ISBN 978-969-8858-29-2

PROCEEDINGS

Vol. 37

20th International Conference on Statistical Sciences

March 9-11, 2023

Theme: **The Role of Data Science in Meeting the Challenges of Environmental Issues**

Jointly organized by

Islamic Countries Society of Statistical Sciences

Imperial College of Business Studies

Pak Institute of Statistical Training and Research

SPONSORS

Pakistan Bureau of Statistics, Government of Pakistan

FutureNow Technologies

murkez IT & Business Consulting

Pakistan Cargo Services

"All papers published in the

PROCEEDINGS

were accepted after formal peer review

by the experts in the relevant field.

Editors:

- **Prof. Dr. Muhammad Hanif**
  National College of Business Administration & Economics Pakistan.
- **Professor Dr. Shahjahan Khan**
  VC, Asian University of Bangladesh and Emeritus Professor, University of Southern Queensland, Australia

Published by: **ISOSS, Lahore, Pakistan.**

Financial support of

Pakistan Science Foundation,

Islamabad

is highly appreciated.

CONTENTS

In the name of Allah, the Most Beneficent, the Most Merciful. All praise is to Allah SWT and peace and blessings to His messenger, Muhammad SWS.

FOREWORD

I am delighted to write this foreword for the Proceedings of the 20th International Conference on Statistical Sciences jointly organized by the Islamic Countries Society of Statistical Sciences (ISOSS) and Pak Institute of Statistical Training and Research (PISTAR) in collaboration with the Imperial College of Business Studies (ICBS), Lahore, Pakistan.

The ISOSS was established in 1988 during the first Islamic Countries Conference on Statistical Sciences (ICCS-1) held in Lahore, Pakistan led by late Professor Dr. Munir Ahmad, may Allah be pleased with him, and his esteemed colleagues. The main purpose of this conference and ISOSS was to promote statistics as an essential tool for planning, development, data analysis, and enhance engagement in statistical research and derive benefit from its wide-ranging applications, as well as provide an international platform to share outcomes of statistical research and build professional network among all statisticians with special focus on those from Islamic Countries.

As a very close associate and fellow traveller of Dr. Munir Ahmad, I find it pleasing that a new generation of statisticians are becoming an active part of ISOSS and have shown very high level of commitment and activism to uphold and continue his noble mission. This is the fruit of his selfless personality, paramount scholarship, unwavering commitment to statistical research, supervising and mentoring a large number of students and young colleagues, and life-long hard work to deliver the best benefit that statistics could offer for his nation and the Ummah.

Statistics has made an enduring and significant contribution in many aspects of the modern society, science, technology, business, trade and commerce. Statistical literacy has become an essential skill for any knowledge society and citizens of contemporary world. Statistics will have no value to the people if statisticians keep themselves confined in their mathematical rigor and sophistication without engaging with the scientists and researchers of other disciplines to apply statistical methods and techniques in real life problem-solving in the way of making statistics beneficial for humanity. Appropriate and timely applications of statistics make statistics relevant to people.

At the age of data revolution statistics has been playing a significant role in the technological development causing rapid changes in the society and science. Increasingly, decision-makers, both in private and public sectors, are relying on evidence-based approach using synthesized results of data analyses via systematic review and meta-analysis.

To solve many complex problems faced by humanity at our time such as climate change, environmental issues, health care services etc. data plays an important role as a result statistics has become indispensable for scientist of many disciplines. Statisticians are required to join experts from other fields to help analyze data, interpret the results and suggest any policy implications of the new data-driven findings.

The conference committee received a total 144 abstracts of full-length research papers of which 101 papers were presented by the participating authors in four parallel sessions in each of the three days. The research areas of the presented papers include but not limited to meta-analysis, statistical modelling, statistical process control, control charts, biostatistics, artificial intelligence, econometrics, agriculture statistics, medical statistics, time series analysis, industrial statistics, machine learning, survey sampling, Bayesian analysis, experimental designs, regression analysis, categorical analysis, multilevel modeling, mathematical statistics, multivariate analysis, cryptography, deep learning and environmental and climate statistics.

In addition to organizing regular international statistics conferences, the on-going initiatives and efforts of ISOSS and PISTAR to present seminars and online workshops are providing much needed opportunities for the new generation of learners and researchers to get benefit from world class experts in statistics.

I thank and congratulate the organizers, sponsors, presenters, participants and volunteers of this conference for making it a success. It may be worth mentioning that success of this conference would inspire the ISOSS team to organise more events in the future.

May Allah continue to help and guide ISOSS and all the dedicated people who work for it.

Professor Dr. Shahjahan Khan
Patron of ISOSS
Vice Chancellor, Asian University of Bangladesh
Emeritus Professor, University of Southern Queensland, Australia
President of ISOSS, 2005-2011
Founding Chief Editor, Journal of Applied Probability and Statistics (JAPS)

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. ix-xvi

THE PROJECT METHOD AS A STRATEGY TO TEACH AND LEARN APPLIED STATISTICS

Manfred Borovcnik[1] and **Teresita E. Terán**[2]
[1] Universität Klagenfurt, Austria
Email: manfred.borovcnik@aau.at
[2] Universidad Nacional de Rosario, Argentina
Email: teresitateran52@gmail.com

ABSTRACT

The project method in teaching applied statistics is derived from didactic and statistical perspectives. Didactically, it focuses on efficient classroom organisation and learning situations. Statistically, it traces roots to applications driving statistical developments. Amid challenges of student engagement, especially in virtual classrooms during the COVID-19 pandemic, traditional methods resulted in high absenteeism and failure rates. Shifting the didactic analysis perspective to knowledge and its didactics, emphasising learning by doing, can be more effective. This paper explores a didactic path grounded on the "project method," offering a motivating approach for meaningful learning by applying statistical concepts in projects. Two project method approaches are presented, along with rationale, challenges, benefits, and drawbacks.

1. INTRODUCTION

The project method for teaching applied statistics has roots in didactic perspectives and the historical applications that influenced statistical development.

Didactic Situation

Modern teaching faces challenges with student engagement, especially in virtual classrooms during the COVID-19 pandemic. Traditional methods resulted in high absenteeism. The proposed didactic trajectory emphasises the "project method" for motivating, interesting, and meaningful learning, applying statistics to real-world projects.

Applied-Statistics Situation

Sir Ronald Aylmer Fisher's view inspires refining a project-based approach for teaching applied statistics. This paper explores variants of the project method, outlining key issues in its application.

Project Method

- We emphasise the project method's innovation for teaching statistics, with a refined research framework.
- We illustrate specific steps for project method implementation, providing examples from past projects.
- We explore key issues in teaching applied statistics, addressing research questions, project topics, and challenges in information acquisition.

- We discuss project evaluation, drawing conclusions about the value and challenges of the project method for teaching applied statistics.

For the future of teaching and learning Applied Statistics, the aspiration of Higher Education is to train active, reflective professionals. The current educational trend focuses on constructivist learning theory, emphasising autonomous learning (Moreno, Martín et. al. 2017). Teaching statistics, whether face-to-face, virtually, or in a mixed approach, involves project-based learning supported by applied statistics, integral to the overall development of students across diverse undergraduate courses globally.

2. THE APPLIED-STATISTICS SITUATION

Sir Ronald Aylmer Fisher's emphasis on applied statistics (e.g., Fisher, & Stock, 1915) inspired the authors to refine a project-based teaching approach. The project method is the natural choice for teaching applied statistics, supported by evidence from diverse university courses and practical experience. Seminars on Applied Statistics using the project method started in the 1990s, addressing challenges in industrial consulting (Borovcnik, 1993, 1995). Completed projects offer advantages, providing reasonable solutions, yet with some loss of authenticity due to data-generation phase cut-off.

Seminar Design:

- Different roles: Students worked collaboratively in teams of four, with the professor acting as the client.
- Modelling phase (systems analysis): Students delivered and revised their models based on client feedback.
- Data collection (production): Students designed methods for generating suitable data, considering challenges in matching systemic views and informed by experiences with data collection.
- Data analysis: Students explored data, checked assumptions, and sought methods to address questions from the initial systems analysis.
- Contextual decision: Teams presented results to the client and other groups, emphasising a holistic interpretation that included decisions and problems throughout the modelling process.

3. THE PROJECT METHOD

"The Project Method" emphasises collaborative learning, creating a dynamic learning community, even in virtual environments. Hiltz (1992) defines collaborative learning as a learning process that emphasises cooperative or group effort between teachers and students with active participation and interaction. Salinas (1999, 2004) defines educational strategy as a plan involving methods, means, and techniques to achieve learning objectives. The proposed path, based on the "project method," facilitates motivating and meaningful learning by applying statistics to real-world problems, aligning with Kilpatrick's (1967) project-based teaching (PBT). PBT fosters global knowledge development, learner-centric interests, and transformative interrelationships between students and teachers, shifting the focus from theoretical to practical issues.

Kilpatrick's "The project method" (1921) gained renewed interest in project-based learning (PBL) around 2000 (Beineke, 1998; Buck Institute, 2010; Nolan & Speed, 1999). PBL aligns with constructivist learning, incorporating educational psychology advancements and theories. Projects, driven by desires or needs, generate autonomy and responsibility, focusing on procedural, conceptual, and attitudinal contents (Álvarez, 1990; Belbin, 1970).

Implementing project-based learning involves four stages: intention, preparation, execution, and evaluation, fostering students' critical sense and individual development. The structured phases—planning, execution, monitoring, control, and evaluation—organise practical work within the project method.

4. KEY ISSUES OF THE PROJECT METHOD FOR TEACHING APPLIED STATISTICS

Investigation cycle for research questions

- Considerations beyond statistical analysis for formulating research questions.
- Illustrate challenges in finding a project topic and acquiring relevant information.
- Emphasise interpreting statistical findings within the contextual framework (Holmes, 1997).

Student skills and teacher role

- Address the high demand on students' argumentative skills.
- Acknowledge the evolving role of teachers as on-demand counsellors.
- Highlight the challenge of balancing direction to ensure comprehensive content without hindering self-learning (Batanero & Díaz, 2004, 2011).

Practical work orientation

- Orient students using the investigation cycle (PPDAC): Problem, Plan, Data, Analysis, Conclusions (Wild & Pfannkuch, 1999).
- Stress the importance of systems analysis in defining problems and planning data (Borovcnik, 1993).
- Emphasise that projects' success relies on anticipating and researching the context and problem of interest.

Benefits and requirements of the project method

- Emphasise the practical significance for applied statistics work (Holmes, 1997).
- Foster understanding of descriptive and inferential statistics with a focus on interpretation (Batanero & Díaz, 2011).
- Discuss the need for early consideration of later issues in the investigation phases (Borovcnik, 2018).
- Recognise the diverse skills required for project methods, including communication, argumentation, collaboration, and leadership (see also Kanji, 1979).
- Address shortcomings in argumentation skills through regular reflections and discussions (Borovcnik & Terán, 2021).

5. EVALUATION OF INDIVIDUAL PROJECTS AND OF THE PROJECT METHOD

Throughout a project, teachers must align with subject matter content, connecting activities and assessments with learning objectives (Salinas, 1999, 2004). The assessment of students has to take into account a much wider variety of achievement than in classical teaching (Holmes, 1997; Starkings, 1997). We will go into some details here, more can be seen from Borovcnik and Terán (2021), or Borovcnik (2018).

1. *Alignment and objectives:*
   - Teachers align activities and assessments with learning objectives.

2. *Assessment categories:*
   - Conceptual understanding.
   - Procedural knowledge.
   - Problem-solving.
   - Formulation and communication in context.
   - Statistical reasoning.

3. *Statistical reasoning:*
   - Essential for project method success.
   - Involves making sense of statistical information for decision-making based on data sets (Holmes, 1997).
   - Connects concepts and combines ideas about data and chance.

4. *Evaluation process:*
   - A dialogue for understanding and improvement (Sanjurjo & Vera, 1998; Santos Guerra, 2000).
   - Instruments capture assessments, motivations, interests, and interpretations (Holmes, 1997; Starkings, 1997).
   - Supports systematic reflection on teaching actions (Casanova, 1995).

5. *Project method success:*
   - Increased approval rates.
   - Areas for improvement include formulation and communication in statistical terminology (Borovcnik & Terán).
   - Resurgence in the 21st century under project-based learning (PBL) principles (Boondee, Kidrakarn, & Sa-Ngiamvibool, 2011).

6. *Definition of PBL (Buck Institute for Education):*
   - Systematic teaching method engaging students in inquiry-based learning.
   - Structured around complex and authentic questions with designed products and tasks (Markham, Larmer & Ravitz, 2003).

7. *Arguments for project work:*
   - Working modes (Álvarez, 1990).
   - General didactical framework (Godino, 2003; Godino, Batanero, & Font, 2007).
   - Action-research framework (Krainer et al., 1998).
   - Problem-solving (Anderson & Loynes, 1987).
   - Project-work methodology (NCTM, 2000; Graham, 1987).
   - Students' motivation (Batanero, 2001).
   - Applied-statistics learning (Borovcnik, 1995, 2018).
   - Guidelines for modelling (Borovcnik, 1993).

8. *Project-method focus:*
   - Collaborative work.
   - Responsibility.
   - Problem-solving.
   - Authentic problems.

It is worthy to note that the project method focuses on collaborative work – responsibility – problem solving – authentic problems.

6. CONCLUSIONS – THE FUTURE OF TEACHING AND LEARNING APPLIED STATISTICS

A transformative teacher profile is shaped by this method, emphasising self-observation, reflective teaching, facilitation, leadership in self-learning, motivation, and fostering intellectual skills. The teacher shifts from content communicator to a supportive guide, aiding students in their project-based problem-solving, method understanding, and contextual application.

Results from Argentinian and Austrian projects indicate positive impacts on immediate success and continued studies, overcoming challenges in areas like sampling and probability distributions. The project method enhances formative and cognitive processes, nurturing statistical reasoning skills and critical thinking.

Aligning with R.A. Fisher's emphasis on applications driving statistical theory, our teaching approach spans Applied Statistics, case studies, collaborative work, industrial projects, and simulation of real-world practice. The project method, rooted in thorough systems analysis and suitable statistical methods, guides students to a realistic understanding of Applied Statistics.

We summarise our endeavour to teach and learn applied statistics via the project method in the following diagram that shows the steps from Applied Statistics up to the Project Method, crossing the steps of case studies, collaborative work, industrial projects, simulating real practice, find the real questions in an applied problem by a thorough systems analysis of the context and answering these questions by the acquisition of suitable statistical methods within the framework of the project method – which leads the students groups back to a realistic view on Applied Statistics (Figure 1).

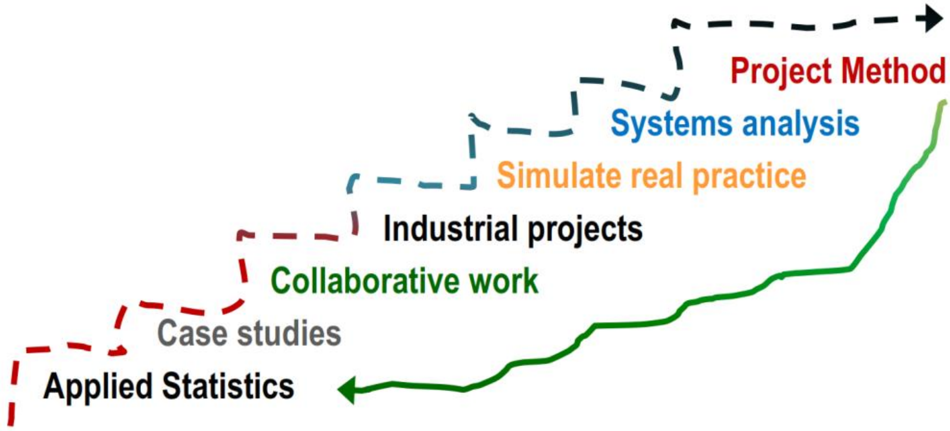


Figure 1: The Project-Method ladder to enhance the steps from Applied Statistics to the Project Method, which highlights the situation in Applied Statistics – which is best perceived by a collaborative project

In conclusion, our recommendation, based on years of exploratory experiments, is to orientate statistics education through the project method. This approach, applicable from introductory to master's level courses (starting even at elementary school, see Harris & Katz, 2001), becomes crucial in the era of evolving traditional statistics into data science, addressing genuinely open problems.

REFERENCES

1. Álvarez, S. (1990). *Proyectos integrados en el aula [Projects integrated in the classroom.]*. Buenos Aires: Kapelusz.
2. Anderson, C.W. and Loynes, R.M. (1987). *The teaching of practical statistics*. New York: Wiley.
3. Batanero, C. and Díaz, C. (2004). El papel de los proyectos en la enseñanza y aprendizaje de la estadística [The role of projects in statistics teaching and learning]. In J.P. Royo (ed.), *Aspectos didácticos de las matemáticas [Didactic aspects of Mathematics]*. Zaragoza: Instituto de Ciencias de la Educación, 125-164.
4. Batanero, C. and Díaz, C. (eds.) (2011). *Estadística con proyectos [Statistics with projects]*. Granada: Universidad de Granada. www.researchgate.net/publication/259974115_Estadistica_con_Proyectos
5. Beineke, J. (1998). *And there were giants in the land: The life of William Heard Kilpatrick*. New York: Peter Lang.
6. Belbin, M. (1970). *Roles de equipo en el trabajo [Team roles in the workplace]*. Belbin Associates (Marian Albaina). www.belbin.es/
7. Boondee, V., Kidrakarn, P. and Sa-Ngiamvibool, W. (2011). A learning and teaching model using project-based learning (PBL) on the web to promote cooperative learning. *European Journal of Social Science*, 21(3), 498-506.
8. Borovcnik, M. (1993). *Checklist for applied-statistics work*. Technical report. Graz: Technical University.

9. Borovcnik, M. (1995). Case studies in applied-statistics work – The role of systems analysis. *Biometrisches Seminar (ROeS)*. International Biometric Society, Rapperswil.
10. Borovcnik, M. (2018). *The potential of statistical consulting for statistics education*. Plenary lecture, Fifteenth International Students Statistics Colloquium, Istanbul. doi.org/10.13140/RG.2.2.35255.01447
11. Borovcnik, M. and Terán, T. (2021). How the project method contributes to the construction of meaning in statistics – also in a hybrid learning environment. In R. Helenius (ed.), *IASE Proceedings 2021. Statistics Education in the Era of Data Science*. The Hague: ISI and IASE. iase-web.org/Conference_Proceedings.php?p=2021_Statistics_Education_in_the_Era_of_Data_Science
12. Buck Institute for Education (2010). *What is PBL*? www.pblworks.org/what-is-pbl
13. Casanova, J. (1985). *El cambio inacabable [The never-ending change]*. Sevilla: Librería y Editorial Renacimiento
14. Fisher, R.A. and Stock, C.S. (1915). Cuénot on preadaptation: a criticism. *Eugenic Review*, 7(1), 46-61.
15. Godino, J.D. (2003). *Teoría de las funciones semióticas en didáctica de las matemáticas [Theory of semiotic functions in mathematics education]*. Granada: Universidad de Granada. www.ugr.es/~jgodino/funciones-semioticas/teoriafs.PDF
16. Godino, J.D., Batanero, C. and Font, V. (2007). The onto-semiotic approach to research in mathematics education. ZDM. *The International Journal on Mathematics Education*, 39(1-2), 127-135.
17. Graham, A. (1987). *Statistical investigations in the secondary school*. Cambridge: The Open University Centre for Mathematics Education.
18. Harris, J.H. and Katz, L.G. (2001). *Young investigators: The project approach in the early years*. New York: Columbia University, Teachers College Press.
19. Hiltz, S.R. (1992). The virtual classroom: Software for collaborative learning. In E. Barrett (ed.), *Sociomedia: Multimedia, hypermedia, and the social construction of knowledge*. Washington: The MIT Press, 347-368.
20. Holmes, P. (1997). Assessing project work by external examiners. In I. Gal & J.B. Garfield (eds.), *The assessment challenge in statistics education*. Voorburg: IOS Press, 153-164.
21. Kanji, G.K. (1979). The role of projects in statistical education. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 28(1), 19-27.
22. Kilpatrick, W. (1921). *The Project method: The use of the purposeful act in the educative process*. New York: Teachers College, Columbia University.
23. Kilpatrick, W.H. (1967). La filosofía de la educación desde el punto de vista experimentalista [Philosophy of education from an experimentalist point of view]. In Kilpatrick, W.H. et al. (eds.), *Filosofía de la Educación*. Buenos Aires: Editorial Losada, 15-74. Original 1942.
24. Krainer, K., Krainz-Dürr, M., Piber, Ch., Posch, P., and Rauch, F. (1998). The CAPE project – Constructing activities for peer-education. In A. Olivier and K. Newstead (eds.), *Proceedings 22nd Conference of the International Group for the Psychology of Mathematics Education, Vol. 1*. Stellenbosch: University of Stellenbosch, 105-112.

25. Markham, T., Larmer, J. and Ravitz, J. (2003). Project-based learning handbook: A guide to standards focused project-based learning for middle and high school teachers. NCTM (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
26. Nolan, D. and Speed, T.P. (1999). Teaching statistics theory through applications. *American Statistician*, 53, 370-375.
27. Santos Guerra, M.A. (2000). *La escuela que aprende [The learning school]*. Madrid: Ediciones Morata.
28. Salinas, J. (1999). El rol del profesorado universitario ante los cambios de la era digital [The role of university teaching staff in the changing digital age]. *Actas del I Encuentro Iberoamericano de Perfeccionamiento Integral del Profesor Universitario*. Caracas: Universidad Central de Venezuela.
29. Salinas, J. (2004). Innovación docente y uso de las TIC en la enseñanza universitaria [Teaching innovation and the use of ICTs in university education]. *RUSC. Universities and Knowledge Society Journal*, 1(1), 1-16.
30. Sanjurjo, L. and Vera, M.T. (1998). *Aprendizaje significativo y enseñanza en los niveles medio y superior [Meaningful learning and teaching at middle and higher education levels]*. Rosario: Homosapiens.
31. Starkings, S. (1997). Assessing student projects. In I. Gal & J.B. Garfield (eds.), The assessment challenge in statistics education. Voorburg: IOS Press, 139-151.
32. Wild, C. and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. With discussion and reply from the authors. *International Statistical Review*, 67(3), 223-265.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 1-8

IMPACT OF HRM PRACTICES ON EMPLOYEE PRODUCTIVITY IN PAKISTAN

Nadia Hanif[1], **Shahid Ghafoor Khokhar**[1], **Noman Arshed**[1], **Tanzila Abdul Karim**[2] and **Madiha Saleem**[1]

[1] Division of Management and Administrative Science
University of Education, Lahore, Pakistan

[2] Federal Urdu University Science and Technology
Islamabad, Pakistan

ABSTRACT

Employee productivity is often cited as the key to an organization's success and expansion. Productivity is the ratio of input to output. The company's inputs include its financial, human, and time resources and its physical, technological, and labour resources; the output is high-quality. Numerous HR procedures may influence employees' productivity. Training, performance reviews, compensation, job definition, teamwork, and employee participation are HRM practices that significantly impact staff productivity in Pakistan's public and private colleges. University and higher education operations must be managed effectively using human resources. Scholars have examined the relationship between various HRM techniques and organizational performance. Therefore, in this study, we discover that HR executives at public and private colleges in Lahore, Pakistan, view HRM practices on Employee Productivity. This study uses quantitative methodology. If numerical data is generated using a sample to evaluate the application of the current study through analysis, the quantitative examination approach is advantageous. The questionnaire includes several questions about how HRM activities like performance reviews, teamwork, and training affect workers' productivity. Teachers from both public and private colleges in Pakistan make up our target sample, and the study's focus was on how HRM policies affected employee productivity. This information comes from three state universities and three private universities in Lahore. Most respondents agreed that HRM practices boost worker productivity. It works well with HRM procedures like training and performance reviews.

1. INTRODUCTION

Highly skilled individuals who work together as a team or individually to achieve a common objective are an organization's highly valued and practical source of human resources (Armstrong 2006). The widely acknowledged source of a company's competitive advantage is its human resources (Tooranloo, Azadi and Sayyahpoor 2017). To enhance educational delivery, every state uses HRM approaches (David and Issahaku, 2013). How a nation's human resources should be managed depends on the ideology and standards of its educational systems (Robinson and Pearce II, 2003). Chang and Chen, 2002 states that current HR research indicates that human resource management strategies are linked with effective operational and high-caliber performance outcomes such as with employee productivity and organizational flexibility.

Performance is the accomplishment of real aims and objectives in comparison to expectations. Employee productivity is one of the performance-related processes (Thuo, Senaji, and Kirimi, 2018). Efficiency and effectiveness are both included in the performance metric known as productivity. Organizations that value culture of employee engagement show high productivity and effectiveness. In such a culture, employees are more enthusiastic to participate in organizational tasks involving establishing goals, making decisions and solving problems that ultimately leading to improved employee performance (Bhatti, and Qureshi 2007).

Performance is a collection of acts or behaviours that support the achievement of long-term goals and objectives. Campbell's (2009). It involves contrasting the intended goals with the actual goals. Productivity among employees is a sign of inspiration (Bhatti, and Qureshi, 2007). Employees who are highly driven are more productive than those whose morale is generally lower.

Huselid (1995) found that the effect of HRM practices determine how effective employees are. When recruiting new personnel various HRM practices including recruitment, selection, and hiring are used. After being hired, an employee must be integrated into the framework of the business through socialization and training. Schuler & MacMillan, (1984) identified other crucial elements of HRM include evaluating an employee's growth, progress, and motivating them through compensation. In fact, there is a long list of human resource management practices that can impact on employee productivity.

Eight human resources practices and their connection to job satisfaction have been examined by Taseem & Soeters (2006). These HR procedures cover hiring and choosing employees, placing them, training them, paying them, evaluating their performance, promoting them, giving them power, and providing them with social security or pension benefits. Huselid (1995) examined eleven practices: hiring processes, performance reviews, incentive pay, task design, decision-making autonomy, sharing of information, evaluation of attitude, involvement of labor, training programs for employees, and standards for promotion.

The association of various HRM techniques with organizational performance is carefully examined by scholars. According to research, incentive pay plans have significant positive effect on worker performance. This effect is significant when measured with innovative human resource management practices such as incorporating flexibility in job design, making employee participate in problem-solving, training employee to enhance their skill set, collaborating, communicating and providing job security. Prevalent human resource policies involve employee compensation, employee training, and employee development. Frye has also investigated this relationship and discovered that equity-based compensation and organizational performance are positively correlated (2004). Consequently, the following are the goals of the current study:

Examine the human resource management practices prevalent in public-sector universities of Pakistan.

Examine the effect of human resource management practices i.e., performance appraisal, training and teamwork on employee productivity as seen by HR executives of both public as well as private universities located in the premises of Lahore, Pakistan.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

The literature talks about how HRM practices including training, performance reviews, teamwork, employee involvement, job definition, and compensation affect employee productivity.

2.1 Performance Appraisal and Employee Productivity

Performance evaluation is understood to be the assessment of a worker's output over a predetermined period as comparable to a worker's report card and authority's assessment of this performance over an academic year (Eldman and Arnold, 2009). The primary goal of a performance evaluation is to guarantee that each representative's skills, interests, and knowledge are fully utilized possible (Arthur, 2008). Fletcher (2001) and Esu and Inyang, (2009) identified performance evaluation system as a system that stands in charge of monitoring the job and the employee fulfilling the explicitly stated organizational objectives. Performance evaluation is considered as a process crucial for employee management and development (Lee, 1985; Eberhardt and Pooyan, 1988). A key factor in determining an organization's competitive edge and success is monitoring and rewarding employee performance (Ployhart et al., 2006). Performance evaluation is referred to a set of actions that organizations employ to evaluate employees, enhance their competence, improve their performance, and distribute rewards. According to Cravens et al. (2015), performance reviews affect employees' motivation, which in turn affects job satisfaction, employee retention, and employee performance.

H1: Performance Appraisal significantly positively affects employee Productivity.

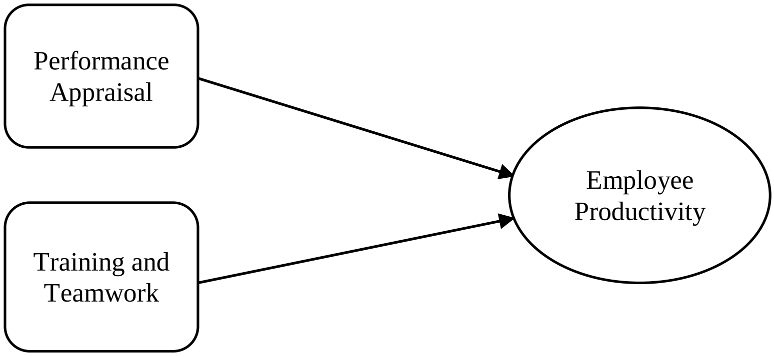
2.2 Training and Teamwork and Employee Productivity

Giving useful skills, ideas, and knowledge through demonstrating how to carry out a practical activity in a professional manner is what is meant by training. Training is a continuous effort and attempts to improve an employee's ability and overall performance (Mahony et al., 2001). Training is used the most common method for improving employee productivity and informing them about organizational goals (Galanou and Priporas 2009). According to Jehanzeb and Bashir (2013), training is also defined as business practice that increases knowledge or enhances skill for the improvement of overall output of human capital, thus, increasing worker productivity results in ideal required performance. Training serves as a systematic strategy that enhances individual, team, and/or organizational performance (Jehanzeb and Bashir, 2013). Investment in employee training in critical areas such as decision-making, problem-solving, cooperation, and interpersonal interactions is rewarding to both i.e., employees and the firm (Rohan & Madhumita, 2012).

Teamwork is the joint effort of two or more people to complete a certain activity, according to Richard (1991). "Teamwork is a collaborative endeavor in which each individual brings their unique talents and perspectives to the whole group's success." The aforementioned statement does not imply that people are less valuable to businesses; rather, it demonstrates that successful teamwork requires more than one person (Ahmad and Manzoor, 2017).

Because it can increase individual production through coordination, teamwork is critical to improve boost the performance of both the individual and the firm. As a result, teamwork among employees becomes the norm for the business (Alie, Bean and Carey, 1998). Teamwork is important to improve personnel as well as organizational productivity, but it must be sustained over time (Ingram, 2000). As per previous discussion, therefore, H2 is

H2: Training and teamwork significantly positively affect employee Productivity.



3. DATA, METHODOLOGY AND RESULTS

In this study, we aimed to assess and examine how HRM practices affect employee productivity in Pakistan's public and private institutions. This study uses quantitative methodology. If numerical data is generated using a sample to evaluate the application of the current study through analysis, the quantitative examination approach is advantageous. The questionnaire includes several questions that gauge how HRM practices affect workers' productivity.

3.1 Research Design

Primary goal of current study is to explore how HRM methods affected employee productivity in Pakistan's public and private universities. This study examines the effects of HRM strategies on employee productivity using information collected from instructors. We only have one type of variable, one that is university-specific, and we are dealing with it on a micro level. This study is quantitative in nature, and it was conducted utilizing a research methodology that was requested as part of the request. The questionnaire used in this study was modified from one used in a study by Qureshi and Ramay (2006) that looked at how HRM practices affected employees' performance.

3.2 Population and Sample

Current study focused on the relationship between employees' productivity and human resource management practices, and target sample includes professors at Pakistan's public and private universities. These statistics come from Lahore universities i.e., three public and three private universities. About 150 questionnaires were given out to people in Lahore's universities, and 114 of them were properly completed and used in the analysis of the study. These questionnaires included those from different department faculty members, including those in the arts, sciences, languages, geography, business, and economics, among others.

This study used SPSS version 22 to analyze the data and distributed it using percentages, averages, standard deviations, and frequencies. A total of 114 questionnaires were completed by students at universities in Lahore, including faculty from various departments including art and science, language and geography, business, and economics, etc.

3.3 Operational Model

Employee productivity is a dependent variable in section 2 of our survey questionnaire. Section 1 of the survey is divided into demographic information such as gender, age, university, faculty, and experience. Section 2 of the survey questionnaire is further divided into three parts related to our selected variables. In contrast, the independent variables' components, such as training, performance evaluation, remuneration and job description, teamwork, and employee participation, were questioned in Section 3. The study used five-point Likert scale for survey.

$$EPT = \sigma + \beta_1 \ PA + \beta T_2 \ W + \ e$$

Here

EPT = Employee Productivity
PA = Performance Appraisal
TW = Training and Teamwork

3.4 Results

The data analysis revealed that 114 total respondents are comprised of 33% males and 67% females. Majority of them fall in 25-35 years of age category and the majority is of the view point that employee productivity is enhanced through human resource management practices. They are of the viewpoint that HRM practices of training and appraisal have positive relationship with employee productivity.

Table 1
Descriptive Statistics

	EPT	PA	TW
Mean	3.396	3.526	3.483
SD	0.628	0.753	0.905
Min	1.67	1.29	1.00
Max	4.67	5.00	5.00
Observation	114	114	114

This table shows that the mean of EPT is 3.3962, PA is 3.5263, and TW is 3.425.

Table 2
Correlation Analysis

	EPT	**PA**	**TW**
EPT	1.0		
PA	0.671**	1.0	
TW	0.527**	0.472**	1.0

$**P < 0.01$(2-tailed) $*p < 0.05$ $N = 144$

Employee productivity has positively correlation with training, performance appraisal and teamwork.

Table 3
Regression of Employee Productivity and Training

	Model 1
PA	0.543*** (7.190)
TW	0.271*** (3.586)
F stats	57.162***
R square	.507

$N = 114$, $p < 0.10$*, $p < 0.05$**, $p < 0.01$***

There is positive and significant relationship between performance appraisal and employee productivity. Training and teamwork positively influence employee productivity since it's significant at 0.01 levels.

4. COMMENTS AND CONCLUSION

The findings and discussion led to the following conclusions: The primary goal of pursuing this study was to explore the impact of HR practices on employee productivity in public and private universities in Lahore, Pakistan. This study aims to build a model that suggests a connection between employee productivity, teamwork, performance evaluation, and pay. The survey concentrated on how well human resources can boost productivity and employee performance. Using a study questionnaire, the data was gathered. The computer program SPSS was used to examine the data that was gathered. This study investigated HRM practices and their impact on output and worker performance and concluded that HRM practices of performance appraisal, training & teamwork increase employee productivity.

5. ACKNOWLEDGEMENT

All the acknowledgement goes to our parents and family, who are pillars in our life and achievements. They are a source of encouragement, motivation, direction and support.

REFERENCES

1. Ahmad, S. and Schroeder, R.G. (2003). The impact of human resource management practices on operational performance: recognizing country and industry differences. *Journal of Operations Management*, 21(1), 19-43.
2. Armstrong, M. (2006). *A Handbook of Human Resource Management Practice* (10th Ed.), London: Kogan Page Limited.
3. Chang, P.L. and Chen, W.L. (2002). The effect of human resource management practices on firm performance: empirical evidence from high-tech firms in Taiwan. *International Journal of Management*, 19(4), 622-631.
4. Galanou, E. and Priporas, C.V. (2009). A model for evaluating the effectiveness of middle managers' training courses: evidence from a major banking organization in Greece. *International Journal of Training and Development*, 13(4), 221-246.
5. Eldman, D.C. and Arnold, H.J. (2009). *Managing individual and group behaviour in organism*. Japan: McGraw Hill Book Company. pp: 392.
6. Bhatti, K.K. and Qureshi, T.M. (2007). Impact of employee participation on job satisfaction, employee commitment and employee productivity. *International Review of Business Research Papers*, 3(2), 54-68.
7. Ingram, H. (2000). Linking teamwork with performance. *Journal of Team Performance Management*, 2(4), 5-10.
8. Lee, C. (1985). Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. *Academy of Management Review*, 10(2), 322-331.
9. Schuler, R. and MacMillan, I. (1984). Gaining competitive advantage through human resource management practices. *Human Resource Management*, 23(3), 241- 255.
10. Tooranloo, H.S., Azadi, M.H. and Sayyahpoor, A. (2017). Analyzing factors affecting implementation success of sustainable human resource management (SHRM) using a hybrid approach of FAHP and Type-2 fuzzy DEMATEL. *Journal of Cleaner Production*, 162, 1252-1265.
11. Thuo, R., Senaji, T. and Kirimi, E. (2018). Influence of strategic human resource management (SHRM) practices on employee productivity: A case of the energy sector in Kenya. *International Academic Journal of Human Resource and Business Administration*, 3(2), 430-451.
12. Arthur, D. (2008). *The First-time Manager's Guide to Performance Appraisals*. American Management Association. New York: Amacom.
13. Cravens, K.S., Oliver, E.G., Oishi, S. and Stewart, J.S. (2015). Workplace culture mediates performance appraisal effectiveness and employee outcomes: A study in a retail setting. *Journal of Management Accounting Research*, 27(2), 1-34.
14. David, W. and Issahaku, J. (2013). *The effectiveness of Human Resource Management Policies on Teachers: A case Study of Ghana Education Services, in Tamale Metropolis*, Ghana. Educational Research.
15. Frye, M.B. (2004). Equity-based compensation for employees: firm performance and determinants. *The Journal of Financial Research*, 27(1), 31-54.

16. Jehanzeb, K. and Bashir, N.A. (2013). Training and development program and its benefits to employee and organization: A conceptual study. *European Journal of Business and Management*, 5(2), 243-252.
17. Mahony, D., Trigg, M., Griffin, R. and Pustay, M. (2001). *International Business*; Pearson: London, UK.
18. Ployhart, R.E., Weekley, J.A. and Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal*, 49(4), 661-677.
19. Richard, L.D. (1991). *Managing teams work in organization*. 2nd Edition. USA, the Dryden Press.
20. Ahmad, I. and Manzoor, S. R. (2017). Effect of teamwork, employee empowerment and training on employee performance. *International Journal of Academic Research in Business and Social Sciences*, 7(11), 380-394.
21. Robinson, R.B. and Pearce II, J.A. (2003). *Strategic management: formulation, implementation and control*.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 9-20

EXPONENTIAL RATIO TYPE ESTIMATOR FOR ESTIMATING THE POPULATION MEAN IN RANKED SET SAMPLING

Irsa Afzal[1] and **Hina Khan**[2]

[1] Virtual University of Pakistan, Lahore, Pakistan
Email: irsa.afzal@vu.edu.pk

[2] Government College University Lahore, Pakistan
Email: hina.khan@gcu.edu.pk

ABSTRACT

This study demonstrated the efficiency of an exponentially ratio type estimator under Ranked Set Sampling design. Mathematical expressions of MSE's and Biases are derived. For efficiency comparison mathematical conditions have been derived by comparing various existing estimators under RSS design. It is concluded through numerical comparison that exponentially type ratio estimator is more efficient than all other present estimators in ranked set sampling.

KEYWORDS

Ranked set sampling, simple random sampling, exponential ratio type estimator, relative efficiency

1. INTRODUCTION

The technique of ranked set sampling (RSS) was first advocated by McIntyre (1952) as a cost efficient altered from simple random sampling for situations where quantities are difficult or costly to find but judgment classification of units according to the variable of study say Y, is moderately easy and inexpensive. It is acknowledged that the estimation of the population mean under ranked set sampling design is more efficient than that the one attained using SRS.

McIntyre (1952) and Takahasi and Wakimoto (1968) anticipated perfect ranking of the elements, that is, there is no flaws in ranking the elements. However, in most circumstances, the ranking could not be done flawlessly. Dell and Clutter (1972) demonstrated that the sample mean under the design RSS is a good point estimator of the population mean, although there are mistakes in ordered classification or not it remains unbiased estimator. Stokes (1980) measured the situation where the ranking is formatted on the basis of a supporting (auxiliary) variable X as an alternative of visual judgment of expertise. For example, at a risky unwanted or wastes location, it can be likely to order pollutant level approving to the degree of wastes. Also it is used in estimating the size of trees in a forest by ranking their widths. In these examples, the variable of interest (pollutant level or size of tree) would be highly associated with the associated (auxiliary) variable.

Exponential ratio type estimator gives most minimum results like regression estimator. In previous studies many exponential ratio type expressions have shown its significance. So the objective of this study is to suggest an exponential type ratio estimator in ranked set sampling and make a comparison of this estimator with some existing ratio type estimators in RSS design.

Let Y be the study variable, X the auxiliary variable, c_x, c_y and c_{yx} be the coefficient of variation of X, coefficient of variation of Y and the coefficient of covariance between X and Y respectively.

2. PROCEDURE OF THE RANKED SET SAMPLING

To generate ranked sets we requisite partition the nominated first level sample into sets of equal size. In mandate to plan an RSS design, we must therefore select a set size that is obviously small, about three or four, to diminish ordered error. Call this set size as m, where m is the number of sampled elements allocated to each set. Now proceed as follows:

i) Randomly select m^2 items from the population.
ii) Allocate the m^2 selected items randomly into m sets, each of size m.
iii) Rank the units within each set based on an observation of comparative values for this variable. This may be built on personal judgment or done with measurements of a covariate that is linked with the variable of study.
iv) Choose a sample for authentic study by including the smallest ranked unit in the first set, the second smallest ranked unit from the second set, continuing in this way until the largest ranked unit is selected from the last m^{th} set.
v) Repeat steps 1 to 4 for r cycles until the desired sample size, $n = m * r$, is acquired for analysis.

The sample mean of RSS is thus defined as:

$$\bar{y}_{RSS} = \frac{\sum_{j=1}^{r}\sum_{i=1}^{m} y_{(i;j)}}{mr}$$

To obtain bias and MSE of the estimators, the useful expectations under RSS are:

$$\bar{y}_{RSS} = \bar{Y}(1+e_0), \bar{x}_{RSS} = \bar{X}(1+e_1), E\ (e_0) = 0 = E\ (e_1), E\ (e_0^2) = \gamma c_y^2 - w_{y(i)}^2$$

$$E(e_1^2) = \gamma c_x^2 - w_{x(i)}^2, E\ (e_0 e_1) = c_{yx} - w_{yx(i)}, \text{where } \gamma = \frac{1}{m * r},$$

$$w_{y(i)}^2 = \frac{1}{(m^2 r)\bar{Y}^2}\sum_{i=1}^{m}\left(\mu_{y[i]} - \bar{Y}\right)^2, w_{x(i)}^2 = \frac{1}{(m^2 r) * \bar{X}^2}\sum_{i=1}^{m}(\mu_{x(i)} - \bar{X})^2$$

and

$$w_{yx(i)} = \frac{\sum_{i=1}^{m}(\mu_{y[i]} - \bar{Y})(\mu_{x(i)} - \bar{X})}{(m^2 r)\ \bar{Y}\bar{X}}.$$

3. BACKGROUND THEORY OF ESTIMATORS

In this section various existing estimators in RSS are mentioned. These estimators are used to compare through the proposed estimator for the sake of efficiency. The well-known estimators in RSS are given below.

Let $Y_{(i)j}$ be the independent random variables. All from the same cumulative distribution function $F(x)$. The RSS estimate of the population mean μ is

$$l_1 = \bar{y}_{RSS} = \frac{1}{m * r}\sum_{j=1}^{r}\sum_{i=1}^{m} Y_{(i)j} \tag{1}$$

According to Zhua chan (2009) $\bar{y}_{RSS}$ is an unbiased estimator of population mean.

