Volume 7 Issue 1

ISSN: 2321-1857

International Journal of Business Analytics & Intelligence

April 2019

A Publication of Publishing India Group

International Journal of Business Analytics & Intelligence

Editorial Message

Dear Readers,

It's an immense pleasure to reach out to you all in the era of automated analytics!

In its seventh year with a steadily increasing readership, IJBAI is proud to have published works of eminent academicians as well as data science professionals. We strive continuously to bring the emerging aspects of Data Science practices and Artificial Intelligence. In the current issue the focus is on enriched contents that would contribute to the future of analytics.

In his regular column Analytically Yours, Prof. Arnab Kr. Laha has discussed the intriguing case of employing different statistical tests on a single set of data for hypothesis testing. The column from Favio Vázquez, would not only guide the readers into the fascinating relationships of ontology and graphs with data science, but also introduce them to some of the key trends in the world of data science.

Data mining approach is an invaluable tool to analyze large body of data to explore previously unknown patterns in virtually any establishment including industry and academic. In this issue, practitioners from the industry and academics have presented their research on use of data mining in diverse scenarios. Hamid Ravan Paknoodezh has compared the use of data mining approach in environmental accounting of Public and Private Sector Companies in India. Shome & Khatri, in their study set in an academic environment, have reported the statistical analysis of examination and evaluation of results. Digitization is the new way of monetization and information mining. Through an empirical approach, Deshpande has studied whether digitization and customized ERP applications in the manufacturing process have affected the overall efficiency of an engineering firm.

In the era of big data, the economic research, too, is experiencing a transformative shift in several dimensions. Data mining is enabling the researchers to make use of the vast data of economic effect across a range of topics. Some of the papers published in this issue of IJBAI are focused on various application of data science in the economic research. Jagadeeshwaran & Basuvaraj have examined the impact of non-performing assets on operational performance of foreign banks of India. While the paper by Roy & Bhattacharyya has explored the use of data mining approach to identify homogeneous clusters of Indian companies in the small-cap segment, Bala & Gupta have studied tests the existence of long-term memory with reference to structural changes/ breaks in Indian Stock Market. In a study involving behavioural finance, Ms. Sharma has employed structural equation model to investigate the impact of behavioural disposition on decisions making process of individual investors.

We would like to thank the researchers and data scientists who have honored us by choosing our young journal to publish some of their research. We are indebted to the esteemed reviewers, whose excellent and anonymous work have made the publication of this issue possible. We hope our readers would find the articles interesting and enriching. It would be great to have valuable feedback from our learned readers about the current version of IJBAI.
Sincerely yours,

Madhumita Ghosh
Joint Editor-in-Chief
&
Dr. Tuhin Chattopadhyay
Editor-in-Chief

Dated: 30th June 2019

International Journal of Business Analytics and Intelligence

Volume 7 Issue 1 April 2019

ISSN: 2321-1857

Columns

Research Papers

Journal is also available online at www.publishingindia.com

International Journal of Business Analytics and Intelligence
7 (1) 2019, 01-02
http://publishingindia.com/ijbai/

Analytically Yours:

One Data, Many Tests

Arnab Kumar Laha*

In many areas of research in management, social science, medical science, genomics, business studies and psychology it is a common practice for researchers to test their theoretical understanding of a phenomenon through formulation of appropriate hypotheses which can be proved or disproved on the basis of data. These researchers argue that if the data provides support to the formulated hypotheses then it can be concluded that the theory based on which these hypotheses were derived is also empirically validated. Typically a piece of research may depend on testing more than one hypothesis and the empirical validation of the theory requires all these hypotheses being supported.

The subject of statistical test of a hypothesis has a long history going back to the work of John Arbuthnot more than three centuries ago. Formally Sir R. A. Fisher introduced what is now termed as the *P-value approach* to hypothesis testing in 1925. The main idea of this approach is as follows: Let H be a hypothesis that the researcher wants to test. Towards this goal she decides on a test statistic T and obtains its probability distribution assuming the hypothesis H to be true. Then she collects the required data and computes the value of T for this dataset. Suppose the obtained value of T is τ. The P-value of the test - which is the probability of obtaining a value of T more extreme than τ when H is true - is next obtained using the probability distribution of T computed earlier. If the P-value turns out to be "small" then it is concluded that the hypothesis H is false and is rejected.

J. Neyman and E.S. Pearson gave a different formulation of testing a hypothesis which is sometimes called the *fixed-α approach*. In this formulation the researcher has to specify two hypotheses H_0 and H_1 which are called the *null hypothesis* and *alternative hypothesis* respectively. H_0 here plays the same role as H in the Fisher's formulation with the additional understanding that if H_0 is rejected then H_1 is assumed to hold true. The researcher now decides on a suitable test statistic T and obtains its probability distribution assuming H_0 is true. She also decides on a value α and determines a "rejection region" (R) such that $P(T \in R | H_0 \text{ is true}) = \alpha$. She then collects the data D and computes the value of T say τ. If $t \in R$ then H_0 is rejected. α is called the level of significance of the test. Also, note that $P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$.

Now suppose there are m hypotheses H_i, $t = 1, \ldots, m$ that are to be tested on the same dataset. Of these m hypotheses suppose m_0 are actually true and the remaining $m_1 - m - m_0$ are actually false. Without loss of generality suppose that the hypotheses $H_1, \ldots, H_{m_0}$ are actually true. Each of these hypotheses are tested separately and the p-values obtained are $q_i < \alpha$. In the fixed-α approach any hypothesis for which $q_i < \alpha$ is rejected. The problem with this approach in the multiple testing context is that chance of rejecting at least one of the m_0 true hypotheses is larger than α. Hence we need to use a different strategy to ensure that the overall probability of rejecting a true hypothesis is not larger than α.

The following Table 1 summarizes the situation after the m-tests are carried out. Note that only N, D and m are observed.

Table 1

	H_0 retained	*H_0 rejected*	
H_0 True	True Negative (TN)	False Discovery (FD)	m_0
H_0 False	False Negative (FN)	True Discovery (TD)	m_1
	N = TN+FN	D = FD+TD	m

The simplest strategy is to carry out the individual tests of hypothesis at a much lesser level of significance $\frac{\alpha}{m}$. This is known as the Bonferroni method. If the overall probability of rejecting a true hypothesis is to be restricted

* Indian Institute of Management Ahmedabad, Gujarat, India. Email: arnab@iima.ac.in

to 5% i.e. α = 0.05 and the study involves testing 10 hypotheses then each hypothesis is tested at 0.5% level of significance. In other words in Bonferroni method only those hypotheses for which q_i < 0.005 are rejected. To understand why this approach works one needs to recall the Boole's inequality which states that for any set of events $A_1, \ldots, A_k$, $P(\bigcup_{i=1}^{m_0} E_i) \leq \sum_{i=1}^{k} P(A_k)$. Let E_i be the event that $q_i \leq \frac{\alpha}{m}$. Then $P(FD > 0) = P(a\ least\ one\ true\ hypothesis\ is\ rejected)$

$$= P\bigcup_{i=1}^{m_0} E_i \leq \sum_{i=1}^{m_0} P(E_i) = m_0 \frac{\alpha}{m} \leq \alpha$$

The Bonferroni procedure is extremely simple but it suffers from a major disadvantage that it has a very high rate of false negatives.

We examine this further using a simulated example. We generate 50 datasets each of size 10 from a normal distribution with mean(μ) = 0 and standard deviation(σ) = 1 and another 50 datasets each of size 10 from a normal distribution with μ=1 and σ=1. We want to test H_0 : $\mu = 0$ against H_1 : $\mu \neq 0$ by applying the t-test to each of the datasets maintaining P(FD > 0) = α = 0.05. The Bonferroni method requires us to carry out the individual tests at $\frac{0.05}{100}$ = 0.0005 level of significance. When this method is carried out we get the following:

Table 2

	H_0 retained	*H_0 rejected*	
H_0 True	49	1	50
H_0 False	47	3	50
	96	4	100

From Table 2 we find that only 3 of the 50 false hypotheses are rejected by the Bonferroni procedure substantiating the conservativeness of the procedure.

Benjamini and Hochberg (1995) introduced the idea of controlling the expected proportion of falsely rejected hypotheses which they call the False Discovery Rate (FDR). They developed a procedure now widely known as the Benjamini-Hochberg (BH) procedure the controls the FDR = $E\left(\frac{FD}{D}\right)$. The BH-procedure ensures that FDR ≤ α. The mechanics of the BH-procedure is given below:

- We first sort the p-values of the m individual tests in ascending order.
- Next, we assign ranks corresponding to the p-values with the smallest p-value having a rank of 1, the second smallest p-value having a rank of 2 etc.
- Calculate the Benjamini-Hochberg (BH) critical value corresponding to each individual p-value, using the formula $\frac{1}{m}Q$, where i = the individual p-value's rank, m = total number of tests, Q = the false discovery rate.
- The original p-values are now compared with the BH critical values obtained in Step 3. Let p^* be the largest p value that is smaller than the BH critical value.
- All hypotheses having p-values less than p^* are rejected.

Let us now examine what happens when the BH-method with FDR control of 5% is applied to the p-values obtained in the simulated example discussed above. Table 3 below provides a summary. We find that there has been a substantial increase in the number false hypotheses that are rejected by BH method compared to that rejected by the Bonferroni method while the number of true hypothesis that are rejected by both the methods remain the same. Thus the BH-method successfully overcomes a critical shortcoming of the Bonferroni method.

Table 3

	H_0 retained	*H_0 rejected*	
H_0 True	49	1	50
H_0 False	15	35	50
	64	36	100

Reference

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 03-07
http://publishingindia.com/ijbai/

Ontology, Graphs and Data Science

Favio Vázquez*

Abstract

The present and future of data science depends on us being alert of the trends and advances in the scientific and technological community. In this article the fundamentals of ontology and graphs are introduce with their relations with data science, also presenting the concepts of knowledge-graphs and data fabric.

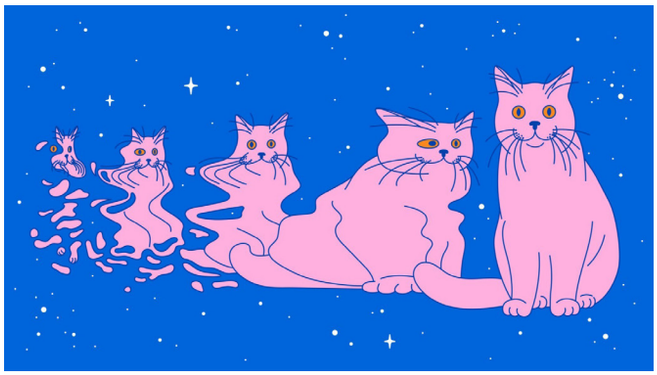
Ontology

If you are new to the word ontology don't worry, I'm going to give a primer on what it is, and then why it matters for the data world. I'll be explicit in the difference between philosophical ontology and the ontology related to information and data in computer science.

Ontology (the philosophical part)

In simple words, one can say that ontology is the study of what there is. But there is another part to that definition that will help us in the following sections, and that is ontology is usually also taken to encompass problems about the most general features and relations of the entities which do exist.

Ontology open new doors for what there is too. Let me give you an example.



Quantum mechanics opened a new view of reality and what "exists" in nature. For some physicists in the 1900s there was simply no reality expressed in the quantum formalism. At the other extreme, there were many quantum physicists who took the diametrically opposite view: that the unitarily evolving quantum state completely describes actual reality, with the alarming implication that practically all quantum alternatives must always continue to coexist (in superposition). And thus opening the whole world to a new view and understanding of the "things" that "exists" in nature.

But let's come back to the relation of entities part of the definition. Sometimes when we talk about entities and their relation, ontology is referred as formal ontology. These are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities. Such theories usually propose axioms about these entities in question, spelled out in some formal language based on some system of formal logic.

And this will allow us to do a quantum jump to next part of the article.

Ontology (the information and computational part)

If we bring back the definition of formal ontology from above, and then we think of data and information, it's possible to set up a framework to study data and its relation to other data. In this framework we represent information in an especially useful way. Information represented in a particular formal ontology can be more easily accessible to automated information processing, and how best to do this is an active area of research in computer science like data science. The use of the formal ontology here is representational. It is a framework to represent information, and as such it can be representationally successful whether or not the formal theory used in fact truly describes a domain of entities.

* Data Science Course Instructor, Business Science University, Pennsylvania Area.

Now it's a good moment to see how ontology can help us in the data science world.

An important concept we need to introduce right now is the one of linked data. The goal of linked data is to publish structured data in such a way that it can be easily consumed and combined with other Linked Data. That allow us to talk about the concept of the knowledge graph which consists in integrated collections of data and information that also contains huge numbers of links between different data.

Ontology is important here because it's the way we can connect entities and understand their relationships. With ontology one can enable such a description, but first we need to formally specify components such as individuals (instances of objects), classes, attributes and relations as well as restrictions, rules and axioms.

Databases Modeling and Ontologies

Currently, most of the technologies that employ data modeling languages (like SQL) are designed using a rigid "Build the Model, then Use the Model" mindset.

For example, suppose you want to change a property in a relational database. You had previously thought that the property was single-valued, but now it needs to be multi-valued. For almost all modern relational databases, this change would require you to delete the entire column for that property and then create an entirely new table that holds all of those property values plus a foreign key reference.

This is not only a lot of work, but it will also invalidate any índices that deal with the original table. It will also invalidate any related queries that your users have written. In short, making that one change can be very difficult and complicated. Often, such changes are so troublesome that they are simply never made.

By contrast, all data modeling statements (along with everything else) in ontological languages for data are incremental, by their very nature. Enhancing or modifying a data model after the fact can be easily accomplished by modifying the concept.

Ontological languages, linked data and all of that exist in the realm of graph databases. So it's time to discuss some ideas and concepts of graph databases, what they are, what are their advantages and how they can help us in our daily tasks.

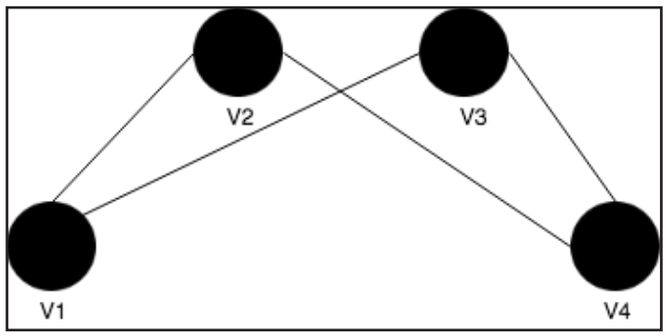
What is a Graph?

I'm going to give two definitions of a graph. First the mathematical one, and then a more simplistic one.

A graph G is a finite, non-empty set V together with a (possibly empty) set E (disjoint from V) of two-element subsets of (distinct) elements of V. Each element of V is referred to as a vertex and V itself as the vertex set of G; the members of the edge set E are called edges. By an element of a graph we shall mean a vertex or an edge.

One of the most appealing features of graph theory lies in the geometric or pictorial aspect of the subject. Given a graph, it is often useful to express it diagrammatically, where by each element of the set is represented by a point in the plane and each edge by a line segment.

It is convenient to refer to such a diagram of G as G itself, since the sets V and E are easily discernible. In the figure bellow, a graph G is shown with vertex set V = {V1, V2, V3, V4} and edge set E = {V1V2, V1V3, V2V4, V3V4}



As you can see the set V contains the number of vertex or points in the graph and E the relationships between them (read V1V2 like V1 is connected to V2).

So in simple words, a graph is a mathematical representation of objects (or entities or nodes) and their relationships (or edges). Each one of those points can represent different things depending on what you want. By the way, here nodes and vertices mean the same, we'll use them interchangeably.

What is a Database?

A database (DB), in the most general sense, is an organized collection of data. More specifically, a database is an electronic system that allows data to be easily accessed, manipulated and updated.

In other words, a database is used by an organization as a method of storing, managing and retrieving information. Modern databases are managed using a database management system (DBMS).

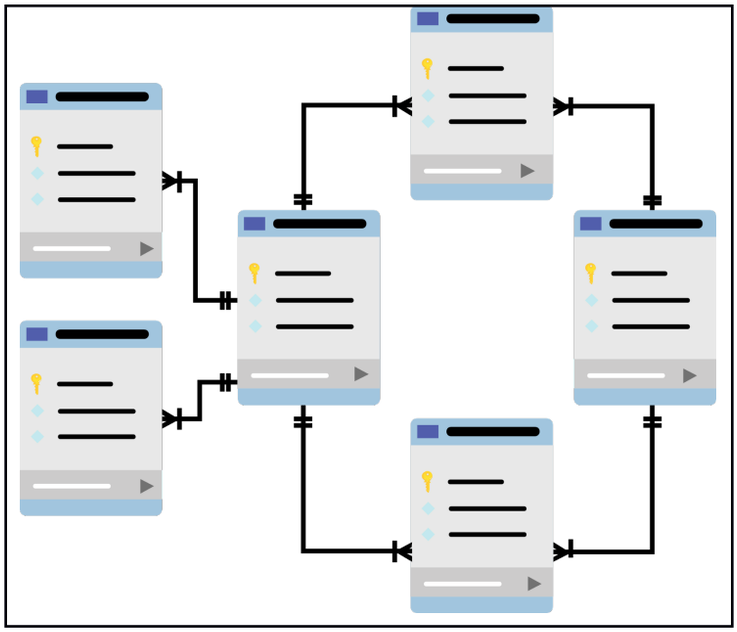
Do you want to know the truth? From my experience most databases are:

- Not organized
- Not easily accessed
- Not easily manipulated
- Not easily updated

When we talk about doing data science, in older years it was easier to maintain a database because the data was simple, smaller and slower. Nowadays we can save almost whatever we want in a "database", and that definition I think is stuck with another concept, the relational database.

In a relational database we have a set of "formally" described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Basically we have schemas in where we can store different tables, and inside of those tables we have a set of columns and rows, and inside of an specific position (row and column) we have an observation.

We also have a relationship between those tables. But they're not the most important thing, the data they contain is the most important thing. Normally they are pictured like this:



What is a Graph Database?

Based upon the concept of a mathematical graph, a graph database contains a collection of nodes and edges. A node represents an object, and an edge represents the connection or relationship between two objects. Each node in a graph database is identified by a unique identifier that expresses key value pairs. Additionally, each edge is defined by a unique identifier that details a starting or ending node, along with a set of properties.

A graph database stores the same sort of data, but is also able to store linkages between the things. I don't have to run JOINs to understand how I should market to each individual customer. I can see the relationships in the data without having to make a hypothesis and test it.

Whereas relational databases store highly-structured data in tables with predetermined columns and rows, graph databases can map multiple types of relational and complex data. Thus, graph databases are not rigid in their

organization and structure, as relational databases are. All relationships are natively stored within the vertices of the edges, meaning that the vertices and edges can each have properties associated with them. This structure allows for a database that can depict complex relationships between unrelated data sets.

Did you know that 2018 was touted as "The Year of the Graph"?, as more and more organizations both large and small have recently begun to invest in graph database technology. So we aren't on a crazy path here.

I'm not saying that everything we know from relational databases, and SQL will not work anymore. I'm saying that there are some cases (surprisingly a lot of them) where you are better using a graph database than a relational database.

I'm going to give you right now an idea on when you should be using a graph database instead of something else:

- You have highly related data.
- You need a flexible schema.
- You want to have a structure and build queries that are more similar to way people think.

Instead if you have a highly structured data, you want to do a lot of grouping calculations and you don't have that many relationships between your tables, then you may be better with a relational database.

A graph database has another, not obvious advantage. It allows you to build a knowledge-graph. Because they are graphs, knowledge-graphs are more intuitive. People don't think in tables, but they do immediately understand graphs. When you draw the structure of a knowledge graph on a whiteboard, it is obvious what it means to most people.

And then you can start thinking on building a data fabric, which then can allow you to re-think the way you do machine learning and data science as a whole. But that's material for a next article.

From RDBS to the Knowledge Graph and the Data Fabric

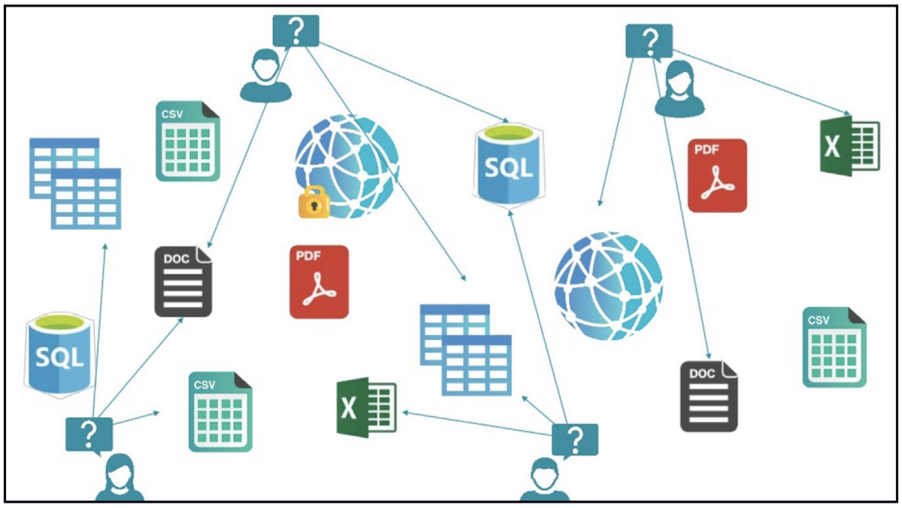
The Data Fabric is the platform that supports all the data in the company. How it's managed, described, combined and universally accessed. This platform is formed from an Enterprise Knowledge Graph to create an uniform and unified data environment.

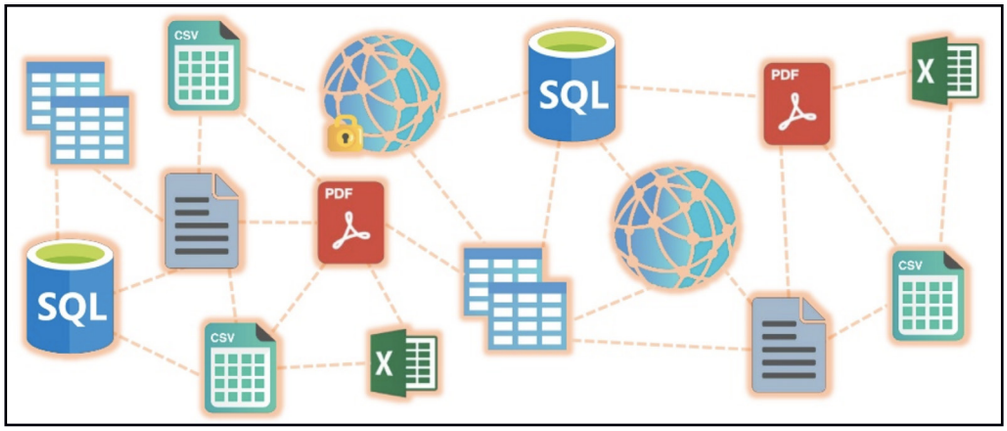
The formation of this data fabric first need to create ontologies between the data you have. This transition can also be thought of as going from traditional databases to graph databases + semantics.

There are three reasons why graph databases are useful:

- Graph Database offerings are showing maturity in capability and diversity
- Graph is being used beyond classical graph problems
- Digital Transformation of complex data requires graph

So if you tie this benefits with a semantic layer, built on ontologies, you can go from having your data like this:





Where you have a human-readable representation of data that uniquely identifies and connects data with common business terms. This layer helps end users access data autonomously, securely and confidently.

And then you can create complex models for discovering patterns in your different databases.

There are lots of advances in automation regarding machine learning, deep learning and deployment. But data is an important asset (maybe the most important one) for companies right now. So before you can apply machine learning or deep learning, at all, you need to have it, know what you have, understand it, govern it, clean it, analyze it, standardize it (maybe more) and then you can think of using it.

We need automation for data storage, data munging, data exploration, data cleansing and all the things we actually spend a lot time doing. That's why my bet it's that semantic technologies are the way to go here. With them you have automatic query generation and using them against the complex graph makes extracting features easy and eventually fully automated.

Conclusions

Graph databases provide an excellent infrastructure to link diverse data. With easy expression of entities and relationships between data, graph databases make it easier for programmers, users and machines to understand the data and find insights. This deeper level of understanding is vital for successful machine learning initiatives, where context-based machine learning is becoming important for feature engineering, machine-based reasoning and inferencing.

For me the key trends for 2019 and 2020 are:

- AutoX: We will see more companies developing and including into their stack technologies and libraries for automatic Machine and Deep Learning. The X here means that this auto-tools will be extended to data ingestion, data integration, data cleansing, exploration and deployment. Automation is here to stay.
- Semantic technologies: On the most interesting discoveries for me this year was the connection between DS and semantics. It's not a new field in the data-world but I see more people getting an interest in the field of semantics, ontologies, knowledge-graphs and its connection to DS and ML.
- Programming less: This is a hard thing to say, but with automation in almost every step of the DS process we will program less and less everyday. We will have tools for creating code and that will understand what we want with NLP and then transform that into queries, sentences and full programs. I think [programming] it's still a very important thing to learn, but it will be more easy soon.

This is one of the reasons why I'm creating this article, trying to follow what's happening across the industry, and you should be aware of this. We will program less, and will use semantics technologies more in the near future. It's closer to the way we think. I mean do you think in relational databases? I'm not saying we think in graphs, but it's much easier to pass information between our heads and a knowledge graph than creating weird database models.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 08-23
http://publishingindia.com/ijbai/

The Use of Data Mining Techniques in Environmental Accounting: A Comparison of Public and Private Sector Companies in India

Hamid Ravan Paknoodezh*

Abstract

Environmental accounting as an essential tool for understanding the role played by the natural environment in the economy is useful for business decisions, especially for proactive environmental management activities. Since recent technological advancement paved the way for the use of big data to assist companies in the decision-making process and one of the best methods to the exploited large dataset is data mining, this study aims to examine the level of data mining techniques in environmental accounting within public and private sector companies in India. This study covering the States of Haryana for North, West Bengal for East, Maharashtra for West and Kerala as representative of South. With the use of structured questionnaires in the soft and hard copy at last 100 managers and 243 accountants randomly were selected. Six hypothesis has been considered that were evaluated in the statements with a five-point Likert scale. In this study, we used the classification and regression techniques as appropriate tools for data mining. The results of this study confirmed the null hypothesis 'there is no significant difference existed in the level of data mining between the public and private sector industry in India' rejects. Therefore, the levels of data mining within environmental accounting in terms of most aspects significantly is higher in the private sector than public sector companies in India.

Keywords: Environmental Accounting, Environmental Management, Data Mining, Industry in India

Introduction

Background of the Study

Today environmental accounting perspective has attracted much attention to reducing or avoiding those costs of the corporate product while at the same time improving environmental quality Yu (2007). Modern accounting is not only concerned with record-keeping and reporting of information to the investors but also aims at fulfilling the information needs of a wide range of internal and external shareholders. It is now considered as a service activity that is estimating and accounting the costs of environmental impacts and rapidly developing an area of management, accounting, and finance. Environmental accounting is a function to provide quantitative information primarily of financial nature, about economic activities that are intended to be useful in making an economic decision Alshhadat (2018). Due to growing public concern about the alarming impact of industrial activities on nature, companies are under force from both the government and the society to reduce adverse impacts of their activities on the environment Chinchuluun (2010). The performance of an institute is now being judged not only by its financial results but also with regard to its contribution to protect and improve the environment Common (1996). In environmental accounting, check the quality of the large size of the estimated data matrix is a difficult task. But the data mining system at the same time is capable of dealing with the temporal and spatial data simultaneously that have unique characteristics for determining. Environmental accounting with the use of proper methodologies like data mining able to correctly measure the environmental impact thus in order to report the environmental cost of the activity of an organization Gilberto (2008). Data mining is defined as the process of sorting through large data sets to recognize patterns and establish relationships to solve problems through data analysis Peter (2013). Data mining tools allow enterprises to predict future trends. In many research areas, data mining techniques are used including mathematics, cybernetics, genetics and

* Department of Commerce, Kerala university, Kerala, India. Email: Hamid.Ravanpakwin7@gmail.com

marketing Shah (2014). While data mining techniques are a means to drive efficiencies and predict customer behaviour, if used correctly, a business can set itself apart from its competition through the use of predictive analysis. In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact businesses Rouse (2017). The growth and development of business and increasing of competition are the cause of increasing complexity in business. Also, managers and decision-makers need to predominate on the details of the business, and therefore in this regard data mining became necessary. The criteria for evaluating performance and usefulness of the Environmental Accounting System are the ability of the system to provide sufficient and necessary information from data mining. After all, the objectives of this research assess the level of data mining within an Environmental Accounting System of public and private sector companies in India.

Literature Review

Over the past years, there has been a dramatic increase in the number of studies which have focused on data mining in environmental accounting. Seifert (2004) examined that data mining is a useful tool at the disposal of all companies. Data mining involves the use of sophisticated analytical tools to discover valid patterns and relationships within the massive dataset Seifert (2004). In other words, data mining is an operation beyond the collection and management of data. This operation also involves analysis as well as forecasting. Pechenizkiy (2005) have argued that some of the experts believe that concepts of data mining techniques are a stage of the comprehensive process, which is called knowledge discovery within the database. The other steps in this extensive process involve data cleaning, data integration, data selection, transformation, and pattern evaluation. Data mining technology makes the connection between many technical areas including databases statistics, machine learning, and human-computer interface Pechenizkiy (2010). In the business world, financial data are considered as a strategic asset. Financial data are collected and stored by institutions such as banks, stock exchange, tax agencies, specialized databases related to auditors, accountants, etc. Data mining methods in financial information can solve the problems of classification and prediction used to facilitate the decision-making process. The importance of data mining in finance and accounting can have a wide range that includes fraud detection in increasing the profitability of the business unit. Kirkos (2007) mentioned that many professional organizations recognize the importance of data mining. American Institute of Certified Public Accountants (AICPA), introduced data mining as one of the top ten technologies of the future. The American Association of Internal Auditors also included data mining in the list of their research priorities. The possible application of data mining within environmental accounting information leads to the identification of the level of utilization of data mining in market segmentation, customer churn, fraud detection, market basket analysis, interactive marketing and trend analysis Kirkos (2007). In terms of data mining application areas refer to surveys such as those conducted by Sirikulvadhana in (2002) have shown that classification and prediction capabilities of data mining techniques are used to facilitate the audit process, also help assess the credit risk, and to predict the performance of the company. Auditing is known as an old profession that has realized, in recent years, sharp increases in trading volume and complexity of financial and non-financial data to their client. To audit these companies, auditors are dealing with bulk data and complex structured data. As a result, auditors cannot only rely on tools for reporting the audit process Sirikulvadhana (2002). Data mining can be a useful tool to automatically extract the information from large volumes of data, although an adaptation of data mining in the audit process is a relatively new technique. However, data mining has shown, in many commercial applications related to the audit such as fraud detection and forensic accounting, to have a great advantage. The tasks of data mining have been widely investigated by Sumathi (2006) that noted data mining is in the early stages of its life. A large part of industries including the financial, healthcare, production, transportation industry are using data mining tools and techniques to achieve their goals. For this purpose, they are using old data. They are taking advantage of shifting techniques and technology, mathematics and statistical analysis of information available in the database. With the help of data mining, the analyzers can detect critical events, business communication processes, and abnormalities that need not be considered Sivanandam (2006). In

business, data mining is used to discover the relationships between the data and also used to make better decisions. Data mining also helps to determine the spot sales trend, more efficient marketing and help in predicting customer loyalty. In the context of data mining in the environmental accounting information, Thuraisingham (2000) pointed out that data mining within environmental accounting information is particularly the process of collecting and analyzing environmental accounting data, and presenting them in a way that generates environmental information and knowledge. The knowledge gained through the analysis of this information can be used to further goals of decision-making within organizations. Data mining activities include summarizing, comparing, analyzing, forecasting and estimating Thuraisingham (2000). Data mining enables organizations to apply updated statistical tools and software for analyzing the environmental data as well as using knowledge management by database management for the extraction of information from large databases. Mendes (2017) observed that the adoption of this technology increases the responsibility of accountants and auditors. With the help of this technology, environmental accounting system will be able to produce complete, timely and related environmental information required for decision-making. The users of accounting information mainly need current and ongoing information Mendes (2017). Crespo (2005) investigated the aspect that data mining is not just a tool to keep track of transactions, but it can also complete testing of the system, as well as undertaking the necessary controls to ensure that the organization can produce accurate financial statements according to environmental information Crespo (2005). The main survey in an issue like A. Amani (2017) studies examined that data mining increases the ability of environmental accounting information to take an effective role in the collection of transaction data, providing information to decision-makers, and collaborate on internal controls. Environmental accounting information maintains volumes of transactions, and it's suggested to use them as primary sources of environmental information to carry out organizational purposes Amani (2017).

