ISSN : 0032-9924
e-ISSN : 0976-3902

NATIONAL PRODUCTIVITY COUNCIL

Productivity

Vol. 59 January - March 2019 No. 4

Focus : Data Mining

Productivity

NATIONAL PRODUCTIVITY COUNCIL

A QUARTERLY
JOURNAL OF THE
NATIONAL PRODUCTIVITY COUNCIL

PRODUCTIVITY published since Oct-Nov 1959 is the principal journal of National Productivity Council of India

Subscription Details :
Subscription and orders for Productivity should be sent to the **Prints Publications Pvt Ltd,**
New Delhi, India

Annual Subscription rates for subscribers are :

National	**Rs. 3000.00**
International	**US$ 350.00**

Online version free with print subscription

Electronic Version of the journal can be accessed at :

www.printspublications.com

All remittance must be paid in favour of :
Prints Publications Pvt Ltd
payable at New Delhi, India

All business correspondence should be addressed to :

Prints Publications Pvt Ltd
"Prints House", 11, Darya Ganj
New Delhi-110 002
Telephone : 91-11-45355555

ISSN : 0032-9924
e-ISSN : 0976-3902

E-mail : contact@printspublications.com
www.printspublications.com

Productivity

A QUARTERLY JOURNAL OF THE NATIONAL PRODUCTIVITY COUNCIL

Vol. 59 • January – March 2019 • No. 4

The Journal **"Productivity"** has been included in the UGC Approved list of Journals for Promotion and Recruitment in Academic posts.
EBSCO, ProQuest, Genamics (Journalseek), Indian Citation Index, NAAS (Rating 3.49), Journal DOI No. https://doi.org/10.32381/PROD

Prints Publications Pvt Ltd
New Delhi

Publishers & Distributors:

Prints Publications Pvt Ltd
"Prints House", 11, Darya Ganj,
New Delhi-110 002
Phone : +91-11-45355555
E-mail : contact@printspublications.com
website : www.printspublications.com

ISSN : 0032-9924
e-ISSN : 0976-3902

Published and printed on behalf of National Productivity Council, by Mr Pranav Gupta, **Prints Publications Pvt Ltd,** at Printext, New Delhi, India.

Contents

Data Mining Techniques and its Applications in Agriculture and Health Care Sectors for the Benefit of Mankind

JHARNA MAJUMDAR, MAMATHA BAI B G AND SHILPA ANKALAKI

We live in an era of rapid technological advancement that has facilitated communication and dissemination of information in various sectors. However, in the Indian context, two sectors that are integral for the country's growth still have a long way to go in terms of getting linked with technology that would ultimately benefit people and fuel the country's development.

The study focuses on agriculture and health sectors in this context. To achieve the result, Data Mining algorithms are implemented and the proposed work mainly focuses on the cultivation of cotton and health condition of the foetus and mother for a better tomorrow.

Dr. Jharna Majumdar, Dean R&D, Professor, Department of M. Tech CSE, Nitte Meenakshi Institute of Technology, Bangalore, India.

Mamatha Bai B G, Assistant Professor, Department of M. Tech CSE, Nitte Meenakshi Institute of Technology, Bangalore, India.

Shilpa Ankalaki, Assistant Professor, Department of M. Tech CSE, Nitte Meenakshi Institute of Technology, Bangalore, India.

DOI : - https://doi.org/10.32381/PROD.2019.59.04.1

I. Introduction

In the digitized world, rapid changes occur every moment and technology has helped people lead a more comfortable life. Two main objectives of a human being for simple living is food and good health. Though India is digitized and is growing economically there is lack of awareness about the facilities/schemes given to people by the government, especially those below poverty line. The study attempts to highlight how with the right push technology can help even the BPL category people as well as farmers improve their lives by raising awareness levels.

Agriculture is the backbone of India's economy and even today our farmers have very little knowledge about crop cultivation and appropriate climatic conditions that leads to better yield and vegetation index. The paper focuses on cotton crop and various factors responsible for its cultivation.

Due to the busy lifestyle and stressful environment people do not have adequate time to spend on their health. With this concern, an analysis of the health conditions of the foetus and mother is carried out. The study will highlight various conditions and the time when precautions and proper medications should be taken and to consult the doctor if required.

II. Literature Survey

A. Literature Survey on Agriculture

As the advancement in the technology is increasing in a broader way, the usage of technology in agriculture sector is very limited. Agriculture is the backbone of the Indian economy. But in last few years, agriculture has taken a downturn due to many reasons. Lack of knowledge of seasons, crops and price are also one of them i.e. not

knowing which season for which crop and which crops are receiving higher payment from the government [Ravishankar, 2017]. Many researchers have worked towards application of Data Mining techniques to analyse agriculture data and predict crop yield.

Prediction of crop methodology has been done using Data Mining techniques like Regression model, Bi-clustering Methods, Support Vector Machines etc. Artificial Neural Network. (ANN) is used for prediction of productivity of different crops like cotton, sugarcane, wheat, rice, bajra using agriculture parameters like PH, Nitrogen (ppm) Depth (ppm), Temp (°C) and Rainfall [Poonia, 2017].

Traditional DBSCAN Algorithm is modified by adding KNN distance graph where 'K' is the number of clusters to automatic detection of DBSCAN parameters those are Epsilon and Minpoints. The number of clusters is determined by Batchelor Wilkins clustering algorithms. This Modified DBSCAN algorithm is used to analyse the agriculture data [Ankalaki, 2016] and applied on the dataset to cluster the different districts of Karnataka, which are having similar rain fall, temperature and soil type using optimal Eps value [Majumdar, 2017].

Some researchers like B. Milovic and V. Radojevic have investigated the difficulties of applying Data Mining in Agriculture and obstacles faced by Data Mining in Agriculture [Milovic, 2015]. Over the last few years [Sharma, 2012], several investigations have been carried out on handling data uncertainty associated with databases because of measurement of invalid data sources, errors in the dataset, sampling discrepancy and accuracy measurement. These problems should be reduced before applying any of the traditional data mining techniques because data uncertainty directly influences results of the data mining techniques used to analyse data. Many of the supervised machine learning algorithms are used for agriculture data analysis. Support Vector Machine is one of the supervised machine learning algorithms where it has been used in a number of applications in agriculture domain. SVM has been used to for rice crop yield prediction in India [Gandhi, 2016] by considering parameters like precipitation, minimum, average and maximum temperature, reference crop evapotranspiration, area, production and yield.

B. Literature Survey on Healthcare

In today's world, with the increase in population and advancement of technology there is a change in the lifestyle of every individual that leads to severe problems if not diagnosed properly at an early stage. Early diagnosis helps patients, doctors and medical staff to analyse the medical data and prescribe appropriate treatment [Khatib et al, 2015]. By adopting and utilizing the relevant medical technologies, the decision making can be improved.

Exploring facts from the massive medical dataset is quite a challenging task including effectiveness of the treatment, management of healthcare and customer relationship, protecting data from fraud and abuse etc. Comparative study of various data mining algorithms is analysed for different diseases by applying the Rough set theory and Artificial Neural Network (ANN) [Durairaj, 2013].

Clustering technique like Expectation Maximization Algorithm and Classification methods such as ANN and C4.5 are applied on the attributes of Chronic Kidney Disease and the accuracy of the Data Mining Techniques is discussed [Tabassum, 2017].

[Dharmawardana, 2017] presents the study of Dengue in the tropical regions of Sri Lanka spread by a virus and its presence is predicted by using Machine Learning applications. A conceptual architectural framework designed for clinical data in big data analytics is proposed by the researcher in [Wullianallur, 2014], which is very similar to traditional health informatics. The methodology of this framework comprises establishing a concept statement, proposal, conceptual model and deployment.

Prediction of the presence of diabetes is done using Data Mining Techniques such as Gaussian Naïve Bayes, OPTICS and BIRCH [Bai B G, 2018]. Gaussian Naïve Bayes is used to classify the patients into diabetic and non-diabetic. OPTICS and BIRCH are used to classify the diabetic patients into different clusters based on the stage of the diabetes they are suffering from.

[Majumdar, 2018] describes the applications of CURE algorithm for Healthcare and Video Summarization. For healthcare analysis, Thyroid dataset is used and the clusters formed predict the stage of the disease for early medication. Similarly, for Video Summarization, different category videos are fed as input to obtain the final summary of the respective video sequence.

The various techniques used for Medical Data Mining are summarized in [Khaleel, 2013], which throws light on the utility of local frequent patterns and the presence of the disease. A model to diagnose the heart disease is

predicted in [Shouman, 2012], which uses the Single Data Mining Techniques and Hybrid Data Mining Techniques for the systematic ease of further investigation.

III. Proposed Methodology

The proposed work focuses on applying the same Data Mining Techniques for analysis of applications such as Agriculture and Healthcare thereby educating people about crop cultivation and significance of health conditions. The Data Mining techniques considered in this paper are G Means and Chameleon algorithms.

A. G Means Algorithm

Algorithm: G-means

Input: k: The number of clusters, D: A data set containing n objects

Output: A set of k clusters is obtained.

Steps:

Step 1 : Let C is said to be the initial set of centers like (C <- {x}).

Step 2 : By applying k-means, C<- kmeans(C, X).

Step 3 : The set of data points assigned to center cj .

Step 4 : A statistical test is used to detect that, if each center cj follow a Gaussian distribution.

Step 5 : If the data obtained is Gaussian then keep cj, else replace cj with two new centers.

Step 6: Repeat the step 2, until no more centers are added.

B. Chameleon Algorithm

Algorithm: Chameleon

Input: k : The number of clusters, D: A data set containing n objects

Output : k Clusters

Steps:

Step 1 : Specify a number k as the number of clusters

Step 2 : Randomly select initial centers as the centroids of k clusters

Step 3 : Generate a new partition by assigning each point to the closest cluster centers

Step 4 : Compute the new cluster centers of present partition

Step 5 : If it is a new partition, then go to Step 3 otherwise stop.

C. Input Dataset

Agriculture- Cotton is one of the most important cash crops of India and it plays a vital role both in the agriculture and industrial sector.

Cotton crop productivity and weather data of Karnataka state has been collected from the Indian Government website along with following websites https://data.gov.in/, http://raitamitra.kar.nic.in/ statistics, http://dmc. kar.nic.in/trg.pdf. Cotton productivity of the different regions in Karnataka is analysed by considering the area under cultivation, production, minimum and maximum temperature, rainfall, soil types and PH value.

Healthcare - Medical Dataset is a very confidential data and to obtain it in real time is a high-risk job. For the purpose of research and successful implementation, the Cardiotocographic data from https://archive.ics.uci.edu/ml/datasets.html is considered. This dataset mainly focuses on conditions of foetal movements and accordingly specifies the precautions to be taken by both patients and doctors. The dataset consists of different attributes such as baseline foetal heart rate (FHR), uterine contractions, accelerations, decelerations, repetitive decelerations, abnormal short term variability, Abnormal long term variability, Foetal movements, symmetric tendency, calm sleep and stress situation concerning the health condition of the mother and foetus.

IV. Experimental Results

A. Experimental Results for Agriculture Application

Cotton crop productivity data of the India is gathered from ICAR-Central Institute of Cotton Research. The data has been collected for the year 2004-05 to 2013-14 which consists of area cultivated for the cotton crop, production and cotton productivity.

G Means Clustering

G Means Clustering Algorithm is applied to the dataset to cluster the states according land used for cotton. The dataset consists of area (lakh ha) used for cotton cultivation in different states during 2004-05 to 2013-14. Table 1 shows the clusters formed as a result of G Means Algorithm.

TABLE 1. Clusters - State wise area used for cotton cultivation

Clusters	States	Area (Lakh ha)
Cluster 1	Gujarat, Maharashtra	34.724 – 24.557
Cluster 2	Andhra Pradesh, Madhya Pradesh	15.556 – 6.301
Cluster 3	Punjab, Haryana, Rajasthan, Karnataka	5.493 – 3.915
Cluster 4	Tamil Nadu, Odisha, Others	1.237 – 0.521

Source : Authors own Research Work

By analysing the results in Table 1, Maharashtra and Gujarat states are having the largest land i.e 34.724 lakhs ha and 24.557 lakh ha respectively for cotton cultivation. Karnataka is having an average of 4.721 lakhs ha area for cotton cultivation.

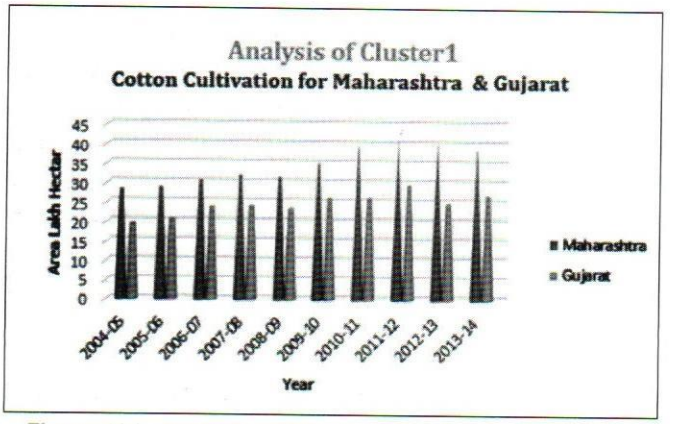


Figure 1(a). Analysis of Cluster 1 for Cotton Cultivation during 2004-05 to 2013-14 for Maharashtra and Gujarat states

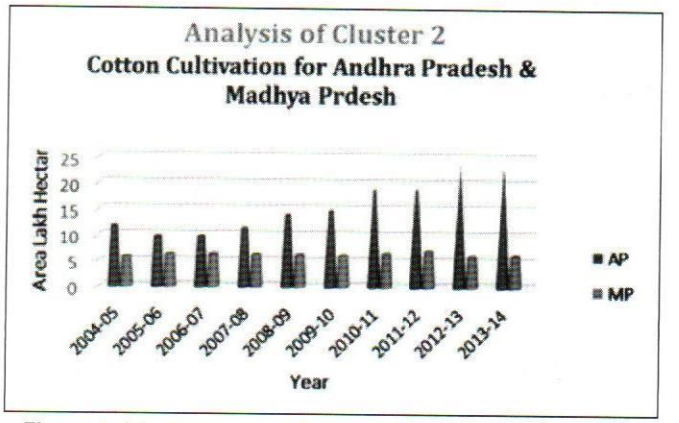


Figure 1 (b). Analysis of Cluster 2 for Cotton Cultivation during 2004-05 to 2013-14 for Andhra Pradesh and Madhya Pradesh states

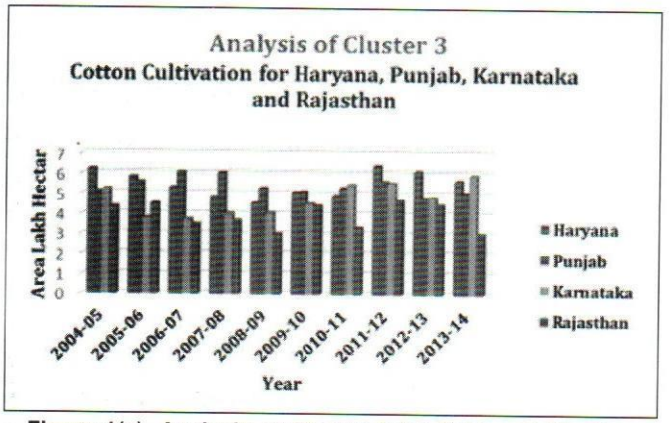


Figure 1(c). Analysis of Cluster 3 for Cotton Cultivation during 2004-05 to 2013-14 for Haryana, Punjab, Karnataka and Rajasthan states

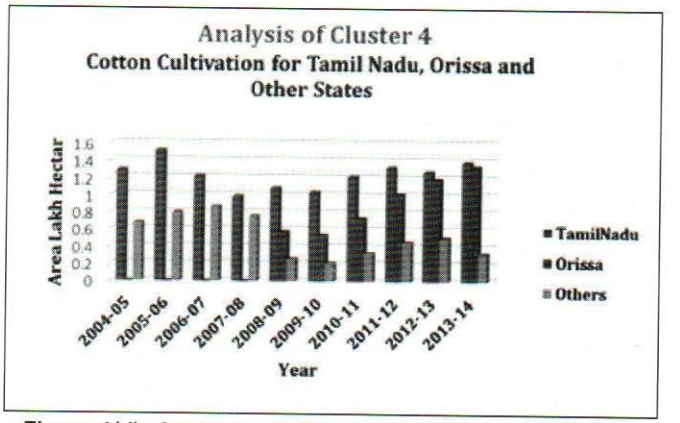


Figure 1(d). Analysis of Cluster 4 for Cotton Cultivation during 2004-05 to 2013-14 for Tamil Nadu, Odisha and other states of India

Figure 1 (a-d) Cluster analysis for Cotton Cultivation in different states of India

Figure 1(a-d) depicts the analysis of each cluster for Cotton Cultivation data. Figure 1(a),(b),(c) and (d) represents detailed analysis of cluster 1, 2, 3 and 4 for cotton cultivation data during the 2004-05 to 2013-14.

G Means Clustering Algorithm is applied to the cotton productivity dataset to cluster the states according to the cotton productivity. The dataset consists of state wise cotton productivity (kg/ha) from 2004-05 to 2013-14. Table 2 depicts the clusters formed as a result of G Means Algorithm.

TABLE 2. Clusters - State wise cotton productivity

Clusters	States	Productivity Range (kg/ha)
Cluster 1	Tamil Nadu, Gujarat, Punjab	605.468– 760.739
Cluster 2	Rajasthan, Others, Haryana, Andhra Pradesh	506.27 –580.741
Cluster 3	Madhya Pradesh, Odisha	439.178 – 484.538
Cluster 4	Maharashtra, Karnataka,	317.273– 390.14

Source: Authors own Research Work

From the analysis of data in Table 2, Tamil Nadu, Gujarat and Punjab states represents the good cotton productivity with the range of 605.468 kg/ha to 760.739 kg/ha. Rajasthan and Haryana states gives average cotton productivity with the range of 506.27 kg/ha – 580.741kg/ha. Madhya Pradesh and Odisha contributes medium cotton productivity with the range of 439.178 kg/ha– 484.538 kg/ha. Maharashtra and Karnataka give low cotton productivity with the range of 317.273kg/ha– 390.14kg/ha.

Fig. 2(a-d) depicts the detailed analysis of cotton cultivation in different states of India during 2004-05 to 2013-14.

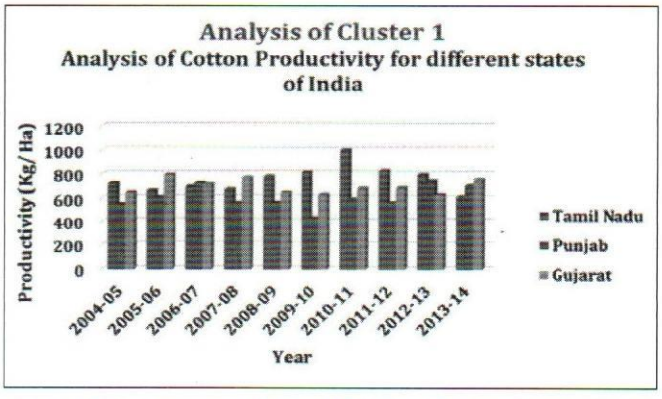


Figure 2(a) Analysis of Cluster 1 for Cotton Productivity during 2004-05 to 2013-14 for different states of India

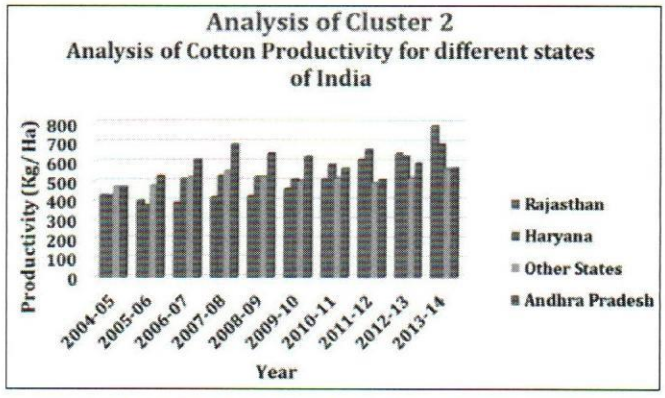


Figure 2(b) Analysis of Cluster 2 for Cotton Productivity during 2004-05 to 2013-14 for different states of India

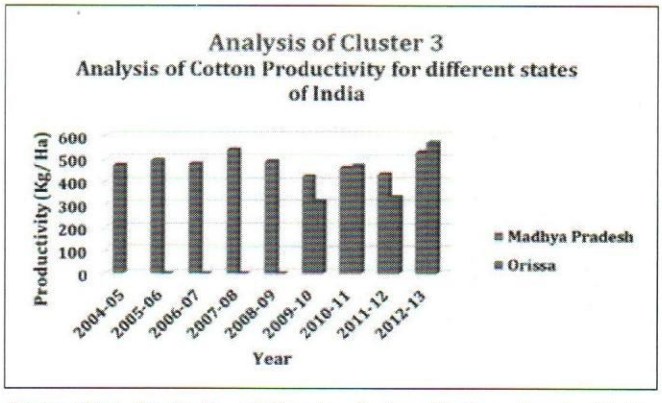


Figure 2(c) Analysis of Cluster 3 for Cotton Productivity during 2004-05 to 2013-14 for different states of India

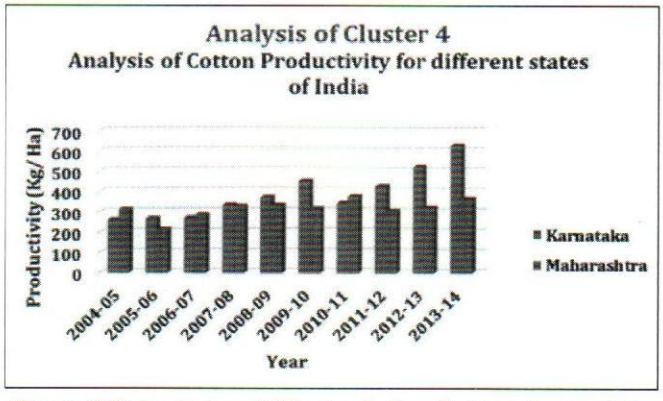


Figure 2(d) Analysis of Cluster 4 for Cotton Productivity during 2004-05 to 2013-14 for different states of India

Fig. 2 (a-d) Cluster analysis for Cotton Productivity for different states of India

Chameleon Clustering

This work focuses on the analysis of cotton crop productivity in different districts of Karnataka. The dataset consists of cotton productivity of different districts of Karnataka during 2001-2009. Table 3 depicts the results of cotton productivity clusters obtained by chameleon clustering.

As result of the Chameleon clustering algorithm 3 clusters are formed. Cluster 1 represents the districts having the high cotton productivity in the range of 1.593-3.032 tones/hectare. Districts belongs to the high production cluster are Gulbarga, Tumkur, Bijapur, Shimoga, Bagalkot, Belgaum, Davangere and Raichur. Cluster 2 represents the districts having the medium cotton productivity in the range of 1.007- 1.483 Tones/ Hectare. Districts belongs to the medium production cluster are Chitradurga, Uttarakannada, Bellary, Haveri, Mandya, Chikmagalur, Hassan, Mysore and Bidar. Cluster 3 represents the districts having the low range cotton productivity in the range of 0.544 – 0.986 Tones/ Hectare. Districts belongs to the low productivity cluster are Dharwad, Gadag, Koppal and Chamarajnagar.

TABLE 3. Clusters District wise cotton productivity

Clusters	Districts	Productivity (Tones/Hectare)	Temperature Range	Soil Type	Ph Value
Cluster 1 (High Productivity)	Gulbarga, Tumkur, Bijapur, Shimoga, Bagalkot, Belgaum, Davangere, Raichur	1.593- 3.032	26.99- 30.2	Black, Red, Laterite, Brown, Mixed	6 - 7.5
Cluster 2 (Medium Productivity)	Chitradurga, Uttarakannada, Bellary, Haveri, Mandya, Chikmagalur, Hassan, Mysore, Bidar	1.007- 1.483	24.1- 31.2	Black, Red, Sandy, Laterite and Mixed	5- 7.8
Cluster 3 (Low Productivity)	Dharwad, Gadag, Koppal Chamarajnagar	0.544 – 0.986	27- 31.5	Black, Red, Granitic, Sandy, Mixed	6- 8.2

Source: Authors own Research Work

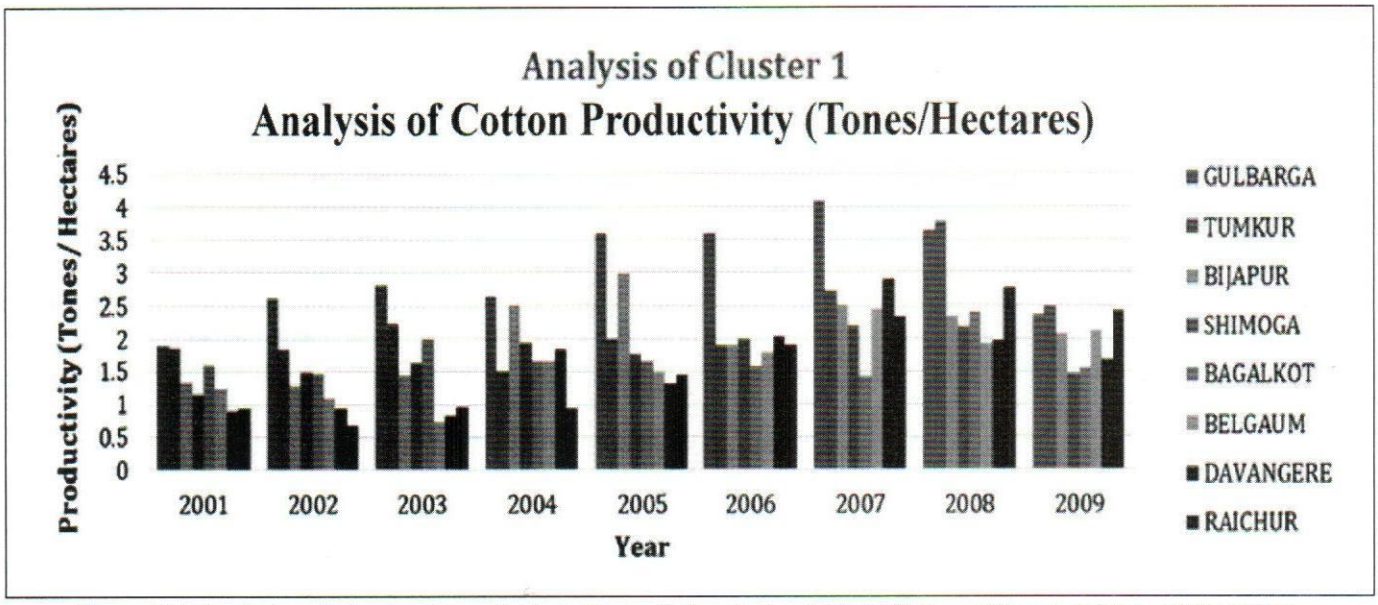


Figure 3(a) Analysis of Cluster 1 for Cotton Productivity during 2001-2009 for different districts of Karnataka

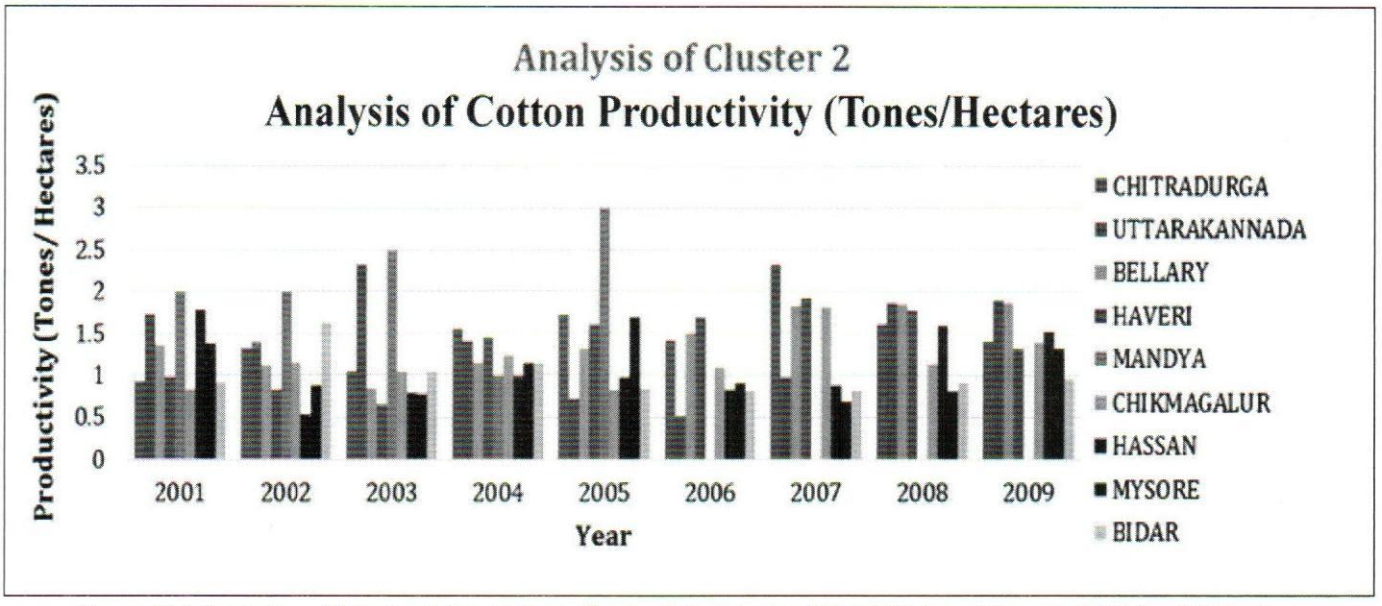


Figure 3(b) Analysis of Cluster 2 for Cotton Productivity during 2001-2009 for different districts of Karnataka

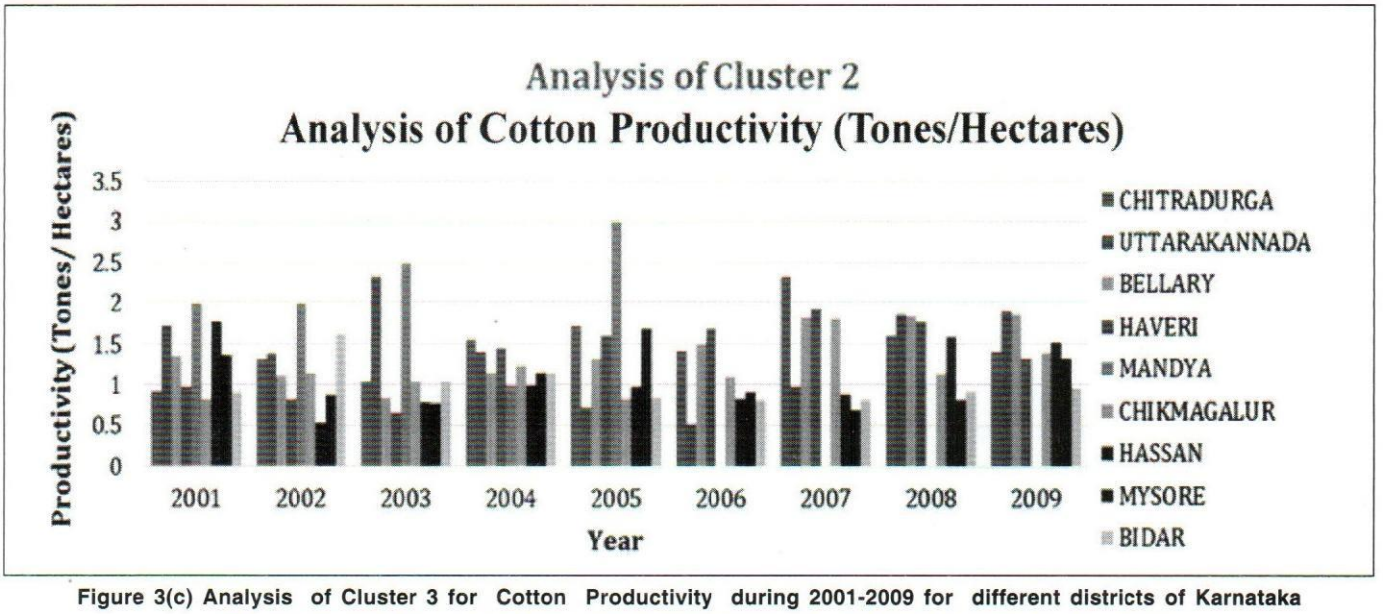


Figure 3(c) Analysis of Cluster 3 for Cotton Productivity during 2001-2009 for different districts of Karnataka

Fig. 3 (a-c) Cluster analysis for Cotton Productivity for different districts of Karnataka

Fig. 3 (a-c) depicts the analyses of cluster wise cotton productivity for different districts of Karnataka during 2001-2009. By using the above clusters, results are studied and analysed to obtain the optimal parameters to produce high cotton productivity. The optimal temperature, rainfall, soil type and soil PH values to produce high cotton productivity is given in the Table 4.

TABLE 4. Optimal Parameters for High Cotton Productivity

	Temperature	Rainfall	Soil Type	PH Value	Soil Depth
Range	25°C	150 cm – 200 cm	Medium Black to Deep Black	6 to 8	20 to 25 cm

Source: Authors own Research Work

B. Experimental Results for Healthcare Application

The Cardiotocographic dataset experimented in this work consists of the observations recorded for about 2130 patients, which keeps track of health conditions of the mother and foetus and educates the patient as well as the doctor to take relevant precautions. The medical dataset is given as an input to the G Means and Chameleon Clustering Algorithms and their respective outputs are obtained. Fig. 4 and Fig. 5 depicts the three different clusters formed based on the health conditions for both the clustering techniques used in this work.

When the same dataset is given as input to both the clustering algorithms, each clustering techniques output is different based on their principle and their outputs are compared to get the optimal resulting value stating the exact health conditions of the patient and the foetus. Fig 6 illustrates the statistics of the clusters formed for different Data Mining Techniques.

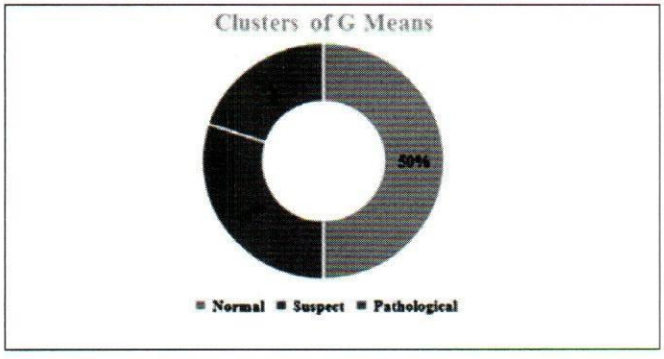


Figure 4: Clusters of G Means

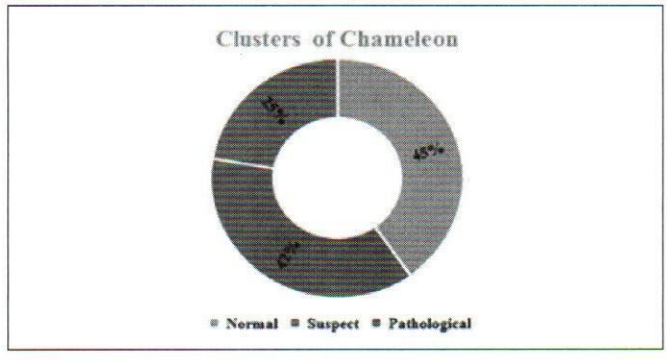


Figure 5: Clusters of Chameleon

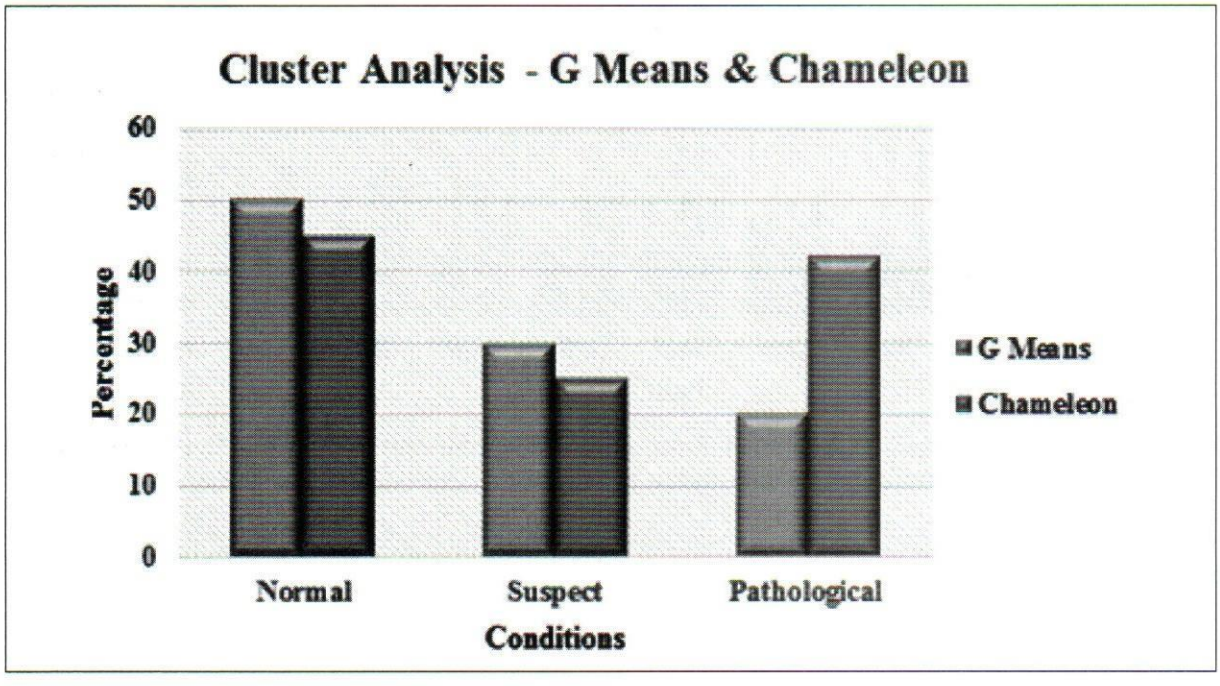


Figure 6: Cluster Analysis

The Heart Rate of the Foetus and Uterine Contractions of the Mother is considered as the basic attributes to determine the health conditions of both the foetus and the mother. Table 5 gives an overview of the clusters formed according to the health conditions w.r.t the readings of FHR and Uterine Contractions for G Means and Chameleon Algorithms.

TABLE 5: Clusters formed based on FHR and Uterine Contractions

Clusters Formed	G Means Algorithm		Chameleon Algorithm	
	Baseline FHR (beats per minute)	Uterine Contractions (% in seconds)	Baseline FHR (beats per minute)	Uterine Contractions (% in seconds)
Normal	100 – 200	50	110 – 160	30 - 60
Suspect	90 – 110	<50	<110	< 30
Pathological	>200	>50	>160	>60

Source: Authors own Research Work

The Foetal Movements are directly proportional to the Accelerations inside the womb which categorizes the health conditions of both the mother and the foetus. Fig. 7 shows how the Foetal movement varies accordingly with the accelerations under Normal health conditions.

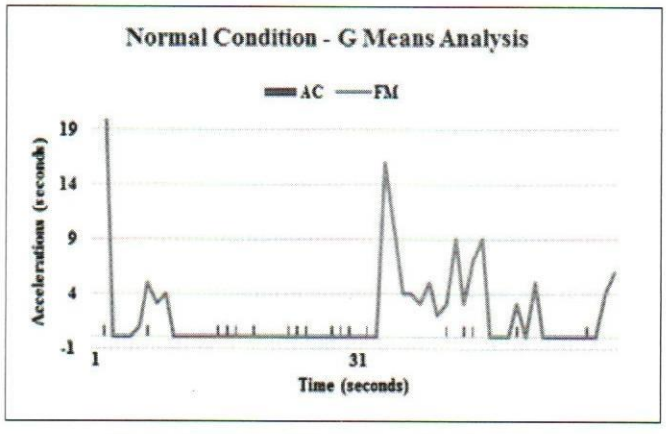


Figure 7a: Foetal Movement for G Means

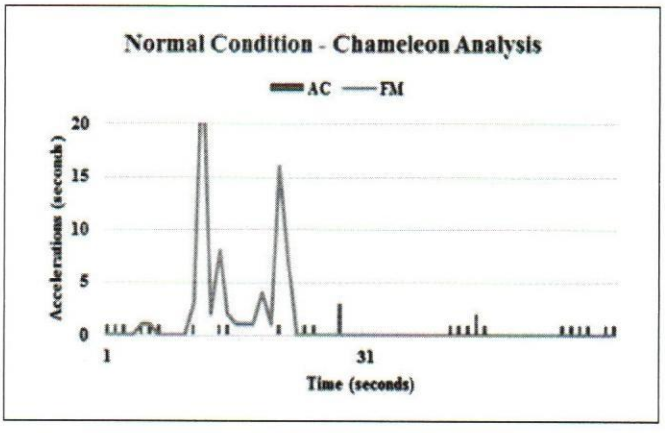


Figure 7b: Foetal Movement for Chameleon

Figure 7: Foetal Movements under Normal Health Conditions

Similarly, the analysis of Foetal movements with respect to the accelerations under the suspicious health condition is shown in Fig. 8, which indicates that the Foetal movements are not synchronising with the Baseline FHR and the accelerations are not at regular intervals.

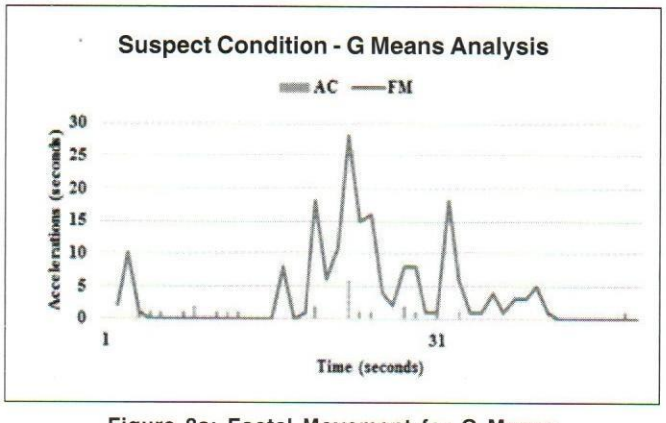


Figure 8a: Foetal Movement for G Means

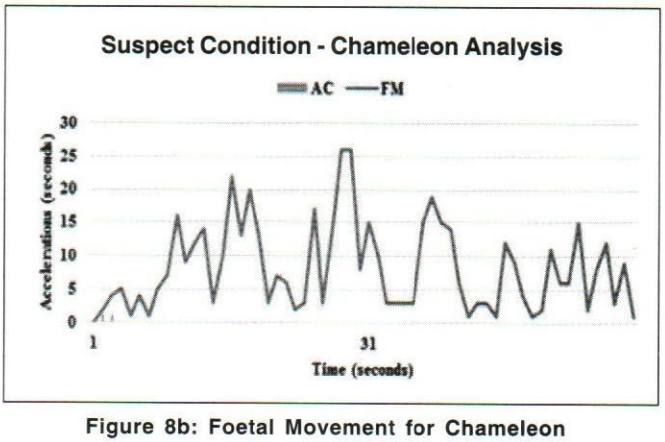


Figure 8b: Foetal Movement for Chameleon

Figure 8: Foetal Movements under Suspect Health Conditions

Figure 9 represents the Foetal movements and the accelerations under pathological health conditions where there is absence or less acceleration and there is less or no movement of the foetus in the womb, which may lead to immediate diagnosis and treatment in order to save both the mother and the child.

With reference to the analysis from Fig 7 to Fig 9, the symmetry of Foetal position is determined by the graphs shown in Fig 10.

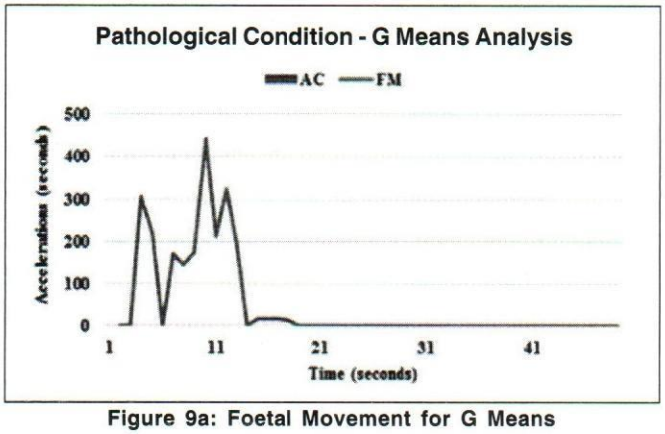


Figure 9a: Foetal Movement for G Means

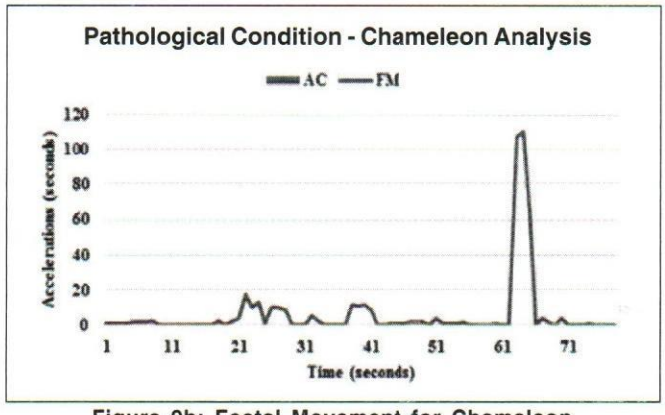


Figure 9b: Foetal Movement for Chameleon

Figure 9: Foetal Movements under Pathological Health Conditions

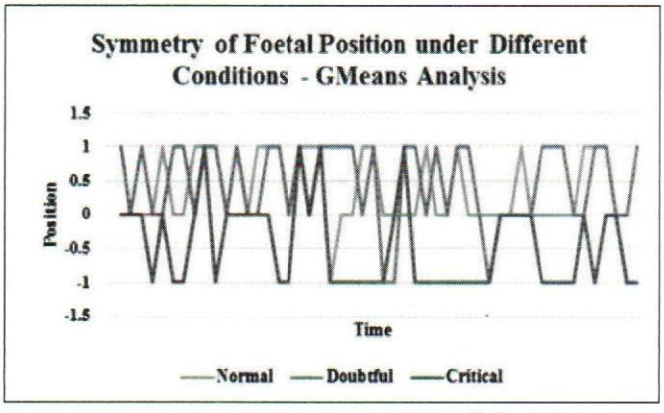


Figure 10 a: Foetal Symmetry for G Means

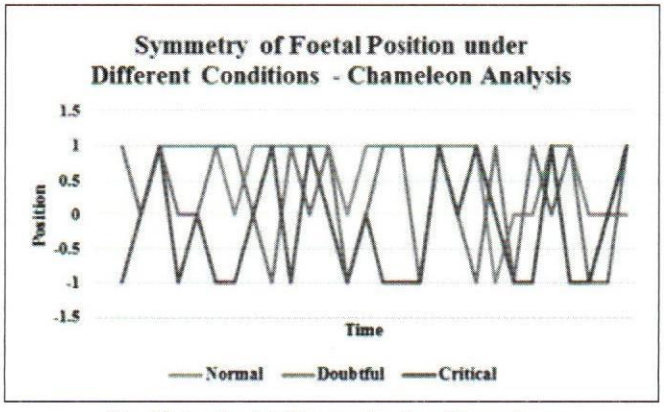


Fig 10 b: Foetal Symmetry for Chameleon

Figure 10: Symmetry of Foetal Position under different Health Conditions

It is observed that the Chameleon Algorithm results are more accurate with the medical reference values and these results will be considered for further analysis of the proposed work. The Chameleon Algorithm is applied for next level analysis by considering the attributes such as Accelerations, Decelerations, Repetitive Decelerations and Symmetric Tendency. Table 6 shows the clusters formed accounting to the internal observations of the foetus in the womb.