The traditional Classical Ratio estimator (1940) for the population μ under simple random sampling is:

$$l_2 = \frac{\bar{y}}{\bar{x}}\bar{X} \tag{2}$$

The classical ratio estimator of population mean μ in RSS given by Khan and Shabbir (2016) is

$$l_3 = \frac{\bar{y}_{RSS}}{\bar{x}_{RSS}}\ \bar{X} \tag{3}$$

The Sisodia and Dwivedi (1981) estimator in RSS.

$$l_4 = \bar{y}_{SD(RSS)} = \bar{y}_{RSS}\,\frac{(\bar{X} + C_x)}{(\bar{x}_{RSS} + C_x)} \tag{4}$$

The Kadilar and Cingi estimator under RSS proposed by Khan and Shabbir (2016),

$$l_5 = \bar{y}_{KC(RSS)} = \bar{y}_{RSS}\ \frac{\bar{X}C_x + \rho}{(\bar{x}_{RSS}C_x + \rho)} \tag{5}$$

The Singh and Taylor (1999) ratio type estimator in Simple Random Sampling proposed in Ranked Set Sampling.

$$l_6 = \bar{y}_{ST(RSS)} = \bar{y}_{RSS}\,\frac{(\bar{X} + \rho)}{(\bar{x}_{RSS} + \rho)} \tag{6}$$

The Bhal and Tuteja (1991) exponential ratio type estimator under RSS:

$$l_7 = \bar{y}_{BT(RSS)} = \bar{y}_{RSS}\ \mathrm{exp}\left[\frac{(\ \bar{X} - \bar{x}_{RSS})}{(\bar{X} + \bar{x}_{RSS})}\right] \tag{7}$$

The Kadilar and Cingi (2009) exponential ratio type estimator in RSS:

$$l_8 = \bar{y}_{KC(RSS)} = \bar{y}_{RSS}\ \mathrm{exp}\left[\frac{(a\ \bar{X} + b) - (a\bar{x}_{RSS} + b)}{(a\ \bar{X} + b) + (a\bar{x}_{RSS} + b)}\right] \tag{8}$$

In the next section the bias and MSE’s of the above estimators debated in section 3.

3.1 Bias and Mean Square Error Expressions

The MSE of l_1 by using the expectations under RSS is obtained as

$$\text{MSE}\ (l_1) = \bar{Y}^2\left(\gamma c_y^2 - w_{y[i]}^2\right) \tag{9}$$

The MSE and Bias of l_2 are given as

$$MSE(l_2) = \alpha\bar{Y}^2\left\{c_y^2 + c_x^2 - 2c_{yx}\right\} \tag{10}$$

where $\alpha = 1 - \frac{f}{n}$

$$Bias\ (l_2) = \alpha\bar{Y}\{c_x^2 - c_{yx}\} \qquad 11)$$

The mean square error (MSE) and Bias of l_3, l_4, l_5, l_6, l_7 and l_8 up to 1st and 2nd leading terms respectively, under RSS design are given as

$$MSE(l_3) = \bar{Y}^2\left[\left(\gamma c_y^2 - w_{y[i]}^2\right) + \left(\gamma c_x^2 - w_{x(i)}^2\right) - 2\left(\gamma c_{yx} - w_{yx(i)}\right)\right] \tag{12}$$

$$\text{B}ias(l_3) = \bar{Y}\{\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right) \tag{13}$$

$$MSE(l_4) = \bar{Y}^2\{\left(\gamma c_y^2 - w_{y[i]}^2\right) + g^2\left(\gamma c_x^2 - w_{x(i)}^2\right) - 2g\left(\gamma c_{yx} - w_{yx(i)}\right) \tag{14}$$

where, $g = \frac{\bar{X}}{\bar{X}+C_x}$

$$Bias(l_4) = g\bar{Y}\left[\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right)\right] \tag{15}$$

$$\begin{aligned} MSE(l_5) = \bar{Y}^2\Bigg\{&\left(\gamma c_y^2 - w_{y[i]}^2\right) + \left(\frac{\bar{X}C_x}{\bar{X}C_x + \rho}\right)^2\left(\gamma c_x^2 - w_{x(i)}^2\right) \\ &- 2\left(\frac{\bar{X}C_x}{\bar{X}C_x + \rho}\right)\left(\gamma c_{yx} - w_{yx(i)}\right)\Bigg\} \end{aligned} \tag{16}$$

$$\text{Bias}\ (l_5) = \left(\frac{\bar{X}C_x}{\bar{X}C_x+\rho}\right)\bar{Y}\left\{\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right)\right\} \tag{17}$$

$$\begin{aligned} MSE(l_6) = \bar{Y}^2\Bigg\{&\left(\gamma c_y^2 - w_{y[i]}^2\right) + \left(\frac{\bar{X}}{\bar{X} + C_x}\right)^2\left(\gamma c_x^2 - w_{x(i)}^2\right) \\ &- 2\left(\frac{\bar{X}}{\bar{X} + C_x}\right)\left(\gamma c_{yx} - w_{yx(i)}\right)\Bigg\} \end{aligned} \tag{18}$$

$$Bias(l_6) = \frac{\bar{X}}{\bar{X}+\rho}\ \bar{Y}\left[\frac{\bar{X}}{\bar{X}+\rho}\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right)\right] \tag{19}$$

$$MSE(l_7) = \bar{Y}^2\left[\left(\gamma c_y^2 - w_{y[i]}^2\right) + \frac{1}{4}\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right)\right] \tag{20}$$

$$Bias(l_7) = \bar{Y}\left[-\frac{\gamma C_{yx} - W_{yx(i)}}{2} + \frac{3\left(\gamma c_x^2 - w_{x(i)}^2\right)}{8}\right] \tag{21}$$

$$MSE(l_8) = \bar{Y}^2\left[\left(\gamma c_y^2 - w_{y[i]}^2\right) + \delta\left(\gamma c_x^2 - w_{x(i)}^2\right)(\delta - 2H_{yx}\ \right] \tag{22}$$

$$\text{Bias}(l_8) = \bar{Y}\left\{-\delta\left(\gamma c_{yx} - w_{yx(i)}\right) + (1+\delta^2)\left(\gamma c_x^2 - w_{x(i)}^2\right)\right\} \tag{23}$$

where

$$H_{yx} = \rho_{yx}\frac{\left(\gamma c_y^2 - w_{y[i]}^2\right)^{\frac{1}{2}}}{\left(\gamma c_x^2 - w_{x(i)}^2\right)^{\frac{1}{2}}}, y \neq x$$

$$\rho_{yx} = \frac{\gamma c_{yx} - w_{yx(i)}}{\left\{\left(\gamma c_y^2 - w_{y[i]}^2\right) * \left(\gamma c_x^2 - w_{x(i)}^2\right)\right\}^{1/2}} \text{ and } \delta = \frac{a\bar{X}}{2(a\,\bar{X} + b)}.$$

4. PROPOSED ESTIMATOR IN RANKED SET SAMPLING

In this section we proposed a class of exponentially ratio type estimator for estimating population mean in ranked set sampling design is given by

$$l_9 = \bar{y}_{RSS}\exp\left[\theta\left\{\frac{\bar{X}^{\frac{1}{h}} - \bar{x}^{\frac{1}{h}}{}_{RSS}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}^{\frac{1}{h}}{}_{RSS}}\right\}\right] \tag{24}$$

where θ, a, h are some auxiliary constants. These constants are used to make a family of estimators by giving some numerical values.

Converting (24) in to e_i's, it becomes

$$l_9 = \bar{Y}(1+e_0)\exp\left[\theta\left\{\frac{\bar{X}^{\frac{1}{h}}\left(1-(1+e_1)^{\frac{1}{h}}\right)}{\bar{X}^{\frac{1}{h}}\left(1+(a-1)(1+e_1)^{\frac{1}{h}}\right)}\right\}\right] \tag{25}$$

Expanding up to 1st order and after simplifications it becomes:

$$l_9 = \bar{Y}(1+e_0)\exp\left[-\frac{t}{h}e_1 + \frac{t}{h^2}{e_1}^2 - \frac{t}{ah}{e_1}^2\right] \tag{26}$$

where $t = \dfrac{\theta}{a}$

Taking expectation on both sides of equation (26)

$$E(l_9 - \bar{Y}) = \bar{Y}E(e_0) - \bar{Y}\frac{t}{h}E(e_1)$$

$$E(l_9) = \bar{Y} \tag{27}$$

which shows that l_9 is an unbiased estimator up to 1st order expansion.

For MSE taking square and applying expectation on both sides of (26)

$$E(l_9 - \bar{Y})^2 = \bar{Y}^2E\left(e_0 - \frac{t}{h}e_1\right)^2$$

By using the results of expectations mentioned in section 3

$$MSE(l_9) = \bar{Y}^2\left[\left(\gamma c_y^2 - w_{y[i]}^2\right) + \frac{t^2}{h^2}\left(\gamma c_x^2 - w_{x(i)}^2\right) - 2\frac{t}{h}\left(\gamma c_{yx} - w_{yx(i)}\right)\right] \quad (28)$$

For obtaining minimum MSE put $\frac{\partial(MSE(l_9))}{\partial t} = 0$ which gives

$$t = h\frac{\left(\gamma c_{yx} - w_{yx(i)}\right)}{\left(\gamma c_x^2 - w_{x(i)}^2\right)} \quad (29)$$

By putting the value of *t* in (28) we get

$$MSE(l_9)_{opt} = \bar{Y}^2\left(\gamma c_y^2 - w_{y[i]}^2\right)[1-\rho^2] \quad (30)$$

where, $\rho = \sqrt{\frac{\left(\gamma c_{yx} - w_{yx(i)}\right)^2}{\left(\gamma c_x^2 - w_{x(i)}^2\right)\left(\gamma c_y^2 - w_{y[i]}^2\right)}}$

To obtain 2nd order bias expanding (25) up to 2nd order and after simplification we get.

$$l_9 - \bar{Y} = \bar{Y}e_0 - \bar{Y}\theta(2-a)\frac{1}{h}e_1 - \bar{Y}(\theta)\frac{1}{h^2}{e_1}^2 - \bar{Y}\theta(2-a)\frac{1}{h}(e_0 e_1) \quad (31)$$

After applying expectation on both sides of (31), we get

$$Bias(l_9) = -\bar{Y}(\theta)\frac{1}{h}[\frac{1}{h}\left(\gamma c_x^2 - w_{x(i)}^2\right) + (2-a)\left(\gamma c_{yx} - w_{yx(i)}\right) \quad (32)$$

4.1 Family of Proposed Estimator

A family can be determined by giving some numeric values to auxiliary constants.

Remark 1:

When $\theta = 1, h = 1, a = 2$ then the proposed estimator turns in to existing estimator l_7 mentioned in section 3.

$$l_7 = \bar{y}_{RSS}\exp\left[\left\{\frac{\bar{X} - \bar{x}_{RSS}}{\bar{X} + \bar{x}_{RSS}}\right\}\right] \quad (33)$$

With MSE

$$\text{MSE }(l_7) = \bar{Y}^2\left[\left(\gamma c_y^2 - w_{y[i]}^2\right) + \frac{1}{4}\left(\gamma c_x^2 - w_{x(i)}^2\right) - \left(\gamma c_{yx} - w_{yx(i)}\right)\right] \quad (34)$$

5. NUMERICAL ILLUSTRATION

To observe the demonstrations of the estimators we use the succeeding three real life populations in which ranked set samples are taken for checking efficiency of the proposed estimator.

The descriptions are given below

Table 1
Population Sources with Characteristics

Parameters	Population 1		Population 2		Population 3	
	Source: US Environmental Protection Agency, 1991		Source: William. G, Cochran Sampling Techniques, 1976		Source: Applied Linear Statistical Models 2004, pg 1348, data set 1	
	X	Y	X	Y	X	Y
	Weight	MPG	Weekly Family Income	Weekly expenditure	Average No. of beds	Average No. of nurses
	N=		N=		N=113	
$\bar{Y}$	2.64217		27.4909		173.248	
$\bar{X}$	3.4712		72.5454		252.168	
ρ	0.58344		0.2521		0.9155	
c_y	0.402		0.36288		0.80028	
c_x	0.2919		0.14357		0.7613	
c_{yx}	0.0685		0.013138		0.5578	
V_y	1.130287		99.5226		19223.21	
V_x	1.0269		108.4904		36859.2	

Table 2
Different Samples and MSE's Results for Population 1

MSE \ n	24 (*m*=3) (*r*=8)	50 (*m*=10) (*r*=5)	15 (*m*=3) (*r*=5)	30 (*m*=3) (*r*=10)	50 (*m*=5) (*r*=10)	100 (*m*=5) (*r*=20)	20 (*m*=10) (*r*=2)	100 (*m*=10) (*r*=10)	20 (*m*=5) (*r*=4)
MSE(l_1)	0.0251	0.00426	0.0412	0.0198	0.00831	0.003742	0.01135	0.00192	0.02217
MSE(l_2)	0.0305	0.0138	0.0497	0.02413	0.0138	0.0062	0.0369	0.006211	0.0369
MSE(l_3)	0.02889	0.01266	0.04707	0.0228	0.01286	0.0057	0.03368	0.00566	0.03423
MSE(l_4)	0.0249	0.00969	0.04066	0.01968	0.0105	0.0047	0.0257	0.0043	0.02805
MSE(l_5)	0.0235	0.00594	0.03852	0.01856	0.0085	0.00385	0.0158	0.00267	0.02287
MSE(l_6)	0.0244	0.00922	0.0399	0.01933	0.01022	0.00458	0.02453	0.00413	0.02723
MSE(l_7)	0.0236	0.00790	0.0385	0.01863	0.00942	0.00423	0.0210	0.003544	0.0251
MSE(l_8)	0.0206	0.00351	0.0338	0.01628	0.00677	0.00305	0.00936	0.00159	0.0181
$MSE(l_9)_{opt}$	0.01659	0.00281	0.0277	0.01308	0.00548	0.00246	0.00748	0.00126	0.0146

Table 3
Different Samples and MSE's Results for Population 2

MSE \ n	9 (m=3) (r=3)	30 (m=3) (r=10)	15 (m=3) (r=5)	28 (m=4) (r=7)	16 (m=4) (r=4)	20 (m=4) (r=5)	15 (m=5) (r=3)	20 (m=5) (r=4)	30 (m=5) (r=6)
MSE(l_1)	4.4422	0.1665	1.977	0.243	1.4818	0.894	1.424	0.758	0.116
MSE(l_2)	7.696	0.2886	3.463	0.515	3.066	1.876	3.463	1.876	0.288
MSE(l_3)	6.483	0.2432	2.910	0.415	2.489	1.515	2.714	1.464	0.225
MSE(l_4)	5.068	0.1901	2.265	0.301	1.817	1.101	1.869	1.002	0.154
MSE(l_5)	4.658	0.1747	2.077	0.264	1.602	0.963	1.586	0.847	0.130
MSE(l_6)	5.067	0.1900	2.265	0.301	1.817	1.101	1.869	1.002	0.154
MSE(l_7)	5.307	0.1991	2.375	0.321	1.937	1.175	2.024	1.087	0.167
MSE(l_8)	4.317	0.1619	1.922	0.237	1.443	0.870	1.388	0.739	0.113
$MSE(l_9)_{opt}$	4.159	0.1560	1.852	0.228	1.387	0.837	1.333	0.710	0.109

Table 4
Different Samples and MSE's Results for Population 3

MSE \ n	60 (m=3) (r=20)	15 (m=3) (r=5)	30 (m=3) (r=10)	20 (m=5) (r=4)	50 (m=5) (r=10)	100 (m=10) (r=10)	100 (m=5) (r=20)	30 (m=10) (r=3)	50 (m=10) (r=5)
MSE(l_1)	86	636	265	318	85	5.06	8.89	107	49.10
MSE(l_2)	25	181	77	129	35	3.60	3.60	76.7	34.9
MSE(l_3)	24	180	76	127	35	3.55	3.56	75.7	34.4
MSE(l_4)	27	200	84	125	34	3.03	3.48	64.6	29.4
MSE(l_5)	25	185	78	121	32	3.07	3.36	65.5	29.8
MSE(l_6)	27	200	84	125	34	3.038	3.49	64.6	29.4
MSE(l_7)	36	264	110	150	40	3.16	4.19	67.0	30.6
MSE(l_8)	44	325	134	161	43	2.57	4.49	54.4	24.9
$MSE(l_9)_{opt}$	14	103	43	51	14	0.82	1.43	17.2	7.97

6. GRAPH COMPARISONS ON DIFFERENT SAMPLE SIZES WITH THE SEPARATE ESTIMATORS FOR POPULATION 1

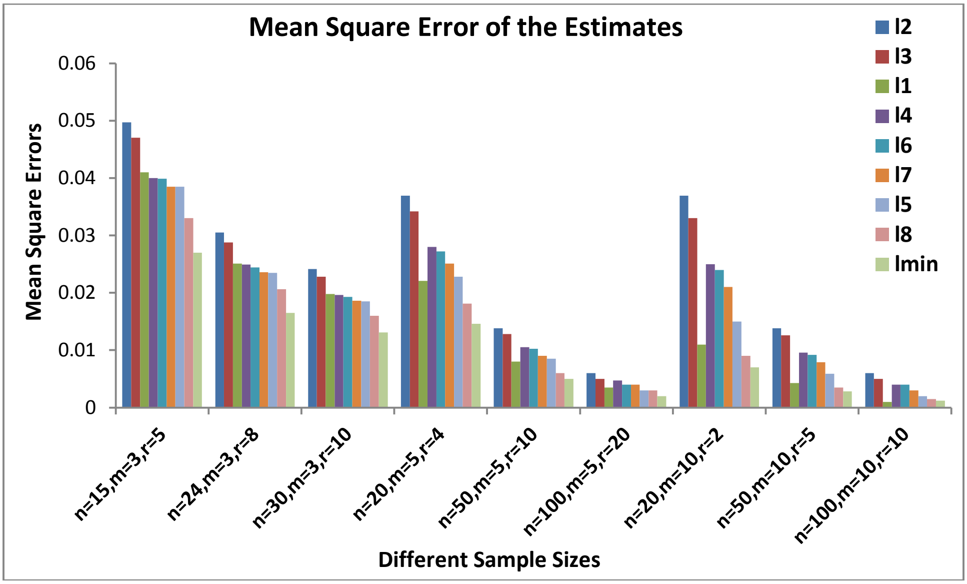


Figure 1: Mean Square Error Comparisons on Different Sample Sizes with the Separate Estimators for Population 1

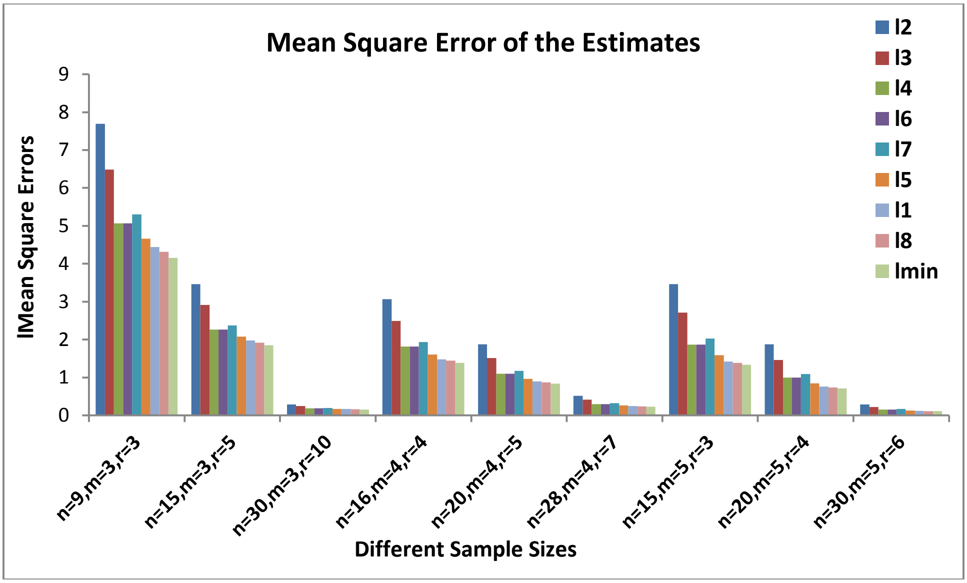


Figure 2: Mean Square Error Comparisons on Different Sample Sizes with the Separate Estimators for Population 2

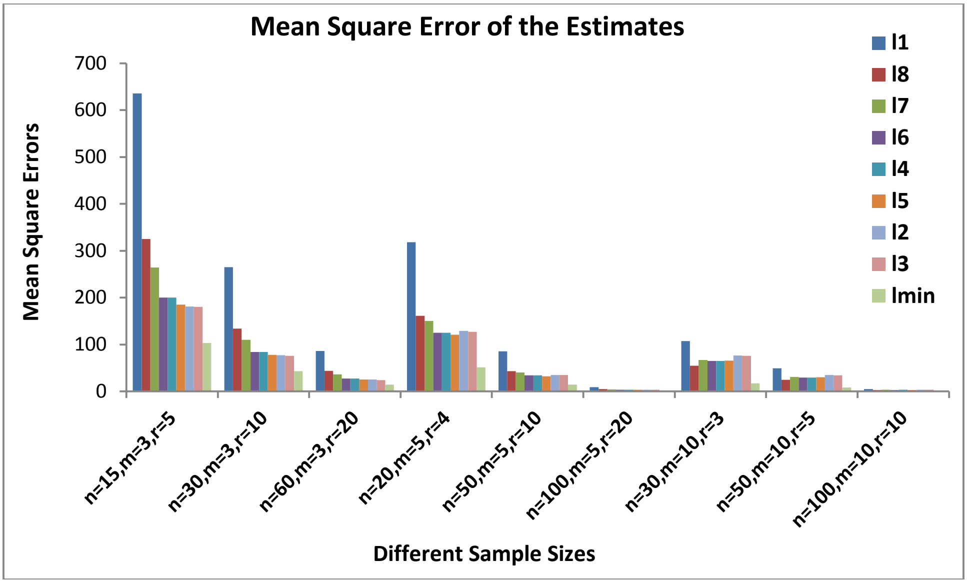


Figure 3: Mean Square Error Comparisons on Different Sample Sizes with the Separate Estimators for Population 3

7. CONCLUSION

It is concluded that derived proposed estimator has minimum results for all sample sets of given three real populations. So from Tables 2, 3, 4 it is numerically determined exponentially ratio type estimator is most preferable over its competitive estimators under RSS. Thus if exponentially ratio type estimator is used it gave better results than other ratio type under RSS design. In this study ratio type existing estimators by Bhal and Tuteja, Kadilar and Cingi, Singh and Taylor unbiased HR ratio type estimators and Sissodia and divedi were experienced with high positive, high negative and with moderate correlations. It is illustrated by using these three correlation levels that MSE of proposed estimator is minimum than all other existing MSE’s under RSS also it is revealed by above numerical results that RSS design gives minimum MSE’s that SRS design, so that RSS design is better than SRS design.

APPENDIX A

In this section different sample sets with some useful information are given.

Table A.1
Different 12 Sample Sets of Population 1 with Characteristics

n	15	24	30	20	50	100	20	50	100
m	3	3	3	5	5	5	10	10	10
r	5	8	10	4	10	20	2	5	10
$w^2_{x(i)}$	0.0023	0.0014	0.0011	0.0023	0.0008	0.00039	0.00323	0.0012	0.00054
$w^2_{y[i]}$	0.00457	0.002833	0.0046	0.00460	0.0017	0.00077	0.00616	0.0023	0.0010
w_{yx}	0.00325	0.0020	0.00159	0.00327	0.00123	0.00054	0.0044	0.0016	0.00075

Table A.2
Different 9 Sample Sets of Population 2 with Characteristics

n	9	15	30	16	20	28	15	20	30
m	3	3	3	4	4	4	5	5	5
r	3	5	10	4	5	7	3	4	6
$w^2_{x(i)}$	0.000844	0.000384	0.00003	0.000409	0.00025	0.00006	0.00051	0.00027	0.00004
$w^2_{y[i]}$	0.00476	0.00217	0.000178	0.0022	0.0014	0.00038	0.0029	0.0015	0.0002
w_{yx}	0.00200	0.000912	0.00007	0.000962	0.00059	0.00016	0.0012	0.00066	0.00010

Table A.3
Different 12 Sample Sets and their Characteristics for Population 3

n	15	30	60	20	50	100	30	50	100
m	3	3	3	5	5	5	10	10	10
r	5	10	20	4	10	20	3	5	10
$w^2_{x(i)}$	0.014	0.0061	0.0019	0.0142	0.0038	0.0003	0.010	0.0049	0.0005
$w^2_{y[i]}$	0.0158	0.0068	0.0021	0.015	0.0042	0.0004	0.012	0.0055	0.0005
w_{yx}	0.0151	0.0064	0.0020	0.0149	0.0040	0.0004	0.011	0.0052	0.0005

REFERENCES

1. Al-Omari, A.I. and Bouza, C.N. (2014). Review of ranked set sampling: modifications and applications. *Revista Investigación Operacional*, 3(35), 215-240.
2. Bahl, S. and Tuteja, R. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12(1), 159-164.
3. Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 545-555.
4. Jeelani, M.I., Mir, S.A. and Pukhta, M.S. (2014). A Class of Modified Ratio Estimators Using Linear Combination of Quartile Deviation and Median of Auxiliary Variable under Rank Set Sampling. *Univ. J. Appl. Math.*, 2(6), 245-249.
5. Khan, L. and Shabbir, J. (2016). An efficient class of estimators for the finite population mean in ranked set sampling. *Open Journal of Statistics*, 6(3), 426-435.
6. Kadilar, C., Unyazici, Y. and Cingi, H. (2009). Ratio estimator for the population mean using ranked set sampling. *Statistical Papers*, 50(2), 301-309.
7. Khan, L. and J. Shabbir, J. (2016). Population mean using ranked set sampling, *Hacettepe Journal of Mathematics and Statistics*, 45(3), 917-928.
8. McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3(4), 385-390.
9. Mandowara, V.L. and Mehta, N. (2013). Efficient generalized ratio-product type estimators for finite population mean with ranked set sampling. *Austrian Journal of Statistics*, 42(3), 137-148.
10. Kadilar, G.Ö. (2016). A new exponential type estimator for the population mean in simple random sampling. *Journal of Modern Applied Statistical Methods*, 15(2), 15.
11. Singh, H.P. & Tailor, R. (2005). Estimation of finite population mean with known coefficient of variation of an auxiliary character. *Statistica*, 65(3), 301-313.
12. Sisodia, B.V.S. and Dwivedi, V.K. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics, 33, 13-18.
13. Singh, H.P., Tailor, R. and Singh, S. (2014). General procedure for estimating the population mean using ranked set sampling. *Journal of Statistical Computation and Simulation*, 84(5), 931-945.
14. Brar, S.S. and Malik, S.C. (2014). Generalized Ratio Type Estimator of Population Mean under Ranked Set Sampling. *International Journal of Statistics and Reliability Engineering*, 1(2), 178-192.
15. Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41(5), 627-636.
16. Wolfe, D.A. (2004). Ranked set sampling: an approach to more efficient data collection. *Statistical Science*, 19(4), 636-643.
17. Wolfe, D.A. (2012). Ranked set sampling: its relevance and impact on statistical inference. *International Scholarly Research Network ISRN Probability and Statistics*, doi:10.5402/2012/568385
18. Zamanzade, E. and Al-Omari, A.I. (2016). New ranked set sampling for estimating the population mean and variance. *Hacettepe Journal of Mathematics and Statistics*, 45(6), 1891-1905.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 21-30

INSTITUTIONAL DEVELOPMENT AND LEVERAGE EFFECT: A PERSPECTIVE FROM CORPORATE SECTOR OF PAKISTAN

Abida Hafeez[1], Sahar Latif Rana[2], Unbreen Arif[2] and **Shamsheer ul Haq[1]**

[1] Department of Economics, Division of Management and Administrative Science, University of Education, Lahore, Pakistan
Email: abida.hafeez@ue.edu.pk; drshamsheer@ue.edu.pk

[2] UE Business School, Division of Management and Administrative Science, University of Education, Lahore, Pakistan
Email: sahar.latif@ue.edu.pk; unbreen.arif@ue.edu.pk

ABSTRACT

This paper looks into how institutional development has affected the leverage effect in Pakistan's corporate sector. The study's goal is to investigate the relationship between volatility change and stock returns both before and after deregulation. The data set in this study covers the time period from 2006 to 2020. The hypothesis developed in this study is that leverage effect after the institutional development is positive and less significant as compared to the period without institutional development. The return-risk theoretical relationship is estimated through the model of the return volatility caused by leverage proposed by Cheung and Ng (1992), Schwert (1989), French et al. (1987), and Christie (1982). The findings highlight that industry-level leverage historically appeared high and also supported by Nishat (2000b). Therefore, relationship of stock return to volatility change is consistently significantly and negative, thereby, explaining leverage effect better during the period of non-reforms/regulation.

KEYWORDS

Institutional Development, Reforms, Volatility, Leverage, Emerging Market.

1. INTRODUCTION

At the firm level, leverage significantly predicts risk premium (Zimmer, 1990). Theoretically, leverage and systemic risk are related. A leveraged firm's systematic risk of is equal to 01 (one) multiplied by the leverage ratio (i.e., debt-to-equity) and the firm's systematic risk in the absence of leverage (Bowman, 1979; and Hamada, 1969). Bowman (1979) asserts that systematic risk, leverage, and accounting beta all have a causal connection (covariability between firm accounting earnings and market portfolio accounting earnings).

One reason to explain the time-variation in stock volatility is where changes/variation in leverage took place as an outcome of changes in the relative values of stocks and bonds. Schwert (1989) argues in which a change in the firm's leverage is necessary for a change/variation in the volatility of stock returns. Haugen and Wichern (1975) used actuarial science to investigate the connection between leverage and the relative stability of stock value. They discovered that the length of the debt is significant in analysing the leverage influence on volatility. The risk of owning shares will fluctuate as the leverage varies over the time period because of the issuance of new debt or appears constant, or when the firms make an effort to repay the loan. An apparently established standard for measurement regarding the benefit of debt financing, according to Kane, Marcus, and McDonald (1985), is equal to the variation in rate(s) of return optimally gained through both types of firms, unlevered, leveraged, along with net of a return premium to account for potential costs of bankruptcy.

By examining firm-specific stock prices, Cheung and Ng (1992), Christie (1982), and Black (1976) find that equity volatility rises after falling stock prices. Leverage along with time-varying risk premia are likely causes. The leverage effect states that as a company's stock price declines, its financial leverage rises, increasing stock return volatility (Christie, 1982; Black, 1976).

According to the hypothesis of leverage, the volatility of log changes in a firm’s net asset value of (which includes debt and equity) is time-invariant, and followed by the volatility of log changes in equity fluctuates over the time period in relation to debt-to-equity ratio of firm. A decrease in a company's assets diminishes (almost) all of its equity value, raising the volatility of future return(s) (Christie, 1982). For highly leveraged organizations compared to less highly leveraged firms, there must be a negative and significant correlation between return and volatility, according to the underlying idea behind the leverage effect. An inverse relationship between changes in stock return volatility from period to $t+1$ and a firm's stock returns in period t. The results further support the idea that the adverse link is stronger for small businesses and those with high debt-to-equity ratios (Cheung and Ng, 1992).

Leverage alone cannot adequately capture the responsiveness of volatility to the direction of returns, according to Black (1976), French et al. (1987), and Schwert (1989). The leverage impact states that a fall in stock value leads to a rise in the debt-to-equity ratio, which raises future volatility as a representation of the riskiness of the company. In contrast to the majority of the time, the industries with minimal leverage show a substantial positive link between change/variation in volatility and stock return(s). The findings support the hypothesis put forward. As a result, the non-reform phase rather than the reform period is a better time to explain the leverage impact. According to Mirza et al. (2016), financial leverage, size, and value show more consistently behaviour than market premium on the PSX. Analysis of the stochastic behaviour of stock prices in Pakistan by Hafeez (2017), Waqar (2014), and Mahmud and Mirza (2011) reveal that the Pakistan Stock Exchange exhibits the leverage effect.

Leverage has an impact on stock return, according to several studies, both at the levels of business and industry (Chorro et al., 2018, Gogia, 2012; Chelley-Steeley and Steeley, 2005; Hull, 1999; and Baker, 1973). According to several research (Baker and Martin, 2011; Dhaliwal et al., 2006; Hamada, 1972; and Bhandari, 1988), stock returns enhance leverage, whereas other studies demonstrate a decrease in leverage as a result of stock gains (Wanjala 2012; Korteweg, 2010; Muradoglu and Sivaprasad, 2009; Dimitrov and Jain, 2008; and Penman et al., 2007). Many academics discover an inverse relationship between stock returns and leverage (Andersson, 2016; Nayeem et al., 2015; Acheampong et al., 2014; Ait-Sahalia, 2013; Wanjala, 2012; Adami et al., 2010, 2013; George and Hwang, 2010; Penman et al., 2007; Hou and Robinson, 2006; Ardatti, 1967; and Hall and Weiss, 1967). Muradoglu and Sivaprasad (2012) discover a favourable correlation and between leverage and stock returns. Kalantonis et al. (2021) highlights negative role of economic agents in leverage in short term.

Hypothesis:

H0: Leverage effect in the period of reform is positive and less significant.

H1: Leverage effect in the period of no-reform is negative and significant.

2. DATA AND MODEL FOR LEVERAGE

Theoretical notions that leverage increases return volatility capture the leverage effect, and it is expected that when the stock market is heavily controlled, the link between stock return and change in volatility is strongly negative. In order to ascertain the leverage effect throughout both the non-reform and reform periods, I estimate the following model in accordance with the aforementioned premise.

$$R_{it} = \alpha_i + \beta_i \Delta(\sigma_{it}) + \mu_{it} \tag{1}$$

where, R_{it} indicates the daily industrial return. $\Delta(\sigma_{it})$ indicates the change in return industrial volatility. A significant as well as negative value of coefficient i.e. β_i implies that leverage effect leads to volatilities in industrial return(s). The industries include textile, sugar, cement, chemical, engineering, food products, fuel & energy, glass and ceramics, ICT, paper and Board, Automobile, pharmaceutical, synthetic and rayon, and miscellaneous. The dataset of fourteen industries is used to explore the leverage affect that has been obtained from the Pakistan stock exchange from 2006 through 2020.

3. RESULT DISCUSSION

One of the theories put up in the literature to explain the time-varying volatility of stock returns seen in the previous section is leverage. According to Schwert (1989), a change in a firm's leverage will affect how volatile stock returns are. For instance, the variance of firm asset returns affects the covariance and variance of stock and bond returns. The variation of the assets is time-invariant for a business with riskless loans. The St. Dev. (standard deviation) of returns is correlated with the St. Dev. of leverage from the previous year, showing a change/variation in the leverage, which affects the volatility of stock return.

Table 1
The Leverage Effect for the Overall Period (2006 to 2020)

Industries	*α*	*t(α)*	*β*	*t(β)*	R^2-adj	F-stat[a]	F-stat[b]
Tex.	0.010	0.470	-0.010	-1.300	0.000	16.098*	8.020*
Sug.	0.030	1.440	0.000	-3.060	0.000	6.388*	2.940**
Cem.	0.040	1.180	0.000	-2.550	0.000	3.111*	3.020*
Chem.	0.010	0.660	0.000	-2.770	0.000	4.401*	3.002**
Eng.	0.020	1.070	-0.010	-6.860	0.020	0.870	4.899*
FP	0.050	2.460	0.000	-20.000	0.120	7.730*	3.650*
FE	0.010	0.350	0.010	4.750	0.010	9.762*	9.230
GC	-0.030	-0.820	0.000	-5.390	0.010	6.824*	4.507*
ICT.	0.010	0.360	0.010	6.280	0.010	9.763*	6.642*
PB	-0.040	-1.370	0.000	-2.760	0.000	12.146*	14.980*
Auto.	0.040	1.490	0.000	-4.170	0.010	18.265*	12.650*
Pharm.	0.030	1.170	-0.020	-7.060	0.020	9.327*	8.620*
SR	-0.020	-0.720	0.000	-10.910	0.040	6.901*	9.405*
Misc.	0.040	0.020	0.000	-2.370	0.000	12.152*	7.302*

The Chow test is used to assess the stability of regression coefficients between the reform and pre-reform eras. The coefficients of regressions are different in the two time periods, according to the alternative hypothesis. The calculated risk premia are established to be regulated by a different relationship in both reform- and pre-reform-periods by the null hypothesis' rejection. The values of F-stat[a] compare periods of both non-reform and reform. The values of F-stat[b] differentiate pre-reform period and the second reform sub-period. *Level of significance at 5%, ** level of significance at 10%.

We calculate the return-risk relationship on the theory that leverage is what causes the stock returns' volatility (Cheung and Ng, 1992; Schwert, 1989; French et al., 1987; and Christie, 1982). I also run a Chow test to show that the connection governing the leverage generated by changes in return volatility is similar in both periods of reform and pre-reform. As the Securities Brokers (Licensing & Operations) Regulations, 2016 established a risk disclosure document to manage various risks, including those associated with leveraged products and derivatives in the later stages of reform, I also check to see if there is a substantial difference between the latter (second) phase of reform and the time of no reform in terms of the relationship between return and change in volatility of returns.

Table 2
The Leverage Effect for the Non-Reform Period (2006 to 2011)

Industry	*α*	*t(α)*	*β*	*t(β)*	R^2-adj
Tex.	-0.080	-3.470	-0.010	-6.040	0.040
Sug.	-0.040	-0.980	0.000	-3.370	0.010
Cem.	-0.080	-1.270	-0.010	-2.580	0.010
Chem.	-0.070	-1.670	-0.010	-1.430	0.000
Eng.	-0.010	-0.310	-0.010	-5.400	0.030
FP	0.030	0.820	0.000	-12.380	0.130
FE	-0.060	-1.230	0.000	-1.100	0.000
GC	-0.160	-3.390	0.000	1.040	0.000
ICT.	-0.090	-1.310	0.010	1.870	0.000
PB	-0.200	-3.910	0.000	-3.770	0.010
Auto.	-0.120	-3.090	-0.010	-5.190	0.030
Pharm.	-0.110	-2.820	-0.020	-5.160	0.030
SR	-0.110	-2.870	-0.010	-3.850	0.010
Misc.	-0.010	-0.400	-0.010	-5.630	0.030

Table 3
The Leverage Effect for the Overall Reform Period (2012 to 2020)

Industry	*α*	*t(α)*	*β*	*t(β)*	R^2-adj
Tex.	0.050	2.150	0.000	-1.010	0.000
Sug.	-0.020	-0.790	0.010	3.900	0.010
Cem.	0.000	-0.050	0.010	3.650	0.010
Chem.	0.090	3.520	0.000	-3.950	0.010
Eng.	0.040	1.530	-0.010	-4.330	0.010
FP	0.070	2.760	0.000	-5.570	0.020
FE	-0.100	-2.580	0.010	7.860	0.030
GC	0.130	2.420	0.000	-4.320	0.010
ICT.	-0.260	-5.530	0.040	11.870	0.070
PB	0.030	0.680	0.000	1.360	0.000
Auto.	0.070	1.840	0.000	1.430	0.000
Pharm.	0.210	6.340	-0.030	-6.080	0.020
SR	0.090	3.090	-0.010	-14.750	0.100
Misc.	0.080	2.900	0.000	-1.230	0.000

Tables 1 to 5 present the results regarding the relationship between return and change in return volatility of industry. The Chow tests show that there are significant differences between the non-reform and reform eras in the relationship between the return and the variation in volatility of returns (induced by the leverage impact) (Table 1). Furthermore, both during the initial (first) time of reform and the later/second reform period, the connection between both returns, and change in volatility (induced by leverage) is noticeably different. Eleven of the fourteen industries with the highest leverage during the entire study period were identified as such by the results shown in Table 1, which generalizes a substantial negative relationship between return and change in volatility. These outcomes are consistent with the actual data gathered from other marketplaces. ICT and the fuel and energy industries show a strong but positive association between return and change in volatility, while the textile business shows a negative but negligible relationship. Except industries for glass & ceramics (GC) and the Information, Communication & Transport (ICT), which appear with insignificant but positive coefficients in the pre-reform/non-reform period, many of the industries (ten (10) out of fourteen (14) industries) appear negative but significant relations of return to change/variation in volatility (Table 2).

Table 4
The Leverage Effect for the Reform Sub-Period (2012 to 2015)

Industry	α	$t(\alpha)$	β	$t(\beta)$	R^2-adj
Tex.	0.050	1.200	0.000	-1.090	0.000
Sug.	0.040	1.120	0.000	-3.080	0.010
Cem.	0.050	0.760	0.010	2.770	0.010
Chem.	0.020	0.660	0.000	-1.030	0.000
Eng.	-0.060	-1.980	-0.010	-4.260	0.020
FP	0.040	1.300	0.000	-10.180	0.090
FE	0.010	0.120	0.000	1.570	0.000
GC	-0.020	-0.260	0.000	-4.030	0.020
ICT.	0.040	0.670	0.020	7.000	0.050
PB	-0.030	-0.570	0.000	1.010	0.000
Auto.	0.020	0.400	0.010	1.780	0.000
Pharm.	-0.020	-0.520	-0.020	-5.050	0.020
SR	0.040	0.830	0.010	3.230	0.010
Misc.	0.010	0.260	0.000	2.400	0.010

Table 5
The Leverage Effect for the Reform Sub-Period (2016 to 2020)

Industry	*α*	*t(α)*	*β*	*t(β)*	R^2-adj
Tex.	0.060	2.070	0.000	1.460	0.000
Sug.	0.880	11.820	0.070	9.200	0.080
Cem.	0.140	3.050	0.000	2.450	0.010
Chem.	0.090	2.770	0.000	-2.720	0.010
Eng.	1.160	16.780	0.010	1.700	0.000
FP	0.080	2.090	0.000	-12.510	0.140
FE	0.090	1.750	0.010	7.130	0.050
GC	0.090	1.610	-0.010	-3.900	0.010
ICT.	0.090	1.750	0.010	2.000	0.000
PB	0.110	2.310	-0.010	-6.070	0.030
Auto.	1.670	16.080	0.080	5.460	0.030
Pharm.	0.200	4.490	-0.010	-2.470	0.010
SR	0.030	0.660	0.000	-11.360	0.110
Misc.	0.800	16.770	0.040	9.000	0.070

However, there is a weak but negative correlation between the chemical and fuel & energy industries. Nishat (2000a) concurs with the large and negative relationship between change in return volatility and return for the PSX enterprises. Six (6) industries showed a substantial negative relationship between change in volatility and return throughout the entire reform period (Table 3). Table 4 lists five industries in the first reform sub-period and six in the second reform sub-period that showed a substantial and negative relationship between change in volatility and return (Table 5). These findings corroborate my alternative theory, which holds that the leverage effect is less important and beneficial during the reform phase than it is during the pre-reform period. However, the leverage effect that causes volatility is more understood during the non-reform phase than it is during the reform period.

4. CONCLUSION

This paper looks into how institutional development has affected the leverage effect in Pakistan's corporate sector. The study's goal is to investigate the relationship between volatility change and stock returns both before and after deregulation. The data set used in this study which spans the years 2006 through 2020.The hypothesis developed in this study is that leverage effect after the institutional development is positive and less significant as compared to the period without institutional development. The findings are consistent with Nishat's (2000b) claim that leverage has traditionally been greater in Pakistan at the industry-level and that has led to a consistently negative and significant link between volatility change and return. However, compared to both reform periods, the leverage effect that causes volatility is better understood during the non-reform period. This is study is of great importance for researchers, investors, portfolio managers and policy makers who want gain insightfulness while trading and predicting the trends in the Pakistan stock exchange.

