine the relationship between intellectual capital performance and financial performance of Malaysian financial institutions. In Taiwan several studies are conducted to examine the relationship between intellectual capital efficiency and business performance. Researchers like Shiu (2006), Chen, Cheng and Hwang (2005) measure the intellectual capital efficiency by following VAIC model. Tan, Plowman and Hancock (2007) of Singapore and Chan (2009) of Hong Kong also applied the said model in their studies to measure the value creation efficiency of sample companies. In Russia the model VAIC is used by Molodchik and Bykova (2011) to measure the organisational value creation efficiency of 401 Russian organisations.

Table 1 shows results of some empirical studies, where VAIC model is followed.

Table 1: Empirical Studies Where VAIC Model is Followed

Authors and year	*Country and Period*	*Research Proposition*	*Dependent Variables*	*Significant Relationships*
Firer and Williams (2003)	South Africa 2001	The study investigates whether the performance of a company's intellectual capital can explain organisational performance	ROA = Net Income less preference dividends / BV Total Assets; ATO = Total Revenue / BV Total Assets; M–B Ratio = Market Capitalisation / BV Net Assets;	VAIC only positive with ROA and negative with ATO and M/B Ratio
Firer and Williams (2003)	South Africa 2001	Investigates the association between efficiency of value added of the major components of a firm's resources and three traditional dimensions of corporate performance (ROA, ATO, M/B ratio)	ROA = Net Income less preference dividends / BV Total Assets; ATO = Total Revenue / BV Total Assets; M–B Ratio = Market Capitalisation / BV Net Assets;	HCE negative with ATO, MB SCE positive with ROA CEE positive with MB

Authors and year	*Country and Period*	*Research Proposition*	*Dependent Variables*	*Significant Relationships*
Chen et al., (2005)	Taiwan Stock Exchange 1992 – 2002	Examines the relationship between value creation efficiency of IC and firm's market valuation and financial performance	MB ratio = MV Common Stock / BV Common Stock; ROE = Pre–tax Income / Average Stockholders' Equity; ROA = Pre–tax Income / Average Total Assets; Revenue Growth = ((Current Revenue / Prior Years Revenue) – 1) X 100%; Productivity = Pre–tax income / Number of Employees.	VAIC positive with ROA, ROE, MB, GR, EP; HCE positive with ROA, ROE,MB, GR, EP; SCE positive with ROA, MB CEE positive with ROA, ROE,MB, GR, EP;
Shui (2006)	Taiwanese Listed Companies 2003 Also tests 1 year lag		ROA = Net Income / BV of Total Assets ATO = Total Revenue / BV of Total Assets M–B Ratio = Market Capitalisation / BV of Net Assets	VAIC positive with ROA, MB; HCE negative with ATO, MB CEE positive with ROA, ROE,MB, GR, EP;
Appuhami (2007)	Thailand 2005	Investigates the impact of the value creation efficiency IC on investors' capital gain on shares	Capital Gain on Shares =((Market Price per Share (PPS) – Prior Year's Market PPS) / Prior Year's Market PPS) X 100)	VAIC positive with MR; CEE negative with MR;
Chan (2009b)	Hong Kong Stock Exchange 2001 – 2005	Whether intellectual capital (IC) has an impact on the financial aspects of organisational performance	M–B Ratio = Market Capitalisation / BV Common Stock; ROA = Operating Income / BV Total Assets; ATO = Total Revenue / BV Total Assets; ROE = Net Income / Total Shareholders Equity;	VAIC positive with ROA, ROE; HCE negative with ATO, MB SCE positive with ROA, ROE; CEE positive with ROA, ATO, MB, ROE;
Ting and Lean (2009)	Malaysia 1999-2007 (9 years)	Examines the intellectual capital performance and its relationship with financial performance	ROA = Profit after Tax / Total Assets;	VAIC positive with ROA; HCE positive with ROA; CEE positive with ROA;
P. Kujansivu and A. Lonnqvist (2007)	20000 Finnish companies, for the period 2001-2003	Examines the relationship between the value of IC and efficiency of IC	CIV = calculated intangible value	Weak relationship between CIV and VAIC
Dimitrios Maditinos, et al. (2011)	96 Greek companies, for the period 2006 to 2008	Examines the impact of IC on firms' market value and financial performance.	M–B Ratio = Market Capitalisation / BV Common Stock; ROA = Operating Income / BV Total Assets ROE = Net Income / Shareholder's Equity GR = (Current year's revenue – Last year's revenue) -1 *100	VAIC has no relationship with M/B, ROA, ROE, GR; HCE positive with M/B, ROA, ROE, GR;

