



# International Journal of Business Analytics & Intelligence

October 2019



A Publication of Publishing India Group

---

# International Journal of Business Analytics & Intelligence

---

## Editor-in-Chief

Tuhin Chattopadhyay  
Chief Technology Officer  
QuantFarm, Pune,  
Maharashtra, India

## Joint Editor-in-Chief

Madhumita Ghosh  
Practice Leader - Big Data & Advanced Analysis  
BA & Strategy - Global Business Services  
IBM, India

## Editorial Board

Prof. Anand Agrawal  
Rector at BlueCrest university College, Ghana

Prof. Anandakuttan B Unnithan  
IIM Kozhikode, India

Prof. Arnab Laha  
IIM Ahmedabad  
Ahmedabad, India

Favio Vázquez  
Principal Data Scientist  
OXXO and Chief Data Scientist  
Iron AI, Mexico

Beverly Wright  
Managing Director, Business Analytics Center  
Scheller College of Business, Georgia Institute of Technology  
USA

Prof. U Dinesh Kumar  
IIM Bangalore  
Karnataka; India

Kevin J. Potcner  
Director of Consulting Services  
Minitab Inc., USA

Prof. Santosh Prusty  
IIM Shillong  
Shillong, India

## Editorial Message



Dear Readers,

It's an immense pleasure to reach out to you all in the era of Automation and Digital journey. We have been publishing by welcoming eminent academicians as well as data science practice professionals along with increased readers. In this Digital transformation era, this issue of IJBAI, focuses on publishing five insightful papers on Digital Roadmap.

We have three thought leadership papers from our regular columnists, besides the five research papers. In the popular column, "Analytically Yours", Professor Arnab Laha of IIM Ahmedabad shared his erudite perspective on the measures of association. Ms. Madhumita Ghosh of IBM shared her practical experience in her write-up on digital experience. The common hindrances in taking the digital leap and the success mantras to win the hurdles have been discussed in her article. Favio Vázquez, internationally acclaimed data scientist from Mexico, has shared his valuable thoughts on the evolution of the science of data and the future of it.

Leading businesses are learning how to use the engagement experiences of individuals and communities as the new basis of their value creation for customers. When an organization aspires to take a next leap and gears for Digital Transformation, 'People' is an integral part of the journey as key stakeholder A Study aims to look at this issue on Factors Influencing Employees on Adoption of Digital Transformation Initiation by Employers will establish various factors viz. employee satisfaction, perceived usefulness and easiness of technology are correlated strongly with employee readiness to accept the digital transformation, whereas employee insecurity and discomfort negatively correlated on digital transformation.

Banks failure is a significant concern to the economy as it creates high cost and heavy losses to the individual banks and society. To avoid the risk of bank failure, banks need to find reliable ways to predict bankruptcy. Certain bankruptcy models are not working in the current economic environment. The current study depicts the ways and means to recalibrate and apply Springate, Zmijewski and Grover model to assess the Indian banks' bankruptcy, showcasing a better model performance. The research paper can throw some roadmap towards unscrupulous promoters with the law establishing the primacy of financial lenders on asset proceeds

In the line of digital transformation, Automation of various process appears as key steps. Robotic process Automation is a technology that uses software robots to automate repetitive tasks and manual processes. A paper on Analysis of Impact of Robotics in Accounting, Reporting and Auditing of Business and Financial Information is in conceptual nature and intended to analyses the Impact of Robotics in Accounting, Reporting and auditing of Business and Financial Information with the help of secondary sources such as journals, websites, reports etc.

A paper on Panel Data Analysis of MINT Economies attempts to find out the relationship between foreign direct investment (FDI) and economic growth in MINT economies namely, Mexico, Indonesia, Nigeria and Turkey during begging of 21st century. It investigates the causal link between the two variables. Cointegration technique, Granger causality test and Vector Error Correction Model (VECM) have been used to validate the relationship between the variables.

We spoke about Bank, Economics, Automation, Digital transformation, however we must acknowledge the fact that programming is a key enabler towards the journey. Our next paper 'Recommending Movies Using User-Based and Item-Based Collaborative Filtering on R Platform' on Recommender systems depicts information filtering systems that deal with

the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest, or observed behavior about a particular item. Recommender system has the ability to predict whether a user would prefer an item or not based on the user's profile. It has a wide use for cross sell and upsell in B2C sector.

I would like to thank the researchers and renowned data science practitioners who have honored us by choosing our young journal to publish some of their research.

I am sure that our readers will enjoy and learn a lot from the present issue. Do let us know your wish, suggestions and views to enrich our journal.

Wish you a very Happy 2020!

Sincerely yours,

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

Dated: 25th November 2019

# International Journal of Business Analytics and Intelligence

Volume 7 Issue 2 October 2019

ISSN: 2321-1857

## Columns

- 1. Measures of Association Among Several Variables**  
*Arnab Kumar Laha* 1-3
- 2. The Tale of Digital Democratic Journey**  
*Madhumita Ghosh* 4-5
- 3. On Data and Science**  
*Favio Vázquez* 6-10

## Research Papers

- 4. A Study on Factors Influencing Employees on Adoption of Digital Transformation Initiation by Employers**  
*Pradeep Kumar C., Renjith K. R., Nimal C. N.* 11-18
- 5. Recalibration and Application of Springate, Zmijewski and Grover Bankruptcy Models in Indian Banking Sector**  
*Reshma Prabhu Verlekar, Manoj S. Kamat* 19-27
- 6. Emerging Trends in Accounting: An Analysis of Impact of Robotics in Accounting, Reporting and Auditing of Business and Financial Information**  
*M. L. Ashoka, Abhishek N., Divyashree M. S.* 28-34
- 7. Causal Link Between FDI and Economic Growth? Panel Data Analysis of MINT Economies**  
*Karan Sabharwal* 35-41

**8. Recommending Movies Using User-Based and Item-Based Collaborative Filtering on R Platform**

*Chirag Malik*

42-46

Journal is also available online at [www.publishingindia.com](http://www.publishingindia.com)

## Analytically Yours:

# Measures of Association Among Several Variables

Arnab Kumar Laha\*

In the current times, humongous amount of data is being collected from a variety of sources and are being put to use for various business applications. Such data is currently being referred to as Big Data. Not only are the size of these datasets huge but often they also have information about a large number of variables which may be dependent on one another. This leads us to the question of understanding the extent of association among these variables. When dealing with only two variables, in elementary statistics courses, one discusses a measure of association called the (Pearson product-moment) correlation coefficient ( $\rho$ ). When two random variables  $X$  and  $Y$  have a joint distribution then  $\rho_{X,Y}$  is computed as follows:

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)}$$

where,  $\text{Cov}(X,Y) = E((X - E(X))(Y - E(Y)))$  is the covariance between  $X$  and  $Y$ ,  $SD(X)$  and  $SD(Y)$  are the standard deviations of the random variables  $X$  and  $Y$  respectively. This simple measure of association takes value between  $-1$  and  $1$  where the extreme values  $\rho_{X,Y} = 1$  or  $\rho_{X,Y} = -1$  indicates perfect linear dependence between the two variables i.e.,  $Y = \alpha + \beta X$  where  $\alpha, \beta$  are two constants. When  $\rho_{X,Y} = 0$  we say that the variables  $X$  and  $Y$  are uncorrelated i.e. they are not linearly related. It may be noted that  $X$  and  $Y$  may be non-linearly related even if  $\rho_{X,Y} = 0$ . To see this, consider the case that  $X$  is symmetrically distributed about  $0$  as a result of which  $E(X) = E(X^3) = 0$ . Let  $Y = X^2$ . A simple computation shows that  $\text{Cov}(X,Y) = E(X^3) - E(X)E(X^2) = 0$  implying  $\rho_{X,Y} = 0$ . When the random variables  $X$  and  $Y$  are independent we have  $\rho_{X,Y} = 0$  but the converse is not true as can be seen from the example given above. The sample correlation coefficient is denoted by  $\rho_{X,Y}$ .

It is therefore natural to seek measures of association akin to the correlation coefficient when more than two variables are involved. Suppose  $\mathbf{X} = (X_1, \dots, X_p)'$  is a  $p \times 1$  random vector and  $\mathbf{Y} = (Y_1, \dots, Y_q)'$  is a  $q \times 1$  random vector and we are interested in measuring the association between  $X$  and  $Y$ . In this context, Hotelling (1936) introduced the concept of Canonical Correlation. Consider the linear combination  $Z_X = \mathbf{u}'\mathbf{X} = u_1X_1 + \dots + u_pX_p$  of the variables  $X_1, \dots, X_p$  and the linear combination  $Z_Y = \mathbf{v}'\mathbf{Y} = v_1Y_1 + \dots + v_qY_q$  of the variables  $Y_1, \dots, Y_q$ . To obtain the first canonical correlation Hotelling suggests choosing the vectors  $u$  and  $v$  so that  $\rho_{Z_X, Z_Y}$  is maximised while ensuring variance of  $Z_X$  and  $Z_Y$  are both equal to  $1$ . The maximum value  $\rho_{Z_X, Z_Y}$  so obtained is called the first (or maximum) canonical correlation ( $\rho_1$ ) between the random vectors  $X$  and  $Y$  and the linear combinations  $Z_X^{(1)} = u_1'X$  and  $Z_Y^{(1)} = v_1'Y$  for which the maximum value of  $\rho_{Z_X, Z_Y}$  is obtained are called the first canonical variates. It is possible to obtain upto  $t = \min(p, q)$  canonical correlations  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_t \geq 0$  by repeatedly following a procedure similar to above with the additional constraint that the canonical variates to be obtained in the current step are uncorrelated with those that had been obtained in the earlier steps. The details are omitted here and the interested reader may look at Chapter 10 of Eaton (2007). A measure of affine dependence based on the canonical correlations is  $\phi = \sum_{i=1}^t (1 - \rho_i^2)$ . When  $\phi = 0$  and  $p \leq q$  it can be shown that  $X$  and  $Y$  are related through an affine transformation i.e.  $X_{p \times 1} = C_{p \times q} Y_{q \times 1} + d_{p \times 1}$  for some  $p \times q$  matrix  $C$  and  $p \times 1$  vector  $d$ . Since  $0 \leq \phi \leq 1$ , one may consider  $1 - \frac{\phi}{t}$  which satisfies  $0 \leq \phi \leq 1$  with  $\phi = 1$  giving

\* Indian Institute of Management Ahmedabad, Gujarat, India. Email: [arnab@iima.ac.in](mailto:arnab@iima.ac.in)

the case of affine dependence between the two random vectors.

How does one estimate the canonical correlations from given data? Suppose  $(x_i, y_i)$  for  $1 \leq i \leq n$  be the given dataset of  $n$  observations. It may be noted that  $x_i = (x_{i1}, \dots,$

$x_{ip})'$  and  $y_i = (y_{i1}, \dots, y_{iq})'$ . We define the following:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$S_{XX} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})',$$

$$S_{YY} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})' \text{ and}$$

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})' = S'_{XY}$$

Let the  $k$ -th sample canonical correlation be denoted by  $\rho_k$ . Then  $\rho_k^2$  is the  $k$ -th eigenvalue of  $S_{XX}^{-1/2} S_{XY} S_{YY}^{-1} S_{YX} S_{XX}^{-1/2}$  (or equivalently that of  $S_Y Y^{-1/2} S_Y X S_X X^{-1} S_X Y S_Y^{-1/2}$ ). Another coefficient that has fairly wide usage for measuring association between two sets of variables is the RV coefficient which is a consistent estimator of the coefficient  $\rho_V$  which is discussed below. Let us denote by  $\Sigma_{XY}$  the  $p \times q$  covariance matrix between the random vectors  $X$  and  $Y$  i.e. the  $(i, j)$ -th entry in this matrix is the  $\text{Cov}(X_i, Y_j)$ . We analogously define  $q \times p$  matrix  $\Sigma_{YX}$ ,  $p \times p$  matrix  $\Sigma_{XX}$  and the  $q \times q$  matrix  $\Sigma_{YY}$ . Escoufier (1973)

defined the coefficient  $\rho_V(X, Y) = \frac{\text{tr}(\boldsymbol{\Sigma}_{XY} \boldsymbol{\Sigma}_{YX})}{\sqrt{\text{tr}(\boldsymbol{\Sigma}_{XX}^2) \text{tr}(\boldsymbol{\Sigma}_{YY}^2)}}$

where for any square matrix  $A$ ,  $\text{tr}(A)$  denotes the trace of the matrix  $A$  (i.e. the sum of the diagonal elements of the matrix  $A$ ). It is easy to see that when  $p = q = 1$ ,  $\rho_V = \rho_{X,Y}^2$ . Further,  $0 \leq \rho_V(X, Y) \leq 1$  and  $\rho_V(X, Y) = 0$  if and only if  $\Sigma_{YX} = 0$ . Moreover,  $\rho_V(X, aBX + c) = 1$  i.e. when  $X$  and  $Y$  are related through an affine transformation then  $\rho_V(X, Y) = 1$ .

The RV-coefficient is defined as

$$RV(X, Y) = \frac{\text{tr}(S_{XY} S_{YX})}{\sqrt{\text{tr}(S_{XX}^2) \text{tr}(S_{YY}^2)}}.$$

This expression of the RV-coefficient can be expressed in several alternative ways. We note one of these here. Let  $\Delta_X$  be a  $n \times n$  matrix whose  $(i, j)$ -th element  $(d_{ij}^X)$  is the Euclidean distance between

$x_i$  and  $x_j$  i.e.  $d_{ij}^X = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$  and  $\Delta_Y$  be another  $n \times n$  matrix whose  $(i, j)$ -th element  $(d_{ij}^Y)$  is the Euclidean

distance between  $y_i$  and  $y_j$  i.e.  $d_{ij}^Y = \sqrt{\sum_{k=1}^q (y_{ik} - y_{jk})^2}$ .

Further let  $C_{n \times n} = I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n'$  where  $I_n$  is the  $n \times n$  identity matrix and  $\mathbf{1}_n$  is the  $n \times 1$  vector of all 1's. The RV-coefficient can now be expressed as:

$$RV(X, Y) = \frac{\langle C \Delta_X^2 C, C \Delta_Y^2 C \rangle}{\|C \Delta_X^2 C\| \|C \Delta_Y^2 C\|}$$

where for any two  $n \times n$  real matrices  $A, B$  the Hilbert-Schmidt inner product  $\langle A, B \rangle$  is defined as  $\langle A, B \rangle = \text{tr}(A' B)$  and

$$\|A\| = \sqrt{\langle A, A \rangle}$$

The third coefficient that we discuss in this article is the dCor coefficient which is based on a measure of dependence between random vectors introduced by Szekely, Rizzo and Bakirov (2007) called the distance covariance or dCov. A very important property of dCov is that  $dCov = 0$  if and only if there is independence between the random vectors  $X$  and  $Y$ . The dCov coefficient is defined as the distance between the joint and the product of the marginal characteristic functions of the random vectors in a weighted  $L^2$  sense that ensures that zero-independence property. The dCov coefficient is then scaled to obtain an association measure  $R(X, Y)$  whose sample estimate is the dCor coefficient which is defined as:

$$dCor^2(X, Y) = \frac{\langle C \Delta_X C, C \Delta_Y C \rangle}{\|C \Delta_X C\| \|C \Delta_Y C\|}$$

**Table 1: Systolic and Diastolic Blood Pressure Readings Taken in the Left Arm (SysLH and DiasLH) and the Right Arm (SysRH and DiasRH) Measured (Nearly) Simultaneously for a Person on 19 Days**

LH-RH Blood Pressure			
SysLH	DiasLH	SysRH	DiasRH
123	85	125	83
112	76	112	76
107	75	115	75
108	72	111	72
110	79	115	79
119	85	125	79
112	78	124	78
117	77	113	79
117	83	110	81
112	80	118	77
106	75	113	72
105	74	112	77
114	81	114	82
107	77	112	72
103	73	119	76
107	71	107	73
112	77	118	84
119	81	129	84

It can be shown that when  $p = q = 1$ , then  $dCor(X, Y) \leq |r_{X,Y}|$ . Further  $0 \leq dCor(X, Y) \leq 1$  and  $dCor(X, aBX + c) = 1$  i.e. when  $X$  and  $Y$  are related through an affine transformation then  $dCor(X, Y) = 1$

In Table 1 above we give Systolic and Diastolic blood pressure readings of an individual measured (nearly) simultaneously on the left arm and right arm of the same individual on 19 days. It is expected that the readings in the left arm would be associated with that in the right arm. Table 2 below gives the R-code and output for computing the three measures discussed above. It can be seen that the measures give quite different values for the same dataset. The two canonical correlations are QUOTE  $\rho_1 = 0.824$  and QUOTE  $\rho_2 = 0.199$  yielding the measure of affine

dependence QUOTE  $\phi = 0.359$ . The RV-coefficient is 0.999 while the dCor coefficient is 0.777.

**Table 2: R-code and Output for Computing Canonical Correlation, RV-coefficient and dCor.**

```
#Canonical Correlations
cc=cancor(LH,RH)
cc$cor
[1] 0.8239106 0.1994156

#RV-coefficient
library(MatrixCorrelation)
RV(LH,RH)
[1] 0.9987494

#dCor coefficient
library(energy)
dcor(LH,RH)
[1] 0.7774218
```

The RV-coefficient is 0.999 while the dCor coefficient is 0.777.

The interested reader looking for further information about measures of multivariate association may refer to the survey article by Josse and Holmes (2016).

## References

- Eaton, M. L. (2007). *Multivariate statistics: A vector space approach*, 53, IMS Lecture Notes Monograph Series.
- Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751–760.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321–377.
- Josse, J., & Holmes, S. (2016). Measuring multivariate association and beyond. *Statistics Surveys*, 10, 132-167.
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6), 2769–2794.