REFERENCES

1. Acheampong, P., Agalega, E. and Shibu, A.K. (2014). The Effect of Financial Leverage and Market Size on Stock Return on the Ghana Stock Exchange: Evidence from Selected Stocks in the Manufacturing Sector. *International Journal of Financial Research*, 5(1), 125-134.
2. Adami, R., Gough, O., Muradoglu, G. and Sivaprasad, S. (2010). Returns and Leverage. *Oxford Business and Economics Conference Program*, June 28-29, London.
3. Adami, R., Gough, O., Muradoglu, Y. G. and Sivaprasad, S. (2013). How Does a Firm's Capital Structure Affect Stock Performance? *Frontiers in Finance and Economics*, 12(1), 1-31.
4. Ait-Sahalia, Y., Fan, J. and Li, Y. (2013). The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency. *Journal of Financial Economics*, 109(1), 224-249
5. Andersson, M. (2016). *The Effects of Leverage on Stock Returns*, Bachelors Thesis, School of Economics and Management, Lund University, Sweden.
6. Arditti, F.D. (1967). Risk and the Required Return on Equity. *The Journal of Finance*, 22(1), 19-36.
7. Baker, H.K. and Martin, G.S. (2011). *Capital Structure and Corporate Financing Decision: Theory, Evidence and Practice*, 75-90, John Wiley & Sons, Inc, New Jersey
8. Baker, S.H. (1973). Risk, Leverage and Profitability: An Industry Analysis. *Review of Economics and Statistics*, 55(4), 503-507.
9. Bhandari, L.C. (1988). Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. *The Journal of Finance*, 43(2), 507-528.
10. Black, F. (1976). Studies of Stock Market Volatility Changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-81.
11. Bowman, R.G. (1979). The Theoretical Relationship between Systematic Risk and Financial (Accounting) Variables. *The Journal of Finance*, 34(3), 617-630.
12. Chelley-Steeley, P.L. and Steeley, J.M. (2005). The Leverage Effect in the UK Stock Market. *Applied Financial Economics*, 15(6), 409-423.
13. Cheung, Y.-W. and Ng, L.K. (1992). Intersections between the U.S. and Japan Stock Market Indices. *Journal of International Financial Management, Institutions and Markets*, 21(1), 93-108.
14. Chorro, C., Guegan, D., Ielpo, F. and Lalaharison, H. (2018). Testing for leverage effects in the returns of US equities. *Journal of Empirical Finance*, 48, 290-306.
15. Christie, A.A. (1982). The Stochastic Behaviour of Common Stock Variance: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics*, 10, 407-32.
16. Dhaliwal, D.S., Heitzman, S. and Zhen, L.O. (2006). Taxes, Leverage, and the Cost of Equity Capital. *Journal of Accounting Research*, 44(4), 691-723.
17. Dimitrov, V. and Jain, P.C. (2008). The Value-Relevance of Changes in Financial Leverage Beyond Growth in Assets and GAAP Earnings. *Journal of Accounting, Auditing and Finance*, 23(2), 191-222.
18. French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19, 3-29.

19. George, T.J. and Hwang, C.Y. (2010). A Resolution of the Distress Risk and Leverage Puzzles in the Cross Section of Stock Returns. *Journal of Financial Economics,* 96, 56-79.
20. Gogia, A. (2012). *Is the Leverage Effect Caused by Leverage?* Master Thesis, Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology.
21. Hafeez, A. (2017). Modelling the Stochastic Behavior of the Pakistan Stock Exchange. *Journal of Basic & Applied Sciences,* 13, 207-212.
22. Hall, M. and Weiss, L. (1967). Firm Size and Profitability. *The Review of Economics and Statistics,* 49(3), 319-331.
23. Hamada, K. (1969). Optimal Capital Accumulation by an Economy Facing an International Capital Market. *Journal of Political Economy*, 77(4), Part II, 684-697.
24. Hamada, R.S. (1972). The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks. *The Journal of Finance,* 27(2), 435-452.
25. Haugen, R.A. and Wichern, D.W. (1975). The Intricate Relationship between Financial Leverage and the Stability of Stock Prices. *The Journal of Finance*, 30(5), 1283-1292.
26. Hou, K. and Robinson, D.T. (2006). Industry Concentration and Average Stock Returns. *The Journal of Finance,* 61(4), 1927-1956.
27. Hull, R.M. (1999). Leverage Ratios, Industry Norms, and Stock Price Reaction: An Empirical Investigation of Stock-for-Debt Transactions. *Financial Management*, 28(2), 32-45.
28. Kalantonis, P., Kallandranis, C. and Sotiropoulos, M. (2021). Leverage and firm performance: new evidence on the role of economic sentiment using accounting information, *Journal of Capital Markets Studies*, 5(1), 96-107. https://doi.org/10.1108/JCMS-10-2020-0042
29. Kane, A., Marcus, A.J. and McDonald, R.L. (1985). Debt Policy and the Rate of Return Premium to Leverage. *Journal of Financial and Quantitative Analysis*, 28(4), 479-500.
30. Korteweg, A. (2010). The Net Benefits to Leverage. *The Journal of Finance*, 65, 2137-2170.
31. Mahmud, M. and Mirza, N. (2011). Volatility Dynamics in an Emerging Economy: Case of Karachi Stock Exchange. *Ekonomska Istraživanja*, 24(4), 51-64.
32. Mirza, N., Rahat, B. and Reddy, K. (2016). Financial Leverage and Stock Returns: Evidence from an Emerging Economy. *Economic Research-Ekonomska Istraživanja*, 29(1), 85-100.
33. Muradoglu, Y.G. and Sivaprasad, S. (2012). An Empirical Analysis of Capital Structure and Abnormal Returns. Cass Business School Research Paper. Available at SSRN: https://ssrn.com/abstract=948393. or http://dx.doi.org/10.2139/ssrn.9483 93
34. Muradoglu, Y.G. and Sivaprasad, S. (2012). Using Firm-leverage as an Investment Strategy. *Journal of Forecasting* 31(3), 260-279.
35. Nayeem, A.M., Parvez, K., Karim, T. and Tooheen, R.B. (2015) The Impact of Financial Leverage and Market Size on Stock Returns on the Dhaka Stock Exchange: Evidence from Selected Stocks in the Manufacturing Sector. *International Journal of Economics, Finance and Management Sciences*, 3(1), 10-15.
36. Nishat, M. (2000a). Institutional Development and Risk Premia in Pakistan. Paper presented at Asia-Pacific Finance Association Conference, Shanghai, China, July 2000.

37. Nishat, M. (2000b). The Systematic Risk and Leverage Effect in the Corporate Sector of Pakistan. *The Pakistan Development Review*, 39(4), 951-962.
38. Penman, S.H., Richardson, S.A. and Tuna, I. (2007). The Book-to-Price Effect in Stock Returns: Accounting for Leverage. *Journal of Accounting Research*, 45(2), 427-467.
39. Schwert, G.W. (1989). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, 44(5), 1115-1154.
40. Wanjala, J.B. (2012). *Effect of Leverage on Stock Returns: Evidence from Nairobi Securities Exchange*. A Research Project (MBA), School of Business, the University of Nairobi, Kenya.
41. Waqar, O. (2014) Modelling Stock Market Volatility using Univariate GARCH Models: Evidence from Pakistan, Conference Paper. https://www.researchgate.net/publication/283486742
42. Zimmer, S.A. (1990). Event Risk Premia and Bond Market Incentives for Corporate Leverage. *FRB New York-Quarterly Review*, 15(1), 15-30.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 31-40

EXPLORING THE ASSOCIATION BETWEEN BMI AND HEALTHY LIFESTYLE AMONG UNIVERSITY STUDENTS

Abida Hafeez[1], **Naheed Akhtar**[2] and **Rashida Perveen**[3]

[1] Department of Economics, Division of Management & Administrative Sciences, University of Education, Lahore, Pakistan
Email: abida.hafeez@ue.edu.pk

[2] Department of HPESS, University of Karachi, Karachi, Pakistan
Email: nkhtar54@gmail.com

[3] Department of Education, Kinnaird College for Women University Lahore, Pakistan. Email: rashidaahmad103@gmail.com

ABSTRACT

The present study investigates the relationship between BMI and healthy lifestyle. This study based on the component of physical health such as diet, physical education, sleep, mental health. The respondents of the survey are undergraduate and postgraduate students of the University of Karachi conducted in 2021. The majority of students who take part in the research found between age group of 18-30. The sample size is limited to only 300 respondents/students from the University of Karachi. The data is collected through convenient sampling from male and female students. The findings indicate that various factors such as diet, mental health, sleep, and physical activity affect the lifestyle of the participant, eating behaviour and BMI. The result implies that eating behaviour among university student must be improved. Such information is required to guide health promotion initiatives that promote healthy eating in academic settings. Participants' diet quality and nutrient density may be improved by education about healthier food options during work hours, before and after practice, and in the afternoon and evening. The study implies that there is need to encourage individuals to improve their lifestyle and create a public health environment.

KEYWORDS

Health and lifestyle, Physical Activity, Diet, Mental Health, University Students.

1. INTRODUCTION

The healthy eating practice is typically low for fruit, vegetable and dietary fiber (Harris et al., 2023). Firstly, focus on physical activity of the person for making healthy life by designing the activities based on time and diet (Perez et al., 2017). A healthy living wants a healthy lifestyle for preventing the different diseases which are easily communicable (Macovei et al., 2014).

Growing process of a person from childhood to adulthood is an essential part for making good behavioural pattern for good BMI with healthy lifestyle and decrease the chronic disease risk (Huang et al., 2003). The thoughts change according to environment and the different changes around us which consider our health and BMI. For maintaining BMI in healthy lifestyle there should be balance diet in which includes fruits, vegetable, salts, carbohydrate, whole grain (Macovei et al., 2014), for maintaining person's health status lifestyle factor obviously (Sproesser et al., 2014). The world health organization defines the concept of "health" as follows; health is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (Rascanu, 2010). Different factors of life affect the person physical quality, physiological quality and many more to maintaining health (Rascanu, 2010). The right to health, just as the right to good education is one of the fundamental rights of any human being (Neacso, 2010). Students have very less time to meet their nutritional needs. They need regularly well-balanced diet to support high energy needed for training and competition (Sarah, 2005). Less dietary habits are caused by people does not have enough financial support to purchase healthy food (Ziegler, 2002). The National collegiate athletic association (NCAA) students and students are exposed their intense training, practice, and conditioning. They want to reduce their injuries so they choice benefits from healthy diet and nutritional practice (Toni et al., 2012).

Students have uncertain nutritional knowledge and understanding about nutrition with their dietary intake. There is needed to be trained while understanding and attitude are major factors impacting college students' dietetic intake (Courtney et al., 2008). Students can gain knowledge about nutrition from magazines, internet, webs, coaches, parents and their teammates (Sarah, 2005) Nutrition is very important for performance and also injury prevention for students (Reid, 2003). Mostly students are not understanding about role of vitamins, minerals and protein in the body (Zawila, 2003). According to Khan et al. (2022), a diet high in protein sped up the eruption of permanent teeth, but foods high in calcium, minerals, and carbohydrates slowed it down. Students are very healthy conscious, that's why they can understand the role of diet and nutrition in their performances and central focus. Nutritional knowledge improves choice of nutritional quality food. Ali et al., (2015) study show that there is a requirement to conduct such studies in order to achieve a better understanding of nutritional knowledge and dietary habits. Students have ideal diets that assurance to promotion of health and support good performance. Nutrition and dietary status have a direct effect on the person's physical fitness level, physical performance and much more important for training. The main dietary content for active supplement knowledge from dieticians or nutritionists makes individuals capable to obtain adequate nutrition to suitable for health, wellbeing and performance

Hypothesis

H0: There is no a significant relationship between BMI and healthy lifestyle

H1: There is a significant relationship between BMI and healthy lifestyle.

2. LITERATURE REVIEW

A cross-sectional study assessed the students' lifestyle and health, diet patterns in relation to physical activity. The findings indicate that students' choices and practices in respect to sleep, smoking, and mental health were the most encouraging (Deniozou, 2015). The study (Sperrin et al., 2016) looked at people's perceptions and their peers' healthy eating behaviour. BMI is typically greater in shorter adults, particularly women. The age and calendar time dependency of the BMI-height correlation may provide information on the temporal determinants of BMI. The data indicates that BMI is influenced by weight and height. BMI reduced less as height increased. It is stronger in women than in men.

A cross sectional study was conducted to educate the development in a healthy lifestyle. The results show that regular practice of physical activities has many benefits such as healthier and stronger body. These activities also contribute to achieving better mental health and lead towards the better quality of life (Macovei et al., 2014). The purpose of the study is to analyze adolescent consumers' body mass index (BMI). The whole data supports the correlation between activities, BMI, and low food intake (Kuster and Vila, 2017).

A cross-sectional study was conducted to examine BMI in detail among university students from Han, Tibetan, and Uygur cultures, to discuss the differences in their physical characteristics and physical well-being, and ultimately to offer some theoretical recommendations for enhancing students' physical well-being. The findings indicate that there are overweight rates among university students. (Jingya et al., 2013). This study examines the factors that influence young people's lifestyle and behaviour. The aim of this study is to determine the relationship between body mass index (BMI), physical activity, and eating behaviour. The results suggest that there are 18% of youngsters who are obese, 23% who are overweight, and 12% who are at risk (Sidoti et al., 2009). The factors most strongly linked to rising BMI were high protein and low fiber consumption (Spencer et al., 2003).

3. RESEARCH DESIGN

This is a cross-sectional study. Data is collected from students of University of Karachi. The sample size is 300. The sampling technique is convenient sampling. Sample selection is based on inclusion criteria in which data is collected from students studying in University of Karachi, Age 18 to 30 of both male and female. Under the exclusion criteria include those who cannot understand the questions, disable person, and who are not interested to participate and using medicine without prescription. Data is collected from students of University of Karachi who fulfilled inclusion criteria. The parameters include height, weight, BMI of healthy people. The percentages and frequencies have been used through SPSS version 20 to analyze the data.

4. RESULT DISCUSSION

Demographic information shows that out of 300 participants 150(50%) belong to undergraduates' group, 150(50%) are from postgraduate level. It shows that out of 300 participants 75(25%) are belongs to humanities, 150(50%) are from social sciences, 75(25%) are from science group. It shows that out of 300 participants, 60(20%) are from 18-20 age group, 120(40%) are from 21-24 age group, 120(40%) are from 24-over. It also shows that out of 300 participants 150(50%) are males and 150(50%) are females in this study

Table 1 shows general physical health out of 300 participants, 39(13%) are excellent, 75(25%) are very good, 111(37%) are good, 51(17%) are fair, and 24(8.0%) are poor. Table 2 shows that activity level over past week out of 300 is 102(34%) are 0-2 days a week, 123(41%) are 3-4 days a week, 75(25%) are 5 or more days a week. Table 3 shows that the workout of 150 minutes over past week out of 300 156(52%) said "Yes" and 144(48%) said "No". Table 4 shows 161(53.7%) are more active, 77(25.7%) are less active and 62(20.7%) are no change. Table 5 indicates that they are more active out of 300 261(87%)

said "Yes" and 39(13%) said "No". Table 6 exhibits that their diet out of 300 participant 40(13.3%) are very healthy, 96(32%) are healthy, 127(42.3%) are average, 32(10.7%) are unhealthy, 5 (1.7%) are very unhealthy. Table 7(b) shows the portion of fruits and vegetables in breakfast 30(10%) take 0, 19(6.3%) take 1, 80(26.7%) take 2, 75(25%) take 3, 55(18.3%) take 4, 22(7.3%) take 5, and 19(6.3%) take 6 or more. Table 8 shows that fluid level in participants out of 300, 19(6.3%) are 0-1, 80(26.7%) are 3-4, 75(25%) are 5-6, 44(14.7%) are 7-8, 45(15%) are 9-10, 37(12.3%) are more than 10. Table 9 shows wellness of sleep out of 300 participants 80(26.7%) are very well, 85(28.3.0%) are well, 106(35.3%) are average, 26(8.7%) are badly, 3(1.0%) are very badly. Table 10 indicates that sleeping hours out of 300 participants 19(6.3%) are less than 4, 70(23.3%) are 4-6 hours, 159(53%) are 7-8 hours, 44(14.7%) are 9-10 hours, 8(2.7%) are more than 10. Table 11(a) represents that the emotions effect the study 178(59.3%) said "Yes", 122(40.7%) said "No". Table 12 shows the BMI level 65(21.7%) are underweight, 164(54.7%) are normal weight, 56(18.7%) are overweight, 15(5%) are obese.

Percentage (42.3%) of adult who consumed average diet plan is higher as compared to other diet plans including very healthy, healthy, unhealthy and very unhealthy. More than 50% of adults in Karachi University consumed 2 to 3 portions of fruit and vegetables per day, with 10% reporting that they consumed no portion in the day. A key observation is that consumption of breakfast is done 70% between different age groups. The WHO (2016) states that eating at least five servings of fruits and vegetables each day constitutes a healthy diet. Our sample's intake of fruits and vegetables is quite low compared to this suggestion. This result is consistent with earlier research on university students from many nations (Yahia, Rapley, & Dey, 2016; Martinez- Gómez et al., 2012). The results found a significant relation between average nightly sleep, BMI age and gender; these all factors affect the sleep of 35.3 % participants on average. Initially students are asked for self-assessment for their general mental health. The results also reveal that 32% students have very good and 30.3% have good mental health. However, finding of Deniozou, (2015) was 80% students had very good mental health. Every participant, on the whole, is in good mental health, follows an average diet, is physically active, and has a normal BMI.

Table 1
How would you rate your general physical health?

	Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Excellent	39	13.0	13.0	13.0
Very good	75	25.0	25.0	38.0
Good	111	37.0	37.0	75.0
Fair	51	17.0	17.0	92.0
Poor	24	8.0	8.0	100.0
Total	300	100.0	100.0	

Table 2
In the past week, on how many days have you been physically active for a total of 30 minutes or more?

	Frequency	Percent	Valid Percent	Cumulative Percent
0-2 days a week	102	34.0	34.0	34.0
3-4 days a week	123	41.0	41.0	75.0
5 or more days a week	75	25.0	25.0	100.0
Total	300	100.0	100.0	

Table 3
If 4 days or less, have you been physically active for at least two and a half hours (150 minutes) over the course of the past week

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	156	52	52	52
No	144	48	48	100
Total	300	100	100	

Table 4
Have you been more physically active since you started attending University?

	Frequency	Percent	Valid Percent	Cumulative Percent
More active	161	53.7	53.7	53.7
Less active	77	25.7	25.7	79.3
No change	62	20.7	20.7	100.0
Total	300	100.0	100.0	

Table 5
Would you like to be more physically active?

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	261	87.0	87.0	87.0
No	39	13.0	13.0	100.0
Total	300	100.0	100.0	

Table 6
How healthy do you consider your diet to be?

	Frequency	Percent	Valid Percent	Cumulative Percent
Very healthy	40	13.3	13.3	13.3
Healthy	96	32.0	32.0	45.3
Average	127	42.3	42.3	87.7
Unhealthy	32	10.7	10.7	98.3
Very unhealthy	5	1.7	1.7	100.0
Total	300	100.0	100.0	

Table 7(a)
Do you have breakfast most days of the week?

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	210	70	70	70
No	90	30	30	100
Total	300	100	100	

Table 7(b)
How many portions of fruit and vegetables do you eat in a typical day?

	Frequency	Percent	Valid Percent	Cumulative Percent
0	30	10.0	10.0	10.0
1	19	6.3	6.3	16.3
2	80	26.7	26.7	43.0
3	75	25.0	25.0	68.0
4	55	18.3	18.3	86.3
5	22	7.3	7.3	93.7
6 or more	19	6.3	6.3	100.0
Total	300	100.0	100.0	

Table 8
How many 200ml glasses of fluids do you drink on an average day?

	Frequency	Percent	Valid Percent	Cumulative Percent
0-1	19	6.3	6.3	6.3
3-4	80	26.7	26.7	33.0
5-6	75	25.0	25.0	58.0
7-8	44	14.7	14.7	72.7
9-10	45	15.0	15.0	87.7
More than 10	37	12.3	12.3	100.0
Total	300	100.0	100.0	

Table 9
How well do you sleep?

	Frequency	Percent	Valid Percent	Cumulative Percent
Very well	80	26.7	26.7	26.7
Well	85	28.3	28.3	55.0
Average	106	35.3	35.3	90.3
Badly	26	8.7	8.7	99.0
Very badly	3	1.0	1.0	100.0
Total	300	100.0	100.0	

Table 10
How many hours of sleep do you get on an average night?

	Frequency	Percent	Valid Percent	Cumulative Percent
Less than 4	19	6.3	6.3	6.3
4-6	70	23.3	23.3	29.7
7-8	159	53.0	53.0	82.7
9-10	44	14.7	14.7	97.3
More than 10	8	2.7	2.7	100.0
Total	300	100.0	100.0	

Table 11(a)
Since being at university, have you ever had an emotional/mental health difficulty that has affected your life/studies?

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	178	59.3	59.3	59.3
No	122	40.7	40.7	100.0
Total	300	100.0	100.0	

Table 11(b)
How would you rate your general mental health?

	Frequency	Percent	Valid Percent	Cumulative Percent
Excellent	78	26.0	26.0	26.0
Very good	96	32.0	32.0	58.0
Good	91	30.3	30.3	88.3
Fair	32	10.7	10.7	99.0
Poor	3	1.0	1.0	100.0
Total	300	100.0	100.0	

Table 12
BMI Body Mass Index

	Frequency	Percent	Valid Percent	Cumulative Percent
Underweight	65	21.7	21.7	21.7
Normal weight	164	54.7	54.7	76.3
Over weight	56	18.7	18.7	95.0
Obesity	15	5.0	5.0	100.0
Total	300	100.0	100.0	

5. CONCLUSION

This study is on association of BMI and healthy lifestyle. The education of healthy lifestyle requires many approaches in different variations through the sample questionnaire especially in the University of Karachi students about their diet, mental health, sleep, and physical activity. These factors affect the lifestyle of the participant, eating behaviour and BMI. The result implies that eating behaviour among university student must be improved. Such information is required to guide health promotion initiatives that promote healthy eating in academic settings. Participants' diet quality and nutrient density may be improved by education about healthier food options during work hours, before and after practice, and in the afternoon and evening. A quality assurance program in the institution must go hand in hand with the promotion of physical activity among university students. A healthy lifestyle is based on behavioural phenomena must change the eating behaviour, activity level, healthy diet, proper sleep and avoid smoking along with taking daily serving according to the needs and required liquid level that leads to make participants socio-economic status as well as better for maintaining students' better mental health. We really need to encourage individuals to improve their lifestyle and create a public health environment. We need to make students' life safe and healthy. With a good night rest and well rested people not only cope better with stress, but may also have better control of their appetites.

REFERENCES

1. Ali, A., Al-Siyabi, M.S., Waly, M.I. and Kilani, H.A. (2015). Assessment of nutritional knowledge, dietary habits and nutrient intake of university student athletes. Pakistan Journal of Nutrition, 14(5), 293-311.
2. Deniozou, T. (2015). *Student Health and Lifestyle Survey*. PhD. Dissertation, University of Edinburgh, UK.
3. Harris, J., de Steenhuijsen Piters, B., McMullin, S., Bajwa, B., de Jager, I., Brouwer, I.D. (2023). Fruits and Vegetables for Healthy Diets: Priorities for Food System Research and Action. In: von Braun, J., Afsana, K., Fresco, L.O., Hassan, M.H.A. (eds) *Science and Innovations for Food Systems Transformation*. Springer, Cham. https://doi.org/10.1007/978-3-031-15703-5_6
4. Huang, T.T., Harris, K.J., Lee, R.E., Nazir, N., Born, W. and Kaur, H. (2003). Assessing overweight, obesity, diet, and physical activity in college students. *Journal of American College Health*, 52, 83-86.

5. Jingya, B., Ye, H., Jing, W., Xi, H. and Tao, H. (2013). Quantitative analysis and comparison of BMI among Han, Tibetan, and Uygur university students in Northwest China. *Scientific World Journal*, 180863.
6. Khan, N. Abbasi, S.A., Khan, H., Baloch, M. and Chohan, A. (2022). Effect of Dietary Pattern on the Emergence of Permanent Teeth of the Children of Larkana, Pakistan. *Ibnosina Journal of Medicine and Biomedical Sciences*, 14(1), 28-34.
7. Küster, I. and Vila, N. (2017). Healthy lifestyle and eating perceptions: correlations with weight and low-fat and low-sugar food consumption in adolescence. *Frontiers in Life Science*, 10(1), 48-62.
8. Macovei, S., Tufan, A.A. and Vulpe, B.I. (2014). Theoretical Approaches to Building a Healthy Lifestyle through the Practice of Physical Activities. *Procedia-Social and Behavioral Sciences*, 117, 86-91.
9. Moreno-Gómez, C., Romaguera-Bosch, D., Tauler-Riera, P., Bennasar-Veny, M., Pericas-Beltran, J., Martinez-Andreu, S. and Aguilo-Pons, A. (2012). Clustering of lifestyle factors in Spanish university students: the relationship between smoking, alcohol consumption, physical activity and diet quality. *Public Health Nutr*., 15(11), 2131-9.
10. Neacşu, I. (2010). *Pedagogie socială: valori, comportamente, experienţe, strategii.* Editura Universitară, Bucureşti.
11. Perez, L.F., Miller, C.K. and Groner, J.A. (2017). Adolescents with at-risk eating and lifestyle behaviors are affected by after school schedules across the clinical weight spectrum. *Patient Education and Counseling*, 100(8), 1511-1518.
12. Rășcanu, R. (2010). Psihologia sănătății: de la credințe și explicații la sisteme de promovare a ei. In Zlate, M., (coord) Psihologia la răspântia mileniilor, Editura Polirom. Iași, 199-228.
13. Reid-St. John SU. (2003) Working out? Here's what to eat. *Health*, 17(8), 92.
14. Sarah, L.P. (2005). *Dietary Habits and Nutritional Knowledge of College Athletes*. Master Thesis, Faculty of the School of Graduate Studies and Research, California University of Pennsylvania, USA.
15. Sidoti, E., Mangiaracina, P., Paolini, G. and Tringali, G. (2009). Body Mass Index, family lifestyle, physical activity and eating behavior on a sample of primary school students in a small town of Western Sicily. *Ital. J. Public Health*, 6(3), 205-17
16. Spencer, E.A., Appleby, P.N., Davey, G.K. and Key, T.J. (2003). Diet and body mass index in 38000 EPIC-Oxford meat-eaters, fish-eaters, vegetarians and vegans. *Int. J. Obes. Relat. Metab. Disord.*, 27(6), 728-34.
17. Sproesser, G., Schupp, H.T. and Renner, B. (2014). The bright side of stress-induced eating: eating more when stressed but less when pleased. *Psychol. Sci.*, 25, 58-65.
18. Sproesser, G.; Klusmann, V.; Schupp, H.T.; Renner, B. (2015). Comparative optimism about healthy eating. *Appetite*, 90, 212-218.
19. Tony, T.M., Pritchett, K.L., Zippel, D., Minton, D.M., Cellamare, A. and Sibilia, M. (2012). Sports nutrition knowledge among collegiate athletes, coaches, athletic trainers, and strength and conditioning specialists. J. Athl. Train., 47(2), 205-11.
20. World Health Organization (2012). Obesidad y sobrepeso. World Health Organization. (2016). World health statistics 2016: monitoring health for the SDGs, sustainable development goals. Available from: https://apps.who.int/iris/handle/10665/206498

21. Yahia, N., Wang, D., Rapley, M. and Dey, R. (2016). Assessment of weight status, dietary habits and beliefs, physical activity, and nutritional knowledge among university students. *Perspectives in Public Health*, 136(4), 231-44.
22. Zawila, L.G., Steib, C.S.M. and Hoogenboom, B. (2003). The female collegiate cross-country runner: Nutritional knowledge and attitudes. *J. Athl. Train.*, 38, 67-74.
23. Ziegler, P.J., Jonnalagadda, S.S., Nelson, J.A., Lawrence, C., Baciak, B. (2002). Contribution of meals and snacks to nutrient intake of male and female elite figure skaters during peak competitive season. *J. American College Nutrition*, 21(2), 114-119.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 41-48

FINANCIAL DEVELOPMENT ROLE TO ATTRACT FOREIGN DIRECT INVESTMENT: A MODERATION OF INSTITUTIONAL QUALITY

Nadia Hanif[1], **Noman Arshed**[1], **Zeeshan Rafiq**[2] and **Tanzila Abdul Karim**[3]

[1] Division of Management and Administrative Science
University of Education, Lahore, Pakistan

[2] Department of Business and Management,
Information Technology University, Lahore, Pakistan

[3] Federal Urdu University Science and Technology
Islamabad, Pakistan

ABSTRACT

This article examines the relationship between financial development (FD) and foreign direct investment (FDI), applying institutional quality as a moderator and considering the significance of FDI inflows for the sustainable economic advancement of a host country. The sample comprises BRICS countries as these countries are entering a new era of economic integration, mutual developments, and foreign trade. The study findings show that the FD of BRICS host countries significantly attracts FDI, while institutional quality plays a significant negative moderating role in this relation. Hence, maintaining solid financial institutions is advised by policymakers to increase the country's appeal to foreign investors, while focusing on the development of financial markets may increase the advantages of FDI.

1. INTRODUCTION

FDI is strongly impacted by FD. A more favourable climate for FDI can be created through improved financial infrastructure, including stronger banking systems, more access to finance, and more advanced financial markets.

The institutional theory of FD advocates that a country's FD level is defined by its institutional context. This comprises the regulatory and legal systems, contract enforcement, property rights protection and a country's governance environment and structure. This theory holds that the effective operation of financial markets and institutions depends on strong institutions. They support lower transaction costs, risk mitigation, and stable operating environments for investors and businesses. The development and improvement of the financial sector can be hampered by weak institutions, which increase risk, foster uncertainty, and restrict access to credit for some population groups. This can result in a lack of FDI and sluggish economic expansion. The institutional theory of FD places emphasis on the significance of strong institutions for fostering FD and economic success.

In this study, the authors investigated the role of the FD of BRICS FDI. This study examined whether the institutions of BRICS countries strengthen or weaken the relationship between FD and FDI. Country in attracting FDI. Further, the authors explore the influence of institutions as a context on the relationship between FD and FDI.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

2.1 Financial Development and FDI

FDI is strongly impacted by FD. A more favourable climate for FDI can be created through improved financial infrastructure, including stronger banking systems, more access to finance, and more advanced financial markets.

Research has revealed that nations with advanced financial systems typically receive more FDI. For instance, research by Levine (1997) discovered a link between FDI inflows and FD. More FDI was drawn to nations with more developed financial systems, according to the study by Carkovic and Levine (2002). Market size, political stability, and economic growth are all important elements in luring FDI. However, the findings indicate FD as a key factor in FDI determination.

The literature has extensively examined the connection between FD and FDI, with most research indicating a favourable association between the two. FD enhances the investment environment by lowering transaction costs, expanding loan availability, and streamlining capital market operations. A country may become more appealing to foreign investors due to these advances, increasing FDI flows. Aitken and Harrison (1999), Liu and Hitt (2000), and Demirgüç-Kunt and Levine (2002), among others, have shown evidence of a positive association between FDI and FD, while De Mello (1999) has found evidence that FDI may occasionally result in FD. These studies show that FD can play a significant role in attracting FDI and fostering economic expansion.

The body of prior research generally supports the idea that FD and FDI are mutually beneficial, with FD helping to foster an environment more advantageous to foreign investment and FDI contributing to FD by boosting competition and facilitating access to capital.

By increasing access to credit, numerous studies have demonstrated FD to have a favourable influence on FDI. For instance, a study by Levine and Zervos (1998) discovered that FDI rises due to FD. The research, which included a sample of 45 nations, discovered a strong correlation between increasing levels of FDI and FD, as indicated by the proportion of M2 to GDP.

Similarly, Alfaro et al. (2004) found that FDI was positively impacted by FD as determined by the extent of the banking sector. The study, which included 69 countries as a sample, discovered that nations with larger banking sectors drew more FDI, indicating that access to credit is crucial for luring in foreign investment.

Another study by Beck and Levine (2002) discovered that FDI was positively impacted by FD as determined by the private-credit to GDP ratio. According to the research, which included a sample of 49 nations, an increase in private credit to GDP was significantly correlated with an increase in FDI, indicating that access to credit is essential for luring in foreign investment. Overall, these studies show that access to credit promotes FD because it encourages FDI. Foreign investment is more likely to flow into nations with a more developed financial sector and greater access to credit.

Cash availability is one-way financial growth can enhance FDI. A better-developed financial sector makes access to cash for enterprises possible, aiding in attracting FDI. Empirical investigations have shown evidence for this connection. For instance, a study by Borensztein et al. (1998) discovered that FDI was positively impacted by financial growth, as indicated by the proportion of M2 to GDP. The study, which included 69 nations in its sample, discovered a strong relationship between rising FD and rising FDI. The authors contend that access to cash, which can aid in luring foreign investment, is made possible by a more developed financial sector. Another study by Asiedu (2002) found that the measure of FD that favours FDI is the ratio of domestic credit to private sector GDP. The study, which included 53 nations in its sample, discovered a strong correlation between rising FD and rising FDI. According to the author, a highly developed financial industry can supply the money required to finance FDI initiatives. Agosin and Machado's 2005 study also discovered that financial growth, as shown by the size of the banking sector, had a favourable effect on FDI. The study indicated that a larger banking sector was significantly correlated with an increase in FDI using a sample of 17 Latin American nations. The authors contend that access to cash, which can aid in luring foreign investment, is made possible by a more developed financial sector. This research indicates that financial growth boosts FDI by making capital more accessible. A highly developed financial sector makes access to capital possible, which can aid in attracting international investment.

H1: Financial development increases foreign direct investment.

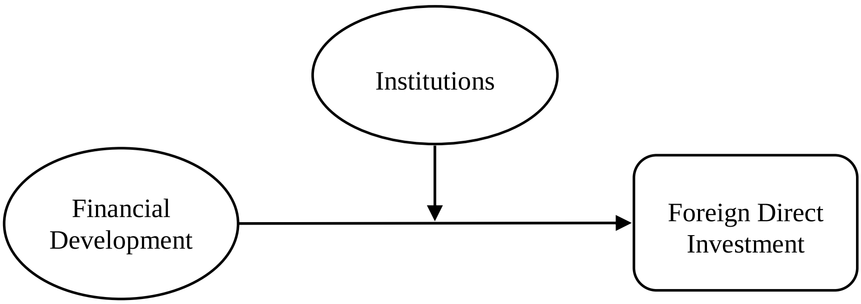
2.2 Institutional Quality and Financial Development

A nation's FD is significantly shaped by its institutions. The degree of FD in a nation is influenced by the advancement and strength of its institutions, including its legal system, property rights, contract enforcement, and regulatory frameworks. For instance, Acemoglu et al. (2005), stated that institutions, including those involved in finance, are the main forces behind long-term economic growth. The authors present data to support their claim that nations with stronger institutions typically have higher levels of FD. Rajan & Zingales (2003) investigate the impact of legal and judicial systems on FD. According to the authors, nations with common law legal systems typically have more advanced and developed financial systems than nations with civil law legal systems. Wei (2000) looks into how corruption affects the development of the financial system. According to Wei (2000), corruption typically discourages international investment through bribery, and FD is hampered. According to a study by Beck et al. (2003), there is a link between FD and the rule of law. They contend that effective legal systems encourage economic expansion by offering a safe and secure environment for business transactions and contractual agreements. Barth et al. (2004) state that well-developed and sound financial regulation can encourage economic development. They contend that by promoting transparency and accountability, a good code can aid in preventing financial catastrophes. Demirgüç-Kunt and Detragiache (1998) discovered a favourable correlation between political stability and FD. They contend that strong political institutions lessen the possibility of arbitrary regulation changes or expropriation, which can damage investor trust. According to Mauro (1995), corruption has a detrimental effect on financial progress. He contends that corruption weakens the rule of law, distorts economic incentives, and damages the credibility and standing of governmental institutions. As a result, access to money is positively correlated with institutional quality. Nations with sound institutional quality typically have more developed financial markets, allowing them to offer a wider variety of financial services to a more significant percentage of the population (Claessens and Laeven, 2005).

H2: Institutional significantly moderates the relationship between financial development and foreign direct investment.

2.3 Theoretical Background

The institutional theory of FD advocates that a country's FD level is defined by its institutional context. This comprises the regulatory and legal systems, contract enforcement, property rights protection and a country's governance environment and structure. This theory holds that the effective operation of financial markets and institutions depends on strong institutions. They support lower transaction costs, risk mitigation, and stable operating environments for investors and businesses. The progress of the financial system can be hampered by weak institutions, which increase risk, foster uncertainty, and restrict access to credit for some population groups. This can result in a lack of FDI and sluggish economic expansion. The institutional theory of FD emphasizes the significance of strong institutions for fostering FD and economic success.



3. DATA, METHODOLOGY AND RESULTS

This study investigated FD's direct influence on attracting FDI in a country while taking institutional quality as a conditional variable. For analysis of the study hypotheses, data regarding financial development and foreign direct investment was collected from World Development Indicators (WDI). The institutional quality data was taken from World Governance Indicators (WGI). The details of the data are provided in Table 1. The study analysis was conducted from 1990 to 2021 on BRICS (Russia, Brazil, India, South Asia and China. The fixed effect and random effect approach was used to study the hypothesized relationships as the study formed panel data.

Table 1
Variables and their Measurements

Variable Name	Proxy	Definition	Source
Foreign Direct Investment	FDI	FDI, net inflows Percentage of GDP	**WDI**
Financial Development	FD	Domestic credit to the private sector by banks' percentage of GDP	WDI
Institutions	Ins	Composite Measure (Regulatory Quality, Corruption, Political Stability, Govt Effectiveness, Voice and accountability and Rule of law)	WGI
Inflation	Inf	Consumer price index (2010 = 100)	WDI
Gross Fixed Capital Formation	GFC	Gross capital formation percentage of GDP	WDI
Bank Capital	BCAR	Bank capital/ assets ratio in percentage	WDI

3.1 Results

Table 2
Descriptive Statistics

	FDI	FD	IQ	Inf	BCAR	GFC
FDI	1.00					
FD	-0.127	1.00				
Ins	-0.167	0.645	1.00			
Inf	-0.147	0.002	-0.184	1.00		
BCAR	-0.213	-0.29	-0.565	-0.08	1.00	
GFC	0.496	0.628	0.979	-0.143	-0.502	1.00

Table 3
Correlations Analysis

	Mean	S.D	Max	Min
FDI	2.218	1.333	.205	5.368
FD	63.869	35.658	16.823	182.868
Ins	4.116	1.5336	1.725	7.256
Inf	93.910	42.158	6.332	186.862
BCAR	8.262	1.6167	5.229	11.948
GFC	26.175	9.911	12.40	46.660

Table 4
Regression Results of Financial Development, Institutions and Foreign Direct Investment

	Model 1	Model 2	Model 3
FD	0.1404*** (0.036)	.116*** (0.042)	0.140*** (0.036)
Ins	1.071** (0.479)	-0.549 (1.379)	1.071** (0.479)
FD*Ins	-0.022*** (0.005)	-0.016** (0.008)	-0.022*** (0.005)
Inf	-0.011** (0.004	-0.015** (0.006)	-0.011* (0.004)
BCAR	0.589*** (0.104)	0.339** (0.149)	0.589*** (0.104)
GFC	0.098 (0.071)	0.237 (0.221)	0.098 (0.071)
Constant	-10.843*** (2.779)	-5.534* (3.279)	-10.843*** (2.779)
F stat	7.52***	2.70**	45.12***
R squared	0.455	0.2709	0.4552
Hausman Test	13.30**		

Model 1 of Table 4 shows the linear regression results, whereas Model 2 and Model 3 show the fixed and random effect analysis results. Model 2 shows that the influence of FD on FDI is significantly positive at $p < 0.01$ level of significance. The moderating influence of institutions is significant and negative at $p < 0.05$ level of significance. This shows that the institutions of BRICS economies are not very strong and feeble, which creates negative conditions and context for attracting FDI. The Hausman test was significant at 5 levels of significance, thus directing to use of the results of the fixed effect equation.

4. COMMENTS AND CONCLUSION

Countries with robust institutions tend to have more developed financial systems, which can promote economic growth and improve the welfare of citizens. Therefore, policymakers should focus on improving institutional quality to foster FD. FD increases the attraction for FDI inflow in a country. Weak institutions negatively influence the relationship between FD and FDI.

5. ACKNOWLEDGEMENT

All the acknowledgement goes to our parents and family, who are pillars in our life and achievements. They are a source of encouragement, motivation, direction, and support.

6. REFERENCES

1. Acemoglu, D., Johnson, S. and Robinson, J.A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of Economic Growth*, 1, 385-472.
2. Agosin, M.R. and Machado, R. (2005). Foreign investment in developing countries: does it crowd in domestic investment? *Oxford Development Studies*, 33(2), 149-162.
3. Aitken, B. and Harrison, A. (1999). Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review*, 89(3), 605-618.
4. Asiedu, E. (2002). On the determinants of foreign direct investment to developing countries: is Africa different? *World Development*, 30(1), 107-119.
5. Barth, J.R., Caprio, G. and Levine, R. (2004). Bank regulation and supervision: what works best? *Journal of Financial Intermediation*, 13(2), 205-248.
6. Beck, T., Demirgüç-Kunt, A. and Levine, R. (2003). Law and finance: why does legal origin matter? *Journal of Comparative Economics*, 31(4), 653-675.
7. Borensztein, E., De Gregorio, J. and Lee, J.W. (1998). How does foreign direct investment affect economic growth? *Journal of International Economics*, 45(1), 115-135.
8. Carkovic, M. and Levine, R. (2002). Does financial development increase economic growth? Evidence from time series and cross-country data. *Journal of Money, Credit, and Banking*, 34(1), 561-586.
9. Claessens, S. and Laeven, L. (2005). Financial dependence, banking sector competition, and economic growth. *Journal of the European Economic Association*, 3(1), 179-207.
10. De Mello, L.R. (1999). Foreign direct investment-led growth: evidence from time series and panel data. *Oxford Economic Papers*, 51(1), 133-151.

11. Demirgüç-Kunt, A. and Detragiache, E. (1998). The determination of banking crises: evidence from developing and developed countries. *International Monetary Fund Staff Papers,* 45(1), 81-109.
12. Demirgüç-Kunt, A. and Detragiache, E. (1998). The determination of banking crises: evidence from developing and developed countries. *International Monetary Fund Staff Papers,* 45(1), 81-109.
13. Demirgüç-Kunt, A. and Levine, R. (2002). *Financial structure and economic growth: A cross-country comparison of banks, markets, and development*. Cambridge, MA: MIT Press.
14. Levine, R. (1997). Financial development and economic growth: Views and agenda. *Journal of Economic Literature*, 35(2), 688-726.
15. Liu, X. and Hitt, M.A. (2000). Financial systems and foreign direct investment. *Journal of International Business Studies*, 31(3), 529-548.
16. Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3), 681-712.
17. Rajan, R.G. and Zingales, L. (2003). The great reversals: the politics of financial development in the twentieth century. *Journal of Financial Economics,* 69(1), 5-50.
18. Wei, S.J. (2000). How taxing is corruption on international investors? *Review of Economics and Statistics*, 82(1), 1-11.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 49-55

IDENTIFYING AFFECTED BRAIN REGIONS OF CHILDREN WITH AUTISM SPECTRUM DISORDER USING INDEPENDENT COMPONENT ANALYSIS AND rs-fMRI DATA

Muhammad Kashif Saleem[1], **Muhammad Farooq**[2]
and **Parwarasha Nazir**[3]
Department of Statistics, GC University, Lahore, Pakistan
Email: [1]ranakashifpu@gmail.com
[2]muhammad.farooq@gcu.edu.pk
[3]parwarashakhan2112@gmail.com

ABSTRACT

Autism Spectrum Disorder is a complex developmental and neurological disability that starts initially in childhood and continues throughout an individual's life. The purpose of this study is to identify the brain regions of children with autism spectrum disorder at early stage using independent component analysis and resting state fMRI data. A sample of 20 rs-fMRI scans of brain of children having problem of autism spectrum disorder is considered along with a sample of size 20 of healthy children. After preprocessing of these images, affected brain regions due to this disorder are identified using independent component analysis technique. It is found that Paracentral Lobule R6402, Postcentral R6002, Frontals up Medial L2601 and Frontals up Medial R2602 are effected regions of brain with autism spectrum disorder. For healthy children the active brain regions are Lingual R5022, Vermis 459120, Suppmotor Areal L8111, Temporals up L8111 and Precuneus R6302.

KEYWORDS

Autism Spectrum Disorder (ASD), Resting-state functional magnetic resonance imaging (rs-fMRI), Independent Component Analysis (ICA).

INTRODUCTION

The brain is a crucial part of the human body and its function can be partially impaired due to genetic factors and lifestyle. This impairment often manifests in the form of diseases such as Parkinson's, Depression, Multiple Sclerosis (MS), Attention Deficit Hyperactivity Disorder (ADHD), and Autism Spectrum Disorder (ASD). While some conditions, such as depression, can be treated with medication, others, such as ASD, cannot be fully cured, but their progression can be managed (Brentani et al. 2013). These diseases add complexity to the brain.