Authors and year	Country and Period	Research Proposition	Dependent Variables	Significant Relationships
Gholamhossein Mehralian et al. (2012)	Listed Iranian pharmaceutical companies (2004-2009)	Examines the relationship between intellectual capital (IC) components (human, structural, and physical capitals) with the traditional measures of financial performance	M–B Ratio = Market Capitalisation / BV Common Stock; ROA = Operating Income / BV Total Assets; ATO = Total Revenue / BV Total Assets	CEE positive with ROA;
Fethi Calisir Et al. (2010)	Listed information technology company of Istambul stock exchange (2005-2007)	Examines VAIC and its components' impact on company performance	M–B Ratio = Market Capitalisation / BV Common Stock; ROA = Operating Income / BV Total Assets; ROE = Net Income / Shareholder's Equity ATO = Total Revenue / BV Total Assets	HCE positive with ROA CEE positive with ROE, ATO
Jose María Díez et al. (2010)	Spanish firms with a staff of 25 employees or more	Explores the possible relation between indicators of human and structural capital and the economic-financial results.	SALES GROWTH (SG) = (current year's sales- Last year's sales)/ Last year's sales	HCE positive with SG SCE positive with SG
Tan et al. (2007)	150 publicly lised companies of Singapore exchange	Investigates the association between the intellectual capital (IC) of firms and their financial performance.	ROE = profit to shareholders / shareholders fund EPS = profit to shareholders / number of shares; ASR = difference between current and last year's share price +dividend / current year's share price	VAIC positive with ROE, EPS, ASR
Stevo Pucar (2012)	134 firms in Bosnia and Herzegovina (B&H)	Analyses the impact of intellectual capital (IC) on export performance	Exports per worker = Total exports /Number of employees Growth of exports per worker = difference between current and last year's export / current tears export	VAIC is positive with growth of exports to some of sample companies
Zeghal and Maaloul (2010)	300 UK companies comprising high-tech, service and traditional sectors	Analyses the role of value added (VA) as an indicator of intellectual capital (IC), and its impact on the firm's economic, financial and stock market performance	Operating income/sales (OI/S) = operating income / total sales M–B Ratio = Market Capitalisation / BV Common Stock; ROA = Operating Income / BV Total Assets;	ICE positive with OI/S and ROA; ICE positive with M/B only to high-tech companies
Mondal and Ghosh (2012)	65 Indian banks for the period 1999 to 2008	Investigates empirically the relationship between intellectual capital and financial performance	ROA = Net Income less preference dividends / BV Total Assets; ROE = Net Income / Shareholder's Equity ATO = Total Revenue / BV Total Assets	VAIC positive with ROA, ROE,ATO; HCE positive with ROA, ROE,ATO
G.B. Kamath (2008)	25 Indian pharmaceutical companies (1996-2006)	Examines the relationship between intellectual capital components and traditional measures of financial performance	ROA = Net Income less preference dividends / BV Total Assets; ATO = Total Revenue / BV Total Assets; M–B Ratio = Market Capitalisation / BV Net Assets;	No significant relationship found by the author

Authors and year	*Country and Period*	*Research Proposition*	*Dependent Variables*	*Significant Relationships*
Makki and Lodhi (2009)	Seven year data set (2001-2007) in Lahore Stock Exchange Index companies (LSE-25)	Examines the relationship between intellectual capital and return on investment (ROI)	ROI = Net income / Total assets	HCE positive with ROI; CEE positive with ROI
Syed Najibullah (2005)	22 banks in Bangaladesh (2005)	Investigates empirically the relation between the value creation efficiency and firms' market valuation and financial performance of 22 Bangladeshi banks	ROE = pre - tax income ÷ average stockholders' equity ROA = pre - tax income ÷ average total assets GR = (current year's revenues ÷ last year's revenues) - 1) × 100%. Employee productivity (EP) = pre - tax income ÷ number of employees	VAIC only positive on GR; CEE has positive impact on ROE and ROA

From Table 1, it is clear that VAIC is very popular among researchers across several countries to measure value creation efficiency of a company's intellectual capital. It very simple and straight forward procedures in the computation of the necessary indexes and coefficients, which may be simple to understand, especially for management and business people who are accustomed to traditional accounting information.