# The Tale of Digital Democratic Journey

Madhumita Ghosh\*

## Introduction

During each Client interaction, solution approach and project execution and interaction with end users and other stakeholder reveals a prime expectation as 'Flawless Digital Experience' during interaction with esteemed Brands. This view is *unanimously* consistent across all age groups and all service takers. However, the journey from conception to implementation is not as smooth for various industries to deliver a strong, digitally-centric value proposition.

If the rate of transformation on the user world surpasses the rate of change on the inside enterprises, the business suffer is very near. We have been witnessing many great brands today are in shrinking situation because of the unpreparedness and lack in embracing the swift turn.

Being in the service of supporting organizations through their digital transformations journey, the key realization and observation is that they struggle digitally do employ sharp. The executives face immense challenges when it comes to *endorsing* the kinds of changes that are necessary in order to make that digital dive.

## Organizational Struggle to Welcome 'Change'

In one side organizations are excited by constantly getting new challenges to attack and new things to learn, yet not more than 20% of people in the world embrace as the 1st hand their perception is 'Change is pain with uncertainty'.

They are reluctant to accept the challenge to their role, job and earning security. The consequences of resistance to change manifest itself in a myriad of ways.

## Blurred Vision for a Digital Customer Journey

Companies that succeed in creating a digital customer value proposition, developed a clear vision of how they will meet their customers' essential digital needs, set objectives against that vision, and execute. Often, companies that are struggling simply haven't painted a clear picture of what they want or rather the need keeping in view of the entire eco system.

To be steady having a handsome market share, its essential to be informed about market and customer trend for self-brand vis-a-vis competitor brands along with the strength and talents of the organization. However, more importantly they need to keep apprised of relevant emerging technology trend and consumer shift and establishing agile processes.

## Ineffective Way of Customer Data Acquisition and Leveraging

Effective gathering, storing and leveraging of customer data at the core acts as the gear and accelerator both to drive. Many organizations today have a numerous of siloed systems containing various scraps of data about customer interactions, but no clear way to pull them together. Others have petabytes of data centralized in an information warehouse that they may use for reporting,

---

\* Practice Leader, Big Data & Advanced Analysis BA & Strategy, Global Business Services IBM, India.

however, they struggle to figure out what to do with all that data in a manner that provides value to the business.

### **Technology Stack and Development Processes are not Flexible Enough with Lack of Cross Functional Collaboration**

Effective digital properties almost always iterate to success via the “test and learn” approach - where new features are being regularly added, measured, adjusted and pruned, based on user feedback and usage data. Leveraging agile processes and technologies that support frequent integration and product releases. Defining ‘MVP’ Minimum Viable Product and roll out each used case after ‘test and learn’ provides agility to the journey.

### **Flexible and Customized Business Model**

Lastly, real success in digital is rarely about providing the exact replica products and services, through a digital channel. Those who have flexibly can often find that the adoption of a digital strategy offers more scale, revenue and profit than the legacy approach, but it takes experimentation, an assumption of risk along the way. This approach is widely accepted among startups, it is one that the management and investors in mature companies generally fear. The digital transformation process is a continuous journey as it embeds itself as a constant improvement project across enterprise. While the process of transforming into its digital version may seem daunting, but one can reduce that anxiety by taking it step by step and considering all options as the plan evolves.

Let's be Digital!

# On Data and Science

Favio Vázquez\*

If we fail to predict the future, they call us a failure; if we do it too well, they call us a sorcerer. But as Poincaré said, “It is far better to foresee even without certainty than not to foresee at all.”

We have been doing science for a while now. I am going to put beginning in 1637 when Descartes published “Discourse on the Method.” The main result of that book is the distinction between knowledge and truth, and that the discourse of the scientist is related to knowledge (that we later discover to be always incomplete), not to look for the truth.

That is a very important point, because it gives us a focus; science wants to know stuff that are not “undisputed truths.”

Again reframing what Poincaré said about Mathematics, if we wish to foresee the future of data science, our proper course is to study the history and present condition of the science.

Let us start with some history. Instead of boring you with paragraphs of text, I build this timeline about data science that will help understand where we are coming from and where we are going.

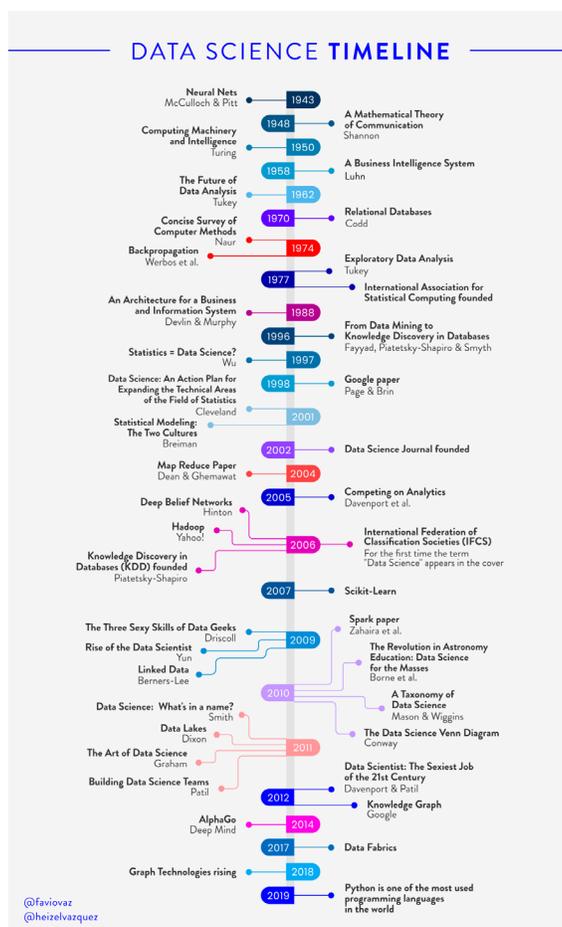


Fig. 1

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

Not a new field at all, but with amazing developments in the past decades. I had to cut some parts and moments because of space, but the most important things are hopefully there.

This timeline will give you an idea of the history of the science of studying data, which we call data science right now, but that may not be the final term we have. So, always be prepared for a change.

## The Role of Data in Science

Data have been close to science in almost all of its history. Sometimes, theory gets us to the data, but for the most of what we are going to discuss here and other upcoming articles, it's going to be in the model data -> theory.

I want to tell you a little story that will make you understand better why data are important, but not the most important part of science.

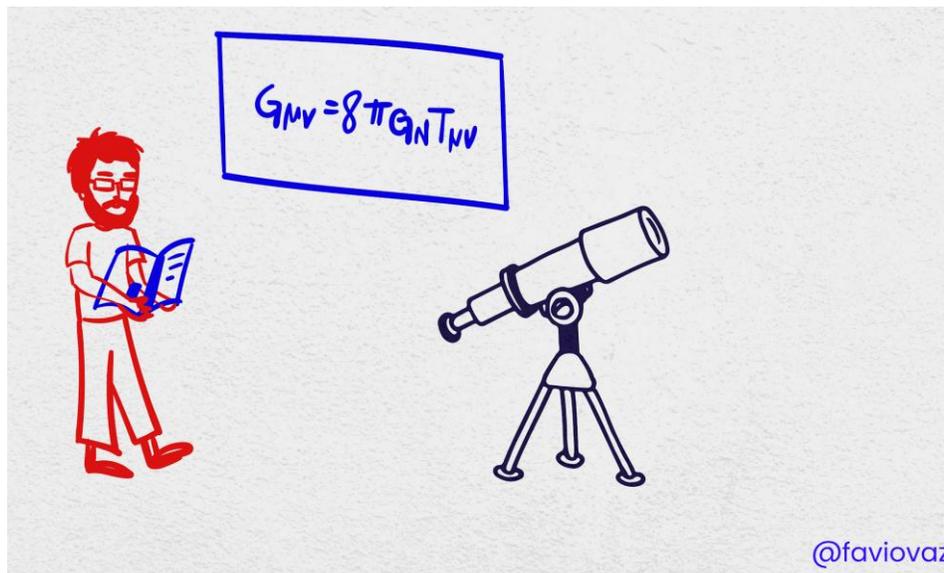


Fig. 2

The story is about a man called Tycho Brahe. He spent almost all his life measuring the positions of the stars, planets, the Moon, and Sun. For what? He wanted to learn how to predict eclipses. Also, he was unhappy with the Ptolemaic system and the Copernican theory was not enough for him either. So, he wanted to find the best way to describe the skies and its moving parts.

Sadly, he was not sure how, but he kept on measuring things until his final days. He died in 1601, and someone named Johannes Kepler, who became his assistant a year before, who was a great mathematician, had access to almost all his data. With that data, Kepler improved the Copernicus theory of the Universe and developed three

laws that described the motion of the planets. Kepler's work served as the basis for the later studies of Isaac Newton about the theory of gravity and the motion of bodies.

The story is much more long and fun, if you want to know more about it, please take a look online. But you could be asking yourself at this point, what does this story has to do with data science?

The biggest learning we have from that story is that having data, and sometimes a lot of data, is worthless unless you have a good question to answer. This is still true nowadays, and the start of the modern love of data began with statistics.

## The Role of Statistics



Fig. 3

I am not going to write a lot about statistics here, but I will point to two specific things that changed the world forever. First is an article called “The Future of Data Analysis” by John Tukey, published in 1962, and the other one is a presentation by Professor Jeff Wu titled “Statistics = Data Science” given in 1997.

These are pretty old references I know, but they are very important. Believe me.

In the article by Tukey, he said this:

“For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. [...] All in all, I have come to feel that my central interest is in data analysis...”

This is a huge statement to make by a statistician. In this time, the words “data science” did not exist as today, but the way Tukey described data analysis is very close to what we call now data science. He even called data analysis a science, because it passes these three tests:

- Intellectual content.
- Organization into an understandable form.
- Reliance upon the test of experience as the ultimate standard of validity.

Saying that this “new science” is defined by a ubiquitous problem rather than a concrete subject. He then goes on and talks about how to learn and get started with data analysis, and how to become a data analyst and also how to teach it. It is an amazing article that we all should read if we want to understand the beginnings of our field.

In the second piece, 35 years later after Tukey’s publication, Jeff Wu said this:

“Statistics = Data Science?”

Where he proposed that statistics should be renamed “data science” and statisticians should be named “data scientists.” In today’s standards, we now that statistics is a part of data science, but why? Because we say that we also need programming, business understanding, machine learning, and more. Maybe, it is just that statistics evolved and now some statisticians became data scientists. But, only some of them.

To understand the portion of statistics and statisticians that became data science and data scientists, we need to read the article “Statistical Modeling: The Two Cultures” by Leo Breiman published in 2001.

Here, he mentions that there is some people in the statistical culture that are driven by data modeling and some by algorithmic modeling. Where the first ones assume that we have a stochastic data model that maps input variables  $x$  to response variables  $y$ . And, the second ones consider that the mapping process is both complex and unknown, and their approach is to find a function  $f(x)$  that operates on  $x$  to predict the responses  $y$ .

He then goes on to discuss why the data modeling culture has been bad for statistics for so long leading to irrelevant theories and questionable scientific conclusions keeping statisticians from using more suitable algorithmic models and working on exciting new problems. Also, he talks about the wonders of the other part of the spectrum, i.e., the algorithmic modeling culture, giving examples from his own work and others on how it can solve hard and complex problems.

## The Role of Data Science

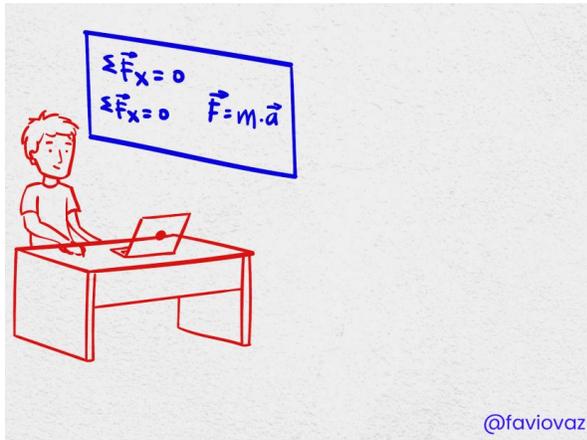


Fig. 4

Data science is the main focus of most sciences and studies right now. It needs a lot of things like AI, programming, statistics, business understanding, effective presentations skills, and much more. That is why, it is not easy to understand or study. But we can do it, we are doing it.

Data science has become the standard problem-solving framework for the academia and the industry and it is going to be like that for a while. But we need to remember where we are coming from, who we are, and where we are going.

## Where Are We Going?

A while ago, I published this chart about the interest on semantic technologies over the years.

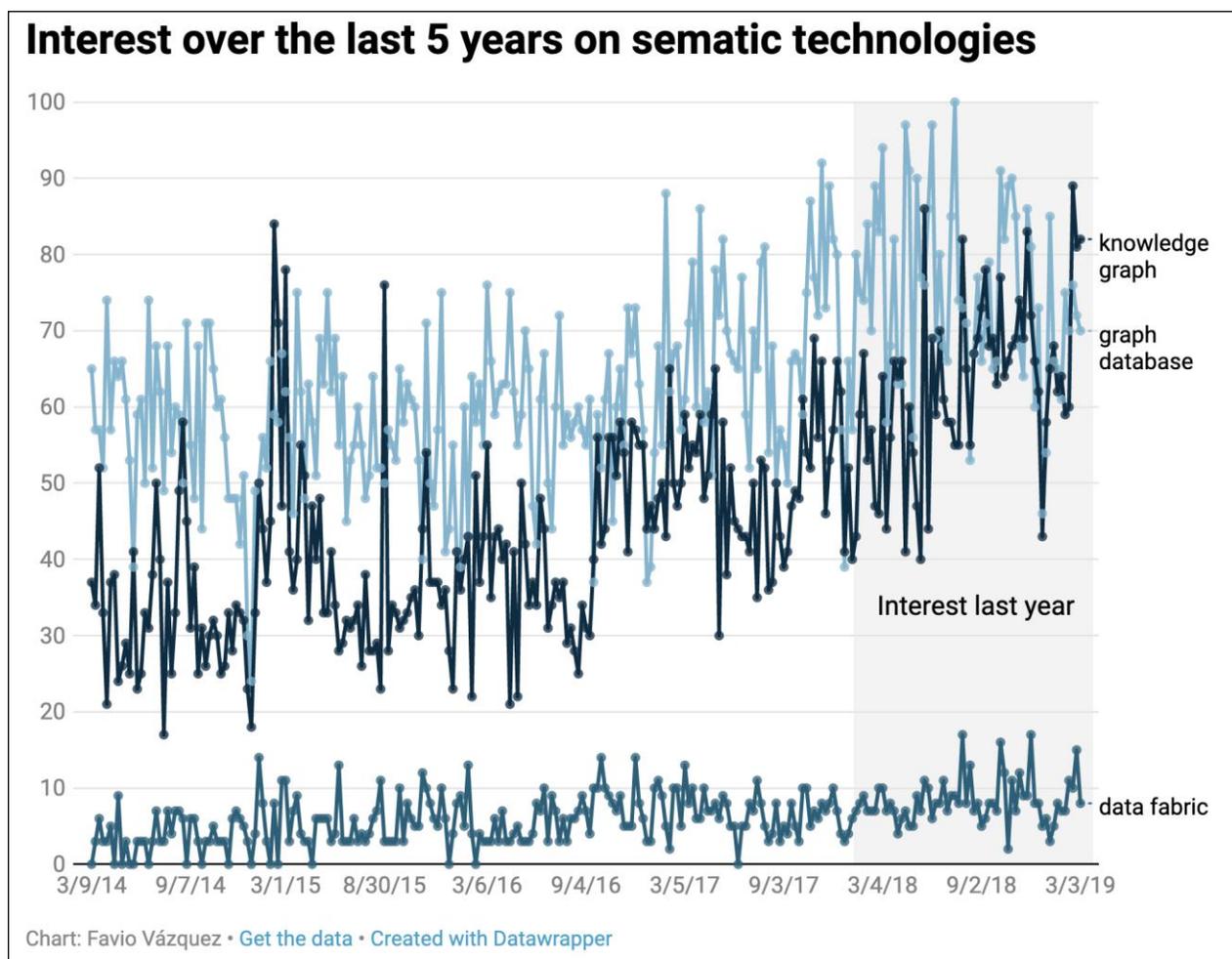


Fig. 5

We can easily see that it is increasing over time. Semantics in this context means the use of formal semantics to give meaning to the disparate and raw data that surrounds us, and also the relationship between signifiers and what they stand for in reality, their denotation.

When we talk about semantics in data, we normally mean a combination of ontology, linked data, graphs and knowledge-graphs, the data fabric, and more. You can read about all of that in the links at the beginning of the article.

But why? Why the shift? The thing is that 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.

We normally store data in graphs in these technologies. 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. And, it is better for what we have right now.

I have been in countless projects right now, and the common thing is we spend a lot of time trying to make sense of the data we have, and one of the reasons may be that we are not storing the data and its relationship in a good format. The promise of the data fabric is just to support all the data in the company, i.e., how it is managed, described, combined, and universally accessed.

Remember, data and context come first, this new paradigm integrates and harmonizes all relevant data sources — structured and unstructured data alike — using a built-in graph database and semantic data layer. The data fabric conveys the business context and meaning of your data, making it easier for business users to understand and properly utilize.

For me, that is the future of data science. We are moving in a direction where semantic technologies are going to be the standard in every company. But, we would not stop there. All the advances in augmented reality, virtual reality, and more will companion these shift.