TABLE 6: Clusters formed with Foetal Movement recordings in the womb

Clusters Formed	Accelerations (seconds)	Decelerations (seconds)	Repetitive Decelerations	Symmetric Tendency
Normal	5 – 120	<120	No	0 (Equal)
Suspect	30 – 60	<30	Indeterminate	+1 or -1 (Left or Right)
Pathological	>120	>600	Yes	+1 or -1 (Left or Right)

Source: Authors own Research Work

The normal condition tends to accelerate in the range of 5 to 120 seconds with decelerations occurring less than 120 seconds. The normal condition does not lead to repetitive decelerations and has equal symmetry denoted by 0. In case of Suspect and Pathological conditions, the accelerations and decelerations vary with high marginal values and their symmetry is either towards the left (-1) or right (+1).

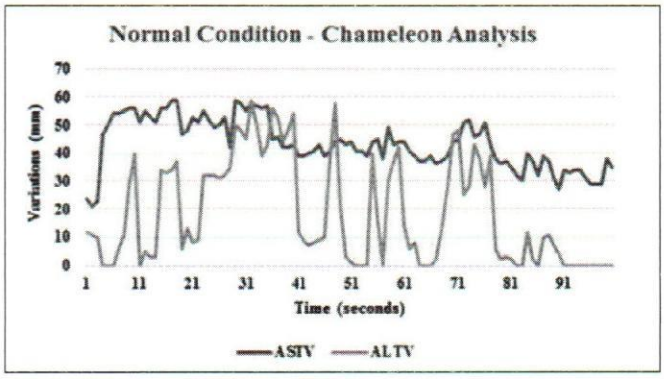


Figure 11a: Abnormal Variability for Normal Condition

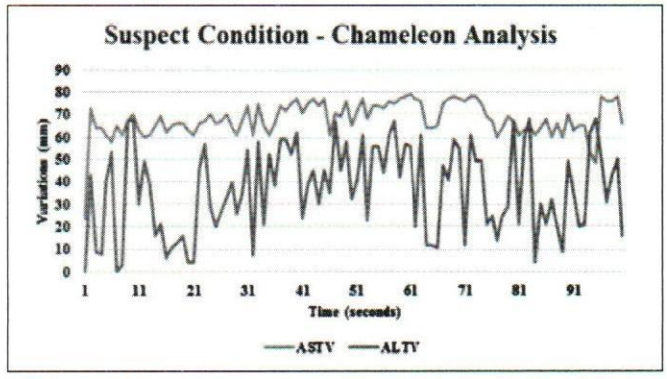


Figure 11b: Abnormal Variability for Suspect Condition

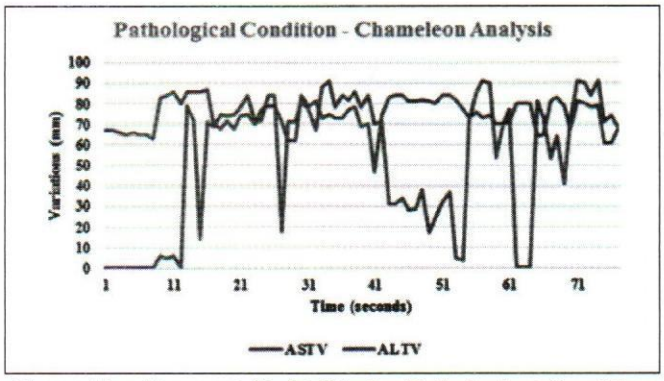


Figure 11c: Abnormal Variability for Pathological Condition

Figure 11: Abnormal Variabilities under different Health Conditions

Fig 11 depicts the Abnormal Short-Term Variability (ASTV) and Abnormal Long-Term Variability (ALTV) variations that take place inside the womb, which is responsible for the health conditions of both the mother and the foetus. If variations are uneven or not at regular intervals, then it is a serious concern that leads to pathological condition.

TABLE 7: Analysis of Mental Health Condition for different Clusters

Cluster Formed	Baseline FHR (beats per minute)	Calm Sleep (of Foetus)	Stress Situation (of Mother)
Normal	110 – 160	Yes	No
Suspect	<110	Indeterminate	Indeterminate
Pathological	>160	No	Yes

Source: Authors own Research Work

The foetus health is proportional to the mother's health and based on the observations and analysis of Table 6, the mental health condition of the mother and foetus is illustrated as shown in Table 7 for various clusters formed as a result of Chameleon Algorithm.

By considering all medical parameters and the above results obtained so far under different categories of clusters for various ranging values, the optimal values of the parameters for normal health condition of the mother and foetus is determined as shown in Table 8.

TABLE 8: Optimal Parameters for Normal Health Condition of Mother and Foetus

Parameters	Baseline FHR (beats per minute)	Uterine Contractions (% in seconds)	Accelerations (seconds)	Decelerations (seconds)	Repetitive Decelerations	Symmetric Tendency
Range Values	110 – 160	30 - 60	5 – 120	<120	No	0 (Equal)

Source: Authors own Research Work

V. Conclusion

In this work, the usefulness of the Data Mining Algorithms like G Means and Chameleon is demonstrated for both the Agriculture and Healthcare domains. It is evident that cotton cultivations depends on a number of factors such as temperature, rainfall, soil type, Ph value and soil depth, which benefit farmers who are dependent on this crop cultivation for their living. Also, on the other hand, the various factors responsible for the normal health condition of foetus and mother is demonstrated, which specifies the range values and occurrence of some of the internal movements through which the mother can be careful regarding the foetus health condition and can seek medical emergency help in suspect or pathological conditions to avoid major impairment.

This work is summarized based on the experimental results obtained by the clustering techniques used for the datasets considered under the study. The proposed methodology emphasises on creating awareness among farmers and the patients in their respective domains to lead a successful life with good health and wealth.

VI. Acknowledgement

The authors express their sincere gratitude to Prof. N. R Shetty, Advisor and Dr. H C Nagaraj, Principal, Nitte Meenakshi Institute of Technology for giving constant encouragement and support to carry out research at NMIT.

The authors extend their thanks to Vision Group on Science and Technology (VGST), Government of Karnataka to acknowledge our research and providing financial support to setup the infrastructure required to carry out the research.

References

Al Khatib et al, "Analysis of Research in Healthcare Data Analytics", *Australasian Conference on Information Systems,* Sydney, 2015, Page 1 – 16.

B. Milovic and V. Radojevic, "Application of Data Mining in Agriculture", *Bulgarian Journal of Agricultural Science,* 21 (No 1) 2015, Page 26-34 Agricultural Academy

Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data", *Journal of Big Data 2017,* https://doi.org/10.1186/s40537-017-0077-4, Page 1 – 15.

Jharna Majumdar, Sumant Dandekar, Mamatha Bai B G, "Implementation of Cure Clustering Algorithm for Video Summarization and Healthcare Applications in Big Data", *Proceedings of Fifth International Conference on 'Emerging Research in Computing, Information, Communication and Applications'* (ERCICA-2018), Springer.

K. Ravisankar, K. Sidhardha, B. Prabadevi, "Analysis of Agricultural Data Using Big Data Analytics", *Journal of Chemical and Pharmaceutical Sciences,* ISSN: 0974-2115, July - September 2017, JCPS Volume 10 Issue 3,Page 1132-1135

K.G.S. Dharmawardana, et. al., "Predictive Model for the Dengue Incidences in Sri Lanka Using Mobile Network Big Data", *International Conference on Industrial and Information Systems* (ICIIS' 2017), IEEE 978-1-5386-1676-5/17, 2017, Page 1 – 6.

Latika Sharma, Nitu Mehta, "Data Mining Techniques: A Tool for Knowledge Management System in Agriculture", *International Journal of Scientific & Technology Research* Volume 1, Issue 5, June 2012, ISSN 2277-8616, Page 67-73

M. Durairaj, V. Ranjani, "Data Mining Applications in Healthcare Sector: A Study", *International Journal of Scientific & Technology Research,* Volume 2, Issue 10, October 2013, ISSN 2277-8616, Page 29 – 35.

Mai Shouman, Tim Turner, Rob Stocker, "Using Data Mining Techniques in Heart Disease Diagnosis and Treatment", *Japan-Egypt Conference on Electronics, Communications and Computers,* 978-1-4673-0484-9/12, 2012 IEEE, Page 173 – 177.

Mamatha Bai B G, Nalini B M, Jharna Majumdar, "Analysis and Detection of Diabetes Using Data Mining Techniques – A Big Data Application in Healthcare", *Proceedings of Fifth International Conference on 'Emerging Research in Computing, Information, Communication and Applications'* (ERCICA-2018), Springer.

Mohammed Abdul Khaleel, Sateesh Kumar Pradham and G.N. Dash, "A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", *International Journal of Advanced Research in Computer Science and Software Engineering,* Volume 3, Issue 8, August 2013, ISSN: 2277 128X, Page 149 – 153.

Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, Amiya Kumar Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines", 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, 978-1-5090-2033-1/16/ $31.00 ©2016 IEEE

Rakesh Poonia, Sonia Bhargava, "Prediction of Crops Methodology using Data Mining Techniques", *International Journal of Innovative Research in Science, Engineering and Technology*, DOI:10.15680/IJIRSET.2017.0610116, ISSN (Online): 2319-8753, ISSN (Print): 2347-6710, Vol. 6, Issue 10, October 2017, Page 19832 - 19838

Shilpa Ankalaki, Neeti Chandra, Jharna Majumdar, "Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka", *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),* DOI 10.17148/IJARCCE, Vol. 5, Special Issue 2, October 2016, Page 25-29.

Tabassum S, Mamatha Bai B G, Jharna Majumdar, "Analysis and Prediction of Chronic Kidney Disease using Data Mining Techniques", *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE),* ISSN (Online) 2394-2320, Vol 4, Issue 9, September 2017, Page 26 – 32.

Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems,* 2014, http://www.hissjournal.com/content/2/1/3, Page 1 – 10.

"Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world."

– Atul Butte

The 10 Vs of Big Data framework in the Context of 5 Industry Verticals

JAYANTHI RANJAN

Big Data is no longer a buzzword that it used to be. It has become part of everyday discussions and has even been added to the English language dictionary. An entire ecosystem of tools and techniques around Big Data has evolved in the form of a multi-billion dollar industry and "Data Scientist" is said to be the best job of this decade. As the tools, technology and know-how of Big Data become mainstream, it is expected that a formula for successful implementation of Big Data initiatives must have become available by now resulting in a large number of companies successfully using Big Data analytics. However, in reality, a huge number of these initiatives continue to fail costing the companies lost investments and missed opportunities.

This paper takes a look at the Big Data challenges mentioned in the literature. After going through various classifications of the problems, it was found that most issues arose due to the very nature of Big Data and the properties which make it different from traditional data. Additionally, it was seen that in some industries some of these properties had a much larger impact on the success of Big Data projects compared to other properties. Thus, a need to study the characteristics of Big Data in the context of different industries was felt and the 10 Vs model describing the ten properties of Big Data was chosen for it.

Jayanthi Ranjan, Professor - Information Management & Systems & Associate Dean - International Relations, IMT Group of Institutions

DOI : - https://doi.org/10.32381/PROD.2019.59.04.2

I. Introduction

Big Data is the business world zeitgeist of this decade. It has been called the innovation story of our time (Brynjolfsson & McAfee, 2011) and most big companies are experimenting, executing and deriving benefits from the use of Big Data. It is nothing short of a revolution and not just a technical one but also a management revolution (HBR, 2012).

Much has been written about the benefits of using Big Data analytics and companies are in a goldrush to extract the maximum value from the data which had been sitting idle with them for years. Some companies have seen tremendous success in improving operational efficiencies, launching new products, decreasing expenses, improving customer engagement and transforming their business for the future (Bean, 2017) through their Big Data initiatives. In fact, some reports suggest that missed opportunities arising out of lack of availability of the right information at the right time can cost businesses making more than £1 billion in earnings up to £20 million yearly in lost revenues (Information Age, 2015).

Therefore, there is a compelling case for businesses across industries to invest resources in leveraging Big Data technologies to generate value for them. This is also reflected in the market, with IDC projecting the Big Data and analytics market to grow to $187 billion in 2019 (CIO, 2016), thus, suggesting that big companies are spending huge money on Big Data tools, technologies and infrastructure.

However, it is not a smooth sailing and the picture is not as rosy as projected in the market. The path to success of Big Data is fraught with risks and failures. Perhaps, the most well-known example of failure of Big Data is the

spectacular failure of Google Flu Trends in 2013 where the predictions were off by 140% when it tried to predict instances of Flu in the US population (Time, 2014). Unlike many other computer science research topics, Big Data receives a lot of interest across industries and even in mainstream media. Headlines like "Big Data: the greater good or invasion of privacy" (Chatterjee, 2013) are commonplace. Considering that it is the public at large that is the one generating this data, everyone is a stakeholder in the Big Data story.

Thus, it makes sense to look beyond the hype and see what is the state of big data project implementations in the industry. The numbers don't look quite promising. A Gartner report (Gartner, 2015) predicted that around 60 % big data initiatives fail to go beyond the piloting stage.This is surprising because Big Data has become so ubiquitous that Gartner dropped it from its hype cycle of emerging technologies back in 2015 (Data Science Central, 2015) and reams has been written about the pervasiveness of Big Data in out lives – *"Big Data is dead, long live Big Data"*(Oliver, 2015) suggesting a slide towards maturity of Big Data. Hence, it becomes imperative to look at the challenges presented by Big Data which makes successful implementation of projects so difficult.

In this paper, an attempt has been made to analyze the problems presented by Big Data in the context of the 10 Vs – Volume, Variety, Velocity, Veracity, Variability, Validity, Visualization, Vulnerability, Volatility and Value characteristics of Big Data to explore if these properties hold the key to establishing success of Big Data project implementations. The rest of the paper is an exploratory study to answer this question. The next section is the research background of this paper which attempts to explain the motivation for choosing the 10 Vs framework, followed by the literature survey which helps in setting the base for the research question and research methodology. The analysis section is next, which tries to assess the impact of each of the 10 Vs, followed by the technical and managerial implications of the previous analysis and finally concluding with the limitations of this study.

2. Background & Motivation

Even though there is a plethora of literature on the subject, there is no one single definition of Big Data (even though the word has been added to the Oxford English Dictionary) as it is a continually evolving field and our understanding of it is improving over time. As a concept Big Data had been around for a long time but the mainstream usage of it in its current context is credited to Roger Mougalas of O'Reilly Media in 2005 (Datafloq, 2017) even though John Mashey of SGI is credited with coining the term in his research paper way back in 1998 (Lohr, 2013).

The most cited definition of Big Data is one of the earliest ones, which uses the 3Vs framework where the Vs are: Volume, Velocity and Variety (Laney, 2001) to identify which data is Big Data. Many corporations have their own definition of Big Data which includes defining it as a process to derive value from data in new ways (Perry, 2017) or in term of deriving value or quantifying it as a median of 300 TB of data in a week (MIT Technology Review, 2013) or going by the Big Data's impact on the consumers and society, some definitions acknowledge this role of Big Data and recognize it as a cultural and social phenomenon in addition to a technological one (Boyd & Crawford, 2012). Even though these definitions keep falling in and out of trend, the 3Vs framework seems to be the most commonly used one. The 3Vs soon became 4 with the addition of "Veracity" by IBM (IBM, 2012) and then became 5 with the Oracle adding another V which was "Value" (Oracle, 2012) and finally, became 7 with SAS adding the next V – Vulnerability (CXO Today, 2016). In common literature, these 7Vs are widespread. However, new developments have led to the expansion of this framework to 10Vs (Firican, 2017) with the addition of Variability, Validity and Visualization. The 10Vs are yet to become commonly accepted in the literature surrounding Big Data and its challenges and this was the first research gap identified.

There is also a lot of research around the challenges presented by Big Data. Many of these studies talk about the organizational and managerial challenges related to Big Data, which include problems with creation of data-driven decision-making culture and talent management (Baloun, 2012), a lack of understanding of how to use analytics to improve the business and lack of management bandwidth due to competing priorities (LaValle et al, 2010).

There are other reports that address the topic of technical challenges surrounding Big Data some of them identified as (Jagadish et al, 2014) data heterogeneity, data incompleteness and inconsistencies in data and lack of data privacy. Interestingly enough, these technical challenges correspond to a lot of the Vs in our 10 Vs

framework for Big Data thus presenting an interesting research opportunity to analyze Big Data challenges in the context of its 10 Vs characteristics. Some attempt has been made to decode these challenges in terms of Vs (Borne, 2014). However, the main drawback of these analyses is that the assumption that the challenges arising out of the Vs have the same impact across every industry which is not true as was found in our literature survey on this topic. Thus, analyzing the challenges of the 10 Vs in context of different industries to understand the problems associated with Big Data and their implications presents a good research opportunity and this is the primary motivation behind this paper.

3. Literature Survey

Big Data is starting to become yesterday's news. Even though IDC has forecasted that big data analytics market is all set be a $203 billion by 2020, analysts at Gartner are already forecasting that Big Data growth story is slowing down with the growth in market expected to decelerate from 63.6% in 2015 to about 19% by 2020 due to Big Data Analytics becoming mainstream (Gartner, 2017). The adoption rates also seem to be on the upswing with a Forrester survey last year indicating that almost 40% of firms are implementing and expanding big data technology while another 30% are planning to adopt it in the next 12 months (Forrester, 2016).

This news seems to suggest that with the technology and adoption headed towards mainstream, Big Data projects should also head towards productivity and see repeated success in companies. However, as mentioned earlier this is not true as Gartner has forecasted that through 2017, 60% of big data projects will fail to go beyond piloting (Gartner, 2015). Even assuming that there is a difference in maturity rates across industry with the same Forrester survey mentioned before (Forrester, 2016) claiming that market growth varies by industry with pharmaceutical, transportation, and primary production industries about to see the highest adoption in the next five years while professional services, telecommunications, government, and financial service sectors being the current highest users, it still doesn't explain the abysmally low rate of success quoted in the Gartner survey.

Thus, this anomaly suggests that we must explore reasons for failure of Big Data projects to get a better understanding. There is no dearth of blogs, news reports and opinions on the subject of Big Data failures as many of them have garnered a lot of negative publicity for companies using Big Data and have spooked many other companies in the industry too. A case in point is the fiasco at Amazon when an automated script at a T-shirt seller published T-shirts with slogans like "Keep Calm and Punch Her" (CIO, 2016).

A lot of corporate white papers have been published to elucidate the challenges and success factors associated with Big Data projects. In their white paper titled *"Big Data: Challenges and Success Factors"*(Deloitte Analytics, 2013), the consulting firm Deloitte identifies challenges around strategy, talent, scalability issues, integration problems and data quality and makes recommendations around these to provide a roadmap to Big Data success. Thus, the challenges faced can be classified into (a) Organizational – which would include vision, strategy and talent acquisition related and (b) Technical – which would include scalability, integration and quality problems. Similar discussion is found in many other reports, for example, in a Gartner blog "Big Botched Data", the author lists many organizational challenges like – management inertia, selecting wrong use of cases, asking the wrong questions, lack of right skills and disagreement on enterprise strategy (Svetlana Sicular, 2014, Gartner). Another example is the best practices whitepaper published by Dell – EMC which talks about aligning business and IT use cases and assessing your preparedness in terms of People, Process & Technology before embarking on Big Data journey (Dell – EMC, 2016).

These papers provide solid recommendations in terms of project management, organizational inputs etc. but do not elaborate the technical aspects of the challenges in much detail nor do they explore the differences in success of initiatives in different industries using Big Data technologies. Most of the data used for setting the context and arriving at recommendations is anecdotal and thus a quantitative framework seems to be missing.

Turning to academic literature on this subject, research papers take a more technical/ data issues oriented approach when talking about challenges of Big Data compared to the more organizational and management oriented approach of the corporate papers. The major types of challenges outlined are classified into *Data Challenges, Process Challenges and Management Challenges* (Thabet & Tariq, 2015). When talking about management challenges, a more general perspective

compared to organization oriented perspective is taken. The management challenges identified are related to data privacy, security and data governance. However, on a closer look one can see that these challenges are due to the intrinsic nature of big data itself and hence management needs to treat this data differently compared to traditional data and not because of any shifts in management paradigm.

Coming to the process challenges identified by the authors, they outline the process from data acquisition to data cleaning then integration to aggregation and finally interpretation. The challenges are identified at each step range from data fidelity issues at acquisition to visualization issues at interpretation stage. Thus, here also it can be said that the challenges are again due to the nature of big data being different from traditional data. Finally coming to data challenges, here the authors directly address the problems arising out of the properties of data and uses the "5 Vs" model and enumerate the problems due to Volume, Variety, Velocity, Veracity and Volatility of Big Data. It is important to note that the other challenges mentioned by him can be accommodated by just expanding the V's model to include Visualization and Vulnerability.

A similar analysis can be done for other literature on the subject where the authors identify challenges arising out of some Vs of Big Data – Volume, Variety, Velocity, Variability & Value while outlining other challenges like *Technical Issues*(fault tolerance, data quality issues (which can be accounted for if the Vs were expanded to include Veracity)), *Skill Requirements* and *Analytical Challenges* (Katal, Wazid and Goudar, 2013).

Sometimes the Vs are indirectly referenced in both academic and research papers. In one case, the challenges identified were classified into *Data Complexity, Computational Complexity and System Complexity* (Jin et al., 2015). Data complexity refers to complex types, structures and patterns – a characteristic which can be represented through the "variety" of Big Data. Computational complexity arises due to the multi-sources, huge volume, and fast-changing nature of Big Data – which is just another way to refer to Volume, Variety and Velocity properties.

Similarly, in a TDWI corporate research survey paper on Big Data Analytics (Russom, 2011), when the survey respondents (belonging to different industries and varying company sizes) were asked *"In your organization, what are the top potential barriers to implementing big data analytics?"*, the responses were as follows (Figure 1):

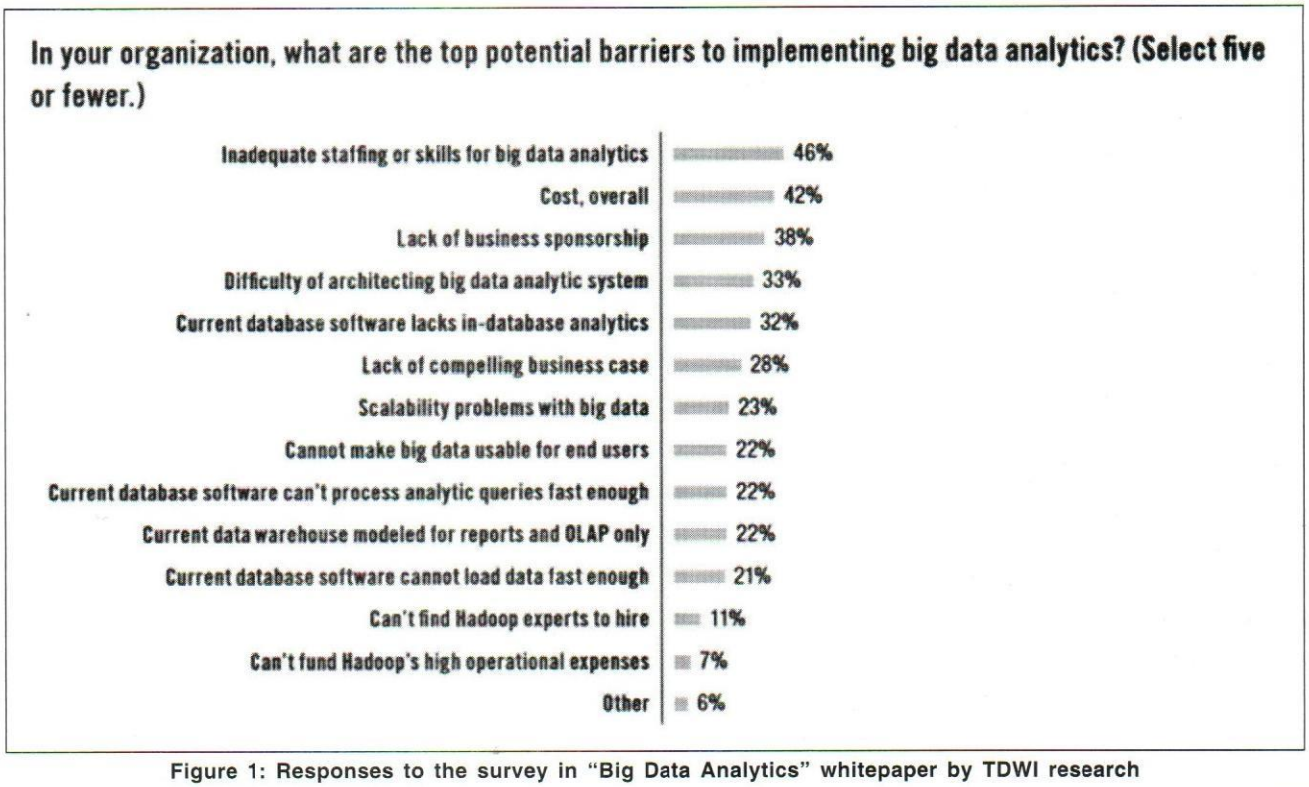


Figure 1: Responses to the survey in "Big Data Analytics" whitepaper by TDWI research

Here, the author identifies scalability issues, visualization issues (cannot make big data usable for end users) etc. under technical/ software issues thus, indirectly referencing some of the Vs. Thus, there is evidence of Vs being used across literature associated with Big Data challenges and issues and therefore, it provides a compelling framework for this analysis. However, since the Vs have been defined anecdotally (Ward & Barker, 2013), thus, any analysis based on the Vs framework will have a similar deficiency in that it will be based on anecdotes and subjective understanding of the concepts and not in quantitative terms.

When one goes ahead with the collection of data on analysis of Big Data challenges in terms of the Vs framework, a few problems are evident. The first is lack of any comprehensive literature, which use all of the 10 Vs mentioned before for their analysis. The authors have their own preferences for which V they consider relevant and some separate security, complexity and quality as other issues not within the Vs.

The other issue is the relative importance placed on some Vs over the others. For example, in the article *"Variety, Not Volume, Is Driving Big Data Initiatives"*(Bean, 2016) the author talks about the relative importance of variety of data in driving good outcomes from Big Data projects over the volume of data. The author says that the discourse on Big Data has unfairly focused on data proliferation and made Big Data synonymous with large volumes of data. In his research, he found that for many of the firms, it is not large data volumes, which is driving successful Big Data outcomes. Rather, it is the ability to integrate more sources of data — new data, old data, structured data, unstructured data, social media data, behavioral data etc. In the report, the survey showed that when asked about drivers of Big Data success, 69% of corporate executives named greater data variety as the most important factor, followed by volume (25%), with velocity (6%) trailing. Thus, as per the authors variety is the king.

But the story changes when it comes to the retail industry. As an industry whose competitive edge relies on the shortest time to market for new products, it might not be a good idea to wait around for integration of maximum number of sources to address the variety challenge. This is the gist of the article "*Big Data: Forget Volume and Variety, Focus On Velocity*" (Forbes, 2016) where the author believes that volume and variety aspects of Big Data receive the lion's share of attention but the velocity dimension of Big Data may have a bigger impact on your business and takes the example of Walmart's use of Big Data for accurate shelf stocking of candies during Halloween. The author believes that velocity is not just about the increased speed in which data is pouring into most organizations today but it also underscores the need to process the data quickly and, most importantly, use it at a fast rate for agile real-time decision making.

Similar observations can be made when the context of which industry is Big Data analytics being applied to changes. For example, consider healthcare, since the source and character of data here (i.e. medical data) is very different from the type of data you get from say retail (e.g. consumer sentiments from social media), the critical factors for success and challenges in front of projects in healthcare would rely on different Vs compared to the Vs which will be relevant in retail. In the article *"Understanding the Many V's of Healthcare Big Data Analytics"*(HealthIT Analytics, 2017) the author explains the importance of various Vs in context of healthcare. Vulnerability of data stands out from other Vs because of the various regulations and issues related to acquisition and storage of medical data. Thus, critical success of healthcare Big Data projects may be dictated by the organization's expertise in keeping the data secure and complying with regulations than with any other V and thus, vulnerability becomes the deciding factor for success.

For applications of Big Data to stock market or applications to a sector like telecom where data like prices and call records remain relevant for only that day or only that week, the dynamics change. Thus, volatility of data becomes the biggest factor to be accounted for when designing the system for Big Data applications in these industries. This was discussed in *"Taming Data Variety and Volatility is Key for Big Data Analytics"*(Lavastorm, 2012).

Going ahead we are staring at a connected, autonomous and exponential future. Machine learning and Artificial Intelligence in conjunction to Big Data technologies stand to deliver massive value across industries. When we use Big Data to solve equally big problems like autonomous cars, the veracity of data becomes sacrosanct. With highly publicized recent failures in this domain like Microsoft's bot Tay becoming racist after twitter users exposed it to such data (TechCrunch, 2016), it becomes imperative that such failure arising due to lack of veracity is avoided in future. In *"Data Veracity: Is it ok*

In summary, the following table lists the sources that were referred to during the literature survey and how they have shaped the thread of discussion in this paper:

Source	Information Obtained	Interpretation/ Understanding	Implication
Gartner (2017), *"Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $18.3 Billion in 2017"*	Growth in Big Data analytics market decelerating due to it becoming mainstream	Big Data is heading towards maturity	Big Data projects should have a high rate of success due to technology heading towards maturity however, this is not the case. Thus, it must be explored as to why the rate of failure is so high.
Forrester (2016), *"Forrester Forecasts Big Data Tech Market Will Grow ~3x Faster Than Overall Tech Market"*	Adoption rates of Big Data are high even though the market growth varies by industry		
CIO (2016), *"5 cases where big data was a big flop"*	Big Data projects at even cutting-edge technology companies are prone to spectacular failures.	These failures bring a lot of bad publicity and cause erosion of brand value and consumer trust	Thus, it is worthwhile to explore the reasons behind Big Data failures to arrive at action points for managers and organizations
TechCrunch (2016), *"Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism"*			
Deloitte Analytics (2013), "Big Data: Challenges and Success Factors"	Identifies challenges around strategy, talent, scalability & data quality and provides recommendations around that to provide a roadmap for Big Data project implementation.	Challenges around Big Data are not just technical but also organizational and managerial.	Some Big Data initiatives will be more successful than others and subjectivity due to organizational and managerial related factors makes it difficult to arrive at a quantitative framework for key success factors.
Svetlana Sicular (2014), "Big Botched Data", Gartner	The author highlights management inertia, silo approach and solution avoidance as reasons for failure of projects.		
DELL-EMC (2016), *"Five Critical Success Factors For Big Data Projects"*	Talks about aligning business and IT use cases along with assessing readiness of People, Technology and Processes as important success factors.		
Thabet & Tariq (2015), *"Big Data Challenges"*	The author talks about data challenges, process challenges and management challenges associated with Big Data	Most of these challenges arise out of the inherent properties of Big Data that makes it different from traditional data.	Thus, a better analysis of the challenges should be done along Big Data's nature because that is what is causing all the difference instead of any process specific or management specific problem, i.e. if you change your management the management challenges won't disappear because they are there due to the nature of Big Data.
Katal, Wazid & Goudar (2013), *"Big data: Issues, challenges, tools and Good practices"*	Authors use Volume, Variety, Velocity, Variability and Value to describe challenges of Big Data along with technical challenges and analytical challenges.		
Jin et al. (2015), *"Significance and Challenges of Big Data Research"*	The author provides a framework for understanding challenges associated with Big Data by dividing them into three types of challenges – Data Complexity, Computational Complexity & System Complexity.	The three types of challenges identified by the author incorporates some of the Vs of Big Data amongst other issues but not all.	The challenges associated with Big Data have also not been assessed in context of different industries in many of the literature surrounding the subject.

Source	Information Obtained	Interpretation/ Understanding	Implication
Philip Russom (2011), *"Big Data Analytics"*, TDWI Research	In a cross-industry survey of companies of varying sizes, the author identified Staffing Problems, Business sponsorship problems and software adequacy issues as the major barriers to implementing big data analytics.	The software inadequacy barriers identified were related to the inability of software to handle scale of data, frequency of change etc. These are the characteristics of big data itself and not a software issue to be thought of as a "technical difficulty".	The challenges in successful implementation of projects which may arise due to the inherent nature of big data itself may mistakenly be thought of as software issues instead of being accounted as separate risks.
Ward, J. S. & Barker, A. (2013), *"Undefined By Data: A Survey of Big Data Definitions"*	There is not a single unified definition of Big Data as its co-evolving from industry, academia and media. Under many definitions it is not clear as to exactly when the term big data becomes applicable it rather provides a means to "know it when you see it".	In most of the literature surrounding Big Data, anecdotal evidence is used to define its characteristics.	There is inherent ambiguity in any exploration of big data's properties in terms of Vs and as such no quantitative frameworks exist to clearly define them. Thus, qualitative analysis and subjective understanding of the concepts has to be applied here.
Randy Bean (2016), *"Variety, Not Volume, Is Driving Big Data Initiatives"*, MIT Sloan	As per the author, even though Big Data is synonymous with volume, it is the variety of data which is the driving force between Big Data initiatives across industries.	In many projects, it is the sheer variety of sources to be integrated which dictates the effectiveness of results instead of the volume of data.	Some authors place more emphasis on a certain V over other Vs of Big Data.
Brent Dykes (2017), *"Big Data: Forget Volume and Variety, Focus On Velocity"*, Forbes	The author uses the example of Walmart's use of Big Data for product stocking on festivals to argue that velocity of data in terms of real-time alerting is more important than either volume or variety.	In retail environment, especially the velocity of data maybe of the greatest value compared to other characteristics (other Vs) of Big Data and thus, managing the velocity of data may be the CSF for the successful implementation of a Big Data projects.	Different industries place different importance on the various Vs of Big Data and thus the critical factors for success in different industries may change depending on which V is more critical.
John Joseph (2012), Lavastorm, *"Taming Data Variety and Volatility is Key for Big Data Analytics"*	Author talks about the importance of Volatility in Telecom and Financial Services sector.		
Jennifer Bresnick (2017) "Understanding the Many V's of Healthcare Big Data Analytics", Health IT Analytics	The author expands the 4Vs Big Data framework to mention the challenges arising out of more Vs (especially Vulnerability & Visualization) due to the nature of healthcare industry. The author makes a case for	Big Data implementation in Healthcare industry faces more challenges than can be explained with just the 3Vs or 4Vs framework.	Later additions to the Vs concept like Visualization and Vulnerability can also significantly impact the value from Big Data initiatives in different industries like healthcare compared to other industries.
Srikanth Kintali (2016), *"Data Veracity: Is it ok to overlook?"*	The importance of data veracity especially as we embark towards a connected future with autonomous devices	Veracity is crucial for the success of any Big Data model.	Veracity is tied to its data source so some industries maybe more susceptible to veracity issues if the data collected by them is from unreliable/ difficult to validate sources when compared to other sources.

to overlook?" (Kintali, 2016), the author makes a case for Veracity being the core tenet at the heart of any Big Data initiative. In conclusion, it can be said that when the industry context changes, some Vs become more important than others and dictate the success of Big Data outcomes. Thus, this paper aims at exploring the 10 Vs of Big Data (Volume, Variety, Velocity, Veracity, Variability, Validity, Visualization, Vulnerability, Volatility and Value) in context of 5 industry verticals (Healthcare, Telecom, Financial Services, Consumer Products and Airlines) to see which Vs are more important and relevant in the context of each industry.

4. Research Methodology

As mentioned previously, the discussion surrounding Big Data definition is beset by a lack of quantitative basis. The definition of Big Data is ever evolving and academicians, analysts and industry experts all co-contribute to extend the understanding of this subject. Because of lack of quantitative methods for definition and because the extension of the definition is done by industry experts based on their experience with the success and failure of Big Data project implementation, the justifications and explanations around the Vs of Big Data are based on anecdotes and narratives from the industry.

In the previous section on literature survey, following three gaps were identified –

- When discussing challenges arising out of Big Data, the 10 Vs framework has not been used comprehensively.
- When the impact of the Vs is discussed, many authors rank one V over the other in importance based on their experience with a certain situation, use or applications. These situations are heavily dependent on the context of the industry they have been taken from.
- The impact of 10 Vs in the context of different industries has not be analyzed.

Based on the above questions, the following research question was identified:

"What are the ways in which the 10 Vs impact the feasibility, success and outcomes of Big Data initiatives across different industries and what technical, management and organizational implications can be drawn from them?"

For the purpose of this paper, the number of industries has been restricted to 5 – Healthcare, Telecom, Financial Services, Consumer Products and Airlines.

Since, the nature of the topic at hand is theoretical and descriptive, therefore an exploratory research based on secondary research was chosen to be the research methodology. The aim of this paper is not to test the impact against outcomes of Big Data initiatives in reality but merely to explore the possibilities and raise questions, which need to be answered in context of different industries when one embarks on a Big Data initiative. An exploratory study is most suitable when "how" questions are to be explored and thus, was the chosen methodology here. Since there is a plethora of information available about Big Data and as academicians as well as industry personnel, all contribute to the expansion of the understanding of this subject, a secondary data research was preferred over a primary one.

The data sources referred for secondary research include: Corporate whitepapers, Industry reports, Experts' blogs, News reports, Research papers published in journals, Magazine articles.

5. Analysis

The 5 industries chosen for this research paper are: Healthcare, Financial Services, Telecom, Airlines and Consumer Products. The data and analysis is not restricted to any one geographical location for any of these industries.

The first step in the analysis would be to identify the various sources of data in each of the industries to be discussed. The sources would dictate the influence of particular Vs in the industry and is hence an important step. After identification of the sources, each of the 10 Vs would be discussed in the context of each of the 5 industries.

5.1 Sources of Data

Many frameworks have been used in literature to classify the data sources for Big Data. Simple division based on the structure of data: *Structured, Unstructured and Semi-Structured* (Assuncao et al, 2014). Structured Data refers to the data which is already stored in databases, in an orderly manner. It accounts for about 20% of the total existing data. Unstructured data has no clear format during its storage and accounts for the bulk of the data (around 80 %) being generated. Semi-structured data

can be defined as information that is not in the traditional database format (like in the case of structured data) but contains some organizational properties which make it easier to process/retrieve. For example, e-mails as you can classify/search on the basis of date/timestamp, subject, sender etc.

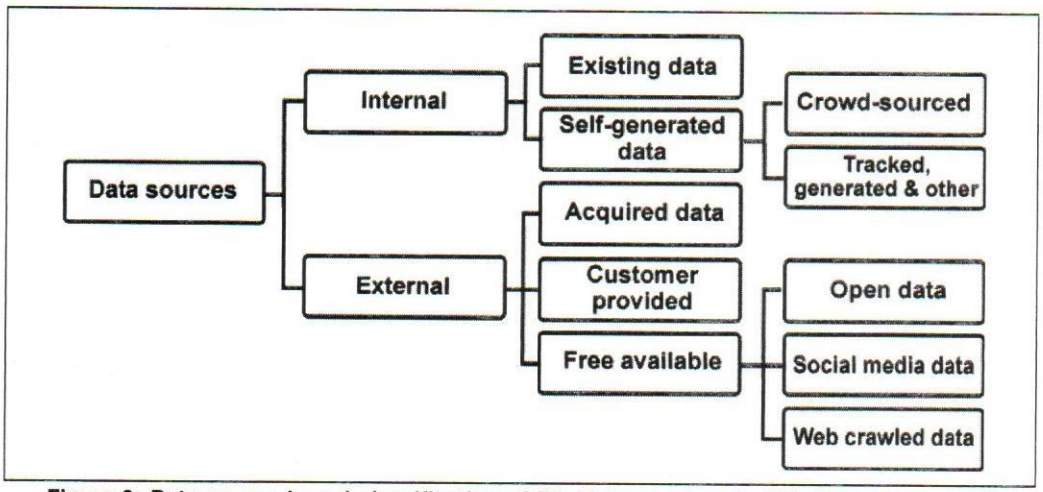


Figure 2: Data source based classification of Big Data as given in "Big Data Taxonomy"

Other types of taxonomies are also available where the classification is done based on the nature of source. One such classification is as follows in Figure 2 (Hartmann et al., 2014) –

- **Internal Sources:** include data that already exists in or is created through existing IT systems (ERP, CRM, etc.) within the enterprise.
- **External Sources:** include data acquired from data providers, freely available data and data shared by the customer.

Another possibility to classify the data on the basis types of sources from which data originates is as follows –

- **Sensors/meters data from electronic devices:** is produced real-time, the number and periodicity of observations is variable - depending on time or on the occurrence of some event. Quality of this data depends mostly on the capacity of the sensor to take accurate measurements.
- **Social data:** is produced by human interactions through a network. E.g. Social networks. This data is most commonly used for sentiment and trend analysis and the quality of the data depends on the accuracy of the algorithms designed to extract meaning from it.
- **Business transaction data:** is produced because of business activities and can be recorded in a structured or an unstructured way.
- **Electronic Files:** These refer to unstructured documents which are stored as electronic files, like Internet pages, PDF files, etc.
- **Broadcastings:** Mainly refers to video and audio produced in real time.

For this paper, I have chosen to classify the data sources into a matrix format with structured, semi-structured and unstructured being the rows and internal and external classification being the columns of the matrix.

Sources of Data – Healthcare

	Internal Data	External Data
Structured	Clinical Information Systems - Patient Registries	Clinical trial/ research data, Public records, Government Databases, Insurance claims data
Semi- Structured	Electronic health records (EHRs)	Web searches, Genetic Data
Unstructured	MRIs, Scans, handwritten notes	Sensor data from devices

Sources of Data – Financial Services

	Internal Data	External Data
Structured	ATMs, internet/ mobile banking data, credit card transactions, loan applications, CRM systems	Trading data, credit rating data
Semi- Structured		Web searches, Analyst and Industry reports
Unstructured	Call centers data	Online forums, social media, News

Sources of Data – Airlines

	Internal Data	External Data
Structured	Equipment/Assets data	Customer purchase history – credit card data, customer location data (arrival and departure time and city)
Semi- Structured	Flight logs, sensors	Web searches
Unstructured	Call centers data	Weather & Geo spatial data, social media

Sources of Data – Telecom

	Internal Data	External Data
Structured	Equipment/Assets data	Customer spending pattern, preferred services (voice/ data/VAS etc.), customer location data, Device information (model, series, technology used)
Semi- Structured	Cellular network data, sensors data	Web searches, Call logs, Apps usage data
Unstructured	Call centers data	Social media

Sources of Data – Consumer Products

	Internal Data	External Data
Structured	Point of Sale data, Supply chain data - Shipment data (manufacturers to distributors, distributors to customers), stocks & orders data	Customer purchase history – loyalty programs data, customer location data (arrival and departure time and city)
Semi- Structured		Web searches, Market Research Data, Customer Surveys data
Unstructured		Social media data

The above compilation of various sources of data will be helpful in the analysis of the 10Vs for these industries.

5.2 10 Vs Analysis

Starting with our analysis of the 10 Vs of Big Data in the context of the previously mentioned industries –

5.2.1 Volume

Big data first and foremost has to be "big" and size in this case is measured as volume. *How big is "big"?* Some experts consider it to be data volume of above 5PB however there is not much consensus on this as the digital universe is expected to grow nearly 20-fold, to

approximately 35 zettabytes of data, by the year 2020 and thus, "big" for big data continues to become "bigger".

- The volume in healthcare industry is expected to grow from 53 exabytes in 2013 to 2,314 exabytes by 2020. (CIO, 2014)
- In the financial services industrya massive push towards digitization post demonetization in India is making over 1.7 billion people with mobile phones, that are currently excluded from the formal financial system, increasingly visible and their data is also valuable (Siddiqui & Qureshi, 2017).
- Coming to the telecom industry, the following should give an idea about the scale of data being generated –
  - There are more mobile devices in the world than people (Independent, 2014)
  - The average person interacts with their mobile phone 1500 times a week (Dailymail, 2014)

Similar is the story for Airlines and Consumer Products industry. To summarize, the volume of data is very high in all the 5 chosen industries:

Industry Vertical	Volume
Healthcare	Very High
Airlines	Very High
Telecom	Very High
Financial Services	Very High
Consumer Products	Very High

5.2.2 Variety

Variety refers to both different types & different sources of data. Integrating different variety of data requires use of multiple tools thus increasing costs & complexity. The more unstructured data an industry collects, chances are that the variety of the data will be very high as it will take significant effort in consolidating that data to get one view. In a Big Data study conducted by TCS, (TCS,2013) the company found that some industries are heavy users of unstructured data and external sources compared to other industries. The results of the study are summarized in figure 3 below:

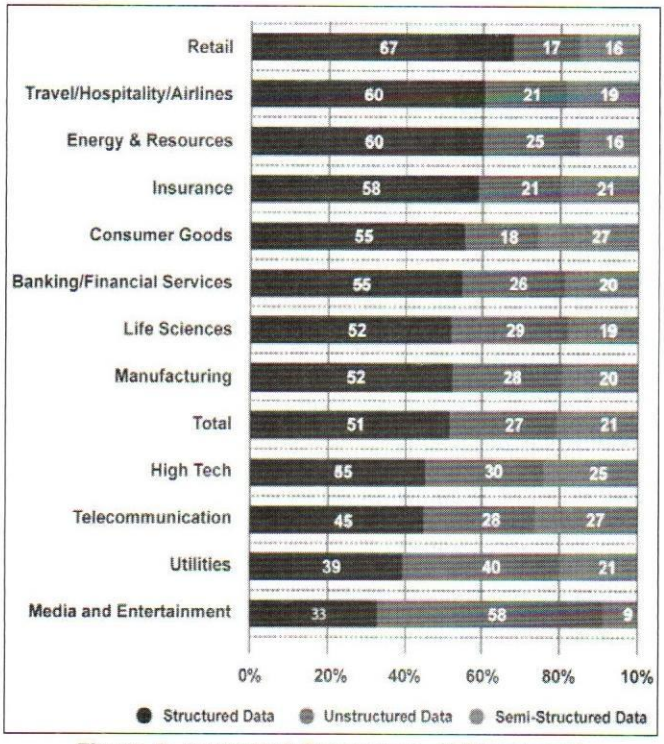


Figure 3: Estimated Percentage of Structured, Unstructured and Semi-Structured Data in TCS Study

From the above, the following conclusions can be drawn:

Industry	Structured Data	Unstructured & Semi structured Data	Variety
Healthcare	Lower compared to others	Higher compared to other industries	Very High
Airlines	High	Low	Moderate
Telecom	Lowest compared to other 4	Highest compared to other 4	Very High
Financial Services	Higher compared to airlines and retail	Lower compared to telecom and Financial services	High
Consumer Products	Highest compared to others	Lowest compared to others	Low

5.2.3 Velocity

Velocity is the measure of how fast the data is coming in. Some sources like social media are called "firehose" sources because of the huge amount of data generated per second. Velocity underscores the need to process the data quickly and thus has cost implications. Many types of data have a limited shelf-life where their value can erode with time—in some cases, very quickly. For example, in retail it's better to know which products are out-of-stock in terms of seconds or minutes rather than days or weeks. The more quickly a retailer can restock the products, the faster it can return to generating product sales. In *"Big Data Taxonomy"* (Cloud Security Alliance, 2016), the author provides the following framework (Figure 4) to understand velocity:

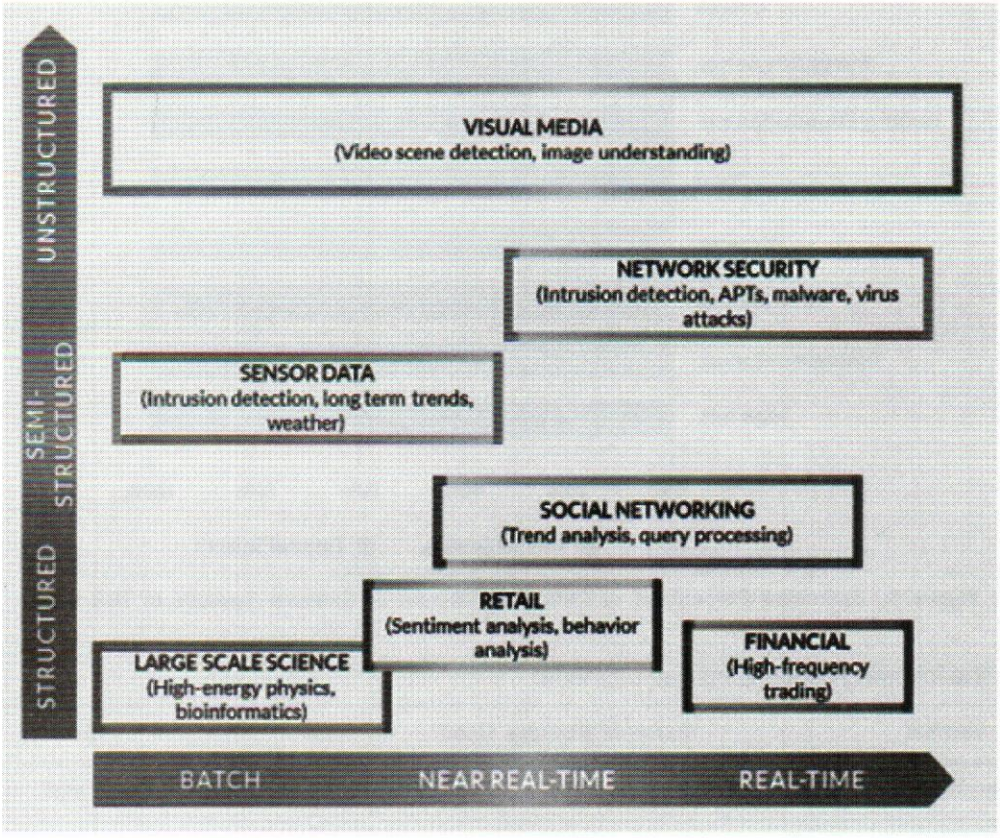


Figure 4: Velocity across verticals in Big Data Taxonomy

On the basis of the above, following conclusions can be made:

Industry Vertical	Velocity
Healthcare	Low
Airlines	Moderate
Telecom	Very High
Financial Services	Very High
Consumer Products	High

5.2.4 Veracity

Veracity refers to reliability of the data and its source, its context, and how meaningful it is to the analysis based on it. It is more difficult to gauge the veracity of external data sources compared to sources which are internal to the company (as the control is more in the case of latter). Thus, if an industry uses more external sources for the purpose of its Big Data analytics, there are more chances of the veracity of the data being low. In the previously mentioned Big Data Study by TCS (TCS, 2013), the study also saw which industries used more external sources than others, the results of which are given in figure 5.