Effectiveness of Intellectual Capital Management

The management of intellectual capital done by the Indian knowledge companies in order to achieve target efficiency is examined in this section. In other words, we examine the speed of achieving target efficiency level by an individual knowledge company using partial adjustment model. It may be mentioned here that the speed of achieving the targeted IC efficiency is viewed as the degree of efficiency of the management of an enterprise in managing its intellectual capital. Thus higher the observed speed, greater the efficiency of the IC management by the firm and vice-versa. The following sections describe about the data source, methodology and results.

Data Source

The present study is conducted on 100 Indian knowledge companies comprising three sectors, namely, information technology sector (32), pharmaceutical sector (32), and banking sector (36). The data used in this empirical study are collected from published annual reports of respective company and from Capitaline Database. Online database of Reserve Bank of India is also accessed to collect bank data. The companies selected in this study are leader in their respective businesses in terms of market capitalisation as on 15th August 2012 and are listed in the Indian stock markets (BSE and/or NSE). However, banks are selected on the basis of availability of 10 years data. Companies with missing data and negative data are not included in this data. In this study 10 years data from the year 2002 to 2011 are used.

Research Methodology

In order to measure the firm's intellectual capital management efficiency in achieving the target level efficiency during study period following OLS model, i.e., partial adjustment model, has been used.

$$y_t - y_{t-1} = \alpha + \beta (Z^* - y_{t-1}) + e$$

where

$y_t - y_{t-1}$ = actual change = Y,

$Z^* - y_{t-1}$ = desired change = X,

Z^* = target efficiency level,

β = coefficient of adjustment/ speed of efficiency, $0 \leq \beta \leq 1$,

Therefore, our OLS equation is as follows:

$$Y = \alpha + \beta X + e,$$

The estimated beta (β) value represents the speed of the individual firm in achieving the target efficiency level. Here, β =1 for a firm indicates the degree of firm's efficiency in the matter of managing intellectual capital is equal to the target level. Similarly, $\beta \leq 1$ speaks for the need for further improvements by the respective firm.

In this study industry intellectual capital efficiency rate (i.e., industry average) is taken as the target efficiency level. In calculating the industry intellectual capital efficiency rate i.e., the target efficiency level, we use 'equal-weighted mean'. However, we excluded firms those have abnormally high or negative intellectual capital efficiency in calculating equal-weighted mean. Median value can also be used in place of equal- weighted mean. In a situation, where individual firm's efficiency level is high above the industry average level, then industry average do not represents the target level to that firm and another target level is selected for those firms. A firm with consistent intellectual capital efficiency throughout the study period is selected among the firms having higher intellectual capital efficiency than industry average. The selected firm's intellectual capital efficiency is considered as the target efficiency level for those firms having higher efficiency level than industry average.

Empirical Results

In this section empirical results are presented. On the basis of observed beta values companies are classified into top performer beta value is above 0.80, good performers have beta score between 0.50 to 0.80, and common performers where beta score below 0.50.

Information Technology sector

Table 2 presents the regression results of 32 information technology (IT) companies. Among the selected companies, two companies namely, TCS and Rolta India Limited have higher intellectual capital efficiency than industry average and Rolta India Limited has consistent intellectual capital efficiency over the study period. Out of the 32 sample IT companies, Geodesic and ICSA are excluded in the study because of abnormally high intellectual capital efficiency and 3i InfoTech also remains outside this study because of decreasing intellectual capital efficiency.

Table 2: Regression Results of Information Technology Companies

Software Companies	*α*	*β*	R^2	*t-value*
TCS	-0.968	0.782	0.612	3.313*
INFOSYS	0.003	0.733	0.538	2.853*
WIPRO	0.138	0.650	0.422	2.262*
HCL INFO	0.280	0.468	0.219	1.402
ORACLE	0.361	0.788	0.621	3.387*
MAHINDRA SATYAM	-0.520	0.832	0.692	3.968*
TECH MAHINDRA	0.775	0.832	0.693	3.974*
MPHASIS	0.245	0.804	0.647	3.580*
HEXAWARE	-0.702	0.992	0.984	20.529*
FINANCIAL TECH	0.638	0.751	0.564	3.009*
CMC	-0.183	0.648	0.420	2.521*
INFOTECH	-0.556	0.867	0.751	4.597*
NIIT	0.035	0.514	0.264	1.340
POLARIS	0.181	0.560	0.314	1.788*
ZENSAR	-0.556	0.935	0.874	6.969*
ROLTA				
HINDUJA VENTURE	-0.491	0.521	0.272	1.617*
HCL INFOSYSTEM	-0.127	0.592	0.769	3.187*
TATA ELAXI	0.133	0.639	0.408	2.195*
GEOMETRIC	-1.029	0.867	0.751	4.599*
RICOH	-1.118	0.963	0.927	9.410*
MASTEK	0.193	0.393	0.155	1.132
SASKEN	-0.276	0.665	0.442	2.356*
SAKSOFT	-0.140	0.717	0.514	2.302*
BLUE STAR	-0.288	0.580	0.337	1.884*
DATAMATICRS	-0.366	0.584	0.341	1.905*
KALE CONSULTANTS	0.069	0.766	0.587	3.153*
SUBEX AZURE	0.302	0.388	0.151	1.115
SONATA SOFTWARE	-0.450	0.792	0.628	3.434*