# A Study on Factors Influencing Employees on Adoption of Digital Transformation Initiation by Employers

Pradeep Kumar C.\*, Renjith K. R.\*\*, Nimal C. N.\*\*\*

## Abstract

India is one of the largest and fastest-growing economies and the organisations in India needs to compete with other developed countries to keep the momentum and achieve the higher growth rate than other countries. Competitive offerings by telecommunications firms have turbocharged internet subscriptions and data consumptions by the individuals and organisations. The competition among the firm is increasing, and it is a fact that the organisation which is capable of exploiting the technology in a better way will outsmart other in the competition. Though the union government and state government prioritise the information technology training, the employees are still not perceived the advantages of digitalisation. The digitalisation effort by few organisations is either not taken off or left at midway due to lack of support. Hence, this study is to evaluate the factor influencing the organisation digitalisation. It is found that all factors like employee satisfaction, perceived usefulness and easiness of technology are strongly correlated with employee readiness to accept the digital transformation. Employee insecurity and discomfort negatively correlated on digital transformation.

**Keywords:** Employee Satisfaction, Employee Readiness, Perceived Usefulness, Perceived Ease of Use, Digital Transformation

## Introduction

Organisations are facing stiff competition from national and international firms and the firm which has the capability of using digital technology in the most efficient

way will take advantage over others. According to Mckinsey Digital India report 2019, the organisations which innovate and digitise more rapidly will be in a better position to take the benefit of India's huge, connected market. It is expected that by 2023, India will have 840 million internet users and 700 million smartphone users. Due to the fast-changing scenario in technology-enabled services and low-cost data availability, business models are becoming pervasive over the next decade. As technology is rapidly progressing, present-day organisations are confronted with massive changes (Colbert, Yee & George, 2016).

Changes brought by digital adoption will disrupt India's labour force as well as its industries. It is estimated that 65 million new jobs could be created from the indirect and direct impact of digital applications which will be helping the productivity enhancement. All the jobs which are done manually will move to technology-enabled process and impact the supporting services also. These jobs could be agriculture, trade and hotel, IT-BPM, media, telecom, manufacturing, logistics and transportation, etc.

To overcome this fast-changing technology, individuals will need to equip themselves constantly by learn and relearn new skills in their lifetime. To take up these challenges, the government needs to facilitate effective and affordable training programmes for fresher's and mid-career courses for existing employees. In addition to that, the organisations are facing problems due to employee's behavioural tendencies for acceptance of new. It is the fact that without the wholehearted support of

\* Professor, Adi Shankara Business School, Kerala, India. Email: [prascropio@yahoo.com](mailto:prascropio@yahoo.com)

\*\* Assistant Professor, Adi Shankara Business School, Kerala, India.

\*\*\* Associate Professor, Adi Shankara Business School, Kerala, India.

each and every employee these transformation will not be fulfilled.

One way of achieving an edge over a competitor and sustaining the competitive advantage is by employing an Enterprise Resource Planning (ERP) system. An ERP system is a well-structured, mandatory, integrated, customised, packaged software-based system that handles most of the system requirements in all business operational functions such as finance, human resources, manufacturing, sales and marketing (Wua, Onga & Hsub, 2008). Though implementation of ERP systems was perceived with high expectations, very few organisations have achieved significant performance enhancement (Soh, Kien & Tay-Yap, 2000) and most ERP projects turn out to be expensive, not cost-effective and a failure (Abugabah & Sanzogni, 2010; Hong & Kim, 2002; Kumar).

There were many reasons for ERP failures in an organisation. Somers & Nelson (2004) found that ERP project failures were due to poor project communication whereas (Al-Mashari, Al-Mudimigh & Zairi, 2003) it is due to lack of senior management support. Presence of ethnic differences (Yusuf, Gunasekaran & Abthorpe, 2004), low user acceptance (Amoako-Gyampah & Salam, 2004; Shih & Huang, 2009), ineffective integration of systems (Al-Mashari et al., 2003), user dissatisfaction (Beyleveld & Schurink, 2005; Yu, Li & Gagnon, 2009) and inadequate system training (Amoako-Gyampah & Salam, 2004) were considered as other reasons.

Understanding the growth potential of IT industry, the Government of Kerala too has implemented few policy directives to catch up with the requirements. The present study evaluates the effectiveness of initiatives at the ground level. The study also assesses the behavioural aspects of existing employees on acceptance of digitalisation by the organisation.

The goal of this paper, therefore, is to review, refresh and structure the literature on factors influencing the employees for acceptance of digital transformation in their organisations. To do so, we conducted an empirical survey at three IT hubs of Kerala namely Trivandrum, Kochi and Kozhikode as major information technology organisations are concentrated on this part of the state.

## **Objective of the Study**

The objective of the study is to evaluate the factors influencing the acceptance of digital transformation in their organisations. The study focus on factors like employee satisfaction, the benefit of perceived use and benefit of perceived easiness due to digitalisation in framing the attitude in accepting the digital transformation initiative. The study also sheds the lights on building a model for the implementation of the organisational digitalisation process.

## **Significance of the Study**

The rationale of this study was to find out and recommend the factors influencing the implementation of digitalisation in the organisations. The study evaluates the employee's perception of technology acceptance while considering the facts on employee satisfaction, advantages of using technology on task easiness and usefulness. The study is of utmost significance to employees as very few organisations only successfully implement the digitalisation process as most of them failed due to the reasons one or other. Kerala with its educated manpower wants to compete with national and international organisations, the organisations need to assimilate the technology to get an advantage over others.

## Proposed Model

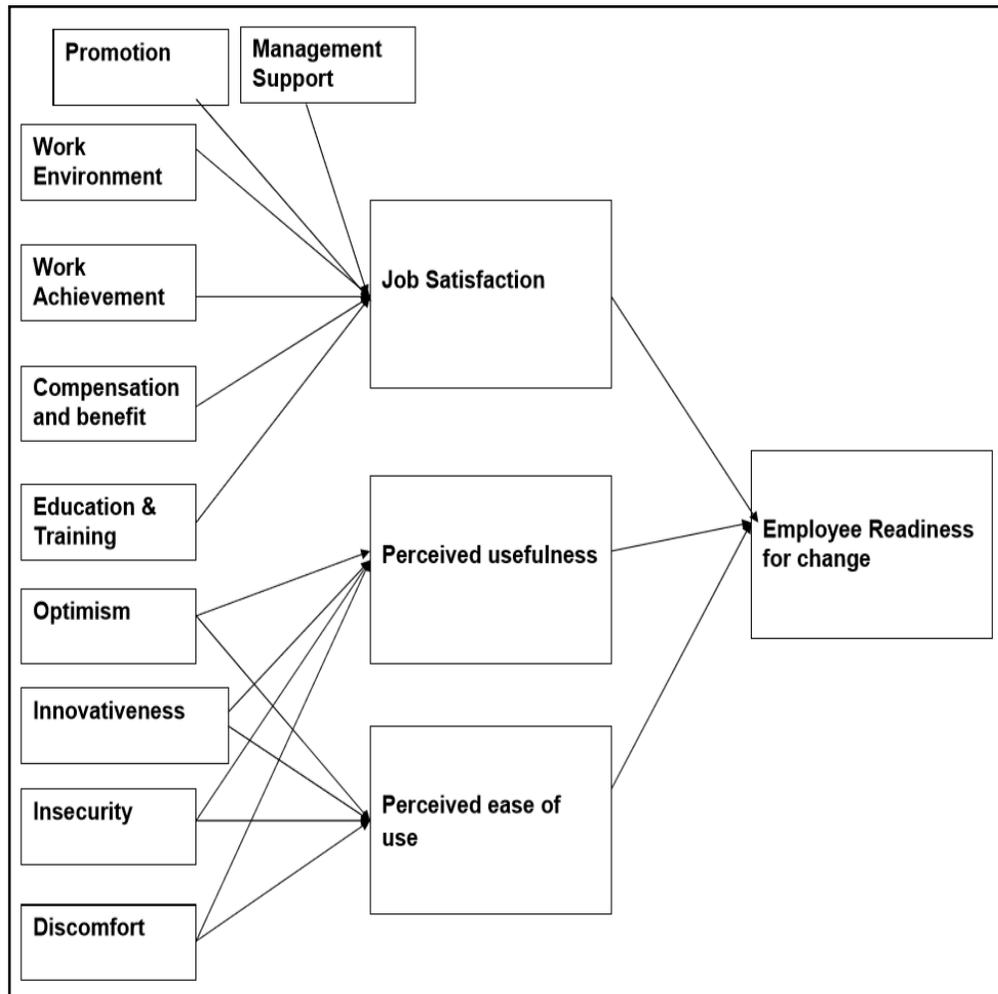


Fig. 1

## Literature Review

The aim of the study is to find out the factor which can influence the employees on adaption of digital transformation by the organisation. Schultz & Schultz (2014) suggests the studies conducted by researchers in the area of digital transformation have given thrust to ergonomic and technology design factors in pre-implementation phases of ERP systems. Meade & Islam (2006) and Venkatesh et al. (2003) found that information technology (IT) acceptance research has explored many competing models for explaining the relation between user attitude, perception and belief towards system usage. People's propensity to embrace new technology is the combination of positive and negative technology-related belief. These beliefs are varied among individuals, and it will effect an individual's pre-deposition to the

technology (Parasuraman & Colby, 2001). Furthermore, findings show that these beliefs can be categorised into four dimensions: optimism, innovativeness, discomfort and insecurity (Parasuraman, 2000).

## Employee Satisfaction

Job satisfaction is defined as the extent to which an employee feels self-motivated, content & satisfied with his/her job. Job satisfaction happens when an employee feels he or she is having job stability, career growth and a comfortable work-life balance. According to Buitendach & de Witte (2005), job satisfaction of an individual is his/her job perception and in turn, it is influenced by the values and expectations. Individuals, therefore, evaluate their jobs on the basis of factors which they regard as being important to them (Sempene et al., 2002). Visser

et al. (1997) define job satisfaction as "...the attitude of workers toward the company, their jobs, their fellow workers and other psychological objects in the work environment."

Camp (1994) defines job satisfaction as the extent to which these needs and values are satisfied in the workplace. In conjunction with this, Robbins (1998) surmises that job satisfaction is based on "the difference between the amount of rewards workers receive and the amount they believe they should receive." Because job satisfaction may be an indicator of whether individuals (a) will be effectively connected to an institution, (b) will merely comply with directives, or (c) will quit (Ma & Macmillan, 1999), principals ought to have some understanding of the factors that influence teachers' satisfaction with their work lives and the impact this satisfaction has on teachers' involvement in their schools, primarily when changes are implemented.

According to Nicolaidis & Katsaros (2011) and Brooks (2000), managerial and organisational performance has a significant effect on job satisfaction. Also, Locke (1976) and Spector (1997) suggest that a positive emotional state resulting from the pleasure an employee derives from the job effects affective, cognitive and emotional attitude held by an employee about various aspects of their work. Vroom defines job satisfaction as "the positive orientation of an individual towards all aspects of the work situation."

### **Perceived Ease of Use**

Nunkoo et al. (2013) defines the perceived ease of use (PEOU) as the degree in which a person believes that using a particular system would be free from effort. Chau (1996) found that perceived ease of use influences usefulness, attitude, intention and actual use. Davis et al. (1989) found that perceived ease of use is an important factor for employee intentions to use computers. Feelings of fear and insecurity related to technology are associated with ambiguity and low usage by employees (Parasuraman & Colby, 2001; Tsikriktsis, 2004). Davis (1989) suggests that perceived ease of use also can improve employee performance due to the fact that he or she can deploy less effort to accomplish the task.

### **Perceived Usefulness**

Perceived value is defined as the degree to which an individual believes that using a particular system would

enhance his or her performance. Davis et al. (1989) found that perceived usefulness and usage have a stronger and more consistent relationship. Also, Chau (1996) found that individuals evaluated the consequences of their behaviour in terms of perceived usefulness and based their choice of behaviour on the desirability of the usefulness.

## **Methodology of the Study**

The objective of the study has been to find out the factors influencing on employees on adaption of digital transformation initiative by the organisation. On the basis of this, the methodology of the study has been designed.

### **Research Design**

The study is descriptive in nature because it has been designed to measure the employee's perception of digitalisation initiatives by the organisation. The data collection has been done by the use of the questionnaire method, which consisted of constructs related to employee satisfaction, perceived ease of use, perceived usefulness and employee readiness to accept the digitalisation. The constructs for employee satisfaction, perceived ease of use, perceived usefulness and employee readiness to accept the digitalisation have been taken from the study of Godeo & Johansen (2012).

### **Sampling Design and Data Collection**

Data required for the study have been collected from both the primary and secondary sources. A structured questionnaire has been prepared and used as an aid to collect the data from the respondents. The research has been carried out in three information technology dominated industries in the state of Kerala, i.e., Ernakulam, Trivandrum and Kozhikode. These districts have the highest IT-enabled industries with various industries. The target population for this study has been primarily the employees of organisations at all level. The study was conducted between Nov. 2017 and Jan. 2018. A total of 600 questionnaires were distributed, and only 354 were returned. Out of these, 323 survey forms from all three places were subjected to statistical analysis. The information collected from secondary sources such as websites, articles, journals, books, magazines, etc., was also used for the study.

## Analysis and Result

Data collected have been analysed using SPSS version 21 in order to find the reliability and the relationship between the dependent and independent variables. The test of reliability using Cronbach’s alpha has been 0.89, which is satisfactory. Correlation and regression analysis have been carried out in order to find out the influencing factors of employee acceptance of digital transformation.

## Demographic Profile of Respondents

The characteristic of the study population is detailed in Table 1. 57.2 percent respondents were below 40 years. 46.7 percent respondents were graduates and the gender of the sample were equally distributed. Sample size from Trivandrum, Kochi and Kozhikode were equal representative of the sample.

**Table 1**

<i>Age of the Respondents</i>	<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cumulative Percent</i>
less than 30 yrs	118	36.5	36.5	36.5
31-40 yrs	67	20.7	20.7	57.3
41-50 yrs	47	14.6	14.6	71.8
51-60	41	12.7	12.7	84.5
Above 60	50	15.5	15.5	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Qualification</b>				
Degree	151	46.7	46.7	46.7
PG	105	32.5	32.5	79.3
Technical	67	20.7	20.7	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Gender</b>				
Male	170	52.6	52.6	52.6
Female	153	47.4	47.4	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Place of Survey</b>				
Kochi	101	31.3	31.3	31.3
Trivandrum	108	33.4	33.4	64.7
Calicut	114	35.3	35.3	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Professional Level</b>				
Management Staff	178	55.1	55.1	55.1
Office Staff	69	21.4	21.4	76.5
Field Staff	76	23.5	23.5	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Years of Service</b>				
Less than 5 yrs	30	9.3	9.3	9.3
5-10	89	27.6	27.6	36.8
10-15	204	63.2	63.2	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	
<b>Educational</b>				
Engineering	3	.9	.9	9.3
Information Technology	103	31.9	31.9	41.2
Admin/HR	55	17.0	17.0	58.2
Financial	100	31.0	31.0	89.2
Marketing	35	10.8	10.8	100.0
<b>Total</b>	<b>323</b>	<b>100.0</b>	<b>100.0</b>	

### Test of Reliability

A reliability coefficient of (Cronbach’s Alpha) 70% or higher is considered acceptable in most social science research. The result of this test in the current study is 0.890 hence accepted.

**Table 2: Reliability Coefficient**

Reliability Statistics	
Cronbach’s Alpha	No. of Items
.890	59

### Correlation

Table 2 shows that the correlation coefficients between independent variables are greater than 0.36. All the variables except discomfort and insecurity of employee perceptions have shown high positive correlation with employee readiness. All the values are significant at the 0.01 level.

### Regression Analysis

Regression analysis has been used to identify the relationship between independent variables and dependent variable. A model of the relationship has been hypothesised and estimated parameter values are used to develop an estimated regression equation. The estimated regression equation confirmed to predict the value of

the dependent variable from the given values of the independent variables.

### Relationship Model

Table 3 provides with the R and R<sup>2</sup> values. The R-value is 0.827, which represents that there is a fairly good correlation between dependent variable employee attitude and independent variables (perceived usefulness, perceived ease of use and employee satisfaction). Similarly, nearly 69% of the variance is explained (R<sup>2</sup> = 0.684) in the regression model.

**Table 3**

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.827 <sup>a</sup>	.684	.681	3.44361

a. Predictors: (Constant), Employee satisfaction, Perceived use fullness, Perceived ease of use

### Parametric Estimates

Table 4 allows us to specify multiple models in a single regression model. The constant represents that the strength of affiliation (5,668) is positive, thereby indicating a positive relationship. The coefficient of perceived usefulness is 0.381. Perceived ease of use is 0.954. Employee satisfaction is 0.438.

**Table 4**

Coefficients						
Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.668	.611		9.279	.000
	Perceived use fullness	-.381	.131	-.269	-2.911	.004
	Perceived ease of use	.954	.154	.668	6.180	.000
	Employee satisfaction	.438	.049	.448	8.876	.000

a. Dependent Variable: Employee Readiness

### Discussion

The objective of the study was to find the factors influencing the employee perception on adaption of digital transformation initiatives in organisations. For

this purpose, a study model was developed, which included several factors. These factors were divided into four dimensions namely Perceived usefulness (PU), Perceived ease of use (PE), Employee satisfaction (ES) and Employee readiness (ER) .

The result of the study showed that perceived usefulness (PU), Perceived ease of use (PE) and Employee satisfaction (ES) have significant positive relationship with Employee readiness (ER) and their values of correlation coefficient are found to be (0.649\*\*), (0.759\*\*) and (0.783\*\*), respectively. Also, the relationships between Employee satisfaction (ES), perceived usefulness (PU) and Perceived ease of use (PE) are found to be highly positive with (0.654\*\*) and (0.763\*\*).

## Recommendations

From the results, it is found that employee discomfort and insecurity are the factors that employee are hesitant to adopt the digital transformation initiated by the employers. By way of training, awareness camps, etc., these feeling can be removed. The importance of digitalisation in this competitive world should be conveyed to employees for a more straightforward implementation of digital transformation.