Therefore, based on the theoretical basis of our study hypothesis is presented as follows:

H1: The environmental marketing helps to products that are presumed to be environmentally safe.
H2: Environmental accounting predicts which customers are likely to leave your company and go to your competitor.
H3: The environmental accounting system helps to identify the manipulation of financial data, intentional deception, and misappropriation of company natural assets.

H4: Environmental accounting helps to find out which services were accepted and used more by the customer to reduce environmental pollution.

H5: Environmental accounting system provides a basis for interactive marketing in the business.
H6: Environmental accounting system helps to analyses the business trends, according to environmental considerations in the course of time.

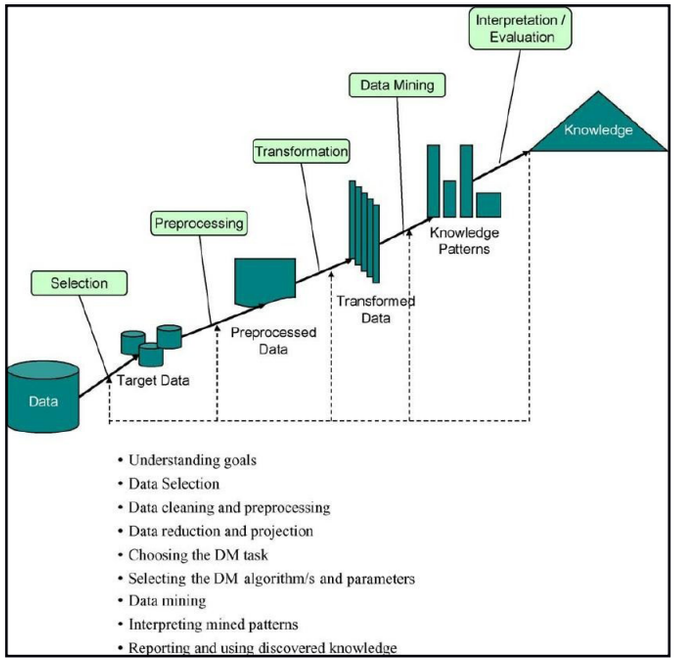
Research Methodology

Sampling Procedure and Data Collection

The study was conducted in India, covering the States of Haryana for North, West Bengal for East, Maharashtra for West and Kerala as representative of South. In this study two separate sets of structured questionnaires were administered, one for managers and one for accountants in the form of soft and hard copy. The sample of study consists of 100 managers, and 243 accountants were selected at random from the public sector (BPCL) and private (AOL) sector companies.

Research Design

Environmental accounting information, manages and exchanges the generated reports and provides services to different systems of the organization. These include the environmental systems, financial systems, management accounting and payroll systems as well as branches of accounting systems. As the model shows, this computerized information system is used as a data storage centre, which contains information about different parts of the accounting system. Using data mining technology involves a combination of information and communications technology (ICT), statistical data analysis tools and knowledge management, which increases the knowledge of decision-makers about accounting A.Rahman (2008).



Source: Utilization of data mining technology within the environmental accounting information in the public sector, Rahman (2008)

Fig. 1: Data Mining used within the Environmental Accounting Information

The goal of data mining is to analyse massive amounts of data to look for patterns that can help predict future results. It can be used to predict customer behaviour, market oscillations, find the best business moves, and so on. The procedure of data mining involves:

Data Mining Process

- *Identification of Objective:* At this stage, it will be determined what the user needs and the level of information he intends to obtain from the database. Also, it is required to understand research objectives clearly and find out what are the needs, on the other hand, the study what want to achieve with these data.
- *Selection of Data:* The data must be based on criteria of objective selection. This involves going over the data to assess its quality and prepare what you need for the following phases. Assess finding the resources, assumptions, constraints and other important factors which should be considered.
- *Data Preparation:* The function of this stage of the process will be the preparation of data and identification of redundant variables used.
- *Evaluation of Data*: The general framework of this stage is the criteria such as the type of data distribution, characteristics and structure of the database and the general condition of the data etc.
- *The Answer Formwork:* The output of this section is to provide formats in the form of images, graphs and neural networks and so on. Then, using mathematics to identify patterns and use them to build different models.
- *Selection Tools:* At this point, appropriate tools for data mining are selected. Data mining is done by techniques like classification, clustering, regression, association rules, outer detection, sequential patterns, and prediction. In this research, we used the classification and regression techniques. Classification analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes. Regression analysis is the data mining method of identifying and analyzing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.
- *Modelling:* Data mining process begins at this stage, which includes search patterns in data sets, classification, and evaluation of the data.
- *Validated Results:* This stage includes the test patterns.
- *Presenting the Results*: The results of this section are the final report to the users.
- *Use the Results:* The purpose of data mining is using the obtained results for decision-making, policy-making, and forecasting to create a better and new situation (Data Mining Tutorial: Process, Techniques, Tools & Examples, 2018; Scott, 2018; Ruxandra, 2013; W. Seifert, 2004).

Results

In this research, we focus on the aspects that are most important components to assess the level of data mining within an Environmental Accounting System in Indian public and private sector companies. In this respect, six

hypotheses has been considered that were evaluated in the statements with a five-point Likert scale, from strongly disagree to agree strongly. The statements measured in classification method are the opinion of managers and accountants regarding the level of data mining within an Environmental Accounting System in Indian companies.

Table 1 shows the number and percentage of public and private sector managers having an opinion regarding the level of data mining in Indian companies. From the table 1, for example, it can be seen that the six aspects of data mining received zero answers for 'strongly disagree' and 'disagree' in both public as well as private sector. In the case of market segmentation, forty per cent of responses received from the public and thirty present from the private sector was 'neutral, ' and sixty per cent of public respondents and seventy per cent of private sector respondents were either 'agree,' or 'strongly agree.'

Table 1: Distribution of Managers of Public and Private Sector in India by Their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	15	30.00
	Agree	25	50.00	30	60.00
	Strongly Agree	5	10.00	5	10.00
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	25	50.00	20	40.00
	Agree	20	40.00	25	50.00
	Strongly Agree	5	10.00	5	10.00
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	15	30.00	5	10.00
	Agree	25	50.00	35	70.00
	Strongly Agree	10	20.00	10	20.00
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	10	20.00
	Agree	30	60.00	25	50.00
	Strongly Agree	0	0.00	15	30.00
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	20	40.00
	Agree	30	60.00	30	60.00
	Strongly Agree	0	0.00	0	0.00
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	10	20.00	5	10.00
	Agree	35	70.00	20	40.00
	Strongly Agree	5	10.00	25	50.00

Source: primary (2017)

Also, the responses in regard interactive marketing are as follows; forty per cent of both sectors respondents expressed 'neutral' response, sixty per cent of responses in both sectors is showing 'agree' and none of them responded 'strongly agree' in both sectors. The last aspect measured is the trend analysis, which respectively received twenty per cent and ten per cent 'neutral' response from public and private sector companies. Eighty per cent of the public sector and ninety per cent of private sector respondents were either 'agree,' or 'strongly agree.'

Table 2: Mean Opinion Scores of Managers About the Level of Various Aspects of Data Mining in Companies with One Sample T-Test And Two Independent Samples T-test

Aspects of data mining	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.70*	0.65	3.80*	0.61	-0.798	0.427
Customer churn	3.60*	0.67	3.70*	0.65	-0.759	0.450
Fraud detection	3.90*	0.71	4.10*	0.54	-1.585	0.116
Market basket analysis	3.60*	0.50	4.10*	0.71	-4.096	0.000
Interactive marketing	3.60*	0.50	3.60*	0.50	0.000	1.000
Trend analysis	3.90*	0.54	4.40*	0.67	-4.096	0.000

* Significantly greater than 3.00 as per one sample t-test

Source: primary data (2017)

Table 2 shows the mean opinion scores of managers regarding the level of data mining within the environmental accounting system (EAS) in public and private sector companies in India separately. It can be found that all mean scores are greater than three as the significance levels of one sample t-test with test value 3.00 are smaller than 0.05. It means that according to managers of both public and private sector companies the level of data mining is above average. From the two independent sample t-test, it can be understood that managers have a difference of opinion in two attributes, namely market basket analysis and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the market basket analysis and trend analysis are better in the private sector as the mean scores are 4.10 and 4.40 respectively.

Table 3: Distribution of Accountant of Public and Private Sector in India by Their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	29	23.58	36	30.00
	Agree	94	76.42	60	50.00
	Strongly Agree	0	0.00	24	20.00
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	62	50.41	60	50.00
	Agree	40	32.52	36	30.00
	Strongly Agree	21	17.07	24	20.00
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	10	8.13	0	0.00
	Agree	83	67.48	60	50.00
	Strongly Agree	30	24.39	60	50.00
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	22	17.89	12	10.00
	Agree	101	82.11	96	80.00
	Strongly Agree	0	0.00	12	10.00
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	21	17.07	24	20.00
	Agree	93	75.61	72	60.00
	Strongly Agree	9	7.32	24	20.00
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	11	8.94	0	0.00
	Agree	72	58.54	60	50.00
	Strongly Agree	40	32.52	60	50.00

Source: primary data (2017)

Table 3 shows the number and percentage of public and private sector accountants by their opinion regarding the level of data mining in India. From the table, it can be seen that the six aspects of data mining received zero answers for 'strongly disagree' and 'disagree' in both public as well as private sector. In the case of market segmentation 23.58 per cent of responses received from public and 30.00 per cent from the private sector was 'neutral,' and 76.42 per cent of public respondents were 'agree' and no responses received regarding 'strongly agree' and 70.00 per cent of private-sector respondents were either 'agree', or 'strongly agree'. For the aspect customer churn, the percentage of neutral responses received from public and private sector accountants are 50.41 and 50.00 respectively. It can also be seen that 49.59 per cent of public sector accountants and 50.00 per cent of private-sector accountants were either 'agree,' or 'strongly agree.' The third aspect is fraud detection, which has received 8.13 per cent of neutral responses from public sector accountants and 0.00 per cent of private-sector accountants. At the same time, 91.87 per cent of the public sector and 100.00 per cent of private-sector accountants respectively were either 'agree,' or 'strongly agree. The last aspect measured is the trend analysis, which received 8.94 per cent and 0.00 per cent neutral response respectively from public and private sector companies. It can also be seen that 91.06 per cent of the public sector and 100.00 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The responses of accountants were converted into scores and are presented in Table 4.

Table 4: Mean Opinion Scores of Accountant Regarding the Level of Existence of Various Aspects of Data Mining within EAS in Public and Private Companies with a Two-Sample T-test

Accountant	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.76*	0.43	3.90*	0.70	-1.826	0.069
Customer churn	3.67*	0.75	3.70*	0.78	-0.338	0.736
Fraud detection	4.16*	0.55	4.50*	0.50	-4.997	0.000
Market basket analysis	3.82*	0.39	4.00*	0.45	3.337	0.001
Interactive marketing	3.90*	0.49	4.00*	0.64	1.347	0.179
Trend analysis	4.24*	0.60	4.50*	0.50	-3.712	0.000

* significantly greater than 3.00 as per one sample t-test

Source: primary data (2017)

Table 4 shows the mean opinion scores of accountants regarding the level of data mining within the Environmental Accounting System (EAS) in public and private sector companies in India separately. It can be found that all mean scores are greater than three as the significance levels of one sample t-test with test value 3.00 are smaller than 0.05. It means that according to accountants of both public and private sector companies the level of data mining is above average. From the two independent sample t-test, it can be understood that accountants have a difference of opinion in three attributes, namely fraud detection, market basket analysis, and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores 4.50, 4.00 and 4.50 respectively which is higher than that of public sector companies.

Table 5: Distribution of Total Respondent (both managers and accountants) of Public and Private Sector in India by their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	49	28.32	51	30.00
	Agree	119	68.79	90	52.94
	Strongly Agree	5	2.89	29	17.06
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	87	50.29	80	47.06
	Agree	60	34.68	61	35.88
	Strongly Agree	26	15.03	29	17.06
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	25	14.45	5	2.94
	Agree	108	62.43	95	55.88
	Strongly Agree	40	23.12	70	41.18
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	42	24.28	22	12.94
	Agree	131	75.72	121	71.18
	Strongly Agree	0	0.00	27	15.88
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	41	23.70	44	25.88
	Agree	123	71.10	102	60.00
	Strongly Agree	9	5.20	24	14.12
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	21	12.14	5	2.94
	Agree	107	61.85	80	47.06
	Strongly Agree	45	26.01	85	50.00

Source: primary data (2017)

Table 5 shows the number and percentage of public and private sector respondents (both managers and accountants) by their opinion regarding the level of data mining in Indian companies. From the table 5, it can be seen that in the case of market segmentation, 28.32 per cent of responses received from the public and 30.00 per cent from the private sector was 'neutral', and 71.68 per cent of public and 70.00 per cent of private sector respondents were either 'agree,' or 'strongly agree.' And also, the aspect of fraud detection has received 14.45 per cent and 2.94 per cent 'neutral' opinion from the public sector and private sector respondents respectively. The table also indicated that 85.55 per cent of public and 97.06 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The responses regarding the aspect market basket analysis are as follows; 24.28 per cent of public and 12.94 per cent of private-sector respondents expressed 'neutral' opinion. The responses received have shown 75.72 per cent of the public sector, and 87.06 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The last aspect measured is the trend analysis, which respectively received 12.14 per cent and 2.94 per cent 'neutral' response from the respondents of public and private sector companies. The responses received have shown 87.86 per cent of the public sector, and 97.06 per cent of private sector respondents were either 'agree,' or 'strongly agree.' The responses of accountants were converted into scores and are presented in Table 6.

Table 6: Mean Opinion Scores of the Total Respondent (both managers and accountants) About the Level of Existence of Various Aspects of the Level of Data Mining within EAS in Public and Private Companies with Two Sample T-test

Combined	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.75*	0.50	3.87*	0.68	-1.950	0.052
Customer churn	3.65*	0.73	3.70*	0.75	-0.661	0.509
Fraud detection	4.09*	0.61	4.38*	0.55	-4.738	0.000
Market basket analysis	3.76*	0.43	4.03*	0.54	-5.182	0.000
Interactive marketing	3.82*	0.51	3.88*	0.62	-1.099	0.273
Trend analysis	4.14*	0.60	4.47*	0.56	-5.291	0.000

* Significantly greater than 3.00 as per two-sample t-test

Source: primary data (2017)

Table 6 shows the opinion scores of both managers and accountants regarding the level of data mining within the Environmental Accounting System (EAS) in public and private sector companies in India. It can be found that all mean scores are greater than three as the significance levels of two sample t-test with test value 3.00 are smaller than 0.05. It means that according to respondents of both public and private sector companies the level of data mining is above average in their company. From the two independent sample t-test, it can be understood that respondent has a difference of opinion in three attributes, namely fraud detection, market basket analysis, and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores are 4.38, 4.03 and 4.47 respectively which greater than that public-sector companies. To find the level of data mining within the Environmental Accounting System in India, the elements of data mining were combined by taking a weighted average using Principal Components Analysis (PCA). Factor loadings were taken as weights for each element of the data mining. Factor loading was calculated from the responses collected from the managers of public and private sector companies. Table 7 presents the weights of elements of the data mining.

Table 7: Weight for the Elements of Data Mining

No.	*Elements of Data Mining*	*Weight*
1	Market segmentation	0.596
2	Customer churn	0.559
3	Fraud detection	0.323
4	Market basket analysis	0.805
5	Interactive marketing	0.743
6	Trend analysis	0.600

Source: primary data (2017)

From Table 7 it can be seen that 'market segmentation' with the weight of 0.805 is the most important aspect of data mining within the Environmental Accounting System in India. 'Interactive marketing' with a weighted score of 0.743 is the second important aspect and the 'trend analysis' has got the third rank with the weight score of 0.600. The findings of this table indicate that, as per the view of respondents, the elements which obtained higher

weight score, will have a higher manifestation in the level of data mining within the Environmental Accounting System than other elements with lower factor loadings and vice versa. Using the above weights, the opinion scores of respondents about the level of data mining within EAS was computed, and descriptive statistics of the estimated score are presented in Table 8.

Table 8: Descriptive Statistics of Opinion Score of Respondents about the Level of Data Mining within EAS by the Opinion of Managers and Accountants

Statistic	*Data Mining*	
	Managers	*Accountant*
Mean	3.81*	3.98*
Median	3.86	4.00
Mode	4.00	4.00
Skewness	0.115	0.647
Kurtosis	-0.732	0.526
Maximum	4.80	4.84
Minimum	3.09	3.48
Range	1.71	1.36
SD	0.47	0.32
t	3.850	
Sig.	0.000	

Source: primary data (2017)

From the table 8, it can be seen that the mean scores of the level of data mining are more than three for both managers as well as accountants. The mean score obtained for the managers is 3.81, and that for accountants is 3.98 which is significantly different as per one sample t-test since the significance levels are found to be lesser than 0.05. The result indicates that the level of data mining is above average in both public and private sector companies in India. The two sample t-test shows that the mean opinion scores of accountants are significantly higher than that of the manager as the significance level is lesser than 0.05. From this result, it can be concluded that the accountants have the opinion that the level of data mining is higher within the Environmental Accounting System as compared to that of the manager.

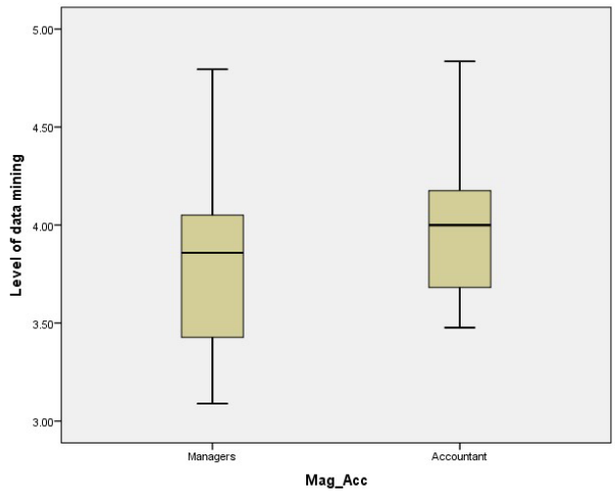


Fig. 2: Box Plot Representing the Mean Score of Elements of the Level of Data Mining by Designation

From Fig. 2 it can be inferred that the level of data mining as per the opinion of managers in companies to an accountant is different. It can be seen that the median of box plot showing that the fifty per cent of responses with positive skewness is acquired the score greater than 3.86, which means that they rely on an environmental accounting system for data mining. The top whisker shows the maximum score of 4.80, and down whisker show, the minimum score of 3.09. Table 8 discerned that the difference between the maximum and a minimum score of manager's opinion is 1.71. The box plot of accountants showing the median score of 3.98, which indicates that managers are less relied on data mining within environmental accounting compared to an accountant. The box plot of accountants showing the positive skewness of responses. The top whisker shows the maximum score of 4.84, and down whisker show, the minimum score of 3.48 and the range difference between a maximum and minimum score of accountant's opinion is 1.36. From the comparison of a median score of managers and accountants, it can be concluded that accountants are more rely on data mining within the environmental accounting system than managers. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that managers are less consistent than the accountants in relying on data mining within environmental accounting.

Table 9: Descriptive Statistics of the Opinion of Respondents from Public and Private Sector Companies about the Level of Data Mining within EAS by Ownership

Statistics	*DM*	
	Public	*Private*
Mean	3.84	4.03
Median	4.00	4.00
Mode	4.00	4.00
Skewness	-0.552	0.081
Kurtosis	-0.279	-0.122
Maximum	4.41	4.84
Minimum	3.09	3.09
Range	1.32	1.75
SD	0.31	0.41
t	4.670	
Sig.	0.000	

Source: primary data (2017)

Table 9 presents descriptive statistics of the level of data mining within an environmental accounting system as per the opinion of public and private sector companies in India. From the table 9, it can be seen that the mean score obtained for the public sector is 3.84 and that for the private sector is 4.03 which is significantly different from the significant level of t-test is less than 0.05. From this result, it can be concluded that the private sector companies are more relying on an environmental accounting system for data mining.

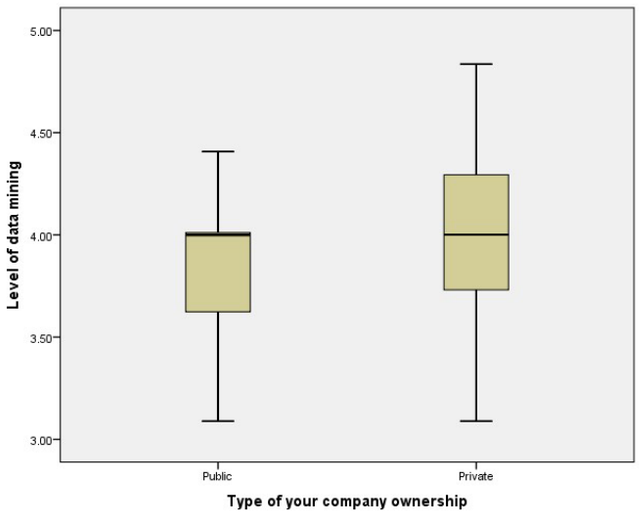


Fig. 3: Box Plot Representing the Mean Score of Elements of the Level of Data Mining by Ownership

From Fig. 3 it can be inferred that the level of data mining as per the opinion of public and private sector companies are different. The box plot of public sector showing the negative skewness of responses regarding the level of data mining within environmental accounting. The median is found to be very close to the upper part of the box indicating that twenty-five per cent of public sector employees has average scores are the median value or little higher than medium. The top whisker shows the maximum score of 4.41, and down whisker show, the minimum score of 3.09. The box showing that fifty per cent of responses is placed between the upper quartile and lower quartile, also the length of the box-and-whisker plot linked to the range of opinion given to the level of data mining within the environmental accounting system. From Table 9 it can be seen that the difference between a maximum and minimum score of respondent's opinion is 1.32. The box plot of private sector companies showing the median score of 4.00, which shows that the fifty per cent of private sector companies is expressing the opinion score greater than this value, which indicates that they strongly rely on data mining within the environmental accounting system. Also, the positive skewness of responses regarding the level of data mining and the top whisker shows the maximum score of 4.84 and down whisker show the minimum score of 3.09. The interquartile range of responses shows that fifty per cent of responses is placed between the upper quartile and lower quartile, also the length of the box-and-whisker plot linked to the range of opinion given to the level of data mining within an environmental accounting system. From Table 9 it can be seen that the range is 1.75, which shows the difference between the maximum and a minimum score of private sector companies' opinion. From the comparison of a median score of public and private sector companies, it can be concluded that private sector companies are more reliant on data mining within environmental accounting. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that private sector companies are less consistent than the public sector companies in relying on data mining within the environmental accounting system. Table 10 presents the result of the Regression Coefficients of perceptions of respondents about the level of data mining within the environmental accounting system in India. The independent variables selected were gender, age, knowledge, education, ownership, and experience. All independent variables were represented as a dummy variable. So the assumptions of multiple regressions were not verified.

Table 10: Coefficients of the Multiple Regression Model for the Level of Data Mining within EAS

Variables B		Unstandardized Coefficients		Standardized Coefficients		
		Std. Error	Beta	t	Sig.	
Name	Dummy variables					
(Constant)		4.576	0.196		23.298	0.000
Designation	manager_d	- 0.091	0.056	-0.11	-1.627	0.105
Type of company ownership	Public_d	-0.09	0.047	-0.12	-1.913	0.057
Gender	male_d	0.093	0.047	0.115	1.99	0.047
Age	Ageless25_d	0.243	0.165	0.239	1.473	0.142
	Age25_d	0.313	0.132	0.388	2.376	0.018
	Age36_d	0.211	0.127	0.239	1.658	0.098
	Age46_d	0.025	0.125	0.028	0.202	0.840
Knowledge about EA	kno_L_d	- 0.691	0.173	-0.324	-3.999	0.000
	kno_A_d	- 0.617	0.115	-0.794	-5.36	0.000
	kno_G_d	- 0.564	0.103	-0.749	-5.468	0.000
Qualification	Edu_prof_d	- 0.546	0.158	-0.538	-3.466	0.001
	Edu_Mas_d	-0.38	0.131	-0.501	-2.892	0.004
	Edu_Bac_d	-0.29	0.131	-0.373	-2.214	0.028
Years of experience	Exp_less5_d	- 0.002	0.099	-0.002	-0.018	0.986
	Exp_5_d	0.098	0.081	0.125	1.207	0.228
	Exp_11_d	0.04	0.08	0.039	0.507	0.612

Source: primary data (2017)

From the table 10, it can be seen that eight dummy variables are found to be significant as their significance levels are less than 0.05. It means that the effects of all eight predictor dummy variables are significant in predicting the perception of respondents on the level of data mining within the environmental accounting. Since out of sixteen dummy variables, fifty per cent of them are significant; this model can be used to predict the level of perception of respondents about the level of data mining.

The validity of this regression model can also be seen from the result of ANOVA given in Table 11 by the fact that the significance level of F-Value is less than 0.05. The R square of the regression model is found to be 0.282. It means that the selected independent variables determine 28.2 per cent variation in the perception of respondents about the level of data mining within the environmental accounting.

Table 11: Result of ANOVA of the Regression Model for Perception of Respondents on the Level of Data Mining within EA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	13.616	16	0.851	7.987	0.000
Residual	34.736	326	0.107		
Total	48.352	342			

Source: primary data (2017)

The results of the Regression Coefficients presented in Table 10 provide detailed information about the relative effects of each predictor variable about the respective reference categories (the category with assigned value zero is called the reference category). The variables which have a significant effect include dummy variable representing gender, the respondent 25 to 35 years old, knowledge and qualification in three categories, namely, professionals, masters and bachelors. The table shows that the regression coefficient B obtained is negative which means that the managers' perceptions about the level of data mining within the environmental accounting are lower than that of accountants by an amount of 0.091. As far as ownership is concerned the dummy variable representing public sector (Sig. = 0.057) is not significant in determining the variation in the perceptions of respondents about the level of data mining in concerning the private as the reference category. The respondent with the gender category, the male dummy variable (Sig. = 0.047) is significant in determining the variation in the perceptions of respondents, concerning the female (reference category). However, the table shows that the regression coefficient B obtained the positive perception (B= 0.093) which means that the perception of male respondent about the level of data mining is higher than female respondents by an amount of 0.093. In the case of age, the second age category of respondents is significant 0.018 in determining the variation in the perceptions of respondents is less than 0.05. It can be inferred that the respondent with age between 25 to 35 years received a significant score (B=0.313). Therefore statistically, they have an effect on the level of data mining within environmental accounting. As the regression coefficient B shows the positive value in all categories, it can be said the respondents in all age categories have a high perception of the level of data mining within an environmental accounting compared to the reference category (55+years). It can also be seen from Table 10 that the respondent in all knowledge levels is significant in determining the variation in the perceptions of respondents about the level of data mining. Therefore, statistically, they have an effect on the level of data mining within environmental accounting. As the regression coefficient B shows negative scores in all categories (little knowledge -0.691, average knowledge -0.617, good knowledge -0.564), it can be said that the all respondent categories have a low perception of the level of data mining within an environmental accounting compared to respondents having rich knowledge (reference category). All qualification categories of respondent are significant in determining the variation in the perceptions of respondents about the level of data mining within the environmental accounting. As the regression coefficient B shows negative scores in all categories (professionals -0.546, masters -0.380, bachelors - 0.290), it can be said that the respondents in those categories have a low perception of the level of data mining within an Environmental Accounting compared to less than Bachelor education (reference category). At the same time, respondents' perception of all experience category is not significant as their significance level is higher than 0.05.

Table 12 presents the MCA (Multiple Classification Analysis) adaptations of the perceptions of respondents about the level of data mining within environmental accounting companies derived from the above regression model. The MCA table provides a clear understanding of the effects of predictive characteristics of respondent on the distinction of their perceptions about the level of data mining within an environmental accounting. From Table 12, it has been observed that the unadjusted and adjusted R-values of knowledge about EA (0.335 and 0.553), qualification (0.128 and 0.294), and gender (0.059 and 0.115) differ very much compared to other predictor variables. Unadjusted R values are found to be higher in three variables, namely designation, ownership and experience while adjusted R values are higher for other variables such as gender, age, knowledge, and qualification. It means that the effect of knowledge about environmental accounting, qualification, gender, and age on the perceptions of respondents about the level of data mining within an environmental accounting is augmented by the effect of other variables. At the same time effect of designation, ownership and experience are explained away by the effect of other variables.

Table 12: MCA Table of the Opinion Score about the Level of Data Mining

			Unadjusted		*Adjusted*	
Characteristics		*N*	*Mean Score*	*R*	*Mean Score*	*R*
Designation				0.204		0.110
1	Managers	100	3.815		3.870	

Characteristics		N	Unadjusted Mean Score	Unadjusted R	Adjusted Mean Score	Adjusted R
2	Accountant	243	3.983		3.960	
Type of company ownership				0.245		0.120
1	Public	173	3.843		3.889	
2	Private	170	4.027		3.979	
Gender				0.059		0.115
1	Male	234	3.949		3.964	
2	Female	109	3.902		3.871	
Age				0.266		0.309
1	Less than 25years	56	3.844		3.982	
2	25-35 years	109	4.011		4.052	
3	36-45 years	82	3.954		3.950	
4	46-55 years	75	3.960		3.764	
5	Above 55 years	21	3.607		3.739	
Knowledge about EA				0.335		0.553
1	Little knowledge	11	3.624		3.752	
2	Average knowledge	127	3.885		3.826	
3	Good knowledge	157	3.908		3.879	
4	Rich knowl-edge	48	4.221		4.443	
Qualification				0.128		0.294
1	Professional quality	56	4.028		3.748	
2	Master's degree	149	3.891		3.915	
3	Bachelor degree	126	3.938		4.004	
4	Others	12	3.979		4.295	
Years of experience				0.273		0.120
1	Less than 5years	82	3.860		3.891	
2	5-10 years	124	3.995		3.991	
3	11-15 years	52	4.102		3.933	
4	Over 15years	85	3.814		3.893	
Full Model				0.531		0.282

Source: primary data (2017)

Since the absolute values of the difference between unadjusted and adjusted R are comparatively lower for age, gender and designation it can be inferred that these variables have an independent effect on the perception of respondents about the level of data mining within environmental accounting of public and private sector in India. Table 12 shows that the *R Square* of the regression model was found to be 0.079, which means that 7.95 percent (R Square = R^2 (0.282) *100= 7.95) of variation in the perceptions of respondents about the level of data mining within environmental accounting in the public and private sector was determined by their designation, type of company ownership, gender, age, knowledge about environmental accounting, qualification, and years of experience. From table 12 it is found that the mean opinion score of accountants about the level of data mining is higher in comparison to the managers. The effect of designation on the level of data mining within environmental accounting is found to be dependent on the adjusted R-value 0.110 and unadjusted R- value 0.204. The adjusted R-value is lesser than the unadjusted R-value, these parameters indicating that the effect of the designation on the level of data mining was explained away by the other independent variables. It means that the effect of the designation on the level of data mining within the environmental accounting is not independent. As per the type of company ownership, it can be understood that the mean opinion score of public sector respondents about the level of data mining is lower in comparison to the private sector respondents. The effect of type of ownership at the level of data mining is found to be dependent on other predictor variables as the difference between adjusted R-value 0.120, and unadjusted *R-value* 0.245 is comparatively higher. The adjusted R-value is lesser than the unadjusted R-value, these parameters indicating that the effect of type of ownership on the perception of the level of data mining was explained away by the other independent variables. It means that the effect of type of ownership on the perception of the level of the data mining within the environmental accounting is not independent. As far as the gender of the respondents is concerned, male respondents are more satisfied with the level of data mining than female respondents as their adjusted mean opinion scores are 3.964 and 3.871 respectively. The adjusted R-value is not very much different than the unadjusted R-value which indicates that the effect of

gender on the perception of the respondent about the level of data mining was independent. In the case of age, the respondents with the age group of 25-35 years are more satisfied with the level of data mining companies with an adjusted mean score of 4.052 followed by the youngest age group with an adjusted mean score of 3.982. The opinion scores of respondents in the higher age groups are found to be decreasing steadily. As the difference between adjusted and unadjusted R-value is the least, it is evident that the effect of age on the perception of the respondent about the level of data mining within EA is independent. According to the MCA table the respondents with qualification less than a bachelor (mentioned as 'others') with the highest mean score of 4.295 indicating that they are more satisfied with the level of data mining within an environmental accounting in Indian companies. The effect of qualification at the level of data mining within the environmental accounting is found to be independent of the difference between adjusted R-Values 0.294 and unadjusted R- Values 0.128 are comparatively is high. This indicates that the effect of qualification at the level of data mining within the environmental accounting is dependent on other independent variables. The effect of knowledge about Environmental Accounting on the level of data mining within the environmental accounting is found to have an independent effect on the difference between adjusted R-value 0.553, and unadjusted R-Value 0.335 is the highest. Respondents with rich knowledge are more satisfied with the level of data mining within an environmental accounting (4.443). The satisfaction levels of respondents in the other levels of knowledge are founded to be more or less equal. As per the years of experience, it can be seen that the respondent with experience between five to ten years has highest mean opinion score of 3.991 about the level of data mining within Environmental Accounting followed by those with age between eleven to fifteen (3.933). The effect of years of experience on the level of data mining within Environmental Accounting is found to be dependent on the difference between the adjusted R-value (0.120), and unadjusted R-value (0.273) is comparatively higher. The adjusted R-value is lesser than the unadjusted R-value, indicating that the effect of years of experience on the level of data mining was explained away by the other independent variables. It means that the effect of type of years of experience at the level of data mining within the environmental accounting is not independent.