Autism Spectrum Disorder (ASD) is a complex neurological and developmental condition that begins in childhood and persists throughout an individual's life. It is characterized by repetitive behaviors, challenges with verbal and nonverbal social communication, and restricted interests. The causes of ASD are believed to involve both environmental and genetic factors (Hallmayer et al., 2011). The severity of symptoms can vary widely among individuals with Autism. Symptoms typically begin to emerge in children as young as two years old (Ecker et al., 2015). According to the World Health Organization, on average, 1 in every 100 children is diagnosed with ASD. Boys are four times more likely to be diagnosed with autism than girls. The latest Global Burden of Disease Study (GBD) shows that 4.57 million children under five are affected by ASD, with the highest rates occurring in Central Asia, the Middle East, and North Africa (Zeidan et al. 2022).

Neuroimaging is a computational method used to study the structure and function of the brain and central nervous system by scanning brain images in a non-invasive manner. Resting-state fMRI (rs-fMRI) is a popular approach in traditional fMRI neuroimaging, where active brain regions are identified through changes in BOLD signals when the subject is not performing any tasks or responding to stimuli. Structural neuroimaging focuses on the analysis of the anatomical properties and visualization of the brain. This method is useful for detecting abnormalities and brain damage. Several studies have been conducted that use a combination of rs-fMRI and structural MRI (s-MRI) to classify ASD (Rakic et al. 2020; Mellema et al. 2019; Akhavan et al. 2018).

Independent Component Analysis (ICA) is a technique used to analyze brain functional data by separating it into two independent components: a spatial component and a temporal component. The advantage of using ICA is that it can identify subnetworks with different spatial and temporal patterns, without the need for predefined regions of interest (ROIs). The study of Autism Spectrum Disorder (ASD) has particularly been investigated using ICA in various research studies. The purpose of this literature review is to discuss some of the significant studies that have utilized ICA and fMRI data to investigate ASD. One such study by Uddin et al. (2013) focused on identifying functional connectivity differences between typically developing individuals and individuals with ASD using ICA. It was found that individuals with ASD exhibited decreased connectivity within the default mode network, a group of brain regions that become active when an individual is at rest, and not engaged in any specific task. Additionally, the study revealed increased connectivity within the salience network, a network of brain regions involved in detecting and attending to important stimuli. Another study by Di Martino et al. (2014) used ICA to identify brain networks disrupted in individuals with ASD. The study discovered that people with ASD showed decreased connectivity within the frontoparietal network, which is responsible for attention and working memory. Furthermore, the study found increased connectivity within the default mode network. Moreover, Chen et al. (2018) utilized ICA to examine the connection between symptom severity and functional connectivity in individuals with ASD. The study concluded that higher symptom severity was associated with decreased connectivity within the default mode network and increased connectivity within the sensorimotor network, which is responsible for motor processing.

The aim of the present study is to identify the problem of autism spectrum disorder in children at early stage using rs-fMRI data and independent components analysis approach. The two objectives are:

- To identify significant components of brain in ASD subjects and health control subjects
- To explore network of identified brain regions for ASD and health control subjects.

METHODOLOGY

The research is done in three stage; collection, and preprocessing of the data and feature extraction using ICA.

Data

The rs-fMRI and s-MRI scans were obtained from the "Autism Brain Imaging Data Exchange" (ABIDE: https://fcon_1000.projects.nitrc.org/indi/abide/), a publicly accessible database that houses data from 24 global brain imaging facilities. The subjects selected for analysis were between 5 to 10 years of age (Aghdam et al., 2019), with a sample of 53 subjects each for both the autism spectrum disorder (ASD) and control groups, yielding a total of 180 volumes per subject.

Preprocessing

The preprocessing of the fMRI data is a process to eliminate artifacts that are not relevant and transform the data into a standardized form. This preprocessing involves several steps to standardize and clean the data before statistical analysis, and the CONN toolbox in MATLAB v2021 was used for this purpose. The preprocessing pipeline consists of five steps: (1) realignment, (2) slice-time correction, (3) normalization and segmentation, (4) outlier detection, and (5) smoothing. All images were realigned to an average image, which contains more information about all images. The fMRI scans were collected in slices, using an ascending slice acquisition method. Sinc-interpolation was used to time shift and resample the functional data to match the time in the middle of each TR (e.g., 2 seconds). Normalization transforms each brain into a standard template, and the Montreal Neurological Institute (MNI) template was used to normalize both the functional and anatomical scans. The SPM12 unified normalization and segmentation process was used for segmenting the scans into white matter, gray matter, and CSF tissue (Ashburner and Friston, 2005). For normalization, the average BOLD signal image and raw T1-weighted image were used as reference images for the functional and anatomical images, respectively. A standard 180x216x180mm box was used to resample both the anatomical and functional data, which were then resampled using 1mm and 2mm isotropic voxel sizes, respectively, using 4th-order spline interpolation. Possible outlier scans were identified based on subject motion in the scanner and observed global BOLD signal changes. By using a 140x180x115mm bounding box around the brain, frame-wise displacements were calculated at each time, and scans with frame-wise displacements greater than 2mm or global BOLD signal changes more than 9 standard deviations were labeled as possible outliers. The final step of preprocessing was smoothing, which helps to reduce the influence of error variability and increase the BOLD signal-to-noise ratio. This smoothing involved using a Gaussian kernel of 8mm FWHM (Full Width Half Maximum) to smooth both the functional and anatomical scans.

RESULTS

Independent Components Analysis (ICA) is a method used to analyze large datasets consisting of multiple variables by reducing it into a smaller number of dimensions that can be understood as self-organized functional networks. This process helps to simplify and make sense of complex data, allowing for more in-depth and meaningful analysis. The assumptions for using the ICA are the components should be statistically independent and non-Gaussian.

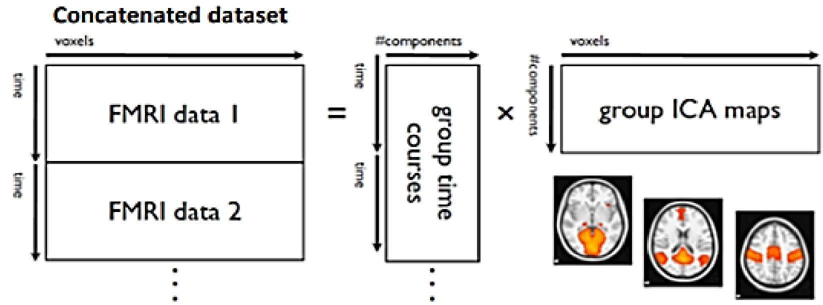


Figure 1: Procedure of ICA

Let us represent fMRI signal with space-time data matrix of measurements $X_{j,t}$. In the linear mixing case we assume that the matrix can be modelled as follows:

$$X_{jt}=\sum_{k=1}^{K} A_{jk}S_{kt} + E_{jt}$$

where A and S are formed by the K independent components of the process, and E is white noise.

The 25 components for ASD and control are derived from ICA model by using "GIFT v3.0c" software.

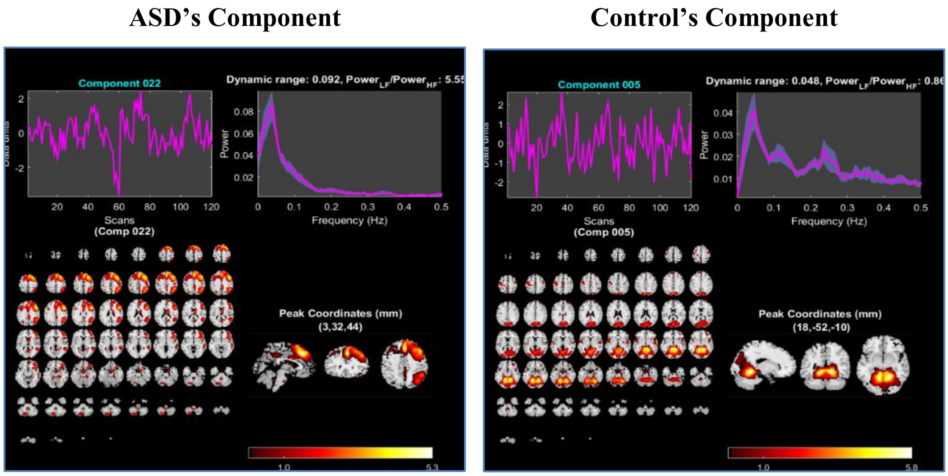


Figure 2: Significant Components for ASD and Controls.

The 4 Components numbered as 6, 16, 21 and 22 are found significant for ASD fMRI. The 5 Components numbered as 5, 1, 14, 2 and 21 are found significant for Control. The above graph of time series to show that pattern of activation changed over time which is ranging from -2 to +2. In z score scale, the red colour indicate some activation and more activation is indicated with brighter yellow colour.

Activated Brain Regions

The activated brain regions are showed by "Connectogram". Warm colours represent decreased and cool colours represent increased connectivity in fMRI brain regions.

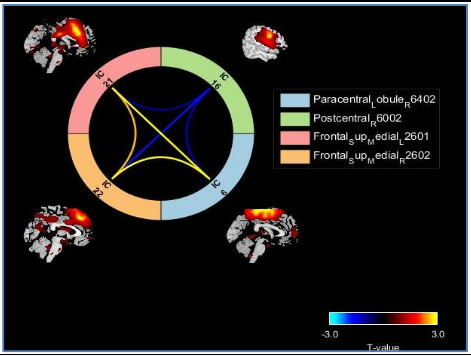
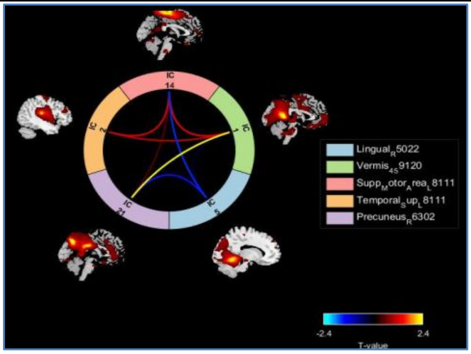
ASD's fMRI Activated Brain Regions	Control's fMRI Activated Brain Regions
	

Figure 4: Connectogram for ASD and Controls

For ASD, Postcentral_R_6002 brain region is negatively correlated with three identified brain regions, which means when the Postcentral_R_6002 region of the brain shows more activation, the rest of correlated regions shows lower activation. And Remaining three brain regions are positively correlated. That shows, with the activation of one region is also caused more activation of the remaining regions.

For Controls, the connectogram indicates the positive correlation between Precuneus_R_6302 and Vermis_45_9120 brain regions. Lingual_R5022 is the only brain region which is negatively correlated with Supp_Motor_Area_L8111 and Precuneus_R_6302 in control Subjects.

DISCUSSIONS

The article is about using independent component analysis (ICA) and resting-state fMRI data to identify the brain regions of children with autism spectrum disorder (ASD). ASD is a developmental and neurological disorder that begins in childhood and persists throughout an individual's life. The disorder is characterized by repetitive behaviors, challenges with verbal and nonverbal social communication, and restricted interests. The causes of ASD are believed to involve both environmental and genetic factors, and the severity of symptoms can vary widely among individuals.

The aim is to identify the brain regions affected by ASD in children at an early stage, using rs-fMRI data and ICA. The study involved a sample of 20 rs-fMRI scans of children with ASD and a sample of 20 rs-fMRI scans of healthy children. The rs-fMRI and structural MRI (s-MRI) scans were obtained from the Autism Brain Imaging Data Exchange (ABIDE), a publicly accessible database that houses data from 24 global brain imaging facilities.

The research used the CONN toolbox in MATLAB v2021 for preprocessing the rs-fMRI data. The preprocessing pipeline consisted of five steps: (1) realignment, (2) slice-time correction, (3) normalization and segmentation, (4) outlier detection, and (5) smoothing. The data was then analyzed using ICA, which is a technique used to analyze brain functional data by separating it into two independent components: a spatial component and a temporal component. ICA can identify subnetworks with different spatial and temporal patterns without the need for predefined regions of interest (ROIs).

The ICA found that the paracentral lobule, postcentral gyrus, and frontal lobe were affected brain regions in children with ASD. In contrast, lingual gyrus, vermis, supplementary motor area, temporoparietal junction, and precuneus were the active brain regions in healthy children.

In conclusion, the study demonstrates the usefulness of rs-fMRI and ICA in identifying the brain regions affected by ASD in children. The study's findings could potentially lead to better diagnostic tools and treatments for ASD. However, it is important to note that the sample size in this study was relatively small, and further research with larger sample sizes is needed to confirm the findings. The other analysis can also be used to find the affected brain regions of the brain and comparisons can be made between ASD patients and healthy controls.

REFERENCES

1. Aghdam, M.A., Sharifi, A. and Pedram, M.M. (2019). Diagnosis of autism spectrum disorders in young children based on resting-state functional magnetic resonance imaging data using convolutional neural networks. *Journal of Digital Imaging*, 32(6), 899-918.
2. Akhavan Aghdam, M., Sharifi, A. and Pedram, M.M. (2018). Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. *Journal of Digital Imaging*, 31(6), 895-903.
3. Brentani, H., de Masi, T.B., Amaral, D., Sato, J.R. and Bressan, R.A. (2013). Autism spectrum disorders: an overview on diagnosis and treatment. *Revista brasileira de psiquiatria*, 35(suppl 1), S62-S72.
4. Chen, C.P., Keown, C.L., Jahedi, A., Nair, A., Pflieger, M.E., Bailey, B.A. and Müller, R.A. (2015). Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage: Clinical*, 8, 238-245.
5. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M. and Deen, B. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6), 659-667.
6. Ecker, C., Bookheimer, S.Y. and Murphy, D.G. (2015). Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *The Lancet Neurology*, 14(11), 1121-1134.
7. Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K. and Lotspeich, L. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*, 68(11), 1095-1102.

8. Mellema, C., Tsuchiya, N., Smith, R.G., Walker, M.E., Strathearn, L. and Kurian, J.R. (2019). Machine learning classifiers applied to resting state fMRI predict individual autism diagnosis. *Neural Computing and Applications*, 31(9), 4557-4568.
9. Rakic, M., Ray, S., Ugrasen, S. and Rasipuram, S. (2020). Classification of Autism using Resting-State fMRI and Brain Network Metrics. *Procedia Computer Science*, 170, 992-996.
10. Uddin, L.Q., Supekar, K., Lynch, C.J., Khouzam, A., Phillips, J., Feinstein, C. and Menon, V. (2013). Salience network-based classification and prediction of symptom severity in children with autism. *JAMA Psychiatry*, 70(8), 869-879.
11. Zeidan, R.K., Dib, A.S., Alothman, A.S., Alhammadi, R.A., Alswaidi, F.M. and Al-Hadidi, K.A. (2022). Prevalence and Associated Factors of Autism Spectrum Disorder among Children under Five Years of Age in the Middle East and North Africa Region: Systematic Review and Meta-Analysis. *Children*, 9(1), 6.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 57-72

ENVIRONMENTAL SITUATION OF OIL INDUSTRY IN PAKISTAN

Mohammad Rafiq Khan[§], **Afaq Amjad Chaudhry,**
Mahlika Javed and **Shanzay Tariq**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

The importance of the Cooking oil Producing Business in Pakistan is the primary subject of this research article to study the manner in which it contributes to the overall deterioration of the environment with pollution generated by the Oil Sector and the extent to which it contributes to the environmental situation in Pakistan. To this end, the inquiry was focused on Kohinoor Oil Mills as a model for interpretation of the situation at the national level. The required secondary data were collected consulting the related literature in the libraries, visiting concerned websites of different oil companies and reading the documented material of the concerned establishments such as reports, etc. For collecting the primary data, interviews with employees and executives of Kohinoor Oil Mills were conducted and responses of the employees and public against well designed questionnaires were taken. The data were sorted, computed and analyzed statistically applying Google. Form software. The findings were then utilized to determine the degree to which the oil industry has been responsible for the deterioration of the environment t in Pakistan.

1. INTRODUCTION

Pakistan is a country with a population of 235 million people and is primarily agrarian (Knoema, 2022). The most recent reported population growth rate is 1.8%, making the country the sixth most populous in the world. This rate is the highest among countries that are geographically close to it. The high pace of population expansion is a cause for concern for the economic planners of the country because the rate of economic growth and development has not been able to keep up with it. The majority of Pakistan's population, approximately 70–80 percent, resides in rural areas and is either directly or indirectly connected to the agricultural industry. Even though there has been significant progress made in the economy of the country, Pakistan is still unable to produce enough edible oils to meet domestic demand. As a result, the country spends a significant amount of its foreign currency on the import of vegetable oils. The total consumption of vegetable oils in Pakistan comes to 27, 73 million metric tons annually (Worldometer, 2019). This equates to a consumption rate of 19.5 kg per person, taking into account both edible and inedible uses. According to sources from the industry, approximately 2.5 kilograms per capita were consumed for non-nutritive purposes, primarily for soap. The remaining 17 kilograms were put to use in food applications. This per capita consumption level of 19.5 kg for edible oils and fats in Pakistan is relatively high in comparison to its neighboring countries. It is also higher than the world's average of 18 kg for edible oils and fats consumption. The cotton seed crop is responsible for the majority of the contribution. The local demand for edible oil is fulfilled not only by the crushing of oil seeds but also by the importation of cooking oil.

Some research studies have been conducted to highlight the problems of the oil industry in Pakistan, a few are outlined below.

The Edible Oil & Ghee Sector Environmental Report (Online) was an attempt to address the environmental contamination problems that are associated with the Edible Oil industry. It has been put together based on the results of three environmental audits that were carried out in three edible oil mills as part of the Environmental Technology Program for Industry (ETPI). The purpose of this report was to evaluate the type and scope of environmental issues that are brought by the edible oil sector and to propose potential remedies to lessen the negative effects of those issues. The report highlighted that the disposal of waste water presents the production of edible oils with its greatest challenge, both in terms of quantity and quality. It did split the generation of wastewater of the industry that deals with edible oils can be split into two categories::Wastewater produced directly by the processes themselves, such as neutralization washings etc. and Wastewater produced by auxiliary systems, such as cooling and vacuum systems etc. The pollution load and concentration of the waste water that is produced by each of these sources is highly variable. The levels of BOD, COD, oil and grease, TSS, TDS, and other contaminants, such as nickel, that are found in process wastewater are rather high. The temperature of the wastewater that is produced by the auxiliary systems is substantially greater, and the quantity of it that is produced is enormous. Sometimes it has residues of volatile organic compounds. Some boiler condensate recovery systems are inefficient, while most of them are so ineffective. In addition to liquid waste, other types of garbage, including solid trash and air pollution, are also produced. The majority of the solid waste that is produced takes the form of spent earth, spent filter cloth, and spent catalyst.

Afzal, et al. conducted a study of the pollutants in wastewater from edible oil/ghee industries and their impacts on plant life, Islamabad. In order to accomplish this goal, 15) samples were taken at random from five 5 different industrial units in duplicate, and then numerous physical and chemical parameters were examined. The data were analyzed were compared with Pakistan's National Environmental Quality Standards (NEQS), It was observed that the majority of the wastewater samples contained contaminants in excess of the permitted limits. The conclusions were as under:

1. The environmental problem of underground and surface water quality around the industrial area was severely deteriorated over the past several years as a direct result of the open disposal of untreated wastewater that resulted into the deoxygenation of the water, which could be harmful for aquatic life.
2. The industrial effluent was fed to the Nala Lye, which eventually makes its way to the River Indus and the agricultural fields via the River Sawan. Even Pakistan's capital city, Islamabad, does not have a proper water pretreatment system, so wastewater from industrial units in Pakistan is not treated.
3. Islamabad is a well-planned city, but Pakistan lacks a proper water pretreatment system. The disposal of industrial effluents has been linked to an increase in the risk of a variety of diseases.

4. Wastewater from industrial processes typically includes both organic and inorganic contaminants. Although some of them, in very small amounts, are necessary for biological functions, there is a possibility that they may become poisonous if they were present in very high concentrations.
5. Both process wastewater and auxiliary wastewater are produced in an edible oil manufacturing facility. The auxiliary wastewater has a high temperature and contains only trace amounts of volatile organic compounds (VOCs), whereas the other wastewater is abundant in oil and grease, chemical oxygen demand (COD), biochemical oxygen demand (BOD), total dissolved solids (TDS), total suspended solids (TSS), sulphates, chloride, phosphate, and nickel, among other contaminants. The aqueous weeds and algal development in waterbodies is promoted by the oil in wastewater flowing from edible oil and ghee factories, which ultimately results in the degradation of water bodies for agricultural use coupled with the loss of dissolved oxygen (DO) caused by eutrophication. DO levels in the range of 4-6 mg/L are highly significant for aquatic organisms; but, due to bacterial activity, these levels are no longer sufficient.
6. The erosion of soil causes sediments and suspended solids, which are a major cause for concern since they prevent photosynthesis from occurring by blocking the sunlight. The discharge of unregulated and non-selective industrial waste poses the largest hazard posed by industrial pollution and is responsible for the degradation of the environment in the Islamabad Industrial Estate as a result of the release of wastewater, solid wastes, and odors. Insofar as the pH of the water is concerned, it plays an essential part in the existence of aquatic life, and any large deviation from its neutral value could create changes in the normal biological processes that occur in the environment. Additionally, acidic or basic water does not help the process of biodegradation of organic contaminants in any way. The pH readings that were measured ranged from 7.70 to 8.30, which is well within the permitted limit of 610 that is established by Pakistan's National Environmental Quality Standards (NEQS). According to the NEQS, the temperatures of all of the effluent samples fall inside the allowable limit of 40 degrees Celsius, and they range from 28.5 degrees Celsius to 31.3 degrees Celsius. The dissolution of ions from dissolved solids results in an increase in the value of electrical conductivity (EC), which in turn results in an increase in water's capacity to conduct more electric current. The measured values of EC in the project area range all the way up to 2136.5 S/cm in the highest point.
7. Edible oil/ghee industries in Islamabad are considered to be the major source of environmental pollution due to negligence of the concerned authorities which has resulted in the ever-increasing stress of pollutants both on soil and groundwater.
8. Demand for industrialization is need of the time but our environment is already much absorbed with industrial wastes which is causing the degradation of living standard of man as well as flora and fauna and natural ecosystem. Water pollution due to slower rate of self- purification and regeneration than that of air is considered to be more complex.
9. The concentrations of BOD, COD, oil & grease and nickel are much higher than the permissible limits of NEQS. Although temperature of all the samples is within the permissible limit i.e. 40°C but it may still accelerate the lowering of dissolved oxygen levels which is fatal for fish.

10. The pH, conductivity, chlorides and sulfates of the samples are within the range. The total suspended solids can be decreased by settling the effluent in ponds. Such water with high concentrations should be avoided for agricultural and irrigation purposes and the whole system should be examined and a workable model be prepared for the treatment of effluents.

The studies reviewed above indicate that it was purely scientific work and the social dimension such as how the pollutants from oil and ghee industries impact the public in the vicinity of the industrial units and impact on employees working in the industrial units was not targeted. Here the study has been extended to target the social dimension also.

2. RESEARCH METHODOLOGY

In order to carry out this research work, the collection of both secondary data and the primary data was requisite. It was collected as under:

Collection of Secondary Data: The secondary data was collected after accumulating a number of research papers and reports and by visiting the websites of the concerned establishments on Internet. The information gathered provided the base line on the study model was to be erected. Moreover it equipped the researchers about the detail of the material used and sub-process involved on the production line to complete the process and exhausting of different types of the pollutants by the industry under study. Above all, the periodic reports published by the industrialists also enlightened the research group how the control of the effluents as being managed.

Collection of Secondary Data: In order to obtain primary data, we relied mostly on two different main modes: interviews of the executives and surveys based on the feedback of employees and the public to the questionnaires served to the physically or online. (Appendix). The first mode was an interview with an executive working in the industry the Kohinoor Oil Mills while the second was the employees of the Kohinoor Oil Mills. The second kind of questionnaire that was used was designed for the general public so that their perspective on the ways they believe the mills were contributing to their safety from the pollutants from the environment.

The Sample: The term "sample" refers to a component or a single object that is representative of a larger whole or group, particularly when it is submitted for examination or displayed as proof of the quality of the whole. This is especially true when the sample is shown as evidence of the quality of the whole.

Interview of the Mill Executive: During the course of our research conducted the interview of the executive of the mill provided fundamental foundation for the data collection. The executive was directly questioned about his ideas on the problem of pollution that is caused by the rice growing industry as well as his thoughts on the subject matter. These questions focused on both the executive's and the executive's thoughts on the issue. Another question that was asked of him involved the actions that his company is taking to stop the further spread of pollution and to create a future that is cleaner, more sustainable, and free of pollution. This question was posed in reference to the future.

Responses from the Employees of Oil Mill: In the second stage of the procedure, an interview with a worker who was employed at the A One rice mills, which has an excellent reputation, was carried out using a questionnaire. The worker was asked a series of questions.

Responses from the General Public: In addition, a questionnaire that was intended for the general public was drafted and then sent to a selection of people for their responses. The people in the sample were questioned about the challenges they are facing as a direct result of the increase in pollution levels, and they were also questioned about the ways in which the continuous exposure to polluted air has damaged their health. Not only were they questioned about their perspectives on pollution, but they were also questioned about what they think the future has in store for them. Additionally, an interview with Mr. X, who is a well-respected executive working at Kohinoor oil mills, was carried out.

3. RESULTS AND DISCUSSION

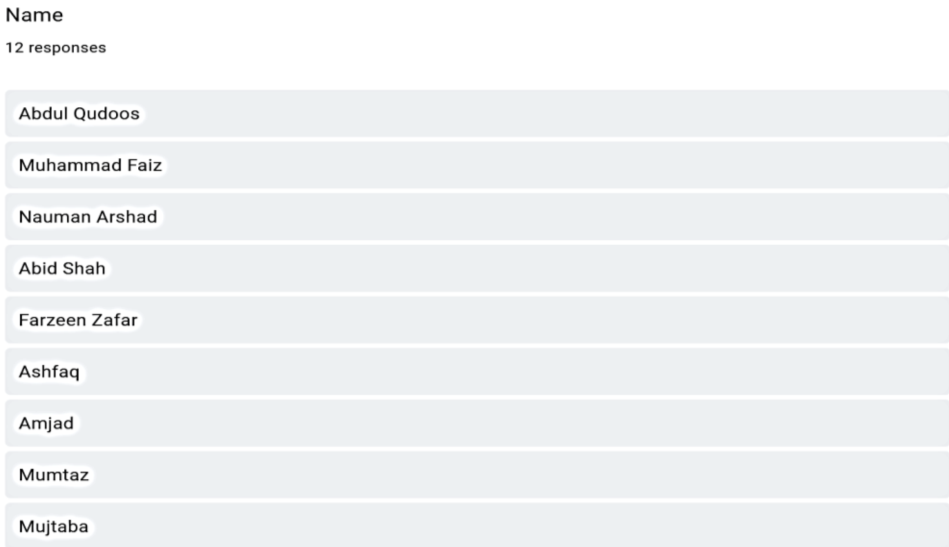
Descriptive Research Results of the Executive Interviewee: After thorough research from questionnaires of the public and employees, to get a better view of the state of the industry we opted for another way of researching. Interviewing executives from the industry broadened our perspective toward oil-producing industries. We interviewed two members of the executive staff, the <u>general manager</u> and the <u>employee manager</u>. We are going to include the analysis of 1 member only (general manager) as both of the interviews were not very different from one another. After asking about his name, health, and occupation we started with our main interview questions. Upon asking how much important the oil industry is for Pakistan he told us all of its importance majorly in terms of Pakistan's economics. Oil industries constitute around 1% of the GDP every year and manufacturing GDP of 12.4% is the highest manufacturing GDP in all industries. Next, we asked what is the potential of this industry to which he replied that it is the source of many incomes for the low class, it plays a great role in the country's economy as discussed above and it also plays significantly in export businesses.

The third question was where do you see the industry in the next 10 years on which he told us the company goals and its vision plans which were to benefit the owners, clients, suppliers, the areas where we do business, and the environment, to foster sustainable growth and coexistence. He agreed to our next question replying yes, he was aware that the oil and ghee industry is causing harm to the environment but then he added that unlike before now the government is taking strict precautions and have implemented hard regulations to save the environment. He also added that their company is strictly abiding by the rules.

For our next two questions, he agreed that it is entirely wrong to dump waste in water bodies as it causes the greatest destruction to sea life but he also added that there are many industries he included no names that still dump their leftover in water bodies.

Upon asked about his take on the spree of smog he replied yes it exists but not entirely because of industries. Yes, industries are a contributing factor but other factors like the burning of forests to clear up space for land, vehicle combustion, and the burning of fossil fuels add to it. Before even asking our next question, he added in his previous statement that if we as a country would be developed, we would have been using advanced technology and methods of disposing of waste, and Insha Allah they will be opting for more environmentally friendly methods in the future which answered our question that if they will opt for more suitable ways of production near future. After analyzing this interview, we came to the conclusion that if governments started taking the right measures industry wouldn't be so bad for the environment but to entirely shut down industry is wrong in many ways mainly economically.

Quantitative Research Results Based on the Reponses of Employees


Name

12 responses

Abdul Qudoos

Muhammad Faiz

Nauman Arshad

Abid Shah

Farzeen Zafar

Ashfaq

Amjad

Mumtaz

Mujtaba

Figure 1: These are 12 Workers from Whom We Collected Relevant Data

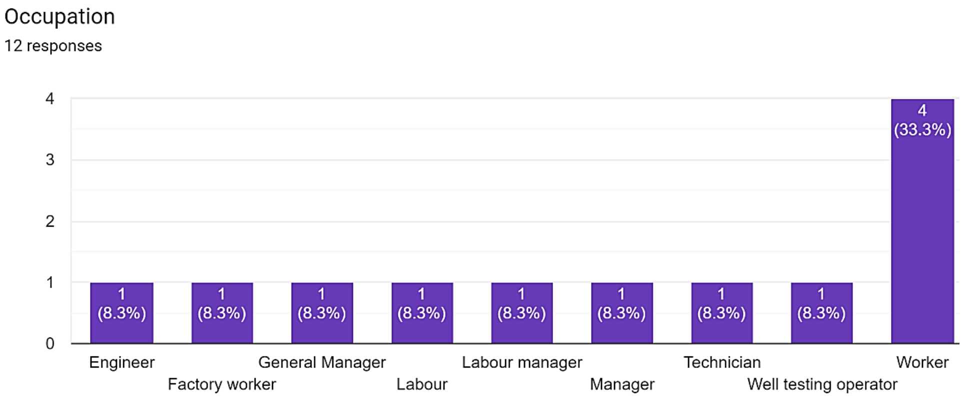


Figure 2: This Bar Chart Shows Different Role of People have in the Company

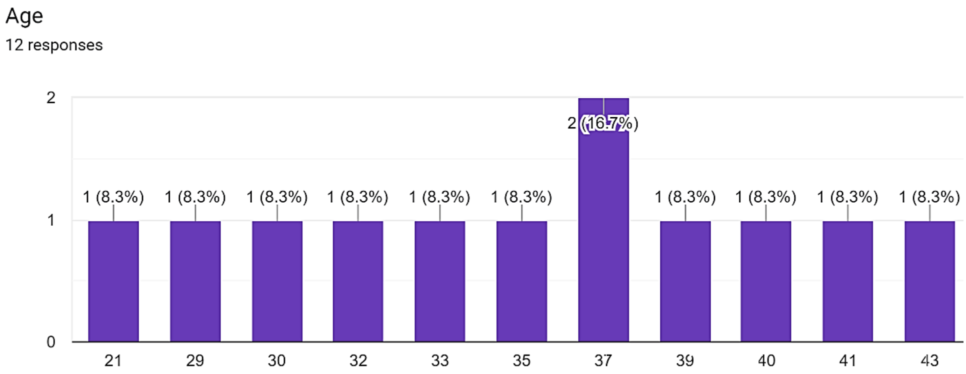


Figure 3: This Bar Chart Shows the Age of Workers

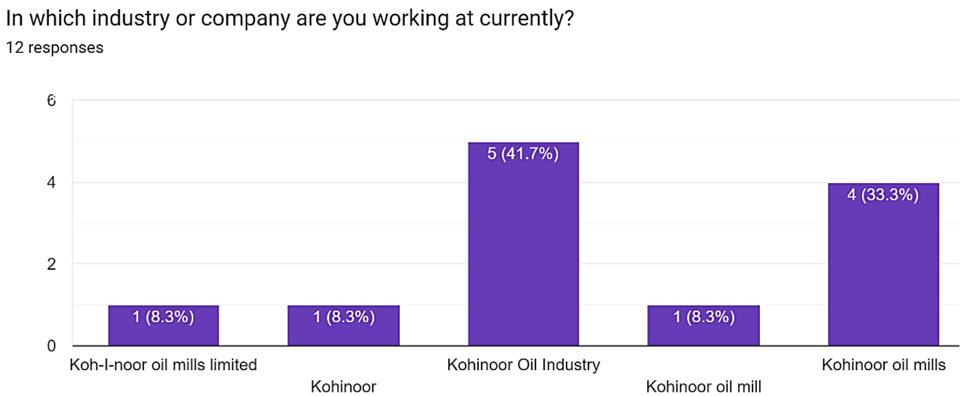


Figure 4: This Bar Chart shows that what Percentage of Workers are Working in Kohinoor Industry

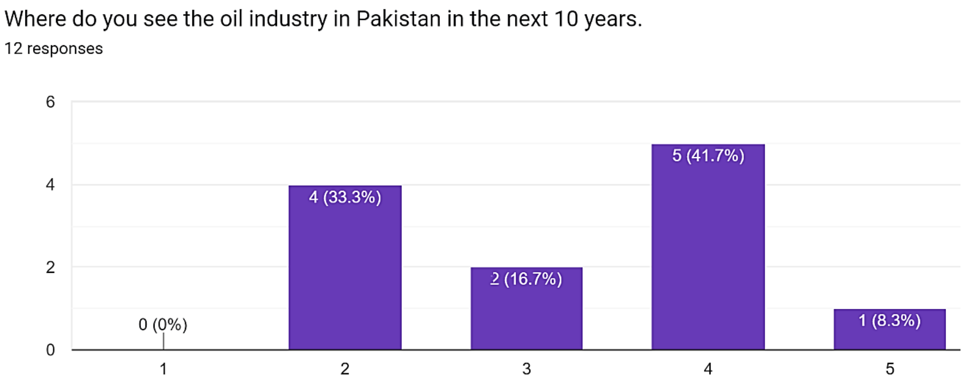


Figure 5: Showing Results of Responses to the Question, "Where Do You See the Status of Oil Industry in Pakistan?"

This question was asked to know the take of oil industry workers on where they see the steel industry of Pakistan in the next 10 years.

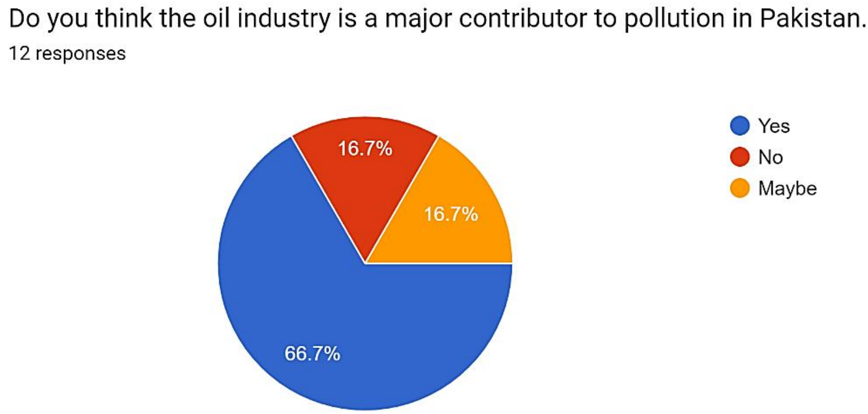
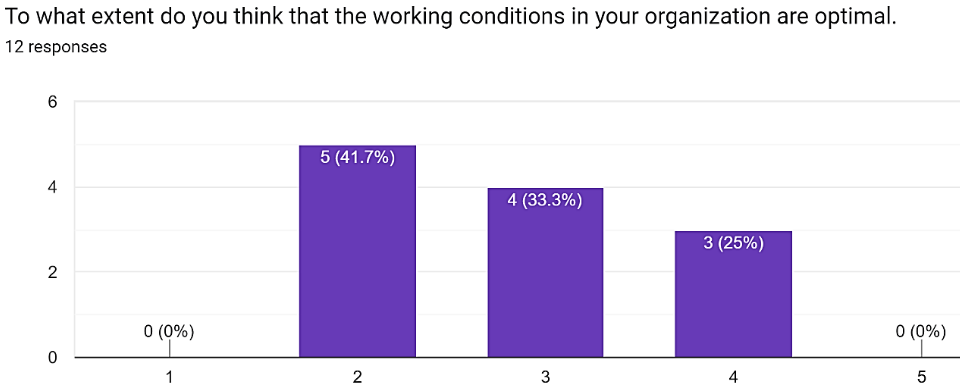


Figure 6: Showing Responses to the Question, "Is the Oil Industry a Major Contributor to Pollution in Pakistan"

This question was asked from the workers to know whether they think that the oil industry is a major contributor to pollution in Pakistan. To this question 66.7 percent of the workers agreed whereas 16.7 disagreed. 16.7 percent of the workers were unsure about this issue.

Figure 7: Exhibiting Results of Responses to the Question, "Are You Provided with Safety Gear by your Government?"

This was question was asked in order to obtain information about the working conditions in the oil industry and whether the employees working there are treated in a positive manner or not. 66.7 percent of the employees claimed to be treated well at their organizations whereas 25 percent claimed that they weren't treated well.



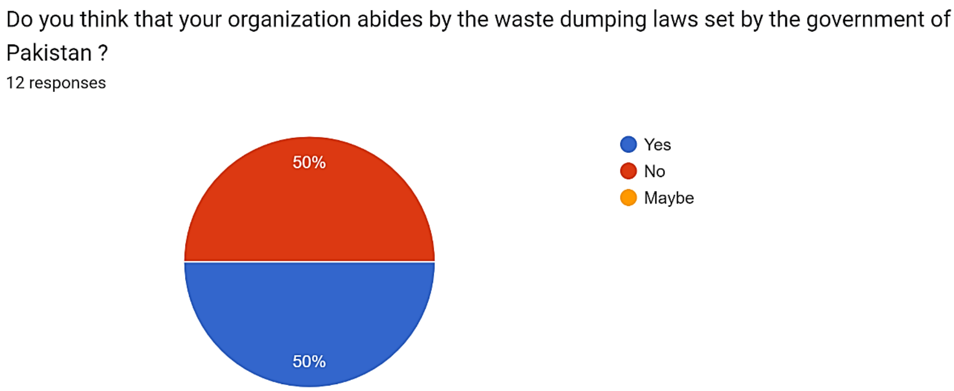


Figure 8: Exhibiting Results of Responses to the Question, "Do You Think that Your Organization Abides by the Waste Dumping Laws Set by the Government of Pakistan?"

This was question was asked in order to obtain information about the working conditions in the oil industry and whether the laws are being followed or not. 50 percent of the employees claimed that laws are being followed whereas 50 percent claimed that they weren't followed.

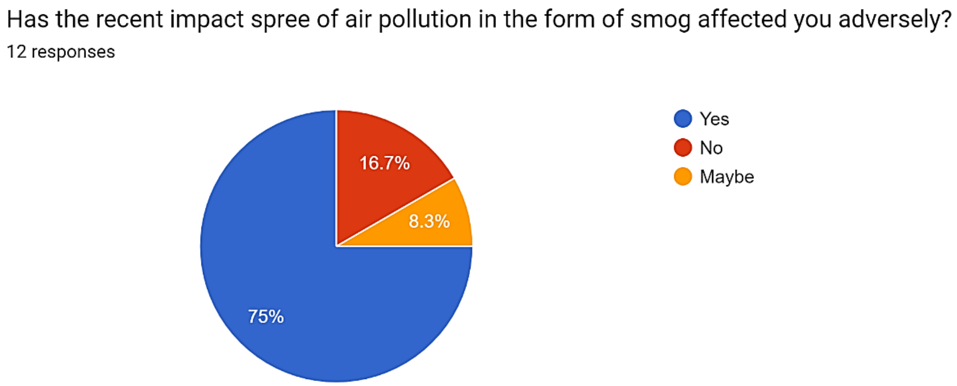


Figure 9: Exhibiting Results of 12 Responses to the Question, "Has the Recent Spree of Air Pollution in the form of smog affected you adversely?"

This was question was asked in order to obtain information about how bad smog has affected and its contribution to the air pollution. 75 percent of the employees claimed that smog has played a major role in air pollution whereas 16.7 percent claimed that it has not affected.

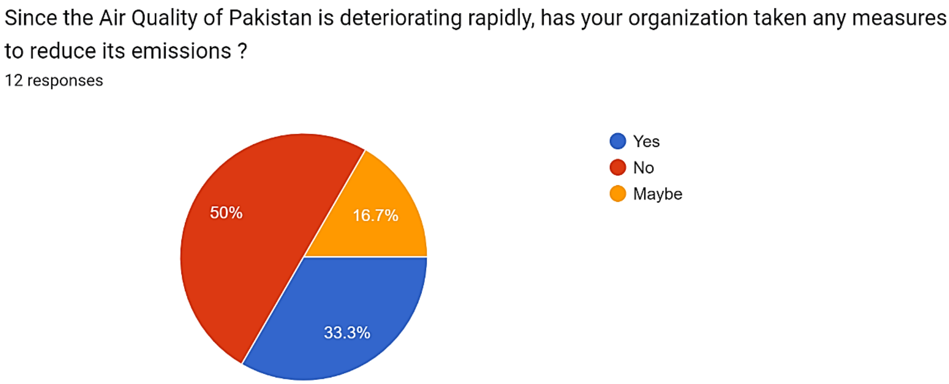


Figure 10: Exhibiting Results of Responses to the Question, "Since the Air Quality of Pakistan is Deteriorating Rapidly, has your Organization Taken any Measures to Reduce its Emissions?"

To this question 33.3 percent of the employees claimed that their organizations are taking relevant measures, whereas 50 percent of them claimed that no measures were being taken.

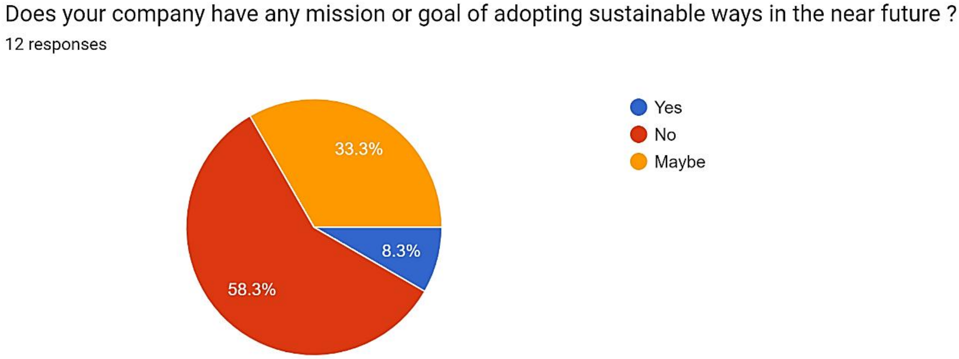


Figure 11: Exhibiting Results of Responses to the Question, "Does Your Company have Any Mission or Goal of Adopting Sustainable Ways in the Near Future?"

To this question 8.3 percent of the employees claimed that their organizations have missions or goals of adopting sustainable ways in the near future, whereas 58.3 percent of them claimed that no such goals existed.

Quantitative Research Results of General Public

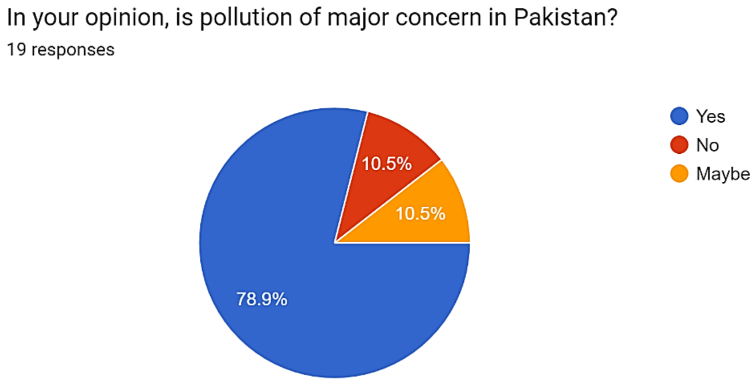


Figure 12: Exhibiting Results of Responses from the General Public to the Question, "In Your Opinion, is Pollution a Major Concern in Pakistan?"