## References

- Abugabah, A., & Sanzogni, L. (2010). Enterprise resource planning (ERP) system in higher education: A literature review and implications. *International Journal of Human and Social Sciences*, 5, 395-399.
- Buitendach, J., & de Witte, H. (2005). Job insecurity, extrinsic and intrinsic job satisfaction and affective organisational commitment of maintenance workers in a parasternal. *South African Journal of Business Management*, 36(2), 27-37.
- Camp, S. D. (1994). Assessing the effects of organisational commitment and job satisfaction on turnover: An event history approach. *Prison Journal*, 74(3), 279-306.
- Chau, P. (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems*, 13, 185-220.
- Davis, F., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35, 982-1003. doi:10.1287/mnsc.35.8.98
- Godoe, P., & Johansen, T. S. (2012). Understanding the adoption of new technology: Technology readiness and technology acceptance as an integrated concept. *Journal of European Psychology Students*, 3.
- Hong, K. K., & Kim, Y. G. (2002). The critical success factors for ERP implementation: An organizational fit perspective. *Information and Management*, 40, 25-40. Retrieved from [http://dx.doi.org/10.1016/S0378-7206\(01\)00134-3](http://dx.doi.org/10.1016/S0378-7206(01)00134-3)
- Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of Industrial and Organisational Psychology* (pp. 1297-1350). Chicago: Rand McNally.
- Ma, X., & Macmillan, R. B. (1999). Influences of workplace conditions on teachers job satisfaction. *Journal of Educational Research*, 93(1), 39-48.
- Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation: A 25 year review. *International Journal of Forecasting*, 22, 519-554. Retrieved from <http://dx.doi.org/10.1016/j.ijforecast.2006.01.2005>
- Nicolaidis, C., & Katsaros, K. (2011). Tolerance of ambiguity and emotional attitudes in a changing business environment: A case of Greek IT CEOs. *Journal of Strategy and Management*, 4(1), 44-46.
- Nunkoo, R., Juwaheer, T. D., & Rambhunjun, T. (2013). *Applying the extended technology acceptance model to understand online purchase*. The behaviour of traveller, Proceeding of 21<sup>st</sup> International Business Conference 10-11 June (2013) Ryerson University, Toronto, Canada, ISBN 978-1-922069-25-2
- Parasuraman, A. (2000). Technology readiness index (TRI): A multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*, 2, 307-320. doi:10.1177/10946705002400
- Parasuraman, A., & Colby, C. L. (2001). *Techno-ready marketing: How and why your customers adopt the technology*. New York: Free Press.
- Robbins, S. P. (1998). *Organisational behavior* (8<sup>th</sup> ed., p. 25). New Jersey: Prentice-Hall.
- Spector, P. E. (1997). *Job satisfaction: Application, assessment, cause, and consequences*. Thousand Oaks, CA: Sage Publications, Inc.
- Schultz, D., & Schultz, S. E. (2014). *Psychology and work today*. Essex, United Kingdom: Pearson Education.
- Soh, C., Kien, S., & Tay-Yap, J. (2000). Cultural fits and misfits: Is ERP a universal solution? *Communication of the ACM*, 43, 47-51. Retrieved from <http://dx.doi.org/10.1145/332051.332070>
- Tsikriktsis, N. (2004). A technology readiness-based taxonomy of costumers: A replication and extension. *Journal of Service Research*, 7(1), 42-52.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Towards a unified view. *MIS Quarterly*, 27, 425-478.

Victor Vroom, H. (1964). *Work and motivation* (pp. 28-52). New York, John Wiley and Sons.

Visser, P., Breed, M., & Van Breda, R. (1997). Employee satisfaction: A triangulation approach. *South*

*African Journal of Industrial Psychology*, 23(2), 19-24.

Wua, L.-C., Onga, C.-S., & Hsub, Y.-W. (2008). Active ERP implementation management: A real options perspective. *Journal of Systems and Software*, 81, 1039–1050. Retrieved from <http://dx.doi.org/10.1016/j.jss.2007.10.00>

# Recalibration and Application of Springate, Zmijewski and Grover Bankruptcy Models in Indian Banking Sector

Reshma Prabhu Verlekar\*, Manoj S. Kamat\*\*

## Abstract

**Objectives:** Banks' failure is a significant concern to the economy as it creates high cost and heavy losses to the individual banks and society. To avoid the risk of bank failure, banks need to find reliable ways to predict bankruptcy. Certain bankruptcy models are not working in the current economic environment. Therefore, the objective of the present paper is to recalibrate and apply Springate, Zmijewski and Grover models to assess the Indian banks' bankruptcy. Method the current study tries to recalibrate the said models by considering the fact of models criticism in past studies with respect to their predictive power, selection of variables, time factor, accuracy rate, change in the economic environment, etc. Models are recalibrated by changing coefficients of the original models using current data through multiple regression technique. Further, this study applies the recalibrated bankruptcy models such as Springate, Zmijewski and Grover to assess the banks' bankruptcy. Results the result shows that the recalibrated Grover model outperforms the original model; however in case of Springate and Zmijewski model, original model performs better than the recalibrated model. Conclusion the implications of the study direct the Reserve Bank of India to make a policy of using advanced modes such as Multiple Discriminant Analysis technique, logit, Probit models along with the CAMEL model for the financial health assessment of banks.

**Keywords:** Recalibration, Bankruptcy Models, Springate Model, Zmijewski Model, Grover Model, Banking Sector

## Introduction

In the Indian financial system, banking plays a vital role in the up-liftment of the society. Banks' distress is a major concern to the economy all over the world, as it has direct and indirect impacts on the individual bank and the global economy. It affects individual bank in terms of direct and indirect cost. Direct cost will be in terms of legal and administration cost associated with bankruptcy proceedings and indirect cost will be in terms of loss of depositor's confidence; withdraw of amount from the bank, avoidance of investment by customer. Similarly, it affects global economy in different context. The banking system affects the economic performance of different country. Failure of giant banks may traumatize not only the domestic economy but also put the global economy at stake (Mayanka & Sharma 2013). The failure of one bank has a spillover effect on the other bank and throughout the world economy. According to (Lawrence, Pongsat, & Lawrence, 2015), the failure of business organization has significant economic effect for its owners, creditors and the overall society. In other words, the effect of bank becoming insolvent often leads to adverse consequences for many stakeholders, thus the prediction of bankruptcy may be highly beneficial. Some predictions may have large utility for the individual bank, government and the other stakeholders. If bankruptcy could be predicted with reasonable accuracy ahead of time, banks could better protect their business and take action to minimize risk and loss of business, perhaps even to prevent bankruptcy (Pongsatats, Ramage & Lawrence, 2004). Bankruptcy study is important for auditors as they

\* Assistant Professor, Department of Accountancy, S.V.S. Sridora Caculo College of Commerce and Management Studies, Khorlim, Goa, India. Email: riddhiprabhu17@gmail.com

\*\* Principal, Shree Mallikarjun College, Goa, India. Email: mskamat@gmail.com

are acknowledging the probability in the going concern of the bank. Bankruptcy study will provide information on variables that will influence the health of bank (Stinjak & Warastuti, 2014). A proper prediction of firm bankruptcies might therefore be extremely important and of great interest wide range of relevant financial actors. There is a need of early warning of financial distress due to changes happening in the recent economic environment. To avoid the risk of bank failure, banks should find the reliable ways to predict bankruptcy.

Literature review shows different methods available for predicting bankruptcy. One of the methods for bankruptcy study is scoring model. The scoring model is a linear combination of the factors (accounting variable), weighted by coefficients which provides relevant score. The scoring model has become very popular during the last 40 years in credit risk applications, forming a vast and fast-growing literature. There are many models available for bank failure predictions such as Standard and Moody's financial ratios, Beaver model, Altman Z-score model, Ohlson's model, CAMEL model, Grover model, Springate model, Neural Network and Zmijewskis model to predict the health of the bank. Among many techniques available for evaluating the financial performance of banks, literature survey shows that majority of international failure prediction studies employ Grover model, Springate model and Zmijewski model. In the past, (Vaziri O, 2012) used Z-score model for predictability of failure of financial institution and the model is found to have high predictive ability. Since model has high predictive ability, the present study applied these models on the Indian banking sector. The existing literature showed that a single bankruptcy prediction model faces limitations and multiples bankruptcy prediction model improved the prediction of accuracy in bankruptcy prediction. Therefore, the present study will apply multiple models to assess the banks bankruptcy. In the past, some of the researchers applied multiple models and its comparison to find the accuracy and predictive power of models. These reviews are: (Imanzadeh P., 2011), (Kodlar A.), (Vaziri O., 2012), (Karamzadeh, 2013), (Avenhuis, Nov 2013), (Hussein F., Dec 14), (Timmermans, 2014), (Kumar & Kumar, April 2012), (Sajjan, 2016), (Aminian A., July 2016), (Stinjak & Warastuti, 2014), (Wati M. & Hidayat R., Nov 2015), (Primassari, 2017), (Monousaridis) and (Syamni G., 2018), who used multiple models and their comparison in their studies. After the comparison, some of

the studies criticized certain models with respect to their predictive power, selection of the variables, time factor, accuracy rate, change in the economic environment, etc. According to Timmerman (2014), when an old original model is applied to a more recent sample the predictive power of the model is very low and the bankruptcy is over predicted. Due to this fact, the present study will recalibrate these models by changing coefficients of original models using current data. A study by (Timmermans, 2014) found better results in its study after recalibration of the models.

## **Models and Literature Review**

### **Springate Model (Model I)**

The Springate model was the first model to be introduced by Gordon LV Springate (1978). This model is a revolution of the Altman model developed by Multiple Discriminant Analysis (MDA). Springate model development process initially used 19 financial ratios that have been frequently used. However, after testing Springate finally chose four financial ratios to be used to determine whether the company is said to be either a healthy company or potentially insolvent. Springate used 40 companies as the sample for this research. Springate test shows that the model has an accuracy rate of 92.5%.

### **Grover Model (Model II)**

This model was created by restoration or redesigns of the model of the Altman Z-Score. It takes X1 and X3 of the Altman model and then adds profitability ratios which are indicated by Return on Asset (ROA). As a result, Grover model is the most appropriate predictive model applied to companies in the food and beverage sector. The study shows that the Grover model has the highest degree of accuracy that is equal to 100% (Grover 2003). On the contrary, the Altman model of Z-Score has an accuracy rate of 80%, Springate model, 90% and Zmijewskis models, 90%.

### **Zmijewskis Model (Model III)**

This model uses a ratio measure of performance, leverage and liquidity applied in companies is already bankrupt, and the company can survive. If Zmijewski's models

have exceeded predictions of 0, then the company could potentially be bankrupt; otherwise, if the company has a

score of less than zero, then the company is not potentially bankrupt. This model has an accuracy rate of 90%.

**Table 1: Description of Bankruptcy Models**

Sr. No	Model	Equation	Description	Score Category
1	Springate (1978)	$SS = 1.03X1 + 3.07X2 + 0.66X3 + 0.4X4$	SS = Springate Score X1 = Working capital/Total asset X2 = Net profit before interest taxes/total asset X3 = Net profit before Taxes/Current liabilities X4 = Sales/Total asset	SS > 0.862 = healthy SS < 0.862 = bankrupt
2	Grover (2001)	$G = 1.650X1 + 3.404X3 - 0.016ROA + 0.057$	GS = Grover Score X1 = Working capital/Total assets X2 = Earnings before interest and taxes/Total assets ROA = net income/total assets	GS ≤ -0.02 = bankrupt GS ≥ 0.01 = health
3	Zmijewski (1983)	$Z = -4. -4.5X1 + 5.7X2 - 0.004X3$	ZS = Zmijewski Score X1 = ROA (Net income/total assets) X2 = Leverage (Total liabilities/total assets) X3 = Liquidity (Current assets/current liabilities)	ZS > 0.5 = healthy ZS < 0.5 = bankrupt

Source: Secondary Data

## Literature Review

In the past, some of the researchers applied multiple models and their comparison to find the accuracy and predictive power of models. These reviews are: (Imanzadeh P., 2011), (Kodlar A.), (Vaziri O, 2012), (Karamzadeh, 2013), (Avenhuis, Nov 2013), (Hussein F., Dec 14), (Timmermans, 2014), (Sajjan, 2016), (Aminian A., July 2016), (Stinjak & Warastuti, 2014), (Wati M. & Hidayat R., Nov 2015), (Primassari, 2017), (Monousaridis) and (Syamni G., 2018), who used multiple models and their comparison in their studies. After comparison, some of the studies criticised certain models with respect to their predictive power, selection of variables, the time factor, accuracy rate, change in the economic environment, etc. According to Timmerman (2014), when an old original model is applied to a more recent sample, the predictive power of the model is very low and the bankruptcy is over predicted. Due to this fact, the present study will recalibrate these models by changing coefficients of original model using current data. A study by (Timmermans 2014) found better results in their study after recalibration of models.

Further, the recalibrated model will be compared with the original model to identify whether there exists a significant difference between these two models. If there exists a significant difference between these two models, the result of both these models will be compared with the robust test; then based on this comparison, the accuracy of the model will be discovered.

Literature review on Springate model shows the predictive ability of this model. A study by Sajjan (2016) aimed at presenting a theoretical foundation and compares the result of investigating two models Zavgren and Springate. Result indicates adjusted Springate model was efficient than other model in the bankruptcy year. A study by (Imanzadeh P., 2011) applied Springate and Zmweskis models on the firms of Tehran stock exchange. Its study concludes that Springate is more conservative model then Zmweskis model. Thus, most of the other reviews express that other models are better than the Springate model. This depicts the need for revision of Springate model. Zmijewski's model uses a ratio measure of performance, leverage and liquidity applied in companies is already

bankrupt and the company can survive. This model is also criticized by the reviewer such as (Timmermans, 2014). Grover model is a model developed by restoration of the Altman Z-model by Jeffrey S. Grover. According to (Primassari, 2017), during the development of model he used the sample according to Altman Z-score by adding 13 financial ratios. He used a sample of 70 companies with 35 bankrupt and 35 non-bankrupt companies. He considered a data period from 1982 to 1996. Literature review on Grover model shows the following facts: a study was conducted by (Qamruzzaman & Jianguo, Dec. 2016) on Grameen bank in Bangladesh using the same models. Its study concludes that G-score provides conflicting predictions. (Primassari, 2017) found Grover as a least accurate model. Majority of the reviews on Grover model reflects negative views on its predictions. This suggests the need for the revision of this model.

## Research Methodology

The sample used for the recalibration of bankruptcy's model is 39 public and private banks by changing coefficients of original bankruptcy model using current data. The data period used to find the new coefficients of the recalibrated model is from 2005 to 2010 in a sample consisting of 39 banks. The multiple regression technique is used to find the coefficients of recalibrated bankruptcy model. In total, 255 observations are used to find the coefficient of recalibrated model. The variables of 41 banks are chosen as per original bankruptcy model. The recalibrated model consists of same variable as the original model but the coefficients differ. To develop

a recalibrated bankruptcy model, all the assumptions required for the regression were checked. These assumptions are correlation coefficients, independence of errors, multicollinearity, homoscedasticity, normality, normality of residuals, presence of outliers and data fit for the model. The recalibrated model is applied on 49 public and private banks covering a period of 2011–2017. Three non-working banks were not taken in the sample of recalibrated model as the recalibrated model was applied for a period 2011–2017 and these three non-working banks are not working before 2011; hence, the data was not available for a period of 2011–2017. There are several useful criteria for measuring the goodness of fit of the multiple regression models. One such criterion is to determine the multiple correlation coefficients denoted as  $R$  and the square of the multiple correlation coefficients denoted as  $R^2$ . After the application of recalibrated models on Indian banks, there is need to find the accuracy of these models. The model accuracy was found out by detecting Type I and Type II errors. Type I error is rejecting a null hypothesis which is true while a type II error is fail to reject a null hypothesis which is false. In other words, type I error in prediction model means predicting a failed firm will not go bankrupt and type II error is predicting a non-failed firm to be bankrupt.

## Results and Discussion

Springate, Grover and Zmijewski models were recalibrated using the above-mentioned methodology. The results of the same are presented in Table 2.

**Table 2: Summary of Different Scoring Models**

Sr. No	Model	Original Equation	Recalibrated Equation	Description	Types of Variable	Bases of Discrimination
1	Springate (1978)	$S = 1.03X_1 + 3.07X_2 + 0.66X_3 + 0.4X_4$	$S\text{-scores} = .093 + 1.37X_1 + 7.124X_2 + 0.585X_3 - 1.988X_4$	S = Springate Score X <sub>1</sub> = Working capital/ Total asset X <sub>2</sub> = Net profit before interest taxes/ total asset X <sub>3</sub> = Net profit before Taxes/Current liabilities X <sub>4</sub> = Sales/Total asset	X <sub>1</sub> = Liquidity X <sub>2</sub> = Profitability X <sub>3</sub> = Profitability X <sub>4</sub> = Profitability	S > 0.862 = healthy S < 0.862= bankrupt

Sr. No	Model	Original Equation	Recalibrated Equation	Description	Types of Variable	Bases of Discrimination
2	Grover (2001)	$G = 1.650X_1 + 3.404X_3 - 0.016ROA + 0.057$	G-scores =.074 $+1.65X_1 + 3.09X_2 - .014X_3$	G = Grover Score $X_1 =$ Working capital/Total assets $X_2 =$ Earnings before interest and taxes/Total assets ROA = Net income/Total assets	$X_1 =$ Liquidity $X_2 =$ Profitability ROA = Profitability	$GS \leq -0.02 =$ bankrupt $GS \geq 0.01 =$ healthy
3	Zmijewski (1984)	$X = -4. -4.5X_1 + 5.7X_2 - 0.004X_3$	$X = -2.534-3.79X_1 + 4.246X_2-.081$	X = Zmijewski Score $X_1 =$ ROA (Net income/Total assets) $X_2 =$ Leverage (Total Debt/Total assets) $X_3 =$ Liquidity (Current assets/current liabilities)	$X_1 =$ Profitability $X_2 =$ Leverage $X_3 =$ Liquidity	X-score < 0 = Safe X-score > 0 = bankrupt Prob (X-score) = $\leq 0.5$ "Healthy Prob (X-score) = > 0.5 "bankrupt

Source: Secondary Data

### Application of Recalibrated Models

The recalibrated Springate model, Zmijewski model and Grover model were applied on 41 public, private and non-

working banks covering a period of 2011–2017. Tables 3 and 4 show the result of recalibrated models.