Summary and Conclusion

In our study, these facts confirmed that from the comparison mean scores of by two independent sample t-test in the public and private sector, it could be seen that the market basket analysis and trend analysis are better in private than public sector. The result of two independent sample t-test also shows that accountants in the public and private sector, in comparison of mean scores of objective such fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores 4.50, 4.00 and 4.50 respectively which is higher than that of public sector companies. In the terms weight, the most important aspect of determent elements of the level of data mining within public and private sector companies in India respectively belong to; market segmentation, Interactive marketing, trend analysis, market segmentation, customer churn, fraud detection. The two sample t-test display that the accountants have the opinion that the level of data mining is higher within the environmental accounting system as compared to that of the manager. Accountants are more rely on data mining within the environmental accounting system than managers. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that managers are less consistent than the accountants in relying on data mining within environmental accounting. According to the result of t-test analysis for comparison the opinion of data mining in public and private sector companies; the private sector is more relying on an environmental accounting system for data mining. The results of the multiple regression presented that according to predictor variables like effect of designation, ownership, gender, age, knowledge about environmental accounting, qualification, and experience on the perception of respondents about the level of data mining; the managers' perceptions about the level of data mining within the environmental accounting are lower than that of accountants by an amount of 0.091. Also, all respondents in all age categories have a high perception of the level of data mining within an environmental accounting compared to the reference category (55+years). In terms of knowledge the all respondent categories have a low perception of the level of data mining within an environmental accounting compared to respondents having rich knowledge (reference category). Our study indicated that all qualification categories of respondent are significant in determining the variation in the

perceptions of respondents about the level of data mining within the environmental accounting. The respondents in those categories have a low perception of the level of data mining within an Environmental Accounting compared to less than Bachelor education (reference category). In addition, respondents perception of all experience category is not significant as their significance level is higher than 0.05. Finally; the result indicates that according to the opinion of respondents the level of data mining within environmental accounting is not the same as the mean opinion scores of the respondents of public and private sector companies are different and the effect of the type of company ownership depends upon predictable variables such as designation, age, gender, knowledge about environmental accounting, experience, qualification which varies in public and private sector companies. Hence the result rejects the null hypothesis that 'there is no significant difference existed in the level of data mining between public and private sector industry in India' and this fact confirmed that the level of data mining within environmental accounting is significantly higher in private sector companies in India.

Limitation of Study

The present study was concentrated on public and private sector companies in India, where Environmental Accounting has been in use. The result of the study indicated the need for a control group of companies without using Environmental Accounting to get exhaustive information about the advantages and disadvantages of Environmental Accounting in management function, decision-making and data mining using a comparative study. So by the felt limitation, future researchers are recommended to include a control group of companies without using Environmental Accounting for conducting studies on the similar topic.

References

Yu, T., Lenzen, M., Gallego, B., & Debenham, J. (2006). *A Data Mining System for Estimating a Large size Matrix in Environmental Accounting*. International Conference on Artificial Intelligence Application and Innovation, 2007; Sydney: Australia Faculty of Information Technology, University of Technology, Sydney, NSW.

Alshhadat, A. R., Ghaleb, E. R., & Amer, Q. (2018). The use of data mining techniques in accounting and finance as a corporate strategic tool: An empirical investigation on banks operating in emerging economies. *International Journal of Economics and Business Research*, *15*(4), 442.

Chinchuluun, A., Xanthopoulos, P., Tomaino, V., & Pardalos, P. (2010). Data mining techniques in agricultural and environmental sciences. *International Journal of Agricultural and Environmental Information Systems*, *1*(5), 26–40.

Common, M. (1996). *Environmental and resource economics: An introduction* (2nd ed.). Harlow, Essex: Addison Wesley Longman Ltd Publishers.

Gilberto Carvalho, P., Ricardo, C., & Nelson Francisco Favila, E. (2008). Data mining for environmental analysis and diagnostic: A case study of upwelling ecosystem of Arraial do Cabo. *Azilian Journal of Oceanography*, *56*(3), 1–12.

Peter, R. (2013). Data mining solutions for the business environment. *Database Systems Journal*, *4*(5), 21–29.

Shah, K., Chauhan, P., & Potdar, M. (2014). Data mining techniques in parallel environment - a comprehensive survey. *International Journal of Computer Applications*, *108*(1), 36–41.

Rouse, M. (2017). Data Mining. *TechTarget*. Retrieved August 26, 2018 from https://searchsqlserver.techtarget.com/definition/data-mining

Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, *21*(4), 461–480.

Pechenizkiy, M., Puuronen, S., & Tsymbal, A. Why data mining research does not contribute to the business?. In *Proc. of Data Mining for Business Workshop DMBiz (ECML/PKDD'05)*, Porto, Portugal.

Kirkos, E., Spathis, C. H., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003.

Sirikulvadhana, S. (2002). *Data mining as a financial auditing tool*. M.Sc. Thesis in Accounting, Swedish: School of Economics and Business Administration.

Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to data mining and its applications* (2nd ed.). New York, NY: Springer-Verlag Publishers.

Thuraisingham, B. (2000). A primer for understanding and applying data mining. *IEEE Access*, *2*(1), 28–31.

Mendes, R. P., & Vilela J. (2017). Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, *5*(9), 10562–10582.

Crespo, F., & Weber, R. A methodology for dynamic data mining based on fuzzy clustering. *Fuzzy Sets and Systems*, *150*(2), 267–284.

Amani, F. M., & Fadlallah, A. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *4*(5), 32–58.

Rahman, H. (2008). *Data mining applications for empowering knowledge societies* (2nd ed.). New York, NY: Information Science Reference (an imprint of IGI Global) Publishers.

Data Mining Tutorial: Process, Techniques, Tools & Examples. 2018. Retrieved September 1, 2018 from https://www.guru99.com/data-mining-tutorial.html

Scott, P. D., & Wilkins, E. (1999). Evaluating data mining procedures: Techniques for generating artificial data sets. *Information and Software Technology*, *41*, 579–587.

Ruxandra, P. (2013). Data mining solutions for the business environment. *Database Systems Journal*, *5*(4), 21–29.

Seifert, J. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, *21*, 461–480.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 24-35
http://publishingindia.com/ijbai/

Long Term Memory: Evidence from Major Sectoral Indices of India

Anju Bala*, Kapil Gupta**

Abstract

This paper tests the existence of long term memory with reference to structural changes/breaks in Indian Stock Market. Furthermore, the present paper applied Hurst Exponent in Rescaled Range Analysis as suggested by Hurst (1951) and Lo (1991) and structural breaks detected by using Multiple Break Test (Balcilar et al., 2015) by using daily returns of sectoral indices from January 2010 to May 2018. Empirical evidence shows the predictable structure in all sectoral indices (2010-2018) except Nifty Private Bank with H value 0.4972. The findings imply that existence of long memory would be useful for the investors, practitioners, academicians, and policymakers.

Keywords: Emerging Market, Long Term Memory, Hurst Exponent, Structural Breaks, Market Efficiency

Introduction

Forecasting the economic series like stock returns, prices, interest rate, inflation rate and trade rate are quite certain areas to be investigated. Prediction the time series could be possible through the testing of long-range dependence. When any new information or shock comes into the market and does not adjust quickly and has a long-term impact on prices or returns, is called long memory component (Peters, 1994). If returns series demonstrate the presence of persistence behavior, previous returns can be utilized to estimate the upcoming returns. Presence of long range property provides support for taking investment decision which may also help to generate abnormal profits from the financial investment. This paper has two major contributions; first, to analysis the structural break dates on sectoral index of the National Stock Exchange of India (NSEI). The break seems to be related to the causes of long term property in sectoral indices. Second, the study also tested for long memory component for returns of indices and found the degree of persistence level with respect to Hurst Exponent.

The study also tries to pursuit whether long memory effect is contingent upon structural breaks. There is a paucity of literature to studying the occurrence of persistence behavior of Indian Stock Market. Therefore, this paper is considered to plug this gap. Paper is further organized in the following sections. The second section concerns with reviews of the literature. Section third covers the database and research methodology. The empirical observations of the study are discussed in the fourth section. The conclusion of the study has been presented in the last section.

Literature Review

The seminal research on long range dependence in stock market initiated by Hurst (1951). Furthermore, Greene and Fietlitz (1977), and Aydogan and Booth (1988) demonstrate that long memory behavior exhibits in US stock returns. However, Lo (1991), did not report significant long-term persistence in US stock returns. Nonetheless, Mandelbrot (1971) confronted that the arbitrage may not be negotiable when a long-term memory is exhibited. Thereafter, Lo (1991), investigated that the dynamic behavior in financial market considerable reason for long memory dynamics Furthermore, Hiremath and Bandi (2010) examined that researchers, academician, investors, and practitioner are more concerned to explore the nature of Indian equity market. They propose that high

* Research Scholar (Finance), Department of Management, I.K. Gujral Punjab Technical University, Jalandhar.
Email: anjubala_attri@yahoo.in
** Assistant Professor, Department of Management, I.K. Gujral Punjab Technical University, Jalandhar, Punjab, India.
Email: kapilfutures@gmail.com

volatility, anomalous behavior, and market resistance are the characteristics of developing economies. This affirms that due to these shortcomings and inadequacy long memory behavior might arise in the developing country.

Moreover, Badani (2006, 2008), explored the persistence behavior in India and found that stock returns do not report the presence of long term behavior whereas; absolute returns and squared returns (which represent volatility) exhibit the persistence behavior. On the other hand, subsample covering the duration from March 2001 to December 2007, affirms that volatility does not contain the persistence behavior.

In addition, Goudarzi (2010) observed long memory in BSE 500 returns. Evidence contended that BSE 500 returns and volatility is more significant with leverage. Moreover, Ma et al. (2006) examined little serial correlation in returns of Chinese stock market. Likewise, Verma (2008) affirms that only three companies out of sixty entail the persistence behavior in returns. In a related study, Killic (2004) reported a serial correlation in the volatility process in Turkey. Furthermore, Edgar E. Peters (1989), observed the stronger influence of investors sentiments on bond and stock returns of standard & Poor 500 by using the Hurst Exponent. In addition, Bala and Gupta (2018) reported significant long term memory in Sensex and Nifty returns series. However, volatility series does not contain any persistence behavior but exhibit clustering.

Similarly, Nath and Reddy (2012) examined the persistence behavior in Rupee-Dollar exchange rates. Evidence indicates there are chances of random walk in three month, while for other time period it may have mean reverting or persistence tendency. Furthermore, Mahalingam et.al (2012) observed high degree of persistence behavior in Indian stock market.

Furthermore, Persistence behavior exhibited in absolute returns along with volatility in ten international markets (Bhattacharya and Bhattacharya (2012). However, evidence did not hold the Taylor effect. Moreover, Chen and Diaz (2013) observed significant persistence behavior in green exchange traded funds. Whereas, non-green exchange traded funds did not advocates serial correlation in volatility. In similar study, Hiremath and Kamaiah (2010) advocates that long range dependence has major inference for the literary work of finance and factual applicability, which has little consideration toward Indian framework. Henry (2002) examined the persistence behavior in Taiwanese, German and South Korean International stock markets. Moreover, literature has emphasized on the concern of observed long memory attribute have a real or spurious effect. Evidence indicates that observed persistence behavior insinuate to be genuine not due to shift enhancement and structural breaks of Africa (MENA) region and Middle East in variance.

On the contrary, Jayasuriya (2009) advocate that structural changes and persistence behavior in volatility does not contain any significant relation with each other. Similarly, Chung, et al (2000), found that Asia-Pacific markets hold the spurious serial correlation due to shift enhancement in variance series. However, Cevik and Emec (2013) observed persistence behavior in returns series of Turkish stock market.

Moreover, the spillover-effect an indispensable part of price movement from one place to another. Lee (2001) examined the spillover effect among developed countries (Germany, US and, Japan) and developing countries in the MENA region. Result indicates price and volatility series exhibited the spillover impact from developed to emerging markets, but not return back. In similar study, Hamao et al. (1990) support the spillover effects, which vary from New York to Tokyo, but not vice versa.

Furthermore, Turkyilmaz and Balibey (2014) examined the Pakistan security exchange is inefficient in a weak form of market and contain serial correlation structure in volatility series. The Brazil financial market reported the serial correlation behavior in volatility and return series Brazil stock market (Cavalcante and Assaf (2002)). Likewise, Danilenko (2009) discussed that industrial sector report the significant long memory behavior whereas, healthcare and utilities sector entail the weak long range dependence. However, Badani (2008, 2009), support the significant persistence behavior in volatility but not in returns series.

Moreover, Serial correlation and persistence of shocks studied in Egypt, Tunisia, and Morocco stock markets. Evidence advocates that spillover effects and shocks in these markets are not persisted for a longer period of time, and lagged returns could be used for forecasting the future prices (Onour, 2010). In similar study, Tolvi (2003), observed 24% to 64% significant persistence behavior in Finnish stock markets returns.

In nutshell, a plenty of literary work is accessible in developed along with emerging markets. Moreover, there is a paucity of empirical observation on inspecting the presence of long range dependence in Indian context. The current study endeavors to plug this gap.

Research Design

Sample and Period of Study

The sample of the study has used daily data of Sectoral Indices namely; Nifty Media, Nifty Auto, Nifty Bank, Nifty Financial Services, Nifty FMCG, Nifty IT, Nifty Metal, Nifty Pharma, Nifty Private Bank, Nifty PSU, Nifty Reality, etc. from National Stock Exchange of India (NSEI). This article computed long term property for every year from January 2010 to May 2018 to ensure existence of persistence behavior occurred because of market microstructure, occasional events, dynamics behavior, market resistance, and peculiar characteristic etc.

Table 1: Description of Data

Sr. No	*Index*	*No. of Observations (N)*	*Period Covered*
1.	Nifty Media	2168	January, 2010 to May, 2018
2.	Nifty-Auto	2085	January, 2010 to May, 2018
3.	Nifty Bank	2086	January, 2010 to May, 2018
4.	Nifty Financial Services	2085	January, 2010 to May, 2018
5.	Nifty FMCG	2063	January, 2010 to May, 2018
6.	Nifty IT	2086	January, 2010 to May, 2018
7.	Nifty Metal	1459	July, 2011 to May, 2018
8.	Nifty Pharma	2087	January, 2010 to May, 2018
9.	Nifty Private Bank	2086	January, 2010 to May, 2018
10.	Nifty PSU	2078	January, 2010 to May, 2018
11.	Nifty Reality	2088	January, 2010 to May, 2018

Methodology

Hurst Exponent

To computing the long-term property, present study has used Hurst exponent. A Hurst exponent lies between 0 and 1. It deals with three types of patterns in economic series; persistence, randomness, or mean reversion. H= 0.5 means time series did not contain persistence behavior. There is no long term dependence, and the financial market might be efficient. When H is more than 0.5 and less than 1, time series is persistent, there is long-term memory. The market is inefficient indicating persistence effect. When H is greater than 0 and less than 0.5, there is a short-term memory and suggesting an anti- persistence effect (Tripathy, 2015).

The estimation of the Hurst exponent lies between 0 and 1:

$0.5 < H < 1$	Persistence
$H = 0.5$	Random walk
$0 < H < 0.5$	Anti-persistence

Multiple Break Point Test

The present study has used multiple break point test for checking the structural breaks in the data.

Table 2: Structural Breaks in the Data

Indices	*Number of Breaks*	*Break date*
Nifty Media	1	April 26, 2013
Nifty-Auto	2	May 25, 2015, and April 10, 2013
Nifty Bank	3	April 18, 2012, May 1, 2014 and February 11, 2016
Nifty Financial Services	1	February 5, 2010
Nifty FMCG	1	August 3, 2011
Nifty IT	2	August 6, 2014, February 10, 2016
Nifty Metal	1	June 4, 2014
Nifty Pharma	1	March 18, 2011
Nifty Private Bank	1	April 12, 2013
Nifty PSU	1	March 22, 2018
Nifty Reality	3	April 18, 2012, September 9, 2014, and March 15, 2016

Notes: This table reports the break date from January 2010 to May 2018 of Sectoral Indices of National Stock

Exchange of India and breaks date of each series estimated from Multiple Break-Point Test (Balcilar et; al, 2015).

Results and Analysis

One Year Analysis of Each Index

Table 3: Nifty Media

Nifty Media	*Coefficient of Hurst Exponent*	*Findings*
2010	0.6103	**Long Memory is Present**
2011	0.5292	**Long Memory is Present**
2012	0.6361	**Long Memory is Present**
2013	0.4546	Long Memory is Absent
2014	0.5207	**Long Memory is Present**
2015	0.5199	**Long Memory is Present**
2016	0.6330	**Long Memory is Present**
2017	0.5776	**Long Memory is Present**
2018(31st May)	0.5020	**Long Memory is Present**

Table 3 presents an analysis of long memory in Nifty Media from 2010 to 2018 by using Hurst exponent. Results advocate that long memory was exhibited in each year except 2013 with H value 0.4546.

Table 4: Nifty Auto

Nifty Auto	*Coefficient of Hurst Exponent*	*Findings*
2010	0.5693	**Long Memory is Present**
2011	0.5014	**Long Memory is Present**
2012	0.6519	**Long Memory is Present**
2013	0.5334	**Long Memory is Present**
2014	0.4816	Long Memory is Absent
2015	0.4498	Long Memory is Absent
2016	0.6210	**Long Memory is Present**
2017	0.4387	Long Memory is Absent
2018 (31st May)	0.6268	**Long Memory is Present**

The above table presents that long memory component was exhibited in Nifty Auto in 2010, 2011, 2012, 2013, 2016, 2018. But the similar behavior was not found in 2014 and 2015.

Table 5: Nifty Bank

Nifty Bank	*Coefficient of Hurst Exponent*	*Findings*
2010	0.5868	**Long Memory is present**
2011	0.4878	Long Memory is absent
2012	0.6024	**Long Memory is present**
2013	0.6166	**Long Memory is present**
2014	0.6270	**Long Memory is present**
2015	0.4499	Long Memory is absent
2016	0.6590	**Long Memory is present**
2017	0.5346	**Long Memory is present**
2018 (31st May)	0.7337	**Long Memory is present**

Table 5 present the findings of Nifty Bank that long memory were observed in 2010, 2012, 2013, 2014, 2016, 2017 and 2018. But in 2011 and 2015 persistence behavior not reported.

Table 6: Nifty Financial Services

Nifty Financial Services	*Coefficient of Hurst Exponent*	*Findings*
2010	0.5838	**Long Memory is Present**
2011	0.4786	Long Memory is Absent
2012	0.5926	**Long Memory is Present**
2013	0.6154	**Long Memory is Present**
2014	0.5996	**Long Memory is Present**
2015	0.4627	Long Memory is Absent
2016	0.6842	**Long Memory is Present**
2017	0.5522	**Long Memory is Present**
2018 (31st May)	0.6953	**Long Memory is Present**

Table 5 shows the result of Nifty Financial Services. Empirical evidence shows that long memory was reported in the year 2010, 2012, 2013, 2014, 2016, 2017, 2018. Similar behaviors were not found in 2011 and 2015.

Table 7: Nifty FMCG

Nifty FMCG	*Coefficient of Hurst Exponent*	*Findings*
2010	0.6121	**Long Memory is Present**
2011	0.5385	**Long Memory is Present**
2012	0.4756	Long Memory is Absent
2013	0.5581	**Long Memory is Present**
2014	0.4907	Long Memory is Absent

Nifty FMCG	Coefficient of Hurst Exponent	Findings
2015	0.4620	Long Memory is Absent
2016	0.6099	**Long Memory is Present**
2017	0.5787	**Long Memory is Present**
2018 (31st May)	0.6848	**Long Memory is Present**

Table 7 shows that long memory component were detected on the year 2010, 2011, 2013, 2016, 2017, 2018 except 2012, 2014 and 2015 with H value 0.4756, 0.4907, 0.4620 as respectively.

Table 8: Nifty IT

Nifty IT	Coefficient of Hurst Exponent	Findings
2010	0.5101	**Long Memory is Present**
2011	0.5788	**Long Memory is Present**
2012	0.5324	**Long Memory is Present**
2013	0.5984	**Long Memory is Present**
2014	0.6128	**Long Memory is Present**
2015	0.5101	**Long Memory is present**
2016	0.4990	Long Memory is Absent
2017	0.4513	Long Memory is Absent
2018 (31st May)	0.6113	**Long Memory is Present**

The above table presents the result of Nifty IT, advocates that persistence behavior were found in 2010, 2011, 2012, 2013, 2014, 2015 and 2018. But similar behavior was not detected in 2016 and 2017 with H value 0.4990, 0.4513 as respectively.

Table 9: Nifty Metals

Nifty Metal	Coefficient of Hurst Exponent	Findings
2011	0.5168	**Long Memory is Present**
2012	0.5510	**Long Memory is Present**
2013	0.7258	**Long Memory is Present**
2014	0.7277	**Long Memory is Present**
2015	0.5321	**Long Memory is Present**
2016	0.5204	**Long Memory is Present**
2017	0.5449	**Long Memory is Present**
2018 (31st May)	0.6298	**Long Memory is Present**

The above table shows that long memory was exhibited for Nifty Metal from 2011 to 2018 in each year, which indicates the serial correlation behavior in return series.

Table 9: Nifty Pharma

Nifty Pharma	Coefficient of Hurst Exponent	Findings
2010	0.5860	**Long Memory is Present**
2011	0.5489	**Long Memory is Present**
2012	0.5081	**Long Memory is Present**
2013	0.4914	Long Memory is Absent
2014	0.6551	**Long Memory is Present**
2015	0.5127	**Long Memory is Present**
2016	0.4773	Long Memory is Absent
2017	0.4979	Long Memory is Absent
2018 (31st May)	0.6023	**Long Memory is present**

The above table presents that the results of Nifty Pharma, finding shows that long memory behavior was exhibited for 2010, 2011, 2012, 2014, 2015 and 2018. While persistence behavior was not observed for the year 2013, 2016, 2017 and 2018 with H value 0.4914, 0.4773 and 0.4979.

Table 10: Nifty Private Bank

Nifty Private Bank	Coefficient of Hurst Exponent	Findings
2010	0.5292	**Long Memory is Present**
2011	0.4978	Long Memory is Absent
2012	0.6060	**Long Memory is Present**
2013	0.6237	**Long Memory is Present**
2014	0.5957	**Long Memory is Present**
2015	0.4813	Long Memory is Absent
2016	0.6683	**Long Memory is Present**
2017	0.5396	**Long Memory is Present**
2018 (31st May)	0.7240	**Long Memory is Present**

The above table presents the results of Nifty Private, indicates that long memory was detected in 2010, 2012, 2013, 2014, 2015, 2016, 2017 and 2018. But Long memory was not observed in 2011 and 2015.

Table 11: Nifty PSU

Nifty PSU Bank	*Coefficient of Hurst Exponent*	*Findings*
2010	0.6269	**Long Memory is Present**
2011	0.4927	Long Memory is Absent
2012	0.6167	**Long Memory is Present**
2013	0.6380	**Long Memory is Present**
2014	0.7245	**Long Memory is Present**
2015	0.4372	Long Memory is Absent
2016	0.5555	**Long Memory is Present**
2017	0.5305	**Long Memory is Present**
2018 (31st May)	0.6395	**Long Memory is Present**

The above table presents that long memory were exhibited in 2010, 2012, 2013, 2014, 2016, 2017 and 2018, while in 2011 and 2015 similar behavior was not detected with H value 0.4927, 0.4372 as respectively.

Table 12: Nifty Reality

Nifty Reality	*Coefficient of Hurst Exponent*	*Findings*
2010	0.6110	**Long Memory is Present**
2011	0.5232	**Long Memory is Present**
2012	0.6504	**Long Memory is Present**
2013	0.5543	**Long Memory is Present**
2014	0.6868	**Long Memory is Present**
2015	0.5378	**Long Memory is Present**
2016	0.6657	**Long Memory is Present**
2017	0.5612	**Long Memory is Present**
2018 (31st May)	0.5577	**Long Memory is Present**

Table 12 presents the results of Nifty Reality that persistence behavior was reported from 2010 to 2018 in each year.

Two Year Analysis of Each Index

Table 13: Nifty Media

Nifty Media	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	**0.5697**	**Long Memory is Present**
2011-2012	**0.5826**	**Long Memory is Present**
2012-2013	**0.5826**	**Long Memory is Present**
2013-2014	0.4876	Long Memory is Absent
2014-2015	**0.5203**	**Long Memory is Present**
2015-2016	0.5764	**Long Memory is Present**
2016-2017	0.6053	**Long Memory is Present**
2017-2018	0.5398	**Long Memory is Present**

In two year analysis, of Nifty Media for long memory indicates that persistence behavior exhibited in all the analysis except 2013-2014 with 0.4876.

Table 14: Nifty Auto

Nifty Auto	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5353	**Long Memory is Present**
2011-2012	0.5766	**Long Memory is Present**
2012-2013	0.5926	**Long Memory is Present**
2013-2014	0.5075	**Long Memory is Present**
2014-2015	0.4657	Long Memory is Absent
2015-2016	0.5354	**Long Memory is Present**
2016-2017	0.5298	**Long Memory is Present**
2017-2018	0.5327	**Long Memory is Present**

In two year analysis of long memory, the finding indicates that persistence behavior was reported for 2010-2011, 2011-2012, 2012-2013,2013-2014, 2015-2016, 2016-2017 and 2017-2018. But did not observe long memory for 2014-2015.

Table 15: Nifty Bank

Nifty Bank	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5373	**Long Memory is Present**
2011-2012	0.5451	**Long Memory is Present**
2012-2013	0.6095	**Long Memory is Present**
2013-2014	0.6218	**Long Memory is Present**
2014-2015	0.5384	**Long Memory is Present**
2015-2016	0.5544	**Long Memory is Present**
2016-2017	0.5968	**Long Memory is Present**
2017-2018	0.6341	**Long Memory is Present**

The above table, two-year analysis indicates that persistence behaviors were strongly observed in all years.

Table 16: Nifty Financial Services

Nifty Financial Services	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5312	**Long Memory is Present**
2011-2012	0.5356	**Long Memory is Present**
2012-2013	0.6040	**Long Memory is Present**
2013-2014	0.6075	**Long Memory is Present**
2014-2015	0.5311	**Long Memory is Present**
2015-2016	0.5734	**Long Memory is Present**
2016-2017	0.6182	**Long Memory is Present**
2017-2018	0.6237	**Long Memory is Present**

In two-year analysis of Nifty Financial Services, indicates that long memory exhibited in all the year from 2010-2018.

Table 17: Nifty FMCG

Nifty FMCG	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5753	**Long Memory is Present**
2011-2012	0.5070	**Long Memory is Present**
2012-2013	0.5168	**Long Memory is Present**
2013-2014	0.5244	**Long Memory is Present**
2014-2015	0.4763	Long Memory is Absent
2015-2016	0.5359	**Long Memory is Present**
2016-2017	0.5943	**Long Memory is Present**
2017-2018	0.6317	**Long Memory is Present**

Above table presents the two-year long memory analysis, indicates that persistence behavior was reported each year except 2014-2015.

Table 18: Nifty IT

Nifty IT	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5444	**Long Memory is Present**
2011-2012	0.5556	**Long Memory is Present**
2012-2013	0.5654	**Long Memory is Present**
2013-2014	0.6056	**Long Memory is Present**
2014-2015	0.5614	**Long Memory is Present**
2015-2016	0.5045	**Long Memory is Present**
2016-2017	0.4751	Long Memory is Absent
2017-2018	0.5313	**Long Memory is Present**

Above table presents the two-year analysis, indicates that Long memory component is reported in every year except 2016-2017.

Table 19: Nifty Metal

Nifty Metal	*Coefficient of Hurst Exponent*	*Findings*
2011-2012	0.5339	**Long Memory is present**
2012-2013	0.6384	**Long Memory is present**
2013-2014	0.72675	**Long Memory is present**
2014-2015	0.6299	**Long Memory is present**
2015-2016	0.5262	**Long Memory is present**
2016-2017	0.5326	**Long Memory is present**
2017-2018	0.5873	**Long Memory is present**

Above table shows that long memory was strongly exhibited in Nifty Metal return series.

Table 20: Nifty Pharma

Nifty Pharma	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5674	**Long Memory is Present**
2011-2012	0.5285	**Long Memory is Present**
2012-2013	0.4997	Long Memory is Absent
2013-2014	0.5732	**Long Memory is Present**
2014-2015	0.5839	**Long Memory is Present**
2015-2016	0.4950	Long Memory is Absent
2016-2017	0.4876	Long Memory is Absent
2017-2018	0.5501	**Long Memory is Present**

Above table shows that long memory was exhibited in 2010-2011,2011-2012, 2013-2014,2014-2015, 2017-2018. But similar findings did not observe in 2012-2013, 2015-2016 and 2016-2017.

Table 21: Nifty Private Bank

Nifty Private Bank	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5135	**Long Memory is Present**
2011-2012	0.5519	**Long Memory is Present**
2012-2013	0.6148	**Long Memory is present**
2013-2014	0.6097	**Long Memory is present**

Nifty Private Bank	*Coefficient of Hurst Exponent*	*Findings*
2014-2015	0.5385	**Long Memory is present**
2015-2016	0.5748	**Long Memory is Present**
2016-2017	0.6039	**Long Memory is Present**
2017-2018	0.6318	**Long Memory is Present**

Above table presents that long memory was observed in all whole the analysis of Nifty Private Bank series.

Table 22: Nifty PSU

Nifty PSU Bank	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5598	**Long Memory is Present**
2011-2012	0.5547	**Long Memory is Present**
2012-2013	0.6273	**Long Memory is Present**
2013-2014	0.6812	**Long Memory is Present**
2014-2015	0.5808	**Long Memory is Present**
2015-2016	0.4963	Long Memory is Absent
2016-2017	0.543	**Long Memory is Present**
2017-2018	0.585	**Long Memory is Present**

In two year analysis of Nifty PSU shows that persistence behavior was exhibited in the whole sample except 2015-2016.

Table 23: Nifty Reality

Nifty Reality	*Coefficient of Hurst Exponent*	*Findings*
2010-2011	0.5671	**Long Memory is Present**
2011-2012	0.5868	**Long Memory is Present**
2012-2013	0.6023	**Long Memory is Present**
2013-2014	0.6205	**Long Memory is Present**
2014-2015	0.6123	**Long Memory is Present**
2015-2016	0.6017	**Long Memory is Present**
2016-2017	0.6134	**Long Memory is Present**
2017-2018	0.5594	**Long Memory is Present**

Above table presents that long memory was strongly observed in the entire sample of Nifty Reality.

Table: 24: (Full period Analysis of All Indices)

Sr. No	*Period Covered*	*Indices*	*Coefficient of Hurst Exponent*	*Findings*
1.	January 1, 2010-May 31,2018	Nifty Media	0.5537	**Long memory is present**
2.	January 1, 2010-May 31,2018	Nifty Auto	0.5415	**Long memory is present**
3.	January 1, 2010-May 31,2018	Nifty Bank	0.5886	**Long Memory is present**
4.	January 1, 2010-May 31,2018	Nifty Financial Service	0.5849	**Long Memory is present**
5.	January 1, 2010-May 31,2018	Nifty FMCG	0.5567	**Long Memory is present**
6.	January 1, 2010-May 31,2018	Nifty IT	0.5449	**Long Memory is present**
7.	January 1,2011-May 31,2018	Nifty Metal	0.5935	**Long Memory is present**
8.	January 1, 2010-May 31,2018	Nifty Pharma	0.5421	**Long Memory is present**
9.	January 1, 2010-May 31,2018	Nifty Private Bank	0.4972	Long Memory is absent
10.	January 1, 2010-May 31,2018	Nifty PSU	0.5846	**Long Memory is present**
11.	January 1, 2010-May 31,2018	Nifty Reality	0.5942	**Long Memory is present**

The above Table presents the full period analysis of Long term memory in sectoral indices of the National Stock Exchange of India by using Rescaled Range Statistics (Hurst Exponent, 1951). All Indices have shown long memory component in returns series. However, Nifty Private Bank does not contain the persistence behavior in returns series.