Here * represents significance level at 1% level.

From the regression results in Table 2 it is revealed that beta values are not significant in 3 companies out of 28 companies and beta values of these three companies do not exceed 0.50. However, eight companies have beta values more than 0.80 ($0.80 \leq \beta = 1.00$) and these beta values are significant at 1% significance level. The beta

values of 17 companies, which are significant at 1% significance level, lie between 0.50 and 0.80 ($0.50 \leq \beta \leq 0.80$) and three companies have values less than 0.50. From the beta values it is also appeared that Hexaware technologies is the most successful company among the 32 sample IT companies with regard to achieving target intellectual capital efficiency level and followed by Ricoh India Ltd., Zensar Technologies Ltd. and so on.

Pharmaceutical Sector

Table 3: Regression Results of Pharmaceutical Companies

Name of Company	*α*	*β*	R^2	*t-value*
Sun Pharma	-0.270	0.596	0.355	1.965*
Cipla	-0.584	0.621	0.386	2.904*
Dr. Reddy	0.132	0.800	0.640	3.526*
Lupin	-0.014	0.708	0.502	2.656*
Ranbaxy	-0.415	0.591	0.349	1.937*
Cadila Health	-0.990	0.528	0.379	1.646*
Glaxo	2.427	0.654	0.434	2.346*
Divis Lab				
Glenmark	0.120	0.622	0.386	2.100*
Phiramal Healthcare	-0.750	0.773	0.597	3.223*
Torrent Pharma	-0.124	0.789	0.623	3.400*
Ipca Labs	-0.238	0.397	0.157	1.143
Sanofi India	0.076	0.626	0.392	2.124*
Biocon	2.000	0.464	0.216	1.387
Strides Acrolab	-0.113	0.637	0.406	2.189*
Astrazencea	0.041	0.886	0.784	5.045*
Pfizer	-0.367	0.769	0.592	3.186*
Aurobindo Pharma	-1.298	0.660	0.436	2.327*
Jubilant Life	0.322	0.531	0.382	1.659*
Novertis India	-0.297	0.897	0.805	5.307*
Wyeth	0.495	0.323	0.105	0.902
FDC limited	0.114	0.631	0.398	2.150*
Unichem	0.027	0.465	0.216	1.389*
Fresuniues Kabi	-0.377	0.44	0.193	1.094
Alembic	-0.601	0.770	0.593	3.196*
Plethico Pharma	1.193	0.433	0.188	1.272
Natco Pharma	-0.268	0.80	0.631	3.461*
Merck Limited	0.041	0.354	0.125	1.001
Ajanta Pharma	-0.343	0.808	0.654	3.634*
Orchid Chemicals	-0.566	0.865	0.749	4.571*
Dishman Pharma	-1.000	0.801	0.636	3.496*

Here, * denotes 1% significance level

Table 3 shows the regression results of 31 pharmaceutical companies. One company (Abbott India Ltd.) is excluded from the study having decreasing intellectual capital efficiency level over the period. Out of 31 sample companies, eight companies i.e. Cipla, Glaxo, Divi's Laboratory, Biocon, Aurobindo Pharma, Wyeth, Plethico Pharma, and Dishman Pharma have higher intellectual capital efficiency than the industry average level. Among the eight pharmaceutical companies 'Dishman pharma' has consistent intellectual capital performance over the 10 years study period and selected as target level to those eight outperforming companies. However, the beta values of seven companies are above 0.80 ($0.80 \leq \beta = 1.00$) out of 30 sample firms. Sixteen companies have beta values between 0.50 and 0.80 ($0.50 \leq \beta \leq 0.80$) and seven companies have beta values less than 0.50 ($0 \leq \beta \leq 0.50$). From the empirical results it is seen than Novertis India (β= 0.897) is the most successful company with regard to achieving target efficiency level and followed by Astrazencea (β= 0.886), Orchid Chemicals (β= 0.865).