To this question 78.9 percent of the general public claimed that pollution is a major concern in Pakistan, whereas 10.5 percent of them claimed that no, it is not a major concern.

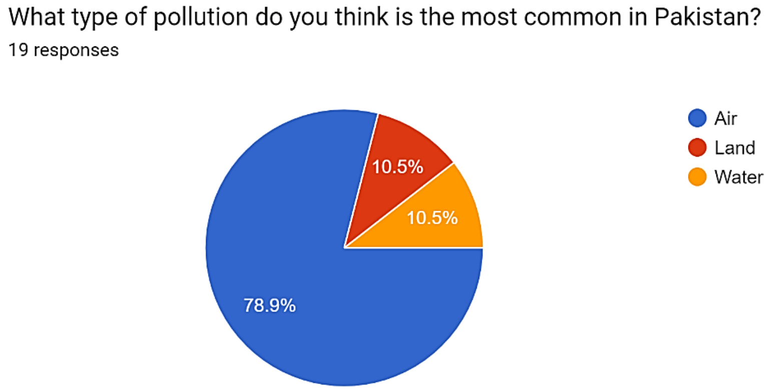


Figure 13: Exhibiting Results of Responses from the General Public to the Question, "What Type of Pollution do you think is the most common in Pakistan?"

To this question 78.9 percent of the general public claimed that Air Pollution is the most common, whereas 10.5 percent of them claimed that Land Pollution is common, and the rest 10.5 percent claimed that Water Pollution is the most common type of pollution in Pakistan.

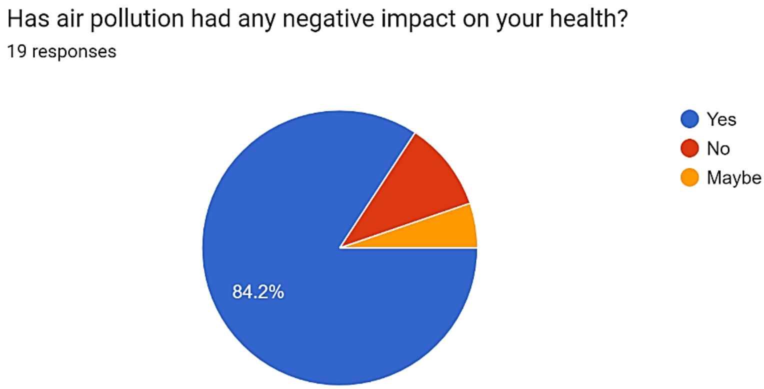


Figure 14: Exhibiting Results of Responses from the General Public to the Question, "Has Air Pollution had Any Negative Impact on Your Health?"

To this question 84.2 percent of the general public claimed that yes it has an effect, whereas 7 percent of them claimed that no it does not have any impact on their health.

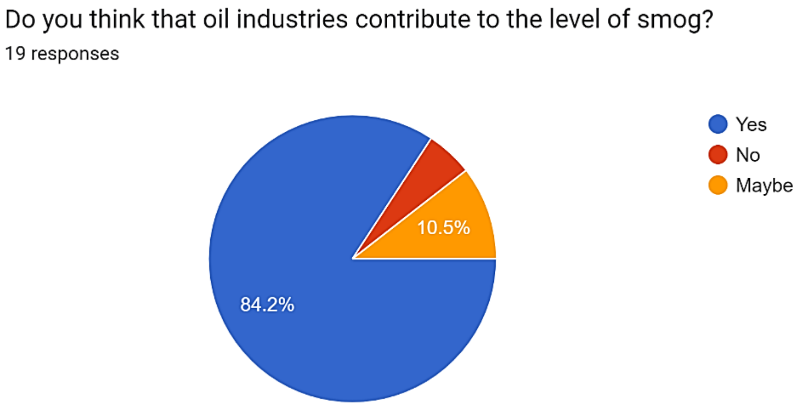


Figure 15: Exhibiting Results of Responses from the General Public to the Question, "Do You Think that Oil Industries Contribute to the Level of Smog?"

To this question 84.2 percent of the general public claimed that yes oil industries contribute to the level of smog, whereas 10.5 percent of them were indifferent.

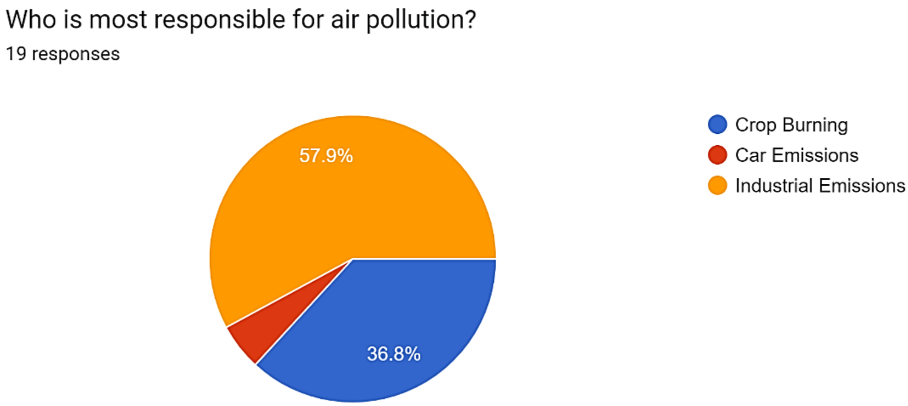


Figure 16: Exhibiting Results of Responses from the General Public to the Question, "Who is Most Responsible for Air Pollution?"

To this question 36.8 percent of the general public claimed that crop burning is responsible for air pollution, whereas 57.9 percent of them claimed that industrial emissions are responsible, and the rest said that car emissions are responsible for air pollution.

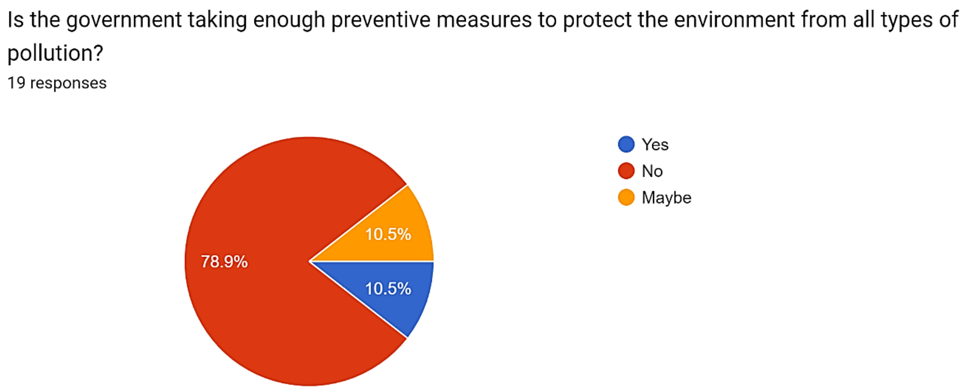


Figure 17: Exhibiting Results of Responses from the General Public to the Question, "Is the Government taking Enough Preventive Measures to Protect the Environment from All Types of Pollution?"

To this question 78.9 percent of the general public claimed that government is not taking any preventive measures to protect the environment , whereas 10.5 percent of them claimed that yes they are taking preventive measure, and the rest 10.5 percent claimed that maybe they are taking any measures.

4. COMMENTS AND CONCLUSION

From the results and discussion it is evident that wastewater from edible oil and ghee factories contains by-products from processes such as oil, nickel, and other similar substances and has high COD. This wastewater is released into watercourses and rivers without undergoing any treatment and with no restrictions. It was also used for irrigation purposes by farmers in an effort to save money, but the negative repercussions of this practice last for a number of years. In addition. It turns the surface of the vegetation into a sort of plate on which pollutants and parasites might interact. It has an effect not just on bodies of surface water and also on bodies of groundwater at shallow depths. It was observed that the plants grown in places that are irrigated with industrial effluents have a lower amount of chlorophyll and carotene than those grown in areas that are not polluted. This was compared to the plants grown in locations that are not polluted. Several types of medicinal plants, including Lepidium sativum, Silybum marianum, and Ammvisnaga, were discovered to have higher quantities of cadmium, nickel, lead, iron, and copper when they were found to be growing in highly polluted and industrial locations. A significant amount of nickel that is discharged into the environment has cumulative effects, which can lead to it making its way to sediments or soil, where it becomes strongly attached to particles containing iron or manganese, and, if it is provided with acidic conditions, it eventually finds its way to ground water. The development of woody plants is similarly stunted when there is a high concentration of nickel in the soil. It is also possible for it to enter the food chain [4]. The presence of oil and grease in wastewater disrupts the process of photosynthesis by forming a thin film on the surface of the water. This film prevents the transmission of sunlight through the water, which is necessary for photosynthesis. This ultimately results is the destruction of the aquatic ecosystem and creation of an imbalance in the system. The interference of some constituents of the oil and grease, in excessive proportions, with aerobic and anaerobic biological processes is another factor that contributes to the lower efficacy of waste water treatment. In addition, a high concentration of it in the wastewater can prevent the receiving bodies of water from properly absorbing oxygen, which can lead to a decrease in the amount of dissolved oxygen and an increase in the risk to aquatic life. The total suspended solids (TSS) are composed of both settleable and non-settleable materials, both of which contribute to the turbidity of the water and consequently reduce the percentage of light. As a consequence, the process of photosynthesis is disrupted, which leads to the death of plant life, which, in turn, upsets the aquatic life because it is dependent on these plants. When these floating particles settle, not only do they harm fish reproduction and organisms that fish need as their primary food source, but they also have the potential to obstruct the fish's gills, which ultimately results in the fish's death.

The study thus concludes that the effluent from oil industry has significant negative effects on the plants and the population in the surrounding area and strongly recommends that the environmental control authorities and oil industrialist should take essential steps to save both employees and public from its impacts.

5. ACKNOWLEDGEMENT

We are extremely grateful to Dr. Shahid Amjad Chaudhry, Rector Lahore School Economics for patronization of our research and sanction of expenditure involved and providing us requisite facilities to participate in 20th ISOSS Conference held in Lahore

Appendix: Interview Questions and the Responses against the Questionnaires

Interview questions for Executive:

1. How important do you think the oil industry is for Pakistan
2. What is the potential of this industry
3. Where do you see the industry in the next 10 years?
4. Do you think the Oil and Ghee industry is causing some sort of harm to the environment?
5. Do you think dumping of leftovers in Water bodies is a common practice in the industry?
6. Do you think the dumping of waste in water should be stopped?
7. What is your take on the ongoing spree of smog?
8. What are you doing as an organization to preserve the environment
9. Do you plan to switch to sustainable ways of oil production in the near future?
10. Does your organization prioritize corporate social responsibility?

Questionnaires

Employees of Kohinoor Oil Industries

- Name
- Occupation
- Age
- In which industry or company are you working at currently?
- Where do you see the oil industry in Pakistan in the next 10 years

- Do you think the oil industry is a major contributor to pollution in Pakistan
- Are you provided with safety gear by your organization?
- To what extent do you think that the working conditions in your organization are optimal.
- Do you think that your organization abides by the waste dumping laws set by the government of Pakistan?
- Has the recent impact spree of air pollution in the form of smog affected you adversely?
- Since the Air Quality of Pakistan is deteriorating rapidly, has your organization taken any measures to reduce its emissions?
- Does your company have any mission or goal of adopting sustainable ways in the near future?

General Public

- In your opinion, is pollution of major concern in Pakistan?
- What type of pollution do you think is the most common in Pakistan?
- Has air pollution had any negative impact on your health?
- Do you think that oil industries contribute to the level of smog?
- Who is most responsible for air pollution?
- Is the government taking enough preventive measures to protect the environment from all types of pollution?

REFERENCES

1. Afzal, M.Z., Ikram, M. and Israr, S. (2016). Study of the pollutants in wastewater from edible oil/ghee industries and their impacts on plant life, Islamabad, Pakistan. *Environmental Science, An Indian Journal*, 12, 112-117.
2. Environmental Technology Program for Industry (ETPI) (1999). *The Edible Oil & Ghee Sector: Environmental Report*. Online available at https://pdfcoffee.com/the-edible-oil-amp-ghee-sector-environmental-report-pdf-free.html

Websites Visited

- https://www.tsijournals.com/articles/study-of-the-pollutants-in-wastewater-from-edible-oilghee-industries-and-their-impacts-on-plant-life-islamabad-pakistan.pdf
- https://www.pacra.com/sector_research/Edible%20Oil%20-%20PACRA%20Research%20-%20Feb'21_1612876391.pdf
- https://www.researchgate.net/publication/279545575_The_vegetable_oil_industry_in_Pakistan
- https://www.tsijournals.com/articles/study-of-the-pollutants-in-wastewater-from-edible-oilghee-industries-and-their-impacts-on-plant-life-islamabad-pakistan.pdf
- https://toaz.info/doc-view
- https://www.ukessays.com/essays/environmental-sciences/oil-and-ghee-industry-pollution-environmental-sciences-essay.php
- https://link.springer.com/article/10.1007/s13201-019-0920-1
- https://pdfcoffee.com/the-edible-oil-amp-ghee-sector-environmental-report-pdf-free.html

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 73-82

MINIMIZING THE CARBON FOOTPRINT AS WELL AS THE PROBLEMS RELATED TO GREENHOUSE GAS EMISSION BY ASIAN CHEMICALS

Mohammad Rafiq Khan[§], Bisma Nawaz, Mohammad Ahmad Taha
Ammar Asad and **Haider Khokhar**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

The article being presented here is a component of the pilot project, "The Environmental Problems of the Industrial Sector of Pakistan" conceived by our Class of Researchers composed of different groups. Each group was deployed to focus on one industrial unit to highlight its environmental problems and make recommendations to solve them. The article being presented here focused on, "Minimizing the carbon footprint as well as the problems related to greenhouse gas emission by Asian Chemicals." Both secondary and primary data were collected. A survey was conducted using Google docs with a sample size of 52. Questionnaires were circulated among the students of The Lahore School of Economics. An employee survey was also crafted and circulated among the employees of Asian Chemicals and interviews were conducted from the manager and the employees of the industry. This helped gather deeper insights of the company and gauge their understanding of the carbon footprint of Asia Chemicals. The results concluded that the public had knowledge regarding pollution and GHGs emission while the employees didn't possess enough knowledge regarding the problem. It is yet to be defined if this was an act of denial or actual naivety. A few recommendations were proposed with respect to the research conducted that the company should plant more trees in the surrounding area to counter the Greenhouse gas production. Renewable resources of energy can be used for the fuel and the industry should switch to green hydrogen.

1. INTRODUCTION

Asian Chemicals Ltd. was founded in 1981 as a pioneer in Pakistani chemical manufacture. They were the first company to create a Star Brand which had branded high-quality precipitated calcium carbonate. Their goal was to develop a high-quality product utilizing the most advanced tools and methods at the time in order to supply the vital raw materials to a variety of sectors of the Pakistani economy. This was done to help reduce expenses and provide a product of an international standard for the local market.

It is one of the top producers of precipitated calcium carbonate (PCC), hydrated lime, and calcium oxide in the country today. They have managed deliver superior products to be used in a variety of applications by basing their business operations on the concepts of innovation, unwavering commitment to continuous improvement, and rigorous adherence to quality standards. They provide chemicals to paper, detergent, paint, plastic, glue, and shoe industries. The company has done a number of certifications to ensure that quality standards are met while simultaneously reducing hazardous gases from the factory. They have also been regarded as one of the most responsible chemical factories in terms of employee safety.

These industrial processes generate voluminous hazardous waste, toxic gaseous pollutants and other health-injurious emissions like smoke and dust. Some earlier studies conducted in the current context have also been reported in literature (Burningham, & Thursh, 2010 and Hosseine, et al., 2013, Jaspal, et al., 2014, Awan, et al., 2019). Sadly, there is no understanding of pollution prevention and improving control of pollutants as far as the industry, primarily in private sector, is concerned. As a result, toxic waste is dumped and released to cleaner Industrial or in water bodies, unchecked by the authorities and other stakeholders (Mehdi, 2019).

In a study by (Hussain & Rehman, 2022) the impact of CO_2 emission on livestock was studied whereby it was found out how CO_2 emissions in Pakistan affected cattle output by using annual data that from 1972 to 2019. Another study was carried out in which industrial effluent of Hattar Industrial Estate was analyzed. According to accepted practices, all the samples were evaluated for pH, temperature, TS, TDS, TSS, EC, DO, BOD, COD, turbidity, nitrates (NO_3-N), phosphates (PO_4), ammonia (NH_4-N), and other heavy metals. Results showed that these samples exceeded the quality standards and were heavily contaminated (Mahmood, et al., 2019). A case study in a heavy industrial area, China was carried out in Taiyuan as to quantitatively investigate the links between industrial pollution and ecosystem service value. Findings demonstrated a downward-upward-downward trend in the pollution level of industrial wastewater, a downward-upward-downward trend in the pollution level of industrial waste gas, and a downward-upward trend in the pollution level of industrial solid waste (Ji & Ma, 2022). Apart from the studies referred above, some studies have been conducted in context of climatic change (Minovi, et al., 2022), Cleaner production of Pakistan's chemical industry: Perspectives of energy conservation and emissions reduction (Lin and Raza, 2021) and of Environmental impacts of hazardous waste, and management strategies (Zhang, et al 2022),

2. RESEARCH METHODOLOGY

The research paper primarily focuses on the increased carbon footprint and greenhouse gas emission by Asian chemicals, and its impact on the environment. For this purpose, a mixed method approach was utilized. In the first half, quantitative research was conducted to have a clear understanding of the problem. In the second half, quantitative research was conducted to support our qualitative findings.

Secondary data was collected in which a number of articles were retrieved from Jstor, Google Scholar, Tandfoline and Research Gate. Moreover, newspaper articles were critically analyzed to ensure that reliable data was used. The research gap, findings and conclusion were considered in writing the literature review.

Surveys and interviews were used in the primary research. The two additionally acted as our study's research tools. The survey was further categorized into two categories i.e. employee survey and public survey. The survey, which was also the first primary tool, served as the preliminary step for the quantitative section of the research paper. We gained valuable insight into people's perspectives on the subject through to the surveys.

Using Google Docs, a questionnaire with 14 questions was created. While keeping a sample size of 52 persons, the survey was circulated via WhatsApp. The survey was sent at random and did not target any specific gender. As there was no specific age restriction for this survey, it was anticipated that the majority of respondents would be between the ages of 18 and 22. The survey asked both judgment-based and subjective questions. This aided us in exploring the topic from a wider angle. The respondent was questioned on his or her awareness of pollution, the carbon footprint, and its effects. There were also questions about pollution causes and health- related concerns. The questionnaire's results were used to create statistical pie charts and other tools. This also aided us in justifying our research.

Interviews were the second tool utilized. Ten semi-structured questions were asked of the interviewee during a single interview. This was done to generate detailed responses and to gain more knowledge about the company's hazardous gas emission and its effects. The semi-structured interview questions also served the purpose of getting more in-depth responses from the interviewee rather than generic ones. Over the phone, the interview was recorded. To draw a conclusion, the interview was transcribed and critically analyzed. It helped us gain a deeper understanding of the subject and was an important tool for gathering qualitative data, while the surveys' sample sizes helped us achieve significant generalizability.

Employee and Executive Survey:

As for the employee survey, gender was restricted to males and females. Their respective designations were also asked in the survey to evaluate their level of knowledge regarding the matter. A single interview was conducted of the factory manager which had 15 years of experience in the field.

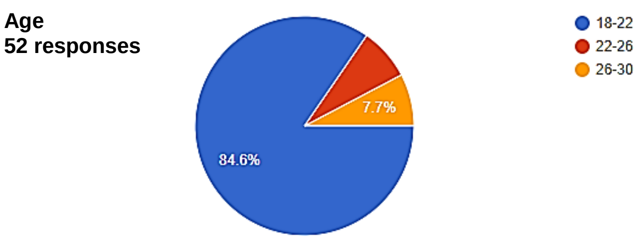
Public Survey:

For the public survey the sample size of 55 respondents was kept for the survey that was in the age bracket of 18-30 years. The target audience for this study was 18-24 year olds. Furthermore, possible gender biases were eradicated by keeping the survey open for all genders.

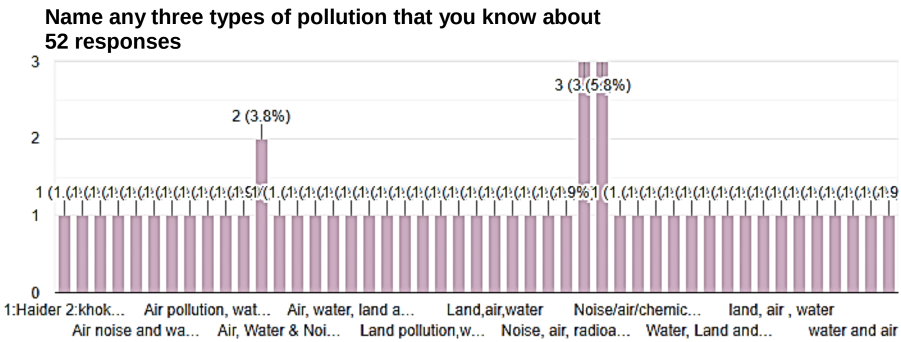
3. RESULTS AND DISCUSSION

1. Survey Analysis- Public

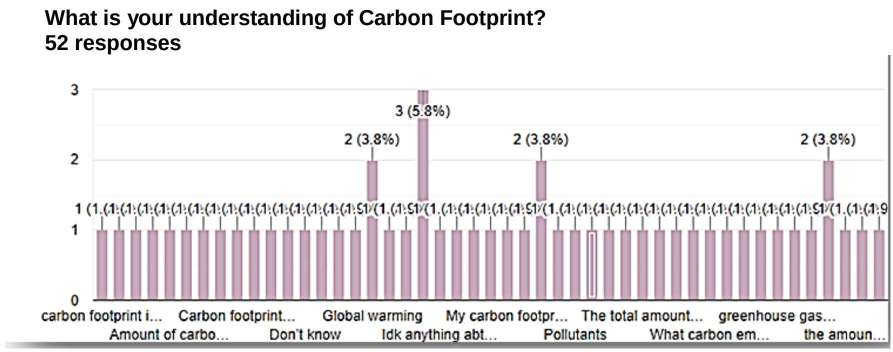
As far as the results are concerned, the findings from our research were almost on the same lines as we expected them to be. Following are the results from the public survey which we conducted.



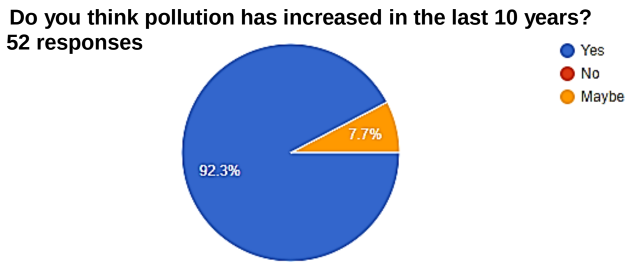
The public survey had 52 respondents where, an equal ration of males to females was observed. This equal number of males and females allowed us to eradicate any potential gender biases from our study. Majority of the respondents fell under the age bracket of 18-22. Majority of our respondents, 57.7% were currently enrolled in their undergrad with 9.6% pursuing their postgraduate.

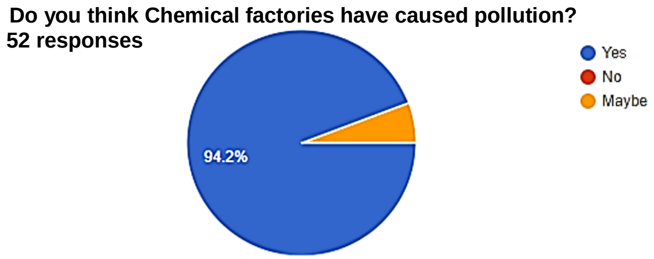


The public was asked about pollution. A variety of responses were recorded which showed signs of understanding among the respondents. Each respondent believed that pollution was "something harmful that disturbed the environment" and event poisonous to some degree. When asked to list down three types of pollution, all the respondents listed air, land and noise, this shows that the sample was well aware of the environmental problems and issues in the environment.

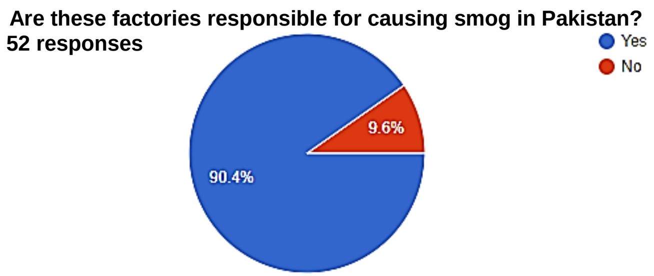


The respondents were asked about their knowledge regarding carbon footprint. 32 out of 52 respondents were well are of the word. The common statement was that it is the release of CO_2 or greenhouse gases which deteriorates the environment. They also showed sound understanding of the negative impact related to carbon footprint.

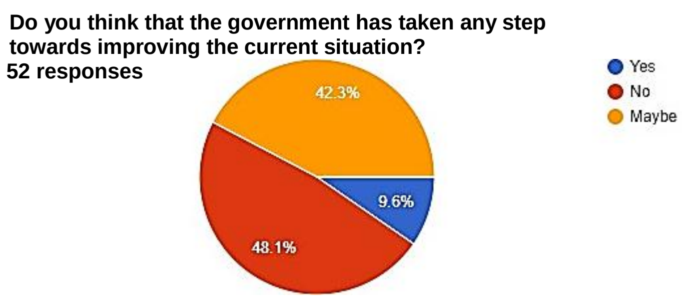




Furthermore, we asked the public about the increase in pollution over the last ten years, 92.3% people believed that pollution has increased significantly in the last 10 years, while 7.7% people chose the option “maybe” as they weren’t sure if pollution had increased or not.



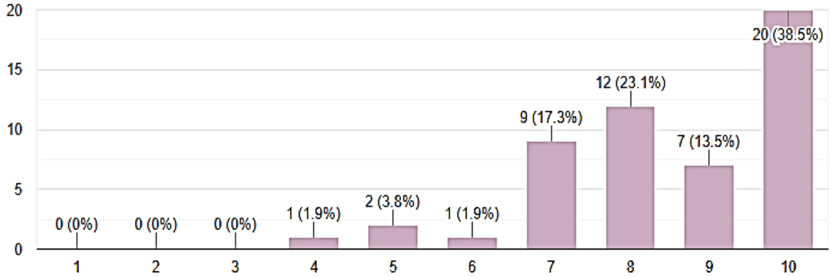
Another question was designed to find out that does chemical industry contribute to the pollution of our country? About 94.2% of the respondents said that chemical factories were causing pollution. The same percentage of respondent believed that these factories released harmful gases to the environment which the remaining 6% unsure.



Respondents were asked whether the factories are causing smog in Pakistan. 90.4% of them believed that these factories are responsible for causing smog. 9.6% of the respondents didn't agree with it.

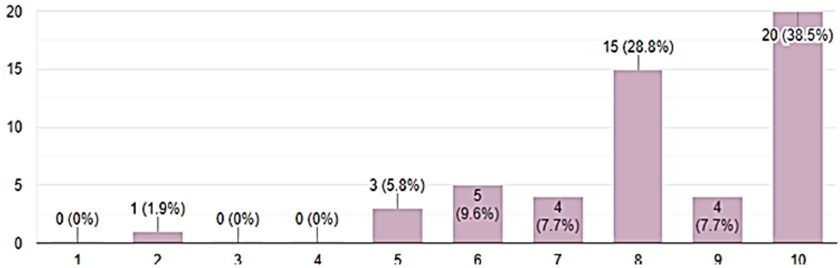
When asked if the respondents were aware of any measures that the government is taking towards improving the current situation in terms of deteriorating the environment, only 9.6% of the people believed that the government was taking steps however, 48.1% of the respondents said no followed by 42.3% of the respondents who said maybe.

On a scale of 1-10, how badly has waste from these factories effected the marine life?
52 responses



The respondents were asked to rate the extent to which chemical factories had effected marine life on a scale of 1 to 10, 38.5% of the respondents rated it 10. Only 4% of the respondents rated it below 5 with nobody rating it at 1, 2 or 3.

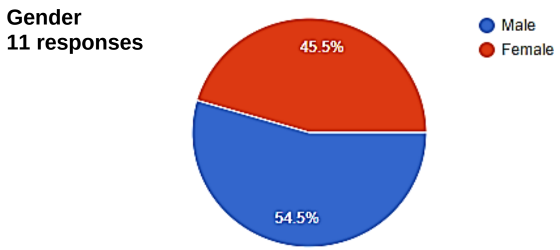
On a scale of 1-10, how badly has waste from these factories effected the land?
52 responses



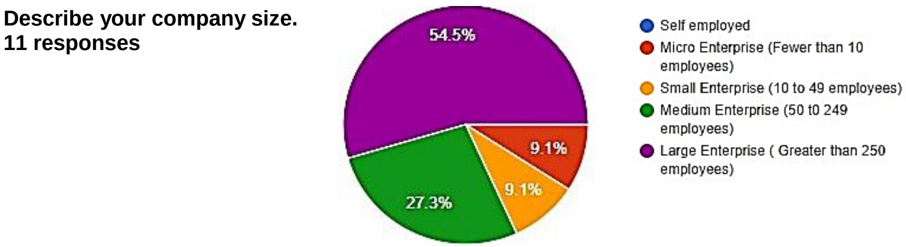
Similarly, the effect of land was asked where the exact same results were found.

2. SURVEY ANALYSIS- EMPLOYEES

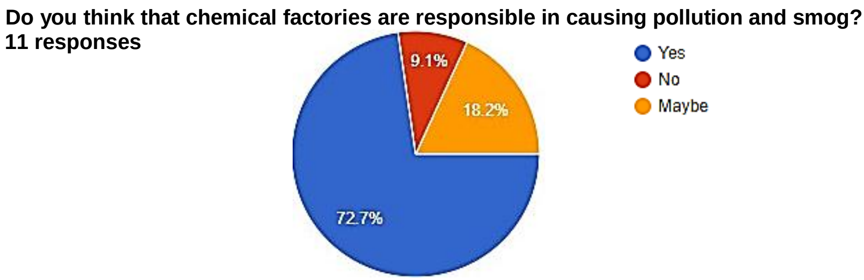
An employee survey was conducted for the employees at Asian Chemicals. We were only able to get the survey filled from 11 employees.



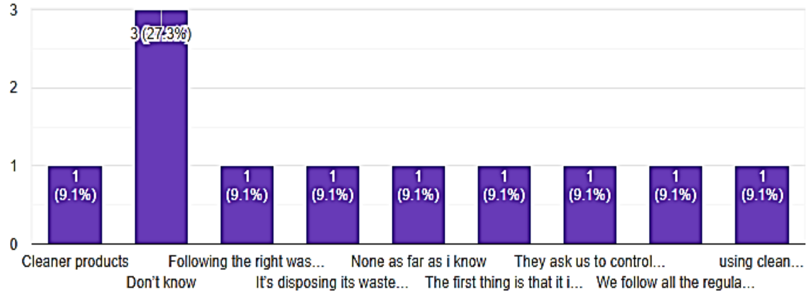
The first question was regarding the employee's gender. Since, it is a factory the percentage of males was higher than females. 54.5% of the employees were males while 45.5% females. The designations of the respondents were: 2 machine operators, production plant assistant, manager, 5 factory workers, 1 technician and 1 secretary.



The respondents were further asked about their assumption of the company size to which, 54.5% said that the company had greater than 250 employees, while 27.3% people said that the company had 50 to 249 employees less than 27.3% and 9.1% thought that Asian chemicals had less than 50 employees.

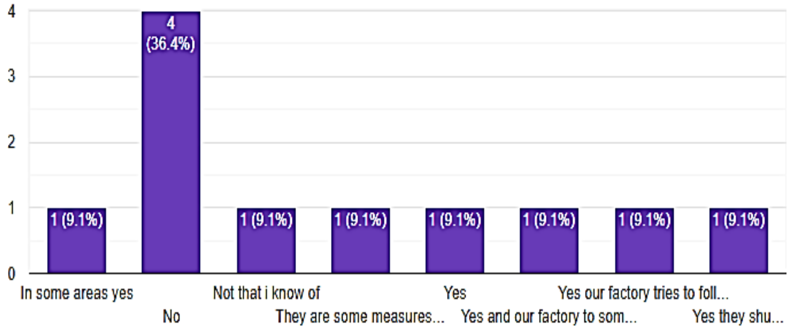


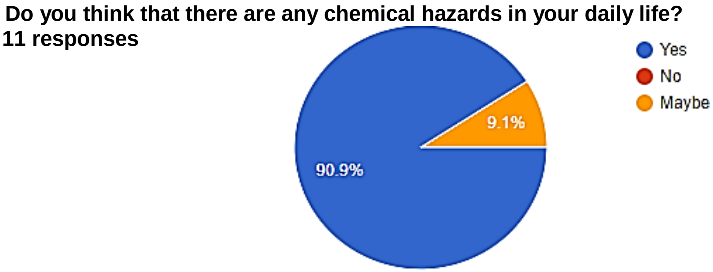
What measure is the management taking in reducing the carbon footprint and greenhouse gas emission from your factory?
11 responses



When asked about the employees understanding of carbon footprint, none of the employees showed proper understanding of the word. This result was shocking as these people were working in a chemical factory and showed no understanding of such an important principle. However, majority of the employees, 90.9% of the respondents believed that they were strongly exposed to chemical hazards themselves. What measures is the management taking in reducing the carbon footprint and greenhouse gas emission from your factory? 9.1% of the employees said that the factory was using cleaner products which weren't impacting the environment in a negative manner. 36.2% of the respondents believed that the factory was taking extensive measures in terms of waste management. 27.3% of the respondents were not aware of the steps being taken towards reducing carbon footprint. 18.6% of the employees added that the location of the factory was an answer itself as the factory was located far away from urban area.

Are there any government regulations in minimizing the hazardous toxic emission and is your factory following them?
11 responses





The next question was whether chemical factory was responsible for causing smog or other kinds of pollution. 72.7% of the employees responded yes, this shows that the employees at least accepted that the chemical wastage and gases were the causing pollution.

The employees were asked about the chemical hazards in their daily life and about 90.9% said that there were chemical hazards in their daily life and 9.1% said maybe.

We then asked them if there are any government regulations in minimizing the hazardous toxic emissions and is your factory following them. 36.4% of the respondents said no while 64.1% of the respondents said yes and added that if the factory didn't follow the regulations set by the government then the factory can be shut down thus, Asian chemicals strictly follows the regulations listed by the government of Pakistan.

An interview of the manager of Asian chemicals was conducted. The Manager was asked about his understanding of carbon footprint. The manager showed sound understanding of the term and emphasized that the carbon footprint should be taken into account on an individual basis than blaming companies only as we use our mobile phones and our equally responsible is harming the environment.

The interviewee was asked if he was aware of the ways through which Asian chemicals could reduce the emission of hazardous gases and reduce its carbon footprint. Asian chemicals weren't currently doing anything in this capacity however, it could shift from coal as a source of energy to biomass and it could also shift to green hydrogen.

4. COMMENTS AND CONCLUSION

As a result of how greenhouse gases affect people's lives and the environment, it is clear from our research that they have caused people a great deal of suffering. Chemical industries are ignorant of how they contribute to environmental degradation and climate change. In comparison to these factories, the government is more accountable for environmental damage because there are no established SOPs by the government. Even if there are any, the factory owners are not informed about them. The negative effects of the techniques used in chemical industries are not commonly known to the workers. Government agencies and industrial owners working together can reduce risks and improve people's quality of life.

RECOMMENDATIONS

A number of steps can be taken to minimize the impact of greenhouse gases being released by Asian Chemicals. They are mentioned as follow:

- The company can plant more palm trees around the factory area as well as the nearby village so that the same amount of oxygen is produces to counter the GHG's production.
- Renewable resources of energy can be used for fuel.
- Invest in research and development to find energy saving plants to go green.
- Switch to green hydrogen completely.
- As most employees were unaware of the hazards, thus conduct regular training sessions to enlighten the employees regarding potential hazards and ways to reduce them.
- A minimum education requirement should be set and a thorough screening process should be used while recruiting people as many of the employees at Asia Chemicals do not possess the right abilities to be at this job.

REFERENCES

1. Mehdi, A.M. (2019). Industrial pollution in Pakistan. Published in *The Nation* dated January 7, 2019.
2. Awan, A., Lodhi, M.U. and Hussain, D. (2018). Determinants of firm value: A case study of Chemical Industries of Pakistan. *Global Journal of Management, Social Sciences and Humanities*, 4(1), 46-61.
3. Burningham, K. and Thursh, D. (2010). Pollution concerns in context: a comparison of local perceptions of the risks associated with living close to a road and a chemical factory. *Journal of Risk Research*, 7(2), 213-232.
4. Hosseini, E.S., Wahid, M.A. and Aghili, N. (2013). The scenario of greenhouse gases reduction in Malaysia. *Renewable and Sustainable energy reviews*, 28, 400-409.
5. Hussain, I. and Rehman, A. (2022). How CO_2 emission interacts with livestock production for environmental sustainability? Evidence from Pakistan. *Environment, Development and Sustainability*, 24(6), 8545-8565.
6. Jaspal, Z.N. and Haider, N. (2014, July). Management of chemicals in Pakistan: Concerns and challendes. *A Research Journal of South Asian Studies*, 29(2), 497-517.
7. Ji, S. and Ma, S. (2022). The effects of industrial pollution on ecosystem service value: a case study in a heavy industrial area, China. *Environment, Development and Sustainability*, 24(5), 6804-6833.
8. Mahmood, Q., Shaheen, S., Bilal, M., Tariq, M., Zeb, B.S., Zahid Ullah, and Ali, A. (2019). Chemical pollutants from an industrial estate in Pakistan: a threat to environmental sustainability. *Applied Water Science*, 9, 1-9.
9. Minovi, D. (2022). *Why the chemical industry is an overlooked climate foe and what to do about it*. Online available at https://therevelator.org/chemical-industry-climate/
10. Lin, B. and Raza, M.Y. (2021). Fuels substitution possibilities and the technical progress in Pakistan's agriculture sector. *Journal of Cleaner Production*, 314, 128021.
11. Zhang, Z., Malik, M.Z., Khan, A., Ali, N., Malik, S. and Bilal, M. (2022). Environmental impacts of hazardous waste, and management strategies to reconcile circular economy and eco-sustainability. *Science of the Total Environment*, 807, 150856.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 83-96

ENVIRONMENTAL SITUATION OF HEAVY ENGINEERING INDUSTRY (PUNJAB STEEL INDUSTRY)

Mohammad Rafiq Khan[§], **Aeman Ashraf, Zohaib Javed**
and **Ghulam Murtaza**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

Pakistan is facing n number of environmental problems for which appropriate solutions are quite requisite. The industrial sector particularly needs attention for disposal of industrial wastes on priority in context of seeking solutions of the environmental problems of Pakistan. Because for solving a problem, it is the basic requirement what is the situation of the subject for which the solution is to be worked out, a pilot project," The Environmental Problems of the Industrial Sector of Pakistan" was conceived by our Class of Researchers composed of different groups while each group was deployed to focus on one industrial unit to highlight its environmental problems and make recommendations to solve them. The article being presented here studied, "Environmental Situation of Heavy Engineering Industry (Punjab Steel Industry). Both secondary and primary data were collected. The former was gathered by, the use of previously published works in this area of concern and later by collecting responses against the questionnaires served to the respondents and interviews conducted through successive visits to the establishments under study. The data were computed and statically analyzed by application of Google form. The results indicating the extent the steel industry has contributed to the environmental degradation are reported. The study also come up with innovative solutions that encourage the steel manufactures to give up the traditional furnace technology and switch towards the new electrical furnaces which are considered to be environment friendly to a great extent.

1. INTRODUCTION

The Steel Industry is deemed to be an integral part of Pakistan economy as it immensely contributes to the GDP of the country (4%) along with the high rate of employment (around 100,000 employees nationwide). The steel industry is mostly located in the outskirts of Karachi, situated near to the Bin Qasim seaport. This is followed by Lahore, which is said to have the second largest steel industry, located in Shahdara, Multan Road and the Badami Bagh area. The state-owned steel mill, known by the name of Pakistan Steel Mill is said to produce around a million ton per annum whereas the setups from the private sector produce over 30 million tons of steel per annum. The country's economic and infrastructural growth heavily depends upon the outputs of the steel industry

In spite of all the benefits that steel industry brings to the country, it seems unfortunate that, on the contrary, the steel industry turns out to be a major polluter of the environment and the natural ecosystems. It would not be take long until the environmental implications caused by the industry will overweigh the benefits it brings to the country's economy. The carbon emissions released from the industry have significantly increased with the growing industrial sector. The steel industry requires heavy burners to mold the raw material which exhaust poisonous gasses into the earth's atmosphere without any prior treatment. Steel production leads to the release of iron oxide, hydrocarbons, nitrogen dioxide, and carbon monoxide which all are well proved pollutants of the environment. The story does not end at air pollution because these industries also contribute to water pollution as organic matter, suspended solids, sulfate cyanides and oil metals that are dumped into the nearby water bodies, rendering the quality of water unfit for human use and the marine life in danger.

The economic challenges faced by Pakistan have left a huge question mark on its future and in order to counter this state of affairs, it has become obligatory to encourage an increase in the establishments of steel industries. This urge for quick economic growth is leading the country towards unplanned industrial growth, with the industrial units emitting toxins into the atmosphere and the water bodies. In Pakistan, the large players of steel, use coal and natural gas to mold steel and many small-scale steel industries depend upon the burning of tires as source of fuel. The burning of tires puts forth environmental hazards to a greater extent as harmful substances such as chromium, mercury and lead are released into the air (Ali, M). Tire smoke is much more harmful than the smoke created by coal as it contains 407 percent more chromium and is 1448% more arsenic than coal 1997 (An, et al. 1997).

Currently, Pakistan has been placed in the list of countries that are severely affected by smog. Smog refers to a kind of air pollution in which fog or haze is intensified by smoke and other atmospheric pollutants. This reduces the visibility and deteriorates the quality of air that we breath in. Lahore and Karachi considered to be the two largest cities of Pakistan, both in terms of land and population. However, this is not the only similar characteristic that they both have in common as both the cities are locking horns in to become the most polluted cities in the nation. Recently Karachi unlocked a reprehensible achievement by becoming the most polluted city in world followed by Lahore which stood fifth on the list. This has caused serious health concerns, such as lungs, eye, skin, and heart disorders. Although the government has taken some steps to lessen these harmful pollution impacts, these have turned out to be insignificant. The public's assistance will be crucial in overcoming this problem, and major efforts are still required. This rapid rise in pollution if not contained will aggravate health hazards and their consequences would be catastrophic.

In the process of steel making manufacturers use natural gas without which it is difficult to complete many stages of steel production. Natural gas is 90% methane that burns at a temperature of 650 Celsius. The combustion of natural gas also produces a high amount of Sulphur, mercury, and nitrogen dioxide which cause smog. These pollutants create numerous health and environmental hazards. Methane is present in the earth's atmosphere, If present above the standard limit, it can cause many health hazards. It is a greenhouse gas that causes an increase in the earth's temperature and is considered 80 times more effective at warming over a 20-year period than is carbon dioxide. A high exposure to methane often leads to adverse health outcomes that have the potential to take multiple lives. Sulfur dioxide is the chemical compound with the formula SO_2. It is a toxic gas released naturally by volcanic activity and is produced as a by-product of copper extraction and the burning of sulfur-bearing fossil fuels. It is a major polluter of the environment with numerous negative effects such as wheezing along with shortness of breath. The tightening of the chest and the distortion of the lung function to cause asthma. Nitrogen dioxide, or NO_2, is a gaseous air pollutant formed when fossil fuels such as coal, oil, gas or diesel are burnt at high temperatures. Nitrogen dioxide can produce nitrate secondary particles, which can lead to haze and decreased visibility.