Conclusion

In this paper, the presence of long memory in sectoral indices of National Stock Exchange of India has been examined by using the Rescaled Range Analysis. (Hurst Exponent,1951). Analysis for the full sample period indicates that all sectoral indices show long memory effect (2010-2018) except Nifty Private Bank with H value 0.4972. On examining the data for individual years, all indices exhibit long memory effect, except Nifty Media in 2013 with H value 0.4546, Nifty Auto in 2014, 2015, and 2017 with H value 0.4816, 0.4498 and 0.4387, Nifty Bank on 2011 and 2015 with H value 0.4878 and 0.4499, Nifty Financial Services in 2011 and 2015 with H value 0.4786 and 0.4627, Nifty FMCG in 2012, 2014 and 2015 with H value 0.4756, 0.4907 and 0.4620, Nifty IT in 2016 and 2017 with H value 0.4990 and 0.4513. Similar findings are reported for Nifty Pharma in 2013, 2016 and 2017 with H value 0.4914, 0.4773 and 0.4979, Nifty Private Bank in 2011 and 2015 with H value 0.4978 and 0.4813 and Nifty PSU in 2011 and 2015 with H value 0.4927 and 0.4372, where they do not indicate persistence behavior. Furthermore, Nifty Metal (2011-2018) and Nifty Reality (2010-2018) reported long memory component for each year with respect of Hurst Exponent coefficient. Moreover, in the two-year analysis, Nifty Bank, Nifty Financial Services, Nifty Metal, Nifty Private Bank, Nifty Reality are significantly reported the persistence behavior of sectoral indices of NSE. With reference to structural break dates, it can be inferred that all returns series from 2010 to 2018 reported a significant long memory effect whereby confirming the significant persistence behavior after the occurrence of break date in the data Findings of Hurst Exponent indicates that market exhibits a degree of persistence, which means "mean" reverting nature is slower than a random walk. Evidence suggests that Indian equity market returns pursue serial correlation with the presence of long range dependence. There is little evidence that market persists random walk process. The findings would be useful to investor, broker, mutual funds, etc., in undertaking the investment decisions.

References

Barkoulas, J. T., & Baum, C. F. (1996). Long-term dependence in stock returns. *Economics Letters*, *53*(3), 253–259.

Bali, T. G., Peng, L., Shen, Y., & Tang, Y. (2014). Liquidity shocks and stock market reactions. *Review of Financial Studies*, *27*(5), 1434–1485.

Balcilar, M., Ozdemir, Z. A., & Cakan, E. (2015). Structural breaks, long memory, or unit roots in stock prices: Evidence from emerging markets. *International Econometric Review (IER)*, *7*(1), 13–33.

Barkoulas, J. T., Baum, C. F., & Travlos, N. (2000). Long memory in the Greek stock market. *Applied Financial Economics*, *10*(2), 177–184.

Baillie, R. T., & Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control*, *33*(8), 1577–1592.

Backus, D., & Zin, S. E. (1993). *Long-memory inflation uncertainty: Evidence from the term structure of interest rates*. Working paper No.133, National Bureau of Economic Research.

Bhattacharya, S. N., & Bhattacharya, M. (2012). Long memory in stock returns: A study of emerging markets. *Iranian Journal of Management Studies*, *5*(5), 67–88.

Badhani, K. N. (2012). Does Nifty have a Long Memory? Semi-parametric Estimation of Fractional Integration in Returns and Volatility. *Decision*, *39*(3), 86-100.

Badhani, K. N. (2008). Long memory in stock returns and volatility in India: A nonparametric analysis. *ICFAI Journal of Applied Finance*, *14*(12), 34–53.

Baum, C. F., Barkoulas, J. T., & Caglayan, M. (1999). Long memory or structural breaks: can either explain nonstationary real exchange rates under the current float? *Journal of International Financial Markets, Institutions and Money*, *9*(4), 359–376.

Barkoulas, J. T., & Baum, C. F. (1997). Long memory and forecasting in Euroyen deposit rates. *Financial Engineering and the Japanese Markets*, *4*(3), 189–201.

Booth, G. G., & Tse, Y. (1995). Long memory in interest rate futures markets: A fractional cointegration analysis. *Journal of Futures Markets*, *15*(5), 573–584.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of econometrics*, *73*(1), 5–59.

Bala, & Gupta (2018) Testing the long memory feature in Indian equity market. *Journal of Research in Management,10*(2), 24–33.

Chung, H., Lin, W. T., & Wu, S. (2000). An analysis of long memory in volatility for Asian stock markets. *Review of Pacific Basin Financial Markets and Policies*, *3*(3), 309–330.

Chen, J. H., & Diaz, J. F. (2013). Long memory and shifts in the returns of green and non-green Exchange-Traded Funds (ETFs). *International Journal of Humanities and Social Science Invention*, *2*(10), 29–32.

Cont, R. (2005). Long range dependence in financial markets. In *Fractals in Engineering* (pp. 159–179). London: Springer.

Christodoulou-Volos, C., and Siokis, F. M. (2006). Long range dependence in stock market returns. *Applied Financial Economics*, *16*(18), 1331–1338.

Caporale, G. M., & Gil-Alana, L. A. (2008). Long memory and structural breaks in the Spanish stock market index. *Open Operational Research Journal*, *2*, 13–17.

Cheung, Y. W., & Lai, K. S. (1995). A search for long memory in international stock market returns. *Journal of International Money and Finance*, *14*(4), 597–615.

Chen, J. H., & Huang, Y. F. (2014). Long memory and structural breaks in modelling the volatility dynamics of VIX-ETFs. *International Journal of Business, Economics and Laws*, *4*(1), 54-63.

Cajueiro, D. O., & Tabak, B. M. (2005). Possible causes of long-range dependence in the Brazilian stock market. *Physica A: Statistical Mechanics and its Applications*, *345*(3), 635–645.

Cevik, P., & Emec, H. (2013). Long memory properties in return and volatility: An application of the impact of arab spring in turkey financial market. *Current Research Journal of Social Sciences*, *5*(2), 60–66.

Dicle, M. F., Beyhan, A., & Yao, L. J. (2010). Market efficiency and international diversification: Evidence from India. *International Review of Economics & Finance*, *19*(2), 313–339.

Caccioli, F., & Marsili, M. (2010). Information efficiency and financial stability. *Economics: The Open-Access, Open-Assessment E-Journal*, *4*(2010–20), 1-20.

Figlewski, S. (1978). Market "efficiency" in a market with heterogeneous information. *The Journal of Political Economy*, 581–597.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance*, *46*(5), 1575–1617.

Gayathri, M., Murugesan, S., & Gayathri, J. (2012). Persistence and long range dependence in Indian stock market returns. *International Journal of Management and Business Studies*, *2*(4), 72–77.

Granger, C. W., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S and P 500 absolute stock returns. *Journal of Empirical Finance*, *11*(3), 399–421.

Gurgul, H., & Wojtowicz, T. (2006). Long memory on the German stock exchange. *Czech Journal of Economics and Finance (Finance a uver)*, *56*(9–10), 447–468.

Greene, M. T., & Fielitz, B. D. (1977). Long-term dependence in common stock returns. *Journal of Financial Economics*, *4*(3), 339–349.

Ho, T. S., Schwartz, R. A., & Whitcomb, D. K. (1985). The trading decision and market clearing under transaction price uncertainty. *Journal of Finance*, *40*(1) 21–42.

Hassler, U., & Wolters, J. (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics*, *13*(1), 37–45.

Hurst, H. E. (1951) *The long-term storage capacity of reservoir*. Transactions of the American Society of Civil Engineers 116, Paper 2447, Published in 1950 as Proceedings-Separate No. 11.

Hiremath, G. S., & Kamaiah, B. (2011). Testing long memory in stock returns of emerging markets: Some further evidence. *Economics, Management, and Financial Markets*, *6*(3), 136–147.

Hiremath, G. S., & Kamaiah, B. (2010). Long memory in stock market volatility: Indian evidences. *Artha Vijnana*, *52*(4), 332–345.

Heidarpoor, F., Rouhi, A., & Mashaei, K. F. (2012). Relationship between stock liquidity and stock price changes rate. *African Journal of Business Management*, *6*(37), 10178–10184.

Huang, B. N., & Yang, C. W. (1999). *An examination of long-term memory using the intraday stock returns*.

Technical Report 99–03. Clarion University of Pennsylvania, Clarion.

Jennings, R. H., & Barry, C. B. (1983). Information dissemination and portfolio choice. *Journal of Financial and Quantitative Analysis*, *18*(1), 1–19.

Joshipura, M. (2008). Price and liquidity effects of stock split: Empirical evidence from Indian stock market. *NSE Research initiative*.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and quantitative Analysis*, *22*(1), 109–126.

Kumar, A. (2004). *Long memory in stock trading volume: Evidence from Indian stock market*. This paper has been downloaded from Social Science Research Network (SSRN) on May 19, 2015.

Kumar, D. (2014). Long-range dependence in Indian stock market: A study of Indian sectoral indices. *International Journal of Emerging Markets*, *9*(4), 505–519.

Kasman, A., Kasman, S., & Torun, E. (2009). Dual long memory property in returns and volatility: Evidence from the CEE countries' stock markets. *Emerging Markets Review*, *10*(2), 122–139.

Kasman, A., & Torun, E. (2012). Long memory in the Turkish stock market return and volatility. *Central Bank Review*, *7*(2), 13–27.

Kasman, S., Turgutlu, E., & Ayhan, A. D. (2009). Long memory in stock returns: Evidence from the major emerging Central European stock markets. *Applied Economics Letters*, *16*(17), 1763–1768.

Kang, S. H., & Yoon, S. M. (2007). Long memory properties in return and volatility: Evidence from the Korean stock market. *Physica A: Statistical Mechanics and its Applications*, *385*(2), 591–600.

Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, *77*(2), 411–452.

Lo, A. (1991). Long term memory in stock market prices. *Econometrica, 59*(5), 1279–1313.

Limam, I. (2003). Is long memory a property of thin stock markets? International Evidence using Arab countries. *Review of Middle East Economics and Finance*, *1*(3), 251–266.

Mukherjee, I., Sen, C., & Sarkar, A. (2011). Long memory in stock returns: Insights from the Indian market. *The International Journal of Applied Economics and Finance*, *5*(1), 62–74.

Nath, G. C., & Dalvi, M. (2004). *Day of the week effect and market efficiency-evidence from Indian equity market using high frequency data of national stock exchange*. This paper has been Downloaded from Social Science Research Network on June 2nd, 2015.

Nath, G. C., & Reddy, Y. V. (2002, December). Long memory in rupee-dollar exchange rate: An empirical study. In *Capital Market Conference*.

Onour, I. A. (2010). North Africa stock markets: Analysis of long memory and persistence of shocks. *International Journal of Monetary Economics and Finance*, *3*(2), 101–111.

Prasanna, K., & Bansal, B. (2014). Foreign institutional investments and liquidity of stock markets: Evidence from India. *International Journal of Economics and Finance*, *6*(6), 103–118.

Raj, M., & Kumari, D. (2006). Day-of-the-week and other market anomalies in the Indian stock market. *International Journal of Emerging Markets*, *1*(3), 235–246.

Pati, P. C., & Rajib, P. (2011). Intraday return dynamics and volatility spillovers between NSE S&P CNX Nifty stock index and stock index futures. *Applied Economics Letters*, *18*(6), 567–574.

Souza, S. R., Tabak, B. M., & Cajueiro, D. O. (2008). Long-range dependence in exchange rates: the case of the European monetary system. *International Journal of Theoretical and Applied Finance*, *11*(2), 199–223.

Souza, S. R., Tabak, B. M., & Cajueiro, D. O. (2008). Long memory testing for Fed Funds Futures' contracts. *Chaos, Solitons & Fractals*, *37*(1), 180–186.

Suresha, B. (2014). Ownership structure and market Liquidity-Sectorial Evidence from India. *Research Journal of Finance and Accounting*, *5*(19), 127–133.

Sadique, S., & Silvapulle, P. (2001). Long-term memory in stock market returns: International evidence. *International Journal of Finance & Economics*, *6*(1), 59–67.

Soofi, A. S., Wang, S., and Zhang, Y. (2006). Testing for long memory in the Asian foreign exchange rates. *Journal of Systems Science and Complexity*, *19*(2), 182–190.

Sensoy, A. (2013). Time-varying long range dependence in market returns of FEAS members. *Chaos, Solitons and Fractals*, *53*, 39–45.

Taqqu, M. S., Teverovsky, V., & Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals*, *3*(4), 785–798.

Turkyilmaz, S., & Balibey, M. (2014). Long memory behavior in the returns of Pakistan stock market: ARFIMA-FIGARCH models. *International Journal of Economics and Financial Issues*, *4*(2), 400–441.

Tayal, R., & Thomas, S. (2012). *Measuring and explaining the asymmetry of liquidity*. This paper has been downloaded from Social Science Research Network (SSRN) on May19, 2015.

Verma, A. (2008). Long memory of the Indian stock market. *The IUP Journal of Financial Economics*, *6*(3), 74–83.

Willinger, W., Taqqu, M. S., & Teverovsky, V. (1999). Stock market prices and long-range dependence. *Finance and stochastics*, *3*(1), 1–13.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 36-45
http://publishingindia.com/ijbai/

Impact of Non-Performing Assets on Operational Performance of Foreign Banks in India and Macro-Economic Determinants

M. Jegadeeshwaran*, M. Basuvaraj**

Abstract

Banks are prime intermediaries in mobilizing resources and channelling resources to various sectors of the economy, free and adequate flow of bank credit has a positive impact on the growth of the sector and directly contributes towards increased national income, national production and employment. The main objective of the study is to examine the Impact of non-performing assets on operational performance of foreign banks and macro-economic variables. The study period covers ten years from 2007-08 to 2016-17. For the purpose of the study, ten banks have been selected based on the banks which have the highest share in Non-performing Assets. The banks are Citi Bank, Standard Chartered Bank, HSBC In India, Deutsche Bank, Development Bank of Singapore, Barclays Bank, AB Bank, SBM Bank (MAURITIUS) Ltd, CTBC Bank, Abu Dhabi Commercial Bank. The data analysis is done using ratio analysis, descriptive statistics like mean, standard deviation, coefficient of variation, compound annual growth rate, multiple regression and Ordinary least squire. The study concludes that the banks should take essential steps to curtail the mounting NPAs. This will allow the overall development of the economy and carry hope among the investors across the world in the Indian economy.

Keywords: Foreign Banks, Descriptive Statistics, Operational Performance, Non-performing Assets and Macro Economic Variables

Introduction

The British initiated the process of setting up of foreign banks in India after 1850. Foreign banks are those banks whose branch offices are in India, and have their head office in a foreign country. The banks were allowed to set up their subsidiaries in India from the year 2002. Foreign banks play a relatively minor role in the Indian economy, as reiterated in Global Development finance 2008. Globalization has compelled the banking sector to reach out to more customers in order to expand their business, especially in the priority sectors. They are opening banking business even in foreign countries. During the war period the turnover of foreign banks was affected adversely due to the lower demand for credit and reduction in general industrial activity. The operational performances of the foreign banks have measured through the ratio analysis and descriptive statistic for the study. The profitability ratios has been used in order to assess the relative efficiency of the foreign banks, different ratios of return on total assets and various heads of expenditure to total assets of foreign banks for the study.

Operational Performance of Select Foreign Banks in India

(i) Gross profit to total assets (GP to TA), (ii) Net return on total assets (NR to TA), (iii) Interest income to total assets (I I to TA), (iv) Interest expended to total assets (IE to TA), (v) Net interest income or margin (Spread) to total assets (NIM to TA), (vi) Provisions and contingencies to total assets (P&C to TA), (vii) Operating expenses to total assets (OE to TA), (viii) Capital adequacy ratio (CAR).

Review of Literature

Vani Shree Sah (2017) in their article attempted to study the NPAs of Indian Banks. NPAs have impact on bank's

* Assistant Professor, Department of Commerce, Bharathiar University, Coimbatore, Tamil Nadu, India. Email:
** Senior Research Fellow, Department of Commerce, Bharathiar University, Coimbatore, Tamil Nadu, India. Email: basuvarajm93@gmail.com

the profitability and the net-worth. The banks must to take initiatives to reduce the NPAs. Gross NPA reflects the quality of the bank loans and the Net NPA reflects the bank's actual burden. It was concluded that the trends of gross NPAs, net NPAs are impact of NPAs and the recent government actions to reduce the NPAs.

Prasanth Kiran and Marry Jones (2016) in their study entitled to "Effect of NPAs on the profitability of banks-A selective study". It has been made on public sector banks in India to assess the effect of NPAs on the profitability of banks. For this study, SBI and 5 nationalized banks were selected and the relation between their gross NPAs and net profit was considered. The result highlights that except for SBI all the other banks demonstrate a negative correlation between their gross NPAs and net profits. But for SBI the net profit is not affected by gross NPAs and also it is in continuous profits. These days government is focused in mounting economy which wants vast financial resources.

Vivek Rajbagadur Singh (2016) in his study investigated the NPAs in Scheduled Commercial Banks which includes public sector banks, private sector banks and foreign banks which are listed in the second schedule of the Reserve Bank of India Act, 1934. It is based on secondary data. The conceptual framework of NPAs and it also highlights the trends, status and impact of NPAs on scheduled commercial banks during the period of 14 years i.e. from 2000 to 2014. The NPAs level of the banks is high when compared to the foreign banks. It is not possible to bring as zero NPAs and also study suggests to the bank management should speed up the recovery process.

Statement of the Problem

India has been growing in the last few years even as lenders have come under presser due to unpaid dues. Non-performing assets is an emerging issue in banking system all over the world. Especially in foreign banks in India have facing NPAs risk and also foreign banks are reducing their branches count. They are facing stiff competition from the all other banks and the banks will burden of growing NPAs if the capital becomes low. If there is very demand for credit and banks lend indiscriminately. The problem of higher NPAs persists for years without getting recognized as well as hampering normal bank lending. In 2015-16, the number of foreign banks reduced from 46 to 44 last year. Spectra of NPAs have continued to haunt the foreign banks in India meanwhile, 50 per cent of foreign bank respondents stated their bad loans is increased In the terms of profit the banks it's weighed down by higher provision and a write-off of NPAs which had risen sharply in the last year. Bad loans as a per cent of total loans rose from 10.2 per cent in 2017. The NPAs could go worsen it will immense economy, if the macroeconomic conditions deteriorated, among the bank groups foreign banks have the highest provision coverage with 88.7 per cent when compared to other sector banks. The health of the banks will be affected if the growth of NPAs is accumulated and the subsequent NPAs resolution has effect on the bank capital as well as the profitability will be affected in terms of operational performance of banks. The study focused on the status of NPAs of foreign banks in India, recovery of NPAs and operational performance of NPAs. The following research questions are,

- What extent effect of the NPAs on operational performance and macroeconomic variables?

Objective of the Study

- To examine the impact of non-performing assets on operational performance of foreign banks and macro-economic determinants.

Hypothesis of the Study

- There is no significant impact of Net NPAs to total assets on Operational Performance of select foreign banks in India and macro-economic determinants.

Methodology

- The present study is an analytical in nature.

Sources of Data

The study is based on secondary data only. The relevant data have been collected from the RBI publications like "Annual Report on Trends and Progress of Banking in India", 'Annual Report of RBI', and the publications of RBI like RBI bulletin, IBA (Indian Banks' Association), websites and magazines.

Period of the Study

The present study covers a period of the ten years from 2007-08 to 2016-17.

Sampling Design

There are 44 foreign banks in India. In this sector, ten banks have been selected on the basis of the level of NPAs for the study. The ten banks are Citi Bank, Standard Chartered Bank, HSBC In India, Deutsche Bank, Development Bank of Singapore, Barclays Bank, AB Bank, SBM Bank (MAURITIUS) Ltd, CTBC Bank, Abu Dhabi Commercial Bank are taken for the study.

Statistical Tools for Analysis

The analysis of data are done using statistical tools like, Mean, standard deviation, coefficient of variation, Compound annual growth rate, Augmented dickey fuller test and ordinary Least Squire.

Table 1: Descriptive Statistics of Operational Performance of Select Foreign Banks in India

Gross profit to total assets of select foreign banks										
Banks	*AB*	*ABU*	*BARCLAYS*	*CITI*	*CTBC*	*DBS*	*DEUTSCHE*	*HSBC*	*SBM*	*SCB*
MEAN	8.30	1.71	2.29	5.18	1.98	1.73	4.18	3.16	2.12	4.05
SD	1.79	0.77	1.24	3.93	1.07	0.81	0.88	0.72	1.30	0.41
CV	21.54	45.26	54.31	75.96	53.95	46.58	21.07	22.80	61.33	10.07
CAGR	-0.01	-0.12	0.10	0.00	0.02	-0.01	-0.01	-0.02	-0.07	-0.02
Net Return to total assets of select foreign banks										
MEAN	8.30	1.71	2.29	5.18	1.98	1.73	4.18	3.16	2.12	4.05
SD	1.79	0.77	1.24	3.93	1.07	0.81	0.88	0.72	1.30	0.41
CV	21.54	45.26	54.31	75.96	53.95	46.58	21.07	22.80	61.33	10.07
CAGR	-0.01	-0.12	0.10	0.00	0.02	-0.01	-0.01	-0.02	-0.07	-0.02
Net Return to total assets of select foreign banks										
MEAN	4.37	1.40	0.28	1.90	0.25	0.58	1.95	1.47	0.12	1.85
SD	1.24	1.41	1.31	0.51	1.98	0.80	0.55	0.33	1.84	0.57
CV	28.25	100.68	468.54	26.85	778.55	138.00	28.62	22.45	569.54	30.71
CAGR	-0.01	-0.21	0.43	0.01	0.01	-0.27	-0.02	0.02	0.06	-0.03
Interest income to total assets of select foreign banks										
MEAN	4.22	7.28	6.73	6.88	7.98	6.19	7.22	6.42	8.43	7.14
SD	0.94	1.16	2.22	0.59	1.61	1.03	0.69	0.65	1.53	0.61
CV	22.25	16.05	32.92	8.50	20.16	16.71	9.58	10.17	18.15	8.52
CAGR	-0.07	-0.04	-0.04	-0.01	0.00	-0.04	0.00	-0.02	-0.02	-0.01
Interest expended to total assets of select foreign banks										
MEAN	0.63	3.97	2.97	2.52	2.62	3.45	2.25	2.67	4.85	3.06
SD	0.19	0.91	0.73	0.37	0.95	0.88	0.67	0.39	1.82	0.74
CV	30.56	22.94	24.69	14.64	36.40	25.57	29.88	14.52	37.60	24.19
CAGR	0.03	0.00	0.01	-0.01	-0.05	0.00	0.03	0.01	-0.03	0.00
Net interest income to total assets (net interest margin) of select foreign banks										
MEAN	3.56	2.98	3.44	4.26	5.18	2.44	4.82	3.64	2.86	4.10
SD	0.94	1.28	1.26	0.33	1.30	0.64	0.78	0.49	0.97	0.25
CV	26.31	42.99	36.69	7.70	25.17	26.28	16.27	13.45	34.10	6.04
CAGR	-0.08	-0.13	-0.03	-0.01	0.05	-0.06	-0.01	-0.02	-0.02	-0.01
Provisions and contingencies to total assets of select foreign banks										
MEAN	3.95	0.69	1.76	2.00	2.29	1.14	2.07	1.64	2.52	2.15
SD	0.97	0.51	1.16	0.59	1.75	0.42	0.45	0.84	1.82	0.50
CV	24.47	74.54	65.84	29.44	76.48	36.96	21.81	51.02	72.43	23.22
CAGR	0.00	-0.14	0.02	-0.02	0.26	0.07	0.00	-0.05	0.10	-0.01

Operating expenses to total assets of select foreign banks										
MEAN	6.41	1.79	2.69	3.06	3.62	1.42	3.19	2.26	1.07	2.50
SD	1.41	0.99	1.31	2.12	1.34	0.57	0.90	0.24	0.31	0.22
CV	22.02	55.24	48.56	69.28	37.04	40.20	28.23	10.43	29.11	8.98
CAGR	0.07	-0.13	-0.12	0.00	0.05	0.07	-0.08	-0.03	0.00	-0.03
Capital adequacy ratio of selected Foreign Banks										
MEAN	32.86	45.09	17.86	15.78	39.01	15.91	14.96	16.21	41.48	12.25
SD	8.61	18.21	1.97	1.91	7.58	1.87	0.81	2.35	6.24	1.02
CV	26.20	40.39	11.04	12.12	19.44	11.72	5.43	14.49	15.04	8.29
CAGR	-0.03	-0.07	-0.02	0.04	0.07	-0.01	0.00	0.06	-0.02	0.03

Source: RBI Published Data and Annual Reports

Impact of Non-Performing Assets on Operational Performance of Foreign Banks

Relationship between Net NPAs to Total Assets And Other Dependent Variables of Foreign banks

The following table describes the relationship between independent variable like Net Non-performing assets to total assets and other dependent variables like gross profit to total assets, net return to total assets, interest income to total assets, net interest income to total assets provision and contingencies to total assets, operational expenses to total assets and capital adequacy ratio using Regression for foreign banks.

Relationship between Net NPAs to Operating Performance of Select Foreign Banks.

H_0 – There is no significant impact of Net NPAs to total assets on Operational Performance of select foreign banks.

Table 2: Model Summary of Net NPAs to Total Assets on Operational Performance of Select Foreign Banks

Ratios	*R*	*R Square*	*Adjusted R Square*	*Std. Error*	*F*	*Sig.*
Net NPAs to total asset on gross profit to total assets	.38	.14	.04	.145	1.37	0.27
Net NPAs to total assets on net return to total assets	.838	.701	.664	.10472	18.793	.002*
Net NPAs to total assets and interest income to total assets	.229	.053	-.066	.03310	.444	0.524
Net NPAs to total assets and interest expended to total assets	.985	.969	.966	.03219	253.39	.000*
Net NPAs to total assets and net interest income to total assets	.717	.514	.453	.04677	8.457	.020*
Net NPAs to total assets and provision and contingencies to total assets	.218	.047	-.072	.09795	.398	0.546
Net NPAs to total assets and operational expenses to total assets	.992	.983	.981	.02694	466.16	.000*
Net NPAs to total assets and capital adequacy ratio	.990	.980	.977	.02591	389.56	.000*

Source**:** Calculated data

*Indicates statistical significance at 5 per cent level

The table 2 shows the significant impact of Net NPAs to total assets on Operational Performance of select foreign banks. Net NPAs to total assets significantly impact on return to total assets (0.002), interest expended to total assets (0.020), net interest income to total assets (0.000), operational expenses to total assets (0.000), and capital adequacy ratio (0.000) at 5 percent level. Hence, the null hypothesis is rejected it refers there is significant effect of the NPAs on operational performance of select foreign banks in India.

Effect of NPAs on Macroeconomic Variables

In this section, attempts to analyze the effect of NPAs on macro economic variables using ordinary least square and analyze the growth rate of macroeconomic variables of India.

Growth of Macroeconomic Variables in India

Table 3: Growth of GDP (In percentage)

Year	*Gross domestic product*	*Inflation rate*	*Gross Domestic Savings*	*Exports*	*FER*
2007-08	22.30	22.40	6.37	14.60	43.51
2008-09	17.50	19.90	8.35	-4.69	48.41
2009-10	16.90	17.20	10.88	19.62	45.73
2010-11	21.50	15.90	11.99	15.58	46.67
2011-12	17.00	13.50	8.86	6.81	53.44
2012-13	14.10	14.20	9.31	7.79	58.60
2013-14	13.90	14.10	10.91	1.78	61.03
2014-15	9.00	10.70	6.65	-5.59	64.15
2015-16	10.90	9.30	4.91	5.37	67.19
2016-17	8.20	15.30	4.94	5.15	65.12
MEAN	15.13	15.25	8.32	6.64	55.38
SD	4.83	3.92	2.53	8.29	8.92
CV	31.92	25.73	30.41	68.72	79.53
SKEW-NESS	0.01	.370	-0.04	0.03	-0.025
KUR-TOSIS	-0.99	.071	-1.37	-0.725	-1.85
CAGR	0.10	0.04	0.03	-0.10	0.04

Source: computed data

The above table 3 demonstrates that Mean, Standard Deviation, CV, Skewness, Kurtosis, Compound annual growth rate of macroeconomic determinants in the study period. The mean value of GDP is 15.13. The CAGR of GDP is 0.10 per cent. The GDP is highest in the year 2007-08 by 22.30 per cent followed by in the year 2010-11 is 21.50 per cent and low in the year of 2016-17 is 8.20 per cent. The mean value of Inflation rate is 15.25. The CAGR value of Inflation rate is 0.04 per cent. The Inflation rate is high in the year 2007-08 by 22.40 per cent followed by in the year 2008-09 is 19.90 per cent and less in the year of 2015-16 is 9.30 per cent. The mean value of Gross Domestic Savings growth rate is 8.32. The CAGR value of Gross Domestic Savings growth rate is 0.03 per cent. The Gross Domestic Savings growth rate is high in the year 2010-11 with 11.99 per cent which is followed in the year 2013-14 with 10.91 per cent and lowest in the year of 2015-16 is 4.91 per cent. The mean value of Exports is 6.64. The CAGR value of Exports growth rate is -0.10 per cent. The Exports is high in the year 2009-10 by 19.16 per cent which is followed in the year 2010-11 is 15.58 per cent and less in the year of 2014-15 with -5.59 per cent. The mean value of FER is 55.38. The CAGR value of FER is 0.04 per cent. The FER is high in the year 2015-16 with 67.19 per cent which is followed in the year 2016-17 with 65.12 per cent and lowest in the year of 2007-08 with 43.51 percent.

Augmented Dickey Fuller (ADF) test

The NPAs of foreign banks on GDP

Table 4: Augmented Dickey Fuller Test of the NPAs on GDP

Variables	*Level*		*1st Difference*		*2nd Difference*	
	t-statistic	*Probability*	*t-statistic*	*Probability*	*t-statistic*	*Probability*
NPAs	-3.780	0.083	-3.990	0.025*		
GDP	-6.161	0.007*				

The above table 4 portrays that the augmented dickey fuller test results of NPAs and GDP. NPAs are stationary at first difference $I(1)$ with the probability value of 0.025 and GDP is stationary at level $I(0)$with the probability value of 0.007. Hence, it is concluded that the above variables are stationary.

H_{01}: There is no significant effect of the NPAs of foreign banks on GDP.

Table 5: Ordinary Least Square Model of NPAs on GDP

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	1.807647	8.223123	0.219825	0.8315
GDP	2.915700	1.137271	2.563769	0.0335
R-squared	0.451036	Mean dependent var		22.33800
Adjusted R-squared	0.382416	S.D. dependent var		7.521248
S.E. of regression	5.910688	Akaike info criterion		6.568258
Sum squared resid	279.4898	Schwarz criterion		6.628775
Log likelihood	-30.84129	Hannan-Quinn criter.		6.501871
F-statistic	6.572912	Durbin-Watson stat		2.232979
Prob (F-statistic)	0.033450			

From above table 5, it can be observed that the result of ordinary least square regression for impact of NPAs of foreign banks on Macro economic variables for the study period. The calculated t-value is 0.220 whose significant value is statistically significant at 5 per cent level. Therefore, the null hypothesis is rejected. It can be concluded that there is a significant difference between the NPAs of foreign banks and GDP of India. The value of Durbin Watson statistics 2.23 is that the model indicates there is no auto correlation problem.

The NPAs of Foreign Banks on Inflation Rate

Table 6: Augmented Dickey Fuller Test of the NPAs on Inflation Rate

Variables	*Level*		*1st Difference*		*2nd Difference*	
	t-statistic	*Probability*	*t-statistic*	*Probability*	*t-statistic*	*Probability*
NPAs	-3.780	0.083	-3.990	0.025*		
INFL	-2.103	0.478	-2.402	0.169	-5.244	0.037*

The above table 6 portrays that the augmented dickey fuller test results of NPAs and Inflation rate. NPAs are stationary at first difference *I*(1) with the probability value of 0.025 and Inflation rate is stationary at second difference *I*(2)with the probability value of 0.037.Hence, it is concluded that the above variables are stationary.

H_{01}: There is no significant effect of the NPAs of foreign banks and Inflation rate.