Banking Sector

The regression results of 36 banks are presented in Table 4. Here it is seen that Corporation Bank, Oriental Bank of Commerce, Andhra Bank, HDFC Bank, IDBI Bank, ICICI Bank, Karur Vysya Bank, and Axis Bank have higher intellectual capital efficiency than the industry average and among them Axis Bank is selected as leader because of having consistent intellectual efficiency than others throughout the study period. From the table it is seen that beta values of seven banks are above 0.80 ($0.80 \leq \beta = 1.00$). However, beta values of 22 banks fall between 0.50 to 0.80 ($0.50 \leq \beta \leq 0.80$) and six banks have beta values less than 0.50 ($0 \leq \beta \leq 0.50$). From the sample of 36 sample banks it is seen that Kotak Mahindra Bank (β =0.907) is the most successful bank in achieving the target intellectual efficiency level.

Table 4: Regression Results of Banks

Bank Name	*A*	*β*	R^2	*t-value*
Allahabad Bank	-0.099	0.628	0.3950	2.136*
Bank of Baroda	0.05	0.299	0.1390	1.536
Bank of India	-0.478	0.787	0.6190	3.375*
Canara Bank	-0.082	0.61	0.3720	1.722**
Central Bank of India	-0.702	0.603	0.3640	2.002**
Corporation Bank	-0.504	0.433	0.1880	1.271

Bank Name	A	β	R^2	t-value
Dena Bank	-1.481	0.867	0.7510	4.601*
Indian Bank	0.05	0.589	0.3470	1.927**
Indian Overseas Bank	-0.535	0.754	0.5680	3.033*
O B of Commerce	-1.196	0.818	0.6700	3.769*
Punjab National Bank	-0.064	0.516	0.2600	1.594**
Syndicate Bank	-0.688	0.629	0.3960	2.141**
UCO Bank	-0.58	0.66	0.4360	2.325**
Union bank of India	0.169	0.633	0.4000	2.161**
Vijaya Bank	-0.27	0.558	0.3110	1.778
State Bank of B & J	-0.399	0.837	0.7000	4.045*
State Bank of India	-0.279	0.535	0.2870	1.677**
State Bank of Mysore	-0.187	0.564	0.3180	1.808**
S B Of Travancore	-0.088	0.732	0.5360	2.483*
Andhara Bank	-0.036	0.277	0.0770	0.768
Bank of Maharastra	-0.476	0.78	0.6080	3.297*
ING Vysya Bank	-0.898	0.744	0.5530	2.945*
Kotak Mahindra Bank	-0.354	0.907	0.8220	5.687*
INDUSLAND BANK	-0.658	0.891	0.7930	5.185*
City Union Bank	-0.334	0.535	0.2860	1.674**
Dhanalakshmi Bank	-1.444	0.831	0.6910	3.955*
Federal Bank	0.131	0.632	0.3990	2.156**
HDFC Bank	-0.236	0.487	0.2370	1.474
ICICI Bank	-0.11	0.531	0.2820	1.658**
IDBI Bank	-0.184	0.667	0.4450	2.370*
Karnataka Bank	-0.068	0.254	0.0440	0.694
Karur Vysya Bank	-0.083	0.253	0.0640	0.639
South Indian Bank	-0.284	0.648	0.4200	2.250*
Axis Bank				
YES Bank	0.651	0.801	0.6320	2.261*
Lakshmi Vilas Bank	-0.531	0.657	0.4320	2.307*

Here, * denotes 1% significance level and ** denotes 5% significance level.

Discussion of Results

This paper makes an attempt to examine the efficiency of intellectual capital management with regard to target level of Indian knowledge companies during the period 2002 to 2011. For measuring the value creation efficiency of intellectual capital Pulic's (2000) VAIC (value added intellectual capital) model has been used in this study. Partial adjustment model is used in the study in order to measure speed of achieving target level by an individual company during the study period. In evaluating the individual firm's efficiency with regard to the speed of achieving target level of efficiency, industry norm is selected as target efficiency level. From the empirical results it is seen that out of the 100 sample companies, 22 companies are top performers since, beta values are more than 0.80 and beta values of 55 companies fall 0.50 to 0.80. However, 14 companies out of the above 55 companies have beta values nearer to 0.80. All over 13 among hundred sample companies have beta values less than 0.50 and categorised as common performers. From the beta values of regression results it is also observed that IT companies are more efficient in intellectual capital management with regard to target level as compared to banks and pharmaceutical companies.