**Table 3: Application of the Recalibrated Model on Twenty-One Public Sector Banks Covering a Period of 2011-2017**

Sr. No.	Banks	S-Score	Decision	Zm Score	Probability	Decision	G-Score	Decision
1	Allahabad Bank	3.06	Safe	-1.37	0.08	Safe	0.42	Safe
2	Andhra Bank	3.12	Safe	-2.22	0.02	Safe	0.45	Safe
3	Bank of Baroda	2.36	Safe	-2.67	0.01	Safe	0.59	Safe
4	Bank of India	2.57	Safe	-1.79	0.04	Safe	0.46	Safe
5	Bank of Maharashtra	2.64	Safe	-1.32	0.08	Safe	0.43	Safe
6	Canara Bank	2.83	Safe	-1.99	0.03	Safe	0.46	Safe
7	Central Bank Of India	2.49	Safe	-0.72	0.21	Safe	0.41	Safe
8	Corporation Bank	3.17	safe	-1.51	0.06	Safe	0.42	Safe
9	Dena Bank	3.10	Safe	-1.84	0.03	Safe	0.47	Safe
10	I D B I Bank Ltd.	2.84	Safe	1.09	0.87	Safe	0.45	Safe
11	Indian Bank	2.77	Safe	-1.30	0.09	Safe	0.40	Safe
12	Indian Overseas Bank	2.88	Safe	-1.40	0.07	Safe	0.47	Safe
13	Oriental Bank Of Commerce	3.21	Safe	-0.93	0.16	Safe	0.44	Safe
14	Punjab & Sind Bank	3.66	Safe	-1.49	0.06	Safe	0.40	Safe
15	Punjab National Bank	2.82	Safe	-1.94	0.03	Safe	0.43	Safe

Sr. No.	Banks	S-Score	Decision	Zm Score	Probability	Decision	G-Score	Decision
16	State Bank Of India	2.15	Safe	-0.79	0.19	Safe	0.40	Safe
17	Syndicate Bank	2.58	Safe	-5.35	0.01	Safe	0.46	Safe
18	Uco Bank	2.42	Safe	-2.36	0.01	Safe	0.44	Safe
19	Union Bank of India	2.99	Safe	-1.04	0.13	Safe	0.45	Safe
20	United Bank of India	2.49	Safe	-1.05	0.13	Safe	0.47	Safe
21	Vijaya Bank	3.21	Safe	-0.61	0.25	Safe	0.37	Safe

Source: Authors Calculations

**Table 4: Application of Recalibrated Model on Twenty-One Public Sector Banks Depicting the Scores of Individual Models Covering a Period of 2011 -2017**

Sr. No.	Banks	S-Score	Decision	Zm Score	Probability	Result	G-Score	Decision
1	Axis Bank Ltd.	2.76	Safe	-2.94	0.01	Safe	0.46	Safe
2	Catholic Syrian Bank Ltd.	3.08	Safe	-1.28	0.09	Safe	0.35	Safe
3	City Union Bank Ltd.	3.14	Safe	-2.24	0.02	Safe	0.47	Safe
4	D C B Bank Ltd.	0.77	Distress	-0.08	0.46	Safe	0.30	Safe
5	Dhanalaxmi Bank Ltd.	2.73	Safe	1.79	0.96	Distress	0.38	Safe
6	Federal Bank Ltd.	3.12	Safe	-2.51	0.01	Safe	0.48	Safe
7	H D F C Bank Ltd.	5.01	Safe	-2.65	0.01	Safe	0.56	Safe
8	I C I C I Bank Ltd.	2.65	Safe	-2.47	0.01	Safe	0.39	Safe
9	Indusind Bank Ltd.	2.95	Safe	2.20	0.01	Distress	0.38	Safe
10	Jammu & Kashmir Bank Ltd.	2.94	Safe	-4.53	0.00	Safe	0.40	Safe
11	Karnataka Bank Ltd.	3.18	Safe	-1.01	0.14	Safe	0.32	Safe
12	Karur Vysya Bank Ltd.	3.39	Safe	-1.96	0.03	Safe	0.41	Safe
13	Kotak Mahindra Bank Ltd.	0.80	FD	-2.73	0.01	Safe	0.42	Safe
14	Lakshmi Vilas Bank Ltd.	3.11	Safe	-2.23	0.02	Safe	0.41	Safe
15	Nainital Bank Ltd.	2.47	Safe	-5.31	0.00	Safe	0.76	Safe
16	R B L Bank Ltd.	2.55	Safe	-8.83	0.00	Safe	0.38	Safe
17	South Indian Bank Ltd.	3.06	Safe	-1.43	0.07	Safe	0.45	Safe
18	Yes Bank Ltd.	0.45	FD	-1.96	0.03	Safe	0.36	Safe
Merged Banks								
1	Bharat Overseas. Bank Ltd. (M)	1.92	Safe	-1.02	0.14	Safe	0.43	Safe
2	I N G Vysya Bank Ltd. (M)	1.34	Safe	1.83	0.97	Distress	0.43	Safe

Source: Authors Calculations

The tested recalibrated Springate model shows three good banks in the financial distress position and the merged banks in a safe position. In the case of the recalibrated Zmijewski model, three banks are in the distress condition, whose probability of bankruptcy is more than 50%. However, this model also shows the non-working banks in the safe position. The Grover model shows all

the good banks in a safe position and also non-working banks in a safe position. Thus, the results of all the three models show certain discrepancies; hence, the need felt to check the accuracy of these models. The accuracy of models is checked by detecting the Type I and Type II errors in the model.

**Table 5: Showing the Accuracy Rate of Models**

<i>Springate Model</i>	<i>Error Rate</i>	<i>Model Accuracy Rate</i>
Non working Banks	Type I - 5% and Type II	88%
Good Banks	7%	
<i>Zmijewski Model</i>		
Non working Banks	Type I - 3% and Type II	90%
Good Banks	-7%	
<i>Springate Model</i>		
Non Working Banks	Type I - 5% and Type II	95%
Good banks	0 %	

Source: Authors Calculations

## Conclusion

The study of bankruptcy models attempts to find whether the high predictive ability, which bankruptcy models claimed, is still valid in the current business environment. The present study used three bankruptcy models such as Springate (1978), Zmijewski (1984) and Grover (2003). After the application of these models in the Indian banking sector, the results did not hold the models very strongly. The model accuracy rate was low compared to what was claimed by the original models in their own country. With this background of low model accuracy rate and the critiques of models in the past literature review, there was a need to recalibrate these models. The models were recalibrated by changing its coefficients using current data and sticking to the same variables used by original models. Multiple regression techniques were used to find the coefficients of the recalibrated model.

The results of the first model called Springate (1978) model failed to accept the hypothesis that there is no difference in the predictive ability between Springate and recalibrated Springate model. The model accuracy rate of the recalibrated model was low compared to original Springate model and also the results of robust test were matching with the original Springate model. Thus, original Springate model performs better than the recalibrated Springate model. However, it still works better in the country of its origin as it showed 93% accuracy rate over there but it proves a less accuracy in some other country as the economic conditions differ from country to country. Also, the original model shows less accuracy rate due to its sample bias; as originally, it was applied for

the manufacturing firms but in the present study it is used in the service sector. The results Zmijewski (1984) model strongly uphold the hypothesis that there is no difference in the predictive ability between the Zmijewskis and recalibrated Zmijewski's model. There was not much difference in the model accuracy rate of Zmijewski and the recalibrated Zmijewskis model; however, the original Zmijewski model rate is higher. Therefore, original Zmijewski model performs better than the recalibrated Zmijewski model. The results of the third model, known as Grover model, also strongly uphold the hypothesis that there is no difference in the predictive ability between the Grover and recalibrated Grover model. There was not much difference in the model accuracy rate of Grover and the recalibrated Grover model; however, the recalibrated Grover model rate is higher. Therefore, the recalibrated Grover model performs better than the original Grover model.

Above summary of all models shows that the recalibrated Grover model performs better than original model and original Springate and Zmijewski models showed the improved accuracy over recalibrated models.

## Policy Implications and Scope for Further Study

The accounting ratios used in these models are poor predictors. The market-driven variables such as past stock returns, high exchange rate and interest rate, which are strongly related to bankruptcy probability, are neglected in these models. Combining these market-driven variables with two accounting ratios may result in a better development of this model. Thus, models should also include macroeconomic variables. Certain environmental factors such as different GDP, unemployment rate, interest rate environments, tax regimes, wage levels, and access to capital markets contribute to the environment heterogeneity. The results of previous studies show that economic environment can affect the accuracy of bankruptcy prediction model. Hence, a further study should focus on incorporating these factors into the model and evaluating their contribution to model accuracy. This may improve the model robustness. The growth of banking industry is a major contributor for the economic development of the country. Many banks enter in financial distress position for different reasons. This

may cause thousands of employees to lose their jobs and a large number of investors to lose their savings. Also, communities will lose vital services and the economy of the country will suffer. Much of these losses can be avoided with predictions of bankruptcy system, which allows adequate plans and take necessary steps. The RBI till today follows a CAMEL model to give its opinion about the banks merger. Although CAMEL model is a good indicator to highlight the banks health, it has its own limitations. Hence, the RBI should make a policy of using recent models such as MDA technique, logit, Probit models along with the CAMEL model for the financial health assessment of banks.

As a follow-up study, a further suggestion is to investigate the study to different economic periods to observe how the models perform in different environments and time periods. That would mean that a cross-sectional study would be applied. Based on the above findings and policy implications, the suggestions for subsequent research are to test bankruptcy position using data from non-banking companies. Further research can also extend to cover 10 – 20 years period of observation. The bankruptcy models used entirely a historical result. Going forward, the model should encompass not only accounting but also other determinants of bank failure. External environmental factors such as volatile stocks, high exchange and interest rate can have a devastating effect on the health of the bank.

## References

- Aminian, A., Mousazade, H., & Khoshkho, O. I. (July 2016). Investigate the ability of bankruptcy prediction model of Altman, Springate, Zmijewski and Grover in Tehran stock exchange. *Mediterranean Journal of Social Science, Rome, Italy*, 7(4), 208-214.
- Avenhuis, J. O. (2013). *Testing the generalizability of bankruptcy prediction models of Altman, Ohlson and Zmijewski for dutch listed and large non listed firms*. Thesis-University of Twente, School of Management and Governance, Netherlands.
- Hussein, F., & Pambetki, G. (2014). Precision of model Altman, Springate, Zmijewski and Grover for predicting financial distress. *Journal of Economics, Business and Accountancy*, 17(3), 405-416.
- Imanzadeh, P., Maran-Jouri, M., & Sepehri, P. (2011). A study of application of Springate, Zmijewski, bankruptcy prediction model in the firms accepted in Tehran stock exchange. *Australian Journal of Basic and Applied Science*, 5(11), 1546-1550.
- Kumar, R. G., & Kumar, K. (April 2012). A comparison of bankruptcy model. *International Journal of Marketing, Financial Service and Management Research*, 4, 76-86.
- Karamzadeh, M. S. (2013). Application and comparison of Altman and Ohlson model to predict bankruptcy of companies. *Research Journal of Applied Science, Engineering and Technology*, 5(6), 2007-2011.
- Kodlar A., N. (n.d.). *Comparing bankruptcy prediction*. School of Doctoral Studies (European Union Journal).
- Lawrence, J. R., Pongsatit, S., & Lawrence, H. (2015). The use of Ohlson O score for bankruptcy prediction in Thailand. *Journal of Applied Business Research*, 31(6), 2069-2078.
- Maina, F. G., & Sakwa, M. M. (n.d.). Understanding Financial distress among listed firms in Nairobi stock exchange: A quantitative approach using the Z-score multidiscriminant financial analysis.
- Monousaridis. (n.d.). *Z-Altman model effectiveness in bank failure prediction - The case of European bank*. Lund University School of Economics and Management.
- Pongsatit, S., Ramage, J., & Lawrence, H. (2004). Bankruptcy prediction for large and small firms in Asia: A comparison of Ohlson and Altman model. *Journal of Accounting and Corporate Governance*, 1(2), 1-13.
- Primassari, N. (July 2017). Analysis of Altman Z-score, Grover Score, Springate, and Zmijewski as financial distress signalling: Empirical study of consumer goods industry. *Accounting and Management Journal*, 11(1), 23-42.
- Qamruzzaman, M., & Jianguo, W. (Dec. 2016). Analysis of financial distress on micro financial institutions (MFI) in Bangladesh: A case study of Grameen bank. *Commonwealth Journal of Commerce and Management Research*, 3(12), 1-15.
- Sajjan, R. (April 2016). Predicting bankruptcy of selected firms by applying Z-score model. *International Journal of Research Granthalaya*, 4(4), 152-158.
- Stinjak, N., & Warastuti. (2014). Analysis of model based prediction of bankruptcy in banking companies listed in Indonesia stock exchange. *South East*

*Asian Journal of Contemporary Economics and Law*, 15(Dec.), 71-80.

- Syamni, G. M. S. (2018). Bankruptcy prediction model and stock prices of coal mining industry in Indonesia. *Etikonomi*, 17(1), 57-68.
- Timmermans, M. (2014). *U.S. corporate bankruptcy predicting models*. University of Tilburg.
- Vaziri, O. B. R. (2012). Comparative predictability of failure of financial institution using multiple model. *Investment and Financial Innovations*, 9(2), 120-126.
- Wati, M., & Hidayat, R. (2015). The analysis of bank health level using Zmijewski model, Ohlson S-score model, Altman Z-score model. *Journal of Administration Business*, 28(1), 185-192.

# Emerging Trends in Accounting: An Analysis of Impact of Robotics in Accounting, Reporting and Auditing of Business and Financial Information

M. L. Ashoka\*, Abhishek N.\*\*, Divyashree M. S.\*\*\*

## Abstract

Accounting is the window of any business firm. With the increased usage of computers in the arena of the Accounting Information System, we can see a drastic change from paper-journals and ledgers to computer-based formats. So to walk along with this drastic change it is necessary to integrate Robotics (Artificial Intelligence) into the Accounting Database. Application of robotics technology in accounting is nothing but an application of expert system based software and other technologies during the process of recording, reporting and communication of business and financial information and also in the auditing process. This enables the business houses in easy and transparent reporting and communication of the business information and also to cater the needs of decision makers. So the present paper is in conceptual nature and intended to analyse the Impact of Robotics in Accounting, Reporting and auditing of Business and Financial Information with the help of secondary sources such as journals, websites, reports etc.

**Keywords:** Accounting, Auditing, Robotics in Accounting

## Background

Accounting is the mouth of the business as it speaks about operational and financial performance. Success or failure of each and every business entity is totally depended upon how it caters the needs of various stakeholders. To cater the needs of various stakeholders, it is not enough to provide the goods and services as it involving

in production and rendering of goods and services; however, it is also necessary to perform other duties such as undertaking social obligations, environment protection activities, obeying the government and other regulatory authority's requirements. After discharging its duties, the business has to communicate the information on its performance in relation to above-said activities to ensure the accountability to all its stakeholders. In simple, we can say that proper communication of business and financial information is the main concern of the management to get the confidence from all the stakeholders. The technologies used in the business to conduct accounting operations are finding new avenues for development. One of the recent developments in the domain of accounting information dissemination system is the usage of robotics in accounting. This development, in the conduct of accounting operations, directly aids in faster, economical and paperless ways.

The present world is filled with full of technology because we can observe the usage of technology in all the spheres of life. Starting from birth till death, we experience infinite usage of technology. In the business world also, there is intensive dependence on the technology and coping with the dynamic changes due to the new up-gradations to the existing one. In accounting functions also, we can observe the rapid changes from paper-pencil based function to computer-internet and software-based functions. Today is the era of Artificial Intelligence (Robotics). So, many accounting firms working around the world are trying to initiate robotics in accounting function because of its fruitful benefits such as it eliminates the tedious and difficult accounting routine tasks and enables the accountants to perform efficient consultancy services.

\* Professor, DoS in Commerce, University of Mysore, Karnataka, India.

\*\* University of Mysore, Mysore, Karnataka, India. Email: [abhishekalmighty93@gmail.com](mailto:abhishekalmighty93@gmail.com)

\*\*\* GFGC Bilikere, Mysore, Karnataka, India.

The basic objective of usage of robotics is to simplify the more complex human tasks by AI. According to Financial Stability Board publication, both public and private sector organisations are using robotic technologies for regulatory compliance, surveillance, data quality assessment and fraud detection. Today, it is evidently observed that accounting and reporting functions of the business organisations are transferred from paper journals and ledger patterns into a computerised pattern with the introduction of computers, which are supported by robotics technology. Robotics technology works with its own special features like self-management, self-tuning, self-configuration, self-diagnosis and self-healing to ease the operations of accounting.