Table 7: Ordinary Least Square Model of NPAs on Inflation Rate

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	18.76572	9.001792	2.084665	0.0706
INFL	0.429496	1.039844	0.413039	0.6904
R-squared	0.020880	Mean dependent var		22.33800
Adjusted R-squared	-0.101510	S.D. dependent var		7.521248
S.E. of regression	7.893764	Akaike info criterion		7.146880
Sum squared resid	498.4921	Schwarz criterion		7.207397
Log likelihood	-33.73440	Hannan-Quinn criter.		7.080493
F-statistic	0.170601	Durbin-Watson stat		2.043042
Prob(F-statistic)	0.690427			

From above table 7 it can be observed that the result of ordinary least square regression for impact of NPAs of foreign banks on Macro economic variables for the study period. The calculated t-value is 2.085 whose significant value is 0.690 which is greater than five per cent level of significance. Therefore the null hypothesis is accepted. It can be concluded that there is no significant difference between the NPAs of foreign banks and Inflation rate of India. The value of Durbin Watson statistics 2.04 is that the model indicates there is no auto correlation problem.

NPAs of foreign banks of GDS

Table 8: Augmented Dickey Fuller Test of the NPAs on GDS

Variables	*Level*		*1st Difference*		*2nd Difference*	
	t-statistic	*Probability*	*t-statistic*	*Probability*	*t-statistic*	*Probability*
NPAs	-3.780	0.083	-3.990	0.025*		
GDS	-9.044674	0.000*				

The above table 8 portrays that the augmented dickey fuller test results of NPAs and GDS. NPAs are stationary at first difference *I*(0) with the probability value of 0.025 and GDS is stationary at level I(0) with the probability value of 0.000. Hence, it is concluded that the above variables are stationary.

H_{01}: There is no significant effect of the NPAs of foreign banks and GDS.

Table 9: Ordinary Least Square Model of NPAs on GDS

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	48.30274	66.20281	0.729618	0.4864
GDS	-0.816501	2.080366	-0.392480	0.7050
R-squared	0.018891	Mean dependent var		22.33800
Adjusted R-squared	-0.103747	S.D. dependent var		7.521248
S.E. of regression	7.901777	Akaike info criterion		7.148909
Sum squared resid	499.5046	Schwarz criterion		7.209426
Log likelihood	-33.74454	Hannan-Quinn criter.		7.082522
F-statistic	0.154040	Durbin-Watson stat		2.141479
Prob(F-statistic)	0.704954			

From above table 9 it can be observed that the result of ordinary least square regression for impact of NPAs of foreign banks on Macro economic variables for the study period. The calculated t-value is 0.729 whose significant value is 0.705 which is greater than five per cent level of significance. Therefore the null hypothesis is accepted. It can be concluded that there is no significant difference between the NPAs of foreign banks and GDS of India. The value of Durbin Watson statistics 2.14 is that the model indicates there is no auto correlation problem.

The NPAs of foreign banks of Export

Table 10: Augmented Dickey Fuller Test of the NPAs on Export

Variables	*Level*		*1st Difference*		*2nd Difference*	
	t-statistic	*Probability*	*t-statistic*	*Probability*	*t-statistic*	*Probability*
NPAs	-3.780	0.083	-3.990	0.025*		
Export	-3.969	0.067	-3.711	0.035*		

The above table 10 portrays that the augmented dickey fuller test results of NPAs and Exports. NPAs are stationary at first difference *I*(1) with the probability value of 0.025 and Export is stationary at first difference *I*(1) with the probability value of 0.035.Hence, it is concluded that the above variables are stationary.

H_{01}: There is no significant effect of the NPAs of foreign banks and Export.

Table 11: Ordinary Least Square Model of NPAs on Export

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	23.82584	3.199959	7.445673	0.0001
EXPORT	-0.224021	0.310898	-0.720562	0.4917
R-squared	0.060946	Mean dependent var		22.33800
Adjusted R-squared	-0.056436	S.D. dependent var		7.521248
S.E. of regression	7.730570	Akaike info criterion		7.105099
Sum squared resid	478.0937	Schwarz criterion		7.165616
Log likelihood	-33.52549	Hannan-Quinn criter.		7.038712
F-statistic	0.519209	Durbin-Watson stat		2.279885
Prob(F-statistic)	0.491694			

From above table 11 it can be observed that the result of ordinary least square regression to analyze the impact of NPAs of foreign banks on Macro economic variables for the study period. The calculated t-value is 7.446 whose significant value is 0.492 which is greater than five per cent level of significance. Therefore the null hypothesis is accepted. It can be concluded that there is no significant difference between the NPAs of foreign banks and Exports of India. The value of Durbin Watson statistics 2.27 is that the model indicates there is no auto correlation problem.

NPAs of foreign banks of FER

Table 12: Augmented Dickey Fuller Test of the NPAs on FER

Variables	Level		1st Difference		2nd Difference	
	t-statistic	Probability	t-statistic	Probability	t-statistic	Probability
NPAs	-3.780	0.083	-3.990	0.025*		
FER	-3.113	0.172	-2.686	0.122	-8.384	0.005*

The above table 12 portrays that the augmented dickey fuller test results of NPAs and FER. The NPAs are stationary at first difference *I*(1) with the probability value of 0.025 and FER is stationary at second difference *I*(2) with the probability value of 0.005. Hence, it is concluded that the above variables are stationary.

H_{01}: There is no significant effect of the NPAs of foreign banks and FER.

Table 13: Ordinary Least Square Model of NPAs on FER

Variable	*Coefficient*	*Std. Error*	*t-Statistic*	*Prob.*
C	10.32668	16.13903	0.639858	0.5402
FER	0.216873	0.288059	0.752879	0.4731
R-squared	0.066165	Mean dependent var		22.33800
Adjusted R-squared	-0.050564	S.D. dependent var		7.521248
S.E. of regression	7.709055	Akaike info criterion		7.099525
Sum squared resid	475.4363	Schwarz criterion		7.160042

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Log likelihood	-33.49762	Hannan-Quinn criter.		7.033138
F-statistic	0.566827	Durbin-Watson stat		2.128229
Prob(F-statistic)	0.473093			

From above table 14 it can be observed that the result of ordinary least square regression to analyze the impact of NPAs of foreign banks on Macro economic variables for the study period. The calculated t-value is 0.639858 whose significant value is 0.473093 which is greater than five per cent level of significance. Therefore the null hypothesis is accepted. It can be concluded that there is no significant difference between the NPAs of foreign banks and FER of India. The value of Durbin Watson statistics 2.12 is that the model indicates there is no auto correlation problem.

Suggestions

- The high level of NPAs have been found in Citi bank, SCB, Deutsche bank, these banks should take steps to reduce the NPAs in order to increase the efficiency of financial position of the banks as well as the borrowers, through timely declaring of repayment to the credit takers.
- Banks should be careful while making advances to personal loan seekers, industrial credit seekers. The banks should not take risks in giving loan without adhering to the norms with a purpose to raising the credit growth rate.
- Suitable lawful measures should be taken against willful defaulters. The banks should be provided monopoly to take action against the confirmed cases. The legal system must be revamped, to make certain that needless delays in the legal measures can be minimized.

Conclusion

The study concluded that the non-performing assets of foreign banks in India. The level of NPAs is high in Citi bank, SCB, and Deutsche bank among other select foreign banks in India. The foreign banks profitability is satisfactory and also restricted to expand their business in India is the huge priority sector requirement. The capital adequacy is lacking in Deutsche bank, DBS bank, Citi bank, and Barclays bank. So, it will affect the bank's strength, bank's profitability and the operations of the bank. For that reason the study concludes that the banks should take essential steps to curtail the mounting NPAs. This will allow the overall development of the economy and carry hope among the investors across the world in the Indian economy.

References

Alagarsami, T., & Ganapathy, S. (2017). Performance of non-performing assets in SBI. *International Journal of Advanced Research in Management and Social Science*, *6*(8).

AyubAhamed, K. S., & Panwar, V. (2016). A comparative study of non-performing assets in private sector banks and public sector banks in India. *International Journal of Commerce Business and Management*, *5*(6).

Dave, K. K., & Inamdar Meghnad Pushkar Kumar. (2016). A comparative study of non-performing assets in public and private sector banks in India. *International Journal of Scientific Research*, *5*(8).

Jayakkodi, V., & Rengarajan, P. (2016). A study on non-performing assets of select public and private sector Banks in India. *International Journal of Finance Research Review*, *4*(9).

Kiran, K. P., & Jones, T. M. (March 2016). Effect of NPA on the profitability of banks – A selective study. *International Academy of Science Engineering and Technology*, *5*(2), 53–60.

Mishra, U. M., & Pawaskar, J. R. (2017). A study of non-performing assets and its impact on banking sector. *Journal for Research*, *3*(1).

Mittal, R. K., & Suneja, D. (2017). The problem of rising non-performing assets in Banking sector in India: Comparative analysis of public and private sector banks. *International Journal of Management IT and Engineering*, *7*(7).

Miyan, M. (2017). A comparative study-Impact of non-performing assets on public sector banks and private sector. *Middle Eastern Finance and Economics*, Issue 7.

Ojha, A., & Jha, H. C. (2018). A comparative study regarding impact of non-performing assets on working of the SBI and he BOI. *International Journal of Research in Management Economic and Commerce*, *8*(3), 105–113.

Prakash, J., Kumar, R., & Mewada, S. (2016). A comparative study non-performing assets in Indian Banking Industry. *IOSR Journal of Business and Management*, *18*(10), 57–64.

Reeta. (2016). Interest rate risk: A comparative study of public and private sector banks in India. *Journal of Business and Financial Affairs*, *5*(4), 1–4.

Sailaja, V. N., & Bindhu Madhavi, N. (2016). A comparative study on non-performing assets in public sector and private sector banks in India. *The International Journal of Business and Management*, *4*(4).

Singh, V. R. (2016). A study of NPA of commercial bank and its recovery in India. *Annual Research Journal of SCMS*, *4*.

Sivasatyanarayana, D. (2016). Non-performing assets of commercial Banks in India- A study. *International Journal of Trend in Research development*, *3*(6).

Tobak, M. S., & Talu, N. I. (n.d.). Bank specific and macro economic determinants of bank profitability evidence from Turkey. *International Journal of Economics and Financial Issues, 7*(2).

Vanishreesah, P. (2017). A study on non-performing assets of the Indian banks. *International Journal of Advanced Research*, *5*(4).

Vasanthi, G., & Aakashkumar. (December 2017). Trend of gross and net non-performing assets in public sector banks and foreign banks: A comparative analysis. *Indian Journal of Accounting*, *49*(2), 98–102.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 46-52
http://publishingindia.com/ijbai/

The Statistical Analysis of Examination and Evaluation of Results of International Trade Course

Samik Shome*, Urmi Khatri**

Abstract

The statistical analysis of any examination results are important the professors to understand about the magnitude of knowledge students have incurred from their courses. Based on the examination results of International Trade course of International Business specialization students in the third term of MBA 2017-2019 batch of IMS Unison University, the quantitative analysis for several parameters including difficulty, discrimination, reliability and ANOVA are investigated. The results indicate that the distribution of examination scores approximate to normal distribution. The results stated that the exam paper belongs to a moderate level, which was then qualified by the discrimination and reliability tests done in the study. Thus it was concluded that the design of the examination paper was good and dependable.

Keywords: International Trade, Difficulty Index, Discrimination Index, Reliability Test, ANOVA

Introduction

The analysis of the examination results is done by any faculty or professor, interchangeably used in this paper, to know the understanding and grasping ability of the student for a particular subject. On the basis of the scores, student's conceptual clarity of that particular subject can be judged. In order to gain the most benefit from examination, faculty at each institution need to develop their own understanding of the process (Palomba and Banta, 1999). According to Hatchings and Marches (1990), the meaning of examination for a college graduate is best captured by responding to some fundamental questions: (a) What should they know or be able to do and value? (b) Have the graduates of our institutions acquired this learning? (c) What are the contributors of the institution and its programs to student growth? (d) How can student learning be improved? and, (e) When individuals involved in assessment become confused about its purposes, it helps to return to these questions.

Nowadays, in several colleges and universities, the examination analysis has been developed as an exercise after each examination. This exercise helps to identify the trend of marks the class received and its reviewing and reflection on academic standard. This exercise can be done by using statistical techniques and models on the results of the students. In other words, the analysis of the results helps the professors to understand the students' knowledge of a particular subject and also the quality of paper prepared for examination. This type of analytical exercise can serve two purposes simultaneously: first, it helps to know the ability of the students to understand the subject; and, secondly, it also helps to modify the examination paper, if required, and make it more adequate and proficient next time. Hence, through this work out, one gets the loop holes of the teaching method and examination system or pattern of the college.

In this backdrop, the central argument of this paper is to analyze the quality of a question paper in terms of its design and standard. The analysis and conclusion drawn are established on the basis of the marks students received in the course.

* Associate Professor, Institute of Management, Nirma University, Ahmedabad, Gujarat, India. Email: samik@nirmauni.ac.in
** Doctoral Scholar, Institute of Management, Nirma University, DPM Cubicles, Ahmedabad, Gujarat, India.
Email: urmi.khatri@nirmauni.ac.in

Literature Review

Basic statistical analysis of examination results is a method typically used nowadays in many universities as a part of academic audit and requirement for accreditation purpose. However, more detailed analysis can be done individually by professors to examine their realizations of expectations after teaching a course. Every course a faculty teaches has its own preferences and hopes that create its own view. Several researchers and scholars in the last few years have proposed means and the underlying analytical framework behind the learning pattern of students from different courses. One of the pioneering research in this area by Tversky and Kahneman (1981) discussed that the concept of framing effect the professors made up of their own views of the subject. They also showed how the framing effect influence the way in which the information is interpreted. Individual decisions are influenced by the presented information and by the problems formulation (Druckman, 2001).

Yang, et. al (2013) analyzed the examination results of materials research methods course for 100 students of Wuhan University of Science and Technology in China to evaluate the validity, reliability, difficulty, and discrimination power of the examinations. The study concludes that the design of the examination paper was good and dependable. Kaspříková (2011, 2012) did such analysis for mathematics course at University of Economics in Prague and JarkovskáKučera, Vostrá Vydrová, and Varvažovská (2012) for distance programs at Czech University of Life Sciences, Prague.

Borožová and Rydval (2014) used difficulty index, discrimination index and Cronbach's alpha to study the examination results of Applied Mathematics for Information Technology (IT) for a sample size of 615 students. They observed that the students found the theoretical questions to be the hardest among all the sections in the question paper. Again, Yang, et. al (2013) in analysis of the examination in anesthesiology for medical students used the Rasch model to estimate the parameters which creates a difficulty in tests for students. Similarly, Ramya, et. al (2017) used the Apriori algorithm technique to examine the performance of the students in an examination.

The Study and Its Methodology

The current study is done on a paper of International Trade offered in a two year full time MBA program in IMS Unison University, Dehradun (India). Out of several major streams, including Finance, Marketing and Human Resource Management (HRM), the program also offers a specialization on International Business (IB). The course of International Trade is a part of IB specialization which is offered in the third trimester of first year of the program. For the 2017-19 batch, 75 students opted for this course of International Trade. The examination paper was of 100 marks divided into three parts. The first part consists of five very short type questions, the second part includes five theoretical questions and, finally, the third part consists of two large practical examples. The questions of the tests cover all the topics of the course.

- Five very short type questions – these questions have a form of a brief question that requires a written answer not longer than a two sentences. For example, the students have to write the basic definitions or any two functions or the simple principle. Maximum score of each question is 2 points.
- Five theoretical questions – these questions have a form of description questions, which have to be discussed in details and the process or results have to be interpreted. Maximum score of each example used to be 10 points.
- Two essay or long questions – this part of the paper has a form of a case-study or scenario question, which is used to prove that students can understand and integrate key concepts of the course, apply theory to a practical context, and demonstrate the ability to analyze and evaluate obtained effects. These types of questions are practical in nature and the students need to attempt any two questions out of three. Maximum score used to be 40 points (20 points for each answer).

A summary of the question paper format is provided in Table 1.

Table 1: Question Paper Format of International Trade

Part	*Compulsory questions*	*Total questions*	*Marks per question*	*Total marks*
Part 1	5 compulsory	5 questions	2 marks each	10
Part 2	5 compulsory	5 questions	10 marks each	50
Part 3	2 compulsory out of 3	2 questions	20 marks each	40
Part (1+2+3)		12 questions		100

Source: Question paper of International Trade of IMS Unison University

To analyze the examination results, the data was used from student actual scores; the data was collected on number of students, their enrolment numbers and the numbers they scored in the examination. Now to identify the test quality and the level of difficulty the students faced in answering the question paper of International Trade, this paper followed the techniques used by Jacobs (1991), Miller (2012), Wells and Wollack (2003) and Yang, et. al (2013):

- Difficulty Index of the test,
- Discrimination Index of the test,
- Reliability of the test.

In addition, this paper also used the Analysis of Variance (ANOVA) to check the difference in the average marks of the students.

Difficulty Index

Any examination should not be either too easy or too tough. The question paper is required to prepare by keeping two objectives in mind: (a) the examination should be able to assess student's knowledge of the subject; and, (b) it should be able to measure student's knowledge.

The ratio of the average scores to item *i*, to the full scores of item *i*, gives the difficulty index of item *i*. It can also be understood as the proportion of a student who answers the examination question correctly. The formula is as follows:

$$Pi = \frac{Ai}{Ni}$$

where, Pi = Difficulty index of item *i*, Ai = Average scores to item *i*, Ni = Full scores of item *i* for the whole script. The average difficulty index P can be calculated by the formula as below:

$$P = \frac{1}{100}\sum_{i=1}^{N} PiNi$$

Interpretation: P > 0.75 implies exam is easy; P < 0.45 implies exam is difficult; 0.45 < P < 0.75 implies exam is average or passable.

Discrimination Index

Each examination should be able to discriminate between a well prepared student and a not so one. It is expected that the students who are prepared should answer the questions well in the examination. There should be some questions which the not so well prepared students will find difficult to answer. Hence, the discrimination index D gives a variation in the student's performance. In other words, from discrimination index highly intellectual students can be discriminated from less intellectual students.

$$D = \frac{P_H - P_L}{100}$$

where, P_H = Average score for the 27% of students with highest test marks,

P_L = Average score for the 27% of students with lowest marks.

Interpretation: According to R. L. Ebel (1972), if D > 0.39, the quality of the exam paper is excellent. When 0.30 < D < 0.39, the exam paper is qualified. If 0.20 < D < 0.29, it indicates that the quality of the exam paper is passable and has possibility for improvement. The exam paper should be discarded if D is less than 0.20.

Reliability of the Test

Cronbach's alpha is used for estimating the reliability of the test. This alpha was developed by Lee Cronbach (1951) to provide a measure of the internal consistency of a test or scale. Internal consistency describes the extent to which all the items in a test measure the same concept or construct and hence it is connected to the inter-relatedness of the items within the test.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\in S_I^2}{S_X^2}\right)$$

where, k = Total of item, S_I^2 = Variance of scores for item i, and, S_X^2 = Variance of scores for script.

It is also consider as the heart of quality control of examination. Although the value can range between 0 and 1, but it generally falls in the range of 0.60 to 0.80. Table 2 explains the internal consistency with respect to the alpha value.

Table 2: Alpha Value with Respect to Internal Consistency

Internal Consistency	*Cronbach's Alpha*
Excellent	0.9 - 1.0
Good	0.8 - 0.9
Acceptable	0.7 - 0.8
Questionable	0.6 -0.7
Poor	0.5 - 0.6
Unacceptable	0.0 - 0.5

Source: Cronbach (1951)

The examination paper is acceptable if the alpha value is greater than 0.7 and is totally unacceptable if it is below 0.5. The paper is excellent if the alpha value is between 0.9 and 1.

ANOVA

Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences among group means and their associated procedures (such as 'variation' among and between groups). It is developed by statistician and evolutionary statistician and biologist Ronald Fisher (1918). This paper used ANOVA to find out the difference in the average marks of the students, in all the three parts (i.e. Part 1, Part 2 and Part 3) of the test.

Results and Discussions

The descriptive statistics of the scores of 75 students is presented in Table 3. It can be inferred that a few students scored full marks in Part A. It can also be observed that some of the students could not score any marks in Part A and Part C. One possibility of this phenomenon may be that these students may have left the questions in those parts.

Table 3: Descriptive Statistics

Parameters	*Part A*	*Part B*	*Part C*
Minimum	00	10	00
Median	06	28	28
Mean	5.89	27.65	27.45
Maximum	10	43	36

Source: Calculated by the authors

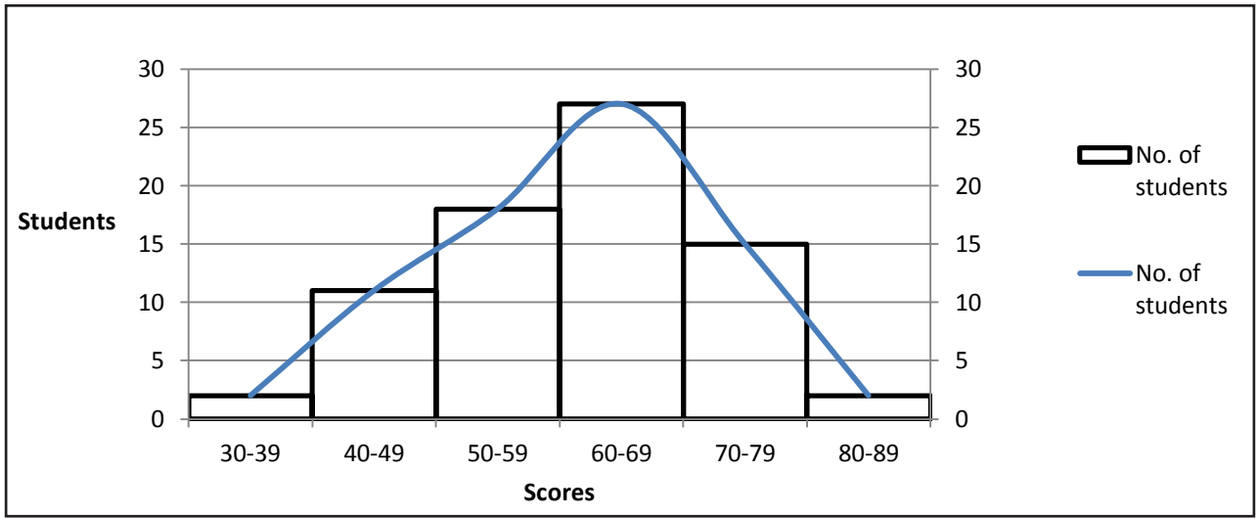
The average marks scored is almost same in Part B and C. However, the maximum score was 10 more in Part C than in Part B. This shows that the students have been able to perform comparatively better in Part C. Thus, it can be inferred that the students were more comfortable in answering essay type questions than the shorter once. The distribution of marks is presented in Table 4.

Table 4: Distribution of Marks

Part A		*Part B*		*Part C*		*Total (Part A + B + C)*	
Marks Interval	*No. of Students*	*Marks Interval*	*No. of Students*	*Marks Interval*	*No. of Students*	*Marks Interval*	*No. of Students*
00	04	00	00	00	01	30 – 39	02
0 – 2	03	0 – 10	01	0 – 10	00	40 – 49	11
2 – 4	10	10 – 20	11	10 – 20	04	50 – 59	18
4 – 6	27	20 – 30	36	20 – 30	48	60 – 69	27
6 – 8	21	30 – 40	25	30 – 40	22	70 – 79	15
8 – 10	10	40 – 50	02	–	–	80 – 89	02
Total	**75**	**Total**	**75**	Total	75	Total	75

Source: Calculated by the authors

Maximum students (27 altogether) have scored between the range of 60-69 marks followed by 15 students between 70-79. There are only two students who scored above 80 marks. However, there are also two students who scored in between 30-39. The distribution of the total marks of all students is depicted in Fig. 1.



Source: Calculated by the authors

Fig. 1: Distribution of Overall Marks

Since the pass mark is 30 out of 100, hence, no students has failed in this subject of International Trade. The results show that the scores are approximate to normal distribution.

Difficulty Index

The difficulty index of each part and the overall question paper of International Trade is presented in Table 5.

Table 5: Difficulty Index of Examination for Part 1, 2, 3 and overall

Particulars	*Part 1*	*Part 2*	*Part 3*	*Total*
Total score	10	50	40	100
Difficulty	0.5893	0.5531	0.6833	0.61
Quality	Average	Average	Average	Average/ Passable

Source: Calculated by the authors

By comparing the different parts of the examination paper, it can be seen that its difficulty index *P* ranges in between 0.55 to 0.69 (Table 5). The easiest part was the essay type or scenario questions in which majority of the students did well. Very short answer questions, as well as the theoretical were the most difficult parts. It implies that although the students were able to relate the basic concepts with the case based practical questions, but it is also evidenced that the ability of students for mastering basic definitions and handle theories was deficient. It shows the students were not able memories definitions and theories and write them but are enough knowledgeable apply the concepts in the practical scenario.

As the overall value of the difficulty index *P* of examination paper is 0.61 it implies that the level of the paper is moderate and therefore it is not difficult for students to pass this paper.

Discrimination Index

The analysis of the discrimination index *D* for the examination results shows that the value is 0.74 (Table 6).

Table 6: Calculation of D value

Particulars	*Value*
P_H	74.2
P_L	46.35
D	0.742666667

Source: Calculated by the authors

According to Ebel's rule, the exam paper is highly qualified and the question paper is designed very well. Although there is always scope for improvement in any examination paper, but this paper also qualifies to be highly acceptable at its current form.

Reliability of the Test

The reliability of the test i.e., Cronbach's alpha is conducted to study the internal consistency of a question paper. The consistency can be achieved by setting a question paper with an appropriate combination of definitions, theories, and practical application questions to determine the level of knowledge of the students. For this paper, the quality is proved to be good as the result of Cronbach's alpha is 0.71.

Table 7: Calculation of Cronbach's Alpha

Particular	*Values*
Variance	39.609
Total variance	130.43243
K	75
Cronbach's α	0.71

Source: Calculated by the authors

ANOVA

The ANOVA test is conducted to find out whether there is any significant difference in average marks scored in the three parts of the examination papers. For that, the scores are converted from absolute value to per cent form and then the single factor ANOVA test is run. The hypothesis of the test is as follows:

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: Not all means are equal.

The result of the ANOVA test is presented in Table 8.

Table 8: ANOVA of the Exam Statistics of Part A, B and C

Summary						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Part 1	75	4420	58.93333	606.955		
Part 2	75	4148	55.30667	176.5939		
Part 3	75	5147.5	68.63333	184.509		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	7121.069	2	3560.534	11.03405	0.000027	3.036524
Within Groups	71636.28	222	322.6859			
Total	78757.35	224				

Source: Calculated by the authors

*Note: All the scores are normalized to 100 for ANOVA

As seen earlier in other analysis, the average marks are relatively higher in Part 3 i.e. 68.63 compared to the other two parts. The average is lowest in Part 2 which is 55.31 followed by 58.9 in Part A. It can also be seen that the variance in marks scored is more in Part 1.

Since, the decision rule of ANOVA test is: Reject H_0 if, $F > F\ crit$, here $11.03405 > 3.036524$. Therefore, H_0 is rejected. It means that the average scores of the three parts of the paper are not equal. It is found that Part C, the average score is higher than the rest of the parts. One probable reason for the students to score more in Part 3 may be because of the availability of choices in the questions. In Part 3, the students got an opportunity to attempt two questions out of three of 20 marks each. This ultimately had a positive effect on the increase in overall average scores of the examination.

Concluding Observations

The statistical analysis of any examination results is an essential for the professors to understand about the magnitude of knowledge students have incurred from their course. It can also act as a feedback mechanism about the

quality of examination papers in terms of its design and standard. This will help the professors to make necessary changes in the questions and improve the standard of the examination.

This paper studies the results of International Trade examination offered in International Business (IB) specialization in a two year full time MBA program in IMS Unison University, Dehradun (India). The analysis suggests that the distribution of examination scores approximate to normal distribution. Several parameters for the examination paper including difficulty index, discrimination index and reliability were calculated. The values are 0.61, 0.74 and 0.71, respectively. The difficulty of the exam paper belongs to moderate level, therefore it is not difficult for students to pass this examination. Thus, it has qualified both the discrimination and reliability. The ANOVA result shows that Part C of the question paper was relatively more scoring as compared to any other parts of the paper. To conclude, the statistical analysis of the result reveals that the design of the examination paper was good and consistent and the standard is dependable.

References

Beveridge, J. R., She, B. D., & Givens, G. H. (2001). Parametric and nonparametric methods for the statistical evaluation of human Id algorithms. In *Proc. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems*.

Borožová, H., & Rydval, J. (2014). Analysis of exam results of the subject applied mathematics for it. *Journal on Efficiency and Responsibility in Education and Science*, *7*(3–4), 59–65.

Cable, D. M., & Judge, T. A. (1996). Person–organization Fit, Job Choice Decisions, and Organizational Entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311.

Cable, D. M., & Judge, T. A. (1996). Person–organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311.

Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, *22*(1), 91–101. http://dx.doi.org/10.1016/S0167-4870(00)00032-5

Ebel, R. L. (1972). *Essentials of educational measurement*. Upper Saddle River, NJ: Prentice Hall.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Hutchings, P., & Marchese, T. (1990). Watching assessment: Questions, stories, prospects. *Change: The Magazine of Higher Learning*, *22*(5), 12–38.

Jacobs, L. C. (1991). *Test Reliability*. Retrieved from http://www.indiana.edu/~best/test_reliability.shtml

Jarkovská, M., Kučera, P., Vostrá Vydrová, H., & Varvažovská, P. (2012). Analysis of students' results in distance-studies centres. *Journal on Efficiency and Responsibility in Education and Science*, *5*(2), 78–91. http://dx.doi.org/10.7160/eriesj.2012.050203

International Journal of Business Analytics and Intelligence
7 (1) 2019, 53-63
http://publishingindia.com/ijbai/

Identifying Homogeneity of Small-Cap Stocks in Indian Market: A Data Mining Approach

Shuvashish Roy*, Rajib Bhattacharya**

Abstract

Investors in equity shares look for two basic aspects while investing i.e. consistently rising returns with a decreasing or at least stabilized level of risk involved. Amidst the numerous stocks available in the market which differ widely on the basis of different aspects i.e. segment, sector, industry, market capitalization etc. it becomes a challenge for the investor to form a diversified portfolio where heterogeneity of the constituent stocks is the main criterion. Thus it is imperative that the basis be finalized on which the heterogeneity of the stocks shall be determined. Traditionally portfolios have been constituted on the basis of low coefficient of correlation of returns from the constituent stocks. The dissimilarity of co-movement of returns from stocks has traditionally been attempted to be maximized in portfolios. Another method of grouping similar stocks by using data mining approach is fast gaining popularity. This approach uses clustering technique to group homogeneous stocks on the basis of a set of selected criteria. These criteria can be financial ratios, indices or any other related matrices. Advanced versions of this technique can group homogeneous time series data as well. This paper attempts to identify homogeneous clusters of companies in the Indian small-cap segment based on valuation ratios. Valuation ratios have been selected to be the grouping criteria as these were not been used in earlier studies by researchers and scholars. The small cap companies in India have been chosen for this study for its better resilience and recovering potential compared to mid cap and large cap companies. The companies constituting the CNX NIFTY Small Cap 50 Index have been considered in the study.

Keywords: Cluster Analysis, Valuation Ratios, Small Cap Sector, CNX NIFTY Mid Cap 50 Index

Introduction

A system of categorization of stock market would be useful to investors and financial analysts, providing them with the opportunity to predict the stock price changes of a company vis-a-vis other companies. In recent years, clustering companies in the stock markets based on their similarities on different aspects has increasingly become a common practice. However stock price data are high-dimensional in nature and the changes in the stock price usually occur with shift, which makes the categorization a complex issue. Clustering Method is an adaptive procedure in which homogeneous objects are clustered or grouped together, based on the principle of maximizing the intra-class similarity and minimizing the inter-class similarity. It is essentially a data mining technique in which similar data are automatically placed into related groups without advanced knowledge of the group definitions. Clustering of companies in the stock market is very useful for managers, investors and policy makers. There are numerous companies listed in the Indian market which vary widely based on a host of aspects i.e. the industry, size, capitalization, business models etc. However, the investors have only two aspects to consider i.e. consistently increase returns with a cap on risk involved. Earlier studies have used financial ratios as clustering factors. To capture the market price of the stocks, which is of prime importance to the investors,

* Financial Advisor, Hazrat Khajar Bashir Unani Ayurvedic Medical College & Hospital Foundation, Jamalpur, Bangladesh. Email: shuvashishroy@gmail.com

** Assistant Professor, Institute of Business Management, The National Council of Education Bengal, Kolkata, West Bengal, India. Email: rajib.conference@gmail.com

into the clustering factors, this paper uses valuation ratios as the grouping factors because valuation ratios link the market prices of stocks with the performance financials of the companies. The small cap companies have been chosen for certain reasons. While CNX Nifty has returned 4 per cent over the past one year, mid-cap and small-cap indices have returned 15 per cent and 24 per cent. Between January 2018 and February 2019, midcap and smallcap indices have seen deeper corrections compared with the largecap index despite registering higher profit growth. The BSE Smallcap index, which fell 32 per cent between January 15, 2018 and February 18, 2019, has recovered 10 per cent since then. Similarly, the BSE Midcap index gained 7.32 per cent since February 2 after sliding 25 per cent between January 9, 2018, and February 18, 2019. The obvious expectation of the market is that the tide is about to turn and the small and mid-cap sector is poised to outperform the large-cap sector. Among small cap and mid cap stocks, the former have exhibited greater resilience and higher recovering potential. Small companies can begin to rebound in growing economies faster than larger companies. Decisions about new products and services and how to bring them to market can also be made and implemented faster with small companies because they have fewer committees, fewer layers of management, and fewer potential obstructions of the kind that exist in the typical large and mid cap companies. When the economy begins to emerge from recession and starts growing again, small-cap stocks can respond to the positive environment quicker and potentially grow faster than large-cap stocks. Small cap companies typically raise most of their capital from investors through the equity route. Larger companies raise capital substantially through debt route which raises their financial leverage positions. Hence this paper focuses on the small cap companies and attempts to identify homogeneity across industries among the companies using valuation ratios as the clustering factors.