Conclusions

In the present chapter attempt is made to examine the intellectual capital performance of 100 knowledge companies. The value creation efficiency of intellectual capital is measured through Pulic's VAIC method. From the average VAIC scores it is revealed that value creation efficiency of pharmaceutical companies comes first and followed by software companies and banks. The average IC efficiency score is four whereas the score is three in case of software companies and banks. In this chapter an attempt is also made to examine the intellectual capital management of sample companies in order to achieve the target efficiency level. Here, industry average IC efficiency is considered as target efficiency level for all companies and companies whose efficiency level are above industry average, industry leader's efficiency is selected as target level. Empirical results show that companies which have better intellectual capital efficiency are able to manage their intellectual capital efficiently to achieve the target efficiency level.

References

Barth, M. E., & Clinch, G. (1998). Revalued financial, tangible, and intangible assets: associations with share prices and non market-based estimates. *Journal of Accounting Research,* 36, 199-233.

Bornemann, M., Knapp, A., Schneider, U., & Sixl, K. I. (1999). Holistic measurement of intellectual capital, International Symposium: Measuring and reporting intellectual capital: Experiences, Issues and Prospects. Retrieved from www.oecd.org/dataoecd/16/20/ 1947871.pdf.

Cabrita, M., & Vaz, J. (2006). Intellectual capital and value creation: Evidence from the por-tuguese banking industry. *The Electronic Journal of Knowledge Management,* 4(1), 11-20, Retrieved from www.ejkm.com.

Chan, K. H. (2009a). Impact of intellectual capital on organizational performance: An empirical study of companies in the Hang Seng Index (Part 1). *The Learning Organization,* 16(1), 4-21.

Chen, M., Cheng, S., & Hwang, Y. (2005). An empirical investigation of the relationship between intellectual capital and firms' market value and financial performance. *Journal of Intellectual Capital,* 6(2), 159-176.

Clarke, M., Seng, D., & Whiting, R. H. (2010). Intellectual capital and firm performance in Australia. *Journal of Intellectual Capital*, 12(4), 505-530.

Firer, S., & Williams, S. M. (2003). Intellectual capital and traditional measures of corporate performance. *Journal of Intellectual Capital,* 4(3), 348-360.

Firer, S., & Stainbank, L. (2003). Testing the relationship between Intellectual Capital and a company's performance: Evidence from South Africa. *Mediatory Accountancy Research,* 11, 25-44.

Ghosh, S. K., & Mondal, A. (2009). Indian Software and Pharmaceutical Sector IC and Financial Performance. *Journal of Intellectual Capital,* 10(3), 369-388.

Goh, P. C. (2005). Intellectual capital performance of commercial banks in Malaysia. *Journal of Intellectual Capital,* 6(3), 385-396.

Guthrie, J., & Petty, R., (2000). Intellectual capital: Australian annual reporting practices. *Journal of Intellectual Capital,* 1(1), 241-251.

Joshi, M., Cahill, D., & Sidhu, J., (2010). Intellectual capital performance in the banking sector: An assessment of Australian owned banks. *Journal of Human Resource Costing & Accounting,* 14(2), 151-170.

Kallapur & Kwan (2004). The value relevance and reliability of brand assets recognized by UK firms. Accounting Review, 79(1), 151-72.

Kamath, G. B. (2007). The Intellectual Capital Performance of the Indian Banking Sector. *Journal of Intellectual Capital,* 8(1), 96-123.

Kamath, G. B. (2008a. Intellectual capital disclosure in India: Content analysis of TecK firms. *Journal of Human Resource Costing and Accounting*, 12(3), 213-24.

Kamath, G.B. (2008b), "Intellectual capital and corporate performance in Indian pharmaceutical Industry", Journal of Intellectual Capital, Vol. 9, No.4, pp. 684-704.

Kujansivu, P., & Lonnqvist, A. (2005). The value and efficiency of intellectual capital in Finnish companies. Retrieved from www.tut.fi/units/tuta/tita/tip/Kujansivu_Lonnqvist.pdf.

Lev, B., & Zambon, S. (2003). Intangibles and intellectual capital: An introduction to a special issue. *European Accounting Review,* 12(4), 597-603.

Maheran, N., Muhammad, N., & Ismail, M. (2009). Intellectual capital efficiency and firm performance: Study on Malaysian financial sectors. *International Journal of Economics and Finance,* 1(2), 206-212.