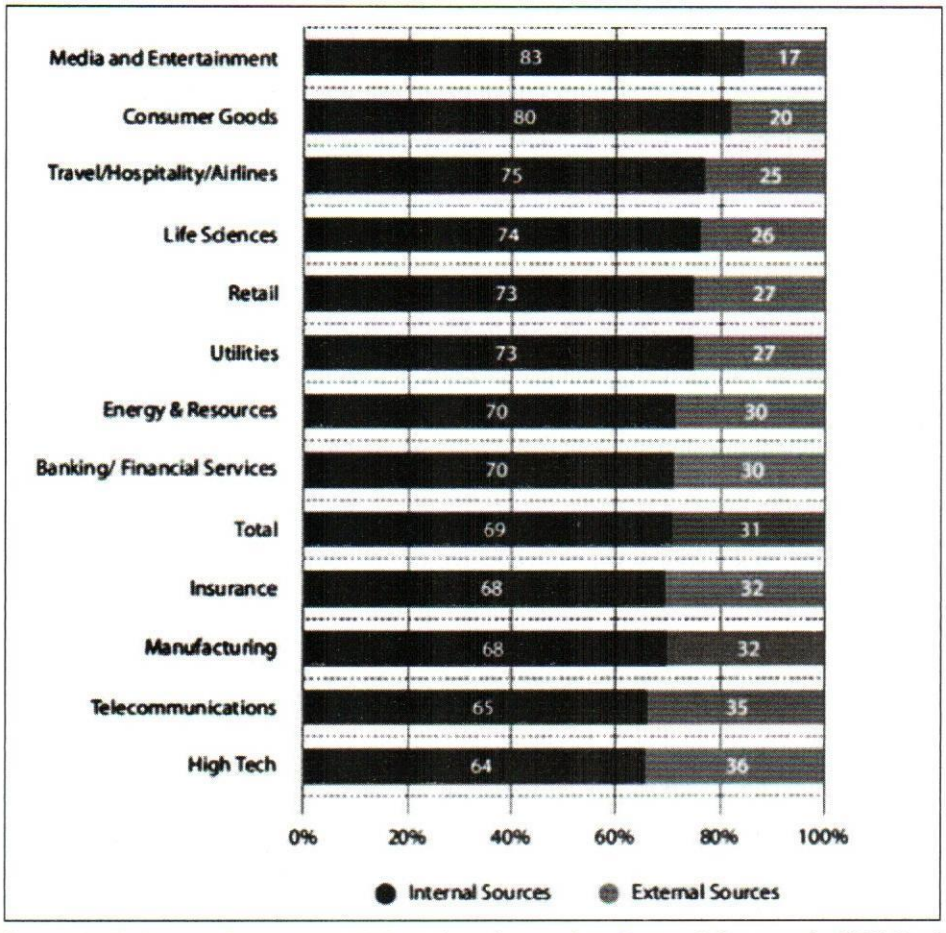


Figure 5: Estimated Percentage of Data from Internal or External Sources in TCS Study

From the above, the following conclusions can be made:

Industry Vertical	External Sources Used	Veracity
Healthcare	Low	High
Airlines	Lowest	Highest
Telecom	Medium	Moderate
Financial Services	Higher	Low
Consumer Products	Medium	Moderate

5.2.5 Variability

Variability is used to measure the consistency of incoming data over time. Variability in Big Data's context refers to a few different things:

- One is the number of inconsistencies in the data. These need to be found by anomaly and outlier detection methods in order for any meaningful analytics to occur.
- Big data is also variable because of the multitude of data dimensions resulting from multiple disparate data types and sources.
- Variability can also refer to the inconsistent speed at which big data is loaded into your database.

Industry	Variability
Healthcare	High - as a lot of the data is unstructured. Seasonal diseases etc. can also add to the variability of data
Airlines	Moderate - Customer data from social sources may be highly variable but flight data is fairly streamlined except for in situations like bad weather, engine problems etc.
Telecom	Moderate - due to variation in the flow of data - iit may be inconsistent with periodic peaks, Daily, seasonal or event-triggered data loads may be challenging to manage.
Financial Services	High – e.g. stock market data will be variable based on the prevailing sentiments etc.
Consumer Products	Very High – due to fluctuation demands, variation in marketing efforts, seasonal sales, inflow of unstructured sentiment data.

5.2.6 Visualization

All industries need good visualization tools to interpret the data. However, visualization might be more challenging in the cases where there is more unstructured data and more number of sources compared to those industries where structured data is more. Therefore, we use the conclusions drawn from the section on variety to arrive at the conclusions for visualization.

Industry	Variety	Visualization
Healthcare	Very High	Very Challenging
Airlines	Moderate	Moderate
Telecom	Very High	Very Challenging
Financial Services	High	Challenging
Consumer Products	Low	Less Challenging

5.2.7 Vulnerability

Big data vulnerability is all about the security measures required to be put in place so that the collected data is stored in compliance with regulations as well as customer's wishes.

Industry	Vulnerability
Healthcare	Extremely vulnerable – Patient reports are confidential and no PIIs should be used. Vulnerability exacerbated by the fact that the data need to be stored for a long time.
Airlines	Extremely vulnerable – as airlines are a big target for terrorism
Telecom	Extremely vulnerable – as call records etc. prone to hacking
Financial Services	Extremely vulnerable – subject to lot of regulations & has PII
Consumer Products	Vulnerable – consumers might not appreciate use of their information for advertising

5.2.8 Validity

Validity refers to the quality of data and the accuracy for the intended purpose for which it was collected. According to Forbes, an estimated 60% of a data scientist's time is spent cleansing their data before doing any analysis.

Since validity of data is related to its intended purpose thus it cannot be studied in isolation of its use or intended application in any industry context. For example, if social media sentiment data is used to judge the effectiveness of a YouTube advertisement then there are high chances that the data is valid, however, if the intended target market of the product is not on social media then the same data becomes invalid.

5.2.9 Volatility

The volatility characteristic of Big Data answers the question – *"How long do you need to store this data?"*. If data needs to be stored for a long time (for say regulatory compliance purposes) then it has a storage cost implication, however, if the data is required near real-time or it becomes irrelevant then it has a computational cost implication. We use the framework used in Figure 5 of the Velocity section to arrive at the following conclusions for Volatility:

Industry	Volatility
Healthcare	Low - As per regulations, healthcare data need to be stored for a long time. Healthcare data is less volatile as patient's entire medical history may be relevant in predicting future illness.
Airlines	Low - Data is less volatile as Airlines must manage multiple sources of data to comply with legal and industry requirements, and anti-terrorism measures. Thus, data needs to be stored for a long time.
Telecom	Moderate – usage data etc. are volatile but regulations require storage of certain data for a long time
Financial Services	High – customer credit history, company reports etc. are less volatile but day to day trading data is quite volatile
Consumer Products	Very High – consumer preferences shift quickly, stocking data needs to be updated atleast hourly, mostly sales data of up to one or two years is only considered relevant

5.2.10 Value

In the report titled *"Big Data: The next frontier for innovation, competition and productivity"* (McKinsey, 2011), the authors have suggested that some sectors are set to gain much more from applications of Big Data compared to some other sectors. The results of the report are summarized in the figure 6 below:

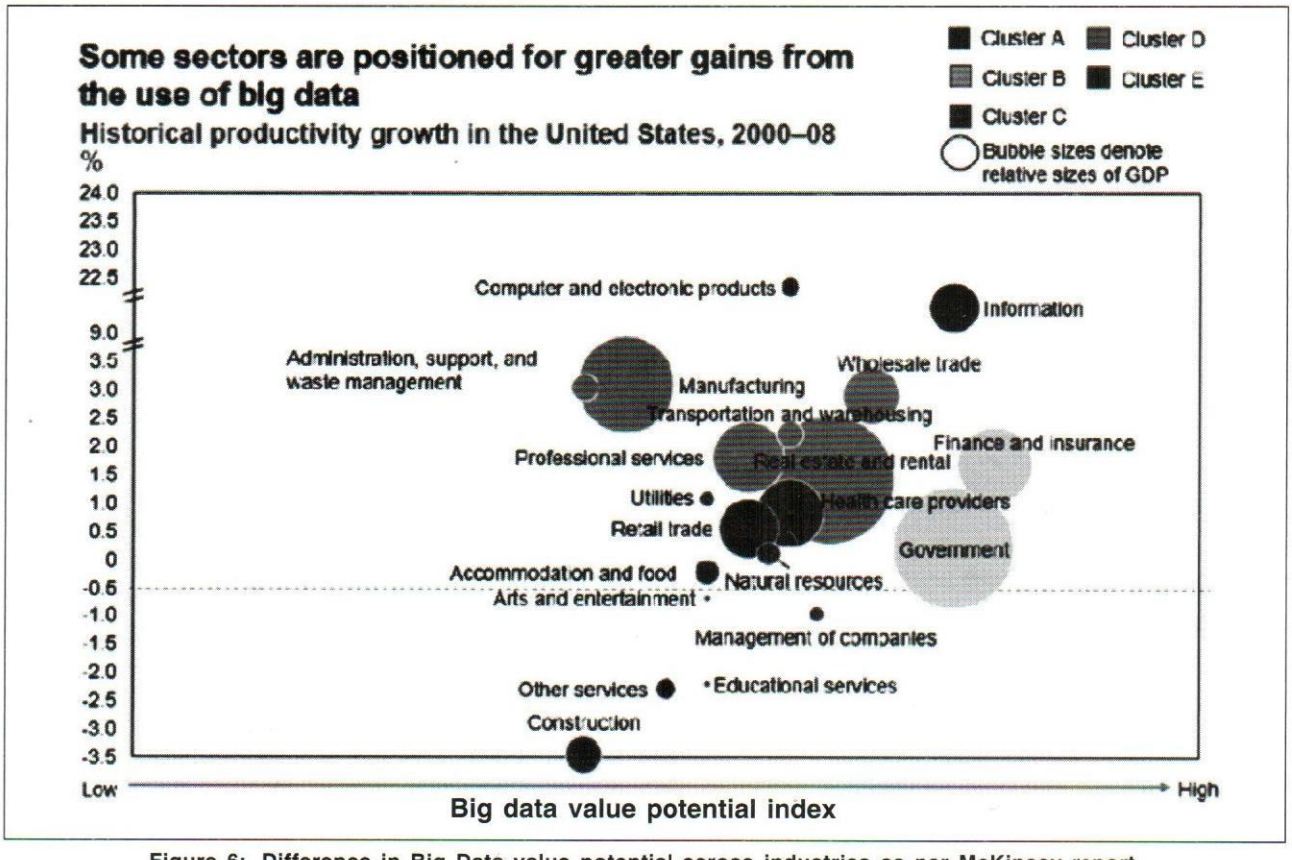


Figure 6: Difference in Big Data value potential across industries as per McKinsey report

Using the results, the following can be concluded:

Industry Vertical	Value Potential
Healthcare	High
Airlines	Moderate
Telecom	Very High
Financial Services	Very High
Consumer Products	Moderate

The above analysis, gives the relative importance of the different Vs of the 10Vs model of Big Data in context of five different industries are extremely important and provided in table below.

Industry	Volume	Velocity	Variety	Variability	Veracity	Validity	Vulnerability	Volatility	Visualization	Value
Healthcare	Very High	Low	Very High	High	High	High	Extremely Vulnerable	Low	Very Challenging	High
Airlines	Very High	Moderate	Moderate	Moderate	Very High	High	Extremely Vulnerable	Low	Moderate	Moderate
Telecom	Very High	Very High	Very High	Moderate	Moderate	Moderate	Extremely Vulnerable	Moderate	Very Challenging	Very High
Financial Services	Very High	Very High	High	High	Low	Moderate	Extremely Vulnerable	High	Challenging	Very High
Consumer Products	Very High	HIgh	Low	High	Moderate	Low	Vulnerable	Very High	Less Challenging	Moderate

6. Implications

6.1 Technical Implications

The following are the technical implications of the previous analysis of 10 Vs in the context of 5 industry verticals:

- As per the previous analysis, volume is and will be an ubiquitous feature of data across all industries and that the data is growing at a faster rate than what storage technologies are and therefore, challenges rising as a consequence of ever growing volume are going to continue plaguing Big Data initiatives. These challenges are: the complete overwhelming of existing IT systems and infrastructure currently existing in enterprises, long run-time for execution of aggregation queries and lags in report generation. Until in-memory technologies become sufficiently effective at handling such data loads, the reliance on cloud infrastructure is going to continue.
- Some of the challenges arising as a consequence of volume can be mitigated if other Vs are also brought in the picture. For example, if in Consumer Products despite volume being very high, if volatility of data is also very high, then all data may not be required to be stored thus reducing some problems arising out of volume. If decision making is strongly being influenced by the volatility and velocity of data then in-memory technologies may become more suitable if only the relevant amount of data set is used instead of running analytics on the entire data set.
- There may be a relation between the amount of variety of data and the difficulty of visualization in that industry. Thus, in case of Healthcare where the variety is high and visualization is quite challenging, it will be better to rely on niche visualization tool providers who have an expertise in that industry despite of the risks instead of opting for more mature providers who provide broad based solutions across industries.
- High velocity of data also brings with itself loads of challenges especially when it comes to loading the data in the system through batch jobs as the high velocity of data completely overwhelms the system and more computational resources will be spent on loading the data than in running analytics. Consider the case of Telecom from the summary of analysis,

where even though the velocity of data is very high, the variability of data is not that high, thus, it may be possible to use effective sampling techniques to run analytics on sample of incoming high velocity data set instead of running it on the entire set as the variations in data is less.

- Vulnerability of data is critical across all industries and it is making cloud with all its security and compliance risks an increasingly less attractive proposition compared to storing it in-house but as storage techniques are not keeping pace and in-memory technology remains expensive, this dichotomy poses a significant problem to the success of Big Data initiatives and may explain the high rate of failures and abandonment of pilot projects.

6.2 Organizational & Managerial Implications

The following are the organizational and managerial implications of the previous analysis of 10 Vs in the context of 5 industry verticals:

- Big Data initiatives require significant investment in buying new software, hiring expert resources, getting management commitment and diverting resources to work on pilot projects. All this has significant cost implications and the success rate of such initiatives is low even though, when they succeed the business benefits are immense. However, not all industries are going to gain equal benefits, some will face less problems and bring in more benefits compared to others. So, as per the previous analysis, if one finds oneself in too many red zones in the 10Vs framework, unless the value is extremely high (as in the case of Financial Services), it is better to look for alternate analytics solutions than going on board the Big Data train.
- Validity of data is an extremely important condition and must be taken into account even if all the other Vs are in the green-yellow zone. Validity is about whether the data being considered is right for the question being asked or not and as such can make or break the business use case one has chosen to pilot their Big Data initiative on as there are no right answers to a wrong question. Thus, systems and processes should be in place to check the validity of data even when the veracity is high and variety and variability is low.
- As the Big Data framework and the Vs model continues to expand, managers must be on the continuous lookout for those new Vs (and Cs and Ts) which may be specific to their industry and can have a big impact on the success outcome of their initiatives. This point is also elaborated in the way ahead section which follows.

7. Limitations & Way Ahead

One of the major limitations of this analysis is lack of quantitative framework to test the data analysis. From the beginning the 10 Vs model is based on subjective understanding of Big Data by the experts and thus any analysis based on this model becomes subjective and anecdotal in nature. The second limitation is that the discussion has been limited to 10 Vs definition only while the V's framework has been expanded much more beyond the 10 Vs discussed here. In fact, the Vs model has been expanded up to 42 Vs (Elder Research, 2017). However, this limitation is mitigated by the fact that many of the new Vs are yet to be commonly accepted by industry experts and academicians alike and many Vs of the 42 Vs can be clubbed together and hence a factor analysis for the same would have been required.

The third limitation of this research arose during the analysis phase when a suitable framework for the impact of Big Data Validity could not be found because as a construct validity could not be separated from the use case defined for it and thus, the data that is valid in one application may or may not be valid in another application in the same industry context.

Finally, the way ahead for this research would rely on the streamlining of Big Data definition. The Vs framework has also proven to be deficient and many new extensions like C's (Complexity, Cost, Coverage) and T's (Tools, Techniques etc.) are being added to the Big Data definition. Further as adoption rates increase and more projects go beyond the pilot phase so that their success/failure rates and RoIs can be calculated, there will be more scope to do analysis with a quantitative basis once such data is available in the industry.

References

Brynjolfsson E., McAfee A (2011), "The big data boom is the innovation story of our time", *The Atlantic,* http://www.theatlantic.com/business/archive/2011/11/the-big-data-boom-is-the-innovation-story-of-our-time/248215/.

HBR (2012), "Big Data: The Management Revolution", *Harvard Business Review,* https://hbr.org/2012/10/big-data-the-management-revolution

Randy Bean (2017), "How companies say they are using Big Data", *Harvard Business Review,* www.hbr.org/2017/04/how-companies-say-theyre-using-big-data

Information Age, (2015), "Big data's big failure? Buried info costs businesses £20M in missed opportunities", http://www.information-age.com/big-datas-big-failure-buried-info-costs-businesses-20m-missed-opportunities-123460598/

CIO, (2016), "Big data and analytics spending to hit $187 billion", https://www.cio.com/article/3074238/analytics/big-data-and-analytics-spending-to-hit-187-billion.html

Chatterjee, Pratap (2013), "Big Data: the greater good or invasion of privacy", The Guardian, https://www.theguardian.com/commentisfree/2013/mar/12/big-data-greater-good-privacy-invasion

Time, (2014), "Google's Flu Project Shows the Failings of Big Data", http://time.com/23782/google-flu-trends-big-data-problems/

Gartner (2015), "Gartner Says Business Intelligence and Analytics Leaders Must Focus on Mindsets and Culture to Kick Start Advanced Analytics", https://www.gartner.com/newsroom/id/3130017

Data Science Central, (2015), "Big Data Falls Off the Hype Cycle", www.datasciencecentral.com/profiles/blogs/big-data-falls-off-the-hype-cycle

Oliver (2015), "Big data is dead – long live big data", Infoworld, https://www.infoworld.com/article/2907872/big-data/big-data-is-dead-long-live-big-data.html

Datafloq.com, (2017), "A short history of Big Data", https://datafloq.com/read/big-data-history/239

Steve Lohr (2013), "The Origins of 'Big Data': An Etymological Detective Story", *NYTimes,* https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

Laney Doug (2001), "3D Data Management : Controlling Data Volume, Velocity, Variety", Gartner Blog, https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Steve Perry (2017), "What is big data? More than volume, velocity and variety", IBM Blog, https://developer.ibm.com/dwblog/2017/what-is-big-data-insight/

MIT Technology Review (2013), "The Big Data Conundrum: How to define it?", https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/

Danah Boyd and Kate Crawford (2012) 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon', Information, Communication & Society, Volume 15, 2012 - Issue 5

IBM (2012), "What is Big data", https://www-01.ibm.com/software/in/data/bigdata/

Oracle (2012), "Oracle: Big Data for the Enterprise", whitepaper, http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

CXO Today (2016), "Vulnerability Is The Most Concerning 'V' Of Big Data", http://www.cxotoday.com/story/vulnerability-is-the-most-concerning-v-of-big-data/

George Firican (2017), "The 10 Vs of Big Data",https://upside.tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

Brett Baloun (2012), "Big Data – Organizational Challenges", https://blogs.perficient.com/dataanalytics/2012/10/07/big-data-organizational-challenges/

Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz (2010), "Big Data, Analytics and the Path From Insights to Value", *MIT Sloan Management Review,* https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/

H. V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, Cyrus Shahabi (2014), "Big Data and Its Technical Challenges", *Communications of the ACM,* Vol. 57 No. 7, Pages 86-94

Kirk Borne (2014), "Top 10 Big Data Challenges – A Serious Look at 10 Big Data V's", https://mapr.com/blog/top-10-big-data-challenges-serious-look-10-big-data-vs/

Gartner (2017), "Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $18.3 Billion in 2017", Gartner report summary, https://www.gartner.com/newsroom/id/3612617

Forrester (2016), "Forrester Forecasts Big Data Tech Market Will Grow ~3x Faster Than Overall Tech Market", Forrester report summary,www.forrester.com/Forrester+Forecasts+Big+Data+Tech+Market+ Will+Grow+3x+Faster+Than+Overall+Tech+Market/-/E-PRE9484

Noah D'Mello (2016), CIO, "5 cases where big data was a big flop", http://www.cio.in/feature/5-cases-where-big-data-was-big-flop

Deloitte Analytics (2013), "Big Data: Challenges and Success Factors", https://www2.deloitte.com/content/dam/Deloitte/it/Documents/deloitte-analytics/bigdata_challenges_success_factors.pdf

Svetlana Sicular (2014), "Big Botched Data", Ga[illegible]r, https://blogs.gartner.com/svetlana-sicular/big-botched-data/

DELL-EMC (2016), "Five Critical Success Factors For Big Data Projects", http://en.community.dell.com/techcenter/blueprints/data_analytics/m/mediagallery/20443633

Thabet, Nasser & Soomro, Tariq (2015), "Big Data Challenges", *Journal of Computer Engineering & Information Technology*

A. Katal, M. Wazid and R. H. Goudar, (2013) "Big data: Issues, challenges, tools and Good practices," Sixth International Conference on Contemporary Computing (IC3), Noida, 2013, pp. 404-409.

Jin, Xiaolong & Wah, Benjamin & Cheng, Xueqi & Wang, Yuanzhuo, (2015), "Significance and Challenges of Big Data Research", *Big Data Research*

Philip Russom (2011), "Big Data Analytics", TDWI Best Practices Report, TDWI Research, https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx

Stuart Ward, Jonathan & Barker, Adam, (2013), "Undefined By Data: A Survey of Big Data Definitions"

Randy Bean (2016), "Variety, Not Volume, Is Driving Big Data Initiatives", MIT Sloan Management Review, https://sloanreview.mit.edu/article/variety-not-volume-is-driving-big-data-initiatives/

Brent Dykes (2017), "Big Data: Forget Volume and Variety, Focus On Velocity", Forbes, https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-and-variety-focus-on-velocity/#4ee1026c6f7d

Jennifer Bresnick (2017) "Understanding the Many V's of Healthcare Big Data Analytics", *HealthIT Analytics,* https://healthitanalytics.com/news/understanding-the-many-vs-of-healthcare-big-data-analytics

John Joseph (2012), Lavastorm, "Taming Data Variety and Volatility is Key for Big Data Analytics", http://www.lavastorm.com/blog/2012/11/14/taming-data-variety-and-volatility-is-key-for-big-data-analytics/

40. "Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism", *Tech Crunch,* https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/

Srikanth Kintali (2016), "Data Veracity: Is it ok to overlook?", Lochbridge, http://lochbridge.com/blog/data-veracity-is-it-ok-to-overlook/

Assuncao, Marcos & N. Calheiros, Rodrigo & Bianchi, Silvia & Netto, Marco & Buyya, Rajkumar, (2014), Big Data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing*

Philipp Max Hartmann, Mohamed Zaki, Niels Feldmann and Andy Neely, (2016), "Capturing value from big data – a taxonomy of data-driven business models used by start-up firms", *International Journal of Operations & Production Management*

KnowledgeHut, (2016), "Types of Big Data", www.knowledgehut.com/blog/bigdata-hadoop/types-of-big-data

Oracle, (2015), "Big Data in Financial Services and Banking", Oracle Enterprise Architecture Whitepaper, www.oracle.com/us/technologies/big-data/big-data-in-financial-services-wp-2415760.pdf

CIO, (2014), "How CIOs Can Prepare for Healthcare 'Data Tsunami'", https://www.cio.com/article/2860072/healthcare/how-cios-can-prepare-for-healthcare-data-tsunami.html

Siddiqui, Ayesha & Qureshi, Riyazuddin, (2017), "Big Data In Banking: Opportunities And Challenges Post Demonetisation in India"

Independent, (2014), "There are officially more mobile devices than people in the world", http://www.independent.co.uk/life-style/gadgets-and-tech/news/there-are-officially-more-mobile-devices-than-people-in-the-world-9780518.html

Dailymail, (2014), "How often do YOU look at your phone?", http://www.dailymail.co.uk/sciencetech/article-2783677/How-YOU-look-phone-The-average-user-picks-device-1-500-times-day.html#ixzz50W5tFLsV

TCS, (2013), "Which Industries Use More Unstructured Data?", Big Data Study, http://sites.tcs.com/big-data-study/industries-unstructured-data/

McKinsey Global Institute (2011), "Big data: The next frontier for innovation, competition, and productivity",https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Elder Research, (2017), "The 42 V's of Big Data and Data Science", https://www.elderresearch.com/company/blog/42-v-of-big-data

"Big data is at the foundation of all the megatrends that are happening."

– Chris Lynch

Mining and Extracting Emotional Information from Human-Based EEG: An Overview

M. DURAIRAJ AND JANSI RANI R.

Emotion plays a vital role in our lives such as attention, memory and decision making and communication. Emotion recognition is a step towards aiding people such as in care taking and designing brain-computer interfaces. Nowadays, automatic emotion recognition is one of the most popular research topics in the fields of computer vision, speech recognition, brain-machine interface, and computational neuroscience. Recognizing emotions from EEG (Electroencephalogram) signals permits us to evaluate the internal condition of the subject, which is viewed as an important factor in HCI (Human Computer Interaction). One of the most important things in improving accuracy of emotion recognition is based upon choosing the Good Artifact Removal algorithm, Feature extraction, Feature selection method and Classification algorithms. This paper reviews EEG emotion recognition processes in 2017.

M. Durairaj, Department of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli.

Jansi Rani R, Department of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli

DOI : - https://doi.org/10.32381/PROD.2019.59.04.3

I. Introduction

An EEG (electroencephalogram) is a method to measure the electrical activity of the brain, with the aid of electrodes that measure voltage fluctuations on the scalp. While synapses are triggered, a minuscule amount of electric current is produced on the neurons because of the fluctuations between dendrites and axon. This electric field is ephemeral.

The frequency of brain is classified as five, Delta (0-4Hz), Theta (4-8Hz), Alpha (8-13), Beta (13-30Hz), Gamma (above 30) and each frequency bands depict the activities related with various cognitive operations.

EEG contains useful information about the brain, however, it contains a lot of noise. The physiological movements make the EEG an artifact. These noises are due to the procedures in Electrooculography (EOG), Electromyography (EMG) and Electrocardiography (ECG). The technique Electrooculography (EOG) is for measuring the cornea-retinal standing potential exists between the front and the back of the human eye. The diagnostic procedure Electromyography (EMG) is applied to evaluate the health condition of muscles and nerve cells. These nerve cells known as motor neurons transmit electrical signals to make muscles to contract and relax. The process of Electrocardiography (ECG) is to record electrical activity of hearts over a period of time through electrodes that placed on the skin. The noise effects out of these procedures Electrooculogram (EOG), Electromyogram (EMG), and Electrocardiogram (ECG). This can cause the EEG signals to uninformative. To correctly recognize the activity of the brain, the movement must be held to low. The raw EEG is reasonably news less and it needs to be processed before feeding into the classifier. Figure 1 illustrates the procedure of EEG. EEG is not only used in the healthcare industry,

but also in video games, Brain-Computer Interaction (BCI) and for helping physically challenged and impaired persons. In Brain Computer Interface (BCI) emotion recognition, a non-invasive way (using EEG) is used because it is cheap and simple.

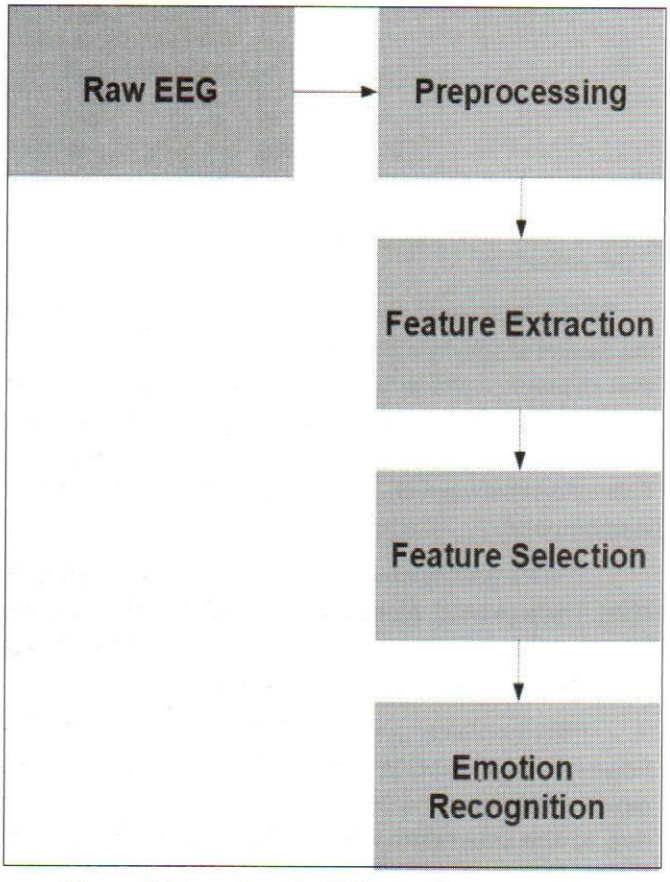


Figure 1. Procedure for EEG emotion recognition

In healthcare, EEG-based emotion recognition system is utilized for helping, observing, improving, or diagnosing patients' emotional states. Emotion state recognition is also useful to identify any neural developmental disorder. A portion of these neuro-developmental issue influence memory, feeling, learning capacity, conduct, and correspondence of people experiencing these conditions. The most well-known cases are Autism, Depression, Schizophrenia, Huntington's ailment (HD). Emotion recognition (Durairaj, 2018) is a step towards aiding people such as in caretaking and designing brain-computer interfaces. Nowadays, automatic emotion recognition is one of the most popular research topics in the fields of computer vision, speech recognition, brain-machine interface, and computational neuroscience.

The advancements of BCI reach various fields, going from computer games to the armed forces. Over the last 10 years, BCI has uncovered its possibilities for severely impaired and secured people who had not very many potential outcomes to collaborate with the ambient and with different subjects (Hochberg, 2006).

The advancement of an electroencephalography (EEG) based feeling acknowledgment framework requires computational models that portray how the emotional state is comprised in EEG signals and how one can appraise internal states from EEG signals. In spite of a long history of looking for EEG research on human emotion recognition, less consideration has been paid to the computational models for emotion state estimation. Henceforth, we feel requirements for an audit of the cutting edge computational models for emotion state estimation to help the improvement of cutting edge emotion classification strategies. This paper will survey the current computational strategies for emotion state estimation from the human EEG with an analysis on difficulties and a future bearings. EEG is not only used in the healthcare industry, it is also used in video games, Brain-Computer Interaction (BCI), and for helping physically challenged and impaired persons. In recent days EEG is used to predict human emotion (Hu, 2017) in the field of Affective Computing (AC) (Rani, 2016), Human Computer Interaction (HCI), neuroscience, and psychology etc.

This paper specifically focuses on the process of emotion recognition methods like preprocessing (removing noises), feature extraction (extracting significant information from raw signals), classification (extracted features are extracted feed into classifier in order to predict respective emotions).

II. EEG Data Acquisition

EEG signals are gathered utilizing EEG cap or EEG gadget. A large portion of the EEG gadget utilizes Ag/Ag-Cl electrodes which are either dry or Gel-based electrodes. The EEG data acquisition segment clarifies how EEG signals are procured from subjects. The investigations are finished utilizing a diverse number of subjects.

EEG Correlates Emotion:

The state space depicts how emotions can be identified using a computational model. This state space can be predominately classified into two, which are a discrete and dimensional model of emotion. The discrete model has six predominant emotions like happy, disgust, fear,

surprise and anger (Russell, 2003). On the other hand dimensional model emotion represents internal states as a vector (Russell, 2003). For example, the circumflex show, created by Russell, portrays an emotion state in a two-dimensional space with the arousal and valence measurements. Different mental models characterize emotion measurements that in this manner constitute the reason for the emotion state space (Russell, 2003).

In light of the development of the emotion state space, an examination of EEG connects of emotion ought to likewise deliver how to decide exploratory boosts to initiate emotions. Ordinarily, emotion inducements are chosen to cover wanted arousal states and valence states, and displayed in diverse modalities including the visual, sound-related, sense of touch, or odor inducement. The reality of the emotion state incited by a stimulus is anchored by abusing the self-ratings of subjects or utilizing the stimuli sets such as the universal emotional picture framework which is called International Affective Picture System (IAPS) or the universal emotional digitized sound framework (International Affective Digitized System (IADS)). The IAPS gives an arrangement of regularizing pictures for emotion inducement to initiate emotion changes and consideration levels (Mikels, 2005). The IADS exemplifies acoustic stimulus to evoke emotion, once in a while together with the IAPS (Redondo, 2008). These models are not dependant on culture or sex. Various neuropsychological investigations have detailed EEG correlates of emotion.

A. EEG device

The devices of EEG can vary with the distinct number of EEG electrode. The most famous gadgets are NeuoScan, BioSemi Active II, Emotive EPOC (wireless and gadget), Nevus EEG (Murugappan, 2010) and these gadgets have an alternate number of electrodes for instance 64, 32, 16, 14, and 8 like that. The quantity of electrodes utilized (and the device) expect the main part because of the time expected to set up the EEG device, the solace level of the subjects who wear the device, and the number of features to process. Hence, in a perfect world, the number of electrodes ought to be reduced. Be that as it may, as we will exhibit in the accompanying passages, the majority of the present works still require a generally huge number of electrodes and costly clinical setup.

B. Electrode Positioning system

Positioning electrodes on the scalp have to be placed according to electrode positioning system, which are 10-10 (Murugappan, 2010) framework, 10-20 system (International system) (Klem, 1999), and American Electroencephalography Society (AEEGS) (Sharbrough, 1991). 10-20 frameworks are generally utilized positioning system. The electrode positioning system is as shown in Figure 2.

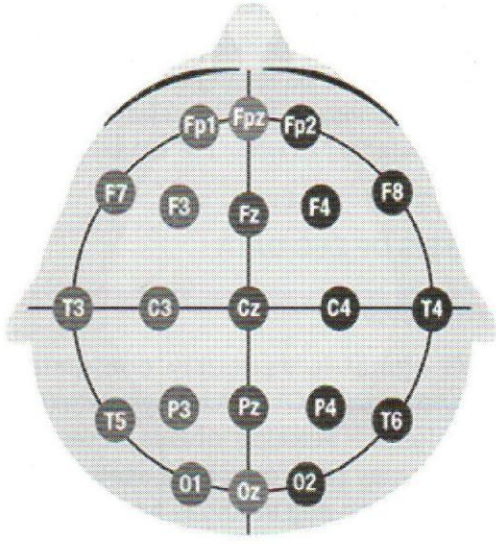


Figure 2. 10-20 Electrode Positioning System

The 10/20 framework or worldwide 10/20 framework is a universally perceived strategy to portray the area of scalp electrodes. The framework depends on the connection between the area of a terminal and the hidden zone of the cerebral cortex. The numbers 10 and 20 allude to the way that separations between neighboring electrodes are either 10 or 20% of the aggregate front-back or right-left separation of the skull. Each electrode has a letter or numerical values to recognize the projection and a number to distinguish the side of the hemisphere area. Four anatomical milestones are utilized for the basic positioning of the electrodes: first the notion which is the point between the forehead and the nose; second the inion which is the most minimal purpose of the skull from the back of the head and is typically shown by a bump; the opposite focuses front to the ear. Additional positions can be included y using the spaces in the middle of the current 10/20 framework.

C. Emotion stimuli

Emotion stimuli are targeted to induce emotion in subjects. The emotion elicitation method could be International Affective Picture System (IAPS), (Lang, 1999),

International Affective Digitized sound (IADS) (Stevenson, 2008), Mahnob HCI (Soleymani, 2012), RafD: Radbound Faces Database (RafD), Geneva Affective Picture Database (GAPED) (Dan-Glauser, 2011), Pictures of Facial Affect (Nasehi, 2012), Chinese Affective Picture System (CAPS) (Lu, 2005), films or self-actuated. While choosing stimuli, the researchers consider a few criteria for inducing emotions on the subjects, which are: I) the stimuli should not be very long; ii) the subject needs to comprehend focused on feeling without clarification; iii) the video should inspire the intended emotion. This is substantial material just when the stimuli are in the form of videos.

III. ANALYSIS OF EEG DATA

Artifact Removal

The EEG signal is seriously contaminated by Electrooculogram (EOG), Electromyogram (EMG), and Baseline signal. Filtering noises completely from EEG signals may pass out some useful information. A lot of noise removing methods are available. They are downsampling, bandpass filter, FIR filter, Independent Component Analysis (ICA), and normalization. Murugappan et al (Murugappan, 2010) used the surface Laplacian (SL) noise removal technique.

Feature Extraction

Keeping in mind the end goal feature extraction and/or feature selection must be performed on the EEG signal. The investigation of EEG signals is an active field of research and the diverse procedures proposed principally rely upon the sort of signals to be investigated and the data to be recovered (Subasi, 2005). A definitive focus of feature extraction is translating raw EEG signal into significant features. This progression assumes a vital part in grouping features. The features are extricated from various frequency bands (delta, theta, alpha, beta, and gamma). We can separate features into time domain {statistical, Event-related Potential (ERP)}, and frequency domain features (Power Spectra, Fourier Transform). The techniques for extricating features like statistical (mean, standard deviation, outright esteem, and standardization) (Yuen, 2011), power spectral features (Ko, 2009), wavelet transform (Murugappan, 2011), Higher Order Crossings (HOC) (Petrantonakis, 2011). The extracted features are fed into classification. To group the EEG signals, helpful attributes (features) are extracted on an appropriate change of the information; for the most part, the wavelet transform is used (Chang, 2011), (Rached, 2013). Specifically, in (Chang, 2011) time-frequency portrayal of the Electroencephalogram signals, with respect to two mental states, was gotten by the Discrete Wavelet Transform (DWT). The feature vectors to be utilized in the classification process were the coefficients of the DWT processed for every preliminary. The most widely recognized features separated from the EEG signals decomposed into various frequency bands by the Discrete Wavelet Transform were reviewed (Subasi, 2005): the entropy, the frequency band power and the standard deviation.

Feature Selection

Feature Selection is the way toward choosing a subset of important features. The feature selection process makes the process of EEG analysis into: a) less training time, b) the most important thing is, EEG is affected by the curse of dimensionality. Feature selection avoids them. The process of feature selection is selecting the features from specific frequency bands, time window, or channel. Principle Component Analysis (PCA), Independent Component Analysis (ICA) or Linear Discriminant Analysis (LDA), Minimum Redundancy Maximum Relevancy (MRMR) (Atkinson, 2016) are some of the examples of feature selection algorithms. These techniques can essentially diminish the number of features, yet there are still some research challenges. Regardless of whether the difference of components meets the prerequisites, the accuracy of classification is as yet inadmissible, likely in light of the fact that they can't expel repetitive features. Due to the multi-dimension of the features vector, the computational cost and complexity can be expanded drastically, and repetitive features will direct lower accuracy.

Emotion Recognition

The final stage of emotion recognition is selecting a good classifier in order to predict emotion. We can use either supervised or unsupervised algorithm to recognize emotions. The most predominant and an off-the-shelf algorithm for emotion recognition is Support Vector Machine (Supervised) and next to it is k-Nearest Neighbour (unsupervised).

Table 1 explains various methods used in the research in 2017. Each column represents different methods like how many subjects involved in the experiment, stimuli used, number of channels used, in the research what emotion this is going to classify, features used in the work (frequency bands: alpha, beta, gamma, delta and theta) and finally the accuracy of work.

Table 1: Comparison of EEG Emotion Recognition on the 2017 year's research articles

Ref	Sub	Stimulus	#channels	Emotions	Features	Accuracy (classifier)
(Becker, 2016)	27	FilmStim	257	Positive, and negative	θ, α, ®, low γ(30-45Hz), high γ(55-80Hz)	70-75% (SVM)
(Sorinas, 2017)	5	video	21	Positive, and negative	Δ, α, β1(14-23Hz), β2 (24-30Hz), γ, θ	70% (SVM)
(Lin, 2017)	12	Music	12	Happiness, and sadness	Ä, θ, α, β, γ	58.31%~64.03%(GNB)
(Huang, 2017)	20	Movie	1	Valence, and arousal	Δ, θ, α1(8-10Hz), α2(11-13Hz), β1(14-20Hz), β2(21-30Hz), γ1(31-40Hz), γ2(41-50Hz)	66.88% (SVM)
(Menezes, 2017)	DEAP	-	4	Valence, and arousal	Δ, α, β, γ, θ	SVM 88.4% (valence), 74% (Arousal)
(Ozerdem, 2017)	DEAP	-	5	Valence, and arousal	θ	77.14% (MLPNN), 72.92% (kNN)
(Chatchinarat, 2017)	DEAP	-	2	Arousal, valence, and dominance	γ	FSVM68.61%(arousal), 63.28%(valence) 68.91%(dominance)
(Chandra, 2017)	DEAP	-	-	Valence, and arousal	θ, low α (8-10Hz), α(8-13Hz), β, γ	SVM RBF 76.8%(Valence), 74.3%(Arousal)
(Hu, 2017)	20	film	32	Amusement, Awe, Gratitude, Hope, Inspiration, Interest, Joy, Love, Pride, Serenity, Arousal, Valence, Familiarity, and Liking	θ, α, β	SVM 82.1% (Amusement) 1.4%(Awe), 80.3%(Gratitude), 79.4%(Hope), 79.1%(Inspiration), 80.3%(Interest), 80.2%(Joy), 80.4%(love), 80.9%(pride), 80.2%(serenity), 80.4%(Arousal), 80.6%(Valence), 79.5%(Familiarity), 80.2%(Liking)
(S. Liu, 2017)	13	IAPS	28	Pleasant/high arousal, neutral, and unpleasant /high arousal	1-45Hz	82.85%(SVM)
(Chai, 2017)	15	film	62	-	Δ, α, β, γ, θ	84.61%(ASFM)
(J. Liu, 2017)	14	film	30	Joy, amusement, tenderness, anger, sadness, fear, disgust and neutrality	Δ, α, β, γ, θ	SVM 86.43%(Joy, amusement, tenderness), 65.09%(anger, sadness, fear, disgust)

(Zheng, 2017)	DEAP	-	-	Valence, and arousal	-	SVM Valence 70.41% (IMF1, 32 channels), 69.10% (8 channels) Arousal (IMF1, 32 channels), 71.99% (8 channels)
(Koelstra, 2017)	15	film	62	Positive, neutral, and negative	Δ, α, β, ã, θ	69.67% (DE + GELM)

SVM : Support Vector Machine; GNB : Gaussian Naïve Bayes; DEAP : Database for Emotion Analysis Using Physiological Signals[30]; RBF : Radial Basis Function; IAPS: International Affective Picture System; ASFM : Adaptive Subspace Feature Matching; IMF: Intrinsic Mode Function; ASFM: Adaptive Subspace Feature Matching; IMF : Intrinsic Mode Function; GELM : Graph Regularized Extreme Learning Machine.

SVM

It is also known as large margin classifier. SVM separates classes with the aid of Support Vectors for linear classes. For non-separable classes use Radial basis function (RBF), polynomial kernel, and Gaussian RBF methods to separate classes. Usually, classifying two classes is simple but multiclass classification is not easy in SVM. Multiclass SVMs are usually implemented by combining several two-class SVMs. The one-versus-all method using winner-takes-all strategy and the one-versus-one method implemented by max-wins voting are popularly used for this purpose. The empirical evidence to show that these methods are inferior to another one versus one method: one that uses Platt's posterior probabilities together with the pair wise coupling idea of Hastie and Tibshirani. The evidence is particularly strong when the training dataset is sparse. The functions of SVM are depicted in Figure 3.

Kernels are utilized to help vector machines to outline learning information (nonlinearly) into a higher dimensional element space where the computational intensity of the straight learning machine is expanded. Each kernel has its own favorable pros and cons (Vapnik, 1998).

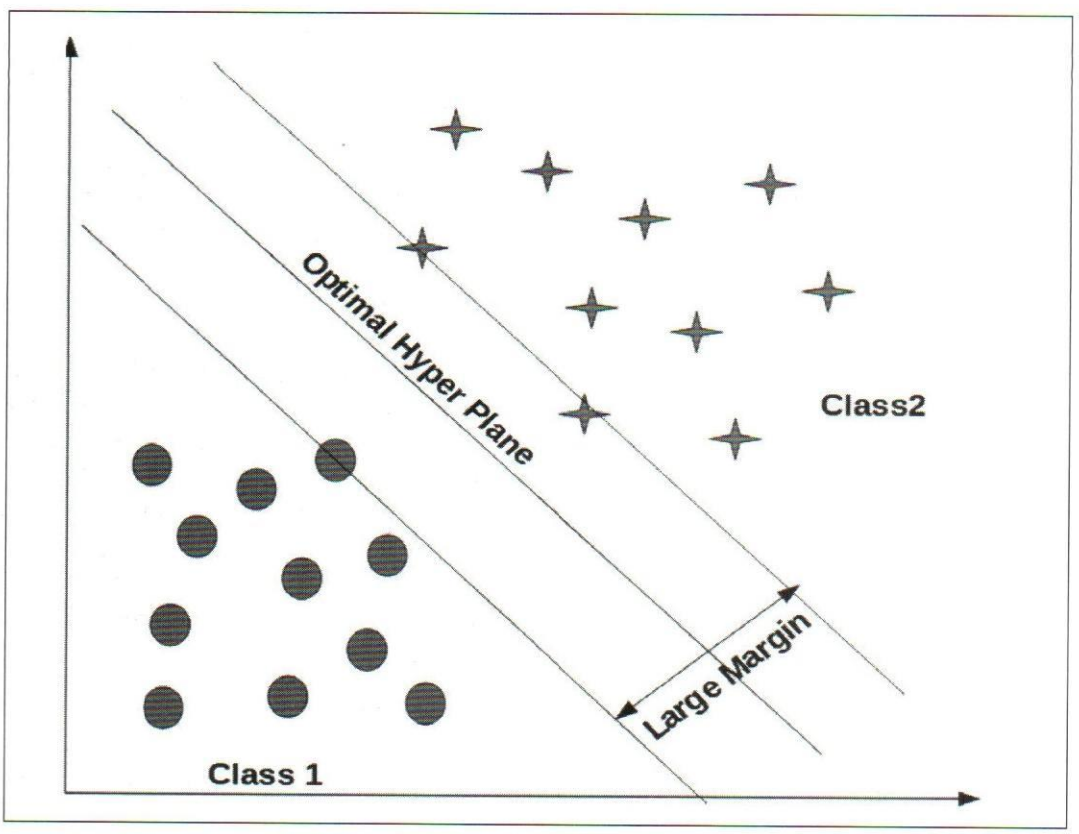


Figure 3. SVM

kNN

kNN is utilized for both predictive models and in classification. It is a basic and instinctive technique. K-closest neighbour is a basic classifier that has been used in many pattern recognition applications. The class dependent variable of another test is resolved with regard to the dependent variable of the closest traning features. k nearest preparing tests to another test are resolved and the name of a test is determined to agree to the most over and again labels of these k nearest tests. The quantity of the closest neighbors (K) is required to be resolved for the characterization procedure. So this calculation is generally utilized in picture and signal processing. This calculation groups in comparing new information and training data. This strategy takes after two sorts of choice principles 1) similarity voting scheme and 2) majority voting scheme. In this method of classification we need to assign the estimation of k, for instance k = 3, 4,7 ... like that.