Survey of Literature and Identification of Research Gap

The survey of concerned literature has been done with the objective of probing into the results of the past researches on clustering companies with focus on three aspects i.e. the method used, the grouping factors used and the results obtained. Tufan and Hamarat (2003) opined that a combination of clustering methods and fuzzy logic methods provided better results in clustering financial ratios of listed companies. Da Costa (2005) successfully applied cluster analysis on companies in North & South America to classify stocks in spot markets according to a risk-return criterion. Lemos et al. (2005) used K-Means clustering analysis technique along with Data Envelopment Analysis method and was successful in reducing the numbers of variables, improving the visualization of variables and making a coherent and a homogeneous comparison. Silva and De Costa (2005) showed how investors in North & South America can take better investment decisions using cluster analysis to select stocks. Wang and Lee (2008), in their research, modified the K-Means clustering method to leverage on its strengths and avoid its weaknesses in classifying companies based on financial ratios. Setty (2010) found encouraging results by using clustering methods to classify NIFTY 50 companies based on financial ratios over the period of two consecutive years of 2008-09 & 2009-10. Li and Sun (2011) used two-step clustering methods and was able to predict business failures successfully. Babu (2012), in their research, proved that Hierarchical clustering algorithm with reverse K means method and Hierarchical Agglomerative Clustering method yielded better results than K-Means clustering methods. Setyaningsih (2012) used cluster analysis to analyze performance of successful entrepreneurs in Indonesia. Temouri (2012) also used cluster analysis techniques while analyzing the performance of 80 selected enterprises across two observation periods i.e. pre-recession (2005-2007) and recession (2007-2009), on through six indicators: i) share of firms aged below 5 years (entrepreneurialism); ii) employment growth; iii) turnover growth; iv) profitability growth; v) liquidity ratio growth; vi) solvency ratio growth and was able to identify the top performing enterprises in the pre-recession as well as in the recession periods. Aghabozorgi and Teh (2014) worked on companies listed in the Kualalampur Stock Exchange. They used time series clustering methods to identify homogeneous movements in returns of stocks in three phases i.e. pre-clustering, purifying and merging. The algorithm developed by them yielded satisfactory results. Gruener (2015) performed clustering technique in risk clustering based on operating leverage. Marvin (2015) proposed an algorithm based on clustering techniques to create diversified Portfolios with S&P 500 companies. Momeni et al. (2015) worked on companies from cement, metal and automobile companies listed in Tehran Stock Exchange. They used K-Means clustering to identify homogeneous stocks across the clusters using five financial ratios i.e. Return on Assets, Return on Equity, Earnings per share, Profit to Sales and Operating Profit Margins.

Their research provided good results. Szucs (2015) worked on the Hungarian Automotive Industry and was able to identify clusters of homogeneous companies based on index numbers and data from the financial statements of the respective companies. Cai et al. (2016) showed that density-based clustering does not suit financial dataset. Dias (2016) used financial ratios to identify clusters which in turned effectively filtered out potential tax-evading companies. Hou (2016) worked on companies listed in the Shanghai Stock Exchange and could satisfactorily predict financial distress by using K-Means Clustering method with 87.50% accuracy. Goudarzi (2017), in his research on companies listed in Tehran Stock Exchange, proved that clustering method can be used to optimize portfolios. Perisa (2017) worked on companies in the Croatian Market and successfully used clustering methods using profitability indicators. Banerjee and Hofmann (2018) used clustering methods to identify firms that were unable to cover debt servicing costs from current profits over an extended period. Ferrando (2018) utilized clustering technique on the basis of five distinct business models to identify access to finance and innovative activity of companies in the European Union. Fodor (2018) was able to form stock groups using cluster analysis of common size statements and was also able to identify homogeneous co-movement of stock returns. Alexandra Horobet et al. (2019) analyzed financial performances of companies from four Central and Eastern European countries covering five industries i.e. financial services, food and beverages, chemicals, energy and pharmaceuticals. They identified natural groups of companies depending on corporate performance. They came to the conclusion that no clear-cut evidence on their grouping according to industries and/or countries can be identified.

The survey of literature revealed that across the globe, past researches have used only financial ratios and indices as clustering factors and have obtained, on an overall basis, satisfactory results. This identified gap in earlier research works prompted the use of valuation ratios as clustering factors in this study.

Objective of the Study

The main objective of this study is to identify homogeneous clusters of companies from the constituent companies of the CNX NIFTY Small Cap 50 Index by using selected valuation ratios as clustering factors. The ancillary objective of this study is to whether the cluster constituents are industry specific.

Methodology of the Study

The constituent companies of the CNX NIFTY Small Cap 50 index have been selected for the study. The index represents top 50 companies selected based on average daily turnover from the top 100 companies selected based on full market capitalization in NIFTY small cap 250 index. The main objective of the NIFTY small cap 50 Index is to capture the movement of the small cap segment of the market. In India, small cap is a term used to classify companies with a relatively small market capitalization i.e. normally below INR 5,000 crores.

Valuation ratios link the market price of an equity shares of a company with the financials of that company for a particular financial year.

Table 1: Ratios Used in Identifying the Significant Financial Ratios & Their Abbreviations

Ratio	*Abbreviation*	*Method of calculation*
Price – Equity ratio	ADPE	Dividing the current market price of the share by the current earnings per share
Price to Book Value	PTBV	Dividing the current market price of the share by the book value per share
Dividend Yield	DIVY	Dividing total annual dividend payments by market capitalization, assuming the number of outstanding equity shares is constant
Enterprise Value to Net Sales	EVNS	Dividing the current enterprise value of the company by its current net sales
Enterprise Value to Capital Employed	EVCE	Dividing the current enterprise value of the company by its current capital employed
Market Capitalization to Total Sales	MCTS	Dividing the current market capitalization of the company by its current total sales
Price to Cash Flow	PTCF	Dividing the current market price per share of the company by its current cash flow per share
Price to Free Cash Flow	PFCF	Dividing the current market price per share of the company by its current free cash flow per share
Free Cash Flow Yield	FCFY	Dividing the free cash flow per share by the current share price

Source: authors' own assignment of abbreviations

The values of the selected nine valuation ratios were collected for three consecutive financial years i.e. 2015-16, 2016-17 and 2017-18 from Ace Equity ® Data Product. The outliers of the ratios were identified and those companies were excluded from the data set. The values of the ratios were then standardized by using the formula $s_{tandardized} = (x -) / \sigma$.

To have an idea about the number of clusters, hierarchical cluster analysis has been done using the nine valuation ratios as the categorizing factors. Based on the outcome of the hierarchical cluster analysis, the K-Means Cluster Analysis has been done to get the final cluster constitutions. The statistical processes have been carried out on R Studio Platform version 3.5.1. The R codes have been appended as an Annexure after the references.

The Financial Year 2015-16

The results of the hierarchical cluster analysis yielded showed that there are two optimum clusters with 5 & 7 members respectively.

The dendogram below gives a visual representation of the clusters

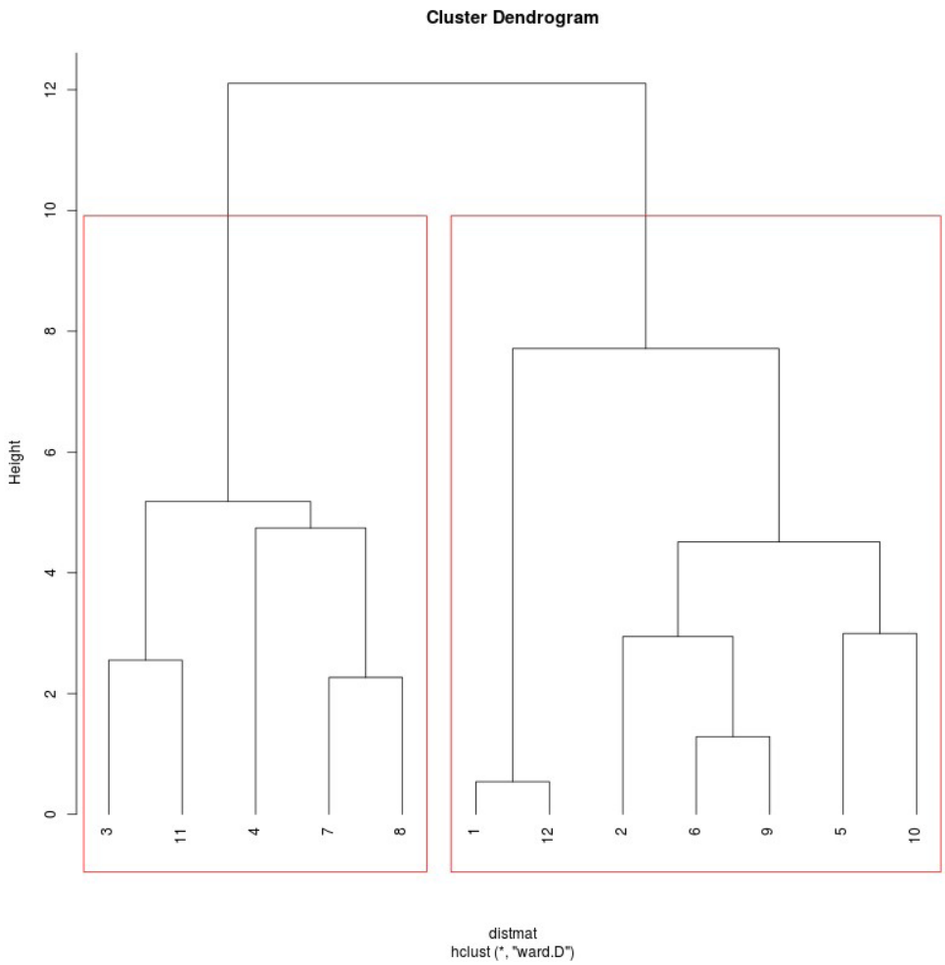


Chart 1: Dendogram – 2015-16

Now K-Means Clustering Analysis is applied taking the number of clusters to be 2. The results are appended below.

Table 2: Cluster Means – 2015-16

	adpe	*ptbv*	*divy*	*evns*	*Evce*	*mcts*	*ptcf*	*pfcf*	*fcfy*
1	-0.58474447	-0.51590575	0.15602953	-0.23982172	-0.4961136	-0.73436711	-0.70395625	0.68174415	0.26749964
2	0.81864226	0.72226805	-0.21844134	0.33575041	0.69455904	1.02811395	0.98553875	-0.95444181	-0.3744995

Table 3: K-Means Clustering with 2 Clusters – 2015-16

Within-Cluster Sum of Squares	
	values
1	36.81984735
2	29.05828865

Sum of Squares		
Total Sum of Squares	*Total Within Cluster Sum of Squares*	*Between Cluster Sum of Squares*
110.91666667	65.878136	45.03853067

Cluster Distribution size	
1	5
2	7

The Financial Year 2016-17

The results of the hierarchical cluster analysis yielded showed that there are two optimum clusters with 2 & 3 members respectively.

The dendogram below gives a visual representation of the clusters

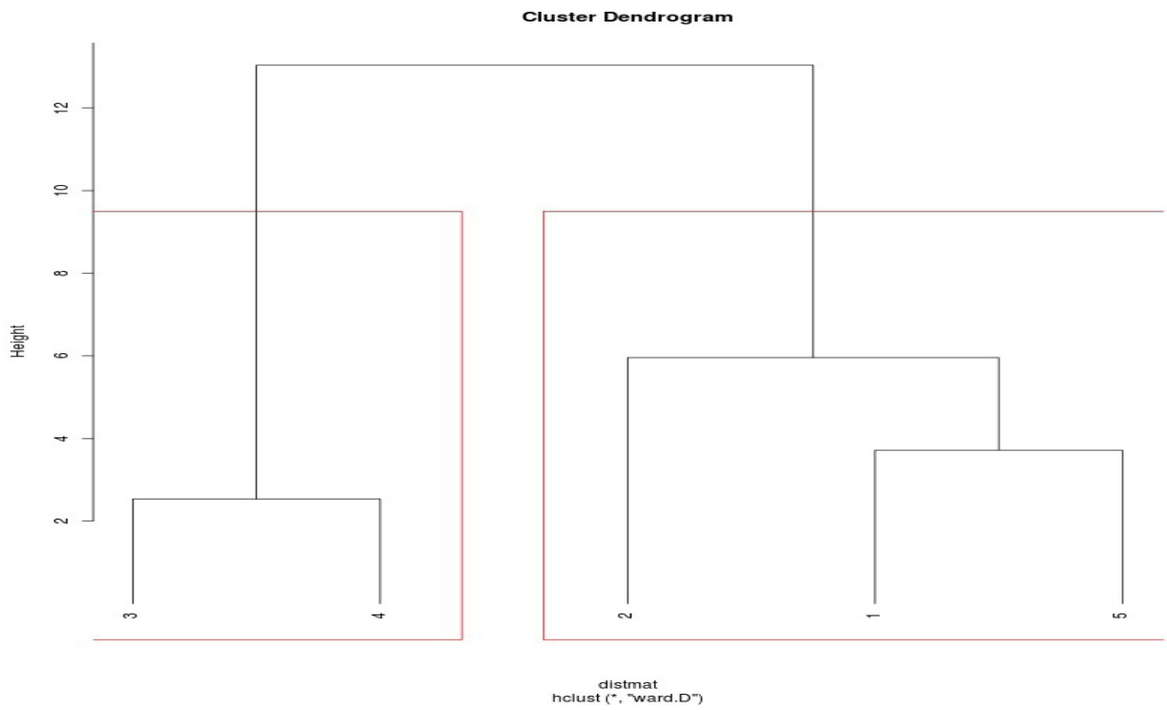


Chart 2: Dendogram – 2016-17

Now K-Means Clustering Analysis is applied taking the number of clusters to be 2. The results are appended below.

Table 4: Cluster Means – Hierarchical Means – 2016-17

Cluster Means									
Cluster	*adpe*	*ptbv*	*divy*	*evns*	*evce*	*mcts*	*ptcf*	*pfcf*	*fcfy*
1	3.0579256	0.58836698	-0.49222891	0.39294344	0.79214206	0.21452671	0.52546139	9.53984275	-0.77878315
2	1.52434337	-0.88255046	0.73834336	-0.58941516	-1.18821309	-0.32179007	-0.78819208	3.65234275	1.16817472

Table 5: K-Means Clustering with 2 Clusters – 2016-17

Within-Cluster Sum of Squares	
	values
1	2.84600579
2	19.93690175

Sum of Squares		
Total Sum of Squares	*Total Within Cluster Sum of Squares*	*Between Cluster Sum of Squares*
81.42422599	22.78290754	58.64131845

Cluster Distribution	
	size
1	2
2	3

The Financial Year 2017-18

The results of the hierarchical cluster analysis yielded showed that there are two optimum clusters with 8 & 6 members respectively.

The dendogram below gives a visual representation of the clusters

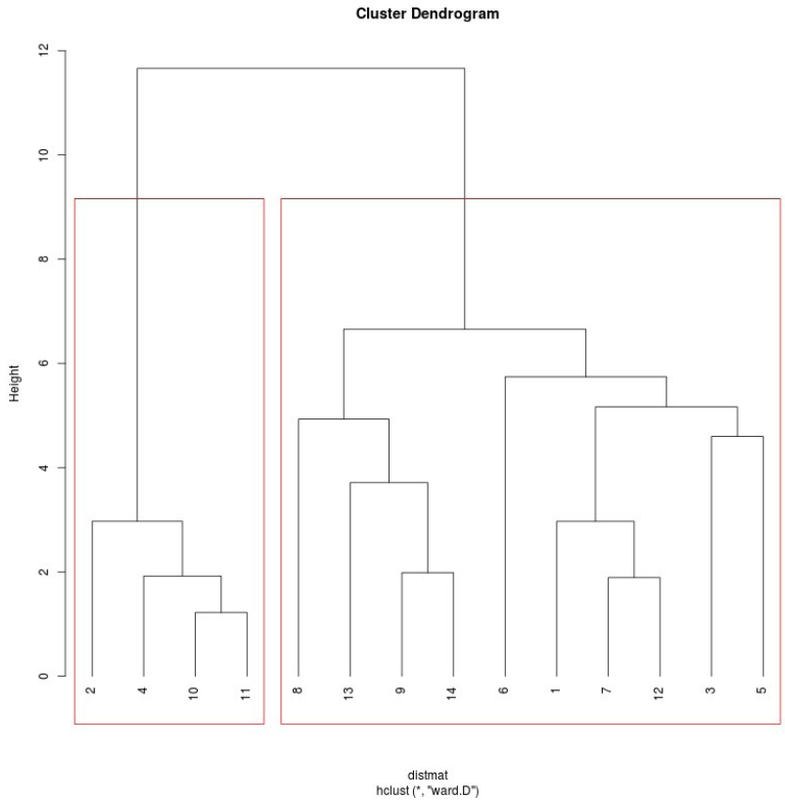
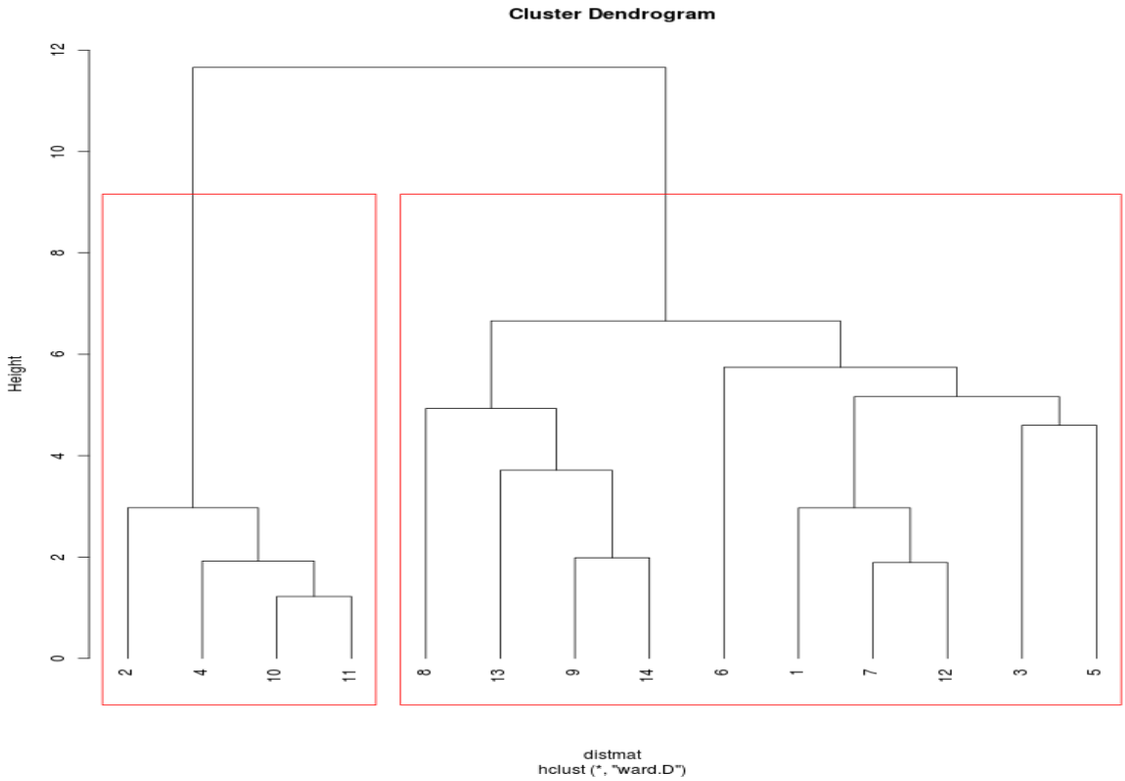


Chart 3: Dendogram – 2017-18



Now K-Means Clustering Analysis is applied taking the number of clusters to be 2. The results are appended below.

Table 6: Cluster Means – Hierarchical Means – 2017-18

Cluster Means									
	adpe	*ptbv*	*divy*	*evns*	*evce*	*mcts*	*ptcf*	*pfcf*	*fcfy*
1	0.26999943	-0.54773092	-0.27070291	-0.70022314	-0.62337297	-0.73792575	-0.10266131	-0.54036934	-0.52026911
2	-0.26999943	0.54773092	0.27070291	0.70022314	0.62337297	0.73792575	0.10266131	0.54036934	0.52026911

Table 7: K-Means Clustering with 2 Clusters – 2016-17

Within-Cluster Sum of Squares	
	values
1	43.85795107
2	47.94217759

Sum of Squares		
Total Sum of Squares	*Total Within Cluster Sum of Squares*	*Between Cluster Sum of Squares*
126	91.80012866	34.19987134

Cluster Distribution	
	size
1	8
2	6

Findings of the Study

In the year 2015-16, number of companies forming the data set was twelve excluding the companies whose valuation ratios contained outliers. There were no overlapping of industries in the two sectors. Crude Oil, chemicals and banks were included in one cluster indicating homogeneity of valuation ratios among these industries.

Table 8: Constitution of Cluster – 2015-16

Company	Sector	Industry	Cluster
Ceat Ltd.	Automobile & Ancillaries	Tyres & Allied	1
CESC Ltd.	Power	Power Generation/Distribution	1
Jain Irrigation Systems Ltd.	Plastic Products	Plastic Products	1
NCC Ltd.	Infrastructure	Engineering - Construction	1
Raymond Ltd.	Textile	Textile - Weaving	1
Allahabad Bank	Bank	Bank - Public	2
Andhra Bank	Bank	Bank - Public	2
Chennai Petroleum Corporation Ltd.	Crude Oil	Refineries	2
Gujarat Narmada Valley Fertilizers & Chemicals Ltd.	Chemicals	Fertilizers	2
Rain Industries Ltd.	Crude Oil	Petrochemicals	2
Rashtriya Chemicals & Fertilizers Ltd.	Chemicals	Fertilizers	2
Syndicate Bank	Bank	Bank - Public	2

In the year 2016-17, the data set contained only 5 companies as most of the companies had to be dropped from the data set due to outliers in the valuation ratios. The findings for this year exhibited overlapping of IT Sector in the two clusters. The results could not lead to any conclusion for this year.

Table 9: Constitution of Cluster – 2016-17

Company	Sector	Industry	Cluster
Gujarat State Fertilizers & Chemicals Ltd.	Chemicals	Fertilizers	1
NIIT Technologies Ltd.	IT	IT - Software	1
Bajaj Electricals Ltd.	Consumer Durables	Consumer Durables - Domestic Appliances	2
Cyient Ltd.	IT	IT - Software	2
Radico Khaitan Ltd.	Alcohol	Breweries & Distilleries	2

In the year 2017-18, the data set comprised of fourteen companies excluding those whose valuation ratios were outliers. The results showed no overlapping of industries in the two clusters except construction engineering companies in the infrastructure sector. Only two companies from the sector were included in the data set, one each of which were members of the two clusters. Except this exception, a clear demarcation of industries were noticed in the two clusters. Information Technology and Fertilizer companies were distinctly classified in the two clusters.

Table 10: Constitution of Cluster – 2017-18

Company	Sector	Industry	Cluster
CESC Ltd.	Power	Power Generation/Distribution	1
Fortis Healthcare Ltd.	Healthcare	Hospital & Healthcare Services	1
Gujarat State Fertilizers & Chemicals Ltd.	Chemicals	Fertilizers	1
Indiabulls Real Estate Ltd.	Realty	Construction - Real Estate	1
Jain Irrigation Systems Ltd.	Plastic Products	Plastic Products	1
NCC Ltd.	Infrastructure	Engineering - Construction	1
Rain Industries Ltd.	Crude Oil	Petrochemicals	1
Rashtriya Chemicals & Fertilizers Ltd.	Chemicals	Fertilizers	1
Cyient Ltd.	IT	IT - Software	2
Godfrey Phillips India Ltd.	FMCG	Cigarettes/Tobacco	2
KEC International Ltd.	Infrastructure	Engineering – Construction	2
NIIT Technologies Ltd.	IT	IT - Software	2
Persistent Systems Ltd.	IT	IT - Software	2
Raymond Ltd.	Textile	Textile - Weaving	2

Conclusion

The findings of the study leads to the conclusion that valuation ratios can be used as categorizing factors in clustering of companies across sectors in the small cap segment of the Indian market. The conclusion has been based on the findings of 2015-16 and 2017-18. The results of 2016-17 has not been considered in forming the conclusion due to very small data set which cannot be used to arrive at a generic conclusion. The results which have been found to hold good in both the years 2015-16 and 2017-18, have identified the following two distinct clusters of industries based on homogeneity of valuation ratios.

Table 11: Clustering of Sectors Based on Homogeneity of Valuation Ratios

Cluster 1	*Cluster 2*
Automobile & Ancillary Units	Bank
Power	Chemicals
Plastic	Crude Oil

Recommendations Based on the Inferences Drawn from the Study

Investors in equity shares may use the information about cluster membership based on valuation ratios in deciding the constitution of their portfolios. Their decision to keep or abstain from keeping the stocks together in a portfolio shall depend on whether the investor wants homogeneity or heterogeneity in valuation ratios of the constituent stocks in the portfolio.

Scope of Future Research

This study can be extended by combining financial ratios and other non-financial factors along with valuation ratios to identify the cluster of companies. This study may also be extended to mid cap companies and large cap companies as well. It may also be extended to sectoral companies. Homogeneity in co-movements of returns of stocks in Indian markets for sectors and segments may also be done by using time series clustering to facilitate construction of better diversified portfolios.

References

Alexandra, H., Joldes, C., & Gabriel, D. D. (2019). *A cluster analysis of financial performance in central and Eastern Europe*, 289–294. Retrieved from https://www.researchgate.net/publication/237262963

Aghabozorgi, S., & Teh, Y. W. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, *41*(2014), 1301–1314.

Babu, M. S., Geethanjali, N., & Satyanarayana, B. (2012). Clustering approach to stock market prediction. *International Journal of Advanced Networking and Applications*, *3*(4), 1281–1291.

Banerjee, R., & Hofmann, B. (2018). The rise of zombie firms: Causes and consequences. *BIS Quarterly Review*, 67–78.

Cai, F., Le-Khac, N.-A., & Kechadi, M.-T. (2016). *Clustering approaches for financial data analysis: A survey*. Retrieved from https://arxiv.org/ftp/arxiv/papers/ 1609/1609.08520.pdf

Costa Jr., N., Cunha, J., & Silva, S. D. (2005). Stock selection based on cluster analysis. *Economics Bulletin*, *13*(1), 1−9.

Dias, A., Pinto, C., Batista, J., & Neves, E. (2016). Signaling tax evasion, financial ratios & cluster analysis. *BIS Quarterly Review.* Working Paper No. 51, 2016.

Ding, K., Hoogduin, L., Peng, X., Vasarhelyi, M. A., & Wang, Y. (n.d.). Clustering based peer selection with financial ratios. *Rutgers, The State University of New Jersey*. Retrieved from http://raw.rutgers.edu/docs/wcars/40wcars/Presentations/KexingXuan Yunsen.pdf

Ferrando, A., & Lekpek, S. (2018). Access to finance and innovative activity of EU firms: A cluster analysis. *European Investment Bank: Working Papers*, 2018/02.

Ferst, R., & Seres, D. (2012). Clustering austrian banks' business models and peer groups in the European Banking Sector. *Financial Stability Report*, 24 December 2012.

Fodor, A., Jorgensen, Randy, D., & Stowe, J. D. (2015). *Forming stock groups with a cluster analysis of common size statements*. Southwestern Finance Association Annual Conference.

Goudarzi, S., Jafari, M. J., & Afsar, A. (2017). A hybrid model for portfolio optimization based on stock clustering and different investment strategies. *International Journal of Economics and Financial Issues*, *7*(3).

Gruener, A., & Schoenenberger, F. (2015). *Risk cluster framework: How to analyse companies by operating leverage*. Retrieved from https://efmaefm.org/0efmameetings/efma%20annual%20meetings/2015-Amsterdam/papers/efma2015_0219_ fullpaper.pdf

Hou, B. (2016). Financial distress prediction of k-means clustering based on genetic algorithm and rough set theory. *Chemical Engineering Transactions*, *51*.

Lemos, C. A. A., Lins, M. P. E., & Ebecken, N. F. F. (2005). DEA implementation and clustering analysis using the K-Means algorithm. *WIT Transactions on Information and Communication Technologies*, *35*, 321–329.

Li, H., & Sun, J. (2011). Mining business failure predictive knowledge using two-step clustering. *African Journal of Business Management*, *5*(11), 4107–4120.

Marvin, K. (2015). Creating diversified portfolios using cluster analysis.

Momeni, M., Mohseni, M., & Soofi, M. (2015). Clustering stock market companies via K-means algorithm. *Arabian Journal of Business and Management Review*, *4*(5).

Perisa, A., Kurnoga, N., & Sopta, M. (2017). Multivariate analysis of profitability indicators for selected companies of Croatian market. *UTMS Journal of Economics*, *8*(3), 231–242.

Setty, D. V., Rangaswamy, T. M., & Suresh, A. V. (2010). Analysis and clustering of nifty companies of share market using data mining tools. *Journal of Engineering Research and Studies*, *1*(1), 152–164.

Setyaningsih, S. (2012). Using cluster analysis study to examine the successful performance entrepreneur in Indonesia. *Procedia Economics and Finance*, *4*, 286–298.

Szucs, G. (2015). The financial analysis of the hungarian automotive industry based on profitability and capital structure ratios. *Central European Business Review*, *4*(1).

Temouri, Y. (2012). The cluster scoreboard: Measuring the performance of local business clusters in the knowledge economy. *OECD Local Economic and Employment Development (LEED) Working Papers*. 2012/13, OECD Publishing.

Tufan, E., & Hamarat, B. (2003). Clustering of financial ratios of the quoted companies through fuzzy logic method. *Journal of Naval Science and Engineering*, *1*(2), 123–140.

Wang, Y.-J., & Lee, H.-S. (2008). A clustering method to identify representative financial ratios. *Information Sciences*, *178*(2008), 1087–1097.

Appendix

R codes for Hierarchical Clustering

```
dataframe<-data.frame(adpe,ptbv,divy,evns,evce,mcts,pt
    cf,pfcf,fcfy)
distmat<-dist(dataframe,
    method="euclidean",diag=FALSE, upper=FALSE)
clust<-hclust(distmat, method="ward.D")
Cluster<-cutree(clust,2)
table(Cluster)
agg<-aggregate(. ~ Cluster,data = dataframe, mean)
dataframe<-cbind(dataframe,Cluster)
rect.hclust(clust,2)
plot(clust,label=,hang=-1)
```

R codes for K-Means Clustering

```
dataframe<-data.frame(adpe,ptbv,divy,evns,evce,mcts,pt
    cf,pfcf,fcfy,hclust_1)
clust<-kmeans(dataframe, centers=2,iter.max=1,nstart=1,
    algorithm="Hartigan-Wong", trace=FALSE)
clust$centers
clust$withinss
clust$totss
clust$tot.withinss
clust$betweenss
clust$size
```

International Journal of Business Analytics and Intelligence
7 (1) 2019, 64-72
http://publishingindia.com/ijbai/

Impact of Behavioural Disposition on Portfolio Investment Decisions of Individual Investors - A Structural Analysis

Renuka Sharma*

Abstract

In the past two decades, behavioural finance, a new criterion of finance gained power on the basis of traditional finance. An ongoing debate between behavioural theorists and traditional theorists provides horizon for interrogation into the changing landscape of investment behaviour. Behavioural finance deals with the influence of psychological; factors on investment decisions. It diverges from the presumption of rationality and can describe how an investor takes investment decisions. In the alternating investment scenario and extreme dynamism in the capital market, investors do not adhere to rational thinking and reflect several dispositions. Therefore, studying how psychology plays a significant role in investment decisions becomes important. A structured questionnaire on impact of behavioral biases on investment decisions of individual investors on 5 point Likert scale was prepared and responses were collected from investors who make their own investment decisions. The paper identified through Confirmatory Factor analysis (CFA) and afterwards through Structural Equation Model (SEM) that Herd Mentality, Overconfidence, Disposition, Mental Accounting Anchoring, Representativeness, Loss Aversion and Regret Aversion are a few biases affects the portfolio investment decisions of individual investors.