Molodchik, M., & Bykova, A. (2011). Applying the VAICTM model to Russian industrial Enterprises. Proceedings of the 3rd European Conference on Intellectual Capital held at University of Nicosia, Cyprus, April 18 -19, pp. 268-277

Mondal, A., & Ghosh, S. K. (2012). Intellectual capital and financial performance of Indian banks. *Journal of Intellectual Capital,* 13(4), 515-530.

Meritum. (2002). MEasuRing Intangibles to Understand and Improve Innovation Management – Final Report, Project funded by the European Community under the Targeted Socio-Economic Research (TSER). Retrieved from http://www.pnbukh.com/site/files/pdf_filer/FINAL_REPORT_MERITUM.pdf

Pal, K., & Soriya, S. (2012). IC performance of Indian pharmaceutical and textile industry. *Journal of Intellectual Capital,* 13(1), 120-137.

Pulic, A. (2000). MVA and VAIC^{TM} analysis of randomly selected companies from FTSE 250. Retrieved from www.vaic-on.net/start.htm

Pulic, A. (2002). Are we Creating or Destroying Value? Retrieved from http://www.vaic on.net/start.htm.

Schneider, U. (1999). The Austrian Approach to the Measurement of Intellectual Potential. Retrieved from http//www.measuringip.at/Opapers/Schneider/Canada/theoreticalframework.html.

Shiu, H.-J. (2006). The application of the value added intellectual coefficient to measure corporate performance: Evidence from technological firms. *International Journal of Management,* 23(2), 356-365.

Sveiby, E. E. (2001, 2007, 2010). Methods for measuring intangible assets. Retrieved from www.sveiby.com/articles/IntangibleMethods.html

Tan, H. P., Plowman, D., & Hancock, P. (2007). Intellectual capital and financial returns of companies. *Journal of Intellectual Capital,* 8(1), 76-95.

Ting, I. W. K., & Lean, H. H. (2009). Intellectual capital performance of financial institutions in Malaysia. *Journal of Intellectual Capital,* 10(4), 588-99.

Tseng, C., & Goo, J. (2005). Intellectual capital and corporate value in an emerging economy: Empirical study of Taiwanese manufacturers. *R&D Management,* 35(2), 187-201

Williams, S. (2001). Is intellectual capital performance and disclosure practices related? *Journal of Intellectual Capital,* 2(3), 192-203.

Zeghal, D. M., & Maaloul, A. (2010). Analysing value added as an indicator of intellectual capital and its consequences on company performance. *Journal of Intellectual Capital,* 11(1), 39-60.

Guidelines for Authors

International Journal of Business Analytics and Intelligence welcomes original manuscripts from academic researchers and business practitioners on the topics related to descriptive, predictive and prescriptive analytics in business. The authors are also encouraged to submit perspectives and commentaries on business analytics, cases on managerial applications of analytics, book reviews, published-research paper reviews and analytics software reviews based on below mentioned guidelines:

Journal follows online submission for peer review process. Authors are required to submit manuscript online at http://manuscript.publishingindia.com

Title: Title should not exceed more than 12 Words

Abstract: The abstract should be limited to 150 to 250 words. It should state research objective(s), research methods used, findings, managerial implications and original contribution to the existing body of knowledge

Keywords: Includes 4–8 primary keywords which represent the topic of the manuscript

Main Text: Text should be within 4000-7000 words Authors' identifying information should not appear anywhere within the main document file. Please do not add any headers/footers on each page except page number. Headings should be text only (not numbered).

Primary Heading: Centered, capitalized, and italicized.

Seconday Heading: Left justified with title-style capitalization (first letter of each word) and italicized.

Tertiary Heading: Left justified and indented with sentence-style capitalization (first word only) in italics.

Equations: Equations should be centered on the page. If equations are numbered, type the number in parentheses flush with the left margin. Please avoid using Equation Editor for simple in-line mathematical copy, symbols, and equations. Type these in Word instead, using the "Symbol" function when necessary.

References: References begin on a separate page at the end of paper and arranged alphabetically by the first author's last name. Only references cited within the text are included. The list should include only work the author/s has cited. The authors should strictly follow APA style developed by American Psychological Association available at American Psychological Association. (2009). Publication manual of the American Psychological Association (6th Ed.). Washington, DC.

Style Check

To make the copyediting process more efficient, we ask that you please make sure your manuscript conforms to the following style points:

Make sure the text throughout the paper is 12-point font, double-spaced. This also applies to references.

Do not italicize equations, Greek characters, R-square, and so forth. Italics are only used on p-values.

Do not use Equation Editor for simple math functions, Greek characters, etc. Instead, use the Symbol font for special characters.