Figure 4 illustrates the function of kNN.

To find a distance in kNN uses different methods. They are as follows:

Euclidien : $\sqrt{\sum_{i=1}^{k}(xi - yi)^2}$

Manhattan: $\sum_{i=1}^{k}|xi - yi|$

Minkowski: $(\sum_{i=1}^{k}(|xi - yi|)^q)^{1/q}$

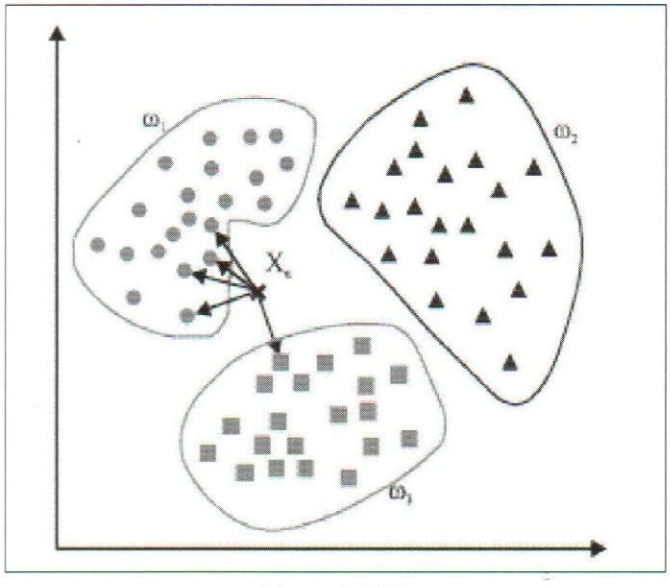


Figure 4. kNN

IV. Result And Discussion

Multi-Modal Emotion Recognition:

Physiological changes in a human body may indicate current emotional status, which is the reason for physiological signals (Takahashi, 2004) and pictures are recorded with a specific end goal to perceive emotions. Some biological signals in the human body and their records are depicted as takes after: 1) Cardiovascular framework: pulse inconstancy (HRV), blood pressure, electrocardiogram (ECG), Cardiac output, and so forth 2) Respiratory framework: breath rate and so forth. 3) Muscular framework: electromyogram (EMG) signals, and so on 4) Brain action: EEG signals, Functional magnetic resonance imaging or functional MRI (fMRI) and so on.

EEG signals because of their straightforwardness to investigate and great time and spatial goals have turned out to be normal and valuable in most BCI applications, for example, feeling acknowledgment. Likewise, EEG recording frameworks are economical and flexible. Previous researches demonstrate that by recording and preparing EEG signals we can accomplish great outcomes in terms of internal (emotional) state identification.

V. CONCLUSION

This paper presents a review of EEG emotion recognition works in 2017. This paper is targeted at people who are new to this field. It may help them understand the basic concepts of EEG emotion recognition and steps involved in the process. The curse of dimensionality (high dimensional space) is one of the biggest problems in EEG signal. Most of the works were SVM, which performs better in emotion recognition. This paper overviews 1) devices used in emotion recognition, 2) emotion stimuli and artifact removal methods 3) the table of 2017 journals, which is related to EEG based emotion recognition, and 4) feature extraction 5) feature selection and 6) classification algorithms used in emotion recognition. SVM performs better recognition rate as compared to other classifiers, but the features vary according to the problem. However, in most cases, beta, alpha, and gamma contains relevant information to emotion recognition (Lu, 2009) (Zengh, 2014).

References

Atkinson. J and Campos. D, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications,* vol.47, pp.35–41, 2016.

Becker. H, Fleureau .J, Guillotel .P, Wendling .F, Merlet .I, and Albera .L, "Emotion Recognition based on high-resolution EEG recordings and reconstructed brain sources", *IEEETransactions on Affective Computing,* doi:10.1109/TAFFC.2017.2768030.

Candra .H, Yuwono .M, Chai .R, Nguyen H.T., and Su.S, "EEG emotion recognition using reduced channel wavelet entropy and average wavelet coefficient features with normal Mutual Information method," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, pp. 463-466

Chai X, Wang Q, Zhao Y, Li Y, Liu D, Liu X, Bai O,"A Fast Efficient Domain Adaptation Technique for Cross-Domain Electroencephalography(EEG)-Based Emotion Recognition," *Sensors,* vo. 17(5), doi: 10.3390/ s17051014, 2017.

Chang, C. .C, Lin .C .J, "LIBSVM: a library for support vector machines," *ACM Transactions onIntelligentSystem and Technology,* vol. 2(27), pp. 1–27, 2011.Software available at http://www.csie.ntu.edu.tw/<"cjlin/libsvm.

Chatchinarat. A, Wang K.W., and, Fung C.C., "Emotion classification from Electroencephalogram using Fuzzy Support Vector Machine, ICONIP 2017, Lecture Notes in Computer Science, vol. 10634, pp.455-462, 2017.

Dan-Glauser .E .S, Scherer .K .R, "The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance." *Behavior Research Methods,* pp.468–77, 2011.

Durairaj. M, Rani .J .R, "A Review on Emotion Recognition," *International Journal of Emerging Technology and Advanced Engineering,* vol. 8, issue.2, pp. 108-112, 2018.

Hochberg. L .R, Serruya. M .D, Friehs .G .M, Mukand. J. A, Saleh. M, Caplan. A. H, Branner. A, Chen. D, Penn R.D, Donoghue. J .P, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature,* vol.442, no.7099, pp.164–171, Jul.2006.

Hu.X, et al, "EEG Correlates of Ten Positive Emotions," *Front. Hum. Neurosci,* vol.11, 2017.

Huang.Y et al, "Fusion of facial expressions and EEG for multimodal emotion recognition," *Computational Intelligence and Neuroscience,* vol.2017, 2017.

Klem G. H., Lders H. O., Jasper H. H., and Elger. C, "The ten-twenty electrode system of the international federation," *The international federation of clinical neurophysiology,* vol.52, pp.3-6, 1999.

Ko, K. E, Yang, H .C, and Sim, K. .B, "Emotion recognition using EEG signals with relative power values and bayesian network," *Int. J. Control Autom. Syst.,* vol. 7, no. 5, pp.865–870, Oct.2009.

Koelstra.S et al., "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," in *IEEE Transactions on Affective Computing,* vol. 3, no.1, pp.18-31, Jan.-March 2012.

Lang .P .J, Bradley .M .M, Cuthbert .B .N, "International affective picture system (IAPS): Technical manual and affective ratings," University of Florida, Center for Research in Psychophysiology, Gainesville, 1999.

Li M, and Lu B. L., "Emotion classification based on gamma-band EEG," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, *MN,* 2009, pp. 1223-1226.

Lin Y. -P., Jao P. -K., and Yang Y. -H., " Improving Cross-Day EEG-Based Emotion Classification Using Robust Principal Component Analysis," *Front. Comput. Neurosci,* 2017.

Liu.S, Zhang .D, Jingjing .T, He .F, Qi .H, Zhang .L, Ming .D, "EEG-based emotion estimation using adaptive tracking of discriminative frequency components," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, pp. 2231-2234.

Liu.Y. J, Yu .M, Zhao .G, Song .J, Ge .Y, Shi .Y, "Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals," in *IEEE Transactions on Affective Computing,* no. 99, 2017.

Lu .B, Hui .M, and Yu-Xia .H, "The Development of Native Chinese Affective Picture System–A pretest in 46 College Students," *Chinese Mental Health Journal,* vol.19, no.11, pp.719–722, 2005.

Murugappan M, "Human emotion classification using wavelet transform and KNN," 2011 International Conference on Pattern Analysis and Intelligence Robotics, Putrajaya, 2011, pp.148-153.doi: 10.1109/ICPAIR.2011.5976886.

Menezes M. L. R,Samara .A, Galway .L, Sant'Anna .A, Verikas .A, Alonso-Fernandez .F, Wang .H, Bond .R "Towards emotion recognition for the virtual environment: an evaluation of EEG features on benchmark dataset," *Pers Ubiquitous Comput.,* 2017.

Mikels .J .A.et al, "Emotional Category data on Images from the International AffectivePicture System," *Behavior Research Methods,* vol. 37, no.4, pp.626-630, 2005.

Murugappan.X, Ramachandran. N, and Sazali.Y, "Classification of human emotion from EEG using discrete wavelet transform "*Journal of Biomedical Science and Engineering,* vol. 3, pp.390- 396, April 2010.

Nasehi .S and Pourghassem .H, "An optimal EEG-based emotion recognition algorithm using Gabor features," *WSEAS Transactions on Signal Processing,* vol. 8, no.3, pp.87–99, 2012.

Ozerdem .M .S, Polat .H, "Emotion recognition based on EEG features in movie clips with channel selection," *H. Brain Inf.,* 2017.

Petrantonakis .P .C, Hadjileontiadis .L .J, "A Novel Emotion Elicitation Index Using Frontal Brain Asymmetry for Enhanced EEG-Based Emotion Recognition," in *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, no.5, pp.737-746, Sept.2011.

Rached .T .S, Perkusich .A, "Emotion recognition based on brain computer interface systems," in: R. Fazel-Rezai (Ed.), Brain–Computer Interface Systems – Recent Progress and Future Prospects, Chapter 13, 2013.

Rani .J .R, Durairaj .M, "A Review on Affective Computing", *IJCTA,* vol. 9, issue. 2, pp. 235-239, 2016

Redondo .J, Fraga .I, Padron I, Pineiro .A, "Affective ratings of sound stimuli," *Behavior Research Methods,* vol.40, no.3, pp.784–790, 2008.

Russell .J .V, "Core affect and the psychological construction of emotion," *Psycological Review,* vol.110, pp.145-172, 2003.

Sharbrough.F, Chatrian. G. E., Lesser, H. Luders, M.Nuwer, and T.W. Picton, "American electroencephalographic society guidelines for standard electrode position nomenclature," *J. Clin. Neurophysiol.*, vol. 8, pp.200–202, 1991.

Soleymani .M, Pantic .M and Pun .T, "Multimodal Emotion Recognition in Response to Videos," in *IEEE Transactions on Affective Computing,* vol.3, no.2, pp.211-223, April-June 2012.

Sorinas.J, Murcia .M .D .G, Minguillon .J, Sanchez-Ferrer .F, Val-Calvo .M, Ferrandez .J .M, Fernandez .E, "Setting the Parameters for an Accurate EEG (Electroencephalography)-Based Emotion Recognition System," Natural and Artificial Computation for Biomedicine and Neuroscience.IWINAC 2017, Lecture Notes in Computer Science, vol. 10337, Springer, 2017.

Stevenson .R .A, James .T .W, "Affective auditory stimuli: characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories," *Behav. Res. Meth.* Vol. 40, pp.315–321, 2008

Subasi .A, Ercelebi .E, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Prog. Biomed,* vol. 78, no.2, pp.87–99, 2005.

Takahashi .K, "Remarks on emotion recognition from bio-potential signals," In: Proc. 2nd Internat.Conf.on Autonomous Robots and Agents, pp. 186–191, 2004.

Vapnik .V .N, "statistical learning theory," Wiley & sons, 1998.

Zheng W. L, Zhu .J .-Y, Peng .Y, Lu .B .-L "EEG-based emotion classification using deep belief networks," *IEEE International Conference on Multimedia and Expo* (ICME), Chengdu, 2014, pp. 1-6, 2014.

Zheng W. L., Zhu J. Y. and Lu B. L., "Identifying Stable Patterns over Time for Emotion Recognition from EEG," in *IEEE Transactions on Affective Computing,* no. 99, 2017.

Zhuang.N et al, "Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain," *BioMed Research International,* vol.2017, 2017.

"Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway."

– Geoffrey Moore

Experiences and Challenges of Dealing with Large Sample

SURASHREE SHOME

The paper narrates the data collection process followed before data mining for a study conducted in 50 schools across two states of India covering more than 14000 students. Objective of the paper is to document the process and challenges of conducting studies with large sample. The paper will help research organizations and/or researchers to plan and organize their data collection process.

Dr. Surashree Shome, Research Head, Grassroots Research Advocacy and Movement (GRAAM), Karnataka

DOI : - https://doi.org/10.32381/PROD.2019.59.04.4

I. Introduction

Datamining and collection is an integral part of any research study, whether it is: field based or lab based; qualitative or quantitative; or for a social science or pure science. In social science studies, adequacy and appropriateness of the collected primary data is a direct function of sampling technique and data collection method. Inadequate and/or inappropriate sampling technique and data collection method not only leads to several academic and research outcomes related inaccuracies but also associated operational/administrative complications. Amongst others, several operational complications such as over-shooting the allocated budget, exclusion of appropriate enumerators in the research team, and/or confusion while collecting data amongst enumerators, could be encountered, which will subsequently affect the quality of the research outcome.

This paper envisages to enumerate the challenges and lessons learnt while collecting data for data mining and analysis of a large study conducted by GRAAM for an Europe based University, covering 50 central government schools spread across the states of Karnataka and Goa, India covering more than 14000 students. This in turn, will help research organizations as well as young researchers to plan and organize their data collection process with better understanding. The paper is divided into two sections. The first section will introduce the nature and objectives of the study succinctly. The second section comprises mainly of two parts: (1) Process of data collection during Baseline Survey and (2) 'Process of data collection during Endline Survey. Although the organization was also part of other activities under the study but this paper is concerned only with the data collection aspect.

I. Brief About the Study

GRAAM was associated as an implementation partner to undertake an action research study to understand the impact of learning method in selected schools. The study covered 45 Central Government schools in Karnataka and five schools from the state of Goa. The data collection for the study was executed in two phases:

1. Baseline Test was conducted before introducing the learning method in the school. More than 14,000 students from 50 schools participated in the test.

2. Endline Test was conducted after the learning method was carried out for more than three months in the schools. More than 11,000 students from 37 schools participated in the test.

II. Process of Data Collection

II.A. Baseline Survey:

Before introducing the learning method in selected schools, all students of the 50 schools from Grade VI, VII and VIII filled their profile detail (*in a format provided*) and took a baseline test in two subjects (*Maths & English*). Each student took about two hours to complete the tests and filled the format (40 minutes' x 3 activities). In total, more than 14,000 students participated from 50 schools in this baseline test. Along with students, English and Mathematics teachers of the selected classes also filled a questionnaire about the teaching methods used in their classes. One of the major aspect (as well as a challenge) of the baseline tests was to conduct it all across 50 schools within 15 working days. This essentially required conducting test in 3 to 4 schools simultaneously, and which in turn required more than four capacitated team working concurrently.

a. Organizing Team for the Survey:

The schools were divided in seven groups on the basis of geographical proximity and a team was assigned to each group. This was primarily done to minimize travel of the team and optimize the time available for the baseline. The Bangalore team covered 14 schools from Bangalore City along with a school in Kolar which is around 100 kilometers away from GRAAM Office, Bangalore. Similarly, all the schools of Goa were covered by a single team. Table 1 below provides details of the groups and number of schools in each group.

TABLE 1: Number of Schools Covered in Each Group

S. No.	Groups for the Survey	Number of Schools
1	All the schools in *Bangalore* City (along with Kolar)	14
2	Mysore and neighboring areas of *Mandya, Chamrajnagar*, and *Kodagu*	5
3	Mangalore and neighboring areas, of *Shimoga, Karwar, Udupi* and *Hassan*	6
4	*Davangere* and neighboring areas of *Tumkur, Haveri, Bellary, Hosepet, Koppal*, and *Raichur*	9
5	*Gulbarga* and neighboring areas of *Bidar Bijapur, Bagalkot, Dharwad,* and *Hubli*	7
6	*Belgaum* and neighboring areas of *Chikkodi*	4
7	All the 5 schools in Goa	5
	Total Number of Schools	**50**

Each team consisted of a team leader and 3 to 18 enumerators (depending upon the number of class-wise sections in a school). It was the responsibility of the team leader to contact the principals, confirm/re-confirm the date and time with the Principal/Contact person from the school, check available question papers as per the number of students in schools in each class, train enumerators for the test, conduct the test with the help of enumerators, and count and seal question papers and deliver it safely to the GRAAM Head Office at Mysore, Karnataka. The enumerator's role was more of an invigilator who distributed the test papers to the students (*of the class allotted to her/him*), addressing the doubts of the students (*if required with the help of team leader or with respective subject teacher from school*), ensuring that students are not copying from each other, collecting the answer sheets, counting and putting them in an envelope and seal the same in presence of the team leader.

We selected five persons from GRAAM and two external professionals as team leaders who were later trained and also participated in the pilot test in a school from Mysuru. The challenges faced during the pilot tests were discussed and accordingly procedural amendments were made. However, at the very last moment, one of our team leader had to drop out because of personal reasons. After a search of four days, we were able to hire another team leader who was trained during the ongoing test in the school by the project holder. This provided us with a lesson that how important it is to have back-up human resources or good networking in the sector in order to conduct such a huge project with a pre-determined time frame.

The number of enumerators for each school also varied according to the section within each grades/classes across all the schools. The study design mandated to engage one enumerator for each section of every class/ grade. For e.g. if a school has three sections in each grade, then nine enumerators were required (*tests were conducted for grade VI, VII, and VIII of all the schools*) for that particular school. The number of sections in each grade differed from school to school. For e.g. - a school from KR Puram, Bangalore has one section for each grade (*meaning three enumerators*), whereas a school from Belgaum has six sections for each grade (*meaning required number of enumerators was 18*). Table 2 below provides a sample list of schools, with their number of sections per grade and required number of total enumerators for the respective school, for better understanding.

Here the grassroots connection of the GRAAM with various NGOs across Karnataka played a major role. GRAAM could organize the team of local enumerators quite smoothly because of its connect with the local NGOs for each of the *study* location. In certain cases, more than two NGOs came together and provided enumerators for the study. This, in turn, evidently has decreased our travel as well as lodging cost. Here, the lesson learnt is about the importance of networking and institutional relationship for organizing human resources (within a defined particular space and time) – which are integral part of large surveys.

TABLE 2: Required Number of Enumerators for Schools (Sample List only)

S. No.	Location of the school	Number of Section in Each Grade			Total number of Classes	Required number of Enumerators
		VI	VII	VIII		
1	Jalahalli, Bangalore	5	5	5	15	15
3	MEG Centre, Bangalore	5	5	5	15	15
4	Malleshwaram, Bangalore	3	3	3	9	9
6	K.R.Puram, Bangalore	1	1	1	3	3
7	Mysore	4	4	4	12	12
8	Chamarajanagar	1	1	1	3	3
9	Belgaum	6	6	6	18	18
10	Mangalore	3	3	3	9	9

On the other hand, we couldn't locate any local organization from Goa for such support. Thus, a team of six enumerators with a team leader travelled to Goa for the study purpose. Only for one school we had hired six local enumerators with the help of teachers from the schools. The team leader trained these local enumerators for the test. This has taught us that the stakeholders can be a major support in the field to organize resources. Therefore, it is important to build a good relationship with stakeholders.

b. *Organizing the test dates with the Schools:*

One of the initial and most complicated aspects of the study was to plan, sequence and confirm the test dates with the schools. This micro-planning was important to synchronize the training of team, travel, logistics, printing and other aspects of the study plan so as to optimize the time and financial resources. Even though the study was sanctioned by the headquarter of the schools located in New Delhi under Ministry of Human Resource Development, no administrative communication was done with the schools in this respect. The regional centre in Bangalore received a letter from the headquarters but they didn't process it further and schools were not informed. Thus, when the study team met the officials at regional centre, they asked us to communicate the matter to the schools. We got a list of email ids with phone numbers of all the 50 schools from the regional centre (however, the same is available in the website of the schools). Based on the mail id provided to us, we started communicating with the schools about the study and then started the long process of confirming/re-confirming dates. However, with all the efforts - we could receive confirmation from only one school. No one else responded to the e-mails sent to them. Then, the research team started calling up the schools for confirming the date. Shockingly, we came to know that the principals rarely check their mail. When we discussed the issues with the internal team of GRAAM, our executive director suggested us to send a courier to the schools as like government offices, the school's office will also put the letter in the 'inward' box of the principal, which will ensure that the letter has been received by the schools. Hence, we started the second phase of communication with the schools. On the next day we couriered (*as regular post would take more than a week to reach in certain cases*), letters to all the 50 schools. However, this also did not provide us 100% delivery assurance as many schools are 'out of the service area' of the courier services. Here, the lesson learnt was 'in most of the cases, IT enabled communication might not work'. And always send letter by post to government institutions', if one really wants them to receive it and process it further.

Moreover, out of 50 schools who received our letter – only two schools acknowledged and replied back with confirmation within the given date. Thereafter, we started calling them again to ensure that they receive, read and confirm the given dates for test in the school. One person, along with the project coordinator, was specially assigned to call up the schools, first on official telephone numbers and then on personal mobile numbers of the principal. However, we realized that calling the personal phone number was not a viable option - as more than 30% of the principals were transferred in the last academic year and neither the website nor the regional office had provided the updated data. Thus, our communication solely depended upon the school landline number/s. There were several challenges of using the landline. Most of the time, it was not picked up. Even if it is picked up, it would be some administrative staff and required to be briefed about the entire study along with the request to pass the phone to the principal. In about 70% of the cases, the principal was not available on the chair or was in a meeting. Thus, we were asked to call back after few hours or a day (*sometimes a week also*). On an average it required more than five to six attempts per school to contact the principal and confirm the test dates. Through this tedious process, the research team could confirm with 23 to 24 schools. Many of the schools that had confirmed the dates requested to change them as they had scheduled another event on the same day, which we obliged without much hesitation. However, in our long and tedious process of getting confirmed dates and developing an over-all time-table for travel, a ray of hope emerged from a principal of a school from Bagalkot. He patiently heard about our study requirement and arranged for the mobile numbers of few of the principals with whom we were unable to make a contact. With this support, we got confirmation from another 4 to 5 schools. However, we were facing a problem from six of the Bangalore based schools as we were unable to reach them through the given landline numbers. The phone was ringing but nobody was picking up the calls. At last, we physically visited those schools, directly met the concerned principals and got the confirmation of dates. When we discussed about not being able to reach them through phone they replied that the phone rings continuously as parents keep on calling and thus they seldom answer it. However, while returning from these schools, we got the mobile numbers of the teachers, who would be our point person for the baseline test. It would be worthwhile to mention here that the last school confirmed the test date about five days before the research deadline. With a great relief, we were able to conduct the test within 15 working days. However, the ordeal was not over with the confirmation of the dates. Some of those are discussed in the next section.

c. Other smaller but relevant issues

We received the final test papers from the concerned University at the last minute and hence printing of thousands of test papers was a major hurdle. Priority was given to the non-Bangalore based schools, and the team worked tirelessly to organize sets for 35 such schools. For Bangalore schools, daily the team of enumerators used to travel back to the coordination office at Bangalore to organize sets of question paper for the next day's test.

The process, might be with small hiccups, were going smoothly until the monsoon started. On one such weekend, rainwater inundated the office and around 5% of the unfilled question papers (sigh!) were damaged. We tried drying most of them but also made extra copies. However, it also increased the workload of the available team.

Other issues, especially human errors also derailed our plan. For e.g. – in few cases the team leader found that question paper available was not enough to conduct the test after reaching the school, thus they immediately rushed to the nearest market areas to make copies of the question papers or ask school admin team to help them to make few copies of the question papers (if the number of copies required was not very high). Similarly, absence/ shortage of required number of enumerators also hampered our progress. In such cases, available enumerators conducted tests in two shifts to cover all the sections of the schools.

II.B. Endline Survey

a. Logistic Arrangements:

After completion of the baseline test, schools were divided into four groups and different learning methods were introduced in each group. After three months of experiment, the end line survey was conducted in February. This time, the number of schools were reduced to 37 and like the baseline survey, all the tests in these schools were to be completed within 15 working days. Thus, based on our previous experiences, we planned the end line survey about two months ahead of the test period.

The question paper was shared by the university almost a month prior to the test. As soon as the test papers were received, the proof reading by subject teachers (from non-program schools) were completed, finalized and sent them for printing. This decision and actions were easy in this round because of our experience in the previous round. Sets of printed question paper were then prepared according to the number of students in each school. The sets were packed and labeled grade wise, section wise and subject wise in an envelope and then all the school level sets were packed in a box and labeled appropriately. This was all done in the GRAAM office at Mysuru. Thereafter, all the boxes were shifted to GRAAM Bangalore office as it was going to be the coordination location.

b. Confirming the dates of the test with the Schools:

As in baseline, we didn't want to repeat the experience of sequencing and confirming the dates with the schools. After preparing a tentative list, we emailed and posted letters to the respective schools for date confirmation, and immediately started calling up the school principals. The contact list prepared in the baseline was very useful to reach out to the concerned principals and teachers. The result was good and immediate. We received 17 confirmations within 14 days. We started reaching out to the remaining principals but they were engaged with the examination. After continuous follow-up, we were able to get the confirmation from all the schools before the starting date of the survey.

On day one of the endline survey, we were informed that as per the central circular all the students have to attend the live telecast of Prime Minister's address to the students on 'Examination Stress' from 1100 to 1200 hours. This certainly disrupted our plan. After much discussion and request, the test in five schools was rescheduled. However, by the end of the first week (the test period was spread across two and half weeks), remaining schools informed us that regional office at Bangalore had asked all the schools under them to conduct pretest for at least in three subjects (viz. Maths, Social Sciences and Science) from standard VI onwards in the very next week, which clashed with the dates of our test. Several schools organized the pretest on different days and several conducted our test after the pre-test. We felt sorry for the students as they were writing tests for five hours at a stretch in a day. This experience taught us a lesson that studies involving government institutions can get disrupted even with all the planning, because of the way system functions. Thus, 'be prepared for the worst, but always hope for the best'.

c. Setting a Team for the Test

We had arranged for team-leaders who could be engaged in case of any contingencies taking our learning into

practice from the earlier round. With all precaution, yet we had to face HR trouble. One of the team leaders had to step out because of personal reasons at the beginning of the test. Another team leader stepped-in and took over the process. By the end of second week, another team leader (of Gulbarga region) had to step out of the process as there was a death in his family. This became more complicated as the information came just 3 hours before the start of the test. Since the travel time from Bangalore to Bidar was more than 8 hours, asking anybody else to go to Bidar was not feasible. Even asking one of the enumerators to lead the team was not possible as they were not aware of the entire process. Thus, started the frantic search for another team leader. At the end, it was decided a teacher from the school, who was also the key contact person (*from most of the school, a teacher had been appointed to facilitate our team*), should lead the team for a day. A decision was made with the help of two facts: 1. All the teachers (contact person for the study) were aware of the project processes and, 2. They have worked as invigilator in school. With the permission of the principal, the teacher (and key contact person) was appointed to be the team leader for the day. The project head again briefed him about his role. With continuous facilitation over telephonic calls at the critical point of the test, the appointed teacher successfully completed the given task. A similar approach was followed for the next day. The team lead rejoined the team next week. However, this incidence reinforced our belief that good relationship with the stakeholders is important to overcome the unforeseen circumstances.

III. Conclusion

The above narration has already highlighted the lessons learnt from different phases of the study and hence will not reiterate the same. However, a few aspects need to be highlighted here for all organizations involved and/or planning large surveys. First important consideration is to plan in advance. We would suggest minimum two to three months (*depending upon the sample size and spread*) for planning, establishing contacts, arranging for questionnaire/test papers in order to successfully conduct such a huge data collection process within a given time frame. Second most important factor is team management and cooperation. As said by the Operation Head of the project "*.... the collective team effort by all the GRAAMMer's[1] made it possible to finish data collection in time...* ". So, working in team, appropriate cooperation and engaging everyone in planning process is important for engaging in primary data collection at large scale.

Another aspect which helped GRAAM to complete the task was the strong networking and institutional relationship with local organization. This not only helped in coordination but also in effective time and financial resource management.

Last but not the least – in case of repeat survey (as was the case here in terms of baseline and end-line surveys), it is important to maintain contact with some principal respondents. In the world of technology, a small mail or a message in a month or two keeps you connected with the important stakeholders. This helps in optimizing and adequate planning of subsequent phases of any study.

The overall learning from the entire experience is that while it is important to design any research with high academic rigor but the operational aspect of conducting research is equally important and requires as much attention as the academic planning of the study.

Notes

[1] Staff of GRAAM

References

Siwale, Juliana. (2015). Why Did I Not Prepare for This? The Politics of Negotiating Fieldwork Access, Identity, and Methodology in Researching Microfinance Institutions. SAGE Open, April-June 2015: 1–12

"You can have data without information, but you cannot have information without data."

– Daniel Keys Moran

Feature

A Study of Impact of Demographic Variables on Quality of Work Life

REETIKA THAKUR AND DINESH SHARMA

A good Quality of Work Life (QWL) is indispensable for the success and survival of any organization as it contributes to employee motivation and satisfaction, which enhances employee performance. QWL enables organisations to enjoy a competitive advantage in the present scenario of increased competition and challenges. Also, people are the most vital resource in an organization as they facilitate optimum and the most effective use of other organizational resources.

Thus, the present study aims at assessing the QWL of employees of Himachal Pradesh Power Corporation Limited (HPPCL) and measuring the impact of various demographic variables on QWL.

Results show that all the eight QWL dimensions taken in the study make a contribution to the overall QWL, and HPPCL employees are quite satisfied with the QWL and its dimensions. Results also reveal that gender, age, hierarchical level and marital status have no impact on QWL. However, annual income was found to have a statistically significant impact on QWL.

Reetika Thakur, Doctoral Research Fellow, Himachal Pradesh University Business School, H.P. University, Shimla.

Dinesh Sharma, Professor, Himachal Pradesh University Business School, H.P. University, Shimla.

DOI : - https://doi.org/10.32381/PROD.2019.59.04.5

Introduction

Every organisation needs five basic organisational inputs/ resources for its survival and growth. These are known as the 5 M's of management- Man, Material, Machine, Money and Methods. Out of these five organisational resources, "Man" is the most important one because an organisation needs "Man" (human resources/workforce) to manage and control the other organisational resources in an effective and efficient manner. People are therefore, the most essential resource in any organisation and they must be provided with a high QWL to ensure their satisfaction and motivation because satisfied and motivated employees perform better. Moreover, such employees are more committed and facilitate optimum and the most efficient utilization of the other organisational resources.

In the present scenario, Quality of Work Life is of paramount importance to the organizations, which are faced with immense competition, workforce diversity, and numerous and gigantic challenges like VUCA (volatility, uncertainty, complexity and ambiguity). Disruptive technologies like Artificial Intelligence, Virtual Reality, 3D Printing, Biometrics and Digital ID etc. have become a common place. In order to sustain and grow in this digital landscape, human skills and capacities must be continually developed. Such opportunities for use and development of human capacities at workplace, and opportunities for continued growth fall under the purview of QWL. In order to successfully capitalize on the opportunities that come to an organisation and to overcome the various challenges it is faced with, it has to ensure that its workforce is satisfied and motivated. Workforce of an organisation must be provided with a conducive and favourable work environment, adequate and fair compensation, opportunities for career growth and development, proper working conditions, adequate training, and recreational

facilities etc. This would enhance the quality of their work life and improve their satisfaction. This is imperative as people are the most important resource in an organisation, enabling the facilitation of optimum and the most effective use of other organisational resources. Therefore, the organisations must ensure that they take all the measures, which are necessary to maintain and improve the QWL of the workforce.

QWL, however, does not merely depend on these factors. Even the demographics of the workforce have a role to play. Hence, it is crucial to study the demographic profile of the workforce of an organisation, and to understand the impact of these demographic variables on the QWL.

The present study, thus, aims at assessing the QWL and impact of demographic variables on QWL of employees working in HPPCL- a promising power generating utility, recent in origin.

Quality of Work Life

The phrase "Quality of Work Life" gained impetus in the late 1960's to assess the effects of job on the overall health and general well-being of employees and as a means to positively influence the quality of a person's job related experience (Taneja & Kumari, 2012). The term "Quality of Work Life" was first conceived by Louis Davis. An International Conference on Quality of Work Life was held for the first time in Toronto in 1972, and the International Council for Quality of Work Life was established in the same year (Bindu & Yashika, 2014). In 1975, Richard E. Walton- a strong proponent of the concept of QWL, explicated the term by acknowledging eight exhaustive conditions of employment which denote a preferable Quality of Work Life. The eight conditions were also recommended by him to be used as standard criteria for measuring QWL. The eight criteria propounded by Walton to measure QWL are as following- safe and healthy working conditions, balanced role of work and personal life, constitutionalism in the work, organisational prospect to use and develop human capability, adequate and fair compensation, future opportunity for continued growth and security, social integration in the workplace and social relevance of work (Walton, 1975). Walton's eight dimensions of QWL can be rarely found to be complied with; however, organisations can seek to enhance their QWL by aiming at complying with these dimensions to the extent possible.

Robbins defined QWL as a process by which an organisation responds to employee needs by developing mechanisms to allow members to share fully in making decisions that design their lives at work (Robbins, 1989, p. 207).

According to Oeij & Wiezer (European Foundation for the Improvement of Living and Working Conditions), Quality of Work Life is a multi-dimensional construct, made up of a number of interrelated factors that need careful consideration to conceptualize and measure. It is related to those working conditions, which are considered ideal by the employees for achieving the four prime objectives, namely; job security, health and well-being, competence development, and work-life balance; and their effect on job satisfaction of the employees (Oeij & Wiezer, 2002). Thus, it is very crucial for the management to understand and fulfill its responsibility of providing the desirable work environment to the people working in an organisation.

Quality of Work Life is concerned with the quality of life at work; including the physical, mental, social, psychological and economic aspects of work. It has, thus, become indispensable for organisations in the present scenario to ensure a high QWL in order to reap the benefits of competitive advantage in the times of fierce competition and huge challenges.

Review of Related Literature

The results of study conducted by Saraji & Dargahi (2006) showed that majority of the employees were dissatisfied with the following QWL dimensions- occupational health and safety standards at work, support to employees by intermediate managers/supervisors, monetary compensation, trust in senior management, balance between work and family, stress experienced at work, and career prospects. The employees were, however, satisfied with the QWL dimensions of support from co-workers and the type of work they did. They also found in their study that dissatisfaction regarding QWL increased with age.

Bolhari, et al (2011) used Walton's eight QWL dimensions to measure QWL of IT workers. Their study revealed a moderate level of QWL among IT workers. QWL was found to be related to age, work experience and income of respondents. Gender and QWL were, however, found to be unrelated.

The findings of study carried out by Islam (2012) indicated that the QWL factors, namely; workload, family life, transportation, compensation policy and benefits,

working environment and working condition, and career growth had an influence on QWL, whereas the factor-colleagues and supervisor had no influence on QWL.

Balachandar, Panchanatham & Subramanian (2013) concluded in their study that the various demographic characteristics like the type of family, education and income of employees had an impact on their QWL. QWL was also found to improve job satisfaction, job performance, and employee involvement and productivity.

Chandra, Lakhawat & Vishwakarma (2013) found that QWL is essential for increasing job satisfaction, organisational commitment and involvement of the employees.

Gupta & Hyde (2013) concluded in their study that QWL was affected by age, experience and income of employees; and gender had no effect on QWL.

The results of study conducted by Nanjundeswaraswamy & Swamy (2013) revealed that demographic factors like gender, designation, department, salary and experience were not related to QWL of employees in private technical institutions. Their study also revealed that all the QWL dimensions were positively correlated with the QWL of faculties.

Raja & Kumar (2013) concluded in their study that the various factors, which influence the satisfaction with QWL in the organisation were- adequate income and fair compensation, safe and healthy working conditions, opportunities to use and develop human capacity, opportunity for career growth, constitutionalism in work organisation, social integration in the workforce, eminence of work life and social relevance of work, cordial relationship with employees and superiors, and remedy for grievance and performance appraisal.

Gupta (2015) revealed in her study that age, income and experience of employees had a significant positive impact on their QWL. Gender was found to have no impact on QWL.

Mehrotra & Khandelwal (2015) concluded in their study that female employees were more satisfied with the QWL in their institutions than their male counterparts. Gender and salary were found to have a relationship with each other and with QWL.

The results of study carried out by Aarthy & Nandhini (2016) showed that the faculty members experienced a moderate level of QWL. The various demographic characteristics, viz. age, gender, marital status, income, experience and number of children were found to significantly affect QWL.

Findings of the study conducted by Anyaoku (2016) showed that the QWL dimensions, which contributed significantly to the overall QWL were - opportunity for continued growth and security, social integration in the work organisation and social relevance of work. The QWL dimensions, namely, fair remuneration, training and retraining, provision for job performance and equal right issues were found to be the areas of dissatisfaction for the librarians. QWL was found to have a significant relationship with age, gender, institution type and experience. No relationship was found between QWL and educational qualification.

Jain (2016) revealed in her study that the workers were mostly dissatisfied with the QWL in the textile industries in Rajasthan. Work-related factors were found to have a significant impact on the QWL of workers. Age and gender were also found to have an influence on QWL. Findings of this study also revealed a significant gap between the existing and expected QWL of workers.

Ahmad (2017) revealed in his study that the relationship between QWL and the demographic variables, viz. age, period of service, income and educational qualification was statistically significant. It was also concluded that gender had no relation with QWL. QWL was found to be favourable for the employees.

Research Gap

A thorough review of literature related to QWL has been carried out. This meticulous review of related literature revealed that there is a paucity of research on the relationship between QWL and the demographic characteristics of workforce in a Public Sector Undertaking (PSU). HPPCL is an upcoming power generating utility in the state of Himachal Pradesh. It is recent in origin, but still it has shown a consistent growth since its inception. Keeping these facts in mind, a need was felt for carrying out a study on the impact of various demographic variables on the QWL of employees of HPPCL, as no such study was found to be carried out earlier.

Objectives of the Study

1. To assess the Quality of Work Life of employees of Himachal Pradesh Power Corporation Limited.

2. To measure the impact of some demographic variables, viz. age, gender, hierarchical level, income and marital status on the Quality of Work Life of employees of Himachal Pradesh Power Corporation Limited.

Research Hypotheses

H_{01} - There is no significant difference in the Quality of Work Life of the employees of HPPCL on the basis of gender.

H_{02} - There is no significant difference in the Quality of Work Life of the employees of HPPCL on the basis of age.

H_{03} - There is no significant difference in the Quality of Work Life of the employees of HPPCL on the basis of hierarchical level.

H_{04} - There is no significant difference in the Quality of Work Life of the employees of HPPCL on the basis of income.

H_{05} - There is no significant difference in the Quality of Work Life of the employees of HPPCL on the basis of marital status.

Research Methodology

For the present study, a sample of 150 HPPCL employees was selected from its various offices- the Corporate office, Design office and various projects in the state of Himachal Pradesh. Convenience sampling technique was used for selecting the sample. Both primary and secondary sources of data collection were used for carrying out the present study. A questionnaire consisting of two parts- the first part consisting of demographic profile of the respondents and the second part consisting of a questionnaire developed by Timossi, et al in their research study on "Evaluation of Quality of Work Life: An Adaptation from the Walton's QWL model"- was used for collecting primary data from the respondents. Reliability of the questionnaire was checked with the help of Cronbach's alpha, which was .921. Journals, publications, internet etc. were used as a source of secondary data. Frequencies, descriptive statistics, independent samples t-test and one-way ANOVA were used as tools for data analysis in SPSS software.

Results and Discussion

Results of descriptive statistics are presented in the following table 1:

TABLE 1: Descriptive Statistics

QWL Dimension	N	Mean	Std. Deviation
1. Adequate and Fair Compensation	150	3.61	.65
2. Safe and Healthy Working Conditions	150	3.77	.42
3. Immediate Opportunity to Use and Develop Human Capacities	150	3.91	.47
4. Opportunity for Continued Growth and Security	150	3.51	.64
5. Social Integration in the Work Organisation	150	3.98	.46
6. Constitutionalism in the Work Organisation	150	3.70	.66
7. Work and Total Life Space	150	3.98	.40
8. Social Relevance of Work Life	150	3.84	.52
Overall QWL	150	3.79	.42

Source: Primary/field data.

Results of Table 1 (Descriptive Statistics) show that the means of all QWL dimensions are more than 3.5, which means that the employees of HPPCL are quite satisfied with the various QWL dimensions as well as with the overall QWL (mean=3.79) in the organisation. This also signifies that all the eight QWL dimensions taken in the study contribute to the overall QWL of employees. "Social Integration in the Work Organisation" and "Work and Total Life Space" were found to have the highest means (3.98). On the contrary, "Opportunity for Continued Growth and

Security" was found to have the lowest mean score of 3.51.

The values of Standard Deviation lie within the range of 0.4-0.7, which indicates that the employees did not differ much in their opinions regarding the various dimensions of QWL.

Demographic profile of the employees of HPPCL is presented in table 2:

TABLE 2: Demographic Profile of Respondents (N=150)

Demographic Determinant	Categories	Frequency	Percentage
1. Gender	Male	128	85.3
	Female	22	14.7
2. Age (in years)	Below 30	27	18.0
	30-50	80	53.3
	Above 50	43	28.7
3. Hierarchical Level	W1-W5	11	7.3
	W6-W11	13	8.7
	S1-S4	36	24.0
	E0-E4	58	38.7
	E5-E9	32	21.3
4. Annual Income (in Rs)	<2 Lacs	12	8.0
	2-5 Lacs	39	26.0
	5-10 Lacs	74	49.3
	>10 Lacs	25	16.7
5. Marital Status	Single	30	20.0
	Married	120	80.0

Source: Primary/field data.

Results of Frequencies and Percentages calculated in Table 2 show that most of the respondents in the present study lie in the age group of 30-50 years (53.3%), 28.7% are above 50 years in age, and the remaining 18% are the young employees of age less than 30 years. The results further reveal that 85.3% of the respondents were males and 14.7% were females.

Results in Table 2 also show that 7.3% employees belonged to worker levels ranging from W1-W5, 8.7% belonged to W6-W11 levels, 24% belonged to S1-S4 supervisor levels, and 38.7% were executives falling in the executive levels ranging from E0-E4. The remaining 21.3% were the middle and higher management executives lying in the E5-E9 levels. Results also show that 8% employees had an annual income less than Rs. 2 lacs, 26% employees earned Rs. 2-5 lacs per annum, 49.3% employees earned Rs. 5-10 lacs per annum, and the other 16.7% had an annual income of more than Rs. 10 lacs.

Results in Table 2 also reveal that 20% of the respondents were single whereas 80% were married.

Results of Independent Samples t-test are presented in the table 3.

Table 3 presents the results of Independent Samples t-test which was used to assess the impact of Gender on QWL of employees. Results in this table show that there was no significant difference in the QWL of male and female employees (t_{148}=1.553, p=.123). The p-value in this case (.123) is greater than .05 level of significance, hence it can be concluded that gender has no impact on QWL.

Results of One-way ANOVA are presented in the following tables. Result of One-way ANOVA for measuring the impact of age on QWL of employees is presented in the following table 4.

The impact of age on QWL was assessed with the help of one-way ANOVA, the results of which are presented

TABLE 3: Independent Samples t-test for measuring the impact of Gender on QWL of employees

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig.* (2-tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Overall QWL	Equal Variances Assumed	.033	.855	1.553	148	.123	.14882	.09583	-.04056	.33820
	Equal Variances not Assumed			1.496	27.896	.146	.14882	.09946	-.05495	.35259

Source: Primary/field data.

TABLE 4: One-way ANOVA for measuring the impact of Age on QWL of employees

Overall QWL				
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.203	2	.101	.580	.561
Within Groups	25.731	147	.175		
Total	25.934	149			

Source: Primary/field data.

in Table 4. In this case, the p-value is .561 which is more than .05 level of significance, thus, it can be concluded that age has no impact on QWL.

Result of One-way ANOVA for measuring the impact of hierarchical level on QWL of employees is presented in the following table 5.

Table 5 presents the results of one-way ANOVA which was used for measuring the impact of hierarchical level of an employee on QWL. In this case, the p-value is .606 which is more than .05 significance level. Hence, it can be concluded that hierarchical level has no impact on QWL.

TABLE 5: One-way ANOVA for measuring the impact of Hierarchical Level on QWL of employees

Overall QWL				
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.478	4	.120	.681	.606
Within Groups	25.456	145	.176		
Total	25.934	149			

Source: Primary/field data.

Result of One-way ANOVA for measuring the impact of income on QWL of employees is presented in table 6.

Results of one-way ANOVA for assessing impact of income on QWL are presented in Table 6. The p-value for impact of income on QWL is .020, which is less than .05 level of significance. This shows that QWL is influenced by income of employees.

The results of Independent Samples t-test which was used to measure the impact of Marital Status on QWL of employees are presented in table 7.

TABLE 6: One-way ANOVA for measuring the Impact of Income on QWL of employees

Overall QWL					
	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	1.683	3	.561	3.378	.020
Within Groups	24.250	146	.166		
Total	25.934	149			

Source: Primary/field data.

TABLE 7: Independent Samples t-test for measuring the Impact of Marital Status on QWL of employees

		Levene's Test for Equality of Variances		**t-test for Equality of Means**						
		F	**Sig.**	**t**	**df**	**Sig.* (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	Upper
Overall QWL	Equal variances Assumed	4.536	.035	.231	148	.818	.01974	.08543	-.14908	.18856
	Equal Variances not Assumed			.204	39.120	.839	.01974	.09670	-.17584	.21532

Source: Primary/field data.

Results in this table show that there was no significant difference in the QWL of single and married employees ($t_{39.120}$=.204, p=.839). The p-value in this case (.839) is greater than .05 level of significance hence it can be concluded that marital status does not affect QWL of employees.

Conclusion

Giving due consideration to the ever increasing relevance of QWL for the organisations as well as employees, the present study has been conducted to study the QWL in an organisation (HPPCL) and to assess the impact of various demographic characteristics of the employees on their QWL.

Results of this study show that all the eight QWL dimensions taken in the study contribute to the overall QWL, which is in consonance with the results of a previous study (Raja & Kumar, 2013). Results also reveal that the employees of HPPCL are quite satisfied with all the QWL dimensions as well as the overall QWL in the organisation. This finding is in tune with the findings of some studies conducted in the past (Bolhari, et al, 2011; Aarthy & Nandhini, 2016; Ahmad, 2017). Findings of Independent Samples t-test show that gender has no impact on QWL. This result is consistent with the earlier findings (Bolhari, et al, 2011; Gupta & Hyde, 2013; Nanjundeswaraswamy & Swamy, 2013; Gupta, 2015; Ahmad, 2017). This result is, however, inconsistent with the findings of some previous studies (Mehrotra & Khandelwal, 2015; Aarthy & Nandhini, 2016; Anyaoku, 2016; Jain, 2016).

The Independent Samples t-test is also used to assess the impact of marital status on QWL and it is concluded that there is no significant difference in the QWL of single and married employees. This result is inconsistent with the finding of a study carried out in the past (Aarthy & Nandhini, 2016). Results of one-way ANOVA show that QWL is not influenced by the age of respondents. This finding is inconsistent with the results of some previous studies (Saraji & Dargahi, 2006; Bolhari, et al, 2011; Gupta & Hyde, 2013; Gupta, 2015; Aarthy & Nandhini, 2016; Anyaoku, 2016; Jain, 2016; Ahmad, 2017).

Results also reveal that hierarchical level of an employee has no effect on QWL. This result is in consonance with a previous finding (Nanjundeswaraswamy & Swamy, 2013). However, QWL has been found to be

influenced by the income of employees. This finding is consistent with the findings of some previous studies (Bolhari, et al, 2011; Balachandar, Panchanatham & Subramanian, 2013; Gupta & Hyde, 2013; Gupta, 2015; Mehrotra & Khandelwal, 2015; Aarthy & Nandhini, 2016; Ahmad, 2017). This finding was, however, inconsistent with a previous finding (Nanjundeswaraswamy & Swamy, 2013).