## Robotics in Accounting

With the advent of computers in the arena of the accounting information system, we can see a drastic change from paper-journals and ledgers to computer-based formats. The accounting information of a business house is being stored in the accounting databases. In these storehouses of the database, only specific accounting transactions are maintained. As a consequence, these systems are not meeting the specific needs of the stakeholders (decision-makers) of the business. Therefore, there is a need to apply a certain type of intelligence into accounting database, which helps to eradicate the flaws of the traditional system. So, one approach to solve this problem is to integrate robotics (AI) into the accounting database, which meets the needs of decision-makers. So, the present paper is intended to analyse the impact of robotics in accounting, reporting and auditing of business and financial information. For this purpose, the next part of the paper is organized as a literature survey, research gap, research questions, objectives, methodology, discussions and conclusions, limitations and scope for further research.

## Literature Survey

Abdolmohammadi et al. (2001) analysed the auditor's perception on the application of expert systems for auditing function and concluded that irrespective of the nature of task complexity auditors more prefer for knowledge-based systems along with human processing.

Amelia et al. (2006) examined the application of artificial intelligence in accounting and auditing and opined that auditing and accounting function can be potentially

improved through the usage of complex AIs such as expert systems, genetic programming, neural networks, fuzzy systems and hybrid systems, should be investigated to the fullest extent possible.

Daniel et al. (1997) empirically analysed the impact of artificial intelligence in accounting, taxation work and influence on other organisational issues and opined that application of artificial intelligence helps top management and reduces the need for supervision because it works based on expert system which allows to solve a complex problem of accounting and taxation and allows the organizations to perform more work with less supervision. Further, it helps to take decisions immediately.

Murphy and Yetmar (1996) conducted the study on the context of use of expert system for the credibility of audited information and found that the application of expert system for the usage by subordinates affects the beliefs of superiors but use of ES by superiors does not affect on their own beliefs and on their own decisions.

O'Leary et al. (1991) discussed the role of artificial intelligence and expert system on accounting function and in solving the problems involved in traditional accounting functions. The researcher concluded that the use of AI in accounting will solve the problems of traditional paper-based accounting.

Odoh et al. (2018) examined the effect of artificial intelligence on the performance of accounting operations among various accounting firms in South-East Nigeria. They concluded that the application of artificial intelligence favourably impacts on the performance of accounting functions. Further, they recommended that accountants and accounting firms should continually upgrade their skills and knowledge in relation to artificial intelligence to enhance the performance of accounting functions, thereby eliminating certain accounting cost.

Patrick B Dorr et al. (1988) studied expert systems based decision aid in AIs and opined that that expert system decision aid was more efficient and effective method than the checklist-based system decision aid on accounting and business information of the business organisation.

## Research Gap

From the review of earlier studies, it is evidently found that the most of the studies have basically focused on the impact of artificial intelligence but only a few studies have focused on accounting and auditing function with the help

of artificial intelligence. So, the present study is to analyse the impact of the application of robotics in accounting and auditing functions.

## Statement of the Problem

The application of technology is one of the important developments in the area of accounting because it provides greater benefits to various categories of stakeholders; especially, nowadays application of robotics is the big debate in the area of accounting research. So, the present study is aimed at analysing the impact of the use of robotics in accounting and auditing operations.

## Research Questions

- What is the importance of the application of robotics in accounting and auditing?
- What are the issues concerning the implementation of robotics in accounting and auditing?
- What are the thrust accounting and auditing areas for the application of robotics?

## Objectives

- To analyse the importance of the application of robotics in accounting and auditing.
- To examine the issues concerning the implementation of robotics in accounting and auditing.
- To identify the thrust areas for the application of robotics in accounting and auditing.

## Methodology

The application of robotics in accounting operations is still in the nascent stage. So, the present study is in conceptual conducted based on secondary sources collected through published sources such as journals, articles, websites, reports, magazines, etc.

## Discussions

### Importance of Application of Robotics in Accounting and Auditing

The development and usage of robotics (AI) in accounting majorly impact the allocation of resources. Because it will

be an effect on the amount of money expended on the accounting and auditing process as it may reduce payroll costs or enhance the decision-making process. Whereas on the other hand, usage of robotics will require an investment of time and monetary resources for procuring robotic technological equipment but the application of robotics in accounting and auditing is cost-beneficial in the long-run.

Further application of robotics in accounting helps in making accounting and auditing-based decisions because it is build based on knowledge and reasoning capacity. Integration of artificial intelligence in accounting and auditing is important due to the following reasons:

- Conventional accounting only dealt with reporting and communicating financial information of the corporate houses but now the environment is in dynamic nature so that the stakeholders need additional volatile information apart from financial. To do this, tradition paper-based and human-readable formats of reports such as PDF, word and excel are not appropriate because these formats are not readable by machines, i.e., computers. So, it is necessary to have reporting languages which are readable by both machines and humans. This can only be done through the use of AI in reporting of business and financial information which is popularly known as Integrated Reporting.
- Traditional financial statements such as income statements, fund flow statements and cash flow statements are the base for getting the data for decision making by the investors. But not all are aware of using these data. To enable them for easy decision-making based on the above statements, AI with knowledge-based expert system can be used.
- Use of robotics that is AI in accounting will not remove the accountants from the job but it eases the accounting and auditing functions to be performed by the accountants and auditors. Preparation and production of paper-based reports with manual processing will be removed because of the reports of accounting and non-financial information of the business will be internet based. Merging expert system with accounting information system helps to gather the huge volume of data with or without the direct effort of the stakeholders. In simple words,

AI or robotics-based system can analyse the data and helps the stakeholders in understanding and interpreting the financial events.

- Filing of various disclosure reports to regulatory authorities was one of the challenges to the business organisation because before filing the reports, it is important to get a compliance audit report. If the Accounting Information System based on AI is used, it will be very easy for the organisations to file reports to regulatory with regulatory authorities with automated verification of standard disclosure requirements with the actual disclosures and which will eliminate the compliance audit. Nowadays, this is carried with the help of XBRL (Extensible Business Reporting Language) in various countries. Further, it helps the firms to file the reports in paper-less mode, i.e., electronic format.
- In auditing context, auditors usually prefer human processing for their function even if it is a complex and time-consuming process. So, nowadays knowledge-based expert systems (Robotics) are most advisable along with human processing because of its efficiency and fruitfulness such as it helps the auditors to replace the checklist verification of accounts system with automated computerised verification system.
- For auditing purpose, auditors need sufficient explanation from their clients; but in all the time, it is difficult to get a proper explanation from the clients on the different contexts. If auditors rely on AI for this purpose, they will get correct explanations as it is a knowledge-based expert system. But to do this, there is a need of effort from academicians and researchers for advanced research on this aspect.

- If the auditing to be considered as efficient, it must ensure the audited accounts are within the four corner points of law. So, here to ensure all the information reported are in “true and fair” state, the auditors must verify the information with legal provisions and it is a highly complex human process. If we apply a knowledge-based expert system, i.e., Robotics, it is very easy to assess the risk of litigation in relation to the accounts of the clients.
- By using AI in auditing function, the auditor can carry the analysis function to verify the vouchers to ensure whether there evidence for the accounts of the clients or not.
- AI can also use the function in which auditor is intended to determine fraudulent insurance claims.
- Ensuring the materiality of transactions reported in the accounts is one of the important tasks to be performed by the auditors and it is also one of the difficult tasks for this also robotics can be applied.
- Application of Artificial Intelligence in auditing is also beneficial in bankruptcy prediction.

### **Issues Concerning the Implementation of Robotics in Accounting and Auditing**

The organisation’s management decisions are totally based on accounting information and also they are impacted based on the functional areas of accounting. In this context, functional areas of accounting can be seen as accounting, auditing, management accounting, tax accounting, information system and financial accounting. Table 1 discusses how robotics is important to perform these functions.

**Table 1: Issues Concerning the Implementation of Robotics in Accounting and Auditing**

<i>Areas of Accounting</i>	<i>Issues to be Resolved by Robotics</i>
Auditing	<p>In auditing, there may be three types of auditors who can perform the auditing task and issues involved are:</p> <p>The external auditor is an auditor who is an outsider of an organisation working for remuneration. He expresses the opinion on the annual financial statements of an organisation with reference to the documents submitted by the management. His opinion speaks about 'true and fair' view of books of accounts and helps in getting the confidence of various stakeholders. Thus, the management, researchers, academicians of both accounting and robotics are concerned with the decision process of external auditors. So, it is required to develop a robotic knowledge-based expert system to make this decision process even more effective.</p> <p>The internal auditor is an auditor who is an employee of an organisation and reviews the operations and their activities of the firm. To review the operation and other activities, an advanced tool based on robotic technology is needed. So, here management, researchers and academicians are had to concentrate to address these issues.</p> <p>Computerized or EDP auditors may be external or internal to the organisation who involves in auditing computerized accounting information system to enable them, management, researchers and academicians are needed to concentrate on developing suitable tools and expert system to assess and control the computerised accounting system.</p>
Management Accounting	Management accounting involves in providing the information to management to aid them for decision-making. For this purpose, today it is necessary to have proper advanced robotic technology for planning and control system.
Tax Accounting	It needs an advanced robotic knowledge-based expert system for getting timely tax advises.
Financial Accounting and Information System	It needs to have a special knowledge-based expert system for following specific GAAP imposed for reporting and disclosure of business and financial information.

Source: Author compiled

### Thrust Areas for the Application of Robotics in Accounting and Auditing

The study identified some of the thrust areas for the application of robotics (AI) in accounting and auditing arena are:

#### Thrust Areas in Auditing

- Tool to review the information needed for the auditing process.
- Tool for assessing the actual amount allowed to bad debts and automated checking of the existence of those debtors who were treated as bad debts.
- Tool for analysing the auditing opinion process and making the decision process of auditors at a faster rate without any flaws.
- Tool for analysing the decisions for the uncertainty of going concerns and projecting the financial position in the future.

- Tool to develop auditing planning process and automated preparation of audit programme which aids in the audit process.
- Tool to analyse, design and evaluating the internal control system and automated application of internal check system.
- Tool to the credit evaluation system.
- Tool to the borrowings evaluation system.
- Tool to litigation risk assessment and detection of fraud.
- Tool to analyse the genuineness of the contract prices.
- Tool to evaluate the reliability of accounting information system.

#### Thrust Areas in Management Accounting

- Tool to inducing the capital budgeting process.
- Tool for implementing cost and management accounting standards or GAAPs.

## Thrust Areas in Tax Accounting

- Tool to analyse the tax implications on the financial and operational soundness of the business.
- Thrust areas in Accounting Information System.
- Tool to prepare and communicate business and financial reports (integrated reports).

With these thrust areas for the application of robotics in accounting and auditing, we can consider accounting and auditing is the domain which is providing greater opportunity for introducing robotics.

## Conclusion

Application of robotics in the area of accounting and auditing is one of the important developments in business research. Some of the debates argued that application of robotics in the field of accounting and auditing leads to loss of jobs by the accountants and auditors but the present study observed that it does not remove the accountants and auditors but it eases their operations in an efficient manner. Further, it helps to process accounting functions with the greater accuracy, speed and efficient manner and this also enables investors to get their required information for timely decision-making. Finally, the study concludes that application of robotics in the accounting and auditing is cost-beneficial in the long run, smoothen its managerial decision process, and enables the firm to easy filing of compliance and other disclosure reports to the regulatory authorities. Importantly, it will build confidence among all the stakeholders by fulfilling their needs at large.

## Suggestions for Policy Implication

The study observed that there is a gap between accounting researchers and AI researchers because accounting and AI researchers are carrying their researches independently. But in reality to make effective implementation of AI in the area of accounting and auditing, there is a need of carrying joint research by these two domain researchers. Further, to make the use of AI in a familiar manner, it is very important to initiate the robotics issues of accounting in the curriculum of accounting education.

## Limitations of the Study

The study is in conceptual in nature and studied the field and which is based on secondary sources. Every methodology applied for the study may suffer from its own limitations and which may influence the results of the study.

## Scope for Future Research

This study not observed the reality of the field and not conducted based on the primary data so future studies may consider this issue.

## References

- Abdolmohammadi, M. J., & Usoff, C. (2001). A longitudinal study of applicable decision aids for detailed tasks in audit. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10(3), 139–154.
- Abdolmohammadi, M. J. (1991). Factors affecting auditors' perceptions of applicable decision aids for various audit tasks. *Contemporary Accounting Research*, 7(2), 535–548.
- Anderson, J. C., Moreno, K. K., & Mueller, J. M. (2003). The effect of client vs. decision aid as a source of explanations upon auditors' sufficiency judgments: a research note. *Behavioral Research in Accounting*, 15(1), 1–11.
- Baldwin, A. A., Brown, C. E., & Trinkle, B. S. (2006). Opportunities for artificial intelligence development in the accounting domain: The case for auditing. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 14(3), 77–86.
- Berton, L. (1995). Big accounting firms weed out risky clients. *The Wall Street Journal*, 26, B1, B6.
- Comunale, C., & Sexton, T. R. (2005). A fuzzy logic approach to assessing materiality. *Journal of Emerging Technologies in Accounting*, 2(1), 1–15.
- Dorr, P. B., Eining, M. M., & Groff, J. E. (1988, March). *Using an Expert Systems Based Decision Aid in Accounting Information Systems*. In *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL Conference* (Vol. 15).

- <https://www.accelirate.com/robotics-in-finance-and-accounting/robotics-in-finance-and-accounting/>
- <https://www.accountexnetwork.com/blog/2018/02/emerging-trends-robotics-accounting/>
- <https://www.cpajournal.com/2018/07/02/how-robotic-process-automation-is-transforming-accounting-and-auditing/>
- <https://www.techfunnel.com/fintech/impact-of-robotics-in-finance-and-accounting/>
- Murphy, D. S., & Yetmar, S. A. (1996). Auditor evidence evaluation: Expert systems as credible sources. *Behaviour & Information Technology*, 15(1), 14–23.
- O’Leary, D. E. (1991). Artificial intelligence and expert systems in accounting databases: Survey and extensions. *Expert Systems with Applications*, 3(1), 143–152.
- O’Leary, D. E., & O’Keefe, R. M. (1997). The impact of artificial intelligence in accounting work: Expert systems use in auditing and tax. *Ai & Society*, 11(1–2), 36–47.
- Odoh, L. C., Echefu, S. C., Ugwuanyi, U. B., & Chukwuani, N. V. (2018). Effect of artificial intelligence on the performance of accounting operations among accounting firms in South-East Nigeria. *Asian Journal of Economics, Business and Accounting*, 7(2), 1–11.
- O’Leary, D. E. (1987, May). *The use of artificial intelligence in accounting* (pp. 83–98). In Expert Systems for Business Addison-Wesley Longman Publishing Co., Inc.
- Viaene, S., Derrig, R. A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk and Insurance*, 69(3), 373–421.
- Zhang, Q., Hu, M. Y., Patuwo, E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: A general framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16–32.

# Causal Link Between FDI and Economic Growth? Panel Data Analysis of MINT Economies

Karan Sabharwal\*

## Abstract

This paper attempts to find out the relationship between foreign direct investment (FDI) and economic growth in MINT economies namely, Mexico, Indonesia, Nigeria and Turkey, during 1981–2015. It investigates the causal link between the two variables, i.e. whether FDI causes economic growth or economic growth causes FDI. To test the relationship between the variables, panel data of four countries is analysed. Cointegration technique, Granger causality test and Vector Error Correction Model (VECM) have been used to validate the relationship between the variables. The analysis of the results shows that there exists cointegration between FDI and economic, i.e. there exists a long-run relationship between FDI and economic growth. Long-run causality was found moving from GDP to FDI in Mexico, Nigeria and Turkey. This is important for policy decision-making as barriers to FDI inflows should be removed, and growth can be maximized by increasing the capacity to absorb FDI inflows.

**Keywords:** JELC01, JELB23, JEL049, Economic Growth, Granger Causality, Foreign Direct Investment, Cointegration, Unit Root

## Introduction

The causal linkage between economic growth and foreign direct investment (FDI) is an issue which has been debated for long and no consensus has been reached in the field of development economics. Many researchers have supported the unidirectional linkage that FDI causes economic growth. FDI integrates economies and promotes internationalization by simulating new production techniques, trade-in technology and resources in terms of administrative skills, capital inflows as well as income

generation. FDI has been found to be one of the chief factors in promoting economic growth and development of the least developed countries (LDCs).

The spillover effect of the transfer of technology (ToT) and new product techniques from multinational corporation to indigenous firms increases economic growth. Thus, economic growth follows FDI. However, the spillover effect will only be visible if the absorption ability of the country of destination is developed. In anticipation of positive effect on returns from capital inflows, technology and skill transfer, job opportunities, developed countries have started seeking FDI by removing all the barriers which were restricting the growth and development of the country. The flow of FDI need not to essentially flow from developed to developing countries or from developing countries to developed countries. Even the less developed sectors provide profitable opportunities to MNCs for R&D and provide huge markets for expansion.

The causal link between FDI and economic growth needs careful investigation, which has major policy implications for developing economies or less developed countries. The linkage between FDI and economic growth can be expressed in three possible ways: Firstly, there can be a causal link from economic growth to FDI. If economic growth of a country attracts FDI, policymakers must make policies which accelerate economic growth rather than making efforts to seek FDI. Secondly, the causal relationship between the two variables can move from FDI to economic growth. If causality happens to flow from FDI to economic growth, FDI not only leads accumulate capital and increase employment opportunities but would also contribute to increasing the economic growth of the host counties. In this scenario, decision makers' decisions must focus on bringing FDI to the host country by removing barriers and making the host country an

\* Assistant Professor in Economics, Institute of Management Studies (IMS) Ghaziabad, Uttar Pradesh, India.  
Email: [karan2mds1987@gmail.com](mailto:karan2mds1987@gmail.com)

attractive destination. Finally, there can be bidirectional causality between economic growth and FDI. If the causal link runs in both ways, both economic growth and FDI strengthen each other and favourably affect one another.