Place tables and figures within the text with titles above the tables and figures. Do not place them sequentially at the end of the text. Tables and figures must also be provided in their original format.

Use of footnotes is not allowed; please include all information in the body of the text.

Permissions

Prior to article submission, authors should obtain all permissions to use any content if it is not originally created by them. When reproducing tables, figures or excerpts from another source, it is expected to obtain the necessary written permission in advance from any third party owners of copyright for the use in print and electronic formats. Authors should not assume that any content which is freely available on the web is free to use. Website should be checked for details of copyright holder(s) to seek permission for resuing the web content

Review Process

Each submitted manuscript is reviewed first by the chief editor and, if it is found relevant to the scope of the journal, editor sends it two independent referees for double blind peer review process. After review, the manuscript will be sent back to authors for minor or major revisions. The final decision about publication of manuscript will be a collective decision based on the recommendations of reviewers and editorial board members

Online Submission Process

Journal follows online submission for peer review process. Authors are required to register themselves at http://manuscript.publishingindia.com prior to submitting the manuscript. This will help authors in keeping track of their submitted research work. Steps for submission is as follows:

1. Log-on to above mentioned URL and register yourself with "International Journal of Business Analytics & Information"
2. Do remember to select yourself as "Author" at the bottom of registration page before submitting.
3. Once registered, log on with your selected Username and Password.
4. Click "New submission" from your account and follow the 5 step submission process.
5. Main document will be uploaded at step 2. Author and Co-author(s) names and affiliation can be mentioned at step 3. Any other file can be uploaded at step 4 of submission process.

Editorial Contact
Dr. Tuhin Chattopadhyay
Email: dr.tuhin.chattopadhyay@gmail.com
Ring: 91-9250674214

Online Manuscript Submission Contact
Puneet Rawal
Email: puneet@publishingindia.com
Ring: 91-9899775880

International Journal of Business Analytics and Intelligence

SUBSCRIPTION DETAILS

Dispatch Address:-
The Manager,
International Journal of Business Analytics and Intelligence
Plot No-56, 1st Floor
Deepali Enclave, Pitampura
New Delhi -110034
Ph - 9899775880

Subscription Amount for Year 2016

	Print	Print + Online
Indian Region	Rs 2700	Rs 3400
International	USD 150	USD 180
Price mentioned is for Academic Institutions & Individual. Pricing for Corporate available on request. Price is Subject to change without prior notice.		

Payment can be made through D.D./at par cheque in favour of "Publishing India Group" payable at New Delhi and send to above mentioned address.

Disclaimer

The views expressed in the Journal are of authors. Publisher, Editor or Editorial Team cannot be held responsible for errors or any consequences arising from the use of Information contained herein. While care has been taken to ensure the authenticity of the published material, still publisher accepts no responsibility for their accuracy.

Journal Printed at Anvi Composers, Paschim Vihar.

Copyright

Printed and published by Publishing India Group, New Delhi. Any views, comments or suggestions can be addressed to – Coordinator, IJBAI, info@publishingindia.com

www.manuscript.publishingindia.com

PublishingIndia.com

Publishing India Group

Plot No. 56, 1st Floor, Deepali Enclave
Pitampura, New Delhi-110034, India
Tel.: 011-47044510, 011-28082485
Email: info@publishingindia.com
Website: www.publishingindia.com

ISSN 2321-1857
02
9 772321 185001

International Journal of Business Analytics & Intelligence

SUBSCRIPTION DETAILS

Dispatch Address:-
The Manager,
International Journal of Business Analytics and Intelligence
Plot No-56, 1st Floor
Deepali Enclave, Pitampura
New Delhi -110034
Ph - 9899775880

Subscription Amount for Year 2016

	Print	Print + Online
Indian Region	Rs 2700	Rs 3400
International	USD 150	USD 180

Price mentioned is for Academic Institutions & Individual. Pricing for Corporate available on request. Price is Subject to change without prior notice.

Payment can be made through D.D./at par cheque in favour of "Publishing India Group" payable at New Delhi and send to above mentioned address.

Disclaimer

The views expressed in the Journal are of authors. Publisher, Editor or Editorial Team cannot be held responsible for errors or any consequences arising from the use of Information contained herein. While care has been taken to ensure the authenticity of the published material, still publisher accepts no responsibility for their accuracy.

Journal Printed at Anvi Composers, Paschim Vihar.

Copyright

Printed and published by Publishing India Group, New Delhi. Any views, comments or suggestions can be addressed to – Coordinator, IJBAI, info@publishingindia.com