Keywords: Behavioural Finance, Behavioural Disposition, Portfolio Investment Decisions, Confirmatory Factor Analysis, Structural Equation Model

Introduction

In the middle of 1950s, the field of finance has been administered by the standard finance model developed by the economist of the University of Chicago. The basic assumption of the standard l finance model is that people are rational. These theories are built on the presumptions that investor behaves rationally and stock and bond markets are efficient. As the financial economist were presuming that people (investors) behaved rationally when making financial decisions, psychologists have found that economic decision are built in an irrational manner, so they challenge this assumption of standard finance. Psychological error and extreme emotional bias can make investors to make bad investment decisions, thereby meaning that they act in irrational manner. Over the past decade, field of behavioral finance has transpired to consider how personal and social psychology impact financial decisions and behavior of investors in general.

Behavioural Finance

Behavioral finance is a new perceptive to financial markets that has become a subject of notable interest to investment group. The field of Behavioural finance seeks to apprehend and define investor decisions by amalgamating topics of Psychology and investing at a micro level, i.e. the decision making process of individuals and groups, and at a macro level where the role of financial markets is contemplated. In everyday business life, people usually take decision, these decisions might have varies outcomes.

* Faculty of Commerce and Management, SGT University, Gurugram, Haryana, India. Email: renuka.sharma1907@gmail.com

Impact of Behavioral Finance in Decision Making

Behavioural finance seeks to find out how investor's psychology affects investment decisions. Investors financial decisions are influenced by various heuristics and prospect biases and their emotions affects their investment decisions profoundly**.** Generally, the investor's behavior deviates from making rational or logical decisions and leans towards being influenced by different behavioral dispositions. These biases impact the investor's rationality in investment decision-making.

Review of Literature

Sewell (2007), Srivastava Aman (2007), Gilovich, Thomas and Griffin Dale (2002), Fromlet (2001), Gervais and Odean (2001), Hirshliefer and Luo (2001), Thaler (1999), Benos (1998), Linter (1998), Olsen (1998), Kyle and Wang (1997) described behavioural finance as a scientific combination of psychology and finance which means that the weight is on psychology based knowledge at the individual level and financial markets findings. It is an integration of classical economics and financial theories within studies investigating psychology and decision making. It studies the influence of psychological happening on financial behavior and tried to explain why and how people make apparently irrational or unreasonable decisions when they save, invest, spend and borrow money. It is an emerging area in the field of financial research, which is at present dominated by theories and theoretical models. This area of behavioral finance borrows the theories from psychology, sociology, and other behavioral sciences to examine the behavioral aspects of the stock market and the investors.

Lakshmi, Visalakshmi, Thamaraiselvan, and Senthilarasu (2013), Iyer and Bhaskar (2002) conducted a research to find as to what degree long term and short term stock investors share different behavioural features. A structural model is employed to compare the features of the investors and examine how investment decision making and behavioural biases are interrelated, as well balanced the relative differences of behavioural biases such as Herding, Representativeness, Overconfidence, Risk Aversion, Disposition Effect and Cognitive Dissonance. Recognition of behavioral characteristics commonly related with investment tenure helps in providing judgments and framing trading strategies. The cognitive effect of investment decision making among investors is measured through a sampling survey of 318 respondents. Structural Equation Modeling [SEM] and path analysis is performed on how investment decision making and the behavioral dispositions are correlated. Analytical results show that the structural path model closely fits to the sample data, implying the role of behavioral biases in investment decision making among individuals. The study also depicts that long term and short term investors prominently differ in behavioral characteristics. Iyer and Bhaskar attempted to identify the various psychological factors affecting the market of Chennai and the various other factors contributing to the market meltdowns and upsurges. Interviews of the leading brokers and the market players in Chennai were done. The authors gave the theoretical explanation of the various psychological facets like fear, greed, overconfidence, overtrading, tickeritis, hope, sentimentality, seeking pride and avoiding regret, self-control, wishful thinking, group think, herd instincts influencing the investor behavior. The researchers stated that the importance of the psychological factors cannot be underplayed as the understanding of the different market participant's psychology gives insight into trading patterns.

Tversky and Kahneman (1974), Slovic and Lichtenstein (1971) evaluated Human beings start evaluating final results by initiating from the beginning values about different situations. That starting point or beginning value may be the partial computation or the formulation of a problem. Adjustments are insufficient in both of the cases. Different initiating points come up with different estimates, which lead to initial values. This phenomenon is called 'anchoring'.

Agarwal (2012), Razek (2011), Hirshleifer and Luo (2011), Oberlechner and Osler (2009) De Bondt and Thaler (1995) defined overconfidence as an misinterpretation of the possibility for a set of events. The concept is significantly reflected by comparing whether the specific probability allocated is greater than the portion that is correct for all assessments allocated to the given probability. It causes people to overvalue their knowledge, undervalue risks and overvalue their ability to control events and the precision of their information. The author claimed that overconfidence originates in people's biased estimation of evidence. Many researchers find evidence for the presence

of the overconfidence bias in different financial decisions. Studies have shown that announcement returns are lower for overconfident bidders as compared to rational bidders. It is the outcome of the strongest identified by psychological research that is to say, investors tend to overvalue the probability of certainty of their information, their successes and capabilities'. They are having past positive outcomes and usually recall only their success than their failures. They find that there are two factors behind this behavior. These are: the underestimation of risk by the investors and overestimation of the success of their own trading program. Investors tend to be overconfident in two areas: they underestimate uncertainty and overestimate their own capabilities.

Kumar and Goyal (2014), Merli and Roger (2013) , Jaiswal and Kamil (2012), Olsen (2008) (Devenow & Welch, 1996), Scharfstein and Stein (1990) stated herding refers to the situation wherein coherent people start behaving illogically by copying the judgements of others while making decisions or it is the tendency of individual to chase the actions (rational or irrational) of a wider group or in other words it refers to the situation where forecasters or investors tend to shade their forecasts or investment decisions in the direction of a reference group (i.e., other market participants). It is the tendency of individuals to imitate the actions (rational or irrational) of a larger group; individually, however, most people would not necessarily make the same choice.

Rekik and Boujelbene (2013), Grinblaltt and Han (2004) explained the main idea underlying mental accounting is that decision-makers tend to divide the different types of gambles they face into separate accounts, and they apply prospect theoretic decision rules to each account by ignoring possible interplay between the accounts. Therefore, mental accounts can be divided not only with respect to time but also according to their subjects. It provides a foundation for the way that decision makers set referral points for the accounts that determine gains and losses.

Waweru et al. (2008), Barberis and Thaler, (2003), Ritter (2003), DeBondt and Thaler (1995), (Kahneman & Tversky, 1974) referred representativeness bias as the degree of resemblance that an event has with its parent population or the degree to which an event similitude to its population. It may result in some dispositions such as people put too much importance on current experience and ignore the average long-term rate. A typical example for this bias is that investors often conclude a company's high long-term growth rate after some quarters of increasing. It also leads to the so-called "sample size neglect" which arises when people try to infer from too few samples In stock market, when investors seek to buy "hot" stocks instead of poorly performed ones, this means that representativeness is there . This behavior is an explanation for investor overreaction.

Fogel and Berry (2006), Lehenkari and Perttunen (2004), (Shiller, 1998 asserted that investors consistently engage in behavior that they regret later. They avoid selling shares that have decreased in value, and readily sell shares that have increased in value Psychologists have found that individuals who make decisions that turn out badly have more regret (blame themselves more) when the decision was more unconventional. For example buying a blue chip portfolio that turns down is not as painful as experiencing the same losses on unknown start-up firm. Any losses on a blue-chip stock can be more easily attributed to bad luck rather than bad decision making and cause less regret. Regret is an emotion occurs after people make mistakes. Investors avoid regret by refusing to sell decreasing shares and willing to sell increasing ones. Moreover, investors tend to be more regretful about holding losing stocks too long than selling winning ones too soon.

Lehenkari and Perttunen (2004), Barberis and Thaler (2003), Barberis and Huang (2001), Odean (1998a) referred Loss aversion as the different level of mental penalty people have from a similar size loss or gain. There is proof showing that people are more disappointed at the anticipation of losses than they are pleased by similar gains. Moreover, a loss coming after previous gain is proved less painful than usual while a loss arriving after a loss seems to be more painful than usual. They further found that both positive and negative returns in the past can raise the negative relationship between the selling trend and capital losses of investors, suggesting that investors are loss averse.

Limitations of the Study

- The researcher has taken her home state as an area for sampling that's why the conclusion drawn from this sample cannot be generalized on whole of India.

- Sample size is small and the researcher has taken her home state as area of research to make the data collection process easy and convenient.

Objectives

The primary objective of the study is to examine the influence of behavioral disposition on portfolio investment decisions of individual investors.

Hypothesis

H0: The various behavioral dispositions and their influence are not related with portfolio investment decisions of individual investors.

H1: The various behavioral dispositions and their influence are related with portfolio investment decisions of individual investors.

Methodology

It is a study which gives the influence of behavioral dispositions on portfolio investment decisions of individual investors. On the basis of the review of literature the following biases has been taken for the study viz. Herding bias, Overconfidence bias, Disposition bias, Mental Accounting bias, Anchoring bias, Representativeness bias, Loss Aversion bias and Regret aversion bias. The study undertakes five heuristic based biases and three prospect theory based biases. Expected Return, Amount of Investment, Time period of Investment has been taken as factors of Portfolio investment decisions.

The overall sample design is as follows:

Universe: Adults above 18 years of age.

Sampling Frame: List of investing clients from equity brokers in Rajasthan.

Sampling Unit: List of male and female investors in Rajasthan.

Sampling Technique: Non-Probability Convenience Sampling method.

Sample Size: 600 respondents.

Tools of Data Collection

The data for the present study has been collected through a structured questionnaire, questionnaire description. In order to study the investor's behavior, the researcher has used various behavioral finance variables. (Questionnaire statement reliability Cronbach's Alpha 0.856 for 64 items statements).

Data Analysis and Interpretation

For the purpose of the study Confirmatory Factor Analysis (CFA) and afterwards Structural Equation Modeling (SEM) has been used.

The CFA model is the focal point of the link between factors and their measured variables, within the framework of SEM, it represents what is called as a measurement model. In this study, the model was developed 'a priori', hence the CFA was used.

Structural equation modeling is exercised for examining the influence of exogenous constructs on investment decisions. It establishes path for concurrent testing of an entire model that covers multiple imaginary relationships. Structural equation model that consists of all constructs and hypothesized relations are examined.

Confirmatory Factor analysis was executed by using AMOS 24.0 software, it was calculated by using maximum likelihood among the variable via correlation. CFA was executed with the objective to define individual constructs and test whether the data fit a hypothesized measurement model. Validity and reliability of measurement model is evaluated.

To estimate model fit, some criteria such as the chi-square to degrees of freedom ratio (CMIN/df ratio), root mean squared residual (RMR), the comparative fit index (CFI), Parsimony adjusted mean and residual mean squared error of approximation (RMSEA) are determined. For CMIN/df value remain < or = 2 or 3. Values of CFI above 0.90 indicate good model fit. Parsimony closer to 1 is generally acceptable for good model fit. For RMSEA and RMR values below 0.05 indicate close fit, 0.08 indicate an adequate fit. Confirmatory factor analysis diagram is as follows:

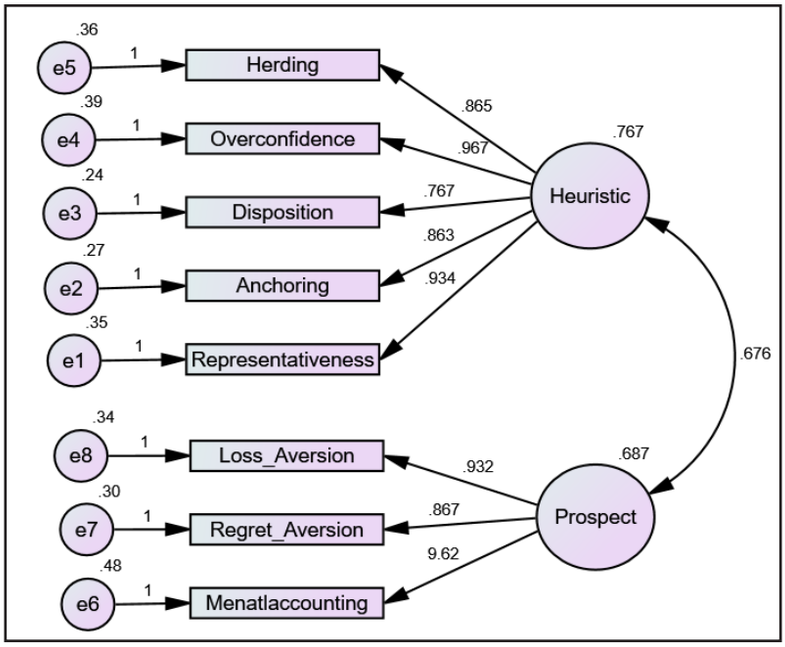


Fig. 1: Confirmatory Factor Analysis

The Measurement Model

Results for each construct are shown in Table 1. From 8 variable 2 latent variables are manifested named Heuristic: Herding (F1), Overconfidence (F2), Disposition (F3), Anchoring (F5) and Representativeness (F6), Prospect includes Loss Aversion (F7), Regret Aversion (F8) and Mental accounting (F4). The CFA model fit the data well with CMIN/df is 2.31; GFI = 0.957; CFI = 0.956; RMSEA = 0.05 based on the statistics, the model fits the data very well, generating a good fit to the current data.

Table 1: Goodness of Fit Test

Latent Constructs and Variables	*Factor Loadings*	*Composite Reliability*	*Average Variance Extracted*
Heuristics			
Herding	0.865		
Over Confidence	0.967		
Disposition	0.767	0.767	0.71
Anchoring	0.863		
Representativeness	0.934		
Prospect			
Loss Aversion	0.932		
Regret Aversion	0.867	0.687	0.643
Mental Accounting	0.962		
Goodness-of-Fit	Statistics		
X2/DF (1to 4)		2.31	
GFI (> 0.90)		0.957	
AGFI (>0.80)		0.955	
CFI (>0.90)		0.956	

Latent Constructs and Variables	*Factor Loadings*	*Composite Reliability*	*Average Variance Extracted*
NFI (>0.90)		0.944	
RFI(>0.90)		0.924	
IFI(>0.90)		0.917	
TLI(>0.90)		0.939	
RMSEA(<0.08)		0.05	
RMR(<0.08)		0.025	

Source: Primary Data

It is apparent that in the above Table 1, latent construct take the factor loading which ranges from 0.863 to 0.962, indicating strong support for construct validity. Correspondingly, the average variance extracted values for heuristics and prospect theory are higher than the benchmark level of 0.50. Composite reliability coefficients are higher than 0.60 for both latent constructs, which shows that high internal reliability. Goodness of fit statistics of the measurement model further confirmed a good fit with the data. Various indexes are used here to measure the fit of the model including Goodness-of-Fit Index (GFI, 0.957); Adjusted Goodness-of-Fit Index (AGFI, 0.955); Normal-Fit Index (NFI, 0.944); Relative Fit Index (RFI, 0.924); Incremental Fit Index (IFI, 0.917); Tucker-Lewis Index (TLI, 0.939); and Root Mean Square Error of Approximation (RMSEA, 0.05) and (RMR.0.025) which are perfectly fit with the data.

Structural Equation Model

Structural Equation Model is a statistical technique for testing and approximating causal relations using a combination of statistical data and qualitative causal assumptions. This model is used to draw relationship between the variable. Model combines both the aspects factor analysis and multiple regression. This study profusely applied structural equation model and it is developed to test the influence of heuristic and prospect biases on the investment decisions. Since, the proposed measurement model is in line with the data, the hypotheses are tested. Relationship among the constructs is showed in the Fig. 2. This study endeavored to test that there is a direct and positive relation between the biases and the investment decision

It is evident from the Fig. 2 below, among the different paths hypothesized in the model, all the paths are found significant at $p < 0.05$ .This study tested the models illustrated in the above chart, which provide the path diagrams for the models on the defined investment decisions .Generally in Structural Equation Modelling the fit of the model is assessed by chi-square measure is not always straightforward .because chi–square is very sensitive to the sample size. Due to this disadvantage, various kinds of fit indexes have been developed that are independent sample size. The goodness of fit is presented in Table 2. To estimate model fit, some criteria such as the chi-square to degrees of freedom ratio (CMIN/df ratio), root mean squared residual (RMR), the comparative fit index (CFI), Parsimony adjusted mean and residual mean squared error of approximation (RMSEA) are determined. For CMIN/Df value remain < or = 2 or 3. Values of CFI above 0.90 indicate good model fit. Parsimony closer to 1 is generally acceptable for good model fit. For RMSEA and RMR values below 0.05 indicate close fit, 0.08 indicate an adequate fit.

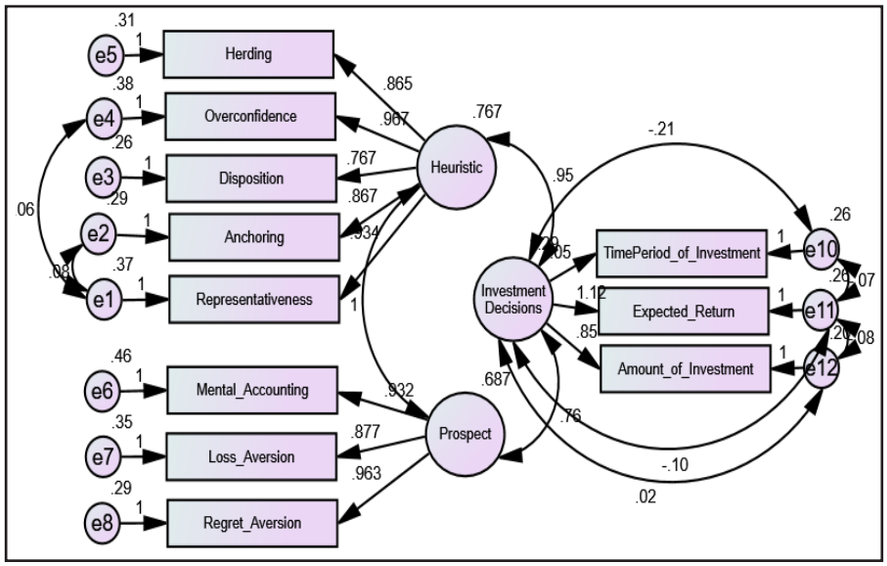


Fig. 2: Structural Equation Model

The Measurement Model

Results for each construct are shown in Table 2. From 8 variables 2 latent variables are manifested named Heuristic (F1, F2, F3, F5 and F6), Prospect (F7, F8 and F4). The CFA model fit the data well with CMIN/ df is 2.14; GFI = 0.955; CFI = 0.966; RMSEA = 0.069 based on the statistics , the model fits the data very well, generating a good fit to the current data.

Table 2: Goodness of Fit Test

Latent Constructs and Variables	*Factor Loadings*	*Composite Reliability*	*Average Variance Extracted*
Heuristics			
Herding	0.865		
Over Confidence	0.967		
Disposition	0.767	0.767	0.73
Anchoring	0.867		
Representativeness	0.934		
Prospect			
Loss Aversion	0.877		
Regret Aversion	0.963	0.687	0.64
Mental Accounting	0.932		
Goodness-of-Fit	Statistics		
X2/DF (1to 4)		2.14	
GFI (> 0.90)		0.955	
AGFI (>0.80)		0.923	
CFI (>0.90)		0.966	
NFI (>0.90)		0.955	
RFI(>0.90)		0.936	
IFI(>0.90)		0.97	
TLI(>0.90)		0.952	
RMSEA(<0.08)		0.069	
RMR(<0.08)		0.034	

It is apparent that in the above Table 2, latent construct take the factor loading which ranges from 0.865 to 0.967, exhibiting strong support for construct validity. Correspondingly, the average variance extracted values for heuristics and prospect theory are higher than the benchmark level of 0.50. Composite reliability coefficients are higher than 0.60 for both latent constructs, which shows that high internal reliability. Goodness of fit statistics of the measurement model further confirmed a good fit with the data. Various indexes are used here to measure the fit of the model including Goodness-of-Fit Index (GFI, 0.955); Adjusted Goodness-of-Fit Index (AGFI, 0.923); Normal-Fit Index (NFI, 0.955); Relative Fit Index (RFI, 0.936); Incremental Fit Index (IFI, 0.97); Tucker-Lewis Index (TLI, 0.952); and Root Mean Square Error of Approximation (RMSEA, 0.069) and (RMR.0.034) which are perfectly fit with the data.

It is identified that this structural equation model accomplished significant improvement in terms of its goodness-of-fit indices as all recommended values are consistent. The result shows the data.

All these measures were found much better.

Conclusion

It can be concluded from the study that there exists a direct relationship between behavioral dispositions and portfolio investment decisions. It is inferred that both heuristic biases namely Herding, Overconfidence, Disposition, Anchoring, Representativeness and Prospect biases namely Loss aversion, Regret Aversion and Mental Accounting bias are considerably impact the portfolio investment decisions of individual investors.

The impact of the behavioral disposition on expected rate of return, amount of investment and time period of investment shows that the herd mentality, Overconfidence, Mental Accounting, Loss Aversion, Regret Aversion, Disposition, Representativeness, Anchoring all these biases have significant impact on portfolio investment decisions.

References

Agrawal, K. (2012). A conceptual framework of behavioural biases in finance. *The IUP Journal of Behavioural Finance*, *8*(3), 7–24.

Benos, A. (1998). Aggressiveness and survival of overconfident traders. *Journal of Financial Markets*, *1*(3–4), 353–383.

Chandra, S., & Sharma, D. (2010). Investment management by individual investors: A behavioral approach. *The IUP Journal of Behavioral Finance*, *7*(1), 7–18.

De Bondt, W. F. M., & Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. In R. Jarrow, V. Maksimovic, & W. T. Ziemba (Eds.), *Handbooks in operations research and management science* (Vol. 9, pp. 385–410). Amsterdam: Elsevier.

De Bondt, W., & Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. *Finance, Handbooks in Operations Research and Management Science*, *9*, 385–410. Amsterdam: North Holland.

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, *40*(3–5), 603–615.

Evans, D. A. (2006). Subject perceptions of confidence and predictive validity in financial cues. *Journal of Behavioral Finance*, *7*(1), 12–28.

Fogel, O., & Berry, T. (2006). The disposition effect and individual investor decisions: The roles of regret and counterfactual alternatives. *Journal of Behavioral Finance*, *7*(2), 107–116.

Fromlet, H. (2001). Behavioural finance – Theory and practical application. *Business Economics*, 36–39.

Gervais, S., & Odean, T. (2001). Learning to be Overconfident, *Review of Financial Studies*, *14*(1), 1–27.

Gilovich, T., & Griffin, D. (2002). *Introduction – Heuristics and biases: Then and now* (pp. 1–18). Cambridge: Cambridge University Press.

Grinblaltt, M., & Han, B. (2004). *Prospect theory, mental accounting, and momentum*, Yale ICF Working Paper No. 00-071, 1–38.

Hirshliefer, D., & Luo, G. (2001). On the survival of overconfident traders in a competitive security market. *Journal of Financial Markets*, *4*, 73–84.

Iyer, B. S., & Bhaskar, K. S. (2002). Investor's psychology: A study of investor behaviour in the indian capital market. *Finance India*, *16*(4), 1357–1375.

Jaiswal, B., & Kamil, N. (2012). Gender, behavioral finance and the investment decision. *Business Review*, *7*(2), 8–22.

Kumar, S., & Goyal, N. (2014). Behavioral biases in investment decision making – A systematic literature review. *Quantitative Research in Financial Markets*, *7*(1), 88–108.

Kyle, A., & Wang, F. (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *Journal of Finance*, *52*, 2073–2090.

Lakshmi, P., Visalakshmi, S., Thamaraiselvan, N., & Senthilarasu, B. (2013). Assessing the linkage of behavioural traits and investment decisions using SEM approach. *International Journal of Economics and Management*, *7*(2), 221–241.

Lehenkari, M., & Perttunen, J. (2004). Holding onto the losers: finish evidence. *The Journal of Behavioral Finance*, *5*(2), 116–126.

Lintner, G. (1998). Behavioural finance: Why investors make bad decisions. *The Planner*, *13*(1), 7–8.

Merli, M., & Roger, T. (2013). What drives the herding behavior of individual investors? *Finance*, *34*(3), 67–104.

Oberlechner, T., & Osler, C. (2009). Overconfidence in Currency Markets, 1–46.

Odean, T. (1998a). Volume, volatility, price and profit when all trades are above average. *Journal of Finance*, *53*(6), 1887–1934.

Olsen, R A. (2008). Cognitive dissonance: The problem facing behavioral finance. *Journal of Behavioral Finance*, *9*(1), 1–4.

Olsen, R. (1998). Behavioural finance and its implications for stock-price volatility. *Financial Analyst Journal*, *54*(2), 10–17.

Razek, Y. H. (2011). An overview of behavioural finance and revisiting the behavioural life cycle hypothesis. *The IUP Journal of Behavioural Finance*, *8*(3), 7–24.

Ricciardi, V., & Simon, H. (2000). What is behavioural finance? *Business Education and Technology Journal*, *1*(1), 1–9.

Rekik, Y. M., & Boujelbene, Y. (2013). Determinants of individual investors' behaviors: Evidence from Tunisian stock market. *Journal of Business and Management*, *8*(2), 109–119.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American Economic Review*, *8*(3), 465–479.

Sewell, M. (2007). *Behavioural finance.* Cambridge: University of Cambridge Press.

Shiller, R. J. (1998). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior and Organization*, *12*(1), 47–66.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behaviour and Human Performance*, *6*(6), 649–744.

Srivastava, A. (2007). An analysis of behavior of investors in India. *The ICFAI University Journal of Behavioral Finance*, *4*(1), 43–52.

Thaler, R. H. (1999). The end of behavioural finance. *Financial Analyst Journal*, *55*(6), 12–17.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 1124–1131.

Waweru, N. M., Munyoki, E., & Uliana, E. (2008). The effects of behavioral factors in investment decision-making: a survey of institutional investors operating at the Nairobi Stock Exchange. *International Journal of Business and Emerging Markets*, *1*(1), 24–41.

International Journal of Business Analytics and Intelligence
7 (1) 2019, 73-87
http://publishingindia.com/ijbai/

An Inquiry into the Impact of Digitization and Customized ERP Applications on Twin Engineers' Overall Efficiency - An Empirical Study Approach

Atul R. Deshpande*

Abstract

This research paper entitled "An Inquiry into the Impact of Digitization and Customized ERP Applications on Twin's Overall Efficiency – An Empirical Study Approach" has examined, through exploratory method and robust statistical analysis, whether 'digitization and customized ERP applications in the manufacturing process has affected overall efficiency at Twin Engineers. With data base (2009-10 to 2017-18), through primary and secondary sources and a 'focus group' survey along with the construction of 'efficiency indices' and 'relevant statistical analysis', that consists of 'regressions' and Granger causality tests, the analysis shows that the overall efficiency after digitization period has improved considerably. The hypotheses testing (parametric and non-parametric) done in this case also supports the conclusion cited above. The research work has used a linear regression model which has used 'turnover data' as proxy for efficiency and a host of factors such as persons employed, machines dispatched, different costs etc. as independent predictors which are highly correlated to turnover as predicted variable. The core discussion in the paper consists of analytical and evaluative aspects which are based on 'findings' from the survey. This discussion is carried out through 'a case study approach'. The paper is concluded with important recommendations along with implications and limitations of this work.

Keywords: Efficiency, Digitization, Customized ERP Applications, Regressions, Granger Causality

Introduction and Company Profile

(1.A) This exploratory case-based study is an attempt to examine whether digitization and customized ERP process application has affected Twin Engineer's overall efficiency. This research, for its, comparative study purpose (pre-digitization versus post-digitization) has looked into Twin's strategic policy of growth from the era of 'local excellence' to the 'era of the application of digitization, customized ERP applications and IoTs (Internet of Things)'. Twin Engineers (Pvt. Ltd.) is an ISO 9001 – 2008 Certified company established in the year 1993. It is a leading manufacturer of 'Adhesive and Sealant Dispensing Machines' and 'Industrial Fluid Filling Machines'. Twin Engineers has developed more than 38 products which cater to a diverse and wide market segment including Flexible Packaging, Automobile, Electronics and other industrial sectors. Although Twin is primarily catering to domestic market, its foreign market sweep in terms of exports stands at nearly 10% of the total value of company turnover in 2012 (Focus Group Survey 2018). Twin has its valued customers that include Maruti Suzuki India Ltd., Honda Siel Cars India Ltd., Yamaha Motors Ltd., Honda Motor Cycles and Scooters India Ltd., Toyota Kirloskar Motor Pvt. Ltd., Ford India Pvt. Ltd., Renault Nissan Automotive India Pvt. Ltd.and many more to add to the list.

* Director and Professor, Institute of Management Education and Research, Belgaum, Karnataka, India.
Email: atdash10@yahoo.com

Objectives of Research

- To examine and understand the digitization and customized ERP application processes functionally.
- To know the integration process between specific ERP application and a particular manufacturing process.
- To establish a correlation between the turnover as proxy for efficiency and other independent variables such as costs, persons employed and machines dispatched.
- To analyze the efficiency scenario prior to the introduction of the digitization process and post digitization process.
- To analyze whether overall efficiency (measured in terms of saving of time, cost, labour productivity and volume of output) has increased or otherwise.
- To identify the loopholes and improvements in the process of implementing the digitization process and the customized ERP application process in the manufacturing activity.

Hypotheses of Research

- There exists a significant correlation between turnover used as 'proxy' for efficiency and other independent variables such as costs, persons employed and machines manufactured and dispatched.
- The performance of efficiency before digitization and after digitization is different from each other.
- Persons employed alone and/or machines dispatched does not Granger cause turnover to change.

Research Methodology

This research is primarily based on 'case study approach' while understanding Twin Engineers' overall efficiency. To meet with the objectives, the research uses 'exploratory method'. While using exploratory method, the research has used 'primary data' in the form of survey conducted with 'focus group'. To carry forward its efficiency analysis, this research has used 'secondary data' in the form of Company documents which consist of data on turnover, costs, persons employed and machines dispatched, used and installed capacity, activity and department-wise data of time consumed etc. The data on these variables has helped construct 'efficiency index', 'time efficiency and labor efficiency indices etc. The personal survey conducted as 'focus group' consists of 'department heads' who are also the owners of 'digitization and customized ERP application process. Since this is a 'focus group' survey only 5 persons as experts are used as 'sample' of this research. A separate interview was conducted with Company MD. For understanding relationship among dependent and independent variables from the model, a correlation analysis is used. To test two of the three hypotheses, the 'T' paired parametric and 'Sign and Wilcoxon non-parametric tests are used. To find out whether the data used from the questionnaire is reliable or otherwise, the Data Reliability test is conducted and the test result on the Cronbach's Alpha based on standardized items has turned out to be .948.

The Model

With a view to examining Twin's efficiency, the researcher has employed turnover as proxy for operational efficiency. Further, it has been understood that efficiency is a function of a host of independent variables along with an error term. Our model, therefore, looks as follows:

TUR = f (PER, MACH, TC, DC, IC)

In this function the meaning of the subscripts used is as follows:

TUR = Turnover, PER = Persons employed, MACH= Machines dispatched, DC = Direct Cost, IC = Indirect Cost

$$TUR = \alpha + \beta_1 PER + \beta_2 MACH + \beta_3 DC + \beta_4 IC + u$$

$$TUR = 1103815.017 - .058\ PER + .412\ MACH + .496\ DC + .183\ IC + 5717336.132\ u$$

(.044) (-.467) (1.377) (3.182) (1.679)

The model summary through SPSS output (Annexure, (C)) states that the R = .996, R Square = .992, adjusted R Square = .985. These numbers suggest that coefficient of correlation between dependent and independent variables is very high. Except the variable 'number of persons employed' all other variables have positive correlation with turnover used as proxy for efficiency. Number of persons employed shows a negative correlation with

turnover. But its coefficient is not very high and therefore insignificant. Rests of the independent factors do have their correlation with turnover but even they also do not show a very noticeable coefficient of correlation. In our model the independent factors do not show multicolinearity since the Durbin Watson statistics has turned out to be 2.17. The regression we have run and used does not show causality between our independent variable and a host of independent variables. To find out whether any causality exists between dependent and independent variables we have done the Granger causality test through 'R'. We have made more relevant use of the outputs of the Granger causality in our analysis of hypothesis testing.