Induction Furnace is a type of system which is used by Steel mills to produce heat which is used to melt the metal. The furnace melts all metals such as iron, steel, copper and aluminum. This is a well-controlled melting process which is highly energy efficient and effective at the same time. Most of the steel industries use coal and oil to operate the furnaces, which cause numerous environmental hazards. The coal plants and oil are responsible for mercury emissions, a toxic heavy metal that is capable of damaging the nervous, digestive, and immune systems, and is a serious threat to the child development. The particulate matter, also known as "soot," is the ashy substance in coal smoke, and is linked with harmful health implications which are caused after it mixes with to cause the risk of chronic bronchitis, aggravated asthma, cardiovascular effects like heart attacks, and premature death.

Some research studies have been conducted on Pakistan that through light on Environmental state of Pakistan and repeatedly warn Pakistan and its nationals to take effective measures to correct the alarming situation to avoid catastrophic results in future.

Younes, et al. (2020) reported the results of their study on how the emissions released from iron and steel industries in Pakistan has negatively impacted the health of the locals. The research paper lays emphasis on the relationship between the iron and steel industries in Pakistan causing air pollution which is adversely affecting the public health. In order to conduct the desired research, the authors targeted two population samples in which 36 industries were included for emission sampling, whereas, around public respondents were chosen from a nearby residential area. The results indicated that the cause of concern was the amount of pollutants was far above the threshold. It was also concluded that emissions from these industries were well above the standard values set by the Environmental Protection Agency in Pakistan. The air quality index of the areas was evaluated and it was found out that the air quality was below standard and at a dangerous level, where locals were prone get affected by numerous illnesses. The results also concluded that most people were affected by the industrial smoke and were suffering from various kinds allergies such as respiratory problems, high pulse rates, shortness of breath and cancer in some serious cases as well.

Mehdi published an article "The news Pakistan" that laid emphasis on Industrial pollution being the worst environmental hazard that the country had ever witnessed. He further raised the concern about industrial pollution creating a havoc both environmentally and in terms of public health. He claimed that industrial pollution contributed to around 60 percent to the total population in Pakistan. He predicted that a massive environmental degradation was ahead for the country if it fails to take a requisite initiative. The writer described how Pakistan's industrial sector was dominated by steel, leather, textile, cement, automobile and pharmaceutical producing industries that all emit poisonous gasses into the earth's atmosphere and toxic waste in water bodies. He raised a question mark upon how the private sector was only concerned about profits and was totally negligent towards the environmental disaster. The writer has also talked about how the social and environmental aspects in Pakistan were worsening as he states that Pakistan is among the 12 out of 15 Asian countries where the rate of industrial pollution has reached to a highly unsafe level. He furthers adds that each year more than 7 million people are admitted to the hospitals due to illnesses caused by pollution.

The writer highlighted the massive surge in water pollution as well as stated that as per the Pakistan Council of Research in Water Resources, the water available for drinking was infected with harmful chemicals due to the dumping of industrial waste in the water bodies, and therefore was not fit for consumption.

Another research study elaborates the challenge that Pakistan is facing in face of air pollution majorly caused by industrial emissions. The article focuses upon Pakistan being the country having the 5^{th} most polluted air in the world caused mainly by industrial emissions and crop leftover burning (Greenstone and Fan, 2019). The authors evaluated that if the rate of pollutions stays the same in Pakistan, each of its citizens would have 2.5 years of their lives cut off of the regular life expectancy, as computed by the World Health Organization. Secondly, The Lahore city suffers the most, with average citizens losing 5.3 years off of their lives in comparison to the WHO estimate. This analysis demonstrates that Pakistan has the chance to significantly improve its air quality only if the government starts to address the issue.

The articles sheds light upon how the population of Pakistan is on the surge and in order to meet the demands of the increasing population, the industrialization and urbanization is also increasing which has polluted the air and is constantly deteriorating its quality. As a result, there has been a noticeable increase in the variety of sources of air pollutants. However, due to a lack of air quality control tools, the nation's air quality has got worse and is on the downward trend. Even though the Pakistan Clean Air Program was initiated, the air quality index in Pakistan remains to be pathetically low as the quality of air is severely degraded due to industrial pollution. The writers repeatedly shed light on sustainable measures to contain air pollution being the need of the hour.

After carrying out the review of literature, it was conceived that the studies targeted the general environmental problems of Pakistan and held some industries of Pakistan responsible for the environmental deterioration in the Country. |Thus the gap encountered was that no comprehensive study was conducted to highlight the Environmental Situation of the Industrial Sector of Pakistan. Thus a large group of researchers was deployed on different industries in Lahore to study the environmental situation of industry as models for later interpretation of results at the national level. This project focused on Punjab Steel Industry as a model to interpret the environmental situation on national level.

The objectives of this research are listed below:

1. To understand the impact of industrial waste from heavy engineering industries, primarily steel industries, on the environment.
2. To understand whether the government or industry owners are aware of their carbon footprint, emission of greenhouse gases and other pollutants.
3. To know whether the government or the factory owners are taking steps towards reducing the toxins being released.

2. RESEARCH METHODOLOGY

In order to conduct this research based on the steel manufacturing industry having a negative toll on the environment and the public, a number of research methods and tools were used to obtain required data. The research included the both primary and secondary data which was used to obtain results regarding the problem and to draft a conclusion based on the results.

Research Tools

As stated above both primary and secondary data s used. In order to get access to prior published work on the topic to collect data, a series of research papers and reports were collected by consultation of literature in the libraries, visiting the websites and reading the reports of the industry under study to collect the secondary data. The primary data were collected by serving different questionnaires to the employees and the public to take their responses. The executives of the establishment were interviewed to know their role in environmental control, health and safety and welfare of the employees and protection of locals from the impact of pollutants released by the industry on the environment. The data were sorted, computed and analyzed applying Google, Form software.

Samples, Interviews and Questionnaires Designed

A sample refers to a representative part or a single item from a larger whole or group especially when presented for inspection or shown as evidence of quality. In our research for the interview was conducted which served as our first basis of data collection. the interview was taken from an executive working in targeted famous steel industry of Pakistan. The executive was clearly asked about his understanding of pollution being caused by the steel industry and his role in the issue. He was also asked about the ways that his organization is doing to curb the spread of pollution and to ensure a greener, pollution free and a more sustainable future.

Secondly, a questionnaire-based interview was also conducted from an employee who was working at a well reputed steel mill. He was asked about his views on whether he thinks that the steel industry is responsible for pollution. He was also questioned about the working condition at his organization and steps that they are taking reduce emissions and to preserve the atmosphere.

The questionnaires for responses of the mill employees and general public (Appendix) were designed which were sent to a group of people who were asked about the problems that they were facing due to the increase in pollution. They were also asked about how the ongoing air pollution has affected their health. They were also questioned about what do they think the future holds for them along with their take on pollution.

Data Collection

Both, qualitative and quantitative data were gathered. The data were collected via an online survey questionnaires, which contained both open, and closed ended questions, and semi structured interviews and also through personal contact.

Data Analysis and Interpretation of Results

The data were statistically analyzed by applying Google. Form Software. The results of analysis were recoded as bar graphs and pie charts to compare the results and investigate the differences between results. Line graphs were also used to study the trends over a short period of time. The open-ended questions in questionnaires and interviews were analyzed manually by the researchers by interpreting and grouping similar answers.

4. RESULTS AND DISCUSSION

The results of analysis of responses to the public questionnaire

The results of analysis of responses to the public questionnaire based upon the effects of industrial pollution on the general public are as follows:

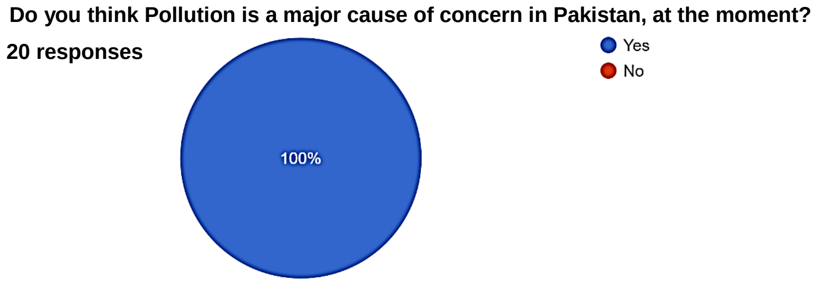


Figure 1: Pie Chart demonstrating the responses of the 20 Participants against the Question, "Is Pollution Cause of Concern in Pakistan?"

Figure 1 shows that all think that the problem pollution is deemed as a major cause of concern for Pakistan, and if not contained would bring about numerous catastrophic implications on the country and its generations to come.

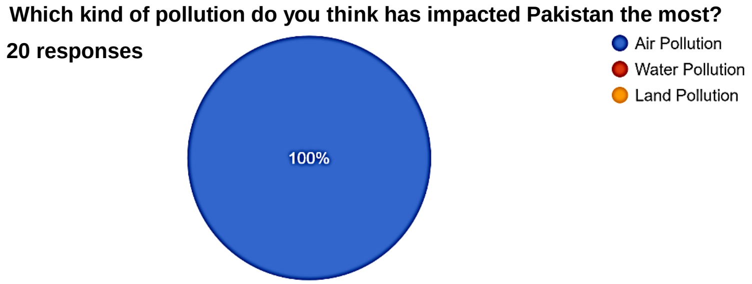


Figure 2: Pie Chart demonstrating the responses of the 20 participants against the question, "Which kind of pollution has affected Pakistan most: Air pollution. Water pollution or Land pollution?"

Figure 2 shows that all the 20 respondents were of the opinion that the air pollution has mostly impacted Pakistan environment.

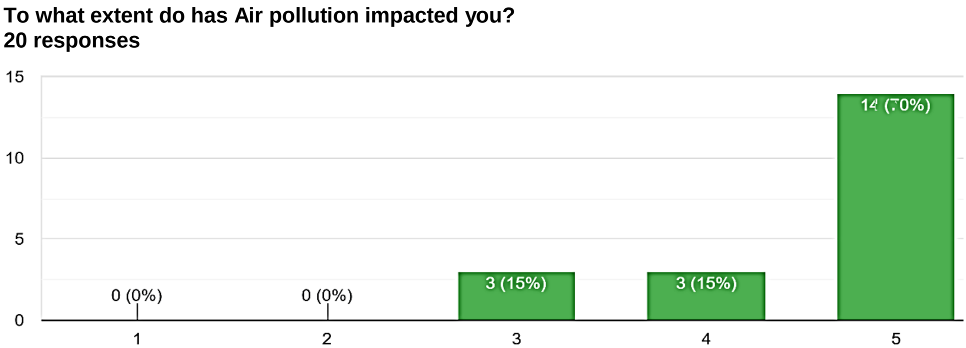


Figure 3: The Bar Chart exhibiting the impact of Air Pollution on Public

Figure 3 portrays how majority of the respondents are immensely impacted by the air pollution. Secondly, a small chunk of respondents claimed to be affected by air pollution but not to a big extent. However, it can be concluded that all the respondents were somehow affected by the air pollution in Pakistan.

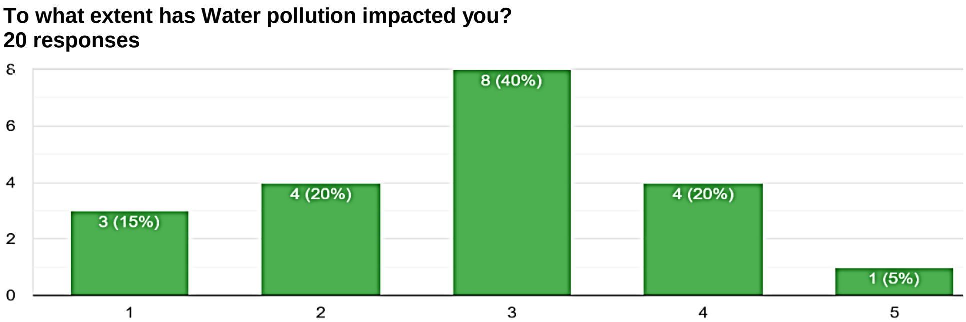


Figure 4: The Bar Chart displaying impact of water pollution on general public

Figure 4 shows that a mix of responses was received against the question showing that the respondents were differently affected by water pollution. Thus the overall conclusion was that majority of the respondents were moderately impacted by water pollution.

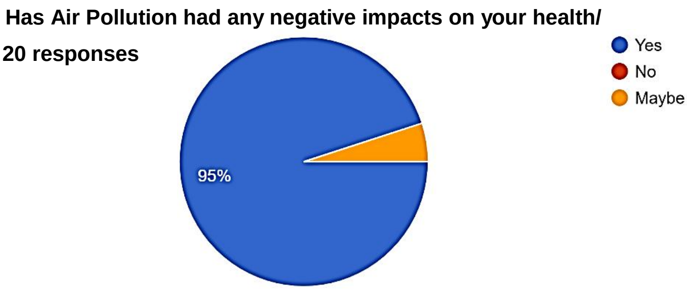


Figure 5: Pi Chart showing responses against the question,
“Did air pollution had any negative impacts on Public health?”

Figure 5 shows that 95 percent of the respondents claimed to have been affected by air pollution in terms of their health while 5% of the respondents were unsure about this question.

Table 1
Public responses to open ended questions about impact smog on life style

Has Smog changed your lifestyle in anyway (if yes, briefly explain) 20 responses
Have to wear a mask
Yes, i made breathing lil difficult
Wearing mask on a regular basis
no
It has drastically impacted my day to day activities. I cant enjoy outdoors activities anymore
It has caused constant flu
Itching, sore throat throughout winters
Yes, we have to wear masks
Yes , health issues , wear mask, Incorporated technology in lifestyle such as Air purifiers , N95 Mask

The responses of respondents against open ended questions were also taken (Table 1) in which they were asked to briefly explain how the smog in the air pollution has forced them to change their lifestyle. To this question majority of the respondents claimed that they have been forced to wear masks all times and have been deprived of enjoying outdoor activities. The respondents also claimed that smog had instigated constant flu, itching throat, shortness of breath and asthma among them.

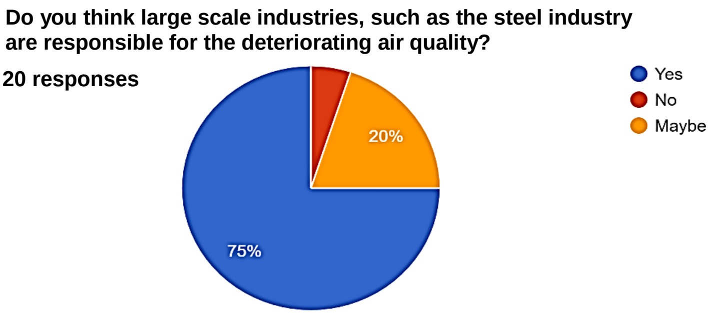


Figure 5: Displaying responses against the question, "Is large scale industry such as steel industry responsible for deteriorating air quality?"

This question was asked in order to know whether the respondents blame the ongoing pollution crisis and the deteriorating air quality the large scale industries such as the steel industry. 75 percent of the respondents believed that the industries were to be held accountable for the air pollution whereas a mere 5 percent disagreed with it. Moreover, 20 percent of the respondents were unsure as they choose the maybe option.

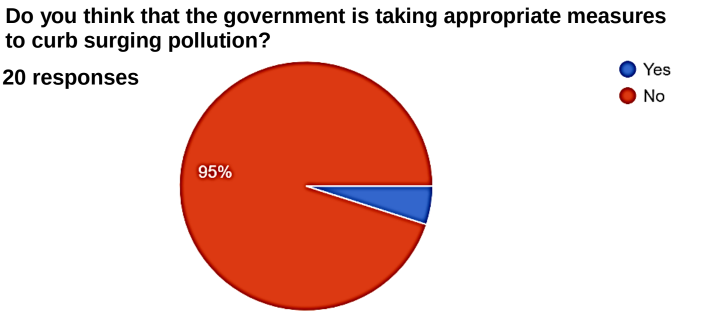


Figure 6: Showing the responses to the question," Is government taking appropriate measures to curb surging pollution?"

Figure 6 shows that 95 percent of the respondents claimed that the government is taking no measures to contain pollution which is rapidly worsening the air quality in Pakistan.

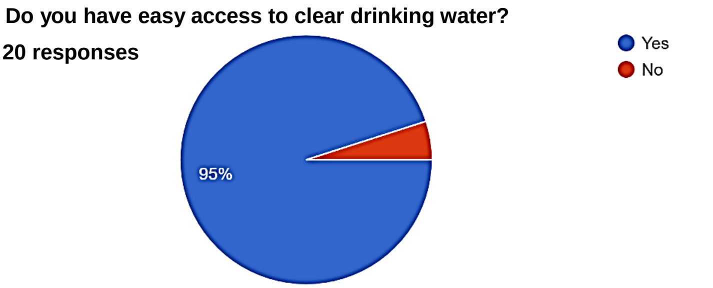


Figure 7: showing responses to the question, "Do you have easy access to clean drinking water?"

This questioned was inclined towards the impacts of water pollution as the respondents were asked whether they have access to clean water or not, to which 95 percent of them claimed to have clean water available.

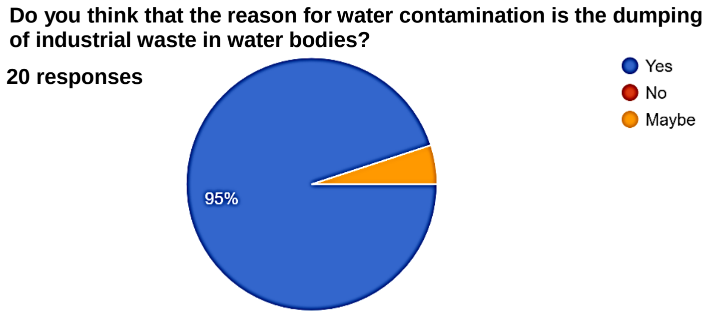


Figure 8: Showing responses to the question, "Do you have the access to the clean drinking water?"

Since industries not only emit poisonous gasses but also dump their toxic waste in the water bodies, this question was asked about industrial dumping being the major cause of water pollution to which the 95 percent of the respondents agreed, with 5 percent of them being unsure.

The results of analysis of responses to the employee's questionnaire

The responses and questions are analyzed as follows:

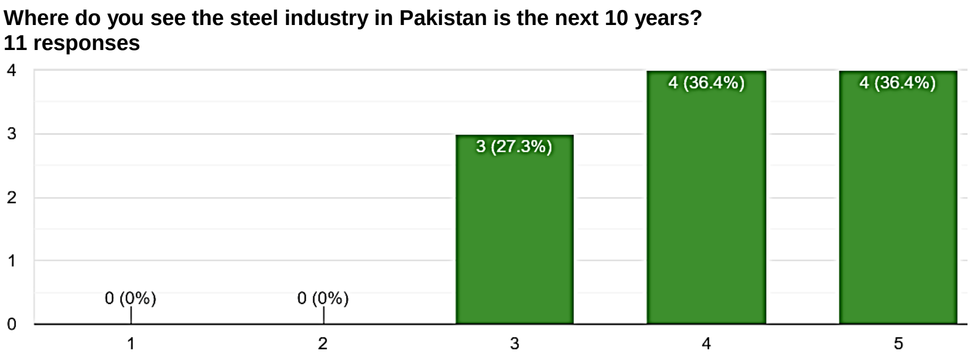


Figure 9: Showing results of responses to the question, "Where do you see the status of steel industry in Pakistan?"

This question was asked to know the take of steel industry workers on where they see the steel industry of Pakistan in the next 10 years. To this majority of the responses were inclined towards the steel industry developing and excelling in the next ten years.

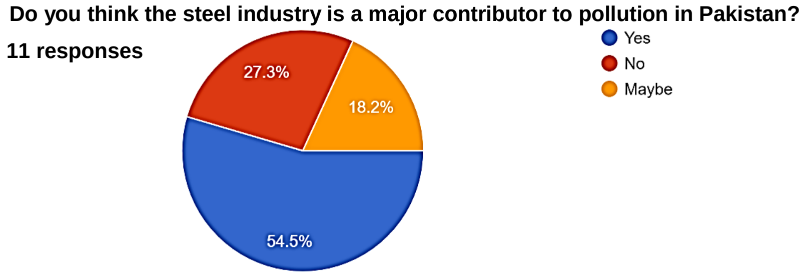


Figure 10: Showing responses to the question, "Is the steel industry is a major contributor to pollution in Pakistan"

This question was asked from the workers to know whether they think that the steel industry is a major contributor to pollution in Pakistan. To this question 54.5 percent of the workers agreed whereas 27.3 disagreed. 18.2 percent of the workers were unsure about this issue.

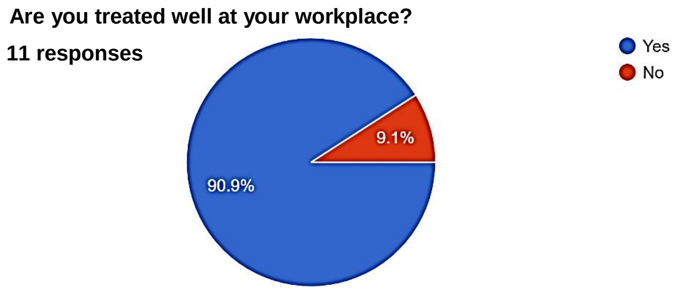


Figure 11: Exhibiting results of responses to the question, "Are you treated well in your workplace?"

This was question was asked in order to obtain information about the working conditions in the steel industry and whether the employees working there are treated in a positive manner or not. 90.9 percent of the employees claimed to be treated well at their organizations whereas 9.1 claimed that they weren't treated well.

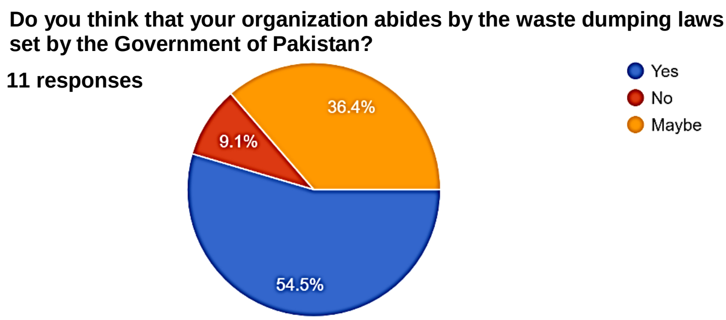


Figure 12: Displaying results of responses of 11 people against the question, "Is your organization abiding by the waste dumping laws set by the Government of Pakistan?"

There are certain waste dumping laws imposed on industrial sector by the government of Pakistan. The employees were asked whether their organizations follow those rules to which 54.5 percent agreed whereas 36.4 percent of the employees were unsure. 9.1 percent of the employees claimed that their organizations didn't not abide by the rules.

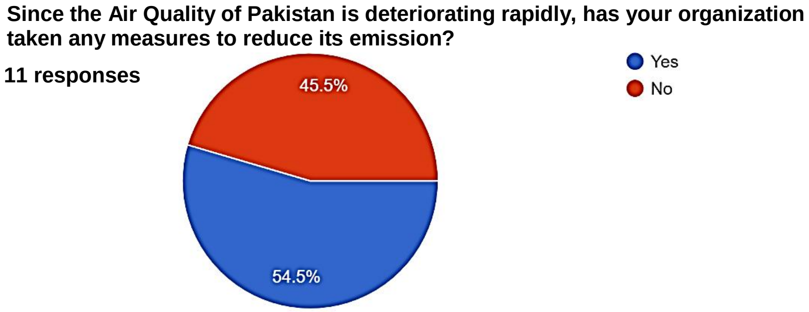


Figure 13: Showing results of the responses to the question, "Whether their organization are taking any concrete measures to curb the increasing air pollution by reducing emission?"

To this question 54.5 percent of the employees claimed that their organizations are taking relevant measures, whereas 45.5 percent of them claimed that no measures were being taken.

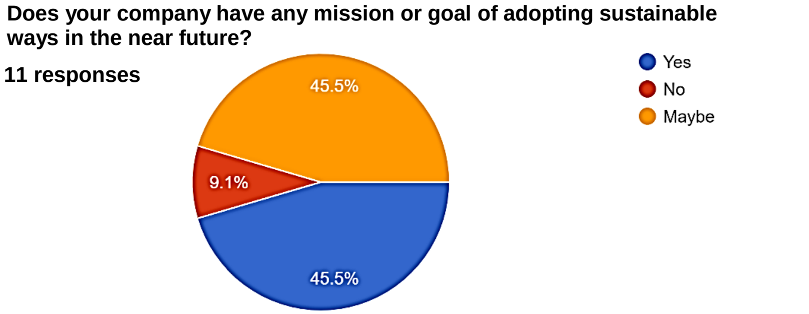


Figure 14: Demonstrating results of the responses to the question, "Whether the company has any mission to adopt sustainable ways in near future?"

This questioned was designed to find out whether the steel manufactures have any mission or plan of action of abandoning traditional methods of steel production and switching to sustainable ways in the future. To this 45.5 percent claimed that their organizations are working towards a sustainable future whereas, 9.1 percent disagreed. 45.5 percent on the other hand remained unsure.

As far as the results are concerned, the findings from our research are almost on the same lines as we expected them to be. But the interviews and questionnaires painted a clearer picture on how the industrial activities are aiding air pollution in Pakistan along with the numerous health impacts they have caused. Results of this study compare very well with views of our predecessors.

As far as the Pakistani steel mills are concerned, they are yet too far from adopting sustainable methods in order to curb air pollution. A major reason behind this negligence is the laziness of the state which has imposed little or no regulation on the pollution caused by the steel industry. No fines are imposed nor any action is being taken against the steel mill owners. In order to find a reasonable solution, the government must take initiative and pass a bill in the parliament which introduces new and strict laws imposed on industries that are polluting the atmosphere to a high extent. The Pakistan Environmental Protection Act 1997 was passed by the National Assembly of Pakistan on September 3, 1997, and by the Senate of Pakistan on November 7, 1997. The Act received the signature of the President of Pakistan on December 3, 1997. The law emphasized upon regulating numerous industries and sectors of the economy which were seen as the major contributors to both air and land pollution. It has been around 25 years until this act was passed, however, no government has penalized the violators. In order for the pollution reduction a concrete action is required by the government and until then there is no light towards the end of the tunnel.

The steel manufactures must shift their crude steel production towards blast furnace or electric furnace technology. Blast furnaces are used to produce pig iron from iron ore for subsequent processing into steel, and they are also employed in processing lead, copper, and other metals. Rapid combustion is maintained by the current of air under pressure. Blast furnaces cannot be considered to be a hundred percent sustainable as they do affect the environment as well, however, to a much lower extent when compared to the traditional furnaces used in Pakistan. Electric furnaces on the other hand perform steelmaking, like the open-hearth process, has a means of charge heating independent of a hot metal charge or the combustion of impurities. The electric powered furnaces bring about numerous benefits and can immensely contribute towards the preservation of the environment. Electric heating to melt steel are a far better option for the steel producers in numerous ways. Firstly, electric heat is much more reliable and convenient as well as being cost effective. These adaptable machines are used for cooling as well as heating, which removes the need for additional equipment and lowers equipment costs. Moreover, this form of heating does not pollute the earth’s environment unlike the fossil fuel powered heating. The other sources of heating like coal and gas are major polluters of the ecosystem. Electrical heating is considered to be sustainable and environmentally friendly as they do not release any sort of emissions keeping the air clean. This helps to contain several harmful emissions such as the carbon monoxide and sulfur dioxide and leaves no leftover debris as well.

5. COMMENTS AND CONCLUSION

The research study shows that greenhouse gases have caused significant human suffering due to their effects on people's lives and the environment. The steel industry is unaware of its role in causing climate change and environmental deterioration. Because there are no clear SOPs by the government, the government is more responsible for environmental damage than these enterprises are. The factory owners are not aware of the consequences. The employees in the chemical industries are not generally aware of the harmful impacts of the methods utilized in such businesses. Together, governmental organizations and business owners can lower risks and raise people's standards of living.

6. ACKNOWLEDGEMENT

We are extremely grateful to Dr. Shahid Amjad Chaudhry, Rector Lahore School Economics for patronization of our research and sanction of expenditure involved and providing us requisite facilities to participate in 20th ISOSS Conference held in Lahore

REFERENCES

1. Ali, S.M. (2018). *Tyre Burning-a Threat to Environment*. Online available at https://www.linkedin.com/pulse/tyre-burning-a-threat-environment-syeda-maheen-aly?trk=public_profile_article_view
2. An, D., He, Y.G. and Hu, Q.X. (1997). Poisoning by Coal Smoke Containing, Arsenic and Fluoride. *Fluoride* 30(1), 29-32 Research Report 29 https://www.fluorideresearch.org/301/files/FJ1997_v30_n1_p029-032.pdf
3. Greenstone, M. and Fan, Q. (2019). *Pakistan's Air Pollution Challenge & Potential for Longer Lives*. Air Quality Life Index, https://aqli.epic.uchicago.edu/wp-content/uploads/2019/02/Pakistan-Report.pdf
4. Younes, I., Liaqut, A. and Sadaf, R. (2020). Consequences of Released Air Pollutants from Iron and Steel Industry on Human Health in Lahore, Pakistan: Consequences of Released Air Pollutants from Iron and Steel Industry. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences*, 57(4), 59-75.
5. Zahra, S.I., Iqbal, M.J., Ashraf, S., Aslam, A., Ibrahim, M., Yamin, M. and Vithanage, M. (2022). Comparison of Ambient Air Quality among Industrial and Residential Areas of a Typical South Asian City. *Atmosphere*, 13(8), 1168.
6. Zia, U. and Hina, H. (2021). LSM–Pakistan Steel Industry Outlook (No. 2021: 43). *Pakistan Institute of Development Economics*.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 97-114

ENVIRONMENTAL SITUATION OF POLYMER INDUSTRY IN PAKISTAN

Mohammad Rafiq Khan[§], **Khadija Shahid**,
Mohammad Nawaz Waseem and **Ameer Mustafa**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

The main objective of the article being presented was to study the significance of the polymer producing business in Pakistan and its subsequent evaluation of its contribution to the environmental deterioration in the Country. The research article begins with comments on the status and significance of the polymer industry in Pakistan and then to its environmental status choosing a very reputable industry, "Pakistan Polymers" as a model study for later generalization. The secondary data was collected by consulting literature in the library, vising concerned websites on Internet and reading reports of different players in this business. The primary data was collected by taking responses of the executives, workers, and public expected to be the victims of the pollutants liberated from the Pakistan Polymers Industry against specifically designed questionnaires. The collected data were sorted, computed and statistically analyzed applying Google. Form to record results in the form of tables/diagrams. The findings were then applied to determine how much contribution the polymer industry has made to the deterioration of the environmental situation in Pakistan. After a comprehensive discussion, it was concluded that plastic waste material has a significant impact on environment.

1. INTRODUCTION

Polymer industries produce polymers, which are large molecules made up of repeating units of smaller molecules. Polymers are used in a wide variety of products, including plastics, textiles, rubber, and adhesives. Depending on the specific industry, polymer industries may produce raw polymers, such as petrochemicals or may process and manufacture polymers into finished products such as polyethylene or polypropylene used to make a wide range of products including packaging materials, plastic bottles, toys, etc. On the other hand, a plastic manufacturing company may take raw polymers and process them into finished products, such as plastic containers or packaging materials and many others: textiles, adhesives, and coatings.

These industries play a crucial role in the global economy and are constantly innovating to develop new products and technologies.

Polymer industries can have both positive and negative effects on the environment. On the positive side, polymers can be used to produce products that are lightweight, strong, and durable, which can help to reduce the environmental impact of manufacturing and transportation. These polymers can persist in the environment for long periods of time, and can cause harm to wildlife and ecosystems. To minimize the negative environmental impacts of polymer industries, it is important for companies to implement sustainable production practices. This can include reducing the amount of waste generated during the production process, using renewable resources instead of fossil fuels, and implementing recycling and waste management programs. Additionally, governments and regulatory bodies can play a role in promoting sustainable practices and reducing the environmental impact of polymer industries.

Some studies have been conducted to address the environmental contamination problems that are associated with the Pakistan Polymers industry. A few are highlighted below.

It has been put together based on the results of three environmental audits that were carried out in Pakistan Polymers Industry as part of the Environmental Technology Program for Industry (ETPI). The purpose of this report was to evaluate the type and scope of environmental issues that are brought on by the synthetic polymers and to propose potential remedies to lessen the negative effects of those issues. The report says that we are now receiving a huge load of information on plastic with less impact to the environment but much of this information is contradictory, not bringing acceptable scientific references on the assertions made. Even the norms for biodegradation tests have been developed under influence of the manufacturers of biodegradable products as a tool to ward off competitors.

Synthetic polymers can come in a variety of forms, such as common plastics, the nylon of a jacket, or the surface of a non-stick frying pan, but these human-made materials have a detrimental impact on ecosystems which U.S. National Institute of Health researchers have called "a rapidly increasing, long-term threat." Understanding the ways that synthetic polymers degrade ecosystems is important in taking steps to eliminate this form of pollution. Biodegradability is a feature that has been highly valued in polymers from the environmental standpoint, but is not the only important one. Sooner or later, all components in a polymer material will be returned to the environment, with the degradation, so it is very important to use pigments, fillers and additives that are not toxic in nature (Ojeda, 2013).

Industrial sources of PBM waste include air-blasting technologies that use microscopic beads to strip paint from metallic surfaces and for cleaning engine parts; when discarded, they enter the environment through foul-water, or via transfer through sewage treatment processes. LDPE films constitute a large volume use of PBMs in agricultural crop production, and consequently they have become an important agricultural emission. Their application is thought to be one of the most important sources of PBM contamination of soils, because they become brittle and easily disintegrate rendering their recovery difficult. Agriculture films can also contain light- sensitive additives, such as ferric and nickel dibutyldithio-carbamates, the ratio of which can be adjusted so that the film is usable during a specific growing season, after which the product begins to photo-degrade. This ultimately results in disintegration of the material, and when coupled with successive precipitation events the disintegrated particles can be washed into the soil where they accumulate (Lambert, 2013).

Natural polymers such as rubber and silk exist in abundance, but nature’s “plastics” have not been implicated in environmental pollution, because they do not persist in the environment. Today, however, the average consumer comes into daily contact with all kinds of plastic materials that have been developed specifically to defeat natural decay processes—materials derived mainly from petroleum that can be moulded, cast, spun, or applied as a coating. Since synthetic plastics are largely non-biodegradable, they tend to persist in natural environments.

Moreover, many lightweight single-use plastic products and packaging materials, which account for approximately 50 percent of all plastics produced, are not deposited in containers for subsequent removal to landfills, recycling centers, or incinerators. Instead, they are improperly disposed of at or near the location where they end their usefulness to the consumer. Dropped on the ground, thrown out of a car window, heaped onto an already full trash bin, or inadvertently carried off by a gust of wind, they immediately begin to pollute the environment. Indeed, landscapes littered by plastic packaging have become common in many parts of the world (Moore, 2023).

The production of synthetic plastics, particularly non-biodegradable ones, is an environmental burden. This is because 'non-biodegradable' plastics take decades or centuries to break down. Non-biodegradability of certain plastics suggests that their chemical structure cannot be adequately modified by naturally occurring microorganisms, water, carbon dioxide or methane to degrade them. Meanwhile, 'biodegradable' plastics are truly compostable materials that can almost entirely be converted into benign trash after a matter of months in a composter (Atiwesh, 2021).

Studies on biological decomposition of plastics by various microorganisms under different environmental conditions have revealed that these decomposition conditions are governed by the physical and chemical characteristics of the type of plastic discarded, such as mobility, crystal structure, molecular weight, functional groups etc. High molecular weight, high degree of crystallinity, high hydrophobicity as a result of linearity of the polymeric carbon chain backbone, and general insolubility in water are some of the factors that typically reduce the degradability of plastics. Indeed, these are the properties that make the petroleum-based plastics polyethylene and polypropylene non-biodegradable. Notably, not all petroleum-based plastics are non-biodegradable. For example, polycaprolactone (PCL) and poly (butylene succinate) (PBS) are both petroleum-based plastics which can undergo microbial degradation. However, the biodegradability of these polymers is affected by their physicochemical properties such as degree of crosslinking, degree of crystallinity, molecular weight and the species of microorganisms used. Indeed, studies have revealed that cross-linked polymers have the lowest rate of degradation, followed by crystalline and then amorphous polymers (Atiwesh et al., 2021).

There are different types of plastics based on their constituents and type of materials used in their production as mentioned:

Researches on worldwide production of plastics and the accompanied environmental pollution have shown that plastic wastes have constituted a major environmental issue. The effect of plastic wastes on marine organisms, humans and the environment at large is of public concern, and calls for the need to salvage the ecosystems and lives therein. Despite the fact that plastics are very useful in everyday life, the toxic chemicals used in the production need to be thoroughly monitored so as to ensure environmental and health safety. Reducing community's exposure to toxicants from plastic wastes will increase the chances of having a clean environment and healthy society. There is an urgent need for government agencies and health authorities to enact and enforce environmental laws that will monitor production, usage and disposal of plastics. In addition, some harmful chemical constituents used in the production of plastics (e.g. phthalates, BPA, etc.) should be banned in consumer goods and in plastic products that are in direct contact with food, beverages and children (Alabi et al., 2021).

The goal of the study was to highlight the impact of plastic polymers on the general society in Pakistan. The objectives were as under

1. Collection of secondary data
2. Collection of primary data
3. Sorting and computation of data.
4. Analysis of data
5. Interpretation of results
6. Conclusion and recommendations

2. RESEARCH METHODOLOGY

The research design and process involved the stages as given in the form of objectives, the procedures are explained below.

Sampling

Sampling is the basic step undertaken to start an inquiry by statistical methods. The sample is defined as a single object that is representative of a larger whole or group, particularly when it is submitted for examination or displayed as proof of the quality of the whole. This is especially true when the sample is shown as evidence of the quality of the whole.

Collection of the Secondary Data

To collect secondary data a number of research papers and reports to acquire access to previously published work on the issue. In addition to that the websites of the concerned establishments were visited. The secondary thus collected formed a good baseline to clear about materials used to make different plastic products, their production lines, technological progress in the focused field.

Collection of the Primary Data

The primary data was collected by taking responses against appropriately designed questionnaires depending upon the status of the respondents in the current study: Executive/Managers, of the establishment targeted, Workforce/Employees and the Locals affected by the pollutants of the polymer industry. Thus, three types of questionnaires designed were Interview Questionnaire, Employee Response Questionnaire and Public Response Questionnaire. The information collected from the executives was used to ease up serving the questionnaire to the employees of Pakistan Polymers to know their perspective on the ways in which they believe the mills was contributing to the degradation of the environment and looking after the employee welfare. The third kind of questionnaire was served to the general public to know their perspective on the ways in which they believe the mills are contributing to degradation of the environment and the extent to which people were being affected by pollutants.

Processing of the Data

The data collected in the form of responses was sorted, computed and analyzed applying Google. Form Software

Interpretation of Results

The results of the analysis were recorded in the form of pi bar diagrams, and bar line graphs. The results of interview of the executives are reported as descriptive research.

3. RESULTS AND DISCUSSION

Results of Public Survey

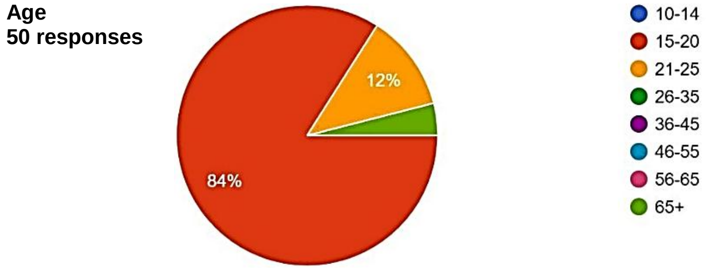


Figure 1: The display of results of the first survey to know which age group is ideal age to respond responsibly to the survey questions

Figure 1 shows that that the majority of people with age range from 15 to 25 years of age have the ideal age to respond responsibly to the survey questions. The reason may be that as they are mostly college or university students who are well aware or have sound awareness of environmental problems. Furthermore, they are mature enough to understand these problems because they are continuously taught about these issues in their universities and many also are aware of these issues through social media as well. So they are able to objectively answer these questions with rationality.

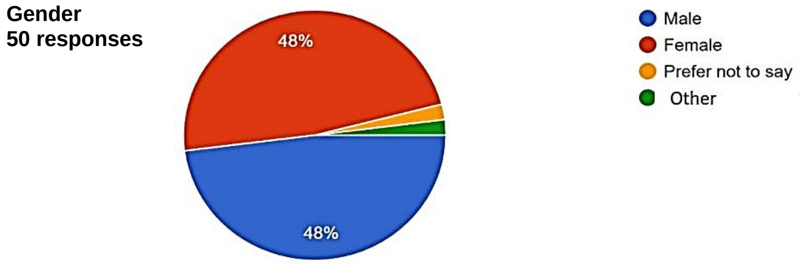
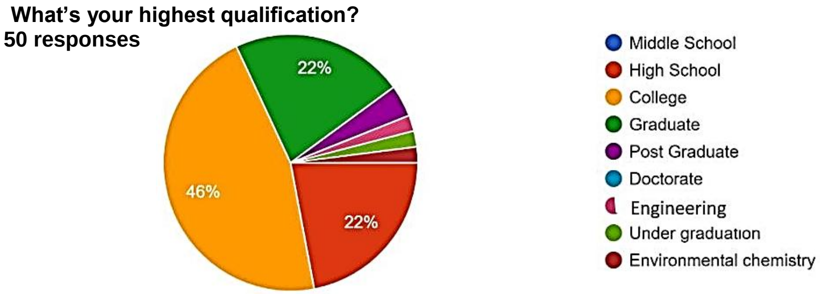


Figure 2: The display of results of the first survey to know which gender group is ideal to respond responsibly to the survey questions

Figure 2 shows that perspectives of both men and women are almost the same. Thus it can be noticed that the gender ratio is almost equal between men and women and that may be an index to a balance and neutrality when considered in the perspectives of both men and women. Even though Men are usually more in the working class as compared to the women but still the responses have come out to be equal.



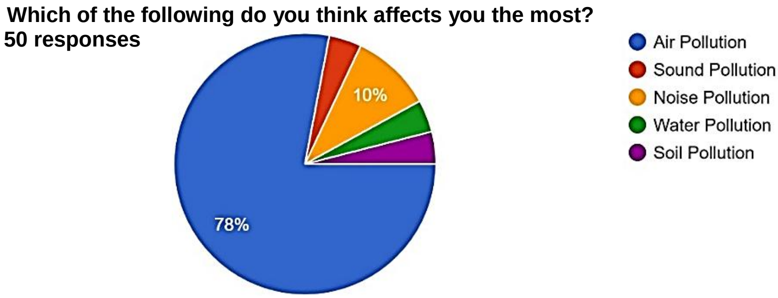


Figure 3: Displaying results in response to the question, "Which of the following types affects you the most?"

Figure 3 exhibits that almost 78 percent of the people are affected by air pollution while 10 percent are affected by noise pollution and the rest are equally affected by soil and water pollution. The case of air pollution is quite understandable as Lahore is a heavily polluted city and throughout the winters there is heavy smog which is a mixture of fog and smoke. Smog is very dangerous for human beings as it adversely affects the lungs and damages them extensively. The main reason for this air pollution is a rapid increase in motorcycles and cars that contribute quite a lot to air pollution. Furthermore, the industrial areas in the vicinity of Lahore are also a major reason for the contribution to the air pollution. If we talk about the noise pollution is talked about, it is noticed that in Lahore there is a lot of noise generated form the traffic in the city plus this city has a lot of open shops and workstations that also contribute to noise pollution.