There is no clarity on the direction of causality between the two variables besides the availability of huge literature. The literature suggests that the direction of causality largely depends on the past developments of the host country. There is no consensus on whether FDI causes economic growth or economic growth causes FDI. This paper examines the direction of causality between FDI and economic growth with respect to MINT economies.

This paper is organized as follows: Section II provides a survey of the literature. Section III discusses the data and empirical methodology. Analysis and interpretation of the results are explained in Section IV.

## Recent Literature

During the last two decades, a large number of studies focused on the role of FDI in stimulating economic growth in the LDC. But there is no consensus with regard to the direction of causality. Mello (1997) surveyed the literature on the impact of FDI on growth in developing countries. The paper concludes that the impact of FDI on the growth of the host country depends on the capacity of the domestic firms to absorb the technology transfer and spillovers due to FDI inflows. Shan et al. (1997) examine the causal link between FDI and economic growth of China over 1985–1996 by using VAR model and Causality test by Toda and Yamamoto (1995). The analysis of this paper shows a robust relationship between FDI and economic growth in both directions.

Reichert and Weinhold (2001) analyses the role of gross domestic investment (GDI) and control for openness to trade as well as the rate of inflation. To test the causal relationship between FDI and economic growth, the paper used data from 1971 to 1995 of 24 developing countries. This paper used Mixed Fixed and Random (MFR) estimator to deal with heterogeneous Panel data. The results of the causal relationship between FDI and growth are highly heterogeneous and differ from one country to the other. Homogeneity of assumptions for all countries should be avoided. Carkovic and Levine (2002) analysed data from 72 countries from 1960–1995 to make

panel data. The results are obtained by using Generalized-Method-of-Moments (GMM) estimators. The paper did not find any strong cause and effect relationship between FDI and economic growth. Chakroborty and Buckley et al. (2002) assess the relationship between growth FDI inflows and economic growth of China over a period of 1989–1999. At the provincial level, FDI positively affects the growth in economically strong and highly competitive provinces. However, the analysis of the paper found that the conditions of the host country have huge impact on the growth of the countries.

Dritsaki (2004) analyses the relationship between trade, FDI and economic growth using time series data over the period 1960–2002 for Greece. The causality test confirms a bidirectional causal relationship between them and cointegration analysis confirms long-run equilibrium relationship. Chowdhury & Mavrotas (2005) tested the causality between FDI and economic growth. To test causality between variables, Toda-Yamamoto test is applied to the data covering the period 1969–2000 for Chile, Malaysia and Thailand. The results for all the countries have been different. In the case of Chile, there run causal relationship from economic growth to FDI only, while the causal relationship existed between the two variables in both directions for both Malaysia and Thailand.

Hansen and Rand (2005) examine causal relationships between FDI and GDP of 31 developing countries of 31 years. The results confirm that FDI causes growth. In addition, using a model for GDP and FDI as a fraction of gross capital formation (GCF), the paper confirms long-run effects from FDI to GDP. Duasa (2007) does not find a causal relationship between FDI and economic growth over the period using Yamamoto's (1995) non-causality test. Applying Generalized Autoregressive Conditional heteroskedasticity (GARCH) model to test volatility between the two variables supports the hypothesis that FDI provides stability to economic growth and vice versa.

Mun et al. (2008) studied the relationship between FDI and economic growth in Malaysia for the period 1970–2005 by applying Ordinary least square (OLS) regression analysis. The analysis suggest that the relationship between GDP and foreign direct investment direct and noteworthy. Nunnenkamp (2008) assess the impact of FDI in post-reform India promotes economic growth.

They used industry-specific FDI and output data to test for Granger causality and cointegration tests. The results show that the results are different in different sectors. FDI and output effects are seen in the manufacturing and service sector. While no causal relation was found in the primary sector.

The relationship between FDI and economic growth was examined for five ASEAN countries only for a period of 1970–2007. The countries were Indonesia, Malaysia, Philippines, Singapore and Thailand. The results confirm bidirectional causality between the two variables at the panel and individual level except for Malaysia (Pradhan, 2009). The empirical analysis gives evidence that the variables are cointegrated, i.e., there exist long-run relationship between them. Sridharan et al. (2009) analyse the direction of the causal relationship between economic growth and FDI. Industrial Production Index (IPI) is used to measure of economic growth of each country. The Vector Error Correction Model (VECM) model confirms the existence of long-run relationship between FDI and economic growth. The results confirm that there exist bidirectional causality between economic growth and FDI for Brazil, Russia and South Africa and FDI cause economic Growth for India and China.

Iqbal et al. (2010) examine the causality relationship between FDI, trade and economic growth in Pakistan over the period 1998 to 2009. On applying the VAR model, the results confirm the long-run relationship between variables and exhibit the bidirectional causal relationship between factors with economic growth. The results confirm that FDI enhances trade growth in Pakistan. ESSO (2010) re-examined the relationship between foreign direct investment and economic growth in the case of ten Sub-Saharan African countries. By using Pesaran et al. (2001) approach to cointegration and Toda and Yamamoto (1995) procedure for the non-causality test, the paper reports that there is a positive and long-run relationship between FDI and economic growth. However, the causal relationship between them is different for different countries. In Angola, Cote d'Ivoire and Kenya, FDI cause economic growth while in Liberia and South Africa, economic growth causes FDI.

Abbas et al. (2011) investigated the impact of foreign direct investment (FDI) on GDP of SAARC countries by using multiple regressions for the period of 2001 to

2010. The results confirm that the correlation between GDP and FDI were positive and considerable for all the countries except the Maldives. The paper concludes that economic growth increases with the increase in FDI. Akpan et al. (2014) examine the determinants of FDI in BRICS and MINT economies using panel data over the period 2001–2011. The paper estimates the model on determinants of FDI for BRICS only, MINT only, and BRICS and MINT combined. The results of this paper show that there are other significant factors which play an important role in increasing the amount of FDI are market size, infrastructure availability and trade openness. The endowment of national resources wasn't that significant.

Rahaman & Chakroborty (2015) examines the causal relationship between FDI and economic growth measured in terms of GDP for Bangladesh. The cointegration test confirms the existence of long-run association between the two variables and the Granger causality test provides strong evidence of FDI causes economic growth and not vice versa. Marobhe (2015) found out a causal relationship and cointegration between FDI and economic growth in Tanzania for a period 1970–2014. The results show that there is a unidirectional causality between the two variables i.e., FDI cause economic growth in Tanzania long term association between them.

The consensus in the literature seems ambiguous. The above earlier findings give the evidence that the nexus between foreign direct investment and economic growth varies from country to country and with different periods in a country. The article aims to investigate causality between FDI and economic growth for four developing economies based on a theoretical framework and estimation methods used in the previous studies.

## Methodology

GDP and FDI data are measured in current dollars and on an annual basis. The paper focuses on data from four countries, namely, Mexico, Indonesia, Nigeria and Turkey. The data has been obtained from the World Bank database source. The sample period runs from 1981 to 2015. In exploring the causal relationship between GDP and FDI, cointegration and Error Correction models are applied.

Before applying these estimation methods, statistical properties of the variables in the time series data are examined.

The order of integration for the two variables of the four countries has been found by using Augmented Dickey-Fuller (ADF) test. The result of the ADF test is provided in Table 1. Then, the cointegration properties of the variables are examined. If the variables are found to be unit root and yet their linear combination may be stationary, they are said to be cointegrated. To test for cointegration, this paper applies two most commonly used tests- the residual-based test of Engle and Granger (EG), (1987) and the VAR based Johansen (1988) test.

If the two variables are found to be cointegrated, there must be at least one-way the causal relationship between the two variables. In addition, the variables have long term relationship with each other and that their deviations from the long-run equilibrium path are corrected. On applying EG or AEG test, if the two variables are cointegrated, there exists long-run equilibrium and there exists a valid error correction mechanism. The null hypothesis that FDI does not Granger causes GDP. On the other hand, if GDP and FDI are not cointegrated, there might be short-term relationship and short-run equilibrium. In other words, there may exist a short-run relationship between the two variables.

## Empirical Results

For time-series data, it is very important for the variables to be stationary or unit root. Thus, to test the stationary condition Augmented Dickey-Fuller unit root test has been used. For the panel data, Levin-Lin (LLC) Tests and Im-Pesaran-Shin (IPS) Test have been used because the IPS test also can be used in association with any parametric unit-root test. However, the condition is that the panel data should be balanced and secondly, all the t-statistics for the unit-root should have the same mean and variance, i.e., they are identically distributed in every cross-section.

The ADF unit root test in Table 1 indicates that both series of GDP and FDI at the level are unit root for all the four countries, Mexico, Indonesia, Nigeria and Turkey. Hence, the values have been found to be significant and hence the behaviour of variables is non-stationary. To make them stationary, the first difference of the variables has been taken. The first difference of both variables rejects the null hypothesis that the variables are unit root at the 5 percent significance level. That is, GDP and FDI series are not unit root (stationary) at first difference, I (1) for all four countries.

**Table 1: Results of Unit Root Test**

Countries	GDP		FDI		Conclusion
	Level	First Difference	Level	First Difference	
Indonesia	3.75	-3.20**	3.38	-4.77*	I(1)
Mexico	1.25	-3.83**	0.38	-3.87*	I(1)
Nigeria	3.55	-3.60**	-3.38	-6.65*	I(1)
Turkey	3.055	-4.12*	-0.91	-4.94*	I(1)
Panel Unit Root test					
LLC	5.70	-4.48*	1.75	-6.62*	I(1)
IPS	7.50	-3.47*	2.25	-8.39*	I(1)

Note: GDP: Economic Growth; FDI: Foreign Direct Investment; LLC: LLC statistics; IPS: IPS statistics; \*: indicates the variable significant at 1% significance level. \*\*: indicates the variable significant at 5% significance level.

**Table 2: Results of Johansen's Cointegration Test**

Countries	Null Hypothesis	Trace Statistics	MEV Statistics
Indonesia	None	11.28 [0.19]	6.19 [0.58]
	At most 1	5.08 [0.02]	5.08 [0.02]
Mexico	None	15.49[0.01]	18.52[0.01]
	At most 1	3.84[0.25]	3.84[0.25]

Countries	Null Hypothesis	Trace Statistics	MEV Statistics
Nigeria	None	13.44[0.99]	12.58[0.09]
	At most 1	0.86[0.35]	0.86[0.35]
Turkey	None	19.90[0.01]	11.23[0.14]
	At most 1	8.67[0.0032]	8.67[0.0032]

Note: Parentheses indicate the probability of significance.

Two or more time series are said to be cointegrated if they have a common stochastic drift. Time series data exhibit a time trend and are therefore generally non-stationary; i.e., the means, variances, and covariances that are not consistent over time. If OLS (Ordinary Least Square) or GLS (Generalized Least Square) is applied to non-stationary data, it results in spurious regressions. Engle and Granger Panel cointegration technique in Table 2 addresses this issue by defining common long-run relationships between a set of variables from individual members of a panel (Engle and Granger, 1987).

The Panel cointegration technique allows pooling of information from individual series of panel data. Pedroni test provides seven different statistics for testing unit roots in the residuals of the variables having common long-run relationship. The first four statistics of Pedroni test are known as panel cointegration statistics; the remaining three are referred to as group mean panel cointegration statistics. Table 3 reports the results of the Pedroni's test which check the possibility of the existence of a long-run relationship between the variables, FDI and economic growth.

It is observed in the test results that five out of the seven statistics are significant which implies that the null hypothesis of no cointegration between variables at 5 percent level of significance is rejected. This implies the existence of a long-run co-integration of FDI and Economic Growth at Panel level. However, these results do indicate prospects of the causal relationship between FDI and economic growth.

**Table 3: Results of Pedroni's Panel Cointegration Test**

Test Statistics	Calculated Value	Probability
Panel v- statistic	0.238235	0.4058
Panel rho-Statistic	-2.978199	0.0014
Panel PP-Statistic	-2.693772	0.0035
Panel ADF-Statistic	-1.519134	0.0644
Group rho-Statistic	-1.902604	0.0285
Group PP-Statistic	-1.665753	0.0479
Group ADF-Statistic	-0.265003	0.3955

The null hypothesis of Engel Granger causality test is that there is no cointegration between variables which is rejected at a significant level of less or equal to 5%. This is supported by both EG residual and Johansen trace tests. The establishment of cointegration between the two series suggests that their exists long-run relation or equilibrium for three countries, Nigeria, Mexico and Turkey. For Indonesia, Johansen trace test indicates that two series (GDP and FDI) are not cointegrated. The null hypothesis that there is no cointegration cannot be rejected. Thus, the result suggests that there exist long-run relationship or equilibrium for Mexico, Nigeria and Turkey, except Indonesia.

Even if variables are not cointegrated, there is a possibility that the variables may have a short-run relationship between them. Thus, the absence of causality or relationship in the short run between the variables in these countries cannot be ignored. The variables are said to be cointegrated if they move together in the short run. If GDP and FDI do not move together, in the long run, they are not cointegrated but it is possible that they may affect each other in the short run. Granger causality test is the most appropriate tool to determine the causal relation between variables.

Vector Error Correction Model (VECM) is applied where GDP and FDI are cointegrated. This helps in determining the causal relation between them. However, if they are not cointegrated unrestricted VAR is the appropriate tool. In Table 4, the results of the Granger causality test have been shown. At lag 2, there exists a causal relationship from GDP to FDI and not vice-versa. In Indonesia, FDI and GDP are not found to be cointegrated. These variables do not affect each other in the short run and long run. The results based on VECM between these variables shows that GDP cause FDI in Mexico, Nigeria and Turkey. Thus, there exists a uni-directional relationship between these variables for these countries. F-statistics for these three countries indicates that the null hypothesis that GDP does not Granger cause FDI can be rejected at a 5 percent level of significance.

Three countries exhibit unidirectional and positive short-run causal relationship from GDP to FDI, namely, Mexico, Nigeria, and Turkey. Thus, GDP growth provides market and attracts foreign investment in these countries.

**Table 4: Granger Causality Test for Short-Run Causality Based on ECM**

<i>Granger Test for Short Run Causality</i>		
<i>Countries</i>	<i>Null Hypothesis</i>	<i>F-statistics</i>
Mexico	GDP does not Granger cause FDI	12.1
	FDI does not Granger cause GDP	0.89
Indonesia	GDP does not Granger cause FDI	2.17
	FDI does not Granger cause GDP	0.92
Nigeria	GDP does not Granger cause FDI	5.03
	FDI does not Granger cause GDP	0.98
Turkey	GDP does not Granger cause FDI	5.77
	FDI does not Granger cause GDP	1.76

**Table 5: Granger Causality Test based on VECM (Vector Error Correction Model)**

<i>Countries</i>	<i>Variables</i>	<i>GDP</i>	<i>FDI</i>
Indonesia	GDP	----	✗
	FDI	✗	----
Mexico	GDP	----	✓
	FDI	✗	----
Nigeria	GDP	----	✓
	FDI	✗	----
Turkey	GDP	----	✓
	FDI	✗	----
Panel Granger			
Causality	GDP	----	✓
	FDI	✗	---

Note: ✓ : Represents the presence of causality; ✗ : Represents the absence of causality

## Concluding Remarks

The present paper tests the direction of causality between FDI and growth for MINT economies in the developing world, namely Mexico, Indonesia, Nigeria and Turkey. All these countries have different macroeconomic history, policies and growth patterns over the period 1981–2015. The empirical findings based on the Engel Granger causality test seem to suggest that it is GDP that causes

FDI in Mexico, Nigeria and Turkey and not *vice versa*. No bi-directional causality between GDP and FDI was found in any of these countries.

The above findings have provided an understanding of the direction of causality between the two variables which is important for formulating policies. In view of findings, the traditional view which suggests that the direction of causality runs from FDI to economic growth is not valid in the case of MINT economies. While in the case of Mexico, Nigeria and Turkey, the causality is uni-directional from GDP to FDI. Consequently, careful attention is required while framing policy guidelines which emphasise the FDI causes economic growth and stability in developing countries. Results also suggest that the causality between the two variables is also country-specific and hence, there is a need for more individual country studies.

## References

- Agrawal, G. (2015). Foreign direct investment and economic growth in BRICS economies: A panel data analysis. *Journal of Economics, Business and Management*, 3(4), 421–424.
- Alici, A. A., & Ucal, M. Ş. (2003, September). *Foreign direct investment, exports and output growth of Turkey: Causality analysis*. In European Trade Study Group (ETSG) Fifth Annual Conference, Madrid (pp. 11–13).
- Buckley, P. J., Clegg, J., Wang, C., & Cross, A. R. (2002). FDI, regional differences and economic growth: Panel data evidence from China. *Transnational Corporations*, 11(1), 1–28.
- Blomström, M., & Sjöholm, F. (1999). Technology transfer and spillovers: Does local participation with multinationals matter?. *European Economic Review*, 43(4), 915–923.
- Belloumi, M. (2014). The relationship between trade, FDI and economic growth in Tunisia: An application of the autoregressive distributed lag model. *Economic Systems*, 38(2), 269–287.
- Chakraborty, C., & Nunnenkamp, P. (2008). Economic reforms, FDI, and economic growth in India: A sector level analysis. *World Development*, 36(7), 1192–1212.
- Chowdhury, A., & Mavrotas, G. (2005). *FDI and growth: A causal relationship* (No. 2005/25). Research Paper, UNU-WIDER, United Nations University (UNU).