References

Aarthy, M. and Nandhini, M. (2016). "A Study on Quality of Work life among the Engineering College Faculty Members in Coimbatore District." *International Journal of Management Research & Review,* 6 (8), 1051-1057

Ahmad, S. (2017). "The Co-relation between QWL and Demographic Factors of Private University Employees in India." *PEOPLE: International Journal of Social Sciences,* 3 (2), 286-305

Anyaoku, E. N. (2016). "Demographic Variables of Quality of Work Life of Librarians Working in Nigeria." *International Journal of Advanced Library and Information Science,* 4 (1), 312-323

Balachandar, G., Panchanatham, N. and Subramanian, K. (2013). "Quality of Work Life- The Power of Insurance Company: Impact of Personal Factors on the Quality of Work Life of the Officers." *Management & Marketing,* XI (1/2013), 123-133

Bindu, J. and Yashika, S. (2014). "Quality of Work Life with Special Reference to Academic Sector." *Research Journal of Management Sciences,* 3 (1), 14-17

Bolhari, A., et al., (2011). "The Relationship between Quality of Work Life and Demographic Characteristics of Information Technology Staffs." *International Conference on Computer Communication and Management,* 374-378

Chandra, A., Lakhawat, P.S. & Vishwakarma, P. (2013). "Study on Measuring the Quality of Work Life among Third Grade Employees in Naini Industrial Area." *International Journal of Technological Exploration and Learning,* 2 (6), 318-321

Gupta, B. (2015). "An Empirical Study of Impact of Demographic Variables on Quality of Work Life among Insurance Sector Employees in Indore Division." *Pacific Business Review International,* 8 (1), 24-32

Gupta, B. and Hyde, A.M. (2013). "Demographical Study on Quality of Work Life in Nationalized Banks." *Vision,* 17 (3), 223-231

Islam, M. B. (2012). "Factors Affecting Quality of Work Life: An Analysis on Employees of Private Limited Companies in Bangladesh." *Global Journal of Management and Business Research,* 12 (18), 22-31

Jain, S. (2016, June 17). "A Study on Quality of Work Life among Workers in Textile Industries in Rajasthan." Retrieved from https://ssrn.com/abstract=2797287 or http://dx.doi.org/10.2139/ssrn.2797287 on April 15, 2018.

Mehrotra, R. & Khandelwal, V. (2015). "Association of Demographic Variables (Gender and Salary) On Quality of Work Life of Teaching Employees in Private Technical Institutions in Bareilly Region." *International Journal of Education and Science Research Review,* 2 (2), 01-06

Nanjundeswaraswamy, T. S. & Swamy, D. R. (2013). "Quality of WorkLife of Employees in Private Technical Institutions." *International Journal for Quality Research,* 7 (3), 431-441

Oeij, P. R. A. and Wiezer, N. M. (2002). "New Work Organisation, Working Conditions and Quality of Work: Towards the Flexible Firm?" (pp. 001-100). Retrieved from https://www.eurofound.europa.eu on April 4, 2018.

Raja, P.J.A. and Kumar, S.A. (2013). "A Study on Quality of Work Life of Employees in Steel Authority of India, Salem." *IJEMR,* 3 (6), 01-15

Robbins, S. P. (1989). *Organisational Behavior: Concepts, Controversies and Applications.* NJ: Prentice Hall, Englewood Cliffs

Saraji, G. N. and Dargahi, H. (2006). "Study of Quality of Work Life (QWL)." *Iranian Journal of Public Health,* 35(4), 08-14

Taneja, G. and Kumari, L. (2012). "Quality of Work life and its Relation with Job Satisfaction among Indian Banks." *International Journal of Research in Commerce and Management,* 3 (2), 097- 106

Walton, R. E. (1975). "Criteria for quality of work life." In Davis, L. E., A. B. Cherns and Associates (Eds.), *The Quality of Working Life.* Vol. 1 (pp. 091-104). New York: The Free Press

"Data are becoming the new raw material of business."

Craig Mundie

Feature

Productivity and Profitability Relationship in Sugarcane Cultivation: A State Level Analysis

P. PADMAVATHI AND A. NARAYANAMOORTHY

This paper analyses the relationship between productivity and profitability of sugarcane cultivation by using state-wise data covering the period of 1973-74 and 2014-15. The descriptive, correlation and regression analyses have been carried out to study the relationship. Descriptive analysis shows that the real value of human labour cost incurred for cultivating sugarcane was found to be high in Tamil Nadu state as compared to other states selected for the analysis. Further, correlation analysis reveals bullock labour cost and yield enhancing input costs, which are positively correlated among the selected variables specifically in Tamil Nadu, Uttar Pradesh and Karnataka states. Finally, the result of regression analysis carried out by using yield enhancing input cost suggests that it plays a dominant role in increasing the productivity of sugarcane, which in turn enhances profitability except Karnataka state for the entire period of analysis.

P. Padmavathi, Research Scholar, Department of Economics and Rural Development, Alagappa University, Karaikudi, Tamil Nadu

Dr. A. Narayanamoorthy, Professor and Head, Department of Economics and Rural Development, Alagappa University, Karaikudi, Tamil Nadu

DOI : - https://doi.org/10.32381/PROD.2019.59.04.6

1. Introduction

Sugarcane is one of the important commercial crops of the world. India ranks second among the sugarcane growing countries of the world particularly in terms of area and production. It is a renewable and natural agricultural resource providing not only sugar but also bio-fuel, fibre, fertilizer and myriad of byproducts/co-products with ecological sustainability. Sugarcane juice is used for making white sugar, brown sugar (khandsari), Jaggery (gur) and ethanol. At present, it is cultivated in 5.07 million hectares across India with annual production of 362 million tonnes during 2015-16 (GOI, 2016). A number of studies have proved that reduction in the incidence of rural poverty from 56.44% to 28.33% has improved production of agricultural commodities during 1972-73 and 2004-05 (Saleth *et al.,* 2003; Hussain and Hanjra, 2003, 2004). Due to green revolution, increased production of agricultural commodities has led to reduction of rural poverty to 28.11% during 1972-73 and 2004-05 (Ahluwalia, 1978; Narayanamoorthy, 2001). A tremendous increase in agricultural production during the post-green revolution era was possible due to the perceptive role, played by the Indian farmers (Swaminathan, 2008). India's agriculture is characterized by indebtedness, poor returns over cost of cultivation, crop failures, low prices for crops and farmers' suicides. A few important studies stated that the major reasons for the farmers' suicides are decline in productivity of crops, inadequate supply of institutional credit and imperfect market conditions (Deshpande, 2002; Deshpande and Prabhu, 2005; Vaidyanathan, 2008). Though some of the studies attribute farmers' suicides to behavioural and social factors, they have not analyzed how those factors incidentally affected the farming community in early nineties in India (Mohanty and Shroff, 2004). Further, few studies indicated that increased demand for human labour and high wage rate have increased the cost of cultivation of sugarcane crop

particularly in major growing areas resulting in negative returns (Ramasamy and Kumar, 2011).

Some studies have analysed the agricultural output with source-wise (tanks, canals and groundwater) impact of irrigation (Dhawan, 1988). However, a study based on cost of cultivation data particular to Maharashtra state proved significant decline in profitability of sugarcane between 1975-76 and 2001-2002 (Narayanamoorthy, 2007). The study analyzed the issue of profitability only for the sugarcane crop by using temporal data (Dev and Rao, 2010). Except those studies mentioned above, there are not many studies that utilised cost of cultivation data covering sugarcane crop for a longer period with a specific focus on productivity and profitability nexus during 1973-74 and 2014-15. Therefore, the present study has been carried out to analyse the relationship between profitability and productivity of sugarcane cultivation by using large coverage of data and selecting important cost variables in order to find out which one is highly influencing whether farmers reap any profit from sugarcane cultivation at all. In order to do an in-depth analysis, the cost of cultivation survey data collected by Commission for Agricultural Costs and Prices (CACP), which contains rich information on the cost and output on various crops on a temporal basis (Sen and Bhatia, 2004) will be utilized. Not many studies are available on productivity and profitability relationship in sugarcane cultivation by using secondary data. Keeping this in view, an attempt has been made to find out the relationship between profitability and productivity in sugarcane cultivation during 1973-74 and 2014-15.

2. Data and Methodology

Secondary data covering period from 1973-74 to 2014-15 has been extensively used for carrying out this study. This study quantifies the contribution of major factor inputs (such as, human labour cost, bullock labour cost, machine labour cost, yield enhancing input cost and fixed cost) over time across the states in India. The main objective of the study is to find out the productivity and profitability relationship of sugarcane crop cultivation, it also studies 4 major states of sugarcane crop cultivation in India. The data for this study has been compiled from the Commission for Agricultural Costs and Prices (CACP) published by the Ministry of Agriculture and Farmers Welfare, Government of India, New Delhi. In order to study the change in profitability of sugarcane cultivated in different states, all the cost related data of the crop has been converted into constant prices using CPIAL deflator at 1986-87 prices. The study also shows factors responsible for productivity and profitability of sugarcane cultivation in India over four and half decades. State-wise data on human labour cost (HLC), bullock labour cost (BLC), machine labour cost (MLC), yield enhancing input cost (YEIC) and fixed cost (FC) have been compiled from various reports of CACP. The reason behind the selection of those variables was that they would greatly influence the profitability of sugarcane cultivation in India. Due to data constraint, the study has utilized in total five variables for the analysis which are presented in Table 1.

The major objective of the study is to find out whether the profitability of sugarcane has a positive

TABLE 1: Descriptive Analysis of the Variables

Variables	Description	Unit	Averages			
			KN	TN	UP	MH
HLC	Human Labour Cost	Rs./ha	6263.11 (2325.25)	9004.18 (2262.50)	2553.36 (793.29)	4533.53 (1420.57)
BLC	Bullock Labour Cost	Rs./ha	435.88 (214.02)	107.26 (63.18)	255.04 (165.36)	649.43 (244.23)
MLC	Machine Labour Cost	Rs./ha	412.03 (305.87)	274.03 (105.43)	174.62 (83.68)	1189.86 (949.52)
YEIC	Yield Enhancing Input Cost	Rs./ha	3635.47 (1196.71)	4483.11 (576.10)	2130.29 (492.93)	5736.35 (1160.25)
FC	Fixed cost	Rs./ha	6003.83 (1034.53)	5561.85 (1112.26)	4246.78 (1112.30)	5126.31 (1719.39)

Sources: Computed using data from CACP (various years) and www.dacnet.nic.in.

correlation with the productivity of the crop. Initially, a total of seven major sugarcane producing states had been chosen for the study namely Andhra Pradesh, Haryana, Karnataka, Tamil Nadu, Maharashtra, Uttar Pradesh and Uttarakhand in India. After the systematic analysis of the data, major four sugarcane producing states had been selected such as Karnataka, Tamil Nadu, Uttar Pradesh and Maharashtra. Therefore, a total of four states belonging to the category of high area with high productivity (Karnataka and Tamil Nadu) and high area with low productivity (Uttar Pradesh and Maharashtra) have been considered for studying the aspect of profitability. Both descriptive, correlation and regression analysis were used for the analysis. The descriptive analysis has been carried out to trace the determinants of profitability of sugarcane cultivation in India. And then the correlation value explains the one to one relationship between profitability and other associated variables. In order to study the simple relationship between productivity and profitability, a linear regression is estimated, taking YIELD as the dependent variable and other factor inputs such as human labour cost, bullock labour cost, machine labour cost, yield enhancing input cost and fixed cost as the independent variable to the entire period of the analysis. The study attempts to find out whether or not the role of factor inputs plays a role in influencing the agricultural profitability that declines over the years. Therefore, after studying the determinants of profitability and correlation in sugarcane cultivation, the regression equation is estimated to know the contribution of each variable in it. Though the main objective of the study is to find out the contribution of various labour costs and other factors in agricultural profitability, it also aims at analyzing the changes that have taken place in agricultural sector since 1973.

3. Results and Discussion

As reported in methodology section, in order to find out the influence of various factors, we have computed descriptive, correlation and regression analysis by taking into account the important cost variables that are expected to have in sugarcane cultivation. The present study attempts to analyze the major changes in productivity and profitability relationship in sugarcane cultivation during 1973-74 and 2014-15. The main objective of the study is to find out the real change in productivity and profitability relationship with cost variables in sugarcane cultivation. According to Table 1, it is observed that the major structural changes have taken place in development of agriculture over the last four and half decades in India.

The variables considered for the analysis including its sources are presented in Table 1. Of the total variables considered for the study such as human labour cost (HLC), bullock labour cost (BLC), machine labour cost (MLC), yield enhancing input cost (YEIC), fixed cost (FC) with productivity have been compiled from the cost of cultivation survey data published by CACP (Commission for Agricultural Cost and Price). The reason behind the selection of these variables was that they would greatly influence the profitability of sugarcane cultivation in India. Therefore, HLC, BLC, MLC, YEIC, FC were chosen as independent variables in the analysis, YIELD was selected as the dependent variable. The descriptive analysis carried out for the above mentioned States from 1973-74 to 2014-15. To study the influence of profitability determinant variables on sugarcane cultivation, a brief discussion is necessary to explain the profitability determining variables in all states.

To begin with, it is observed from Table 1 that on an average human labour cost is found to be significantly high about Rs. 9004.18/ha particularly in Tamil Nadu state. Whereas the same stands about Rs. 6348.88, Rs. 4442.98 and Rs. 2499.61/ha in the remaining states: Karnataka, Uttar Pradesh and Maharashtra respectively during the entire period of analysis. Consequently, standard deviation is also evaluated in the study in order to obtain more clarification as well as clear idea about the data throughout period of analysis. In Uttar Pradesh, standard deviation in human labour cost is found to be very low about Rs. 793.29/ha as compared to its counterpart states. An average of increased human labour cost implies that the bullock labour cost in sugarcane cultivation had declined to Rs. 107.26, Rs. 435.88, Rs. 255.04 and Rs. 649.43/ha in Tamil Nadu, Karnataka, Uttar Pradesh and Maharashtra respectively. While evaluating standard deviation on bullock labour cost, it is found to be very low about Rs. 63.18/ha and Rs.165.36/ha in Tamil Nadu and Uttar Pradesh, respectively. Similarly, the average of machine labour cost in Maharashtra state is very high about Rs. 1189.86/ha. But, the same is found to be low about Rs. 412.03, Rs. 274.03 and Rs. 174.62/ha in the remaining states, namely Karnataka, Tamil Nadu and Uttar Pradesh correspondingly.

The standard deviation of machine labour cost is evaluated to be very low about Rs. 83.68 specifically in Uttar Pradesh state when compared to other states under study. Further, the same parameter is about Rs. 105.43, Rs. 305.87 and Rs. 949.52/ha to the remaining states

namely Tamil Nadu, Karnataka, and Maharashtra respectively. It is very clear from the analysis that human labour cost plays a vital role among the above mentioned labour cost variables in sugarcane cultivation to the entire period of analysis. Subsequently, an average of fixed cost is also found to be very high about Rs. 6003.83/ha in Karnataka, which is followed by other states under study. The standard deviation on fixed cost is very low about Rs. 1034.53/ha in Karnataka state. Finally, an average of yield enhancing input cost in Maharashtra is found high about Rs. 5736.35/ha in sugarcane cultivation. Whereas standard deviation of yield enhancing input cost is estimated to be low about Rs. 492.93/ha in Uttar Pradesh state when compared to other states under study and then in Tamil Nadu state which is about Rs. 576.10/ha. On the whole, the study reveals that the average human labour cost in sugarcane cultivation is found to be very high about 9004.18/ha in Tamil Nadu state. And then the average fixed cost is also evaluated to be high about Rs. 6003.83/ha in Karnataka state to the entire period of analysis. Whereas standard deviation on bullock labour cost in sugarcane cultivation is very low about Rs. 63.18/ha in Tamil Nadu, the standard deviation on machine labour cost is evaluated to be low about Rs. 83.68/ha throughout the period of analysis.

The result of the study shows that the human labour cost in real value incurred for cultivating sugarcane is found to be very high particularly in Tamil Nadu among the states selected for the analysis. The study reveals that the changes taken place in development of total agricultural factor inputs over the years across the states must have played a significant role in increasing the agricultural profitability, that also significantly vary across the states.

Consequently, standard deviation is also evaluated in order to obtain more clarification as well as clear idea about the data throughout the period of analysis. However, this is not a sufficient justification to express that agricultural factor inputs determine the profitability of sugarcane cultivation in India. As mentioned in methodology section, the correlation value explains the one to one relationship between profitability and other associated variables, let us first study correlation before getting into the analysis of regression results. The correlation value between profit and other above mentioned variables presented in Table 2 shows that the intensity of profitability vary considerably. First of all, results obtained out of the correlations presented in Table 2, have shown a significant positive and negative relationship with the percentage of profitability from 1973-74 to 2014-15.

Uttar Pradesh and Maharashtra have been used for the entire analysis of the study. The study also attempts to identify association among the agricultural inputs like HLC, BLC, MLC, YEIC and FC with yield from 1973-74 to 2014-15. It is found that the human labour cost (HLC) in sugarcane cultivation is highly and negatively related to the states namely Tamil Nadu, Uttar Pradesh, Maharashtra except Karnataka. In Karnataka the human labour cost is estimated to be non-significantly correlated to period of the analysis. The corresponding figure for bullock labour

TABLE 2: Correlation Values: Productivity with other Selected Variables

Independent variables	Description of the variables	Unit	Karnataka	Tamil Nadu	Uttar Pradesh	Maharashtra
(1)	(2)	(3)	(4)	(5)	(6)	(7)
HLC	Human labour cost	(Rs./ha)	0.12^{ns}	-0.81^{a}	-0.83^{a}	-0.73^{a}
BLC	Bullock labour cost	(Rs./ha)	0.01^{ns}	0.59^{a}	0.77^{a}	-0.46^{a}
MLC	Machine labour cost	(Rs./ha)	-0.38^{c}	-0.54^{a}	-0.77^{a}	-0.78^{a}
YEIC	Yield enhancing input cost	(Rs./ha)	0.29^{d}	-0.46^{ns}	-0.62^{a}	-0.21^{ns}
FC	Fixed cost	(Rs./ha)	-0.02^{ns}	-0.08^{ns}	-0.79^{a}	-0.45^{a}
N	No. of observations	(years)	20	19	28	30

Sources : Computed using data from CACP (various years) and www.dacnet.nic.in .

Notes : a, b, c and d are significant at 1 per cent, 5 per cent, 10 per cent, 20 per cent level, respectively; ns-not significant.

cost has been highly and positively related in Tamil Nadu and Uttar Pradesh. But, Maharashtra, which is found to be highly and negatively correlated with their counterpart states under study. While evaluating Karnataka, the bullock labour cost is found to be negatively and non-significantly related in sugarcane cultivation. In case of machine labour cost Tamil Nadu, Uttar Pradesh, Maharashtra are found to be highly and negatively related to the entire period of the analysis. Karnataka is also similar to the above mentioned states but there is a slight variation in degree of correlation in sugarcane cultivation. It follows that the yield enhancing inputs cost is positively related particularly to Karnataka when compared to other states, but it is negatively related with a high degree in Uttar Pradesh.

However, in the remaining two states Tamil Nadu and Maharashtra, yield enhancing input cost are negatively non-significantly correlated in sugarcane cultivation. Finally, fixed cost seems to be negatively and significantly related to Uttar Pradesh and Maharashtra but it is not considered in Karnataka and Tamil Nadu, which is found to be negatively non-significant for the analysis period. Correspondingly, state-wise analysis of the correlation results of the following variables like HLC, BLC, MLC, YEIC and FC proved that association of the variables and its significant levels during the period of analysis. In Karnataka, the yield enhancing input cost is found to be positively and significantly related with when compared to other states. Thus, machine labour cost is found to be negatively significant correlated in sugarcane cultivation. At the same time, the remaining variables namely HLC, BLC and FC are non-significantly related to the entire period of the analysis. It is followed that BLC is positively and significantly related in sugarcane cultivation. And then human labour cost and machine labour cost are highly and negatively related to Tamil Nadu. It is surprising to note that HLC, MLC, YEIC and FC seem to be highly and negatively related except BLC in Uttar Pradesh. However, BLC in Uttar Pradesh is highly and positively related to the entire period of the analysis. An interesting point here to be noted is that HLC, BLC, MLC and FC are demonstrated highly and negatively related except YEIC in Maharashtra during 1973-74 and 2014-15.

Therefore, it can be concluded that BLC and YEIC are the two agricultural inputs positively correlated among the variables selected for the study to all the states. The purpose of the study was not only to quantify the impact of profitability but also to examine whether the costs-influencing factors how much it has on the impact of profitability fluctuations over the years. As observed at the national level, structural changes have also taken place across different states in India. The states considered for the analysis, where human labour cost has increased significantly but the use of BLC and MLC have reduced sharply in sugarcane cultivation, confirming the fact that human labour cost is used to substitute bullock and machine labour costs. Similarly, the YEIC and FC incurred for cultivating sugarcane crop has risen substantially in all the states under study. The profitability of sugarcane crop of the four states Karnataka, Tamil Nadu, Uttar Pradesh and Maharashtra are depicted in Fig. 1. The figure clearly shows that fluctuations in the profitability of sugarcane cultivation for the long period during 1973-74 and 2014-15.

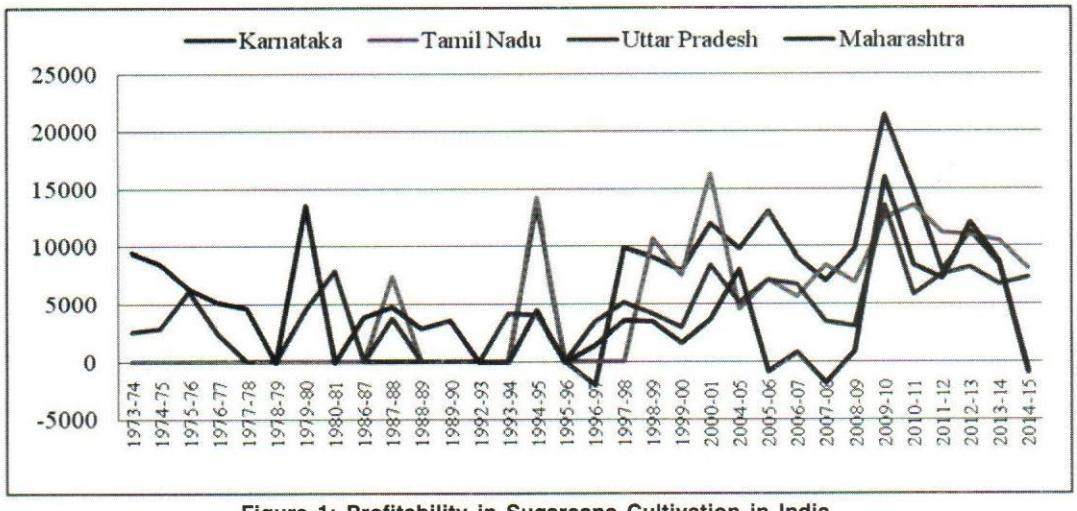


Figure 1: Profitability in Sugarcane Cultivation in India

Sources: Computed using data from CACP (various years) and www.dacnet.nic.in .

One of the major objectives of the study is to capture the real influence of agricultural factor inputs on yield with the conventional method of irrigation of sugarcane cultivation. It is a known fact that productivity of any crop is determined by a number of different agricultural factor inputs. Besides yield enhancing factors, personal characteristics of the farmers (human resource factors) as well as managerial factors also influence productivity of crop. As per the earlier studies, the method of cultivation (conventional method) also considerably influences crop productivity. Keeping this in view, as mentioned in the methodology section, regression equation has been estimated and also taking together all the five factors namely HLC, BLC, MLC, YEIC, FC with yield for the four states in order to find out the contribution of each factor inputs on sugarcane productivity. Therefore, an attempt has been made here to assess the average contribution of agricultural factor inputs on productivity in agriculture across the states. This is done by using statistical analysis (panel regression analysis), which provides us better pictures on the time over changes in the above relationship between agricultural productivity and other factor inputs. For the purpose of ready reference, the regression equation used for studying the impact of conventional method of sugarcane cultivation on productivity is presented below:

$$YIELD = a + b_1HLC + b_2BLC + b_3MLC + b_4YEIC + b_5FC$$

HLC - Human labour cost (Rs./ha)

BLC - Bullock labour cost (Rs./ha)

MLC - Machine labour cost (Rs./ha)

YEIC - Yield enhancing inputs cost (Rs./ha)

FC - Fixed cost (Rs./ha)

YIELD - Yield (Qtl/ha)

In general, it has been understood from regression results that agricultural inputs determine productivity as compared to human labour cost, bullock labour cost, machine labour cost and fixed cost which was expected in the study. As followed in the descriptive and correlation analysis, regression analysis has been estimated separately using yield as dependent variable and other factor inputs such as HLC, BLC, MLC, YEIC, FC as independent variables. It is expected that all the five independent variables included in the regression equation one way or the other are expected to influence sugarcane productivity, the influence of conventional irrigation method of sugarcane cultivation on its productivity is expected to vary under different states. Therefore, the study has analysed the contribution of factor inputs from 1973-74 to 2014-15 with the use of regression estimates which is presented in Table 2. The value of adjusted R^2 estimated using the data of different states varies from -0.012 to 0.86 indicating that the variables included in the equation seem to be appropriate in explaining the variation in sugarcane productivity.

The regression results are satisfactory from a statistical perspective as the values of R^2 estimated to be very high for all the states except Karnataka. Studies carried out using cost of cultivation survey data empirically show that the cost of cultivation for different crops has increased at a faster rate than the rate of increase in farm income (Narayanamoorthy, 2007, 2013 and 2017). Hence, the impact of profitability and productivity relationship was analyzed by using regression analysis. The regression results, displayed in Table 3, show that for Karnataka, the R^2 was found to be 0.25 implying that 25% of variation in the yield could be explained by the independent variables. Similarly, variation in yield about 82% in Tamil Nadu, 88% in Uttar Pradesh and 70% in Maharashtra were influenced by the independent variables together. This shows that all the independent variables together determined most of the yield in Uttar Pradesh followed by Tamil Nadu, Maharashtra and Karnataka.

Firstly, human labour cost is negatively and significantly influencing the yield in Tamil Nadu, Maharashtra and Uttar Pradesh but, the same is insignificant in Karnataka. The study reveals that HLC have a less impact on productivity in sugarcane cultivation. Secondly, BLC in sugarcane cultivation is positively and significantly influencing the yield only in Uttar Pradesh. But it is found to be negative and insignificant in Karnataka, Tamil Nadu and Maharashtra. Thirdly, machine labour cost is found to be negatively non-significant in Karnataka, Tamil Nadu and Maharashtra. However, it is positively and significantly influencing the yield in Uttar Pradesh. Among the labour cost variables such as human labour cost, bullock labour cost and machine labour cost is found to be highly and negatively significant in all the states. And then human labour cost is negatively and significantly influencing the yield in Tamil Nadu, Uttar Pradesh and Maharashtra but, the same is negative and insignificant to Karnataka. In the case of bullock labour cost, it is found to be negatively and insignificantly influencing the yield in all the states except Uttar Pradesh. Fourthly, yield enhancing input cost is the only variable which is found to be highly and positively significant in Tamil Nadu, Uttar

Pradesh and Mahrashtra whereas the same is positive but insignificant in sugarcane cultivation in Karnataka. Finally, fixed cost is negatively and significantly influencing the yield particularly in Tamil Nadu but the same is found to be positively and insignificantly influencing in Karnataka and Maharashtra. Furthermore, fixed cost negatively and insignificantly influenced the productivity in Uttar Pradesh. Almost regression coefficients of all independent variables, used for the analysis, turned out to be negatively significant except yield enhancing input cost of all the states under study. The regression results reveal that human and machine labour cost have turned out to be negative and significant in all the states except Karnataka. Moreover, the study reveals that independent variables are negatively and significantly influencing the productivity in sugarcane cultivation throughout the period of analysis. The regression results also demonstrated the fact that among the selected variables, yield enhancing input cost is the only variable, which determines the productivity positively and significantly in all the states selected for the study except Karnataka. This shows that yield enhancing input cost is very crucial for enhancement of productivity as well as profitability of sugarcane cultivation which was expected in the study. The study has shown that HLC plays a dominant role in productivity than other selected variables in sugarcane cultivation under descriptive analysis. Finally, the regression analysis carried out by using yield enhancing input cost proved that it plays a dominant role in increasing the productivity as well as profitability in sugarcane cultivation to the entire period of the analysis except Karnataka.

TABLE 3: Factors Determining Profitability - Regression Results

Independent Variables	Regression Results			
	Karnataka	Tamil Nadu	Uttar Pradesh	Maharashtra
(1)	(2)	(3)	(4)	(5)
HLC	-0.005 (-0.54)^{ns}	-0.05 (-4.55)^{a}	-0.18 (-3.03)^{a}	-0.17 (-1.68)^{c2}
BLC	-0.12 (-1.03)^{ns}	-0.64 (-0.97)^{ns}	0.34 (2.27)^{b}	-0.14 (-0.47)^{ns}
MLC	-0.11 (-1.62)^{d}	-1.09 (-2.78)^{a}	-1.19 (-3.03)^{a}	-0.26 (-2.09)^{b}
YEIC	0.02 (-1.30)^{ns}	0.14 (2.34)^{b}	0.15 (2.02)^{b}	0.15 (2.30)^{b}
FC	0.00 (0.01)^{ns}	-0.04 (-1.55)^{d}	-0.008 (-0.22)^{ns}	0.02 (0.51)^{ns}
Constant	272.10 (2.02)^{c}	746.35 (2.66)^{b}	553.47 (5.66)^{a}	742.80 (2.38)^{b}
R^2	0.25	0.82	0.88	0.70
Adjusted R^2	-0.012	0.75	0.86	0.64
D-W	0.629	1.026	1.725	0.69
F	0.95	11.93	34.17	11.39
N	20	19	28	30

Sources: Computed using data from CACP (various years) and www.dacnet.nic.in .

Notes: a, b, c and d are significant at 1 per cent, 5 per cent, 10 per cent, 20 per cent level, respectively; ns-not significant.

4. Conclusion

It is widely believed that farming profitability remains poor in most parts of India and this adversely affects sustainable yield. The nature of linkages between productivity and profitability, however, remain debatable. Against this backdrop, this paper attempted to better understand the linkages between productivity (and profitability) and agricultural inputs in sugarcane cultivation by using

secondary data during 1973-74 and 2014-15. Both descriptive, correlation and regression analyses were carried to study the relationship. Descriptive analysis shows that average of human labour cost was found to be very high in Tamil Nadu state as compared to the other factor inputs in sugarcane cultivation over the years. In most states where the HLC has increased substantially, the use of bullock labour cost and machine labour cost has declined sharply in sugarcane cultivation, confirming the fact that HLC is used to substitute bullock and machine labour costs over the years. Further, the correlation analysis has shown that bullock labour cost and yield enhancing input cost are positively correlated among the selected variables particularly in Tamil Nadu, Uttar Pradesh and Karnataka. Finally, the results of regression analysis revealed that yield enhancing input cost is stronger in influencing productivity and profitability relationship than other factor inputs in all the states except Karnataka state to the entire period of the analysis. Therefore, authorities, policy makers and researchers should emphasize on development of yield enhancing input cost through quality seeds, fertilizers, pesticides and improved micro-irrigation method particularly drip irrigation method increases productivity of sugarcane, which in turn enhances profitability of large number of farmers in India.

References

Ahluwalia, M.S., "Rural Poverty and Agricultural Performance in India", *Journal of Development Studies*, Vol.14, No.2, April, pp.298-323, 1978.

Commission for Agricultural Costs and Prices (CACP)), *Report of the Commission for Agricultural Costs and Prices*, Ministry of Agriculture, Government of India, New Delhi, various years.

Dhawan, B.D., "Indian Irrigation: An Assessment", Economic *and Political Weekly*, Vol. 23, No. 19, May 7, pp. 965-967+969-971, 1988.

Deshpande, R.S., "Suicide by Farmers in Karnataka: Agrarian Distress and Possible Alleviatory Steps", *Economic and Political Weekly*, Vol. 37, No. 26, pp. 2610-2610, 2002.

Deshpande R.S. and N. Prabhu., "Farmers' Distress: Proof beyond Question", *Economic and Political Weekly*, Vol. 40, Nos. 44-45, pp. 4663-65, 2005.

Dev S. M. and N.C. Rao., "Agricultural Price Policy, Farm Profitability Security", *Economic and Political Weekly*, Vol. 45, Nos. 26 & 27, 26 June, pp.174-182, 2010.

Government of India., *Agricultural Statistics at a Glance: 2009*, Directorate of Economics and Statistics, Ministry of Agriculture, Government of India, New Delhi, 2016.

Hussain, I and M.A. Hanjra., "*Does Irrigation Water Matter for Rural Poverty Alleviation? Evidence from South and South-East Asia*", *Water Policy*, Vol. 5, No. 5-6, pp.429-442, 2003.

Hussain, I and M. A. Hanjra., "*Irrigation and Poverty Alleviation: Review of the Empirical Evidence*", *Irrigation and Drainage*, Vol. 53, No. 1, pp. 1-15, 2004.

Mohanty, B.B. and S. Shroff , "Farmers' Suicides in Maharashtra", *Economic and Political Weekly*, December 25, pp. 5599-5606, 2004.

Narayanamoorthy, A., "Irrigation and Rural Poverty Nexus: A Statewise Analysis", *Indian Journal of Agricultural Economics*, Vol. 56, No.1, January-March, pp.40-56, 2001.

Narayanamoorthy, A., "Deceleration in Agricultural Growth: Technology Fatigue or Policy Fatigue?", *Economic and Political Weekly*, Vol. 42, No. 25, June 23, pp. 2375-2379, 2007.

Narayanamoorthy, A., "Profitability in Crops Cultivation in India: Some Evidence from Cost of Cultivation Survey Data", *Indian Journal of Agricultural Economics*, Vol.68, No.1, January-March, pp.104-121, 2013.

Narayanamoorthy, A., "Farm Income in India: Myths and Realities"?, *Indian Journal of Agricultural Economics*, Vol. 72, No. 1, January-March, pp. 49-75, 2017.

Ramasamy, C. and D.S. Kumar., "Regional Differences in Farm Profitability in Indian Agriculture", in B.C. Barah and S. Sirohi (Eds.) (2011), *Agrarian Distress in India: Problems and Remedies*, Concept Publishing Company Pvt. Ltd. New Delhi, pp. 21-47, 2011.

Saleth, M.R., M. Samad, D. Molden and I. Hussain., "Water, Poverty and Gender: A Review of Issues and Policies", *Water Policy*, Vol. 5, No. 5-6, pp. 385-98, 2003.

Sen Abhijit and M.S. Bhatia., *Cost of Cultivation and Farm Income in India*, Academic Foundation, New Delhi, 2004.

Swaminathan, M.S., "Ending the Debt Trap and Attaining Food Security", *The Hindu*, March 3, p.12, 2008.

Vaidyanathan, A., "Farm Loan Waiver: A Closer Look and Critique", *The Hindu*, March 6, p. 11, 2008.

"The goal is to turn data into information, and information into insight."

Carly Fiorina

Feature

Macro Mechanisms in the Promotion of Chinese Outward Direct Foreign Investment

QING YANG AND BADAR ALAM IQBAL

This paper analyzes in detail relevant mechanisms that influence the development of Chinese Outward Foreign Investment (OFDI). Since these mechanisms directly affect the development of Chinese transnational corporations, it is of great significance to establish and improve such mechanisms. The mechanisms relating to the development of OFDI mainly include: foreign exchange administration, legal systems, the system of administrative examination and approval, monetary support system, tax system, insurance system and related service support system. After analyzing their current situations in China and the influences they have on the development of OFDI, the paper points out existing problems, discusses the problems' after effects and illustrates the roles of the relevant mechanisms. In addition, the paper examines the practices and experiences of other countries in this field and presents suggestions and ideas about how to establish and improve the mechanisms in the growth of Chinese transnational corporations.

Professor Qing Yang, School of Economics and Management, South China Normal University, Guangzhou City, Guangdong, China.

Prof. Dr. Badar Alam Iqbal, Visiting Professor, Department of Marketing and International Business, Makerere University Business School, Uganda

DOI : - https://doi.org/10.32381/PROD.2019.59.04.7

1. Introduction

In the process of economic globalization, transnational corporations function as carriers of product flow, service flow and capital flow, and they also act as agents of the spread of knowledge, information and techniques. In the course of promoting global economy, transnational corporations themselves net significant profits. Meanwhile, while promoting the development of global economy, they also make a positive contribution to the welfare of their own countries. Along with the development of economic globalization, markets will be integrated into one — the global market. As a result, resources will become global, while enterprises are of apparent national attributes. Developed countries, such as America and Japan, which have large number of transnational corporations, are the biggest winners in the process of economic globalization. Therefore, it is, realistically, an urgent task for China to develop its own transnational corporations to take part in international competitions to strive for a share in global markets and global resources.

China just has more than 30 years' experience in the growth of its transnational corporations, which was characterized by high-speed economic development. Over the past 30 years, the increases in both the total number of Chinese overseas enterprises and the average growth rate of foreign direct investments (FDI) have been more than 50%. In 2011, the Outward Foreign Direct Investment (OFDI) of China is $74.7 billion at No.6 in the world while the OFDI of China is $87.8 billion at No.2 in the world 2012. However, behind the high-speed development in cross-national operations of Chinese enterprises, there exist problems such as the small scale of enterprises, inefficient operational levels, small scale investments, and poor profitability, which are in contrast to the fast-growing Chinese economy. In fact, the stock of Chinese OFDI is

$532 billion at the end of 2012 to be only one-tenth of the United States.

Consequently, besides forming and strengthening competitive advantages to enable enterprises themselves to meet the requirements and to have the competences of cross-national operations, China should also establish and form a mechanism that is beneficial for rapid growth of Chinese transnational corporations. This paper emphasizes the outer requirements and environments for development of OFDI, which amounts to analyzing the growth mechanism of Chinese transnational corporations. Chart 1 illustrates the roles of various relevant mechanisms as follows.

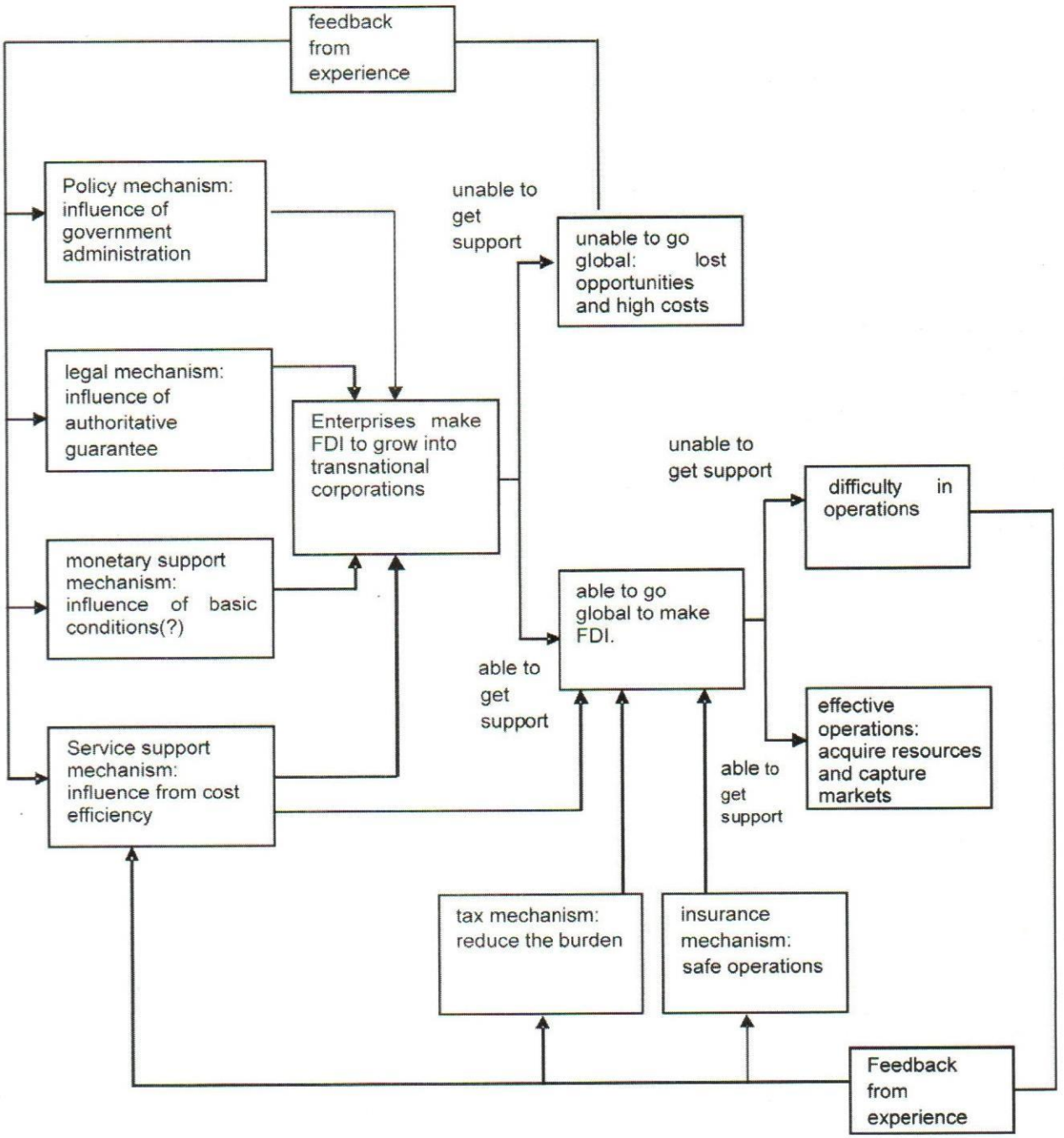


Chart 1. Mechanisms to development of OFDI in China

2. Policy Mechanism —A Functional Mechanism of Government Administration

Governmental policies play a very direct and pivotal role in the development of OFDI. As under current governmental policies, more systematic theories have not been developed yet on the research of the development of transnational corporations, there is no established standpoint. But the policies from any country, which aim at boosting the development of OFDI are out of considerations of overall national strategies. As the behavior corpus of development of OFDI is the enterprise, in countries with a typical market economy, previous cross-national operations of the enterprise were absolutely corporate behaviors with minimal influences from governmental policies. This is the reason why there are few theories relating policies to the development of OFDI in the theories of transnational corporations. In reality, once the government realizes the overall long-term strategic significance brought about by the cross-national operations of its transnational corporations, the promotional and mechanistic role of policies will emerge. But the present relative polices undertaken by the Chinese government are adverse to the development of OFDI and therefore are required a modification.

2.1 Foreign Exchange Administration

The purpose of a foreign exchange administration in a country is for maintenance of its international balance of payment equilibrium, maintenance of an orderly exchange rate arrangement and financial stability. To some degree, strict foreign exchange administration is the outcome of economic underdevelopment and inadequate opening up. When the ability to earn foreign exchange is insufficient, along with disconnection between domestic and foreign price systems and unsteady macro financial systems, it is necessary to establish and strengthen the system of foreign exchange administration. Apparently, existence of foreign exchange administration will inevitably lead to a conflict against practical economic development, especially in the period of faster economic development. At present, there is such a conflict between the development of offshore investments and the foreign exchange administration in China.

2.1.1 The Functions of Foreign Exchange Administration in China

The functions of foreign exchange administration departments in China in relation to offshore investments mainly include:

(1) Evaluate risks and sources of foreign exchange after offshore investment project is put forward;

(2) Foreign exchange administration after offshore investment project is approved;

(3) Financing administration of offshore investment project;

(4) Administration of deposits remitted by offshore investors.

2.1.2 Drawbacks of Current Foreign Exchange Administration

The current foreign exchange administration discourages enterprises from making offshore investments, which is reflected in the following aspects:

(1) The over-strict foreign exchange control, affecting offshore investment.

According to the regulations, when enterprises make offshore investments, initially, those who have foreign exchanges make investments with their own foreign exchange; those who lack foreign exchange can make foreign direct investments with a loan. Enterprises are also encouraged to make real investments or are allowed to export without settlement exchange. However, the overwhelming majority of Chinese enterprises have a limited foreign exchange, with insufficient competences and conditions to raise loans for offshore investments and great difficulty in real investments. Then the investment by product export without settlement exchange, which is surely a good way to solve the problem of funds for offshore investments to some extent, is restricted by the sales capability of the enterprises. In addition, products exported will not turn into investment capital until they are sold out. Consequently opportunities for investments on many projects will be lost as the foreign exchange funds cannot be acquired in time. Obviously, these regulations are unfavorable for enterprises to expand the scale of their offshore investments and to operate efficiently. Moreover, the current regulations, which also prohibit enterprises from purchasing foreign exchange for investment on the ownership of offshore shares, are more likely to cause enterprises to lose valuable opportunities in the respect of holding shares or cooperative projects.

(2) The system of deposits remitted from profits is adverse to enterprises' capital turnover.

According to the rules, in offshore investments, 5% of the total amount of foreign exchange funds remitted out will be kept as a deposit; profits from offshore investments and other foreign exchange incomes should be transferred back to home after 6 months of the end of the local fiscal year. As an important part of China's foreign exchange administration, the system of deposits remitted from profits is to guide enterprises to remit their earnings from offshore investments to home, and to improve the conditions of payments balance. But impersonally, the system results in an overstocking of lots of enterprises' funds, causing enterprises to have trouble in funds turnover and influencing their normal operations. For the enterprises that have just begun their investments from the scratch, it is hard for most of them to make a profit. The overstocked deposits will undoubtedly increase the burden of the enterprise. Most enterprises that are able to make profits after hardships hope to reinvest, expand the scale and further increase their profitability. If the reinvestments are made after profits are remitted, it will go through complicated procedures.

(3) The restriction of international business loans on enterprises causes enterprises to be unable to take full advantage of the international capital market to conduct financing.

Making good use of international capital market to conduct overseas financing is one method that should be adopted by Chinese enterprises lacking enough capital, and is also typically how transnational corporations operate. But the foreign exchange control authorities in China place the international business loan used by domestic enterprises for offshore investments in the category of foreign loan administration. A strict examination and approval system is implemented and domestic enterprises are obstructed from making good use of the outer capital market, it is adverse for domestic enterprises to use international business loans for offshore investments. In reality, by taking advantage of the international capital market, enterprises can often acquire more advantageous financing terms, achieve lower financing costs and obtain important information related to projects to redound to enhancing enterprises' international competitive advantages.

2.1.3 Improvements in Foreign Exchange Administration

In 2003, the Bureau of National Foreign Exchange Administration affirmed 10 provinces as the experimental provinces for the reformation of foreign exchange administration. The experimental provinces possess the foreign exchange quota from US$50 million to US$200 million. Within the quota, projects below $3 million foreign exchange are examined and approved by the local branch of foreign exchange administration. The problem that enterprises purchase foreign exchange for investments is one part of such an experiment. For instance, Beijing foreign exchange administration department acquired a quota of $200 million foreign exchange for domestic enterprises for offshore investments, which helped solve the Shougang's problem of purchasing foreign exchange for the ownership of shares. The problem regarding profit remitted is also one main part of such an experiment. Domestic investors can decide independently whether to remit profits back from offshore investments or to change profits into local capital to expand the investment scale. This can be realized by reporting to foreign exchange administration departments for records. A series of reformation in the foreign exchange administration, an important mechanism for growth and promotion, stimulates the development of Chinese enterprises' cross-national operations and the growth of Chinese transnational corporations. But it needs a time to extend the countrywide.

2.2 Administrative Examination and Approval

The formation and real situations of China's administrative examination and approval about offshore investments are influenced by historical as well as practical factors. Firstly, China's overall economic strength is not strong enough to reach the mature stage for FDI. Secondly, at present, China is not qualified for the conditions of capital items convertibility. Although the level of China's exchange reserve and balance of international payment position are no longer the main factors, which restrict direct offshore investment. Capital items convertibility should be carried out step by step. Otherwise, great risks will be brought about. Finally, the dual-economy characteristics in China's economy are rather obvious and industrialization is still under way while the domestic economy requires a lot of

funds, etc. Therefore establishment of the system of administrative examination and approval about offshore investments is the outcome of the above factors. But in the file in the mid-1980s *About the Authority and Principles of Examination and Approval on the Setup of Non-Commerce Joint Ventures Abroad and in Hong Kong-Macao Areas*, which was issued in May, 1984, and the file in the late-1990s concerning *Encouraging Enterprises to Process Materials and Conduct Assembly with stuff to Overseas*, which was issued in Feb., 1999, many parts of the system of administrative examination and approval of offshore investments have been lagging behind the development of new situations, which are specified as follows:

(1) The excessive administrative levels of examination and approval caused the opportunities to be missed out and the costs to be increased.