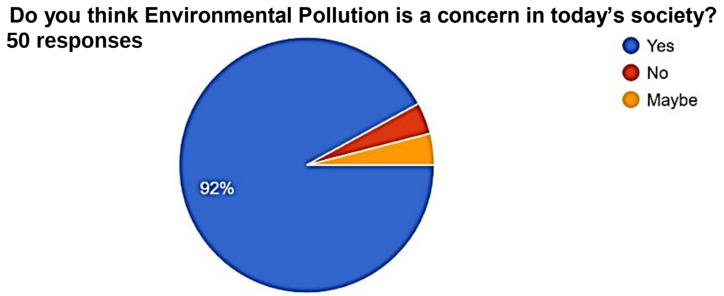


Figure 4: Showing the results of responses to the question, "Is the environmental pollution is a major concern in today's society?"

Figure 4 shows that around 92 percent of the respondents endorse the severity of the situation. It helps to realize that environmental pollution is an important issue to many people, and they have an interest in it as it directly affects them. This also goes on to show us that there is a general awareness in the society in regard to this issue but it may also be argued that many people from the lower classes many may be un aware of this especially the working class and daily wagers. This can be supported by the fact that our research answers are from university students and not from the lower middle class.

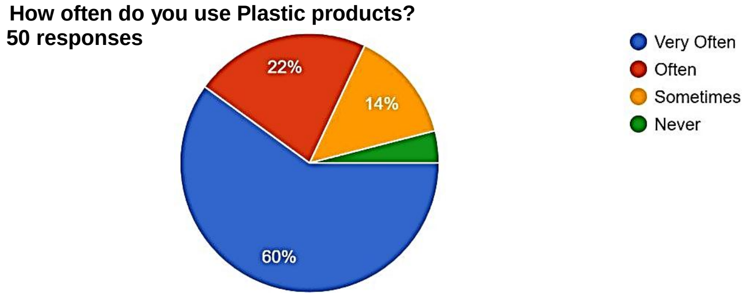


Figure 5: Showing answers to the question, "How often do you use the plastic products?"

From Figure 5, it can be noticed that almost 60 percent of people use plastic products, and this is certainly justifiable since almost all the daily use items nowadays are made from plastic. Some efforts have been made to replace plastic bags with cloth and paper bags, but the implementation will take time. Moreover it may not be an acceptable option for the people.

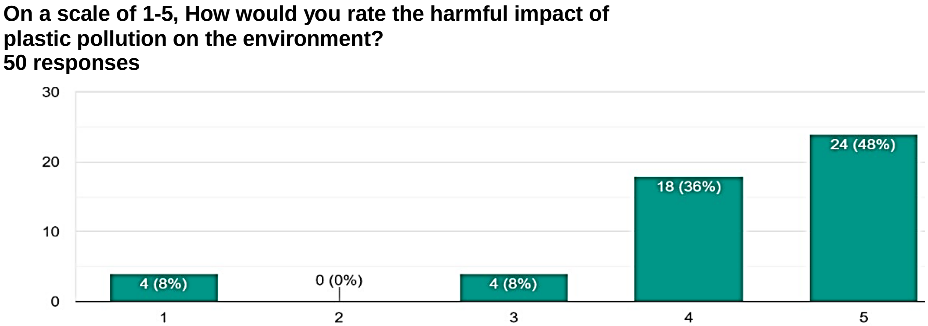


Figure 6: Response to the question, "How would rate the harmful impact of plastic pollution on the environment on the scale 1-5?"

Figure 6 indicates that a large majority of responses, almost half, were people who were aware of the harmful impact of plastics in the society. This was important to determine due to the fact that plastics are of the leading pollutants in the society. They are the permanent waste which harm the environment even if they are incinerated. If the society understands the harms of it on our environment there is better understanding of the problem and the right actions can be taken to prevent the issue from worsening.

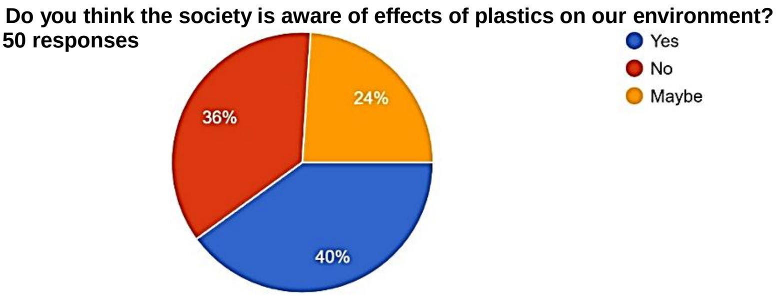


Figure 7: Response to the question that the society is aware of the effects of plastics, on the environment?

Figure 7 shows that out of the 5 responses, 40% said yes, 36% said no and 24% was unsure of it. This quite evidently shows that even though the number of Yeses are in majority, the society is still unaware of the actual harms it causes to the society. There are a lot of mixed feelings about this questions which points to one direction, that is, that people need to be made aware of the harms of plastics in the environment.

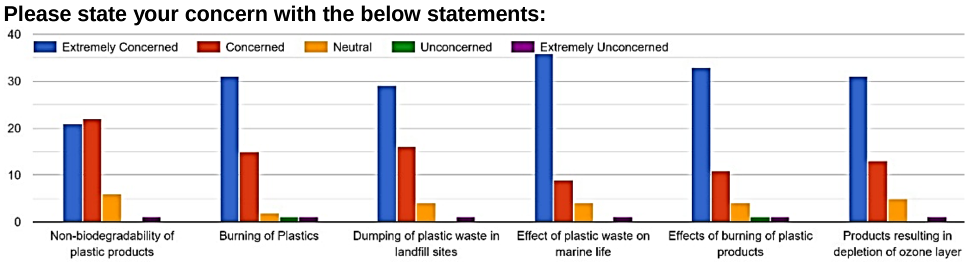


Figure 8. Response to the concern with grading in statement given on the top of Figure 8.

Shows Figure 8 grading regarding plastics, a majority was mildly concerned with the non-biodegradability of plastics, a majority was greatly concerned about the burning of them, and similarly the majority was concerned by all other statements due to their intensity.

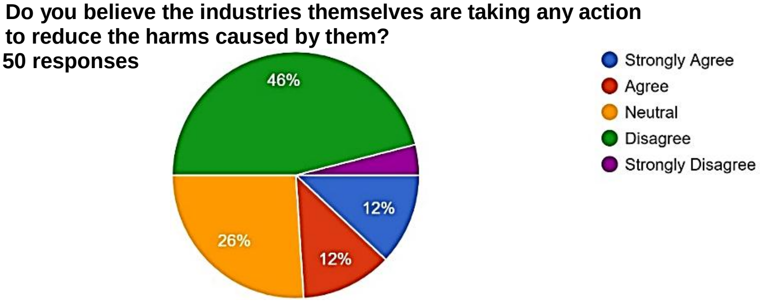


Figure 9: Showing responses to the question, “Do you believe that the industries are taking any action to reduce the harms caused by them?”

Figure 9 exhibits that 46% of respondents think that industries do have a major hand in the increasing of the pollution through its wastes. Only a 12% minority agrees that the industries are taking an action to aid the environment but the stereotype that industries do nothing has already been set.

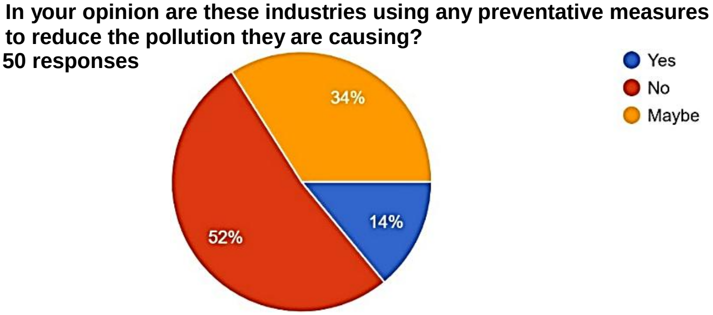


Figure 10: Responses to the question, “In your opinion, are these industries taking any preventive measures to reduce pollution caused by them?”

Figure 10 shows that a whopping 52% of society thinks that the industries in fact do cause harm to society from the improper disposal of their wastes in the form of liquid waste, solid waste or even gaseous emissions that can harm the environment around us. A common stereotype revolves around society that the industries have a direct hand in purposely ruining the environment, but this is not the case as the industries face great legal pressure and are kept in check about the disposal of their industrial wastes.

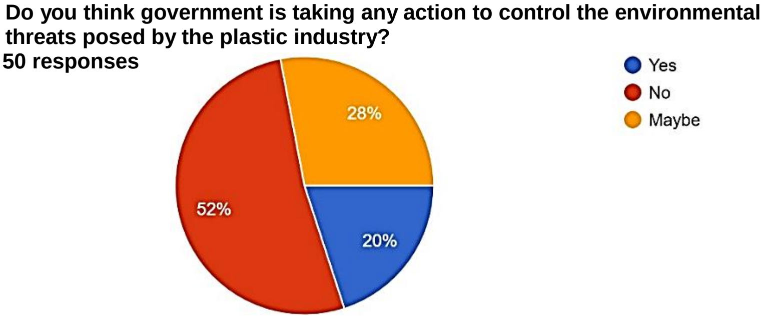


Figure 11: Responses to the question "Do you think that government is taking any action to control the environmental threats posed by the plastic industry?"

The Government plays an important role in the conservation of the environment because they are the entity who makes the rules and enforces them. Since this survey was taken in Pakistan, where the enforcing of rules is a little shabby, a 52% of the majority believes that the government is not taking any proper action to reduce pollution caused through industries. 28% are unsure of the situation and 20% agree that the government is indeed taking action to prevent it (Figure 11). The government of Pakistan does not have any apparent laws that cater to this issue, but some actions like the billion tree plan are all being taken for the conservation of environment.

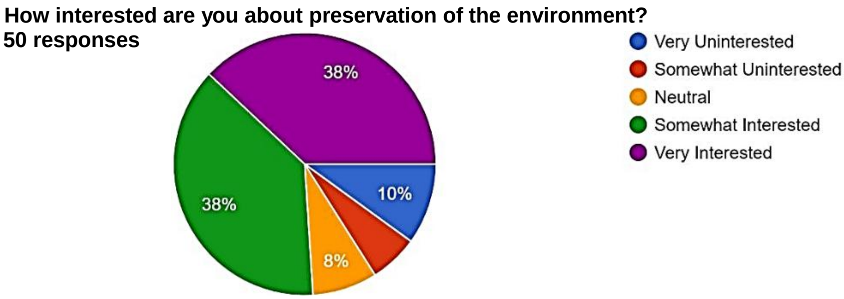


Figure 12: Response to the question, "How much interested you are about the preservation of the environment?"

This question focused on how aware and worried the public of Pakistan in regarding the prevalent issues in our environment and how willing they are to prevent it. The results (Figure 12) shows that 76% of the people are worried about the deterioration of the environment and were interested in its preservation. The people can do so by creating more awareness of the issue among the masses. It is a happy sight to see such a vast majority in favor of the preservation of the issue since it is the future of the planet and adequate actions are needed to protect the future generations as well as the animal and plant life too.

RESULTS OF EMPLOYEES SURVEY

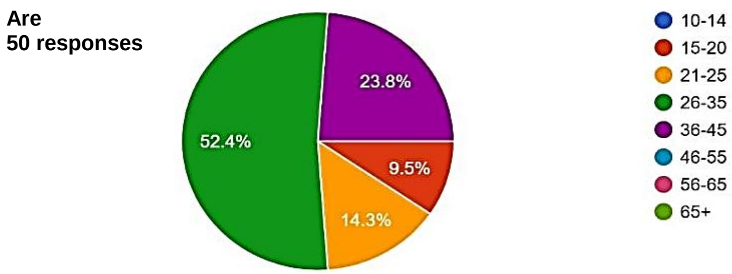


Figure 13: Survey showing the responses to which age group is ideal age to respond responsibly to the survey questions.

The second survey (Figure 13) was for the employees of Pakistan Polyester. A majority lie between the ages of 26-35 which is the ideal age for an average workplace. These people gave accurate answers as they are the most effected by climate change and pollution.

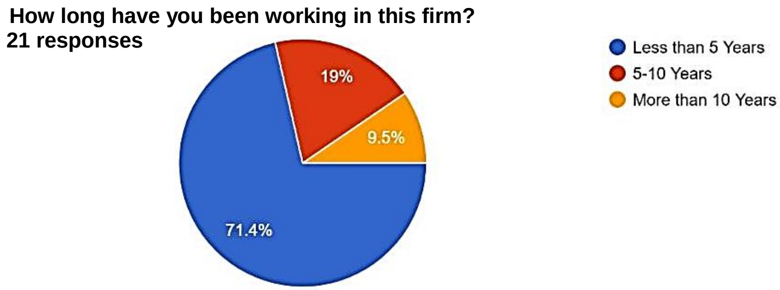


Figure 14: Showing responses to the question, "How long have you been working in this firm?"

Figure 14 indicates that majority of employees (71.4%) has worked in the firm for less than 5 years, while 19% for more than 10 years from 5 to 10 years and only 8.6%.

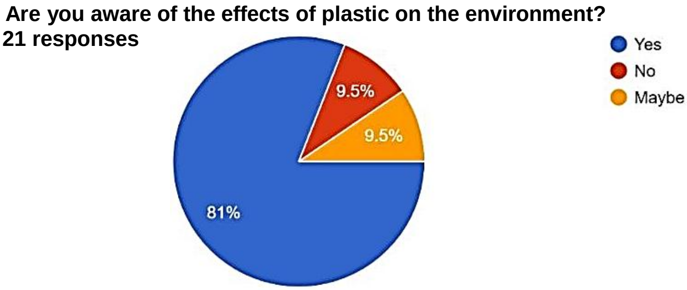


Figure 15: Showing responses against the question, "Are you aware of the effects of plastics on the environment?"

Figure 15 shows that they are well aware of the presence and harms of plastic in their society and are given seminars for the awareness of this problem.

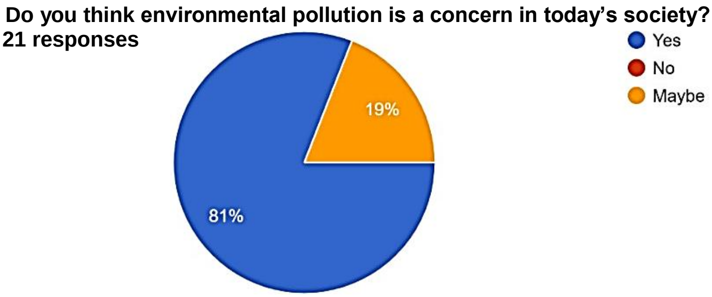


Figure 16: Responses against the question, "Do you think that environmental pollution is a concern in to-day's society?"

When asked in environmental pollution is a concern to the society and majority of 81% (Figure 16) said it was, on a better perspective there were no subjects who said no, that means 100% of the employees consider it an issue.

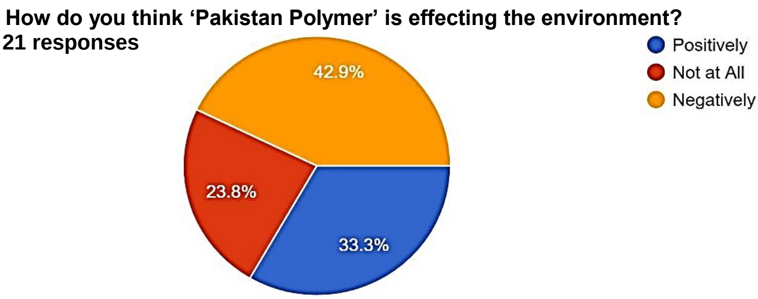


Figure 17: Showing the responses against the question," How do you think that Pakistan Polymers is affecting the environment?"

Figure 17 shows that a mixed response. Of course, 42% firmly said that Pakistan polymers is affecting the environment and stressed that they are indeed hurting it. This may be due to improper disposal or wastes or how the public uses their products.

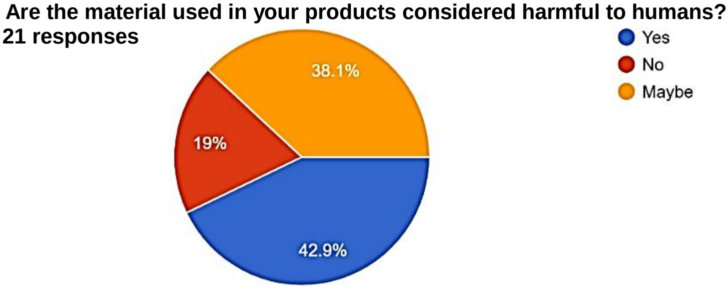


Figure 18: Showing responses to the question, "Are the materials used in your products are harmful to humans?"

A polymer is created when many monomers are linked together. Polymers can be compared to a chain of interconnected paperclips in several ways. A polymer is a big molecule consisting of monomers, which are littler molecules bonded together. Since polymers are made of synthetic material and chemicals it has a permanent life and does not disintegrate easily. Thus the employees in majority (42%) do consider the materials to make the polymer a threat to society.

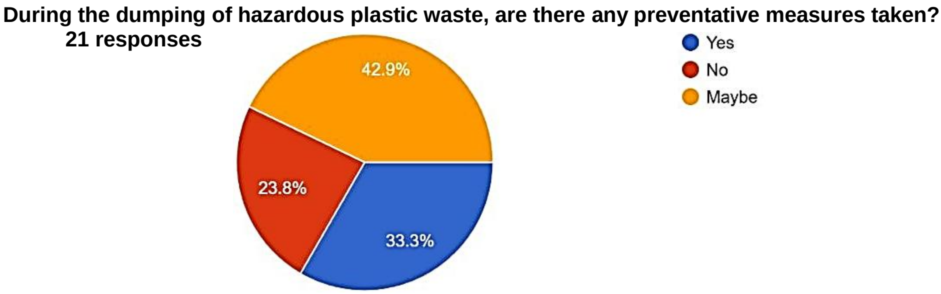


Figure 19: Responses to the question, "During the dumping of hazardous plastic waste, are the any preventive measures taken?"

Figure 19 shows that majority (42%) are unaware while a large portion of the employees almost 33% agreed that preventative measures are taken while 23% said no. This shows that there is an uncertainty among the employees on whether the waste is dumped properly or not. The company should make their position clear on this aspect since it is a very sensitive matter and is directly linked to the environmental wellbeing.

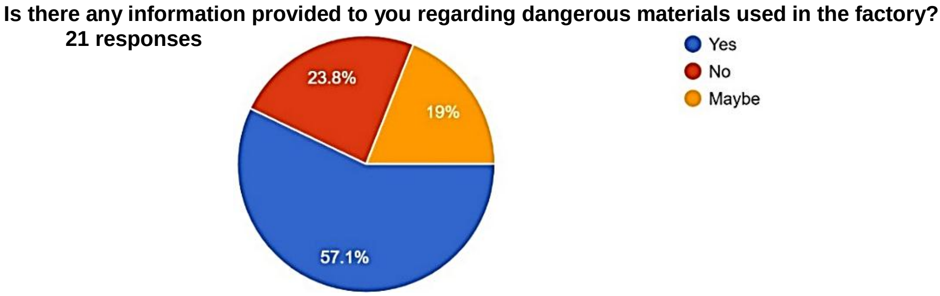
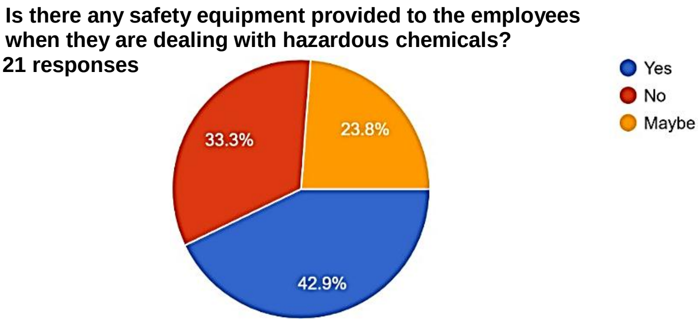


Figure 20: Responses to the question, "Is there any information provided to you regarding dangerous materials used in the factory?"

Figure 20 indicates that almost 57% are aware while 23% aren't aware regarding the handling of these chemicals. This may also be due to improper implementation of rules and regulations at Pakistan Polymers and is also reflective of the mismanagement on the part of the company.



Here we again notice that a large part of the employees are either not given the right equipment to deal with the chemicals or they are either unaware of this fact this again is reflective of the mismanagement and improper implementation of rules and regulation at the factory. This is extremely important as the chemicals can harm the employees to a large extent and also in some cases cause lifelong injuries.

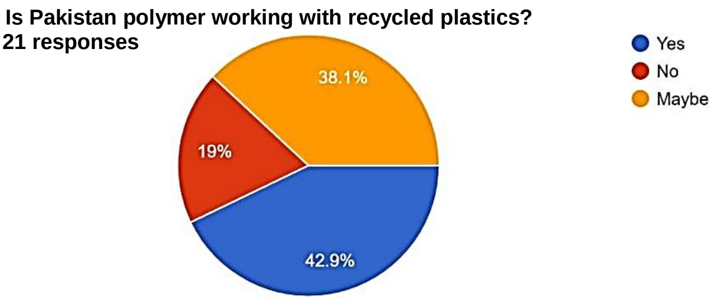


Figure 21: Responses to the question, "Is Pakistan Polymers working with recycled plastics?"

Figure 21 indicates that the majority (almost 42% agrees) that it is certainly good since it shows that the industry is making some effort in trying to reduce the plastic pollution. This is beneficial for the company in two ways firstly it will be helpful In creating a good image of the company for being environmentally friendly secondly it will help the company with the cost cutting as they will be using recycled material.

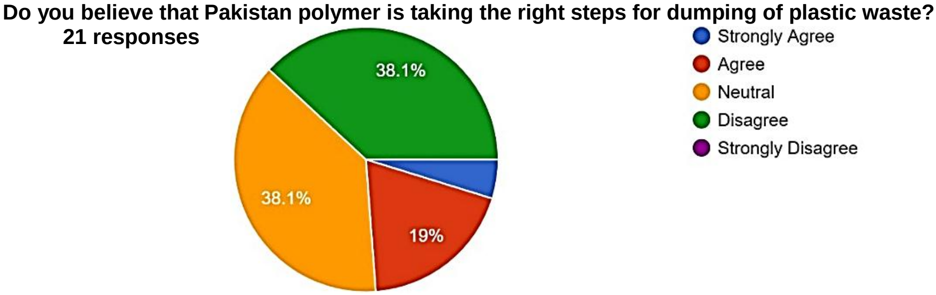


Figure 22: Responses to the question, "Do you believe that Pakistan Polymers is taking the right steps for dumping of the plastic waste?"

Figure 22 indicates that majority are neutral while a large portion of the employees almost 38% disagreed that preventative measures are taken while 23% agree. There is uncertainty among the employees on whether the waste is dumped properly or not, and the company should make their position clear on this since it is a very sensitive matter and is directly linked to the environmental wellbeing.

RESULTS OF THE EXECUTIVE INTERVIEWS

Nine executives were surveyed out of which a majority lied between the ages of 26-35 and occupied managerial positions. They were questioned about certain aspects of their company.

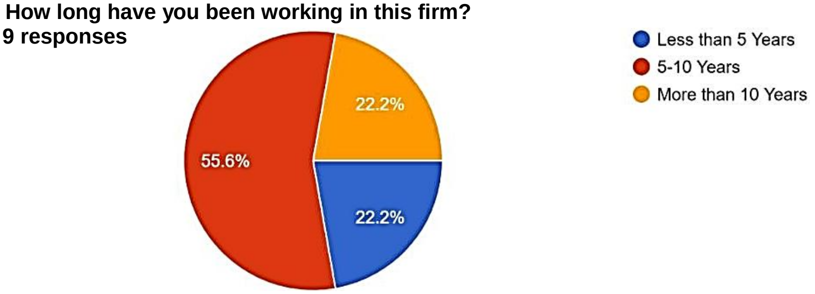


Figure 23: Responses of the executives against the question "How long you have been working in this firm?"

Figure 23 shows that the majority of executives think that environmental pollution is a concern in the society, with almost 89% in favor of this statement. This is because environmental deterioration is one of the leading problems facing the world nowadays and it requires a solution.

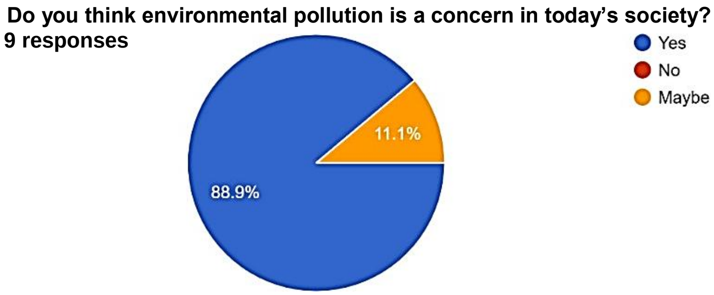


Figure 24: Responses against the question, "Do you think that environmental pollution is a concern in today's society?"

Figure 24 indicates that the majority almost 88% agree with the fact that polymers harm the society and it surprising since this is coming from the top management of the company and thus their honesty should be appreciated but this shows us a real picture of how damaging polymers actually are for the environment.

Figure 25: Responses to the question, "Do you think that polymers harm environment?"

Figure 25 shows that polymers majority (66%) agreed that polymers harm environment.

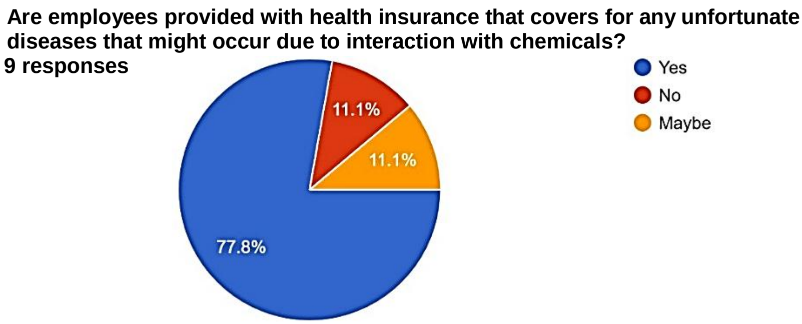


Figure 26: Responses against the question, "Are employees provided with health insurance that covers for any unfortunate diseases that might occur due to interaction with chemicals?"

Figure 26 shows that majority agreed that Pakistan Polymers employees are given proper training and awareness about the negative presence and care for and of the plastics and polymers in the society

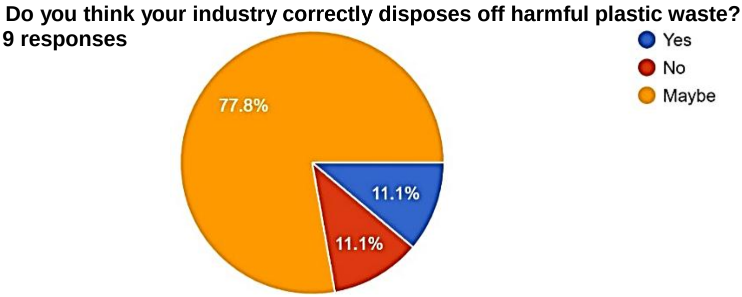
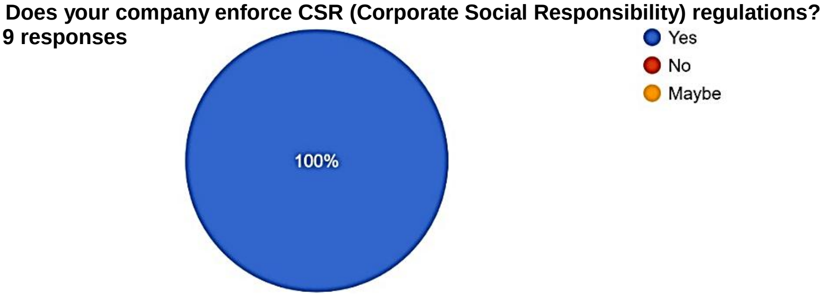


Figure 27: Responses against the question, "Do you think that your industry correctly disposes harmful plastic waste?"

Figure 27 shows that the higher executives could better answer this question as they were aware of the conditions of the employees and their perks. A good healthcare plan covers for the employees in case for any illness caused by working in the company. Here majority agreement is a good gesture,



Enforce Figure 27: Responses to the question, "Does your company enforce Corporate Social Responsibility regulations?"

Figure 27 shows that Pakistan polymers is a strict advocate of following CSR policies and regulations (100% agreement). They maintain a positive role in the society by supporting charities and also work for the environment in terms of proper waste disposal.

4. COMMENTS AND CONCLUSION

After a comprehensive discussion, it was concluded that plastic waste material has a significant impact on environment. One notable result of the study is that Pakistan polymers is a strict advocate of following CSR policies and regulations (100% agreement). They maintain a positive role in the society by supporting charities and also work for the environment in terms of proper waste disposal.

5. ACKNOWLEDGEMENT

We are extremely grateful to Dr. Shahid Amjad Chaudhry, Rector, Lahore School Economics for patronization of our research and sanction of expenditure involved and providing us requisite facilities to participate in 20th ISOSS Conference held in Lahore.

REFERENCES

1. Alabi, O.A., Ologbonjaye, K.I., Awosolu, O. and Alalade, O.E. (2019). Public and environmental health effects of plastic wastes disposal: a review. *J. Toxicol Risk Assess*, 5(021), 1-13.
2. Atiwesh, G., Mikhael, A., Parrish, C.C., Banoub, J. and Le, T.A.T. (2021). Environmental impact of bioplastic use: A review. *Heliyon*, 7(9), e07918.
3. Lambert, S. (2013). Environmental risk of polymer and their degradation products (Doctoral dissertation, University of York). https://etheses.whiterose.ac.uk/4194/1/EnvironPoly.pdf
4. Moore, C. (2023). *Plastic Pollution*. *Encyclopedia Britannica*. https://www.britannica.com/science/plastic-pollution
5. Ojeda, T. (2013). *Polymers and the Environment*. Retrieved from intechopen: https://www.intechopen.com/chapters/42104

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 115-130

THE ENVIRONMENTAL SITUATION OF DIAMOND PAINT INDUSTRY (PRIVATE) LIMITED

Mohammad Rafiq Khan[§], **Tayyaba Ishaq, Sarah Zahid, Hashir Shafique** and **Hasham Khalid**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

The article being presented here is a component of the pilot project," The Environmental Problems of the Industrial Sector of Pakistan" conceived by our Class of Researchers composed of different groups. Each group was deployed to focus on one industrial unit to highlight its environmental problems and make recommendations to solve them. This article is titled as "Environmental Situation of diamond paint Industries". The purpose of this article is to gather information regarding environmental situation of Diamond Paints. For this purpose both secondary and primary data were collected. Secondary data was compiled from various articles on this area of concern while the primary data was collected through Quantitative analysis, such as online questionnaires and structured interviews. Then the data was analyzed through computer software. The result showed that 72.7% employees can assess the risks that exists in the workplace. 81.8% employees were agreed that they are provided with personal protective equipment during work. The most of the public (43.8%) agreed that they are having different types of respiratory diseases that are associated with the chemicals in paint industry. The study reveals that Diamond Paints, is a cooperative firm that abides by the rules and regulations tuning with the environment's protocols. From treating waste disposal to taking care of employees, they are doing their level best. The problem lies with the government as they need to regulate all other industries to follow the rules and regulations.

1. INTRODUCTION

Paint comprises of pigments (the color) carried by a resin that holds the pigment particles together and provides adhesion to the paintedsurface, some additives and a solvent to aid in applying the color, and a dryer. The plastic compounds such as formaldehyde, arsenic, thinners, etc., are present in vinyl and acrylic paints. Pigments add color, conceal imperfections, and control gloss. Dyes are typically classified into two types: Stains such as Titanium Dioxide (white), Chrome Green Oxide (Yellow) and Red Iron oxide, and others are included in the 'Prime Pigments' category while, Calcite (Calcium Carbonate), Talc (Magnesium Silicate), Mica, Barites (Barium Sulphate), and other pigments are classified as 'Extender Pigments.' Thus it is observed that due to an extended spectrum of compounds involved, there is above average contribution of this industry to the environment (Kwaambwa et al., 2013).Apart from the studies cited above, some recent studies have been also conducted in context of sustainability and disposal of wastes (Assaad, 2015, Assaad, 2017, Avei, et al., 2017).

For industries in Pakistan, the Environmental Protection Act (PEPA) of 1997 based on Pakistan's Environmental Policy is applied, which is based on a participatory approach to accomplish long-term development goals through legally, administratively, and technically competent institutions. On December 6, 1997, the Pakistan Environmental Protection Act was passed, abolishing the Pakistan Environmental Protection Ordinance 1983 (Geurts et al., 2008). The PEPA 1997 establishes the framework for implementing NCS, establishing Provincial Sustainable Development Funds, protecting and conserving species, conserving renewable resources, establishing Environmental Tribunals and appointing Environmental Magistrates, Initial Environmental Examination (IEE), and Environmental Impact Assessment (EIA). Environmental Protection Council of Pakistan Thehighest body was established in 1984 under Section 3 of the Pakistan Environmental Protection Ordinance (PEPO), 1983, with the President of Pakistan as the chairperson, as Chairman of the Board. The Ordinance was amended in 1994 to make the Prime Minister or his nominee the Chairman of the Council. The Council was reconstituted following the passage of a new law, the Pakistan Environmental Protection Act of 1997. It is led by Pakistan's Prime Minister (Chief Executive). Trade and industry, influential NGOs, educational institutions, specialist media, and related ministries are all represented on the Council. Section 5 provides for the establishment of the Pakistan Environmental Protection Agency. Under Section 6 (d) of the Pakistan Environmental Protection Ordinance, 1983 (Xiong et al., 2013).

Before, the current study was planned it was deeply felt that the country should start from the first principles that involve a strong convincing baseline to erect on it the brick by brick strategic control model for correction of the environmental situation in Pakistan. The thought led the conclusion that different sectors of Pakistan maybe targeted to highlight their environmental situation (Ohama et al., 2020). Here the target was industrial sector that ultimately led to work on “Environmental Situation of the Industrial Sector of Pakistan.” This being multi-group activity, different groups were created to work on different industrial set ups to highlight their environmental situation and to recommend to the concerned circles essential steps to implement the results with feedback to the recommending experts (Dursun et al., 2006).

The gap the researchers are trying to fill is how the wastage of the paint industry, focusingon Diamond Paints as model, deals with wastage. This will include their effort towards sustainability, theiremployees' health, and their attitude towards the conservation of the environment because of their production.

The Goal and Objectives

The goal of the study was to highlight the environmental situation of Pakistan while the objectives of this research are listed below:

1. To understand the impact of industrial waste from Diamond Paint industries, primarily on the environment.
2. To understand whether the government or industry owners are aware of their carbon footprint, emission of greenhouse gases and other pollutants.
3. To know whether the government or the factory owners are taking steps towards reducing the toxins being released.

2. RESEARCH METHODOLOGY

Justification of the Choice of Industry

As established through the literature review, Diamond Paints is the most popular choice among Pakistanis. In 2018, they opened Sunder Industrial Estate, Pakistan's most extensive integrated paint manufacturing facility. Diamond Paints products are divided into the following categories:

1. Exterior: For exterior reproduction with excellent grade paint that provides outstanding durability and protection. They offer a wide range of colours to give your outside a remarkablefirst impression and to serve as inspiration.
2. Inside: For reimagining the interior with top-quality paint that provides excellent coverage forthe bedroom, living room, dining room, bathroom, and kitchen. Walls will come to life with a breathtaking array of gorgeous hues that create the ideal space .metal surface:
3. They provide long-term protection against dampness, rust, and corrosion for metalworking tasks. Restore and restore metal surfaces using Diamond solutions that provide adurable finish and excellent color retention.
4. Surface Preparation: Any painting project requires surface preparation. Their Base primers aremainly developed to pull out superior quality and colour richness from topcoats with fewer layers, saving time and money
5. Industrial Product: A diverse variety of items to meet your commercial and industrial projectrequirements

Sampling Techniques

It was planned to conduct a structured interview with the General Manager Faheem A. Khan of Diamond Paints. As the questions were all the same, we were able to easily interpret or notice variations in answers. It facilitated the decision-making and analysis of the data gathered.

The Google forms were used to collect data as it was an inexpensive option to collect quantitative data required as aid to the targeted research. The questionnaire was posted on our social media, emailed,and sent to our sired sample. Questionnaires were inexpensive but also provided a speedy approach to obtaining findings. The vast research issue necessitated both primary and secondary data. Quantitative analysis, such as online questionnaires and structured interviews, were used to collect preliminary data. The secondary data was compiled from various articles, published by different experts in the field and posts on the websites.

Methodological Tools

The purpose of the survey was to gather information needed to forecast the environmental situationof Diamond Paints in Kot Lakhpat. The survey questionnaires were designed using Google Forms for both employees to take an insight into knowing how they interact with the chemical waste of industry,and the public to take an insight into how they interact with pollution caused by the industry, and how the breakdown meddles with their daily routine. A total of 10 employees and 15 random people were chosen to accompany the survey. The people were questioned regarding their demographic and socio-economic backgrounds. Apart from the demographical questions, the second section of the questionnaire included questions that helped researchers understand their problem statement.

Analytical Method

The sampling technique used while selecting the sample for interviews was simple random sampling which falls under probability sampling. The sampling frame was middle-aged employees who probably already had the experience of long-term policies implemented by Diamond Paints through the years to improve their environmental situation. Due to ethical considerations, no details regarding the participants privacy were asked.

3. RESULTS AND DISCUSSION

Based on collected data and their results thediscussion and analysis are being done which comprises three parts that are:

a. Interview with Mir Shoaib (CEO of Diamond Paints)
b. Employee Survey
c. Public Survey

Interview with Mir Shoaib (CEO of Diamond Paints)

A one-on-one conversation with Mir Shoaib at Diamond Paints' Head Office at Kot Lakhpat was held. A valuable insight into the industry through his valuable answers was obtained.

Mir Shoaib told us that to him; “Paint is anything used for decoration or protection; It is used in various places, and there are different types of paints. However, our main product is building paint which comes under the decorative segment, which is also the main segment of other paint manufacturers as it shares 70-80% of the industry. Industrial paints are also manufactured in ourindustry, which is used to paint Fans, Gates, and all products made up of metal, and give them abetter look and protect it from rust. Then we have different protection coatings, which include epoxies and rigs used for heavy industries. He also gave us insight into how he sells his products, markets them, and the new items he has introduced.”

The most important thing he told was about how diamond paints are environment-friendly and are being made

He told that the industry has to make sure that they are lead-free, as when they peel off after a certain age, children might put that peal in their mouth so that it does not affect their health as well. Another thing is that in households, paints used are water-based as in foreign countries it isseen VOC (Volatile emission content) and it is controlled quite seriously. However, unfortunately, here there are no regulations like that, but applied, there might be better results. He emphasized fire-resistant Paints as they are essential as paints are highly flammable, and “we have been working with them for quite a long period as one of our customersis government supplies which include the Pakistan army, navy, and ordinance factories.

However, their paints are specially designed as the Pakistan navy used wooden ships, and for that, there is special heat-resistant paint for them, and we have to clear out tests before selling itto them, which is done by giving flame to the paint. However, there is significantly less demandfor it in the local market, whereas in developed countries where insurance policies are made so those companies make sure that fire-resistant paint is used as their houses are made of wood aswell, whereas here concrete houses are made so people also tend to use it less. We are also reaching out to insurance companies here so that this is introduced to people so that wherever wood is used fire resistant paint must be used there to avoid damage in case of fire.

Employee Survey Results

The employee's survey results are exhibited in Figure (1) to Figure (7), discussed below collectively under the pie charts.

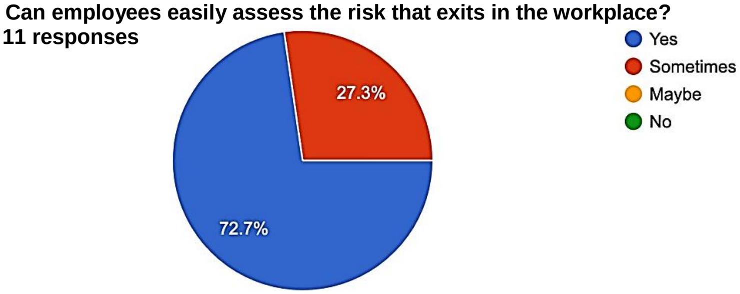


Figure (1)

While conducting our survey at Diamond Paints, specifically with employees, when asked ifthey could quickly assess the risks in the workplace. 72.7% of employees answered with affirmation, as shown in Figure (1). Hazardous substances can significantly threaten employees' health, safety, and well-being. They must be adequately trained and informed about potential hazards they face in their working day to handle any hazardous substances they come into contact with safely.

Under the Diamond paint company policy, they made sure that employees are aware of the proper safety procedures around hazardous substances is a vital part of this, helping workers tostay safe during their working lives and employers to stay on the right side of the law.

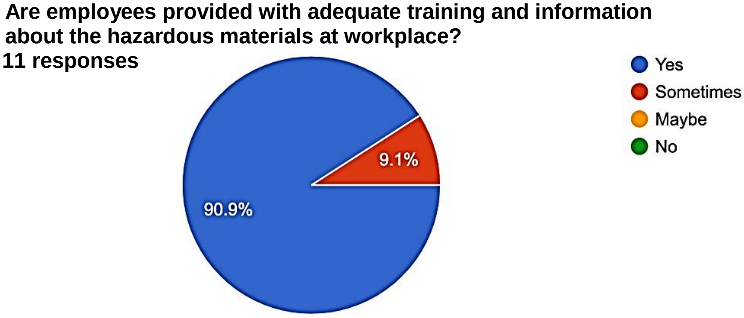


Figure (2)

The second question asked was whether employees were provided adequate training and information about hazardous materials in the workplace. Again, the majority of employees, 90.9%, responded Yes, as shown in Figure (2). Safety is the essential duty that Diamond paints consider to have towards their staff to make sure they return home safe and sound at the end of their working day. When employees are informed of the possible dangers of substances they useduring their work, they are more likely to follow the safety procedures in place and less likely toendanger themselves and others through the improper handling of hazardous substances in an emergency such as a chemical spillage or a colleague ingesting a hazardous substance, every second counts. When staff is adequately trained, they will know the correct way to handle the situation and improve the outcomes for everyone involved (Ohama et al., 2020).

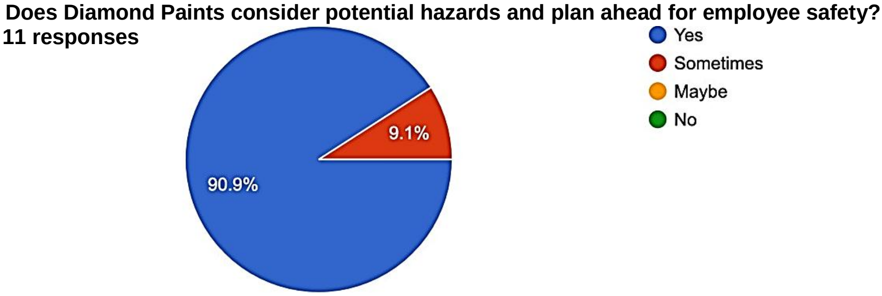


Figure (3)

When the employees were asked, “Does Diamond Paints consider potential hazards and plan ahead for employee’s safety? To this question, 90.9% responded with Yes, as shown in Figure (3). At DiamondPaints, a set of procedures is followed when young workers are employed for the first time.

For many young people, the workplace will be a new environment, and they will be unfamiliar with 'obvious' risks and expected behavior. They may lack experience or maturity. Ensure they understand what is expected of them, and check that they understand and can remember and follow instructions. They may not have reached physical maturity and be more at risk if their muscle strength is not fully developed. They may need to be more skilled in handling techniquesor pacing work according to their ability. In high-risk environments, it is considered the work they will be doing or observing, the risks involved, and how they are managed. Diamond Paints considers specific factors that must be managed for young people, including exposure to:

- Radiation
- Noise and vibration
- Toxic substances
- Extreme temperatures

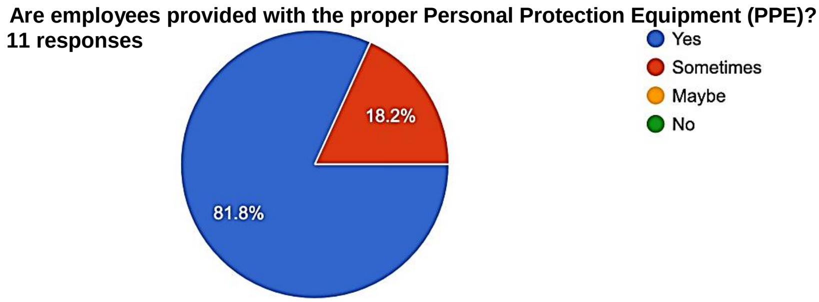


Figure (4)

Employees were also asked What the training timeframe was provided at Diamond Paints, to which multiple answers were 2 hours each Thursday, Weekly training, two months to 8 months.The level of training depended on the task employees were employed to complete. Some tasks were of the nature that required in-depth training, while other tasks were relatively easier tolearn.