- De Mello Jr, L. R. (1997). Foreign direct investment in developing countries and growth: A selective survey. *The Journal of Development Studies*, 34(1), 1–34.
- Dritsaki, M., Dritsaki, C., & Adamopoulos, A. (2004). A causal relationship between trade, foreign direct investment and economic growth for Greece. *American Journal of Applied Sciences*, 1(3), 230–235.
- Duasa, J. (2007). Malaysian foreign direct investment and growth: Does stability matter?. *Journal of Economic Cooperation Among Islamic Countries*, 28(2), 83–98.
- Esso, J. L. (2010). Long-run relationship and causality between causality between foreign direct investment and growth: Evidence from ten African countries. *International Journal of Economics and Finance*, 2(2), 168–177.
- Frimpong, J. M., & Oteng-Abayie, E. F. (2006). Bivariate causality analysis between FDI inflows and economic growth in Ghana. Retrieved from <https://mpra.ub.uni-muenchen.de/351/>
- Hansen, H., & Rand, J. (2006). On the causal links between FDI and growth in developing countries. *The World Economy*, 29(1), 21–41.
- Har, W. M., Teo, K. L., & Yee, K. M. (2008). FDI and economic growth relationship: An empirical study on Malaysia. *International Business Research*, 1(2), 11–18.
- Iqbal, M. S., Shaikh, F. M., & Shar, A. H. (2010). Causality relationship between foreign direct investment, trade and economic growth in Pakistan. *Asian Social Science*, 6(9), 82–89.
- Karimi, M. S., & Yusop, Z. (2009). FDI and economic growth in Malaysia. Retrieved from <https://mpra.ub.uni-muenchen.de/14999/>
- Lee, C. C. (2005). Energy consumption and GDP in developing countries: A cointegrated panel analysis. *Energy Economics*, 27(3), 415–427.
- Marobhe, M. (2015). Do foreign direct investment inflows cause economic growth in Tanzania? The granger causality test approach. *Journal of Economics and Sustainable Development*, 6(24), 144–150.
- Nair-Reichert, U., & Weinhold, D. (2001). Causality tests for cross-country panels: A new look at FDI and economic growth in developing countries. *Oxford Bulletin of Economics and Statistics*, 63(2), 153–171.
- Ozturk, I., & Kalyoncu, H. (2007). Foreign direct investment and growth: An empirical investigation based on cross-country comparison. Retrieved from <https://mpra.ub.uni-muenchen.de/9636/>
- Ozturk, I. (2007). Foreign direct investment-growth nexus: A review of the recent literature. *International Journal of Applied Econometrics and Quantitative Studies*, 4(2), 79–98.
- Pradhan, R. P. (2009). The FDI-led-growth hypothesis in ASEAN-5 countries: Evidence from cointegrated panel analysis. *International Journal of Business and Management*, 4(12), 153–164.
- Qaiser Abbas, S. A., Nasir, A. S., Ullah, H. A., & Muhammad, A. N. (2011). Impact of foreign direct investment on gross domestic product: A case of SAARC countries. *Global Journal of Management and Business Research*, 11(8), 34–40.
- Rahaman, A., & Chakraborty, S. (2015). Effects of foreign direct investment on GDP: Empirical evidence from developing country. *Advances in Economics and Business*, 3(12), 587–592.
- Samad, A. (2009). Does FDI Cause Economic Growth? Evidence from South-East Asia and Latin America. Working Paper 01–09.
- Shan, J., Tian, G. G., & Sun, F. (1997). The FDI-led growth hypothesis: Further econometric evidence from China. Economic Division Working Papers 97/2, NSDS, Australia.
- Tang, S., Selvanathan, E. A., & Selvanathan, S. (2008). Foreign direct investment, domestic investment and economic growth in China: A time series analysis. *The World Economy*, 31(10), 1292–1309.
- Vijayakumar, N. (2009). Causal relationship between foreign direct investment and growth: Evidence from BRICS countries. *International Business Research*, 2(4), 198–203.

# Recommending Movies Using User-Based and Item-Based Collaborative Filtering on R Platform

Chirag Malik\*

## Abstract

Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of a large amount of dynamically generated information according to user's preferences, interest or observed behaviour about an item. Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile. Recommendation algorithms are best known for their use on e-commerce websites, where they use inputs about a customer's interests to generate a list of recommended items. A new emerging sector in India, online movie viewership and subscription of movies, demands an expert and highly technical skill to understand the movie preferences of the viewers spread across India. This paper uses two techniques (run on R platform), user-based collaborative filtering and item-based collaborative filtering, to understand the preferences of people (without giving any reason to it) and recommending mechanism was solely based upon user-user similarity matrix and item-item similarity matrix. A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The robustness of the two techniques is also compared and explained.

**Keywords:** User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Recommendation Engine

## Introduction

The explosive growth in the amount of available digital information and the number of visitors to the internet have created a potential challenge of information

overload which hinders timely access to items of interest on the internet. Information retrieval systems, such as Google, DevilFinder and Altavista, have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for recommender systems more than ever before. Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest or observed behaviour about the item. Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile.

Recommender systems are beneficial to both service providers and users. They reduce transaction costs of finding and selecting items in an online shopping environment. Recommendation systems have also proved to improve the decision-making process and quality. In e-commerce setting, recommender systems increase revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender systems support users by allowing them to move beyond catalogue searches. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasised.

Recommendation algorithms are best known for their use on e-commerce websites, where they use input about a customer's interests to generate a list of recommended items. A new emerging sector in India, online movie viewership and subscription of movies demand an expert and highly technical skill to understand the movie preferences of the viewers spread across India. This

\* Associate Professor, BML Munjal University, Gurgaon, Haryana, India. Email: [chiragmalik@yahoo.com](mailto:chiragmalik@yahoo.com)

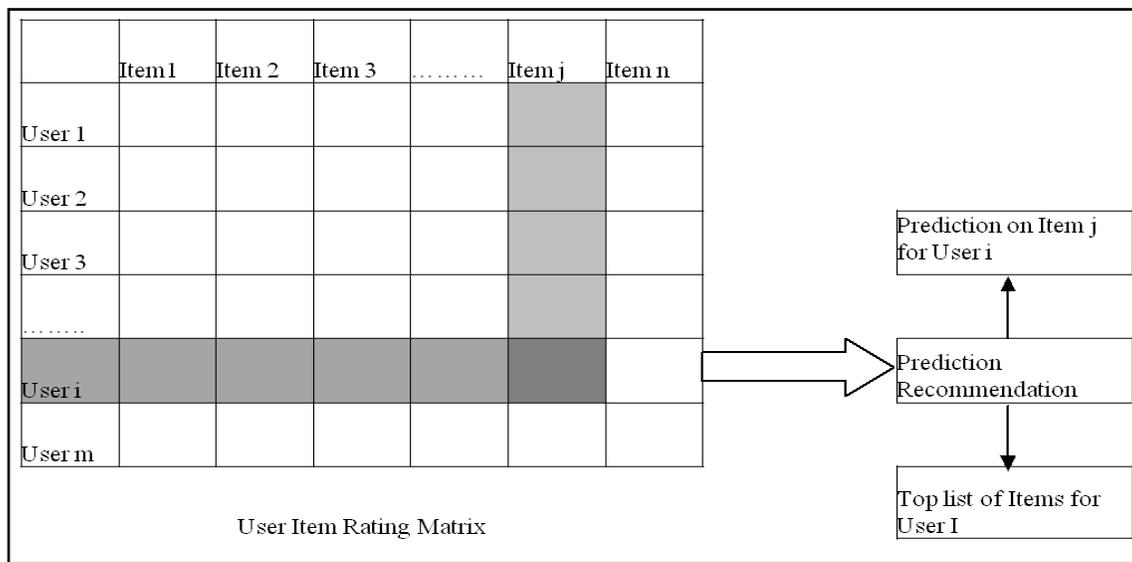
paper uses two techniques (run on R platform), user-based collaborative filtering and item-based collaborative filtering, to understand the preferences of people (without giving any reason to it) and recommending mechanism was solely based upon user-user similarity matrix and item-item similarity matrix. A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The robustness of the two techniques is also compared and explained.

### Literature Review and Research Problem

Previous personalised movie recommendation system focus only on viewers’ historical viewing records or demographic data. This study proposes a new recommending mechanism from a user-oriented perspective. New recommender system, which is

user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF), is free from AIMED (Activities, Interests, Moods, Experiences, and Demographic information) characteristics. This paper aims at using R platform (free resource available for corporate and consultants) for recommending movies on online movie subscription websites using UBCF and IBCF.

Digital TV allows people to access numerous and varying kinds of TV programmes without space-time constraints. Although electronic programme guides (EPGs) can increase the accessibility of TV programmes, they often overload users by providing too many programme options. Therefore, the usefulness of EPG is an important issue in DTV design. To enhance the usefulness of EPGs, researchers have begun to develop TV recommendation



**Fig. 1: Collaborative Filtering Process**

mechanisms (i.e., embedded assistance) to help users access the TV programmes of their choice. Recommendation mechanisms operate by collecting viewing history data over a period of time, gleaning user viewing preferences from the data, mapping user preferences for TV programme attributes, filtering non-interesting programmes, and finally recommending an appropriate programme (Frank et al., 1980). In general, these TV programme recommendation mechanisms

can be divided into two types-content-based filtering and collaborative filtering. Content-based filtering recommendation operates by automatically tracking each user’s TV viewing patterns. Programme descriptors, such as the programme category, the name of the actor or actress, programme keywords, viewing time period, and so on, and a set of similarity metrics of the user profile are collected. Using these data, the recommender builds a user profile that represents the viewing preferences of

each

The recommendation system stores massive amounts of viewer preference data in a database. The system then recommends programmes to users based on programme ratings given by people with similar profiles (Ansari et al., 2000; Ang, 1996). This system effectively forms a viewing community whose members share similar viewing preferences or similar viewing habits. In this virtual community, viewers can share information or recommend programmes to other viewers. Examples of this kind of recommender include the TV-scout system developed the peer-to-peer (p2p) recommender system proposed by Wang et al. (2006). Although both types of recommenders can assist TV viewers with programme selection, they both have some weaknesses. For example, a content-based filtering recommender only follows data from users' viewing histories and their viewing behaviour. Therefore, it cannot extend to other types of programmes or provide new programmes to viewers. This problem causes over-specialization in recommended programmes. On the other hand, a collaborative filtering recommender depends on other viewers' suggestion data, without which the mechanism may not make good recommendations, i.e., AIMED - Activity, Interest, Mood, Experience, and Demographic information of the user (Hsu et al., 2006). The AIMED recommender is a hybrid recommendation system based on the content-based and collaborative filtering methods. It makes programme suggestions using not only viewers' personal profile prediction criteria such as demographics, lifestyle and explicit preference information, from conventional recommendation models, but also inputs users' viewing contextual information, such as mood and viewing behaviour, into a prediction model (Burke, 2002; Alspector, 1998). By doing so, it has two advantages. Firstly, the recommender can avoid the weaknesses of conventional content-based filtering and collaborative filtering while taking advantage of their strengths. At the initial use, programme recommendation can be inferred from the viewer group's preference when the information of the user's viewing history is not available yet (Gena et al., 2001). As the system gathers more information about the user's viewing context, adding viewing context into the prediction of programme preference can fine-tune programme recommendation to match the personal preference of an individual user (Basu et al., 1998; Perse, 1998). Secondly,

a well-trained AIMED recommender could suggest a suitable programme to users by considering both long-term programme preferences and the particular viewing context. Therefore, an AIMED recommender is able to adapt to the viewing context (Balabanovic et al., 1997). But, given the data of AIMED characteristics, the process of running the data and coming out with recommendation is a time-consuming and complex process. Another way to process the data only on the basis of item similarity matrix and user similarity matrix. This method doesn't demand any prior information of AIMED and solely works on the similarity matrix mechanism.

Two most important similarity measures are correlation-based and cosine-based. Pearson correlation coefficient is used to measure the extent to which two variables linearly relate with each other and is defined as:

$$\text{sim}(a,b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}} \quad \dots(1)$$

where;

a, b; users

ra, p: rating of user a for item p

set of items, rated by both a and b possible similarity values between -1 to 1.

The Common Prediction Functions

$$\text{pred}(a,p) = \bar{r}_a + \frac{\sum_{b \in N} \text{sim}(a,b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} \text{sim}(a,b)} \quad \dots(2)$$

The second method is the cosine similarity measure. It produces better results in item-based collaborative filtering. In cosine similarity measures, ratings are seen as vectors in n-dimensional space, and the similarity is calculated based on the angle between the vectors.

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \quad \dots(3)$$

To compare the recommendation outcome of UBCF method and IBCF method, the deviation between predicted rating and actual ratings (also known as prediction accuracy) have been calculated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - r_i| \quad \dots(4)$$

and

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - r_i)^2} \quad \dots(5)$$

and the recommender engine using R was applied. The techniques used were user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF). To measure the robustness of the method, a separate set of data was used and a classification table was made.

### Findings of the Research

It has been found the user-based recommendation was more robust as it has less error (Root Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MEA) or more accuracy percentage in the classification table. However, both methods are equally important to recommend movies to users.

### Objective of the Research

To run the movie recommendation engine using recommendation algorithm on R and recommend movies for the active and non-active users and compare the robustness of user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF).

### Research Methodology

A dataset of 563 movies and 9,985 consumers from Amazon prime has been taken for recommending movies for people who have not watched a particular set of movies. The sample dataset was taken with the due approval and permission of the Amazon India authorities

**Table 1: Comparative Values of RMSE, MSE and MAE for UBCF and IBCF Methods**

	RMSE	MSE	MAE
UBCF	0.341864916	0.116871621	0.237863157
IBCF	0.476725497	0.2272672	0.255404009

**Table 2: Top 10 Scores Using UBCF and IBCF**

TOP 10 Score User Based Rec. Filtering		TOP 10 Score Item Based Rec. Filtering	
The Situation Room with Wolf Blitzer	0.111882811	The Situation Room with Wolf Blitzer	0.064647966
NBC Nightly News	0.077134284	NBC Nightly News	0.267689573
Dancing with the Stars	0.112713567	Dancing with the Stars	0.333699171
The Colbert Report	0.077134284	The Colbert Report	0.017279588
Larry King Live	0.077134284	Larry King Live	0.061431732
Everybody Loves Raymond	0.153878714	Everybody Loves Raymond	0.305660932
NHL Hockey	0.113013825	NHL Hockey	0.223449589
Campbell Brown: No Bias, No Bull	0.077134284	Campbell Brown: No Bias, No Bull	0.172957837
NBA Basketball	0.077134284	NBA Basketball	0.184547026
Two and a Half Men	0.077134284	Two and a Half Men	0.298187223

**Table 3: Top 10 Recommendations Using UBCF and IBCF**

Sr. No.	TOP 10 Recommendation Based Upon	TOP 10 Recommendation Based Upon
	UBCF	IBCF
1	Everybody Loves Raymond	Dancing with the Stars
2	NHL Hockey	Everybody Loves Raymond
3	Dancing with the Stars	Two and a Half Men
4	The Situation Room with Wolf Blitzer	NBC Nightly News
5	NBC Nightly News	NHL Hockey
6	The Colbert Report	NBA Basketball
7	Larry King Live	Campbell Brown: No Bias, No Bull
8	Campbell Brown: No Bias, No Bull	The Situation Room with Wolf Blitzer
9	NBA Basketball	Larry King Live
10	Two and a Half Men	The Colbert Report

## Implications of the Study

As online shopping and movies subscription is an emerging sector in India, it has become all the more important to understand the subscription behaviour and purchase patterns of the consumers. This method will enable the marketers to understand the purchase patterns of online consumers and recommend for non-buyers and existing buyers. This method will also help match the user and item associations.

## References

- Alspector, J., Kolez, A., & Karunanithi, N. (1998). *Comparing feature-based and clique-based user models for movie selection*. In Proceedings of the Third ACM Conference on Digital Libraries, Pittsburgh PA, ACM Press, New York.
- Ang, I. (1996). *Living room wars*. Routledge, London.
- Ansari, A., Essegai, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research*, 37(3), 363–375.
- Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Difino, A., & Negro, B. (2004). User modeling and recommendation techniques for personalized electronic program guides. In L. Ardisson, A. Kobsa & M. Maybury (Eds.), *Personalized Digital Television: Targeting Programs to Individual Viewers* (pp. 3–26). Kluwer Academic Publishers, Dordrecht.
- Balabanovic, M., & Shoham, Y. (1997). FAB: Content-based collaborative recommender. *Communications of the ACM*, 40(3), 66–72.
- Basu, C., Hirsh, H., & Cohen, W. (1998). Recommendation as classification: Using social and content-based information in recommendation. In *Recommender System Workshop*, 98, 11–15.
- Baudisch, P., Brueckner, L., & Scout, T. V. (2002). *Guiding Users from Printed TV Program Guides to Personalized TV Recommendation*. In Proceedings of the 2<sup>nd</sup> Workshop on Personalization in Future TV, (pp. 157–166). Malaga, Spain.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User Adapted Interaction*, 12(4), 331–370.
- Frank, R., & Greenberg, M. (1980). *The public's use of television: Who watches what and why*. Sage, Beverly Hills.
- Gena, C., & Ardisson, L. (2001). On the Construction of TV Viewer Stereotypes Starting from Lifestyle Surveys. In Workshop on Personalization in Future TV, Sonthofen, Germany.
- Hsu, H. S., Wen, M. H., & Lee, C. H. (2006). *An Activity-Oriented Approach to Designing a User Interface for Digital Television*. In Proceedings of the 4<sup>th</sup> Euro iTV Conference, Athens, Greece, (pp. 83–90).
- Wang, J., Polwelse, J., Fokker, J., & Reinders, M. J. T. (2006). *Personalization of a Peer-to-Peer Television System*. In Proceedings of the 4<sup>th</sup> EuroITV Conference, Athens, Greece, (pp. 147–155).

# 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.

**Secondary 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 2020

	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

Copyright – ©2019 Publishing India Group. All Rights Reserved. Neither this publication nor any part of it may be reproduced, stored or transmitted in any form or by any means without prior permission in writing from copyright holder. 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](http://www.manuscript.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](mailto:info@publishingindia.com)

Website: [www.publishingindia.com](http://www.publishingindia.com)



Copyright 2019. Publishing India Group.