(2) Examination and approval are overly detailed and repetitive, and the system of examination and approval is opacity, which lead to inefficient examination and approval.

(3) It is factually difficult for enterprises to acquire the relevant preferential policies.

Administrative examination and approval is necessary, but the criteria to evaluate its effectiveness are whether it is efficient and the formulation of policy is favorable for enterprises to increase their competition advantages in offshore operations.

2.3 Assets Administration

Assets administration is one part of policy mechanism for the growth of Chinese transnational corporations. Its basic logic is that enterprises that go global become transnational corporations and if they want to expand effectively, offshore property should be administered well, kept intact and appreciate. The administration of offshore property of transnational corporations has been a core issue in their offshore growth. The effective management of offshore assets is associated with a series of policies, such as functions of government departments, company's structure, parent company's control over its subsidiaries and exterior supervision on investment corpus etc. Furthermore, when any enterprise, no matter what its character is, becomes a transnational corporation and turns its offshore investments into offshore assets, these assets become national assets according to the *national principle* in national economic accounting, which is different from the attribute demarcation of assets in its domestic operations. Therefore, as to offshore assets, the country claims responsibilities for keeping them intact and avoiding the risks, such as escapes and loss etc. Reflected in assets administration, the responsibilities are supervision over offshore asset in different forms and to different degrees.

At present, large Chinese enterprises are in the mainstream to make offshore investments. In spite of different ownerships of these large enterprises, among them, the most competitive and the largest number are state-owned enterprises. Therefore, administration of offshore state-owned assets is under main consideration. It should be realized that no matter what kind of character of enterprises making offshore investment, their offshore assets are national assets and are one form of national wealth. As enterprises keep their offshore assets intact and promote a rise in value out of their own interests, requirements and motives, the country should regard them as overall national interests and claim the responsibilities for the guarantee of their safety and the supervision in some form.

3. Legal mechanism — An Authoritative Safeguard Mechanism

Laws and regulations are necessary to regulate all relationships between behavior corpus. They can restrict the relevant behaviors of behavior corpus and protect their interests. If enterprises in China hope to grow into transnational corporations to make FDI, their legal status, and the form and behavior patterns acknowledged by laws and regulations need to be affirmed. They need to get the most authoritative safeguard, which is the legal mechanism that promotes the growth of Chinese transnational corporations. In this respect, two problems exist:

(1) The legal system regarding offshore investments of Chinese enterprises has not been formed yet.

Currently, there is no *Offshore Investment Law of the P.R.C.* or *Direct Offshore Investment Law of the P.R.C.*, or other written legal files regarding offshore investments by Chinese enterprises. In this respect, a system of laws and regulations has yet to be developed. Some are outdated, and measures and scope of adjustment have not been suitable for development of offshore investments of Chinese enterprises. For example, though a series of present

legislative files about offshore investments are directed at offshore investment by state-owned enterprises, in practice of Chinese offshore investments, private enterprises have the necessity of direct offshore investments and are doing so now. Until now, China has not a single fundamental law regarding modulation of offshore investments and absence of corresponding legal system, which causes absence of legal evidence and macro guidance in practice. All the above bring about the blindness to certain degrees in China's offshore investments, unreasonable investment location and industrial choices. As a result China's national interests have been harmed in some offshore investments, which cause loss of national assets and sometimes subject corpora that observe disciplines and laws to missing out opportunities and legal advice.

(2) *The protection mechanism from international laws between double-lateral or multilateral investments has yet to be formed.*

Offshore investments involve two or more countries and are typical international affairs. The law of home countries cannot protect the benefits of their enterprises making offshore investments. Internationally, along with large sum of cross-national funds and more and more countries being involved, the legal system of international investments and relevant regulations will be established and be implemented accordingly. Normally, in order to attract and administrate foreign capital, the host country will formulate specific laws and regulations. Similarly, the home country will adopt the policy to encourage and protect its domestic enterprises in their offshore investments. However, the obvious differences in socioeconomic systems and technological levels between nations, the legislation of a sole country cannot play a full role in protecting the safety and interests of investors. When it comes to some major problems regarding interests conflicts, such as treatment of foreign capital, nationalization, settlement of investment disputes etc, domestic laws cannot solve them thoroughly. It is necessary to appeal to international laws and establish double-lateral or multilateral cooperative mechanism in order to promote the economic and technological cooperation based on mutual interests between nations, to maintain a favorable investment environment, and to protect foreign investors internationally.

Anyway, if China establishes a cohesive domestic legal systems regarding offshore investments, takes part in and takes advantage of international double-lateral or multilateral legal systems to protect investments, they will be an important mechanism to protect the interests of offshore investors of China and to promote the growth of Chinese transnational corporations.

4. Monetary Support Mechanism — A Supportive Mechanism at the Initial Stage

The prerequisite for an enterprise to become a transnational corporation to do FDI is to possess funds. On most occasions, it is hard for enterprises to provide investment funds for projects solely depending on their own funds. Even when they have made investments, it is hard for them to conduct the normal turnovers and operations. The most common solution is to acquire financial support or to conduct capital operations in capital markets. The degree of financial support, which is embodied by domestic financial supporting policies and given to enterprises' offshore investments, reflects whether the attitude of governmental policies is encouraging or restrictive. According to a study conducted by former Ministry of Foreign Economy and Trade, among the Chinese offshore processing and trade enterprises that hoped to get policy support were 77%.

In China, the policy of monetary support to offshore investments is mainly embodied in the file *The Guidance of Supporting the Loans to the Business of Processing Materials and Conducting Assembly with stuff to Overseas* issued by People's Bank of China and previous Ministry of Foreign Economy and Trade, and the file *The Administration of Loan Subsidy Interest for the Turnover of Foreign Exchange to Enterprises that Conduct Offshore Processing Trade* issued by the previous Ministry of Foreign Economy and Trade, the Ministry of Finance, previous State Economic and Trade Commission, People's Bank of China, State Administration of Foreign Exchange. The main objects are enterprises that conduct assembly and process materials with stuff to overseas. The main contents include: allowing domestic banks to provide financing to offshore investments that meet the requirements; favorable loan interests; precedent provision of export loan; direct financial support.

Apparently in China, the financial support mechanism about offshore investment is very primary and the support strength is not enough. It is far from forming financial support systems and lags back behind the development of situations, which is manifested as follows:

(1) Limited scope involved, appointed favorable objects and inappropriate guidance.

Two aspects are included. The first is that a favorable financial support policy is directed at the business of processing materials and conducting assembly with local inputs. The second is that the favorable financial policies have obvious limitations on ownership of enterprises.

The financial support policy should firstly abandon the discrimination of ownership and the regulations that financial support projects are restricted with *the business of processing materials and conducting assembly with stuff to overseas.*

(2) Monetary support has little strength and has a sole form.

The benefits in respect of interests and loans are of little substantial significance, that is, there is no strong support. In practice, the sum of loans provided by import-export banks is limited and has a sole form, which is a substantial difference from the monetary strength and forms of many other countries.

(3) Low operational efficiency and high costs, along with insufficient transparency of policies.

With strict standards of examination and approval, there are too many links, complicated and repetitive procedures; therefore, the present financial support is inefficient.

As monetary support is the basic requirement for some enterprises that hope to become transnational corporations and related to whether these enterprises would really be able to go outside, the government should make improvements in the following aspects, such as, expansion of scope in monetary support policies, increase of support strength, and increase of efficiency and transparency. This will lead to an important promotion mechanism for growth of Chinese transnational corporations.

5. Tax support mechanism: An effective operation mechanism to promote

If monetary support mechanism is related to whether enterprises will be able to take the first step during offshore investments, taxation support mechanism will be the promotion mechanism for effective operations of enterprises that have already made offshore investments. Its role is to reduce the burden, financially guarantee and support enterprises to operate normally, expand their scale and increase their competitive advantages.

Giving monetary support and tax preferences is the customary practice of many countries and areas to encourage their enterprises to develop offshore investments and to support the growth of their transnational corporations.

In China, such policies only give financial support and preferential tax measures to the business of processing materials and conducting assembly with overseas inputs, allowing such enterprises to supplement their capital funds with the succeeding five years' profits after the year of making profits with the purpose of expanding the scale of their production. The profits within the succeeding five years after the first year of making profits are free of submission. Enterprises' export equipment, raw materials and scattered parts as physical investments enjoy export drawback. Among them, the duty drawback of second-hand equipment should be calculated according to the balance after depreciation while the duty drawback of new equipment and raw materials should be calculated according to the sum of income taxes listed in the specific invoices for added value tax.

China's tax support mechanism in respect to offshore investments and promotion of development of OFDI has such problems as narrow scope, not enough support strength and lagging behind economic development.

6. Insurance support mechanism — An operational risk-prevention mechanism

Compared with domestic investments, offshore investments tend to face more risks, such as business risks and political risks. Occurrence of these risks will bring about a loss of different degrees to offshore investors' operations. Some of them involve life or death situations and seriously affect operations of investors. The effective way to manage and prevent risks counts on insurance. But many risks related to offshore investments are not covered by insurance companies, which requires a country to establish corresponding offshore investment insurance

mechanism to protect the interests of offshore investors and to promote the development of OFDI.

Along with the growth of China's economy, it is an inevitable trend for enterprises to strengthen their competitive advantages, grow into transnational corporations and make more foreign investments.

7. Service support mechanism —A Promotion mechanism for efficient operations

The service support mechanism for growth of China's transnational corporations is a series elements and systems, which increase operational efficiency, survival capability and competitive advantages, and improve their development conditions. The service support of enterprises' offshore investment has wide content and form, and is a basic requirement for the development of transnational corporations, and it is the promotion mechanism for the fundamental environment. Service support mechanism includes: (1) Information assistance; (2) Technological support of investments and personnel training; (3) Cooperation between governments.; (4) Intermediary organizations.

8. Conclusion

Along with reforms and fast economic development in China, more and more enterprises, in the process of economic globalization, make foreign direct investment, conduct cross-national operations and grow into transnational corporations. This is an inevitable trend. The growth of Chinese transnational corporations is of important and strategic significance to China. This is a necessary step for China to captures international market more extensively, to utilize and allocate global resources, and to upgrade the domestic industry structure. In the process, China will be able to acquire tremendous economic benefits and improve its economic status and influences on the international stage. If Chinese enterprises hope to become transnational corporations, they need to improve their competitive advantages and exert their ownership specific advantages. Furthermore, the exterior environments to nurture their growth also play an important role, which are relevant mechanisms that influence the growth of Chinese transnational enterprises. However, the main aspects of these mechanisms are apparently adverse to the growth of Chinese transnational enterprises and do not exert their roles of facilitation. For example, foreign exchange administration implements a strict administration mode, which means restricting the use of foreign exchange by enterprises, restricting enterprises from spending their own funds purchasing foreign exchange to expand the shares, and carrying out the system of deposits remitted from profits of FDI. As far as legal mechanism is concerned, China should quicken the fundamental legislation on transnational corporations, specify the status and necessity of transnational corporations, standardize their behaviors and safeguard their interests, and thus promote their healthy and fast development. At present, under China's current administrative examination and approval, there exist such problems as excessive links, repetition, inefficiency, insufficient transparency, which lead to increased costs and lost business opportunities. In terms of the administration of offshore assets, assets financed by Chinese FDI, including state-owned and private investment assets, should be regarded as national assets and administered scientifically. Meanwhile, relevant government administration departments should conduct necessary supervisions in order to keep them intact and promote a rise in value, and avoid such risks as escapes and losses. In the respect of monetary support, besides its little strength and sole form, the objects restricted are projects of processing materials and conducting assembly with stuff to overseas and state-owned enterprises, which apparently ignores the investment project of market-seeking, resources-seeking and strategic assets-seeking, as well as support to the offshore investments of private enterprises. In terms of tax support, China should sign more relevant tax agreements to protect the interests of Chinese offshore investors, reduce their burden, promote effective operations and expand the scale. As far as insurance mechanism is concerned, management of offshore enterprises' non-business risk, which business insurances organizations are unwilling to insure, should be established and strengthened in China. On the other hand, international cooperation on insurance should be strengthened and relevant insurance agreements should be signed quickly in order to safeguard the safety of operations of Chinese offshore enterprises. As to information support, Chinese enterprises have not enjoyed premium and complete services, which influence Chinese enterprises to go global smoothly. In this aspect, the following are to blame: insufficient information support and technological assistance, cooperation between governments that cannot better meet the requirements of enterprises to operate in host countries; insufficient intermediary services and low quality. In general, the objective mechanisms to promote the growth of Chinese

transnational corporations still require an improvement and development and great effort should be made, which is of great significance to the growth of Chinese transnational corporations.

This article is subsidized by the fund of Guangdong Science and Technology Department. No.8451063101001599

References

Benwu Xiang. China's Strategies of "Going Global": Achievements, Challenges and Adjustments [J]. *Macroeconomics* 2009(3).

Gordon H. Hanson, Should Countries Promote Foreign Direct Investment? UNITED NATIONS *G-24 Discussion Paper Series* No. 9, February 2001

Jiancheng Guan, Xiaojing Wang. A Research on the Determinants of China's Outward Direct Investment [J]. *China Soft Science,* 2007(2):59-65.

Jing Xu, Qirong Zhu. Problems with Policy Systems of China's Outward Foreign Direct Investment and Solutions [J]. *International Economic Cooperation,* 2008(5).

John A. Mathews, Dragon multinationals: New players in 21st century globalization, *Asia Pacific J Manage* (2006) 23: 5–27

Lei LI, Zhaoyang Zheng,. Is China's Outward FDI Seeking Resources?[J]. *Journal of International Trade,* 2012(2):146-157.

Lee Branstetter Nicholas Lardy, China's Embrace of Globalization, *NBER Working Paper 12373* July 2006

Peter Ping Li, Toward an integrated theory of multinational evolution: The evidence of Chinese multinational enterprises as latecomers, *Journal of International Management* 13 (2007) 296–318

Ping Deng, Investing for strategic resources and its rationale: The case of outward FDI from Chinese Companies, *Business Horizons,* (2007) 50, 71–81

Shiyue Cheng, Min Xu. Comment on the Domestic Institutions for Outward Foreign Direct Investment in China [J]. *Inquiry into Economic Issues,* 2011(10):149-154

Statistical Bulletin of China's Outward Foreign Direct Investment (2012), China Statistics Press 2013

Yansheng Zhang. Speed up the Enforcement of the "Go Global" Strategy[J]. *Internade,* 2011 (8).

Xiaohui Liu, Trevor Buck, Chang Shu, Chinese economic development, the next stage: outward FDI? *International Business Review* 14 (2005) 97–115

Yan Chen, Huan Yang, Bin Zhang, China's Outward FDI Drivers, Institution Moderation and Regional Differences[J]. *Journal of Management Science,* 2012,(6):112-120.

Zhizhong Yao, Zhongmin Li. China's Outbound FDI: Trends and Policy Outlook [J][J]. *International Economic Review,* 2011, 2: 014.

Zhongmin Li. Challenges and Reforming Suggestions for China's Management Systems of Outward Foreign Direct Investment [J]. *Internade,* 2010(10).

"It's easy to lie with statistics. It's hard to tell the truth without statistics."

– Andrejs Dunkels

Feature

Productivity Change in Indian Information Technology-enabled Services (ITeS) Sector: A Malmquist Index Approach

PROSENJIT DAS

This study aims to evaluate the productivity change and its determinants in the Indian ITeS sector during 2004-05 to 2014-15. For this purpose, firm-level data is collected from the Centre for Monitoring Indian Economy (CMIE) Prowess database. The findings of productivity analysis reveal that, on an average, the ITeS industry has experienced improvement in total factor productivity by 3.1% during the entire study period. The decomposition analysis of productivity reveals that innovation is the major contributor to the growth of total factor productivity. The regression results show that export and human capital have positive impact on productivity whereas the US subprime crisis has negative impact on productivity. The policy implications of this study suggest that the firm-level managerial efficiency should be improved for further enhancement of productivity. In other words, policy should be devised in order to encourage the private limited firms for a holistic development of this industry. Moreover, the stakeholders of Indian ITeS industry should reduce dependence on the US and UK for export revenue and needs to explore new markets to keep pace with the ever changing technical and business environment.

Prosenjit Das, Economic Officer, Directorate of Economics & Statistics, Department of Agriculture, Cooperation & Farmers Welfare, Ministry of Agriculture & Farmers Welfare, Govt. of India, New Delhi.

DOI : - https://doi.org/10.32381/PROD.2019.59.04.8

1. Background

Generally, theories of economic development indicate that at the beginning of the development process of a nation, the relative contribution of agriculture sector is high as compared to the other two sectors, namely, manufacturing and services sectors. For instance, as a nation begins to develop, the contribution of agriculture to GDP start to decline and that of the manufacturing sector starts to increase. At the advanced stage of development, the contribution of manufacturing sector begins to fall and that of the services sector rises. Although this path of development process has been noticed for most of the now advanced economies, in India, however, the process of economic development has followed somewhat different trajectory. In case of India, it has been observed that, with a decline in the agriculture sector's share in GDP, the share of services sector in GDP started to increased since the mid-1990s instead of the manufacturing sector. This untimely emergence of services sector in the Indian economy is popularly named as 'services revolution' by the World Bank. In this regard, the IT-ITeS sector has come out as one of the leading sectors in the Indian services sector.

The Indian IT-ITeS sector has been playing a pivotal role in the economic development of India since post liberalization period. On the other hand, India has successfully positioned itself in the global export market of IT-ITeS since the beginning of the 21st century. The rapid growth of the IT-ITeS sector has not only helped India to occupy its remarkable position in the international offshoring market but also to achieve the high growth path. India's competitive advantage in global outsourcing market primarily comes from the abundance supply of low cost IT-skilled and English proficient workforce. In addition, public investment in Indian technical education in a

sustainable manner, difference in time-zone, etc., have also instrumental role in establishing India's brand image in the global IT-ITeS market.

However, in recent time, some global issues have been posing serious threat to the sustainability of Indian ITeS industry. Among those issues, the new US administration's affinity towards protectionist policies, emerging competition from various low-cost destination nations, automation, US sub-prime crisis, etc., are noteworthy. Particularly, uncertainties over the US government's decision towards H-1B visa regulation along with Britain's exit from the European Union have caused havoc in the Indian ITeS industry.

In view of the above discussions, it is pertinent to investigate the performance of the Indian ITeS industry. We have found few studies, which endeavoured to evaluate the performance of either the IT industry or IT and ITeS industries taken together. For instance, Shao and Shu (2004) measured the performance of IT industry of 14 OECD nations by estimating total factor productivity growth (TFPG) during 1978-1990. Shu and Lee (2003) employed stochastic frontier analysis to evaluate productivity as well as productive efficiency of IT industry across OECD countries for the year 1998. Mathur (2007) applied data envelopment analysis (DEA) technique to measure the technical efficiency of 92 Indian software farms for the year 2005-06. Chen et al. (2011) estimated the managerial, scale and overall efficiency scores in 73 Chinese IT companies during 2005-07 by applying DEA method. Bhattacharjee (2012) studied the technical efficiency of Kolkata-based Software Technology Park's IT and ITeS firms during 1993-94 to 2007-08. Sahoo (2013) analyzed the TFPG of Indian software industry during the period 1998 to 2008 by applying DEA-based Malmquist productivity index. Sahoo and Nauriyal (2014) examined the technical efficiency of 72 Indian software firms on the basis of the input and output data collected from CMIE Prowess database for the period 1999-2008. Further, this study also evaluated the determinants of overall, pure and scale efficiency scores by using Tobit regression analysis. Chou and Shao (2014) investigated the productivity growth of IT industries in 25 OECD nations for the period from 1995 to 2007 by applying Malmquist index. Das and Datta (2017) applied DEA technique to assess the technical efficiency of Indian ITeS industry during 2000-01 to 2014-15. This study also assessed the determinants of technical efficiency by using Tobit regression analysis. Das (2017) assessed the TFPG of 70 Indian IT companies for the period 2004-05 to 2014-15 using DEA-based Malmquist index.

However, we have not found any study that exclusively addresses the issue of the performance of Indian ITeS industry by evaluating the TFPG. In order to fill this research gap, the present study aims at analyzing the aspect of TFPG in the Indian ITeS industry.

1.1. Objectives of the Study

The objectives of the study are as follows:

- To evaluate the TFPG in the Indian ITeS industry during 2004-05 to 2014-15.
- To decompose the TFPG into frontier-shift, pure technical efficiency change and scale efficiency change to find out their relative contributions to productivity change.
- To investigate various determinants of TFPG, innovation and managerial technical change.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of India's ITeS sector. The research methodology is explained in Section 3. Section 4 describes the data. Section 5 discusses the results of the study. Finally, Section 6 concludes the study with policy implications.

2. The Indian ITeS Sector: A Brief Overview

The Indian ITeS sector consists of Business Process Outsourcing (BPO), Knowledge Process Outsourcing (KPO), Medical Transcription (MT), Call Centres, Legal Process Outsourcing (LPO) and other IT-enabled services such as document processing, data entry, among others. A typical ITeS company primarily engages with a variety of back office services by using software and computer applications. In other words, an ITeS company is mainly dedicated in outsourcing of low-end IT-enabled services to third-party service provider and not directly related to the development of high-end products such as software applications. Moreover, the ITeS sector is relatively labour intensive than its IT counterparts.

A call center is a centralized facility where telephone calls from customers are operated by an organization with the help of some automated computer systems and telecommunication system. Normally, a call center is capable to attempt a large number of phone calls, screening those calls and lastly transmit the same to the

appropriate individual for further pursuance. A usual call center can handle either inbound or outbound calls or both. A typical BPO firm carries out back office outsourced activities for other organizations. BPO is a generic term of all ITeS activities like MT, KPO, LPO, etc. The activities like call center, MT, KPO, LPO are variants of BPO. In fact, various organizations of advanced western countries outsource their non-core back office works like Payroll, accounting, data transcription, etc., to the low-cost destinations like India. A medical transcription company's job is to prepare a digitally typed document of the medical history/reports of patients. In future, this digital document would help medical practitioners to treat other patient with similar kind of medical record. The major activities of a KPO firm consist of a range of research and information gathering activities such as, domain-based processes, consultancy, business and market research, R&D in pharmaceuticals, equity research, evaluation of intellectual property for patent applications, etc. A LPO firm is usually associated with the outsourcing works of legal support services. For instance, legal research, patent services, legal document review, drafting legal briefs, etc., are some of the key activities of a LPO firm.

3. Research Methodology

In this paper, we apply the efficient frontier approach to measure the total factor productivity change in the Indian ITeS sector. For this purpose, we employ a two stage methodology. In the first stage, a non-parametric DEA-based normative index, namely, the Malmquist productivity index is employed to assess the TFPG. In the second stage, regression method is considered to find out the determinants of TFPG and its constituent components, viz. catch-up and frontier-shift. We discuss these aforementioned methods in the following sub-section.

3.1. Measurement of Productivity Change

The concept of Malmquist index was first introduced by Malmquist (1953). Subsequently, this index has further been developed in non-parametric DEA framework by several researchers. Caves et al. (1982) first develop the distance functions based Malmquist index (MI) in multi-input, multi-output framework. Färe, Grosskopf, Lindgren, and Roos (FGLR) (1992) first made an attempt to decompose the MI into two components, viz. catch-up and frontier-shift under the assumption of constant returns to scale (CRS) technology. However, the assumption of global presence of CRS technology seems to be restrictive and may not be reasonable in the real world. Therefore, the decomposition of MI by FGLR may not be suitable. Subsequently, Färe, Grosskopf, Norris, and Zhang (FGNZ) (1994) extended the FGLR decomposition under variable returns to scale (VRS) technology. In this paper, we evaluate the Malmquist Index on the basis of the output distance function. The output distance function can be defined as

$$D^T(x^t, y^t) = \inf [\, \theta : (x^t, \tfrac{1}{\theta} y^t) \varepsilon\, S^t \,] \qquad (1)$$

where 'x' and 'y' refer to the vectors of inputs and outputs, S denotes the production possibility set, and the superscript 'T' indicates the reference period of technology. Generally, 'T' assumes two time periods (here, years), viz. t and t+1. 'θ' refers to the output oriented technical efficiency score. As proposed by Caves et al. (1982), the change in total factor productivity (TFP) between periods t and t+1 can be measured by Malmquist index in the following way:

$$MI = D^T(x^{t+1}, y^{t+1}) / D^T(x^t, y^t) \qquad (2)$$

The above formula is essentially the ratio of two output distance functions pertaining to the periods t and t+1, respectively. FGLR construct the Malmquist Index between year t and t+1 in the following manner:

$$MI = \left[\frac{D^t(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \times \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^{t+1}(x^t, y^t)} \right]^{1/2} \qquad (3)$$

The formula of Malmquist index presented above is the geometric mean of two indices for the years t and t+1. FGLR further decompose the above expression and represent the Malmquist index as a product of two factors, namely, catch-up and frontier-shift.

$$MI = \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \times \left[\frac{D^t(x^{t+1}, y^{t+1})}{D^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D^t(x^t, y^t)}{D^{t+1}(x^t, y^t)} \right]^{1/2} \qquad (4)$$

where catch-up $= \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)}$, and frontier-shift

$$= \left[\frac{D^t(x^{t+1}, y^{t+1})}{D^{t+1}(x^{t+1}, y^{t+1})} \times \frac{D^t(x^t, y^t)}{D^{t+1}(x^t, y^t)} \right]^{1/2}$$

While the catch-up component provides the measurement of change in technical efficiency between years t and t+1, the frontier-shift component measures the change in technology between years t and t+1. FGNZ further breakdowns the catch-up into two components, viz. Pure Technical Efficiency Change (PTEC) and Scale Efficiency

Change (SEC) under VRS technology. The following Eq. (5) depicts the decomposition of catch-up proposed by FGNZ:

Catch-up (C) =

$$\frac{D_V^t(x^{t+1},y^{t+1})}{D_V^t(x^t,y^t)} \times \frac{(D_C^{t+1}(x^{t+1},y^{t+1})/D_C^t(x^t,y^t))}{(D_V^{t+1}(x^{t+1},y^{t+1})/D_V^t(x^t,y^t))} \quad (5)$$

where PTEC = $\frac{D_V^t(x^{t+1},y^{t+1})}{D_V^t(x^t,y^t)}$

and

SEC = $\frac{(D_C^{t+1}(x^{t+1},y^{t+1})/D_C^t(x^t,y^t))}{(D_V^{t+1}(x^{t+1},y^{t+1})/D_V^t(x^t,y^t))}$

It is noticed that while the catch-up is evaluated under CRS technology, its breakdown into PTEC and SEC is occurred under VRS technology. The subscripts 'c' and 'v' in distance functions in Eq. (5) indicate the technical efficiency under CRS and VRS technologies, respectively. PTEC captures the change in managerial efficiency over time.

Thus the Malmquist Index can be presented as:

$$MI = \frac{D_V^t(x^{t+1},y^{t+1})}{D_V^t(x^t,y^t)} \times \frac{(D_C^{t+1}(x^{t+1},y^{t+1})/D_C^t(x^t,y^t))}{(D_V^{t+1}(x^{t+1},y^{t+1})/D_V^t(x^t,y^t))} \times \left[\frac{D^t(x^{t+1},y^{t+1})}{D^{t+1}(x^{t+1},y^{t+1})} \times \frac{D^t(x^t,y^t)}{D^{t+1}(x^t,y^t)}\right]^{1/2} \quad (6)$$

Hence, MI = PTECSECTC

It is to mention that, a value of MI greater than unity implies improvement in TFP from period t to t+1 while a value equal to one indicates no change in productivity. On the other hand, a value less than unity means decline in TFP from period t to t+1. The same explanations are applicable for its components.

3.2. Regression Method

To evaluate the determinants of catch-up, frontier-shift and TFP; regression method is considered. Since we have a panel dataset, it is essential to find out a suitable regression model for analyzing our dataset. In this regard, we have applied three specification tests, namely, poolability test, Breusch-Pagan LM test and Housman test. While the poolability test helps to find out the suitable model between OLS and fixed-effects models, the Breusch-Pagan LM test indicates the appropriate model between OLS and random-effects models. On the other hand, Housman test helps to choose the correct model between random-effects and fixed-effects models.

4. Data Description and Source

We have chosen one output and three input variables for the computation of Malmquist productivity index. Sales revenue is considered as output and salaries and wages, net fixed assets, and operating expenses are selected as input variables.

To find out the determining factors of frontier-shift, catch-up and TFP; we have considered the independent variables on the basis of Caves (1992). According to Caves (1992), the determinants of industrial productivity can be classified into five categories, namely, structural heterogeneity, organizational features, competitive conditions, regulatory environment and dynamic disturbances. On the basis of this classification, the measurement of the regression variables is presented in Table 1.

5. Empirical Results

5.1 Results Pertaining to the Productivity Change

The trends in TFP and its constituent components pertaining to Indian ITeS industry are discussed in this section. The TFP is measured with respect to adjacent year frontier. In this study, we have considered a balanced panel of 22 firms over a span of 11 years. Table 2 reports the year-wise average TFP and its components, namely, frontier-shift, catch-up, pure technical efficiency change (PTEC), and scale efficiency change (SEC).

From Table 2, it is observed that the average frontier-shift is greater than one during the study years except the years 2008, 2009, 2010 and 2013. The average catch-up scores are found to be greater than one for the years 2006, 2007, 2010, 2011, and 2012 and less than one for the years 2005, 2008, 2009, 2013 and 2014. The PTEC showed improvement during 2005, 2010, 2011 and 2012. Furthermore, the SEC showed improvement during most of the study years except 2005, 2009 and 2013. Finally, the average MI scores demonstrate improvement in TFP in most of the study periods except few study years (2008, 2009 and 2013).

Figure 1 depicts the average TFP and its two components (frontier-shift and catch-up) during the study period. The frontier-shift is found to be highest (1.208)

TABLE 1: Variable Measurement for Regression Analysis

S. No.	Variable	Measurement
1	Export intensity	Total exports/sales revenue
2	Market concentration	Hirshman-Herfindahl index
3	Age	Natural logarithm of years in business
4	Size	Natural logarithm of sales revenue
5	Human capital	Salaries and wages/operating expenses
6	Plant scale dummy	Returns to scale (RTS) dummies.
		a) IRS dummy = 1, if the firm exhibits IRS = 0, otherwise b) DRS dummy = 1, if the firm exhibits DRS = 0, otherwise
7	Research and Development (R&D) dummy	R&D dummy = 1, if the firm spends on R&D = 0, if the firm does not spend on R&D
8	Royalty dummy	Royalty dummy = 1, if the firm pays for royalty = 0, if the firm does not pay for royalty
9	Ownership dummy[1]	= 1, for public limited firm = 0, for private limited firm
10	Group dummy[2]	= 1, if the firm belongs to a group of companies = 0, otherwise
11	US subprime crisis dummy	= 1, for the years 2008 to 2014 = 0, otherwise

Source: Author's construction.

All the data required for this study is collected from CMIE Prowess database[3]. All the nominal variables are transformed into real variables by using GDP deflator (base=2004-05).

TABLE 2: TFP and its Components

Year	Frontier-shift	Catch-up	PTEC	SEC	MI (TFP)
2004-05	1.208	0.963	1.024	0.940	1.163
2005-06	1.022	1.006	0.993	1.013	1.028
2006-07	1.048	1.006	0.979	1.028	1.055
2007-08	0.977	0.995	0.931	1.069	0.973
2008-09	0.982	0.965	0.980	0.984	0.947
2009-10	0.976	1.067	1.028	1.038	1.042
2010-11	1.001	1.020	1.015	1.006	1.022
2011-12	1.025	1.054	1.044	1.009	1.080
2012-13	0.998	0.945	0.997	0.947	0.943
2013-14	1.082	0.996	0.982	1.015	1.078
Average	**1.030**	**1.001**	**0.997**	**1.004**	**1.031**

Source: Author's calculations based on CMIE Prowess database.

during 2005 with a growth rate of 20.8% and lowest (0.976) during 2010 with a negative growth rate of -2.4%. The catch-up is found to be maximum (1.067) during 2010 with a growth rate of 6.7% and minimum (0.945) during 2013 with a negative growth rate of -12.4%. The MI is found to be highest (1.163) during 2005 with a growth rate of 16.3% and lowest (0.943) during 2013 with a negative growth rate of -5.7%. Figure 2 graphically presents the catch-up and its two components (PTEC and SEC). The PTEC is largest (1.044) during 2012 with a growth rate of 4.4% and smallest (0.931) during 2008 with a negative growth rate of -6.9%.

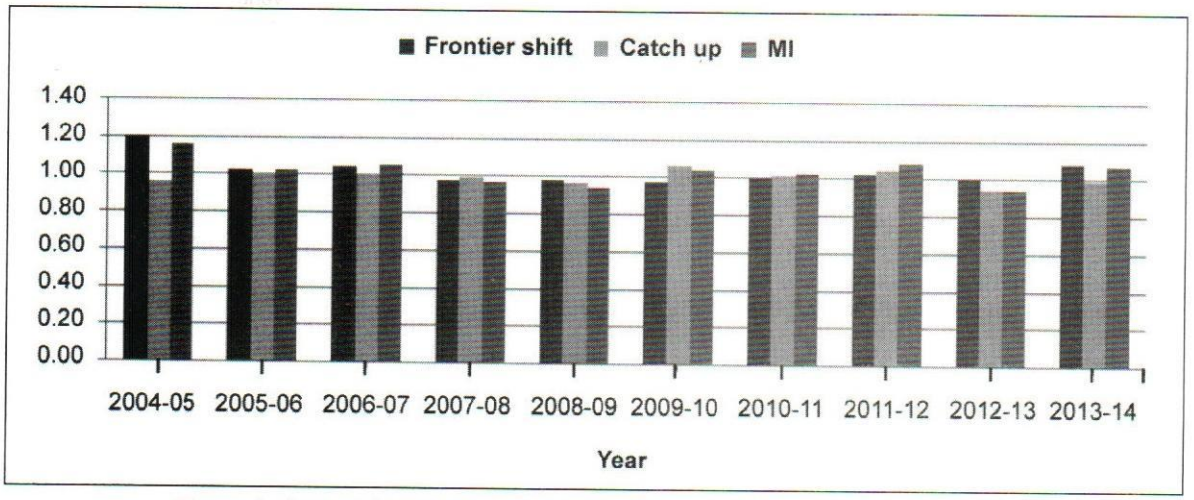


Figure 1. Annual Average Frontier-shift, Catch-up and Malmquist index (MI)

Source: Table 2.

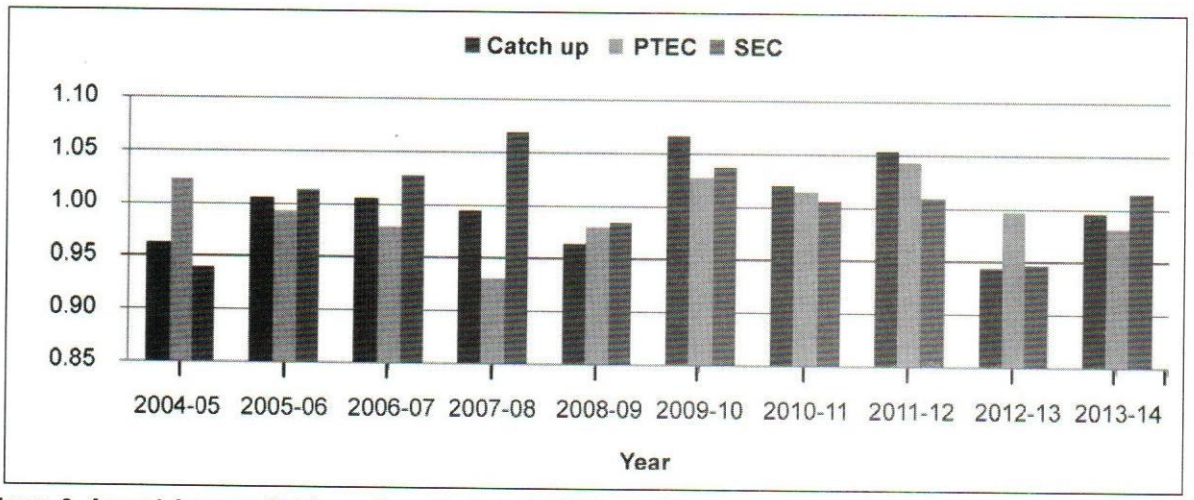


Figure 2. Annual Average Catch-up, Pure Technical Efficiency Change (PTEC) and Scale Efficiency Change (SEC)

Source: Table 2.

The SEC is found to be largest (1.069) during 2008 with a growth rate of 6.9% and smallest (0.940) during 2005 with a negative growth rate of -6%. During the entire study period, the average MI, frontier-shift, catch-up, PTEC, and SEC are found to be 1.031, 1.030, 1.001, 0.997 and 1.004, respectively. Moreover, it can be said that, on an average, MI, frontier-shift, catch-up and SEC have registered improvement with growth rates of 3.1%, 3%, 0.1% and 0.4%, respectively; whereas PTEC has registered negative growth of -0.3% during the entire study period.

Our analysis indicates overall improvement in TFP (MI), frontier-shift and SEC during the study period. On the other hand, PTEC is found to be declining during the

study period. Consequently, it can also be said that the frontier-shift effect has played a significant role not only in determining TFPG but also its improvement over time. However, the PTEC has played a dampening role in catch-up and therefore, in TFPG also. Notwithstanding the negative effect of PTEC on catch-up and TFPG over time, on an average, TFPG had been found to be improving. This happened due to the fact that, the positive impact of frontier-shift was relatively stronger than the negative impact of PTEC on TFPG during the study period.

Table 3 reports company-wise average frontier-shift, catch-up, PTEC, SEC, and MI of 22 ITeS firms during the overall study period.

TABLE 3: Company-wise TFP and its Components

S. No.	Company Name	Frontier shift	Catch up	PTEC	SEC	TFP
1	Aditya Birla Minacs Worldwide Ltd.	1.086	1.037	1.033	1.004	1.126
2	Allsec Technologies Ltd.	1.006	0.958	0.968	0.990	0.963
3	Axis-I T & T Ltd.	1.049	1.023	1.020	1.004	1.074
4	Caliber Point Business Solutions Ltd.	0.990	0.967	0.980	0.986	0.957
5	Coral Hub Ltd.	0.948	1.010	1	1.010	0.957
6	Cosmic Global Ltd.	1.050	0.974	0.965	1.010	1.023
7	Datamatics Financial Services Ltd.	1.053	1.017	1.015	1.002	1.071
8	Firstsource Solutions Ltd.	1.036	1.005	1.010	0.995	1.041
9	H C L Comnet Systems & Services Ltd.	1.150	1.009	1	1.009	1.161
10	I C R A Online Ltd.	1.087	1.010	1.007	1.003	1.098
11	In House Productions Ltd.	1.114	1.023	1	1.023	1.140
12	Informed Technologies India Ltd.	1.004	1.060	0.973	1.089	1.064
13	Microwave Communications Ltd.	1.097	1.067	1.036	1.030	1.170
14	Mold-Tek Technologies Ltd.	1.003	1.016	0.984	1.032	1.019
15	N I I T Smartserve Ltd.	1.118	1.058	1.048	1.010	1.182
16	Omnitech Infosolutions Ltd.	0.974	1.002	1	1.002	0.976
17	Reliance Corporate I T Park Ltd.	1.033	0.959	0.949	1.011	0.990
18	Rev I T Systems Pvt. Ltd.	1.097	1.015	1.032	0.983	1.113
19	Shreejal Info Hubs Ltd.	1.097	1.060	1.045	1.014	1.162
20	T C S E-Serve Ltd.	1.092	0.992	1	0.992	1.083
21	Tricom India Ltd.	1.011	0.982	0.982	1	0.992
22	Virtualsoft Systems Ltd.	0.973	0.988	0.967	1.022	0.962
	Mean	**1.047**	**1.010**	**1.000**	**1.010**	**1.058**
	Median	**1.049**	**1.010**	**1.000**	**1.007**	**1.068**
	Std. Dev.	**0.055**	**0.032**	**0.028**	**0.022**	**0.077**

Source: Author's calculations based on CMIE Prowess database.

It is revealed from this Table that, on an average, 4 firms (Virtualsoft Systems Ltd., Coral Hub Ltd., Omnitech Infosolutions Ltd., and Caliber Point Business Solutions Ltd.) registered technical regress and remaining 18 firms experienced technical progress during the entire study period. Among 22 ITeS firms, on an average, 7 firms (T C S E-Serve Ltd., Virtualsoft Systems Ltd., Tricom India Ltd., Cosmic Global Ltd., Caliber Point Business Solutions Ltd., Reliance Corporate I T Park Ltd., and Allsec Technologies Ltd.) exhibited negative growth and the remaining 15 firms experienced progress in catch-up during the overall study period. As far as SEC is concerned, it is observed that, on an average, 5 firms (Rev I T Systems Pvt. Ltd., Caliber Point Business Solutions Ltd., Allsec Technologies Ltd., T C S E-Serve Ltd., and Firstsource Solutions Ltd.) registered negative growth, one firm (Tricom India Ltd.) experienced no change, and the remaining 16 firms recorded improvement during the study period. Finally, it is revealed from Table 3 that, on an average, 7 firms (Tricom India Ltd., Virtualsoft Systems Ltd., Reliance Corporate I T Park Ltd., Coral Hub Ltd., Caliber Point Business Solutions Ltd., Allsec Technologies Ltd., and Omnitech Infosolutions Ltd.) experienced regress in TFP and remaining 15 firms exhibited progress in TFP during the study period.

5.2. Results Pertaining to the Determinants of Productivity Change

In this section, we intend to find out the factors influencing catch-up, frontier-shift and TFP (or MI) of the ITeS sector during the study period. For this purpose, we will be using regression methods. Prior to the regression analysis, we apply some specification tests to find out the suitable regression method. The specification test results are reported in appendix.

The results of the poolability test (see appendix, Table A1) result shows that the F-test static is statistically significant at 1% level across three regression models, implying that the dataset is not poolable and there exists significant firm-specific heterogeneity. On the other hand, the poolability test results indicate that fixed-effects panel may be suitable for all three regression models.

The test statistic of Breusch-Pagan LM test (see appendix, Table A2) is observed to be statistically significant at 1% level in all the regression models, suggesting thereby the suitability of random-effects model over pooled OLS model. Now, we apply the Housman test to find out the most appropriate model (between fixed effects and random effects panel models) for our ITeS dataset.

Appendix Table A3 reports the test statistics and the corresponding P-values of Housman test pertaining to the three regression models. All the three Housman test statistics are found to be statistically insignificant. It implies that the random-effects model is suitable for all the three regression analyses.

Based on the specification tests results discussed above, we are going to apply the random effects generalized least square (RE-GLS) model for the three regression models having dependent variables, namely, catch-up, frontier-shift, and Malmquist index, respectively. The summary of three RE-GLS regression results is reported in the following Table 4.

The Wald Chi-square statistics are found to be statistically significant at 1% level across three regression models, meaning that the models are overall significant. On the other hand, the non-values of ρ indicate that the panel estimator is significantly different from the pooled OLS estimator. Furthermore, it can be inferred that firm-specific differences are significant in our dataset.

The coefficient of export intensity is found to be significant at 1% level across the three models. The coefficient of market concentration is observed to be statistically significant in all three models. The coefficient of age is found to be statistically insignificant in R1. On the other hand, in R2 and R3, the coefficients of age are found to be statistically significant at 10% level. It is observed that size has no statistically significant impact on catch-up, frontier-shift, and TGP over the study period. Hence, the relationship between size and the three dependent variables cannot be ascertained in the present study. The coefficient of human capital is found to be significant in R1 and R3 and insignificant in R2. In R1, the coefficient of human capital is observed to be 0.563, suggesting that, on an average, there would be 56.3% improvement in technical efficiency as a result of 100% increase in wages and salaries intensity during the study period. On the other hand, the coefficient of salaries and wages intensity in R3 is found to be 0.623, indicating thereby on an average, 62.3% increase in TFP would occur due to 100% rise in salaries and wages intensity during the study period.

Among two RTS dummies, viz. IRS and DRS dummies, IRS dummy is found to be positive and statistically significant in R1 and R3. It implies that, on

TABLE 4: Summary Results of Random Effects Generalized Least Square (RE-GLS) Regression Models

Independent Variables	Regression 1 Dependent Variable: Catch-up		Regression 2 Dependent Variable: Frontier-shift		Regression 3 Dependent Variable: Malmquist Index	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Export intensity	0.421***	0.001	0.639***	0.001	0.485***	0.000
Market concentration	0.012**	0.018	0.020*	0.092	0.038**	0.024
Age	0.005	0.589	0.014*	0.086	0.025*	0.054
Size	0.002	0.162	0.005	0.215	0.001	0.698
Human capital	0.563***	0.009	0.150	0.117	0.623***	0.004
IRS dummy	0.006**	0.025	0.001	0.298	0.010*	0.078
DRS dummy	-0.001	0.487	0.005	0.526	0.002	0.591
R&D dummy	0.008	0.165	0.004	0.284	0.006	0.314
royalty dummy	0.006	0.250	0.001	0.894	0.002	0.365
Ownership dummy	0.009**	0.013	0.014**	0.019	0.017***	0.004
Group dummy	-0.002	0.215	0.006	0.294	0.012	0.651
US subprime crisis dummy	-0.052	0.141	-0.077**	0.036	-0.081**	0.027
Constant	0.823***	0.000	0.712***	0.000	1.291***	0.000
Wald ^{2}-statistic	153.54***	0.000	185.32***	0.000	169.74***	0.000
R Squared	Within=0.320 Between=0.655 Overall=0.251		Within=0.289 Between=0.647 Overall=0.205		Within=0.342 Between=0.695 Overall=0.268	
Number of observations	220		220		220	
sigma υ	0.2876		0.2784		0.2415	
sigma ε	0.2641		0.2136		0.2376	
sigma ρ	0.5425		0.6294		0.5081	

Source: Author's calculations based on CMIE Prowess database.

*, **, *** => Significant at 10%, 5%, and 1% level.

an average, the IRS firms experienced higher improvement in catch-up and TFP with respect to the benchmark CRS firms during the study period. On the other hand, the coefficients of DRS dummy are found to be negative in R1 and positive in R2 and R3 but statistically insignificant across all the three models. Hence, the relationship between the dependent variables and the DRS dummy cannot be established. The coefficients of the R&D dummy and royalty dummy are found to be positive but statistically insignificant across three models.

The ownership dummy is found to be positive and statistically significant across all regression models. Therefore, it can be said that, on an average, the public limited firms are performed better than the private limited firms (i.e., the reference firms) the study period. The group dummy is found to be statistically insignificant in all the three models. Finally, the coefficient of US subprime crisis dummy is observed to be negative across all models but

statistically significant in R2 and R3. Consequently, it can be inferred that due to the US subprime crisis in 2008, on an average, the catch-up and TFP of Indian ITeS companies deteriorated during the post-crisis period (2008-2014) as compared to the pre-crisis period (2005-2007).

6. Conclusion and Policy Implications

The study made a modest attempt to evaluate the productivity change in Indian ITeS sector during 2004-05 to 2014-15. For this purpose, a DEA-based output oriented Malmquist index was applied to measure the TFPG. The average TFPG was found to be 3.1 percent for the entire study period. The decomposition analysis of TFPG showed that technical progress played pivotal role in productivity growth during the study period. The roles of catch-up and scale efficiency change are observed to be positive on TFPG. However, the impact of managerial efficiency change is found to be negative on TFPG. The regression results revealed that the firms with higher export intensity and human capital have experienced higher TFPG. The impact of 2008's US subprime crisis is found to be negative on TFPG. On the other hand, the relationship between expenditure towards R&D and TFPG is observed to be positive but statistically insignificant. This may be due to the fact that most of the activities of Indian ITeS firms occur at the lower end of the value chain, where investment on R&D is not essential.

This study has the following policy implications. First, policy should be devised to improve the managerial efficiency of the ITeS firms as it is found to be the only dampening factor for TFPG during the whole study period. Second, keeping in view of the adverse aftermath of the US subprime crisis on the TFPG, policy recommendation should be towards reduction of dependence on the US and UK markets and exploration of domestic and international markets where opportunities are plenty. Third, policy should be formulated to improve the productivity of the private limited firms. Fourth, the positive coefficients of the variable export intensity across three regression models indicate that, on an average, the firms having higher exports are more efficient, innovative and have realized better TFPG. In other words, this implies that policy should be formulated to encourage the ITeS firms for increasing their exports in future.

Notes

[1] Public limited firms are listed in the stock market, whereas the private limited firms are not listed in the stock market.