(1.F)Testing of Hypothesis (Annexure, (E and F)

In this research out of the three major hypotheses the researcher has tested the following two immediately relevant hypotheses.

Hypothesis 1

- H_{01}: The performance of efficiency before digitization and after digitization is not different from each other.
- $H_{11:}$ The performance of efficiency before digitization and after digitization is different from each other.

To test this hypothesis the secondary data on time taken (time efficiency) to complete a specific activity before digitization and after digitization has been considered. To do that, the two activities have been considered. One consists of 'design' activity and the other consists of 'procurement' activity. The parametric "paired't' test has been conducted on the available data. The test results show that the mean score (4.0833) on the completion of design activity before digitization is considerably higher and the 'p'value (.02) denoting significance is lower than .05. The coefficient of correlation between design to design is .891 which is highly significant. With the similar test conducted on 'procurement' data the results are similar. For example, the mean score (1.8333) on the completion of procurement activity before digitization is considerably higher and the 'p'value (.04) denoting significance is lower than .05. The coefficient of correlation between the two pairs is .867 (considerably higher) at significance level .000. The 'Wilcoxon Signed Ranks Test', a non-parametric, has also been conducted on time taken to complete design activity before digitization and after digitization. The time taken after digitization is greater than time taken before digitization shows all 'negative differences (all 6 are negative differences) and time taken after digitization is greater than time taken before digitization shows positives equal to zero. The ties between the two are also equal to zero. The test statistics is .02 which is less than 05. And the Binomial distribution test statistics is also significant at the level equal to .03 which is also less than .05.

From the above discussion we can reject the null hypothesis and accepts the alternative hypothesis.

Hypothesis 2:

- $H_{02:}$ Persons employed alone and/or machines dispatched does not Granger cause turnover to change.
- $H_{22:}$ Persons employed alone and/or machines dispatched does Granger cause turnover to change.

To test this hypothesis the researcher has run the Granger causality test (Annexure, (D)) through 'R' software. In the causality function it was tested that whether persons employed alone as independent factor granger causes turnover to change as a dependent variable? The test statistics shows the 'p' value equal to 0.20 which is greater than .05 at significance level of 0.006. The 'F'statistics estimated is 2.0726 is far greater than the 'p'value. This result suggests that persons employed alone does not Granger cause turnover to change. This further suggests that it might be a combined effect of digitization and customized applications along with factors from our original model that affect turnover to change. Based on this result we accept the null hypothesis and reject the alternative hypothesis.

Review of Literature and Identification of Gaps

Since this research work is an independent case study which has its own problem area and strategies of growth, there are hardly any previous studies which can be compared with this research on the basis of 'similarities or differences'. In the past there appear a good number of studies which have gone into examining the impact of ERP product and processes on productivity and profitability. But the problem areas of such studies are found to be entirely different from what has been discussed in this paper. Also important is the fact that the previous studies

have not used in multiple numbers the case study form in their development of research studies. Notwithstanding, a few studies need to be mentioned.

- A study (2016) entitles 'Enterprise Resource Planning ……..Operational Efficiency' by Madanhire, Ignatio and Mbohwa, Charles discusses how ERP framework was designed to reduce work in progress on the shop floor and inventory of South African Company.
- In another study (2006) entitles 'Improvement in Operational Efficiency Due to ERP Systems Implementation: Truth or Myth? by Memuri, Vijay and PalviaShailendrathe authors investigated the impact of ERP systems implementation on operational efficiency of medium sized firms in the pharmaceutical and chemicals industry. Their analysis of the data indicates that for a majority of the firms improvement of operational performance expected due to ERP systems did not materialize.

In addition to these major efforts, a number of other studies have just examined the significance of application of either SAP or ERP in the context of specific manufacturing processes without going into any serious analytical research. These many studies have provided a descriptive case approach without making any data-based analysis. This research primarily looks at this major gap and bridges the same through a combined analysis of primary and secondary data with construction of linear regression model and hypotheses testing. The general framework used is a case approach. Most of the previous works in this context have not brought out the limitations of their research and have also not included in their research the implications of their studies for respective companies. This research has also made a conscious effort to take care of these gaps.

Discussion: Analysis and Evaluation

Findings from 'Focus Group Survey

- It has been found from the survey that Twin Engineers has been using digitization process and customized ERP applications in the entire assembly process of adhesive dispensing technology and in all its corresponding and relevant departments for more than 5 years. This seems to be reasonably adequate period to look into the results of the impact of digitization technology and customized ERP on the overall assembly process of adhesive dispensing technology machines.
- Among the reasons that have prompted Twin Engineers to use digitization and customized ERP applications include product suitability, cost saving thought and most prominently efficiency improvement criterion.
- It has also been noticed through the survey that prior to using digitization process and the customized ERP applications; a few of the manufacturing and non-manufacturing processes were carried out manually. Even during this time period, a few of the processes were digitized. When a few important processes were manually carried out, there were important issues related to operational efficiency. In a descending order of importance, the issues observed included high time consumption in implementing processes, loss of resources, high cost, low efficiency and finally quality compromise.
- Through the survey conducted with a 'focus group', it is found that the digitization process and customized ERP products are partially integrated with required select processes. This, therefore, suggested that there exists a greater scope for complete integration of digitization processes and customized ERP products with the existing required processes.
- Since there has been partial integration of digitization processes and the customized ERP variants, the existing processes are subject to multiple issues related to low efficiency. The issues included are the existing processes prior to digitization and /or with partial digitization have been greater time consuming, partial digitization appeared to be technically complicated and therefore required process specific training to technical and non-technical employees.
- In spite of partial integration of digitization and customized ERP variants, a considerable improvement took place in overall efficiency. The variables in which a noticeable improvement is being noticed are considerable time saving in implementing processes, considerable improvement in labor productivity, and improvement in quality of applications and products and considerable saving in cost.
- If we need to use the number to show the average rise in efficiency level, it is found from the focus

group that as compared to pre-digitization situation the efficiency level grew by 20 to 25% after digitization.

- While setting up digitization process and using customized ERP applications, the overall costs have increased to an average level of 10 to 20%. In some exceptional applications the costs have increased by 30 to 40%. The increase in costs has been witnessed since 2014-15 when digitization process and customization of ERP applications are setting their tone.
- Through the survey it has also been found that the RoI, after the implementation process of digitization and customized ERP application, has stood at nearly 10 to 15%. Since customized ERP applications and digitization process have been helping reduce costs and increase turnover, the overall profitability of the Company has been growing. Looking at the present RoI, the company has greater scope to enhance its RoI through the application of ERP processes which are in their initial stages of development and still getting streamlined.
- Since the introduction of digitization and application of ERP customized solutions, it has now been 3 years that the RoI has been growing steadily. There are three specific reasons or factors accountable for the growth in RoI. First, in last three years the number of machines dispatched has been on the rise. Second, labor efficiency has been improving (labor efficiency index) and total cost to turnover ratio has fallen after digitization and customized ERP applications to the existing processes.
- Our survey has also found that though the pace and impact of digitization process and the ERP application drive have been quite satisfactory, there are important issues that need to be taken care of on priority basis. These issues include important problems such as how to make structural adjustments and bring improvements in the initial stage of the development of digitization and ERP application process, how to improve on the present level of employee training specifically required for enhancing the speed of implementation process, how to reduce the cost of establishment and implementation of digitization process and customized ERP applications to the existing processes and with these important issues how to improve upon the overall efficiency after introducing digitization process.
- It has been found through the survey that each crucial department is trying at its best level to adapt to structurally new design of assembling and manufacturing processes. To overcome the problems cited earlier, different improvements are taking place department wise. For example, they include such things as employee training and empowerment, new knowledge addition and enhancement, online help in acquiring new skills and digitized processes and enhancement of reporting system. These steps will take care of improving on time, cost, quality and overall efficiency.
- There is a general feeling among employees and medium and higher cadre management people that the digitization and customized ERP application have been a successful drive from the point of view of bringing about a structural change that has taken care of overall productivity and efficiency.

In spite of the existing problems, there seems to be a general consensus among employees of Twin Engineers that the success level of the positive impact of digitization and customized ERP applications on the overall improvement in operational efficiency has been more than 'least successful' on the Likert Scale of 1 to 5 where 1 being 'least successful' and 5 being 'most successful'

Analysis of Efficiency

To understand overall operational efficiency of assembling and manufacturing processes, we have used 'capacity utilization' as a proxy to analyze operational efficiency. In this case, therefore, we have constructed 'efficiency index' by taking into account a ratio between actual capacity utilized against installed capacity built-up over a variety of product category. Although data is available on product-wise category, we have not constructed product-wise efficiency index. Instead, we have taken a total of installed capacity and actual capacity by simply adding a number of machines in each category of product. Secondly, the efficiency index is constructed over a period from 2015 to 2018. The reason for considering these three years has been the fact that Twin Engineers has started digitization process after 2012 and the robust results of the impact of digitization process and the customized ERP applications

are being witnessed after 2015 onwards. The first three years have been utilized to streamline and strengthen the structural changes introduced in the assembly line and manufacturing processes. A few analytical underpinnings of overall operational efficiency are as follows:

Efficiency Index (Efficiency Through Capacity Utilization) (Annexure, Table 4)

- Owing to improvement in time, labor and cost efficiency the efficiency index (capacity utilization) is ranging in the band of 0.77 and 0.88.As compared to the pre-digitization period (in the years before 2012) when the efficiency index was in the range of 0.55 and 0.65 due to greater structural bottlenecks. After a few years of digitization process, there has been a definite improvement in the efficiency index.
- A product-wise analysis shows that the efficiency index is steady (except for the year 2017-18) primarily due to satisfactory performance on efficiency level by product and process categories such as packaging, MMD, Filling (this is one activity where actual capacity surpasses the installed capacity), SPM, SCDM (actual capacity is greater than the installed capacity) etc. In the year 2017-18, the efficiency index shows a noticeable fall mainly due to unsatisfactory performance in categories of product such as ROBOTIC and SCDM. In that specific year the capacity utilization in the category of ROBOTIC was just 16.66% of the installed capacity (Annexure, LEI)
- During the same period, labor efficiency index has also fallen considerably from its previous level of 1.75 (2016-17) to 1.50 (2017-18). This suggests that labor productivity (machines dispatched to labor employed ratio) experienced a fall in that specific year.

Efficiency Through Costs

To know about operational efficiency from a different dimension we have considered 'total cost to turnover ratio' (Annexure, Table 1) as another proxy. The behavioral pattern of this ratio is analyzed over two time periods. The first period looks at the pattern from 2009-10 to 2012-13; which is essentially a pre-digitization period. And the second period takes into account the behavioral pattern from 2013-14 to 2017-18; which is post-digitization period. Our analysis in this regards is as follows:

- In the pre-digitization period the total cost to turnover ratio was in the band of 91 to 95%. During the same period, the direct cost to turn over ratio was in the range of 71 to 81 percent. The direct costs mainly comprised of the cost of purchases and the cost of salary. This range was, by any standard, very high in any organization. During this period there was hardly any digitization or the use of ERP related customized application to the existing processes. Most of the processes were done manually. What went true with direct costs was also valid in the case of indirect costs. The indirect cost to direct cost ratio (Annexure, Table 2(b)) was in the range of 26 to 39%. The indirect costs mainly comprised of such costs as cost of sales promotion, cost of travel, cost of electricity used, high office expenses, high cost of depreciation including insurance costs, professional, and legal and consultancy charges (The Company hired German Technical Consultancy).
- During post-digitization period the total cost to turnover ratio has declined and now it stands at the range of 86 to 87% specifically since the year 2015-16. During this period, the direct cost to turnover ratio (Annexure, Table 1(a)) has shown a considerable decline and has stood in the range of 62 to 71%. This has been mainly due to effective input utilization owing to digitization and the application of customized ERP processes. Along with this, the indirect cost to direct cost ratio has also declined and has stood at the range of 25 to 31% except for the year 2016-17. In spite of this, we may argue that the present level of indirect costs is high. This is high mainly because of the fact that digitization and application of customized ERP processes have yet to produce optimum results because they are, at present, in the nascent stage of development.
- In the post-digitization period it can also be seen that the rate of growth in total cost as compared to the rate of growth in total revenue (Annexure, Table 3, (a and b)) is found to be marginally low. This has also added to the overall cost efficiency.

Efficiency Through Labor Productivity (LEI, Annexure, Table 5)

Company's overall operational efficiency is also affected by labor utilization process along with machines. To understand the impact of labor productivity on overall operational efficiency, we have constructed 'labor efficiency index' (Table 5) To construct labor efficiency index, we have used a ratio between number of machines manufactured and dispatched and number of persons employed. Our logical analysis in this regard is as follows:

- In the pre-digitization period, most of the manufacturing processes were being done manually. With limited market size and process bottlenecks, the 'labor efficiency index' during this time period was in the range of 1.11 to 1.44.
- In the post-digitization period the labor efficiency index has gone up and has stood between the range 1.50 and 1.83. This upward shift in the labor efficiency index was mainly due to the introduction of digitization and customized ERP application processes.
- The improvement in labor efficiency is also viewed from the point of view of total time taken by labor along with digitization in specific activity carried out in specific department. The time taken is measured in terms of number of days utilized in completing total activities in various departments. The data on time taken to complete activities department-wise comparing pre-digitization period performance with post digitization shows some interesting results. For instance, in Design department, having considered various activities taken together, in the post-digitization period the total time taken to complete activities shows a big fall to the extent of 72%, in the case of Procurement activity the fall in taken has been 57%, in the case of Stores it has been 68% followed by activities such as Sales and Marketing, HR, Administration and Accounts and Finance showing fall in timings to the extent of 58%, 45% and 29% correspondingly. From these findings our analysis shows that in Design, Stores, Sales and Marketing departments the impact of digitization and customized ERP applications on total time saving is much greater than the impact realized in departments such as HR and Admin and Accounts and Finance. In fact, it can be argued that the activities carried out in Accounts and Finance department need a wider level of integration with digitization and customized ERP application processes.

Analysis of Strategy for Growth and Efficiency

Strategic Perspective for Efficiency

After having encountered a few structural issues related to lack of up-gradation of the previous technology, rigid organizational structure, low labour productivity, higher degree of price competition etc., Twin Engineers has made a strategic shift in its overall operational functioning. Its new strategic plan comprises three innovations under 'technological enhancement'. They include such things as "machine up-gradation, complete solution/plant, updated technology- IoT (Internet of Things) and global presence". We shall discuss these innovations in details given as follows:

- *Machine Up-gradation:* During the pre-digitization period, Twin Engineers employed in its manufacturing process the 'basic customized machine' such as basic machine for automation in process. During the post-digitization period, the earlier basic machine has been up-graded. This has helped reduce the existing cycle time of machine to half cycle time for completing the whole process (The Focus Group Survey). Owing to this innovative change, the volume of production has increased and wastages have been reduced. The wider production base has helped customer base grow along with an increase in market size. For example, the Robotic Machine has increased customer production by 35 % and has reduced wastages by 30% (Focus Group Survey and Personal Interview with MD).
- *Complete Solution/Plant:* Earlier, Twin Engineers was engaged in providing its clients with product-specific solution. For example, 'standalone machine'. During the post-digitization period, the Company has started providing its clients with 'complete integrated customized solution'. For example, a complete integrated system is part of complete solution that consists of the MMD machine with Oven, Conveyor System and Robotic Dispensing Machine.

- *Updated Technology – IoT (Internet of Things):* Twin Engineers was earlier working with 'basic machine with MES System'. It is a centralized system with MES/ANDON/ERP. In the operational mechanism of this system the machine used to get connected with the centralized master. This is subsequently followed by the process of the master deciding upon the parameters. In the continuing process the machine sends back to the master a few cycle parameters. At the same time, the machinestores the parameters locally. In this entire mechanism the machine operation is decided centrally and is not fully in control of the operator.

  In the innovative centralized system the machine is IoT compatible. Through this compatibility, the IoT compatible machine can communicate with adjacent machines. The mechanism assures that the machine performs self diagnostically and generates preventive maintenance alerts. The IoT compatible machine can be operated from anywhere with internet connectivity. Through the internet facility the machine can generate e-mails and SMS.
- *Global Presence:* As part of Company's overall growth and expansion strategy, along with technological advancement, the Company has also been making headway towards expanding its global base through enhancing exports. Earlier, Twin Engineers was just confined to domestic market through their participation in domestic product exhibitions. In last couple of years, the Company is expanding its activity base and has been participating in overseas product exhibitions. For example, earlier the Company was exporting only to 1 or 2 countries. At present, its export base has been expanding and the Company is exporting to 10 to 12 countries (Twin's Documents). This suggests two things. One, as its strategic growth plan, the Company is also using actively the option of 'exports' in recent times. Second, as compared to the previous times, the Company is becoming more price competitive.
- *New Product Development (NPD)after digitization:* New product development has been a continuous activity at Twin Engineers. During digitization process, this activity has been scaled up even further. For example, after the initial level of digitization, a few important products have been developed. For example, the products such as MMD, Fluid Filling and Robotic have been developed through extensive R&D. These products have a wide domain of applications in many sectors such as renewable energy, power sector, auto electrical, automotive, wind energy, mining, elevator and solar etc. These, being highly digitized products, have helped speed up the manufacturing process on one hand and reduce labour time used in various activities in number of departments.

Conclusion/s

This research on understanding the impact of digitization and the application of customized ERP processes on the overall operational efficiency at Twin Engineers distinctly recognizes the following conclusions. These conclusions have obviously emanated from the findings and the analysis of the findings.

- We can firmly conclude that the operational efficiency at Twin Engineers, after and during the process of digitization, has gone up considerably.
- The pre-digitization era of 'local excellence' has definitely shown a considerable progress in terms of overall company growth. Albeit this, the Company has had a couple of structural issues related to labour and resources efficiency along with high costs and competitive price. During the digitization era, these issues are being tackled with the introduction of new technology which is based on digitization and customized ERP application processes.
- Since 2012 and more importantly 2015 onwards a variety of new products and processes have been developed whose basis is 'technological advancement' with the ERP and IoT. These newly developed products have improved overall operational efficiency.
- Although digitization and customization through ERP products and processes are very much in place, their activity-wise and department-wise impact is not uniform. For example, the department such as stores, administration and accounts need a greater degree of integration with digitization.
- The improved operational efficiency is a matter of a combined result of various factors such as time, labour and cost which are being satisfactorily integrated with digitization process. In spite of this, the

present pace, degree and level of these new applications have been in nascent stage and they require a considerable amount of time to strengthen their operations andstreamline the overall integration process.

- The original model (1.E) which uses turnover as a proxy for efficiency shows a significant correlation with such predictors as total cost, direct and indirect costs, labour efficiency index and number of machines produced.
- Although this research paper has analysed the impact of digitization on overall operational efficiency, our regression outputs on turnover as predicted variable and direct and indirect costs as predictors suggest multicollinearity (Durbin Watson = 1.73) between cost variables. So, we have room to argue that even predictors need a separate analysis to understand correlation and regression between them.
- As against the suggestion number (7), one more regression between turnover as predicted variable and labour efficiency and machines manufactured as predictors have not shown multicollinearity (Durbin Watson= 2.17) between the predictors. This observation has made us conclude that these two predictors have separate association with turnover used as a proxy for operational efficiency.
- Much greater drive towards digitization of processes will certainly help the Company enhance its present RoI (return on investment).
- Twin's long-term growth strategy may focus on two important aspects. One is widening the base of 'technological advancement' along with greater level of digitization and deeper use of IoTs and the second aspect is exploring better opportunities for strengthening its present export base. Twin has to converge its 'local excellence' policy with better use of AIs, Robotics, and Machine Learning and wide spread of digitization.

Recommendations

- Although, at present, the digitization process and the application of customized ERP have been resulting in positive outcomes, the process is in its nascent stage. Its scope and speed need to be elaborated.
- Our analysis of efficiency based on capacity utilization makes it clear that the efficiency in the categories of product such as ROBOTIC and SCDM should improve in recent future.
- Our analysis related to time efficiency also makes it clear that the activities carried out in Accounts and Finance department need a wider level of integration with digitization and customized ERP application processes.
- Although total cost to revenue ratio has fallen after digitization period, we can argue that the Company should control indirect costs in future.
- On one hand, the Company is focusing on 'technological advance' through digitization and customized ERP applications; it should also lay a greater emphasis on 'human resource skills development'.
- To enhance the participation rate of employees in various activities and in different departments, the employees should be given training which would help them handle digitized and IoT processes smoothly and with greater accuracy.
- Twin's present organizational excellence is weak. To enhance it, employees need flexible deadline schedule, greater technical compatibility, proper training on quality performance of quality products and improvement in the present information network for employees.
- Our earlier survey conducted in 2012 (Employee Survey, 2012) and our observations through our recent survey (Focus Group Survey, 2018) make it clear that a few things on the level of organizational excellence have not changed much. For example, a small section of employees (36% of the surveyed in 2012) feel that the rigid attitude to work culture should change. This will change through better understanding of functional responsibilities (10% of the employees surveyed in 2012) and cooperative learning (12% of the employees surveyed in 2012). It will also improve through enhancing organizational coordination (8% of the employees surveyed) with digitization process.
- Although, at present, Twin Engineers has shown a definite progress towards 'exports' when compared with pre-digitization period, the number of countries to which the Company is exporting may be

increased through 'achieving greater price competitiveness' and 'quality consciousness'.

- Twin has just spent nearly 5 years' of 'digitization and customized ERP process application. In its long-run growth strategy it should spend more on R&D and achieve a much better diversification in the present range of products. This long-term strategy will help Twin Engineers achieve consistent and sustainable growth.

Implications

This research has the following implications so far as Twin's day-to-day operational activities and policy for growth is concerned.

- This research may turn handy for Twin to clearly segregate and better understand the operational efficiency scenario prior to digitization and post digitization.
- Twin can strategize its future plan to enhance integration process of digitization in the case of those activities and departments where the present degree of digitization is less.
- Twin can take definite steps to control its indirect costs and monitor better its competitive price policy and market expansion strategy.
- This research may help Twin to take concrete steps to improve upon its 'organizational excellence' through improving human resource skills and imparting training.

Limitations of Research

The following are some of the important limitations of this research.

- The questionnaire used to conduct survey with 'focus group' is relatively a small sample and the responses to the questions are found to be very uniform.
- The time series data used on such variables as 'turnover, total cost, direct and indirect costs, persons employed and machines dispatched are based on just 10 years' period. Such small time series does not produce robust statistical results.
- Presently, this research has used 'turnover' as proxy for efficiency. If 'profitability' were usedas proxy for efficiency, the results could have been marginally different from the present results.

References

Agrawal, A. (2012). *Customizing materials management processes in SAP ERP*. Retrieved from https://www.sap-press.com

Balance Sheets (2009–2012 and 2012–2018). *Twin Engineers*.

Bansal, V. (2013). *Enterprise resource planning*. Pearson India. ISBN 9789332524583.

Drucker, F. P. (1985). *Innovation and enterprise*. New York: Harper and Row.

Drucker, F. P. (2006). *The practice of management*. Retrieved from www.amezon.com

Employees Survey. (2012). *Questionnaire prepared for employees of Twin Engineers*.

Foley, J. K. (1999). The future of quality management in the contemporary business enterprise, Monash MT. *Eliza Business Review*, *2*(1), 58–71.

Garg, V. K., & Venkitakrishnan, N. K. (2003). *Enterprise resource planning: Concepts and practice books*. Retrieved from google.com

Gerald, I. W., Sondhi, A. C., & Fried, D. (2006). *Financial statements.* Hoboken, NJ: John Wiley and Sons.

Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). *The motivation to work*. New York, NY: John Wiley.

MacArthur, J. (2006). *The book on leadership*. Retrieved from www.amezon.com

Madanhire, I., & Mbohwa, C. (2016). *Enterprise resource planning in improving operational efficiency*. ELSEVIER. Retrieved from www.sciencedirect.com

Magal, S. R. (2011). *Integrated with ERP systems.* Wiley, ISBN978-0-470-47844-8

Memuri, V., & Shailendrathe, P. (2006). Improvement in operational efficiency due to ERP systems implementation: Truth or myth? *IJMR*, *19*(2), 18–36.

Madson, K. B. (1968). *Theories of motivation, institute of general psychology.* The Royal Danish School of Educational Studies.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396.

McGregor, D. (1960). *The human side of enterprise.* New York, NY: McGraw-Hill.

Oakland, S. J. (1999). *Total organizational excellence*. Retrieved from www.amezon.com

Personal Interview. (2012, 2018). *With M.D.Mr.Pradeep Bhalwankar*. Twin Engineers.

'R' software (Version 3.4.1) Causality Analysis

Schultz, H. (2010). Leadership inspiration.....Interview with schultz, Howard. *Harvard Business Review.*

SPSS (17 Version). Regression Outputs.

Survey (2018, Focus Group at Twin Engineers).

Twin Engineers (2018). Various documents such as turn-over data, capacity utilization etc.

Wallace, S. T., & Kremzar, M. H. (2001). *ERP making it happen*. Hoboken, NJ: John Wiley and Sons, Inc.

Westcott, T. R. (Ed.). (2005). *The certified manager of quality/organizational Excellence*. Retrieved from www.amezon.com

www.efqm.org (European Foundation Quality Management Model).

Annexure

Tables

Table 1: a) Total Cost to Turnover Ratio (Cost Efficiency)

Pre-digitization Period

Year	Ratio
2009-10	0.9528
2010-11	0.9257
2011-12	0.9502
2012-13	0.9099

Source: Author's Calculation

Digitization Period

Year	Ratio
2013-14	0.9212
2014-15	0.9258
2015-16	0.8624
2016-17	0.8766
2017-18	0.8773

Source: Author's Calculation

Table 2: a) Proportion of Direct Cost to Turnover Ratio

Year	DC/Turnover Ratio
2009-10	0.81
2010-11	0.73
2011-12	0.71
2012-13	0.65
2013-14	0.72
2014-15	0.71
2015-16	0.68
2016-17	0.62
2017-18	0.66

Source: - Author's Calculation

b) **Proportion of Indirect Cost to Direct Cost**

Year	IC/DC Ratio
2009-10	0.16
2010-11	0.26
2011-12	0.33
2012-13	0.39
2013-14	0.27
2014-15	0.29
2015-16	0.25
2016-17	0.39
2017-18	0.31

Source: - Author's Calculation

Table 3: a) Rate of Growth in Total Revenue

Year	ΔTR/TR
2010-11	0.14
2011-12	0.0053
2012-13	0.22
2013-14	0.25
2014-15	0.0655
2015-16	0.0085
2015-17	-0.0236
2017-18	-0.0281

Source: Author's Calculation

b) **Rate of Growth in Total Cost**

Year	ΔTC/TC
2010-11	0.11
2011-12	0.0320
2012-13	0.17
2013-14	0.27
2914-15	0.0707
2915-16	-0.0605
2016-17	-0.0075
2017-18	-0.0274

Source: Author's Calculation

Table 4: Efficiency Index (Capacity Utilization = Actual Capacity/Installed Capacity *100)

Year	*Efficiency Index*
2015-16	0.88 (88.00%)
2016-17	0.87 (87.68%)
2017-18	0.77 (77.21%)

Source: - Author's Calculation

Table 5: Labour Efficiency Index (LEI)

(No. of machines dispatched / No. of persons employed)

Year	*LEI*
2009-10	1.11
2010-11	1.30
2011-12	1.41
2012-13	1.44
2013-14	1.83
2014-15	1.76
2015-16	1.83
2016-17	1.75
2017-18	1.50

Source: Author's Calculation

Table 6: Efficiency of Time*

Sr No.	*Department/ Activities*	*Pre-digitization time required*	*Post-digitization time required*	*% Change*
1	Design	25	7	72%
2	Procurement	7 (6hours)	4 (6 hours)	57
3	Stores	22	7 (6hours)	68
4	Sales and Marketing	12 (2 hours 40 minutes)	5 (2hours 40 minutes	58
5	HR and Admin.	22	12	45
6	Accounts and Finance	57	40	29

* Time measured in terms of number of days required to complete specific activity

Source: Twin's Activity List and Author's Calculation

(B) Twin Data Table 7

1) Quantitative Details

Year	*Turnover*	*Direct cost*	*Indirect Cost*	*Total Cost*	*No. of Persons*	*No. of Machines Dispatched*
2009-10	124106807	101179817	17074429	118254246	51	57
2010-11	142059255	103862463	27643533	131505996	52	68
2011-12	142825294	101872478	33843745	135716223	55	78
2012-13	175412806	114703659	44921658	159625317	58	84
2013-14	220273447	158636020	44299886	202935906	60	110
2014-15	234714580	167503106	49798953	217302059	65	115
2015-16	236713380	162712664	41429033	204141697	65	119
2016-17	231106340	145383281	57217363	202600644	70	123
2017-18	224596161	150092264	46956162	197048426	85	128

Source: Twin's Data Sheet on Quantitative Details

Table 8: Installed Capacity

Product Category	*2015-16 Installed Capacity*	*2015-16 Actual Capacity*	*2016-17 Installed Capacity*	*2016-17 Actual Capacity*	*2017-18 Installed Capacity*	*2017-18 Actual Capacity*
PACKAGING	60	56	60	36	60	48
MMD	27	16	29	25	36	28
FILLING	9	15	10	8	12	21
ROBOTIC	18	12	19	16	24	4
SPM	9	6	10	8	12	13
SCDM	9	10	10	25	12	7
PROJECT	2	4	2	3	2	1
TOTAL	134	119	138	121	158	122

Source: Twin's Data Sheet on Installed Capacity

Table 9: Activity List

Sr.No.	*Department*	*Name of the Activity*	*No.of Days Pre-digitization Time Required*	*No.of Days Post Digitization Time Required*
1	Design	Design Calculation	0.50	0.25
		DAP document	2.00	1.00
		Drawing 3D	8.00	3.00
		Drawing 2 D	8.00	2.00
		File Release	4.00	0.50
		RPO Release	2.00	0.25
2	Procurement	Study the requirement	1.00	1.00
		Verify the availability	1.00	0.25
		Floating requirement in the market	1.00	0.25
		Costing, comparison negotiation	1.00	0.25
		Release of P.O.	0.25	0.10
		Follow, D/W/M	1.00	1.00
		Updation of the receipt	1.00	0.25
		Payment to vendor	1.00	1.00

Source: Twin Engineers Activity List

Regression Outputs

Model Summary[b]					
Model	*R*	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*	*Durbin-Watson*
1	.996[a]	.992	.985	5717336.132	2.814

a. Predictors: (Constant), Indirect Cost, Persons, Direct cost, Machines

b. Dependent Variable: turnover

ANOVA[b]

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	1.688E16	4	4.220E15	129.096	.000[a]
	Residual	1.308E14	4	3.269E13		
	Total	1.701E16	8			

a. Predictors: (Constant), Indirect Cost, Persons, Direct cost, Machines

b. Dependent Variable: turnover

Coefficients[a]

Model		*Unstandardized Coefficients*		*Standardized Coefficients*			
B		*Std. Error*	*Beta*	*t*	*Sig.*		
1	(Constant)	1103815.017	2.502E7		.044		.967
	Persons	-253437.982	543076.229	-.058	-.467		.665
	Machines	718848.406	521957.588	.412	1.377		.240
	Direct cost	.812	.255	.496	3.182		.033
	Indirect Cost	.689	.410	.183	1.679		.168

Guidelines for Authors

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

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

Title: Title should not exceed more than 12 Words

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

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

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

Primary Heading: Centered, capitalized, and italicized.

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

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

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

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

Style Check

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

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

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

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

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

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

Permissions

Prior to article submission, authors should obtain all permissions to use any content if it is not originally created by them.

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

Review Process

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

Online Submission Process

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

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

International Journal of Business Analytics and Intelligence

SUBSCRIPTION DETAILS

Dispatch Address:-
The Manager,
International Journal of Business Analytics and Intelligence
E-598, Ground Floor
Palam Extension, Sector-7, Dwarka (Near Ramphal Chowk)
New Delhi-110077
Office no. : 011-28082485; 011-47044510, 011-49075396
Mobile No.: +91-09899775880

Subscription Amount for Year 2019

	Print	Print + Online
Indian Region	Rs 2700	Rs 3500
International	USD 150	USD 180
Price mentioned is for Academic Institutions & Individual. Pricing for Corporate available on request. Price is Subject to change without prior notice.		

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

Disclaimer

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

Journal Printed at Anvi Composers, Paschim Vihar.

Copyright

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

www.manuscript.publishingindia.com

PublishingIndia.com

Publishing India Group

E-598, Ground Floor, Palam Extension,
Sector-7, Dwarka (Near Ramphal Chowk)
New Delhi-110077
Tel.: 011-28082485, 011-47044510
Email: info@publishingindia.com
Website: www.publishingindia.com

ISSN 2321-1857
01
9 772321 185001