[2] A group firm is a subsidiary firm of conglomerate/corporations.

[3] CMIE Prowess provides company level aggregate financial data for financial year (FY). For instance, any financial data for FY 2004 indicates from April 2004, to March, 2005. For notational simplicity, we have used 2004 instead of 2004-05 to represent the FY 2004. The same notion is followed for other financial years mentioned in this study.

References

Bhattacharjee, S., "Efficiency dynamics and sustainability of the Indian IT-ITeS industry: an empirical investigation using DEA", *IIMB Management Review*, 24(4), 203-214, 2012.

Caves, R. E., *Industrial efficiency in six nations*, Cambridge, Massachusetts, MIT Press, 1992.

Caves, D. W., Christen, L. R. and Diewert, W. E., "The economic theory of index numbers and the measurement of input, output and productivity", *Econometrica*, 50(6), 1393–1414, 1982.

Chen, Y. and Ali, A. I., "DEA Malmquist productivity measure: new insights with an application to computer industry", *European Journal of Operational Research*, 159(1), 239-249, 2004.

Chou, Y.C. and Shao, B. B. M., "Total factor productivity growth in information technology services industries: A multi-theoretical perspective", *Decision Support Systems*, 62, 106-118, 2014.

Das, P., "An Evaluation of the determinants of total factor productivity growth of Indian information technology industry: an application of DEA-based Malmquist index", *Central European Review of Economics and Management*, 1(4), 175-223, 2017.

Das, P. and Datta, A., "Performance evaluation of Indian information technology-enabled services (ITeS) industry: an application of two-stage data envelopment analysis", *International Journal of Advances in Management and Economics*, 6(2), pp. 52-70, 2017.

Färe, R., Grosskopf, S., Lindgren, B. and Roos, P., "Productivity changes in Swedish pharmacies 1980–1989: a nonparametric Malmquist approach," *Journal of Productivity Analysis,* 3:1/2 (June), 85–101, 1992

Färe, R., Grosskopf, S., Norris, M. and Zhang, Z., "Productivity growth, technical progress, and efficiency change in industrialized countries", *American Economic Review*, 84(1), 66-83, 1994.

Malmquist, S., "Index numbers and indifference surfaces", *Trabajos de Estatistica*, 4, 209-242, 1953.

Mathur, S.K., "Indian IT & ICT industry: a performance analysis using data envelopment analysis and Malmquist index", *Global Economy Journal*, 7(2), 2007.

Sahoo, B. K., "Total factor productivity of the software industry in India", Working paper no. 331, Institute of Economic Growth, New Delhi, 2013.

Sahoo, B. K. and Nauriyal, D.K., "Trends in and determinants of technical efficiency of software companies in India", *Journal of Policy Modeling*, 36, 539-561, 2014.

Shao, B. B. M. and Shu, W. S., "Productivity breakdown of the information and computing technology industries across countries", *The Journal of the Operational Research Society*, 55(1), 23-33, 2004.

Shu, W. S. and Lee, S., "Beyond productivity—productivity and the three types of efficiencies of information technology industries", *Information and Software Technology*, 45, 513-524, 2003.

Websites:

www.prowessiq.cmie.com

www.rbi.org.in

Appendix

Table A1: Poolability Test Results

Regression Equation No.	Dependent Variable	Value of F-test statistic	P-value
1.	Catch-up	18.26***	0.0015
2.	Frontier-shift	14.65***	0.0023
3.	Malmquist Index	12.44***	0.0018

Source: Author's Calculations based on CMIE Prowess database.
***=> Significant at 1% level.

TABLE A2: Breusch-Pagan LM Test Results

Regression Equation No.	Dependent Variable	Value of F-test statistic	P-value
1.	Catch-up	84.52***	0.0002
2.	Frontier-shift	62.57***	0.0004
3.	Malmquist Index	76.18***	0.0000

Source: Author's Calculations based on CMIE Prowess database.
***=> Significant at 1% level.

TABLE A3: Housman Test Results

Regression Equation No.	Dependent Variable	Value of F-test statistic	P-value
1.	Catch-up	4.29	0.3652
2.	Frontier-shift	6.75	0.5414
3.	Malmquist Index	8.13	0.4469

Source: Author's Calculations based on CMIE Prowess database.

"Data is a precious thing and will last longer than the systems themselves."

– Tim Berners-Lee

Feature

Impact of Corporate Social Responsibility in Sustainable Agriculture & Rural Infrastructure

VIKRAM SINGH AND SHAILY DEEWAN

This paper makes an attempt to study the impact of corporate social responsibility (CSR) its process, conceptual and theoretical understanding. It is also based on the assumption that corporate social responsibility is towards sustainable agriculture & rural infrastructure. The paper also aspires to look the widespread adoption of corporate social responsibility (CSR) policies in developing countries have led to call for a concerted effort to better capture CSR effects. The paper considers the practical implications for the effort to assess CSR contribution to community development in developing countries. Rural infrastructure is crucial for agriculture, agro-industries and overall economic development of rural areas. It also, incidentally, provides basic amenities that improve the quality of life. Corporate Social Responsibility has been attracting attention recently by the corporate world worldwide. The corporations discharge their CSR through social development in various ways in varying degree. Methodological considerations, objectives, findings and conclusion have been also discussed.

Vikram Singh, Assistant Professor and Head, Department of Social Work Guru Ghasidas Vishwavidyalaya, Koni, Bilaspur

Shaily Deewan, Student MSW fourth Semester, Department of Social Work Guru Ghasidas Vishwavidyalaya, Koni, Bilaspur

DOI : - https://doi.org/10.32381/PROD.2019.59.04.9

1. Introduction

The word 'sustainable' has become very popular these days and is used to describe many different things. In this lesson, we will explore how sustainability is associated with agriculture and the benefits and issues. Agriculture, with its allied sectors, is unquestionably the largest livelihood provider in India, more so in the vast rural areas. It also contributes a significant figure to the Gross Domestic Product (GDP). Sustainable agriculture, in terms of food security, rural employment, and environmentally sustainable technologies such as soil conservation, sustainable natural resource management, and biodiversity protection, is essential for holistic rural development. Indian agriculture and allied activities have witnessed a green revolution, a white revolution, a yellow revolution and a blue revolution. Furthermore, to give stagnant agricultural growth a boost, a shift must be made from concentrating on the country's food security to focusing on the farmers' income security. The National Agricultural Policy of 2000 stated that private sector participation will be promoted through contract farming and land leasing arrangements to allow accelerated technology transfer, capital inflow and assured a market for crop production.[1]

However, development of agriculture continues to remain critical for India's economic growth, poverty reduction and ensuring food security of the country, as over 58% of rural households depend on agriculture as their principle means of livelihood. Green Revolution, which brought food sufficiency to the country was due to combination of technologies viz., hybrids and high yielding varieties, fertilizers and improved agronomic practices and public policy. This revolution was made possible through an organized and committed agricultural extension system

that ably supported and supplemented the input-intensive production system.

In spite of significant growth in agriculture, Indian agriculture continues to face serious challenges such as declining soil; water and other natural resources; decreasing size of farm holding; Input use inefficiency; costly and scarce agriculture labour; drudgery in farming operations; growing risks in farming; information, knowledge and skill gaps; poor access to credit and investments; slow diffusion of relevant technologies; competiveness of quality and prices in export & domestic markets; inadequate focus on processing and value addition; Low profitability of agriculture; inadequate rural infrastructure; poor access to resources and services for women in agriculture; weak institutional linkages and convergence; extreme events of climate change etc. The real challenge remains in diffusion of technologies generated by the research system to farmers through an effective extension delivery system to address these challenges. Thus, the extension is strategic to the growth of agriculture and allied sectors and enhancing the farmers' income.

Rural infrastructure is generally defined as the physical framework of facilities in rural areas through which, facilities and services are provided to the public. Rural infrastructure assumes great importance in India because of the country's predominant rural nature, the crucial linkages of infrastructure to economic growth, poverty alleviation, and human development. Rural infrastructure covers a wide spectrum of services such as transportation, power generation, transmission and distribution, telecommunication, port handling facilities, water supply, sewage disposal, irrigation, medical, education and other primary services. Rural areas would have a high concentration of poverty given the existence of disguised unemployment in a big way in agriculture. Access to land and ownership of land is the key to income differences since land is the major productive asset in rural areas. Rural areas may be more usefully viewed as the concentration of poor resulting in little value for economic demand for infrastructural services.

The evolution of a relationship between corporate and society has been one of the slow transformations from hard core business to a philanthropic and from philanthropic to stakeholder participation. However, the data shows that very meagre amount of CSR fund has been spent in the agricultural sector by the companies though there is huge potential for investment in the sector. Agriculture cannot be seen in isolation. CSR may be looked in terms of "Creating Shared Value" wherein, the business can help the progress of agriculture and agriculture sector can help business to improve and flourish.

Sustainable agriculture is a type of agriculture that focuses on producing long-term crops and livestock while having minimal effects on the environment. This type of agriculture tries to find a good balance between the need for food production and preservation of ecological system within the environment. In addition to producing food, there are several overall goals associated with sustainable agriculture, including conserving water, reducing the use of fertilizers and pesticides, and promoting biodiversity in crops grown and the ecosystem. Sustainable agriculture also focuses on maintaining the economic stability of farms and helping farmers improve their techniques and quality of life.[2]

Benefits of Sustainable Agriculture:

There are many benefits of sustainable agriculture, and overall, they can be divided into human health benefits and environmental benefits. In terms of human health, crops grown through sustainable agriculture are better for people. Due to the lack of chemical pesticides and fertilizers, people are not being exposed to or consuming synthetic materials. This limits the risk of people becoming ill from exposure to these chemicals. In addition, the crops produced through sustainable agriculture can also be more nutritious because the overall crops are healthier and more natural.[4]

Operational Definitions:

The FAO definition of sustainable agricultural development:

"The management and conservation of the natural resource base, and the orientation of technological and institutional change in such a manner as to ensure the attainment and continued satisfaction of human needs for present and future generations. Such development conserves land, water, plant and animal genetic resources, is environmentally non-degrading, technically appropriate, economically viable and socially acceptable."

Agricultural systems, in both developed and developing countries need to use new approaches to increase food supplies while protecting the resources on which they depend. This can be achieved with

practices that:

- Fully exploit natural processes such as recycling nutrients, using plants that fix their own nitrogen and achieving balance between pests and predators;
- Reduce the reliance on inputs such as mineral fertilizers and chemical pesticides;
- Diversify farming systems, making greater use of the biological and genetic potential of plant and animal species;
- Improve management of natural resources;
- Rotate crops or develop agro-forestry systems that help maintain soil fertility. This can be seen in Figure 1.1 below.

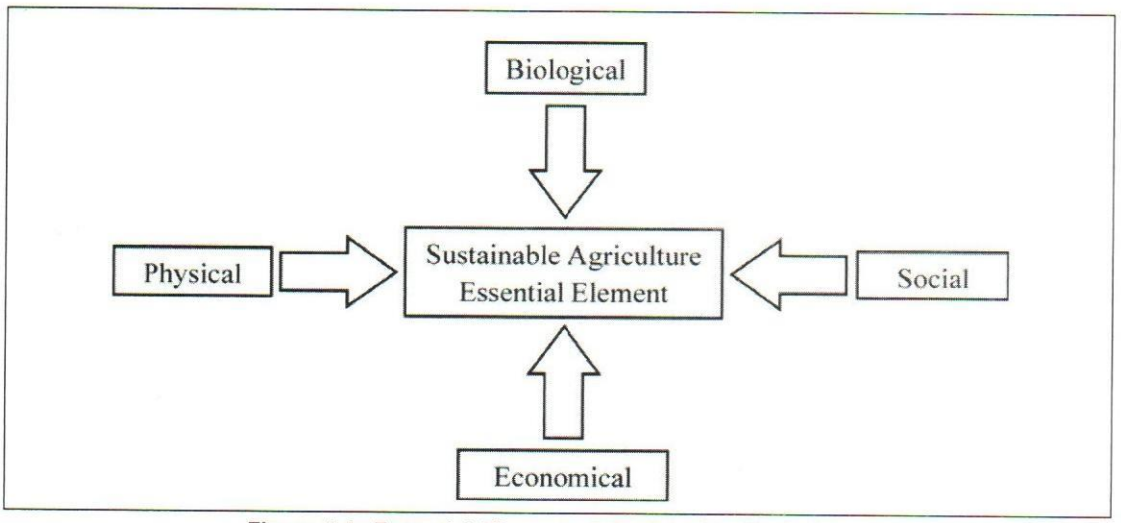


Figure 1.1: Essential Element of Sustainable Agriculture

Corporate Social Responsibility:

According to Carroll (1983), "Corporate social responsibility involves the conduct of a business so that it is economically profitable, law-abiding, ethical and socially supportive. To be socially responsible then means that profitability and obedience to the law are foremost conditions when discussing the firm's ethics and the extent to which it supports the society in which it exists with contributions of money, time and talent". This diversity of conception is a testimony to Moon's (2002) observation that CSR, similar to other important concepts like democracy and justice, is "essentially contested". Moon (2002) also makes the point that CSR "is only one of several terms in currency designed to capture the practices and norms of new business-society relations. There are contending names, concepts or appellations for corporate social responsibility" It is clear from Figure 1.2.

Cells and Walton (1961) described social responsibility in their conceptual foundations of business as follows: When people talk about CSR, they are thinking in terms of the problems that arise when corporate enterprises caste its shadow on the social scene and of the ethical principles ought to govern the relationships between the corporations and society'.

Drucker (1946) in his article – The concept of the corporation published in Business and Society Review has observed that under any circumstance we are moving in the direction of demanding that our institutions take responsibility beyond their performance and beyond their own contribution. We will demand this as from the university, the hospital, the government agency, the school and even from charitable organizations and place of worship.

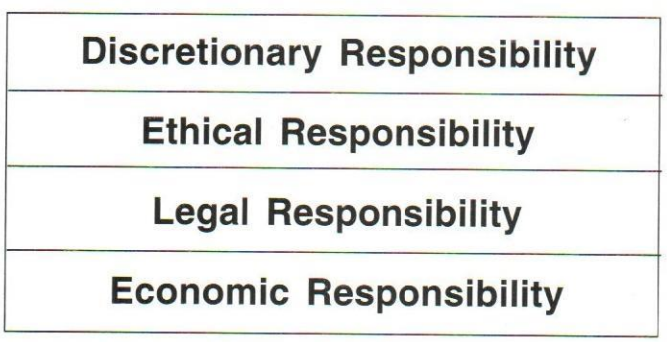


Figure 1.2: Social Responsibility Categories (Carroll:1979)

Rural Infrastructure:

Rakesh Mohan Committee Report (1996) & the Central Statistical Organisation (CSO):. Rakesh Mohan Committee in "The India Infrastructure Report" included Electricity, gas, water supply, telecom, roads, industrial parks, railways, ports, airports, urban infrastructure, and storage as infrastructure. Except for industrial parks and urban infrastructure, all these sub-sectors are treated by CSO also as infrastructure.

Insurance Regulatory and Development Authority (IRDA):

The IRDA (Registration of Indian Insurance Companies) (Second Amendment) Regulations, 2008 defined infrastructure to include following:

- A road, including a good quality of roads, bridges or rail network system.
- Construction of highway, port, airport, inland waterways, water supply project, irrigation management project, water treatment, and management system, sanitation, and sewerage disposal system or solid waste management system is an integral part of the infrastructural activities.
- The industrial park, special economic zone, generation and distribution of electricity by laying a network of a new grid of electricity lines, construction agro-based processing units, and supply of agricultural inputs.
- Construction of educational institutions and hospitals and another public facility of similar nature as may be notified by the Authority in this behalf in the Official Gazette.

The World Development Report of 1994 included the following in its definition of Infrastructure:

- Public utilities - power, telecommunications, piped water supply, sanitation and sewage, solid waste collection and disposal and piped gas.
- Public works - roads, major dam, and canal works for irrigation and drainage.
- Other transport sectors-urban and inter-urban railways, urban transport, ports and waterways, and airports. (World Bank, 1994).

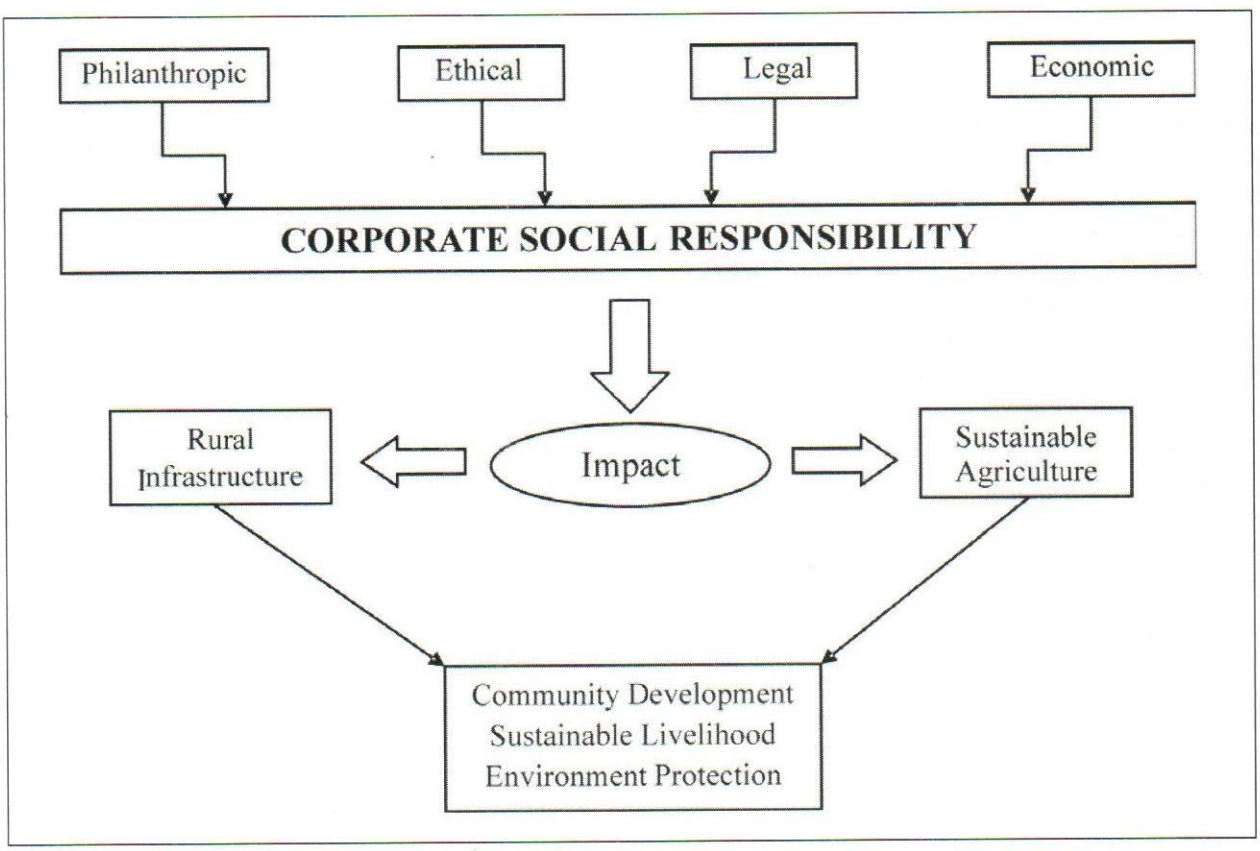


Figure 1.3: Conceptual framework of Corporate Social Responsibility

Objectives of the Paper:

- To study the impact of corporate social responsibility in the field of sustainable agriculture and rural infrastructure.
- To analyse the community perception about the CSR activities on rural infrastructure and sustainable agriculture.
- To study the role of CSR in community development through sustainable agriculture.

Research Design:

The research design means the conceptual structure within which research is conducted. It constitutes the blueprint for collection, measurement and analysis of data. The research design is needed because it facilitates smooth sailing of various research operations, thereby making research as efficient as possible yielding maximum information with minimal expenditure of effort, time and money. In this study, empirical research design is used which means that conclusion drawn in this article is based upon need evidence gathered from information collected from community people experience and researcher observations.

Universe of the Paper:

Population or Universe is the aggregate of all elements possessing certain specified characteristics which need to be studied and defined prior the sample population. The total population of Janjgir block, according to the Census of India 2011, was 40,561. The universe of the study was Lchhanpur and Basantpur community. The total population in Lachhanpur community was 2981 and Basantpur community is 1981. These two communities out of 16 communities has been selected for the paper.

Respondent Group:

The study targeted Lachhanpur and Basantpur communities within the CSR sphere of Marwa Tendubhata power plant that operates in Janjgir Champa and is socially responsible for these community developments. The target respondents comprised of beneficiaries of CSR activities within the Marwa Tendubhata power plant which includes the local community population of both the communities.

Sample Size:

An optimum sample is one which fulfils the requirements of efficiency, representatives, reliability and flexibility. The choice of sample size depends on non-statistical considerations and statistical considerations. The non-statistical considerations may include availability of resources, manpower, budget, ethics and sampling frame. The statistical considerations will include the desired precision of the estimate of prevalence.

The sample size for the study consisted of 40 respondents that have drawn from Lchhanpur and Basantpur community.

Sampling Technique: In this paper purposive sampling technique (non-probability sampling) has been used to select respondents purposefully to achieve the objectives of the paper. Because it is be very useful in situations when researcher needs to reach a targeted sample quickly in the selected villages, and because sampling for proportionality is not the main concern.

Sources of Data:

Generally, data can be collected from two sources, primary source and secondary source. Data collected from a primary source known as primary data and data collected from the secondary source called secondary data. **This paper is based on both primary and secondary sources of data.**

Primary data was collected from the sample of selected 40 members of Lachhanpur and Basantpur community of Janjgir –Champa district in Chhattisgarh. Primary data is collected via -structured Interview schedule. **Secondary data** is obtained from various publications of the central government, journals, books, magazine, newspaper and reports related to the concerned area of article.

Tools for Data Collection:

Interview schedule: - Interview schedule prepared by the researcher which containing the relevant questions on the topic. The researcher conducted face to face interview with respondents with these schedule. Data has been collected by filling up the schedule by the researcher on the basis of responses given by respondents.

Observation: - This method implies the collection of information by way of investigators own observation. Participant observation method has been used for collecting information about the operation/attitudes existing within community by researcher while living in the area for an extended period. In this paper participant observation is used. Hence observation method is one of the several methods of data collection. Simply collection of data

without the actual observation leads to research gap. Therefore, the study adopted participant observation method for data collection.

Characteristic of Interview Schedule Variables:

The Interview Schedule was constructed in four sections comprising:

Section I : Demographic information of respondent.

Section II : Status of social infrastructure in selected villages

Section III : The extent to which the selected community has benefited(sustainable agriculture) from CSR practices of the Marwa Tendubatha Thermal Power Plant.

Section IV : Village Infrastructure/Assets.

Section I: Demographic Information of Respondent

This section was planned to obtain social and occupational status of the community people. In addition number of the question were structured regarding gender, marital status types of houses they owned, drinking water facilities, monthly income, educational status, occupation, type of drainage etc. to know about the basic detail of respondent.

Section II: Status of Social/Rural Infrastructure in selected Villages

This section has been planned to obtain the details related status of social infrastructure in selected villages, which includes social relationships, gender roles, Castes system, Aaganwadi, Mini Aaganwadi, Health Centre, Primary School, Upper Primary School, ASHA worker the major focus of these part is to know about the level of improvement in quality of life of community people through CSR activities of Marwa Tendubhata power plant.

Section III: Extent to which the selected Communities has benefited from CSR practices by the Marwa Tendubatha Thermal Power Plant.

This section was planned to obtain the details related to community people view about the CSR activities performed by Marwa Tendubhata power plant and gather details related to the extent to which local community has benefited from the CSR practices by the Marwa Tendubatha Thermal Power Plant.

Section IV: Village Infrastructure/Assets/Sustainability in Agriculture through CSR.

This section was planned to obtain the details of sustainability in agriculture through CSR intervention along with rural infrastructure, public relation, activities based on preserving the environment like pollution prevention programs, awareness program in the community. It also analyzed how CSR activities have provided significant employment to the local community people and production of crops and enhancement in services

Infrastructure assets such as rural roads, tracks, bridges, irrigation schemes, water supplies, schools, health centres and markets are needed in rural areas for the local population to fulfill their basic needs and live a social and economic productive life.

Community Profile :

I. About Community–Lachhanpur	
Name of the Community	**Lachhanpur**
Block	Baloda
Tehsil	Janjgir
District	Janjgir-Champa
Type of community	Rural Community
Nature of community	Heterogeneous
Special feature	The 2500 MW Marwa Tendubhata Thermal Power Plant (MTTPP)

Community Map :

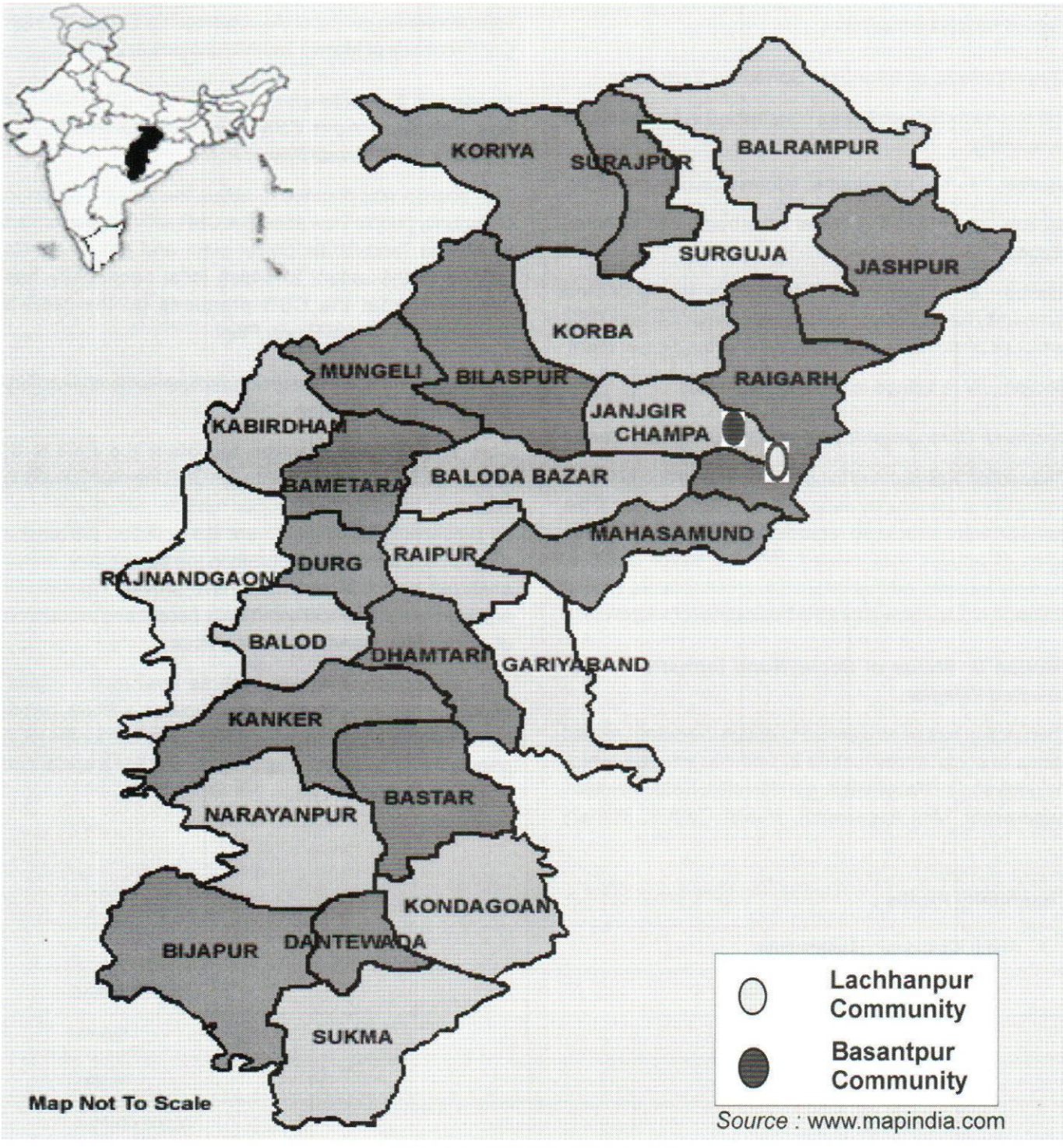


Figure: 1.4 Universe for Research

Total area	212.9 Hectare
The area under Non- Agriculture Uses	1.75 Hectare
Total Irrigated Land Area	35.78 Hectare
Main Language	Hindi and Chhattisgarhi
Total no. of Households	648
II. Resource Persons:	
Name of Sarpanch	Mr Teras Ram Yadav
Source	www.censusindia.gov.in
III. Demographic Profile:	
Total Population	2981
Male	1558
Female	1423
Total SC Population	418
Total SC Male Population	226
Total SC Female Population	192
Total ST population	170
Total ST Male Population	89
Total ST Female Population	81
Total child(0-6) Population	450
Total child(0-6) Male Population	221
Total child (0-6) Female Population	229
IV. Educational Status:	
Total No. of Senior Secondary School	0
Total No. of Middle School	0
Total No. of Primary School	0
Total No. of Teacher in Higher Secondary School	0
Total No. of Teachers in Middle School	0
Total No. of Teachers in Primary School	0
Total Private Primary School	1
Total Private Middle School	1
V. Literacy Rate:	
Total Literate People	73.13%

Male	83.47%
Female	61.56%
VI. Basic Amenities:	
Total No. of Hand Pump	32
Total No. of Tub Well	17
Total No. of Water Tanks	02
VII. Basic Health Facilities:	
Total no. of CHC	0
Total no. of PHC	0
VIII. Basic Infrastructure Facilities:	
Post Office	0
Sub Post Office	0
Telephone (landline)	Available
Common Service Centre	0
National Highway Road	Available
State Highway Road	Not Available
Pukka Road	Available
Water Bounded Macadam	Not Available
ATM Status	Not Available
Commercial Bank	Not Available
Cooperative Bank	Not Available
Agricultural Credit Societies	Not Available
Self - Help Group (SHG)	02
Public Distribution System (PDS) Shop	Not Available
Tap Water-Treated	Not Available
Covered Well	Not Available
Uncovered Well	Available
Hand Pump	Available
Tube Wells	Available
Closed Drainage	Not Available

Open Drainage	Available
River/Canal	Available
Community Toilet Complex (including Bath) for	
General Public	Not Available
IX. Basic Facilities	
The total length of Drainage System	267 mtr.
The total length of Pipeline	623 mtr.
Total no. of Electricity poles	49
The total length of Road	212 mtr.
Total no. of Society	02
Total no. of Aaganwadi Centre	03
X. Poverty Alleviation Programmes/Schemes:	
Total No. of APL card holders	1,235
Total No. of BPL card holders	231
Total No. of AYY card holders	79
XI. Social Welfare Schemes Implemented:	
-	National Swasthya Bima Yojna
-	Indira Gandhi National Disabled Pension Scheme
-	Indira Gandhi Widow Pension Scheme
-	Social Security Pension Scheme
-	Sukhad Sahara Pension Scheme
-	Indira Gandhi Family Benefit Scheme
XII. Main Source Of Income:	
-	Agriculture
Agricultural Commodities	
"	FIRST -Paddy
"	SECOND - Tivra
"	THIRD- Liseed
-	Unorganized Labour
-	Industrial Labour

-	Public/Govt. Services
XIII. Main Source Of Fuel:	
-	Gas Cylinder
-	Fuel Wood and Dung Cake
-	Electric Heater
Basic Profile of the Community Basantpur:[1]	
I. About Community-Basantpur:	
Name of the Community	Basantpur
Block	Baloda
Tehsil	Janjgir
District	Janjgir-Champa
Type of community	Rural community
Nature of community	Heterogeneous
Special feature	The 2500 MW Marwa Tendubhata Thermal Power Plant (MTTPP)
Total area	230.32 hectare
The area under Non- Agriculture Uses	1.41 hectare
Total Un-irrigated Land Area	153.43 hectare
Main Language	Hindi and Chhattisgarhi
Total no. of Households	461
II. Resource Persons:	
Name of Sarpanch	Mr. Laxmi Narayan Sahu
Source	www.censusindia.gov.in
III. Demographic Profile:	
Total Population	1981
Male	979
Female	1002
Total SC Population	47
Total SC Male Population	26
Total SC Female Population	21
Total ST population	66
Total ST Male Population	38
Total ST Female Population	28

Total child (0-6) Population	431
Total child (0-6) Male Population	178
Total child (0-6) Female Population	163
IV. Educational Status:	
Total No. of Senior Secondary School	1
Total No. of Middle School	1
Total No. of Primary School	1
Total No. of Teacher in Higher Secondary School	11
Total No. of Teachers in Middle School	05
Total No. of Teachers in Primary School	08
V. Literacy Rate:	
Total Literate People	82.56%
Male	93.13%
Female	72.47%
VI. Basic Amenities:	
Total No. of Hand Pump	32
Total No. of Tube Well	17
Total No. of Water Tanks	02
VII. Basic Health Facilities:	
Total no. of CHC	0
Total no. of PHC	0
VIII. Basic Infrastructure Facilities:	
Post Office	1
Sub Post Office	1
Telephone (landline)	0
Common Service Centre	0
National Highway Road	0
State Highway Road	0
Pukka Road	Available
Water Bounded Macadam	Not Available

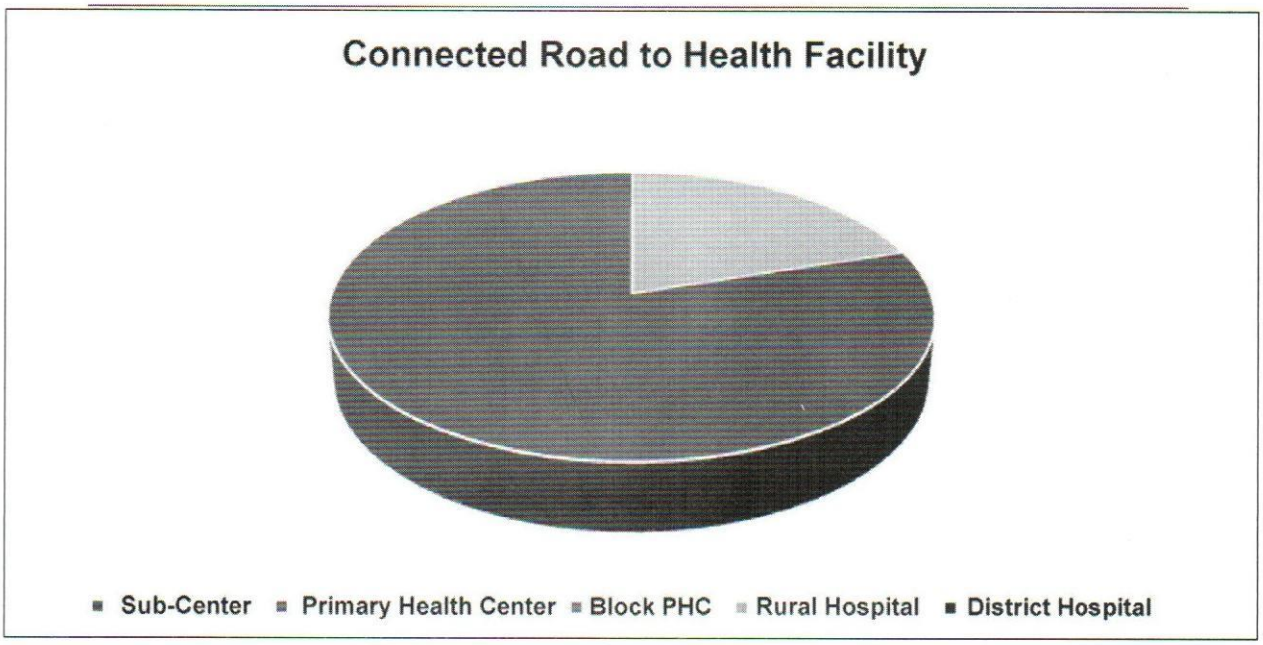
ATM Status	Not Available
Commercial Bank	Not Available
Cooperative Bank	Not Available
Agricultural Credit Societies	Not Available
Self - Help Group (SHG)	02
Public Distribution System (PDS) Shop	Not Available
Tap Water-Treated	Not Available
Covered Well	Not Available
Uncovered Well	Available
Hand Pump	Available
Tube Wells	Available
Closed Drainage	Not Available
Open Drainage	Available
Community Toilet Complex (including Bath) for	Not Available
IX. Basic Facilities:	
The total length of Drainage System	245 mtr.
The total length of Pipeline	575 mtr.
Total no. of Electricity poles	37
The total length of Road	164 mtr.
Total no. of Society	02
Total no. of Aaganwadi Centre	03
X. Poverty Alleviation Programmes/Schemes:	
Total No. of APL card holders	1,235
Total No. of BPL card holders	231
Total No. of AYY card holders	79
XI. Social Welfare Schemes Implemented:	
- National Swasthya Bima Yojna	
- Indira Gandhi National Disabled Pension Scheme	
- Indira Gandhi Widow Pension Scheme	
- Social Security Pension Scheme	
- Sukhad Sahara Pension Scheme	
- Indira Gandhi Family Benefit Scheme	

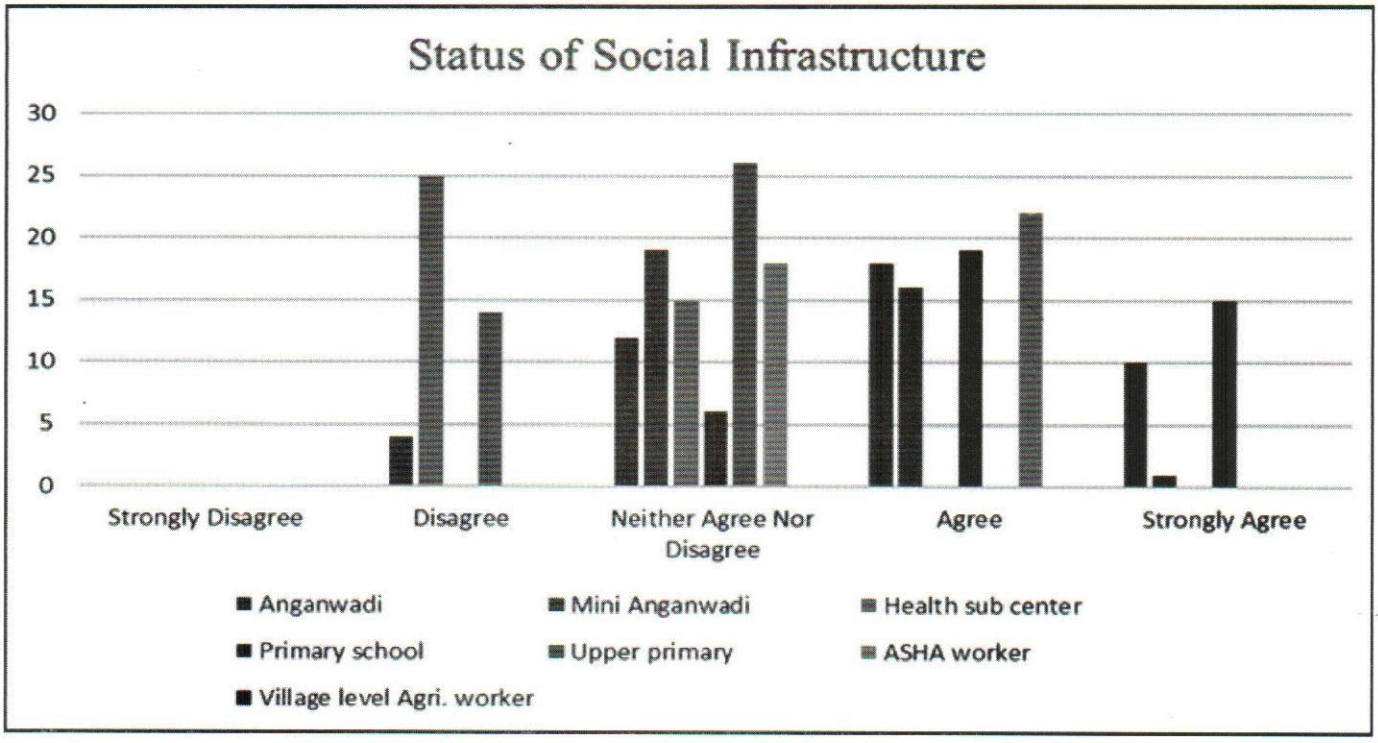
XII. Main Source Of Income:	
Agriculture	
- Agricultural Commodities	
	• FIRST –Paddy
	• SECOND – Tivra
	• THIRD– Linseed
-Unorganized Labour	
-Industrial Labour	
-Public/Govt. Services	
XIII. Main Source of Fuel:	
-	Gas Cylinder
-	Fuel Wood and Dung Cake
-	Heater
The Basic Profile of Respondent:	
Total No. of Respondents	**40**
1. Sex-ratio of the Respondents:a)	
(a) Male	20
b) Female	20
2. Age of the Respondents:	
Age Group	**Number**
a) 18-25 years	04
b) 26-35 years	18
c) 36- 45 years	17
d) Over 56 years	01
3. Educational Status of the respondents:	
Level of Education Received	**Number**
a) Under Secondary School	06
b) Secondary School	26
c) High School	08
d) Higher than High School	0

4. Occupation of the respondents:	
a) Student	0
b) Factory Worker	02
c) Farmer	18
d) Business	08
e) Other	12
5. Monthly Income of the Respondents:	
a) Under 5000	16
b) 5,000-10.000	14
c) 10,001-15,000	08
d) Over 15,000	02
6. Housing Status of the Respondents:	
a) Own House	10
b) Rental House	02
c) Kucha House	19
d) Pukka House	09
7. Agriculture land owned by respondent:	
a) 0-1 hectare	10
b) 1-2 hectare	13
c) 2-4 hectare	10
d) 4 hectare and above	07
8. The main source of drinking water	
a) Pipe water	00
b) Tanker	00
c) River and pond	24
d) Well	16
9. Toilet facility available to the Respondents:	
a) Own Toilet	30
b) Public Toilet	0
c) Open Space/Field	10

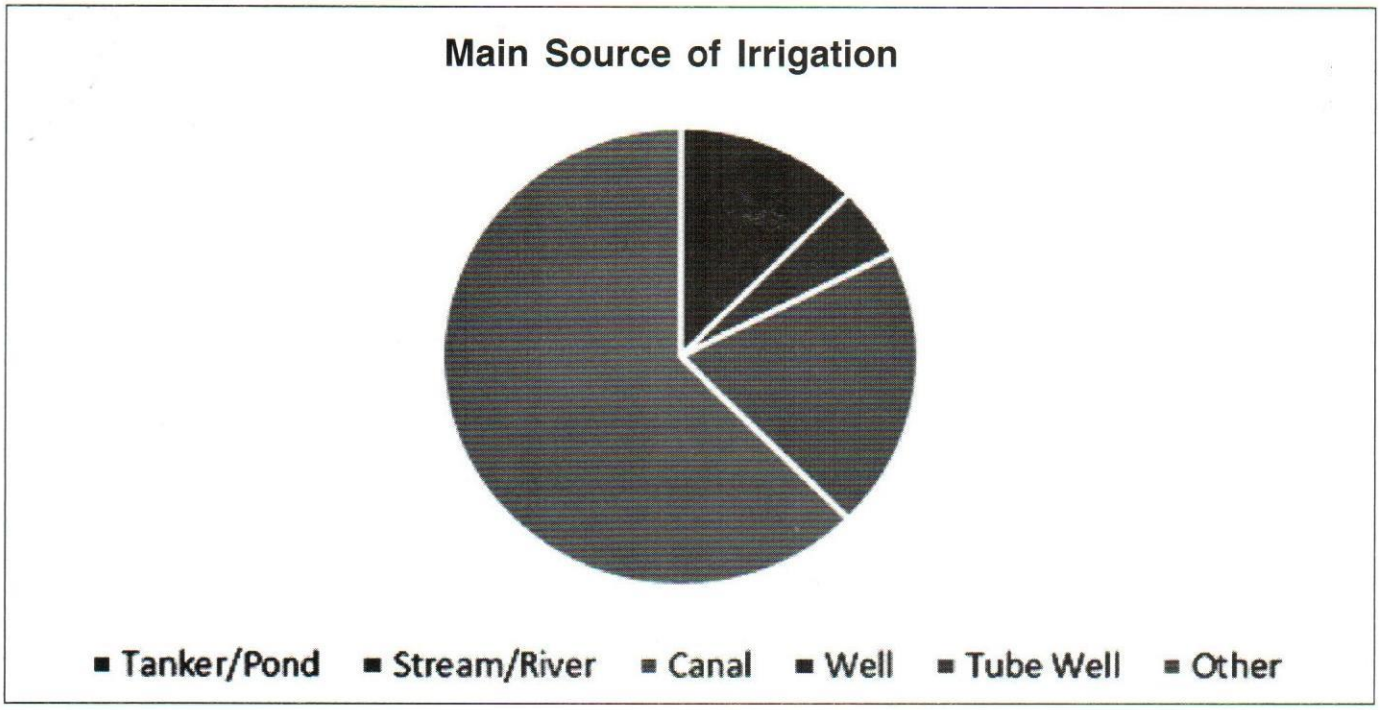
Data Interpretation:

Data Representation used in this paper is the result of surveys conducted by author among 40 people with the help of students placed under his supervision in the Janjgir Champa region, Chhattisgarh.

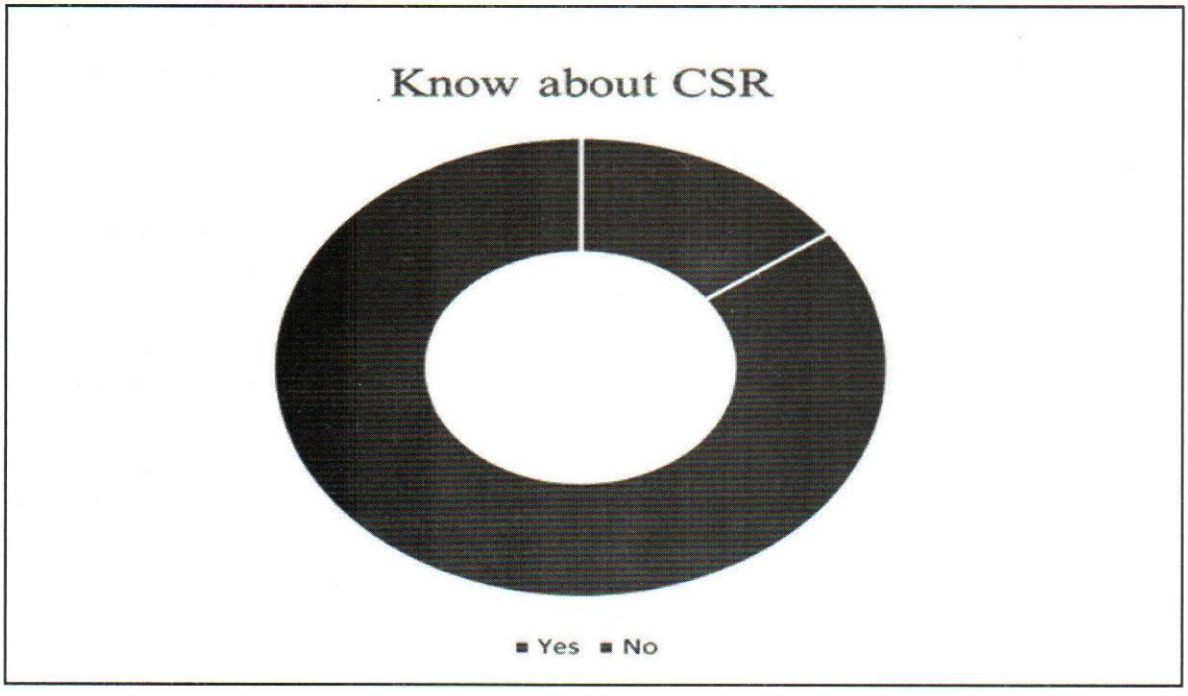




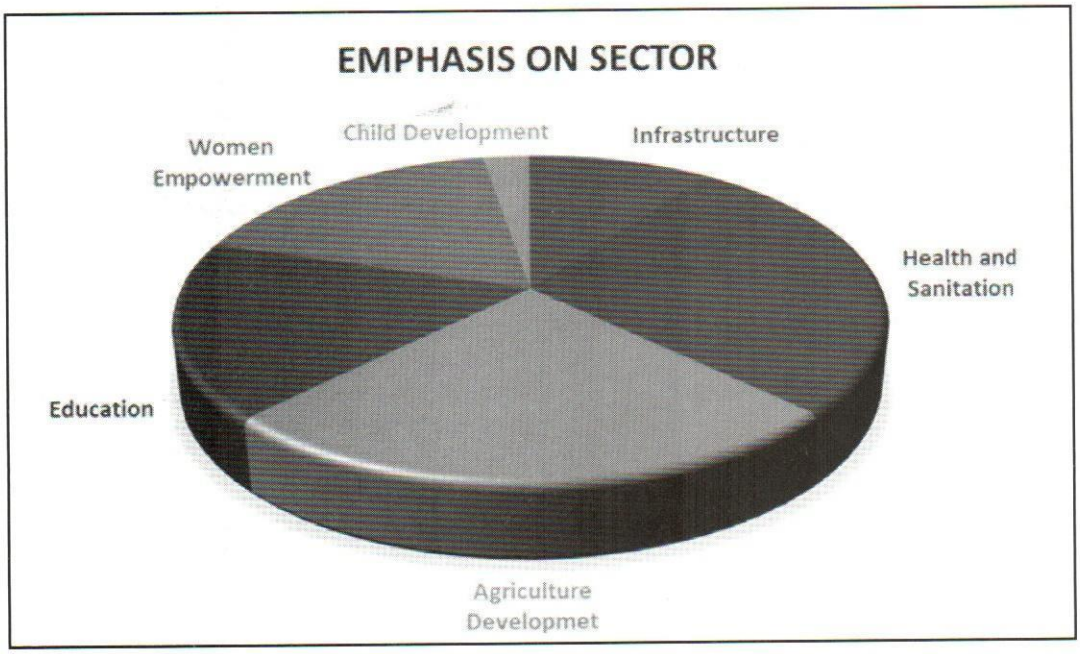
The above bar graph represents to the status of social infrastructure in the community out of total respondent, the majority of respondent have disagreed for the better condition of infrastructure and there is need to focus on infrastructure development.



The above pie chart represents to the main source of irrigation in the community out of the total population, 12% says that Tank/Pond is the major source of irrigation in the community than 20% response to the canal as the main source of irrigation and 63% are the response to the other source.



The pie chart shows there is 85% of the population don't know about the CSR activities carried out by Marwa Tendubatha Power Plant only 15% response to the yes knowledge about CSR in a village, which shows there is need to establish a good public relation between Power Plant authorities and community population.



The above pie chart represents the community view about the sector in which there is the need to emphasize. The above data shows there is a need to emphasize more on health and sanitation sector 27% is a response to the emphasis on health and sanitation sector and 25% are responding to the agriculture development sector and only 3% response to the child development sector.

Analysis & Conclusion:

Hence speaking about inclusive growth, after negligence of 70% Indians who live in rural areas would not be possible. Therefore most of the local bodies including the government started focusing on rural development not just to help the rural masses but most importantly for helping themselves in sustenance of agriculture so the raw product can be provided to secondary sectors. Though improving the fate of rural masses and sustenance in agriculture without creating necessary infrastructure is just a daydreaming that would never happen. Thus, in the backdrop of rural development what lies most sternly is rural infrastructure like rural roads, rural water supply, rural housing, rural electrification, irrigation, etc. The government in India is not affluent enough to cater all necessary infrastructures to rural areas for their development. Meanwhile, the concepts of public-private partnership (PPP) and corporate social responsibility (CSR) have gained popularity in recent times.

There is a need for emphasis on sustainable agriculture for better community development and environmental protection through CSR, which will help the community for sustainable development. Hence, it can be analyzed that there is need to implement the FAO principles framework for sustainable agriculture investment through CSR.

Principle 1: Existing rights to land and associated natural resources are recognized and respected.

Principle 2: Investments do not jeopardize food security but strengthen it.

Principle 3: Processes for accessing land and other resources and then making associated investments are transparent, monitored, and ensure accountability by all stakeholders, within a proper business, legal, and regulatory environment.

Principle 4: All those materially affected are consulted, and agreements from consultations are recorded and enforced.

Principle 5: Investors ensure that projects respect the rule of law, reflect industry best practice, are viable economically, and result in a durable shared value.

Principle 6: Investments generate desirable social and distributional impacts and do not increase vulnerability.

Principle 7: Environmental impacts due to a project are quantified and measures are taken to encourage sustainable resource use while minimizing the risk/ magnitude of negative impacts and mitigating them.

References

Agrawal, S. K. (04/04/2008). *Corporate Social Responsibility in India.* India: SAGE Publications Pvt. Ltd.

Gurbir S Bhullar, N. K. (2012). *Agricultural Sustainability: Progress and Prospects in Crop Research.* United Kindom: Academic Press.

Idowu, S. O. (2014). *Stages of Corporate Social Responsibility.* Salmon Tower Building New York City: Springer publication Pvt Ltd.

R K Mishra, P. S. (2015). *Revolutionising Corporate Social Responsibility: Towards a Sustainable Tomorrow.* New Delhi: Academic Foundation.

Saha, D. (2000). *rural development in a new dimension.* India: Global Vision Publishing House.

Singh, K. (2009). *Rural Development: Principles, Policies and Management.* india: SAGE Publications India Pvt Ltd.

"Things get done only if the data we gather can inform and inspire those in a position to make [a] difference."

– Mike Schmoker

Feature

Effect of Weather Changes on the Credit Risk in Agricultural Microfinance: An Indian Perspective

SUHAIL AHMAD BHAT AND MUSHTAQ A. DARZI

The objective of the paper is to determine the impact of weather changes on credit risk of loans granted to small scale farmers by microfinance institutions. Generally small-scale farmers in developing countries are undersupplied with credit by microfinance institutions, because they believe that lot of risk is involved in loan repayment from farmers. Therefore, the study employs secondary data obtained from Indian Metrological Department regarding average rainfall received during the four seasons pre-monsoon, monsoon, post-monsoon and winter. With regard to credit risk in agricultural microfinance, data was obtained from the Reserve Bank of India depicting annual leading and outstanding towards marginal, small, medium and large farmers by schedule commercial banks in India. Results of the study depicts that rain has insignificant impact on credit risk of loans granted to marginal, small, medium and large scale farmers. The study will guide farmers and policy makers in understanding the real cause of credit risk and non-payment of debt, which is the main cause of farmers' suicide in India. The study has taken data of the last 27 years from 1985-2011. This is the first empirical study that determines the effect of weather changes on credit risk of micro-loans granted to small-scale farmers in India.

Suhail Ahmad Bhat, Research Scholar, Department of Management Studies, University of Kashmir

Mushtaq A. Darzi, Professor, Department of Management Studies, University of Kashmir

DOI : - https://doi.org/10.32381/PROD.2019.59.04.10

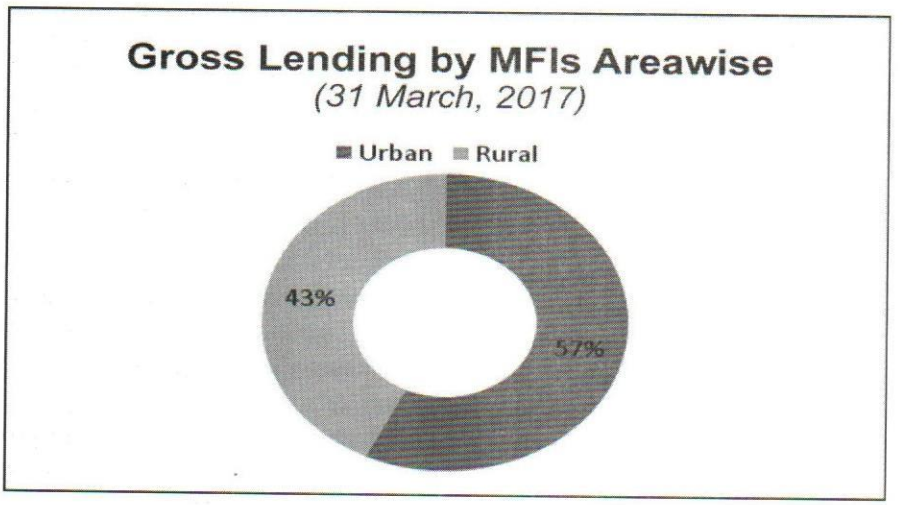
1. Introduction

Development of microfinance institutions worldwide has expanded the frontier of finance by providing loans and other financial services to poor and underdeveloped people. This area has attracted significant attention lately after the UN designated 2005 as the International Year of Microcredit. It has been estimated that approximately 1.2 billion people worldwide are classified as extremely poor (United Nations Development Program, 2013). More than 75% of the people, who are affected by poverty live in rural areas, are in one way or another associated with farm business (Food and Agriculture Organization, 2002). Therefore, advancements in the form of microfinance within agricultural sector, is one of the primary step to fight poverty. Besides, rural financial markets and institutions play an important role in the process of economic development and poverty alleviation. Bardhan and Udry (1999) have opined that lack of access to credit market is one of crucial factors for anti-poverty. Consequently, lenders have emphasized the problems associated with adverse selection, unpredictable returns and moral hazards; have considerable effect on access of credit to poor (Stiglitz and Weiss, 1981). Microfinance Institutions (MFIs) have adopted number of mechanism to reach their services out to the poor while maintaining financial viability, group guarantee, frequent loan repayments, women targeting and sustainable rates of interest (Morduch, 1999).

It has been estimated that around 10,000 microfinance institutions exist worldwide. In developing countries MFIs have mostly focused on the provision of financial services to micro, small and medium-sized enterprises (Godquin, 2004). The number of micro-lenders has increased from 1997 to 2012 by 1.307%, indicating enormous growth in this sector (Maes and Reed, 2012). Even though the rate of financial inclusions has progressed

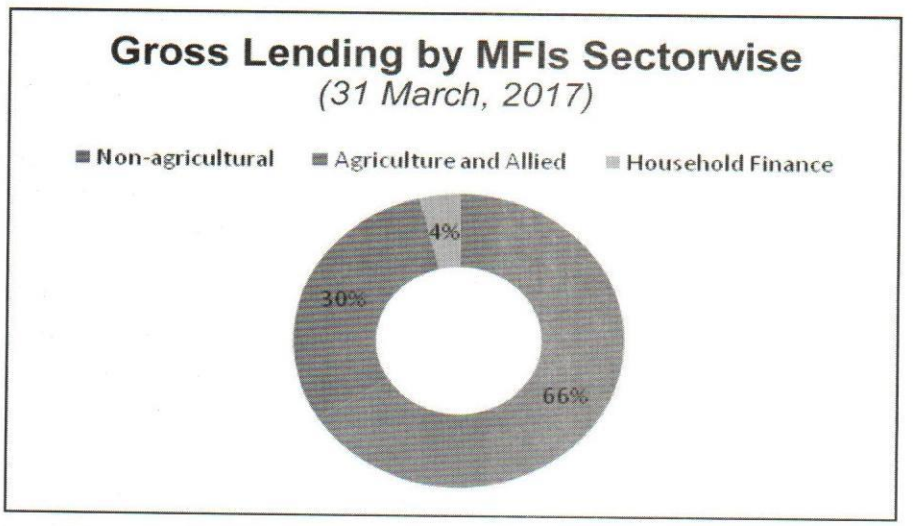
rapidly in recent years but still a large gap between developing and developed countries exists. For instance, the ratio of private credit to GDP in countries with large microfinance markets hovers around 40 percent while as in developed countries the ratio of private credit to GDP is above 100% (World Bank, 2015). This gap indicates that microfinance sector still has plenty of room to catch up growth of the economy of developing countries. India has also witnessed a growth in microfinance industry with a total loan portfolio of Rs 106,916 cr. This represents a growth of 26% over the last year (Micrometer, 2017). The studies conducted by authors namely Andrewa (2006); Hartarska and Holtmann (2006); Christen and Anderson(2013) indicates that everyone does not have the same opportunity for accessing microcredit. Most of the MFIs in developing countries have Gross Loan Portfolio (GLP) consisting primarily of loans for urban businesses. In India 57% of Gross Lending Portfolio by MFIs is towards urban areas and 43% is towards rural areas as shown in Figure 1.

Further, the sector-wise GLP indicated that only 30% investment is being made in the agricultural and allied businesses whereas 66% is towards non-agricultural businesses and 4% towards household finance as shown in Figure 2.



Source: Micrometer, 2017

Figure 1. Area-wise Gross Lending Portfolio of MFIs in India



Source: Micrometer, 2017

Figure 2: Sector-wise Gross Lending Portfolio of MFIs in India

MFIs in India have not made much progress towards the lending of farmers as clients. Consequently, agricultural sector has poor access to the credit (Cabannes, 2012). One of the major problem that deprives this sector from credit financing is their exposure to risk, which makes MFIs difficult to lend them (Diagne *et al.*, 2000; Foltz, 2004; Beck *et al.*, 2006; Simtowe *et al.*, 2008; Weber and Musshoff, 2012). Due to these difficulties farmers in India mostly rely on non-institutions sources of credit than commercial banks. Moreover, non-institutional sources provide some flexibility in repayment and are verbal contracts (less documentation) between the farmers and money lenders thereby making it convenient. However, non-institutional sources have risk in terms of high interest rates, less flexibility in marketability of the agricultural produce and lesser output gains. These problems put additional burden on the sustainability of farmers in terms of the credit availed.

Agricultural sector represents 35% of India's Gross National Product and plays an important role in economic development. Indian agriculture is mostly dependent on monsoon, which is received in the months of June-August. Any change in the weather significantly affects crop yield and thereby increases volatility of annual cash flows (Binswanger and Rosenzweig, 1986). Therefore, weather risk in agriculture yield has been assumed as an important factor affecting agricultural lending (Giné and Yang, 2009; Miranda and Gonzalez-Vega, 2011; Weber and Musshoff, 2012). Weather changes influence profitability of agricultural production and jeopardize the capacity of small-scale farmers to handle the loan repayments in due time. Consequently, weather risks are expected to render the Gross Lending Portfolio (GLP) of MFIs towards rural agricultural sectors considerably riskier than urban businesses (Miranda and Gonzalez-Vega, 2011). Substantial amount of prior literature is available on cost structure and performance of MFIs in European regions (Altunbas *et al.*, 2001; Pastor, 2002; Hartarska and Nadolnyak, 2008; Hartarska and Mersland, 2012). Moreover, there has been little systematic empirical literature available till date on the performance of MFIs, relationship lending and community banking (Berger and Udell, 2002; Berger *et al.*, 2003, 2004). Despite growing importance of MFIs there is very little rigorous empirical evidence on effect of weather shocks on financial outcomes of MFIs. This study contributes to the literature by investigating the possibilities of weather variations on the repayment performance of loans granted by MFIs to small-scale farmers.

Literature Review and Hypotheses

To accelerate the economic development in developing countries, the state-owned banks need to increase the lending rate to the priority sector. However, repayment rate and allocation inefficiencies of loans issued were low hence failed to create sustainable credit access to priority sector (Krahnen and Schmidt, 1994; De Janvry, 2010). Due to these inefficiencies, focus has shifted towards the MFI, which follow both commercial and developmental goals. The microfinance approach addresses these information asymmetries and lack of collaterals by proper client assessments and regular loan repayment schedules (Armendariz and Morduch, 2000). These institutions provide credit to poor farmers and small businesses, however, loan products for farmers are assumed to be more risky due to seasonality and risk in agricultural production. Because of these risks farmers are deprived from profitable investments in the agricultural sector (Field *et al.*, 2011). Therefore, weather related agricultural risk to be the main cause that affects cash flow variance in agricultural than in non-agricultural enterprises.

The effect of weather (precipitation) has been discussed in theory but very rare empirical evidences are available in the economic literature (Skees and Barnett, 2006; Czura and Klonner, 2010; Berg and Schrader, 2012). Collier *et al.* (2011) illustrated that natural disasters can significantly threaten financial institutions serving the poor. The researchers have actually studied the effect of El Niño-related catastrophic floods that destroys infrastructure, production assets and private homes of small borrowers on microfinance lending in Peru. Pelka *et al.* (2015) have found that excessive rain in the harvest period of rice increases the credit risk of loans granted to small-scale farmers in Madagascar. Raghunathan *et al.* (2011) and Weber and Musshoff (2012) have argued against the higher risk perception of farmers while addressing the difference in repayment performance for agricultural micro-borrowers. Berg and Schrader (2012) have found positive effect of volcanic eruptions on loan repayment risk and interest rates of microfinance clients when loans are approved after high volcanic eruption in Ecuador. In view of the above discussions following hypothesis has been propounded:

H: *Weather changes significantly affect repayment performance/credit risk rate of loans granted to small-scale farmers in India.*

TABLE 1: All India Scheduled Commercial Banks' Direct Finance to Farmers According to Size of Land Holdings (Disbursements and Outstanding) Short-term and Long-term Loans

Year	Marginal Farmers (Up to 2.5 acres)			Small Farmers (Above 2.5 acres to 5 acres)			Medium and Large Farmers (Above 5 acres)			Total	
	Disbursements (₹ Billion)	Outstanding (₹ Billion)	O/D Ratoo	Disbursements (₹ Billion)	Outstanding (₹ Billion)	O/D Ratio	Disbursements (₹ Billion)	Outstanding (₹ Billion)	O/D Ratio	Disbursements (₹ Billion)	Outstanding (₹ Billion)
1985-86	6.2	15.26	2.47	5.9	14.83	2.51	10.4	36.78	3.54	22.4	66.87
1986–87	7.6	16.82	2.21	7.1	16.83	2.37	12.8	40.23	3.14	27.4	73.88
1987–88	8.2	20.15	2.44	7.6	20.44	2.68	13.6	50.29	3.69	29.4	90.88
1988–89	8.8	23.24	2.63	8.4	23.12	2.76	14.7	54.6	3.71	31.9	100.96
1989–90	10.3	27.27	2.64	8.9	26.73	3.00	16.1	64.94	4.04	35.3	118.94
1990–91	11.8	28.95	2.45	9.5	28.7	3.01	17.8	66.24	3.71	39.1	123.89
1991–92	11.7	32.39	2.76	10.1	30.5	3.01	18.9	70.58	3.74	40.7	133.46
1992–93	11.7	34.37	2.93	10.3	33.28	3.22	20.0	74.44	3.71	42.1	142.1
1993–94	13.1	35.95	2.73	11.8	34.11	2.90	20.7	79.02	3.81	45.6	149.08
1994–95	16.9	38.89	2.29	14.7	36.59	2.48	29.7	83.59	2.81	61.4	159.06
1995–96	20.0	43.26	2.16	19.5	42.95	2.19	37.0	92.69	2.50	76.6	178.85
1996–97	21.8	48.94	2.24	22.9	50.33	2.19	45.1	104.69	2.32	89.8	203.96
1997–98	22.9	50.58	2.21	24.1	54.42	2.25	48.3	117.52	2.43	95.3	222.52
1998–99	27.9	55.11	1.97	31.8	56.8	1.78	58.6	126.51	2.15	118.3	238.42
1999–00	33.4	61.85	1.85	34.7	64.45	1.85	72.1	147.19	2.04	140.1	273.49
2000–01	37.4	72.15	1.92	36.4	73.08	2.00	71.3	169.63	2.37	145.2	314.86
2001–02	43.5	87.59	2.01	43.7	96.86	2.21	75.8	190.83	2.51	163.0	375.29
2002–03	48.3	98.13	2.02	55.8	113.16	2.02	114.4	238.31	2.08	218.6	449.61
2003–04	79.5	148.05	1.86	73.4	139.74	1.90	165.9	287.86	1.73	318.9	575.65
2004–05	108.3	204.99	1.89	105.5	207.59	1.96	197.4	372.18	1.88	411.2	784.76
2005–06	168.2	297.19	1.76	176.2	292.55	1.66	326.8	527.69	1.61	671.2	1117.43
2006–07	232.5	373.36	1.60	215.9	378.15	1.75	493.4	648.1	1.31	941.7	1399.61
2007–08	253.5	464.57	1.83	232.2	466.31	2.00	481.4	809.56	1.68	967.1	1740.44
2008–09	342.7	601.99	1.75	332.8	597.92	1.79	727.5	993.49	1.36	1,403.0	2193.4
2009–10	426.3	779.52	1.82	443.3	729.16	1.64	730.6	1195	1.63	1,600.2	2703.68
2010–11	460.2	887.12	1.92	574.4	951.2	1.65	854.6	1254	1.46	1,889.1	3092.92
2011–12	897.1	1244.39	1.38	829.2	1263.65	1.52	990.4	1405	1.41	2,717.7	3913.89
Mean	123.3	214.5	2.1	123.9	216.4	2.2	209.8	344.5	2.5	457.1	775.5
Standard Deviation	203.8	317.12	0.388	202.78	321.32	0.499	295.84	419.09	0.914	649.82	1052.33

Source: RBI

TABLE 2: All India Seasonal and Annual Rainfall (in mm)

Year	Annual	Pre-Monsoon	Monsoon	Post-monsoon	Winter
1985	1140	32.8	117.8	830.7	158.8
1986	1126.6	51.5	125.7	803.2	146.3
1987	1078.9	36.1	134.9	749.7	158.2
1988	1351	37.3	159.7	1066.1	87.9
1989	1122.4	28.8	115.9	890.9	86.8
1990	1400.6	59.2	209.7	970.6	161.1
1991	1159.9	40.1	144.6	868.6	106.6
1992	1121.3	38.8	105.1	867.7	109.7
1993	1201.9	41.6	134.9	894.9	130.5
1994	1295.6	50.9	121	998	125.6
1995	1243.6	61.5	140.1	916.2	125.8
1996	1181.8	45.8	118.1	891.6	126.3
1997	1171.4	23.9	122.3	863.2	162
1998	1243.5	53.7	137.2	897.4	155.3
1999	1132	30.4	124.2	837.9	139.5
2000	1050.4	46	127.1	812.5	64.9
2001	1083.3	17.5	128.1	815.4	122.2
2002	920.8	37.8	119.5	689.2	74.3
2003	1174.5	49.3	107.8	897.8	119.5
2004	1071.3	36.9	137	785.7	111.7
2005	1232.5	86.3	131.4	879.9	134.9
2006	1199.4	28.9	145.5	927.8	97.1
2007	1215.6	38.4	118.4	971.6	87.2
2008	1132.1	41.6	115.5	888.7	86.3
2009	959.3	25	97.8	702.4	134.1
2010	1200.5	23.8	127.8	902.4	146.5
2011	1110.1	34	113.9	900.9	61.4
Mean	1160.01	40.66	128.92	871.14	119.27
SD	103.61	14.25	21	83.65	30.28

Data

The unique data set used for empirical analysis was obtained from the RBI handbook of statistics containing credit disbursement and outstanding to marginal, small, medium and large scale farmers by scheduled commercial banks. The credit is directly granted to farmers according to size of land holding with low collateral and only basic documentation. RBI is the central bank of India and regulates the functioning of the entire banking system including schedule commercial, regional rural and cooperative banks. A time series all India data was obtained from RBI from 1985 to 2011 (Table 1). These years represent a gradual increase in the volume of loans disbursed and credit outstanding. The data set was cleaned to correct for any data input errors. Incomplete data was not considered.

Precipitation data was obtained from Indian Metrological Department and data.gov.in, Open Government Data (OGD) Platform. All India seasonal and annual average rainfall data from 1985-2011 was obtained in the same chronology as loan disbursement and outstanding data (Table 2). Rationale for taking rainfall data seasonally like pre-monsoon, monsoon, post-monsoon and winter being was the requirement of rainfall to agricultural crop that varies with respect to the kind of crop grown and the growing season. Further, to get a clearer and comprehensive result a seasonal weather data was taken for the above mentioned period.

Methodology

The study introduces dependent variables as 'credit risk' and specifies 'weather indices' as an independent variable. Average annual rainfall has been regarded as an independent variable. The study attempts to investigate the impact of weather changes on repayment risk of loans granted to farmers. Weather indices and credit risk are operationalized as follows:

Operationalization of Credit Risk and Weather Indices

Repayment risk or repayment performance is determined on the basis of credit risk of loans granted to marginal, small and medium & large scale farmers provided in the data set. Credit risk is defined as whether or not a borrower is able to pay back all the loan installments by the due date. Credit risk was calculated from the data set by taking the ratio of outstanding credit over disbursed credit for each year. The outstanding disbursement ratio gives NPA rate for the three categories of farmers. Higher the credit outstanding, lesser will be the repayment performance. Consequently, greater NPA will be due to lesser lending and higher outstanding of loans to farmers. Therefore, repayment performance forms the dependent variable of the study.

Weather indices or weather changes have been specified into rainfall variations over the period of time. Precipitation data was obtained from Indian Metrological Department contemporaneously to the credit data. Average annual rainfall data has been taken for all the four seasons like pre-monsoon, monsoon, post-monsoon and winter. The study employs accumulated index for obtaining seasonal weather data (Berg and Schmitz, 2008). Accumulated index (I_t^{ac}) is easily applicable to every type of weather variable, and corresponds to the sum of rainfall within a certain accumulated period x in year t where $Ws_{,t}$ indicates rainfall on season s in year t. Rainfall data obtained during a particular season are summarized over the months of accumulated period.

$$I_t^{ac} = \sum_{s=1}^{x} Ws, t$$

Research Model

The present study has time series data available on two variables such as rainfall and disbursement outstanding ratio. A static OLS model relating the two variables is

$$CR_t = \beta_0 + \beta_1 P_t + u_t,\ t = 1, 2, \ldots, n.$$

where ***CR*** is credit risk rate for ***t*** years, P_t is average annual precipitation for t years, β_0 and β_1 are constants designated as parameters of the econometric model, ***u*** is error term or disturbance term. The name static model comes from the fact that we are modeling contemporaneous relationship between dependent and independent variables. Since, the study has taken data of last 27 years from 1985-2011 of outstanding disbursement ratio for marginal, small and medium & large scale farmers and weather indices data for pre-monsoon, monsoon, post-monsoon and winter seasons. Therefore, different static model can be formed for each category as follows:

$$CR_t(\text{Pre-monsoon}) = \beta_0 + \beta_1 P_t + ut,\ t = 1, 2, \ldots, 27.$$

$$CR_t(\text{Monsoon}) = \beta_0 + \beta_1 P_t + u_t,\ t = 1, 2, \ldots, 27.$$

CR_t (Post-monsoon) $= \beta_0 + \beta_1 Pt + u_t$, $t = 1, 2, \ldots, 27$.

$CR_t(Winter) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 P_t + u_t$, $t = 1, 2, \ldots, 27$.

Using these OLS models effect of weather indices on credit risk of four seasons such as pre-monsoon, monsoon, post-monsoon and winter has been established. Such relationships were determined for the three categories of farmers like marginal, small and medium & large scale.

Results

Descriptive Statistics

The descriptive statistics provided in the Table 1 shows mean and standard deviation of the dependent variables for all the loan categories in the data set. The table shows amount of credit disbursed and outstanding by schedule commercial banks towards marginal, small, medium and large scale farmers from 1985-2011. Highest mean (209.8 billion INR) disbursement was observed for medium and large scale farmers and lowest mean (123.5 billion INR) disbursement was observed for marginal scale farmers. Highest mean outstanding (344.5 billion INR) was observed for medium and large scale farmers and lowest mean outstanding (214.5 billion INR) was observed for marginal farmers. High outstanding indicate high NPA rate and low repayment rate. Hence, O/D (outstanding/disbursement) ratio was estimated to determine the repayment performance. It was observed that marginal difference on the O/D ratio exists across the marginal, small and medium & large scale farmers with highest repayment performance for marginal farmers and lowest for medium and large scale farmers. Furthermore, it was observed from the Table that mean total disbursement to all the three categories of farmers being 649.82 billion INR and outstanding being 1052.33 billion INR.

TABLE 3: Results of Linear Static Regression Model of Credit Risk and Weather Indices

Variables	Coefficients			
	Pre-monsoon	*Monsoon*	*Post-monsoon*	*Winter*
CR_{27} (Marginal)	$R^2 = 0.00$	$R^2 = 0.053$	$R^2 = 0.022$	$R^2 = 0.056$
Rainfall	-0.002	0.230	0.149	0.236
CR_{27} (Small)	$R^2 = 0.00$	$R^2 = 0.116$	$R^2 = 0.042$	$R^2 = 0.019$
Rainfall	-0.002	0.340	0.206	0.136
CR_{27} (Medium & Small)	$R^2 = 0.00$	$R^2 = 0.113$	$R^2 = 0.015$	$R^2 = 0.049$
Rainfall	0.011	0.337	0.122	0.222

Table 2 shows mean and standard deviation of independent variable (rainfall) for all the four seasons in the data set from 1985-2011. Highest average rainfall (871.14 mm) for the mentioned period was received in post monsoon season and lowest rainfall (40.6 mm) was received in the pre-monsoon season. Annual average rainfall received for the period was 1160.1 mm. These results are provided in the Table 2.

Static Model Results

The results of linear static regression model are provided in the Table 3. It is clear from the results that 12 linear regression models have been drawn to estimate the effect of weather indices (rainfall) on credit risk of all the three category of farmers. The results depict that rainfall has insignificant effect on credit risk in pre-monsoon season for marginal farmers. The coefficient of determination (R^2) for the above model is 0.00, which means that rainfall has no influence on credit risk. Similar results were observed in monsoon and winter seasons for marginal farmers. It was observed that rainfall has insignificant impact on credit risk for marginal scale farmers in monsoon post-monsoon and winter seasons (β_1= 0.230; β_1 = 0.149; β_1= 0.236 respectively). The coefficient of determination (R^2) for monsoon, post-monsoon and winter seasons is 0.053, 0.022 and 0.056 respectively.

In case of small scale farmers insignificant impact of rainfall on credit risk was observed in pre-monsoon season. R^2 statistics depicts that rainfall does not influence credit risk in pre-monsoon season. In monsoon season insignificant impact of rainfall on credit risk was observed (β_1= 0.340). R^2 depicts that rainfall explain 11.6% of variance in credit risk. Furthermore, insignificant results were observed for rainfall in post-monsoon and winter seasons. The R^2 estimates of the two regression models in post-monsoon and winter seasons are 0.042 and 0.019 respectively. Results in the Table 3 reveal that rainfall has insignificant impact on credit risk for medium and large scale farmers in pre-monsoon season. R^2 statistics reveals that rainfall does not influence credit risk. During monsoon season insignificant impact of rainfall on credit risk was observed. R^2 depicts that rainfall explains 11.3% of variance in credit risk. In post-monsoon and winter season, insignificant effect of rainfall was observed. R^2 statistics reveals that rainfall explains 1.5 and 4.9% of variance in credit risk, which is almost negligible.

Discussions

The present study is an attempt to evaluate the impact of weather changes on credit risk of loans granted to farmers by schedule commercial banks. To enhance generalization and accuracy of results three categories of farmers were identified on the basis of land ownership and were investigated separately. Further, the analysis was carried out seasonally like pre-monsoon, monsoon, post-monsoon and winter. The regression analysis was performed on actual credit and weather data (rainfall) indicating least influence of rainfall on credit risk or repayment performance during these four seasons. The rationale for insignificant impact of rainfall on credit risk being that quantity of rainfall received is limited to agricultural production only (Miranda and Gonzalez-Vega, 2011; Weber and Musshoff, 2012).

Inspite of higher production there is lower credit repayment rate leading to increased credit risk in developing countries like India. The fact is supported by empirical findings that agricultural output will increase in the event of normal rainfall, and will fall in the event of sub-optimal rainfall. But profitability is having inverse relationship with agricultural production, hence balancing the impact of rainfall and stressing the focus on other key factors that influence profitability. Therefore, better profits earned by farmers will increase repayment performance and decrease credit risk. These research findings are supported with the fact that since last two years there has been bumper production of onion in India with no buyers. This excessive supply of onion forced the government to announce a belated decision to procure the onion for Rs 8 per kilogram in Madhya Pradesh. In some other markets the prices of vegetables plummeted as low as Rs 1 to Rs 2 per kg. Thus, excessive production of onion, tomato and potato forced the farmers to sell their crop at throwaway price, bearing heavy loss (Hindustan Times, 2017).

Another key factor that hampers farmers in India from getting better market price is inefficient supply chain management. Farmers can sell their agricultural produce directly to government or can take their produce to nearby government *mandi* (markets) for auction. Central government procures 24 essential food commodities from the farmers through National Agricultural Cooperative Marketing Federation of India Limited (NAFED) and Food Corporation of India (FCI). Central government procures these essential items at higher price than market price and sells them at cheaper price through public distribution systems (PDS). Government has fixed a minimum selling price (MSP) for farmer produce in the event of market price falling below the specified minimum. But in reality farmers are seldom able to sell at MSP because every village does not have NAFED and FCI outlets. About 70% of rice procured by FCI comes from Punjab, Andhra Pradesh, Chhattisgarh and Uttar Pradesh whereas 80% of wheat procurement comes from Punjab, Haryana and Madhya Pradesh. FCI has less number of outlets in the major rice and wheat producing states like Bihar, West Bengal, Assam and Orissa. Furthermore, the procurement dates announced by the governments are time bound therefore making it impossible for farmers to sell their produce at MSP. Marginal and small scale farmers do not have access to cold storage hence cannot store the produce for long time and have sell it to middlemen or traders (Banik and Stevens, 2016).

Agricultural commodities, which are not procured by FCI and NAFED such as vegetables and fruits are taken by farmers to nearby government *mandis*. In India there are about 7700 government designated *mandis*, where commissioned agents and other middlemen purchase from the farmers. To ensure fair price to farmers' produce Agriculture Produce Market Committee (APMC) Act was implemented. Under this act state government official are appointed to look after the activities like auctioning, trading, infrastructure availability and whether the farmer gets the right price. But in reality these middlemen take government

official into confidence and form an alliance with them. At the time of auction, these middlemen offer lower price to the farmers. Moreover, the seller has to bear the additional burden of *mandi* fee which is as high as 13-14% of the produce. Therefore, the profits earned by farmers are less as compared to those earned by middlemen. These minimum earnings influence credit risk and repayment performance of loans provided by schedule commercial banks rather the rainfall received during the period (Banik and Stevens, 2016).

Non-payment of loans before due date is one of the important causes of farmers' suicide in India. This distress is visible in the increase number of farmer suicides. During 1995-2012 National Crime Record Bureau (NCRB) reported 284673 farmer suicides which is 14% of all reported suicide deaths (Mishra, 2014). Moreover, from February, 2016 to February, 2017, 1982 farmers and farm labourers committed suicide which was one-fifth of the total suicides in Madhya Pradesh. NCRB attributed the reasons for suicides being crop failure, sell the produce at unsustainable prices, inability to repay loans, poverty and property disputes (Hindustan Times, 2017).

The results from empirical analysis reveal an unorthodox finding that can act as valid policy input. The research has implications for policy makers that findings of this study will provide insights about the traditional belief regarding relevance of weather changes and credit risk of loans granted to farmers. Policy makers need to think out of box regarding the relationship of weather and credit risk. Nature always tries to maintain a balance and will never coerce farmers to commit suicide because nature is best creator, sustainer and nourisher. The reason for non-payment of credit is man-induced, which vertically slashes the profit earning capacity of farmers.

Conclusion and Suggestions

The study empirically investigated the influence of weather changes (rainfall) on the credit risk (repayment performance) of loans granted to small-scale farmers by schedule commercial banks. The study has employed original lending data of schedule commercial banks, obtained from RBI, and actual rainfall data obtained from Indian Meteorological Department for the period of 27 years from 1985-2011. After analyzing the data, it was found that rainfall has insignificant impact on credit risk during all the four seasons (pre-monsoon, monsoon, post-monsoon and winter). Rainfall actually affects agricultural output, which will increase in the event of normal rains and decrease in the event of sub-optimal rainfall. Rainfall does not influence profitability that ultimately depends on better marketability of agricultural produce. Large production has inverse impact on market price of the agricultural produce. Even though production has increased but farmer has to sell below MSP which reduces profits. Increased production is favourable only if it enhances profitability of farmers. Another factor for low profitability of agricultural produce has been the inefficient supply chain management. Presence of middlemen and commission agents between the producers (farmer) and end consumers reduces the profitability as a farmer has to bear this whole burden. This leads to non-payment of loans before the due date and is one of the key factors of farmers' suicide in India. According to NCRB of the over 3000 farmers who committed suicide in 2015 across India, 2474 have taken loans from the banks or microfinance institutions. These suicide figures indicate that 80% of farmers killed themselves in 2015 because of bankruptcy or debts after taking loan from financial institutions.

Farmers' suicide and repayment of credit is one of the serious issues in India. Nevertheless, lots of political dramas are being played to capture the vote bank. Instead of giving loan waivers to farmers, the problem needs to be addressed at grass-root level. Political parties in power need to rise above party-politics and should bring in concrete reforms in APMC Act that will ensure fair and sustainable prices to agricultural produce.

For timely procurement and better market prices FCI and NAFED should establish their units in leading rice and wheat producing states where farmer suicide rate is high. Further, investments in infrastructure in the form of cold storage warehouse, electricity and sanitation in remote villages should be prioritized rather than giving subsidies and tax evasions. The levels between farmers and end consumers of agriculture produces should be minimized. Emphasis should be given on the direct channel of marketing the agricultural commodities rather through commission agents and middlemen in government designated *mandis*.

References

Altunbas, Y., Gardener, E.P.M., Molyneux, P. and Moore, B. (2001). Efficiency in European Banking. *European Economic Review,* 45, 1931-1955.

Andrews, M. (2006). Microcredit and Agriculture: How to Make it Work, MEDA (Mennonite Economic Development Associates), Waterloo, ON.

Armendariz, B. and Morduch, J. (2000). Microfinance beyond group lending. *Economics of Transition*, 8, 401-420.

Banik, N. and Stevens, P. (2016). Farmer suicides in India and the weather god. ARTNeT Working Paper Series No. 161, Bangkok, ESCAP.

Bardhan, P. and Udry, C. (1999). Development Microeconomics, Oxford University Press, Oxford.

Beck, T., Demirgüc-Kunt, ALaevenL. and V. Maksimovic (2006). The determinants of financing obstacles.*Journal of International Money and Finance*, 25(6), 932-952.

Berg, G. and Schrader, J. (2012). Access to credit, natural disasters, and relationship lending. *Journal of Financial Intermediation*, 21(4), 549-568.

Berger, A. and Udell, G. (1995). Relationship lending and lines of credit in small firm finance.*Journal of Business*, 68(3), 351-381.

Berger, A., Hasan, I. and Klapper, L. (2004). Further evidence on the link between finance and growth: an international analysis of community banking and economic performance.*Journal of Financial Services Research*, 25(2-3), 169-202.

Berger, A., Klapper, L., Miller, M. and Udell, G. (2003). Relationship lending in the Argentinesmall business credit market.Credit Reporting Systems and the International Economy,MIT Press, Cambridge, MA, pp. 255-270.

Binswanger, H.P. and Rosenzweig, M.R. (1986). Behavioral and material determinants ofproduction relations in agriculture.*Journal of Development Studies*, 22(3),503-539.

Cabannes, Y. (2012).Financing urban agriculture.*Environment & Urbanization*, 24(2),665-683.

Christen, R.P. and Anderson, J. (2013). Segmentation of smallholder households: meeting the rangeof financial needs in agricultural families, available at: www.cgap.org/sites/default/files/ Focus-Note-Segmentation-of-Smallholder-Households-April-2013.pdf (accessed May 24, 2017).

Collier, B., Katchova, A.L. and Skees, J.R. (2011). Loan portfolio performance and El Niño, an intervention analysis.*Agricultural Finance Review*, 71(1), 98-119.

Czura, K. and Klonner, S. (2010). The Tsunami and the chit fund – evidence from the Indian Ocean Tsunami hit on credit demand in South India, paper for the German Development Economics Conference, Hanover, June, 18-19.

De Janvry, A. (2010). Agriculture for development: new paradigm and options for success.*Agricultural Economics*, 41(s1), 17-36.

Diagne, A., Zeller, M. and Sharma, M. (2000). Empirical measurements of households' access tocredit and credit constraints in developing countries: methodological issues and evidence. Discussion Paper No. 90, FCND, International Food Policy Research Institute,Washington, DC.

Field, E., Pande, R., Papp, J. and Rigol, N. (2011). Debt structure, entrepreneurship, and risk: evidence from microfinance, Harvard working paper, Yale, available at: http://scholar.harvard.edu/files/field/files/repayment_default_dec19-1.pdf (accessed May 18, 2017).

Foltz, J. D. (2004). Credit market access and profitability in Tunisian agriculture.*Agricultural Economics*, 30(3), 229-240.

Food and Agriculture Organization (2002). Reducing poverty and hunger: the critical role offinancing for food, agriculture and rural development, Paper for the International Conference on Financing for Development, FAO, Monterrey, Mexico, March, 18-22.

Giné, X. and Yang, D. (2009). Insurance, credit, and technology adoption: field experimental evidence from Malawi, available at: http://siteresources.worldbank.org/DEC/Resources/ Insurance_Credit_Malawi.pdf (accessed May 21, 2017).

Godquin, M. (2004). Microfinance repayment performance in Bangladesh: how to improve the allocation of loans by MFIs. *World Development*, 32(11), 1909-1926.

Hartarska, V. and Mersland, R. (2012). Which governance mechanisms promote efficiency in reaching poor clients? Evidence from rated microfinance institutions. *European Financial Management*, 18(2), 218-239.

Hartarska, V. and Nadolnyak, D. (2008). An impact analysis of microfinance in Bosnia and Herzegovinia.*World Development*, 36(12), 2605-2619.

Hartarska, V.M. and Holtmann, M. (2006). An overview of recent developments in the microfinance literature. *Agricultural Finance Review*, 66(2), 147-165.

Hindustan Times (2017). Apathy, heavy losses: Why farmers in Madhya Pradesh are on warpath against Shivraj Singh Chouhan Govt., July 07, 2017.

Krahnen, J.P. and Schmidt, R.H. (1994). Development Finance as Institution Building: A New Approach to Poverty-oriented Banking, Westview Press, Boulder, CO.

Maes, J.P. and Reed, L.R. (2012). State of the microcredit summit – Campaign Report 2012, Microcredit Summit Campaign (MCS), Washington, DC.

Micrometer (2017),http://mfinindia.org/2017/05, Issue 21.

Miranda, M.J. and Gonzalez-Vega, C. (2011). Systemic risk, index insurance, and optimal management of agricultural loan portfolios in developing countries. *American Journal of Agricultural Economics*, 93(2), 399-406.

Mishra, S. (2014). Farmers' Suicides in India, 1995-2012: Measurement and Interpretation, Asia Research Centre, London School of Economics & Political Science, Working Paper No. 62, London, UK.

Morduch, J. (1999). The role of subsidies in microfinance: evidence from the Grameen bank. *Journal of Development Economics*, 60(1), 229-248.

Pelka, N., Musshoff, O. and Weber, R. (2015). Does weather matter? How rainfall affects credit risk in agricultural microfinance. *Agricultural Finance Review*, 75(2), 194-212.

Pastor, J. (2002). Credit risk and efficiency in the European banking system. *Applied Financial Economics*, 12(12), 895-911.

Raghunathan, U.K., Escalante, C.L., Dorfman, J.H., Ames, G.C.W. and Houston, J.E. (2011). The effect of agriculture on repayment efficiency: a look at MFI borrowing groups. *Agricultural Economics*, 42(4), 465-74.

Simtowe, F., Diagne, A. and Zeller, M. (2008). Who is credit constrained? Evidence from rural Malawi.*Agricultural Finance Review*, 68(2), 255-272.

Skees, J.R. and Barnett, B.J. (2006). Enhancing microfinance using index-based risk-transfer products. *Agricultural Finance Review*, 66(2), 235-250.

Stiglitz, J.E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *American Economic Review*, 71(3), 393-410.

United Nations Development Program (2013): Human Development Report 2013 – The rise of the south: human progress in a diverse world, March 15, New York, NY.

Vedenov, D.V. and Barnett, B.J. (2004). Efficiency of weather derivatives as primary crop insurance instruments. *Journal of Agricultural Resource Economics*, 29(3), 387-403.

Weber, R. and Musshoff, O. (2012). Is agricultural microcredit really more risky? Evidence from Tanzania. *Agricultural Finance Review*, 72(3), 416-435.

Weber, R. and Musshoff, O. (2013). Can flexible microfinance loans improve credit access for farmers? *Agricultural Finance Review,* 73(2), 255-271.

World Bank (2015). World Bank Annual Report 2015, www.worldbank.org/en/about/annual-report. Acessed on: 15/05/2016.

"Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom."

– Clifford Stoll

Annual Index of Productivity Vol. 59 (April 2018– March 2019)

Vol. 59, April-June, 2018, No. 1 (Focus: Sustainable Agriculture)

Vol. 59, July-September, 2018, No. 2 (Focus: Workforce)

Guidelines for Contributors

PRODUCTIVITY is the principal journal of the National Productivity Council of India. The Journal aims at disseminating information on concepts of and data on productivity and its growth in India and elsewhere. It also aims at disseminating knowledge on techniques and methods of productivity improvement through effective management of all types of resources. Thus contributions from a large spectrum of disciplines are accepted for publication. Only those manuscripts that present the results of Journal. The managerial/policy implications of the study should be highlighted separately towards the end of the paper.

Manuscript Submission

The text should be addressed to

The Editor, PRODUCTIVITY, National Productivity Council,
Utpadakata Bhawan, Lodi Road, New Delhi - 110003.
Email : npcres@rediffmail.com

Ethical Statement

The cover letter should include a written statement from the author(s) that:

1. The manuscript is an original research work and has not been published elsewhere including open access at the internet.
2. The data used in the research has not been manipulated, fabricated, or in any other way misrepresented to support the conclusions.
3. No part of the text of the manuscript has been plagiarised.
4. The manuscript is not under consideration for publication elsewhere.
5. The manuscript will not be submitted elsewhere for review while it is still under consideration for publication in the Productivity.

The cover letter should also include an ethical statement disclosing any conflict of interest that may directly or indirectly impart bias to the research work. Conflict of interest most commonly arises from the source of funding, and therefore, the name(s) of funding agency must be mentioned in the cover letter. In case of no conflict of interest, please include the statement that "the authors declare that they have no conflict of interest".

General Guidelines

- The manuscript should be of about 5,000 words length.
- Tables, illustrations, charts, figures, exhibits, etc., should be serially numbered and typed in separate pages and should not be mixed with the main text.
- The Text should contain in the following order: an Abstract; Main Text of the Article; References; and Appendices (if appropriate).

Reference Style Guidelines

- Only those references which are actually utilized in the text should be included in the reference list.
- In the text, references should be cited with the surname of the author(s) alongwith the year of publication and the page number, all in brackets.
- If there are more than one reference by the same author during any year, the year may be subscripted with 'a' or 'b'. For instance, reference may be given at the end of the sentence as: (Szendrovits, 1998a, p. 337).
- Reference list should be alphabetically arranged. Each reference should carry the surname of the author, followed by other names, the title of the paper in quotes, the name of the journal underlined, volume and issue numbers, and the year of publication.
- In the event of a book, the title should be followed by the publisher's name and year of publication.
- In the event of a report from an organization, the name of the organization may be cited in the place of the author.

Table Guidelines

- Tables should be numbered e.g., Table 1, consecutively and titled.
- Sources of data need to be given below each table unless otherwise mentioned in the text.
- Each table should contain a short caption explaining the content of the table.
- All tables column should have an explanatory heading.

Figure and Artwork Guidelines

- Figures, Graphs, and Maps should be numbered using Arabic numerals.
- Each figure should contain a short caption explaining the content of the figure.
- At the end of the figure, the source of the figure must be mentioned clearly.

Accompanying Material

The manuscripts should be accompanied by:

- An abstract of the paper not exceeding 100 words.
- A declaration that the paper is original and has not been submitted elsewhere for publication.
- A note about the author(s) not exceeding 50 words.

Copyright Transfer

Once the manuscript is accepted for publication, the corresponding author will receive an E-mail informing about the acceptance of the article. The publication of an article in the "Productivity" means that the author(s) transfer the Copyright of the article to the Journal.

Registered with Registrar of Newspapers REGD. NO. 4757/59

Productivity

Announces

a

Special Issue

Vol. 60, No. 1,

April - June, 2019

on

Entrepreneurship

Place your orders with :

Prints Publications Pvt Ltd
"Prints House", 11, Darya Ganj
New Delhi-110 002
Telephone : 91-11-45355555
E-mail : contact@printspublications.com
www.printspublications.com