At Diamond Paints, many options for training your staff in hazardous substances are available. Increasingly, employers are turning towards eLearning courses, attracted by the effectiveness ofsuch methods, their value for money and their time efficiency compared to alternative methods.Diamond Paints effectively train their employees in the dangers and uses of hazardous substances that are safer to work for and avoid negative personal, financial, and legal repercussions (Avco et al., 2017). Varies from industry to industry, specific equipment types apply to certain types of work. ThePPE kit at Diamond Paints included Eye Protection.

Goggles and protective glasses are the most common equipment for the eyes and facial area nearthem. Even in office and lab work, where there is the possibility of flying debris or splashing, eye protection is a must. Ear Protection, where there is an extreme amount of sound exposure regularly, employees have to be provided ear protection. Hands and Arm Protection Limb protection is frequently provided in arm guards and gloves. People still have to be able to do their work, but paint material is often used to prevent immediate lacerations or temperature- related injuries such as burns. Feet and legs in the case of feet, the protective gear is often in the form of boots with protective soles, rugged toe protection, and thick layer construction to avoid laceration or temperature exposure. Lungs Where employees are exposed to environments with hazardous material that can be inhaled, breathing equipment is a must. This may be as simple as a mask with filters or as elaborate as a complete breathing system and connected oxygen tank.

Whole body in the case of continuous or repetitive exposure to harmful environments, completebody protection is required Assaad et al., 2017).

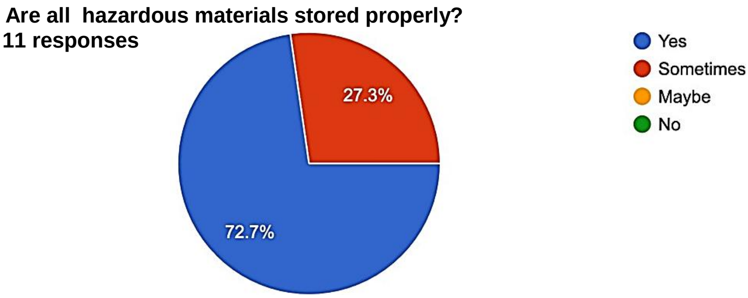


Figure (5)

Employees were also asked whether Diamond Paints considers potential hazards and plans for employee. The questionnaire asked whether our employees had the proper personal protection equipment(PPE). 81.8% of employees responded with Yes, as shown in Figure (5). Personal protective equipment, or PPE, protects its user against any physical harm or hazards that the workplace environment may present. It is important because it exists as a preventative measure for paint industries that are known to have hazardous substances. It is most effective when it meets its employees' correct size, fit, and height. The Different Types of PPE Equipment Because PPE.

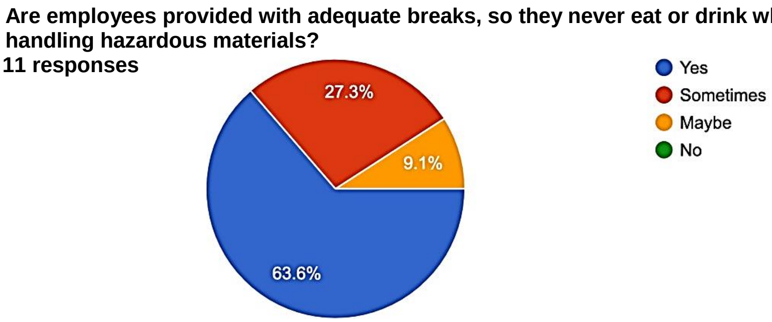


Figure (6)

Furthermore, employees were also asked if all hazardous materials were stored properly. To this, 72.7% responded with Yes, as shown in Figure (6). At Diamond Paints, it has been established that those utilizing hazardous substances will have to store and contain them safely in order to prevent an incident taking place which has the potential to put at risk the health and safety of workers, site visitors, people living and working nearby, plus creatures and plants living in the surrounding environment.

Employees were also asked "Are they provided with adequate breaks, so they never eat or drink while handling hazardous materials?" To this question, 63.6% answered. Yes, while others needed clarification about the question. At Diamond Paints, there is a Designated location for food and drink. Moreover, it is Forbidden to eat and drink around the work area to prevent workers from accidentally ingesting chemicals. They have created a location far from the chemical storage and worksite for eating, drinking, and breaks. Near this spot, they give employees ready access to sinks and soap for hand washing before eating or drinking. This washes off any possible chemicalresidue from the hands and can prevent accidental ingestion.

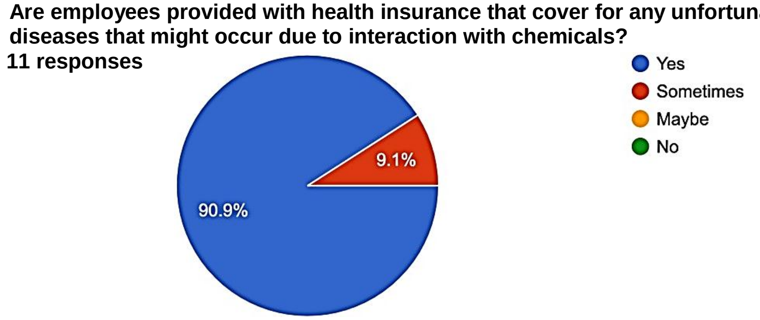


Figure (7)

Lastly, employees were asked if they are provided with health insurance covering any unfortunate diseases that might occur due to chemical interaction. To this, 90.9% responded witha Yes as shown in Figure (7). At Diamond Paints, health benefits are considered necessary as they can also deliver a range of benefits, such as increasing productivity, boosting morale, and helping shape a positive company culture. There is a wide range of options available at DiamondPaints to provide employee medical insurance, but, generally, group health benefits will usually come in the form of:

- Basic health benefits: Outpatient and inpatient care, coverage for dependents, international/regional coverage, and inclusion of pre-existing conditions.
- Popular additional benefits: Maternity and fertility coverage, vision, and dental insurance.
- Extended/comprehensive benefits: Health checks and immunizations, general wellness,and mental health support.

The main point of a medical insurance plan for employees is to protect and support the healthand well-being of staff so they can remain active and productive members of their company (Ismail et al., 2011).

Responses from Public Survey

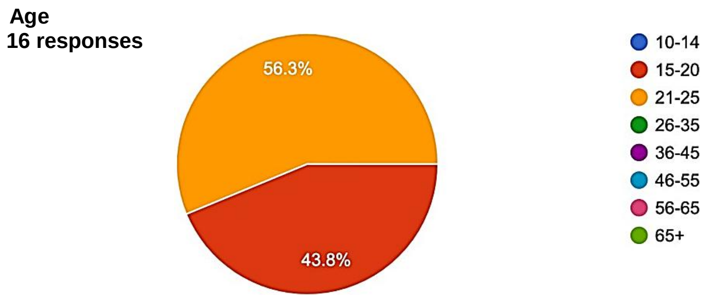


Figure (8)

The public survey began with demographic questions that were asked of 16 people. Most of thepeople who participated in this survey were aged 15-20 and 21-25, as shown in Figure (8).

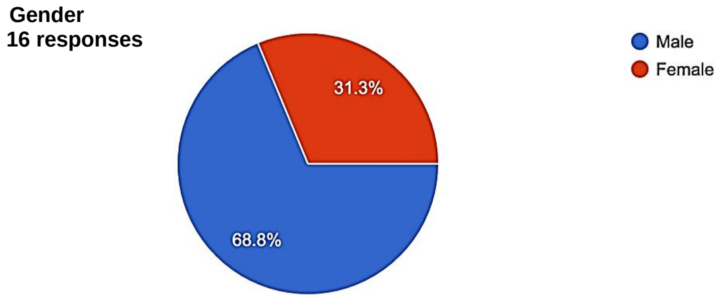


Figure (9)

The participants in this survey were 68.8% Female while 31.3% of them were male, as shown in Figure (9).

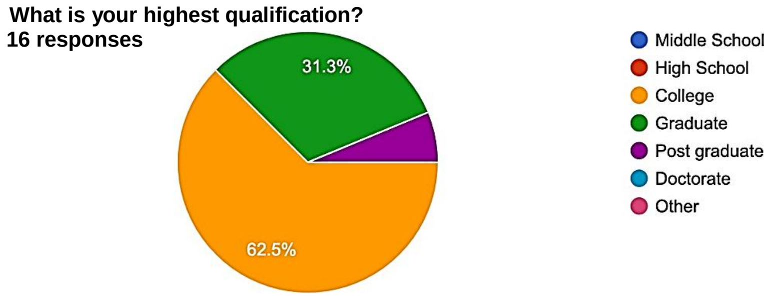


Figure (10)

When asked about the highest qualification of participants, 62.5% of them responded with College, while 31.3% were undergraduate and 6.3% were Post-graduate-d, as shown in Figure (10).

Considering their knowledge about the risks associated with environmental pollution, how likely they will recommend a friend or a colleague to promote activities that will help reduce global warming or to live a clean and green life

Considering your knowledge about the risks associated with environmental pollution, how likely are you to recommend a friend or a colleague to promote activities that will help reduce global warming or to live a clean and green life? 0 Very Unlikely to 10 Very Likely
16 responses

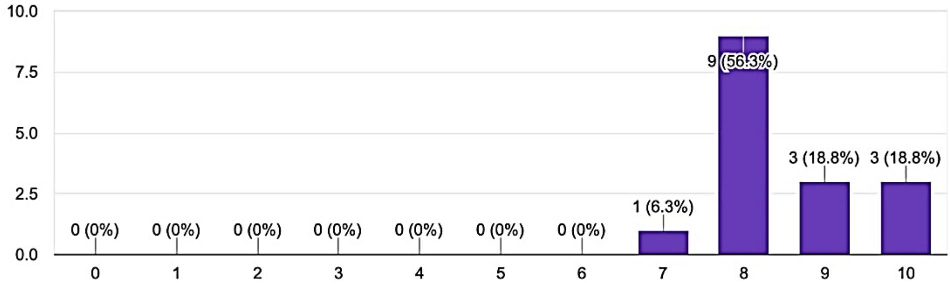


Figure (11)

The participants were questioned, Considering their knowledge about the risks associated with environmental pollution, how likely they will recommend a friend or a colleague to promote activities that will help reduce global warming or to live a clean and green life (Geurts et al., 2008). On a scale from 1to 10, 1 being very unlikely while ten being likely. To this, 56.3% recommended a point 8, as shown in Figure (11). This could be traced to greenhouse gases, deforestation, and everything else that will come back and affect us somehow. Built, and now the pollution is killing thousands or millions worldwide. Mass deforestation leads to biodiversity loss throwing the earth's system off balance. Increased greenhouse gases in the atmosphere are causing climate change, and sooner or later, everyone will be experiencing heat waves, floods, drought, or other extreme events. Even the plastic we invented now makes its way into our bloodstream, hidden as micro-plastics. Who knows what effects this could have on health?

On which of the following platform, have you heard about environmental pollution? Select all that apply:
16 responses

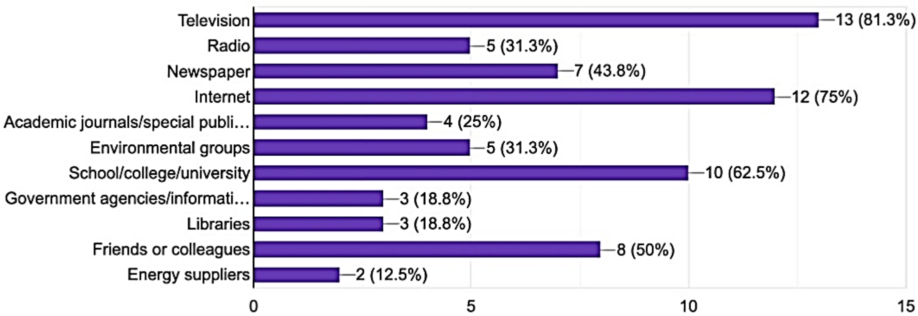


Figure (12)

The participants were asked on which of the platforms have they heard about environmental pollution. To most popular answers were Television (81.3), Internet (75%), Educational institutes (62.5%), and Friends and colleagues (50%), as shown in Figure (12). This can be traced down that these mediums are most commonly used and hence hold power to mold the perceptionof people interacting with them. Another similar question was asked participants about the level of trust in information about environmental pollution if they were to receive it from the following? The top positions were held by Families, Scientific Journals, and Blogs as shown in Figure (13). This can be linked to the fact that people either trust the word of the family as they sharetrust with them or are verified and published sources to trust only.

What is the level of trust on information about environmental pollution,
if you were to receive it from the following?

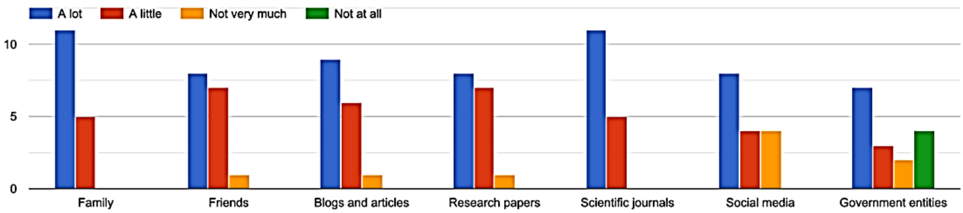


Figure (13)

Participants were asked On a scale of 1-10, please rate how environmental pollution affects their health and safety, where 1 affects them the least and 10 affects the most. This majority respondedon point 9, which makes 43.8%, as shown in Figure (13). This can be linked to the fact that **Environmental pollutants can cause health problems like respiratory diseases, heart disease, and cancer**. Low-income people are more likely to live in polluted areas and have unsafe drinking water. Children and pregnant women are at higher risk of health problems related to pollution, so individuals tend to care about this aspect far more.

On a scale of 1-10, please rate how environmental pollution affects your personal health
and safety where 1 affects you the least and 10 affects you the most
16 responses

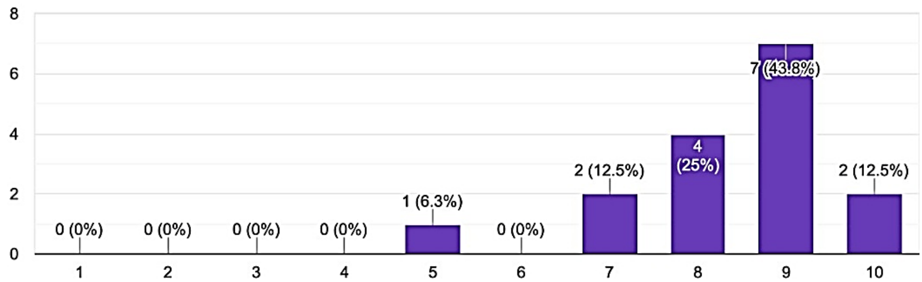


Figure (14)

Furthermore, participants were asked to Please state their concern with the type of pollution. Tothis, the most popular answer among respondents was Air pollution and Ozone layer depletion, as shown in Figure (14). This could be linked to the fact that these are the most prevalent environmental issues in Pakistan; air pollution as Smog is prevalent in Lahore, and Ozone Depletion played a crucial role in the worst floods of 2022 that Pakistan experienced.

Please state your concern with the below statements:

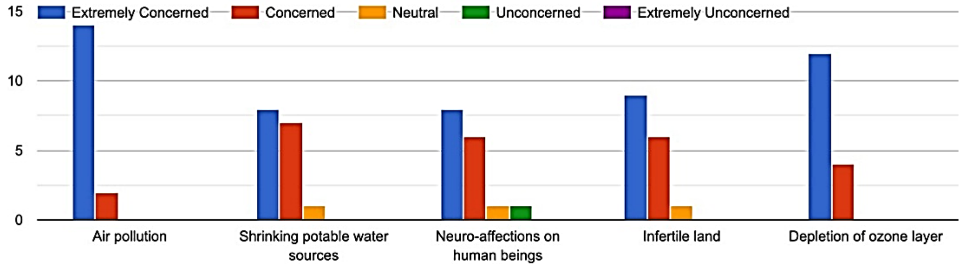


Figure (15)

Participants were also asked Which of the following option given to them they think the impact and are the biggest causes of environmental pollution. To this, the most popular responses were of Industrial activities (87.5), Plastic consumption (68.8%), and vehicle usage (62.5%), as shownin Figure (15).

Which of the following do you think impact and are the biggest causes of environmental pollution?
16 responses

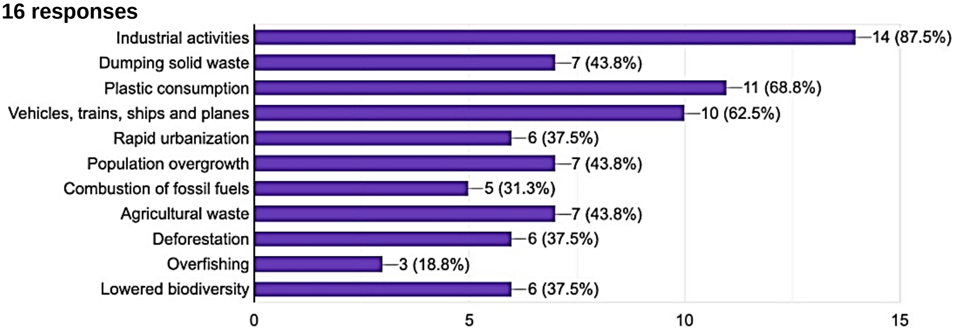


Figure (16)

Moreover, the participants were asked Whether they think the government's laws that restrict pollution are as they are. To this, the majority responded 81.3% responded that the laws are as shown in Figure (16). This may be translated to the fact that it is easier to evade environmental laws in Pakistan; hence the government needs to keep a strict check.

Do you think your Government's laws that restrict pollution be as they are?
16 responses

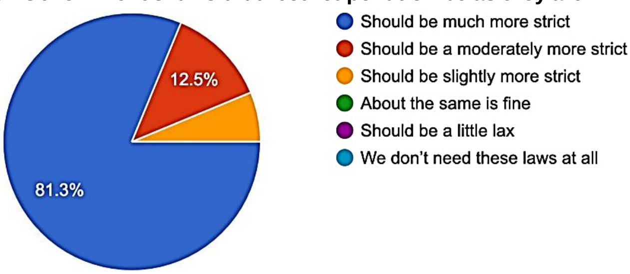


Figure (17)

Another similar question was asked. Do they think the government is spending too much time, money, and resources to reduceenvironmental pollution? To this question, 56.3% responded that the government needs to do more, as shown in Figure (17). This can be linked to the fact that in Pakistan, the effort for the environment made last time was In December 1997. Pakistan Environmental Protection Act (PEPA'97) was signed and promulgated by the President of Pakistan. It provides for the protection, conservation, rehabilitation, and improvement of the environment, for the preventionand control of pollution, and promotion of sustainable development. PEPA'97 covers nearly all issues from pollution generation to pollution prevention, monitoring to confiscation, compliance to the violation, and prosecution to penalization. However, the results of this legislation have been subjected to virtuous and unadulterated implementation and since this, there have yet to be successful efforts.

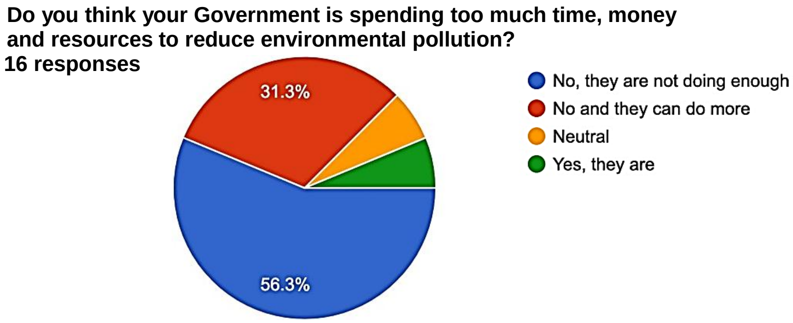


Figure (18)

Towards the end of the questionnaire, it was asked whether they think remedial action on environmental pollution should be the most crucial aspect to take care of. The majority, 56.3%, responded yes, as important as a few other issues, as shown in Figure (18). This means that people are looking for institutional support for environmental problems.

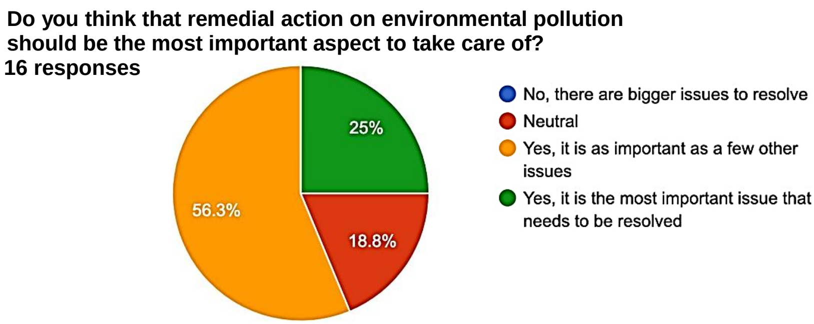


Figure (19)

The survey also explored that participants think the entities are taking initiatives to reduce environmental pollution. The most popular answers were from international organizations and the environmental group, as shown in Figure (19). This can be linked to the fact that the most active organizations for this cause are like WWF. IUCN, UNEP, and SDF, and all of them are internationally registered.

Do you think the following entities are taking initiatives to reduce environmental pollution?

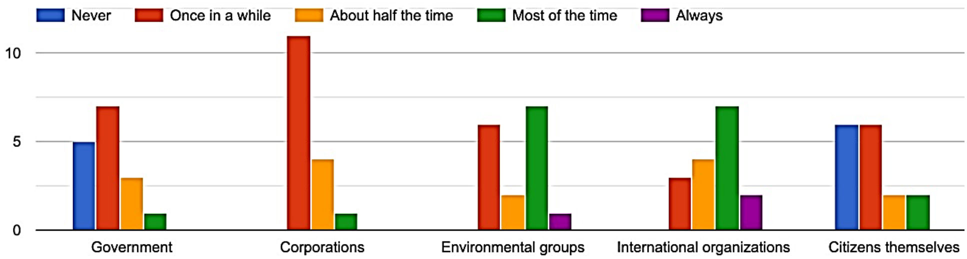


Figure (20)

In the end, participants were asked how interested they were in preserving the environment. To this, 50% responded with interest, while the second most popular answer was exciting, with 31.3%, as shown in Figure (30). This can be translated as people who are educated and mature carefor their environment and aims to preserve it.

4. COMMENTS AND CONCLUSION

The study reveals that Diamond Paints, as a cooperation, abides by the rules and regulations thatsit with the environment's protocols. From treating waste disposal to taking care of employees, they are doing their level best. The problem lies with the government as they need to regulate all other industries to follow the rules and regulations. The environmental situation of Kot Lakhpat and, majorly, Lahore will not be better by only one single company working hard. Rather, a dominant effort is required. The government needs to make strict laws and penalize industries not abiding by them. Furthermore, people's attitude is progressive, which means changes are looked forward to but with institutionalized support only.

5. ACKNOWLEDGEMENT

We are extremely grateful to Dr. Shahid Amjad Chaudhry, Rector, Lahore School Economics for patronization of our research and sanction of expenditure involved and providing us requisite facilities to participate in 20th ISOSS Conference held in Lahore

REFERENCES

1. Assaad, J.J. (2017). Use of polyvinyl acetate polymers in sustainable concrete made with recycled aggregates. In *European Advanced Materials Congress*, DOI: 10.5185/eamc.2017
2. Assaad, J.J. (2015). Effect of waste latex paints on rheological properties of cement pastes: Compatibility with water reducers. *Journal of Materials in Civil Engineering*, 27(12), 04015056. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001343

3. Avci, H., Ghorbanpoor, H., Topcu, I.B. and Nurbas, M. (2017). Investigation and recycling of paint sludge with cement and lime for producing lightweight construction mortar. *Journal of Environmental Chemical Engineering*, 5(1), 861-869. https://doi.org/10.1016/j.jece.2017.01.009
4. Dursun, D. and Sengul, F. (2006). Waste minimization study in a solvent-based paint manufacturing plant. *Resources, Conservation and Recycling*, 47(4), 316-331. https://doi.org/10.1016/j.resconrec.2005.12.004
5. Geurts, J., Bouman, J. and Overbeek, A. (2008). New waterborne acrylic binders for zero VOC paints. *Journal of Coatings Technology and Research*, 5, 57-63. https://doi.org/10.1007/s11998-007-9036-x.
6. Ismail, Z.Z. and Al-Hashmi, E.A. (2011). Assessing the recycling potential of industrial wastewater to replace fresh water in concrete mixes: application of polyvinyl acetate resin wastewater. *Journal of Cleaner Production*, 19(2-3), 197-203. https://doi.org/10.1016/j.jclepro.2010.09.011
7. Kwaambwa, H.M. (2013). A review of current and future challenges in paints and coatings chemistry. *Progress Multidisciplinary Research Journal*, 3(1), 75-101.
8. Ohama, Y. (1995). Handbook of polymer-modified concrete and mortars: properties and process technology. *William Andrew*, 20(1), 28-33.
9. Xiong, J., Wang, L., Bai, Y. and Zhang, Y. (2013). Measuring the characteristic parameters of VOC emission from paints. *Building and Environment*, 66, 65-71. https://doi.org/10.1016/j.buildenv.2013.04.025.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 131-142

ON ESTIMATION OF KUMARASWAMY EXPONENTIATED INVERSE RAYLEIGH DISTRIBUTION USING BAYESIAN ANALYSIS UNDER NON-INFORMATIVE PRIOR AND INFORMATIVE PRIOR

Zulaikha[1], **Wajiha Nasir**[2] and **Muqadas**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]zoyamehar464@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

In this study, four parameters of Kumaraswamy's Exponentiated Inverse Rayleigh Distribution (KEIR) have been utilized. Using complete samples, we have estimated the shape parameter of Kumaraswamy Exponentiated Inverse Rayleigh Distribution using Bayesian analysis. Posterior distribution has been derived by using non-informative and informative priors. We have also studied the performance of posterior distribution by using different loss functions using a complete sample. A simulation study has been utilized to check the performance of the Bayes estimator and its corresponding risks by using different sample sizes. We have also drawn the graphs of the posterior distribution for different values of the scale parameter.

KEYWORDS

Bayesian analysis, MLE, Loss functions, Posterior Distribution under Non-Informative Prior & Informative Prior.

INTRODUCTION

In (1964) Trayer introduced a model named Inverse Rayleigh Distribution. Later, Voda worked on this distribution and discussed its properties and MLE of the scale parameter. Later worked by Gharraph, he finds mean, median, harmonic and geometric mean. Later different authors introduced different distributions from which a few models are Kumaraswamy Gumbel, Kumaraswamy Inverse Weibull, Kumaraswamy Weibull Distribution, Kumaraswamy GP Distribution, Kumaraswamy Pareto Distribution, and Kumaraswamy Birnbaum-Saunders Distribution. Ahsan-ul-Haq studied the properties of Kumaraswamy Power Function Distribution and Transmuted Power Function. Ahsan-ul-Haq also studied Kumaraswamy's Exponentiated Inverse Rayleigh Distribution and discussed its mathematical properties and applied it to real data to check the flexibility of its model.

Hiba et al. (2015), The proposed estimator was compared with the corresponding Bayes estimators obtained under squared error loss function and maximum likelihood estimators through their simulated risks. The exponentiated gamma distribution is one of the important families of distribution in lifetime sets. In this paper, a new generalized version of this distribution which is called Kumaraswamy exponential gamma distribution is introduced.

A new distribution was more flexible and has some interesting properties. A comprehensive mathematical treatment of the KEG distribution was provided. We discuss the maximum likelihood estimation of the distribution parameter. Finally, an application to real data sets is illustrated.

Ahsan-ul-Haq (2016) has studied four-parameter Kumaraswamy Exponentiated Inverse Rayleigh distribution (KEIR) by using mathematical properties to derive distribution including mean, moment generating function, quantiles, and renyi entropy on real data set. Using the maximum likelihood estimation method to illustrate its potential and flexibility. The real data set is better than the other models.

Ahsan-ul-Haq (2016) has studied the generalized form of Inverse Rayleigh distribution derived by using Quadratic Rank Transmutation Map(QTRM) is called Transmuted Exponentiated Inverse Rayleigh(TEIR) distribution. They studied the mathematical properties of TEIR distribution. He has derived its moments, renyi entropy, random number generator, and quantile function. Maximum likelihood equations and the order statistics densities were derived. The utility of the distribution is optimized through a real data set application.

Thobias (2017) has introduced a new family of continuous distributions called the Marshall-Olkin Kumaraswamy (MOKS) distribution. They studied some mathematical properties and derive distribution by using order statistics, record value properties, 10 two types of entropies, and Zenga curves. They also constructed an autoregressive model with a minification structure and estimated it using the maximum likelihood method. An application to a real data set is studied.

Mustapha et al. (2018) have studied The Kumaraswamy Exponentiated U-Quadratic Distribution in which mathematical and statistical properties are derived. This distribution fails to hold maximum likelihood estimates (MLEs). In this situation, they used two different methods: alternative maximum likelihood estimation (AMLE) and modified maximum likelihood estimation (MMLE). These studies were conducted to assess the finite sample behavior of the AMLEs and MMLE. H.F.

Bashiru et al. (2021) have studied Bayes Estimators of Exponentiated Inverse Rayleigh Distribution using Lindley’s Approximation. They use both the frequentist and Bayesian methods Bayes estimators of the unknown shape and scale parameters and have adopted them to obtain the posterior distribution of the shape and scale parameters of the Exponentiated Inverse Rayleigh Distribution (EIRD) have been derived. Lindley approximation was adopted to obtain the parameters of interest. The Bayes estimate for the simulated data-sets and real-life data-sets were obtained. The Bayes estimates obtained under different loss functions are close to the true parameter value of the shape and scale parameters.

Jayalakshmi (2022) has studied Kumaraswamy Exponentiated Rayleigh Distribution by applying a Double Sampling Plan for life test using percentiles. It may be conducted to evaluate the smallest sample size to ensure a certain percentile lifetime of products. The minimum sample size values are calculated for various quality levels. The plan parameters and the measures are studied for a sampling plan suitable for the manufacturing industries for selecting samples. They calculated by using the method Percentiles, Producer’s risk.

KUMARASWAMY EXPONENTIATED INVERSE RAYLEIGH DISTRIBUTION (KEIRD)

The Kumaraswamy Exponentiated Inverse Rayleigh Distribution (KEIRD) has been developed by Ahsan-ul-Haq (2016). The p.d.f of Kumaraswamy Exponentiated Inverse Rayleigh Distribution is as follows:

$$f(x) = \frac{2ab\alpha\theta}{x^3}\left(e^{-\frac{\theta}{x^2}}\right)^{a\alpha}\left[1-\left(e^{-\frac{\theta}{x^2}}\right)^{a\alpha}\right]^{b-1} \qquad x \geq 0;\ a, b, \alpha, \theta > 0$$

Respectively, where one is the shape parameter and one scale parameter of the Kumaraswamy Exponentiated Inverse Rayleigh Distribution respectively according to Ahsan-ul-Haq (2016).

MODEL AND LIKELIHOOD FUNCTION

The pdf is as follows:

$$\mathrm{f}(x) = \frac{2ab\alpha\theta}{x^3}\left(e^{-\frac{\theta}{x^2}}\right)^{a\alpha}\left[1-\left(e^{-\frac{\theta}{x^2}}\right)^{a\alpha}\right]^{b-1}$$

The log-likelihood function is as follows:

$$\mathrm{L}(x; a, b, \alpha, \theta) = \frac{2^n a^n b^n \alpha^n \theta^n}{\prod_{i=1}^{n} x_i^3}\left(e^{-\left(\Sigma\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\left[1-\left(e^{-\left(\Sigma\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right]^{b-1}.$$

PRIOR DISTRIBUTION

The prior distribution is a method that includes informative and non-informative priors.

NON-INFORMATIVE PRIORS

Sometimes, the deviation of the prior distribution based on information other than the current data is impossible. In that case, we use non-informative priors. They are used when we don't have any information about the parameters. A non-informative prior is one in which little new explanatory power about the unknown parameter is 4 provided by intention. It is often improper. Some non-informative priors are given below:

USING UNIFORM PRIOR

The Uniform prior of parameter b is defined as:

$$p(b) \infty 1, 0 < b < \infty$$

The posterior distribution for parameter b is:

$$p(b|x) = \frac{-ln\left[1-\left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right]^{n+1}}{\Gamma(n+1)} b^{n+1-1} e^{-b\left[-ln\left\{1-\left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}\right]}$$

where $\alpha 1 = n + 1$ and $\beta 1 = -ln\left\{1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}$ which is density function of Gamma Distribution with parameter $\alpha 1$ and $\beta 1$.

USING JEFFREY'S PRIOR

The Jeffrey's prior of parameter b is defined as:

$$p(b) \infty \frac{1}{b}, 0 < b < \infty$$

The posterior distribution for parameter b is

$$p(b|x) = \frac{-ln\left[1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right]^{n+1}}{\Gamma(n)} b^{n+1-1} e^{-b\left[-ln\left\{1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}\right]}$$

where $\alpha 2 = n$ and $\beta 2 = -ln\left\{1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}$ which is density function of Gamma Distribution with parameter $\alpha 2$ and $\beta 2$.

INFORMATIVE PRIORS

When we have known the prior information about the parameter, then it will be our prior distribution. Sometimes these priors are not so simple to be used, in these cases, we have to use expert opinion. In this case, the expert opinions about the probability of an event must be determined by a proper form of a probability distribution. This process is called prior elicitation. Sometimes we use the distribution as prior information whose range is the same as the current distribution are as below:

USING EXPONENTIAL PRIOR

The Exponential prior of parameter b is defined as:

$$p(b|y) \propto \frac{1}{\theta} e^{-b/\theta}, \qquad 0 < b < \infty$$

The posterior distribution for parameter b is

$$p(b|y) = \frac{-ln\left[1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right]^{n+1}}{\Gamma(n+1)} b^{n+1-1} e^{-b\left[-ln\left\{1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}\right]}$$

where $\alpha 1 = n + 1$ and $\beta 1 = -ln\left\{1 - \left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}$ which is density function of Gamma Distribution with parameter $\alpha 1$ and $\beta 1$.

USING GAMMA PRIOR

The Gamma prior of parameter b is defined as

$$p(b|y) = \frac{b^{\alpha}}{\Gamma(\alpha)}(\alpha)^{\alpha-1}e^{-b\alpha}, \qquad 0 < b < \infty$$

The posterior distribution for parameter b is

$$p(b|y) = \frac{-ln\left[1-\left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right]^{n+\alpha}}{\Gamma(n+\alpha)} b^{n+\alpha-1} e^{-b\left[-ln\left\{1-\left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}\right]}$$

where $\alpha 2 = n+\alpha+1$ and $\beta 2 = -ln\left\{1-\left(e^{-\sum_{i=1}^{n}\left(\frac{\theta}{x^2}\right)}\right)^{a\alpha}\right\}$ which is density function of Gamma Distribution with parameter $\alpha 2$ and $\beta 2$.

BEs AND PRs UNDER DIFFERENT LOSS FUNCTIONS

This section contains Bayes estimators and posterior risks of different loss functions and also contains their results for informative and non-informative priors. Some loss functions are defined below:

Loss Function (LF)	Bayes Estimator (BE)	Posterior Risk (PR)
Square error (SELF)	$\alpha^* = \mathrm{E}(\alpha)$	$\rho(\alpha^*) = E_{\alpha\|x\{L(\alpha,\alpha^*)\}=\mathrm{Var}(\alpha)}$
Precautionary (PLF)	$\alpha^* = \sqrt{E(\alpha^2\|x)}$	$\rho(\alpha^*) = 2\left[\sqrt{E(\alpha^2\|x)} - E(\alpha\|x)\right]$
DeGroot (DLF)	$\alpha^* = \frac{E(\alpha^2\|x)}{E(\alpha\|x)}$	$\rho(\alpha^*) = \frac{Var(\alpha\|x)}{E(\alpha^2\|x)}$

EXPRESSION FOR BES AND PRS

This section contains expressions for BEs (Bayes estimators) and PRs (posterior risks) given under different loss functions using non-informative priors. They are derived using the expression in the section.

Expression For Bes and PRs Under SELF		
Prior	**Bes**	**PRs**
Uniform	$\frac{n+1}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{n+1}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Jeffrey’s	$\frac{n}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{n}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Exponential	$\frac{n+1}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{n+1}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Gamma	$\frac{n+\alpha}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$	$\frac{n+\alpha}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Expression For Bes and PRs Under PLF		
Prior	**Bes**	**PRs**
Uniform	$\frac{\sqrt{(n+1)(n+2)}}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{\sqrt{(n+1)(n+2)}-(n+1)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Jeffrey’s	$\frac{\sqrt{(n)(n+1)}}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{\sqrt{(n)(n+1)}-(n)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Exponential	$\frac{\sqrt{(n+1)(n+2)}}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$2*\frac{\sqrt{(n+1)(n+2)}-(n+1)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$
Gamma	$\frac{\sqrt{(n+\alpha)(n+\alpha+1)}}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$	$\frac{\sqrt{(n+\alpha)(n+\alpha+1)}-(n+\alpha)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$

Expression For Bes and PRs Under DELF		
Prior	**Bes**	**PRs**
Uniform	$\frac{(n+2)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{1}{(n+2)}$
Jeffrey's	$\frac{(n+1)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{1}{(n+1)}$
Exponential	$\frac{(n+2)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]}$	$\frac{1}{(n+2)}$
Gamma	$\frac{(n+\alpha+2)}{\left[-ln\left\{1-\left(e^{\sum_{i=1}^{n}\left(\frac{\theta}{x_i^2}\right)}\right)^{a\alpha}\right\}\right]^2}$	$\frac{1}{(n+\alpha+2)}$

SIMULATION STUDY

Simulation is the act of taking a model generally expressed as a computer program and running it several times to obtain results, then collecting the data and calculating the estimated probability. This section shows how the simulation can be helpful to approach the problems in the Bayesian analysis. We have conducted the simulation study to investigate the performance of the Bayes estimator and Posterior risks using different priors under three different loss functions. Based on the risk estimator the behavior of the loss functions is evaluated. Bayes estimator and the risks of these estimators have been computed and presented concerning these loss functions. We have simulated the random sample of sizes $n = 30,\ 50,\ 100$ and 500 of Kumaraswamy Exponentiated Inverse Rayleigh Distribution with parameters $b = 3,\ 5;\ a = 1,\ 1;\ \alpha = 2,\ 3$ and parameter $\theta = 1,\ 1$ for 10,000 times. Bayesian Analysis Kumaraswamy Exponentiated Inverse Rayleigh distribution is as follow:

BEs and PRs for SELF when {a, α, θ} ∈ {1, 2, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.2218	0.3462	5.3136	0.9416
50		3.1237	0.1954	5.2007	0.5415
100		3.0494	0.0929	5.0945	0.2594
500		3.0105	0.0181	5.0099	0.0502
30	**Jeffrey's**	3.0977	0.3305	5.1872	0.9286
50		3.0726	0.1930	5.1086	0.5329
100		3.0317	0.0928	5.0395	0.2565
500		3.0096	0.0182	5.0060	0.0502
30	**Exponential**	2.7439	0.2587	4.2060	0.6044
50		2.8382	0.1639	4.5313	0.4173
100		2.9264	0.0864	4.7454	0.2272
500		2.9845	0.0178	4.9359	0.0488
30	**Gamma**	2.9119	0.2780	4.4896	0.6582
50		2.9224	0.1693	4.6228	0.4230
100		2.9486	0.0865	4.7976	0.2292
500		2.9902	0.0178	4.9528	0.0490
BEs and PRs for SELF when {a, α, θ} ∈ {1, 3, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.2145	0.3452	5.3645	0.9607
50		3.1250	0.1957	5.2321	0.5475
100		3.0578	0.0935	5.1185	0.2619
500		3.0209	0.0181	5.0105	0.0502
30	**Jeffrey's**	3.0899	0.3282	5.1565	0.9177
50		3.0650	0.1918	5.1333	0.5385
100		3.0432	0.0935	5.0584	0.2585
500		3.0082	0.0181	5.0063	0.0502
30	**Exponential**	2.8133	0.0914	4.3298	0.1408
50		2.8749	0.0566	4.5957	0.0905
100		2.9213	0.0289	4.7939	0.0475
500		2.9736	0.0059	4.9494	0.0098
30	**Gamma**	2.8787	0.2720	4.4308	0.6425
50		2.9254	0.1697	4.6364	0.4250
100		2.9719	0.0879	4.8109	0.2304
500		2.9908	0.0178	4.9635	0.0492

BEs and PRs for PLF when {a, α, θ} ∈ {1, 2, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.2754	0.1032	5.4372	0.1713
50		3.1494	0.0609	5.2709	0.1018
100		3.0774	0.0302	5.1299	0.0504
500		3.0105	0.0060	5.0287	0.0100
30	**Jeffrey's**	3.1406	0.1021	5.2834	0.1718
50		3.0965	0.0610	5.1806	0.1021
100		3.0467	0.0302	5.0520	0.0501
500		3.0095	0.0060	5.0263	0.0100
30	**Exponential**	2.8133	0.0914	4.3298	0.1408
50		2.8749	0.0566	4.5957	0.0905
100		2.9213	0.0289	4.7939	0.0475
500		2.9736	0.0059	4.9494	0.0098
30	**Gamma**	2.9255	0.0911	4.5282	0.1410
50		2.9390	0.0564	4.6768	0.0897
100		2.9613	0.0290	4.8261	0.0472
500		2.9932	0.0059	4.9712	0.0099
BEs and PRs for PLF when {a, α, θ} ∈ {1, 3, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.2316	0.1018	5.3895	0.1698
50		3.1317	0.0605	5.1802	0.1001
100		3.0823	0.0303	5.1394	0.0505
500		3.0177	0.0060	5.0324	0.0100
30	**Jeffrey's**	3.1425	0.1022	5.2702	0.1714
50		3.0915	0.0609	5.1546	0.1015
100		3.0442	0.0302	5.0924	0.0505
500		3.0023	0.0059	5.0127	0.0100
30	**Exponential**	2.7995	0.0910	4.3056	0.1400
50		2.8549	0.0562	4.5689	0.0900
100		2.9225	0.0290	4.7718	0.0473
500		2.9841	0.0059	4.9580	0.0099
30	**Gamma**	2.9278	0.0912	4.4610	0.1390
50		2.9561	0.0567	4.6737	0.0897
100		2.9884	0.0292	4.8678	0.0472
500		2.9937	0.0059	4.9814	0.0099

BEs and PRs for DELF when {a, α, θ} ∈ {1, 2, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.2956	0.0313	5.4809	0.0313
50		3.1775	0.0192	5.2846	0.0192
100		3.0902	0.0098	5.1673	0.0098
500		3.0145	0.0019	5.0336	0.0019
30	**Jeffrey's**	3.2186	0.0323	5.3188	0.0323
50		3.1386	0.0196	5.2345	0.0196
100		3.0560	0.0099	5.0932	0.0099
500		3.0163	0.0019	5.0152	0.0019
30	**Exponential**	2.8514	0.0322	4.3874	0.0322
50		2.9124	0.0196	4.6167	0.0196
100		2.9486	0.0099	4.8183	0.0099
500		2.9919	0.0019	4.9483	0.0019
30	**Gamma**	3.3536	0.0309	5.5422	0.0309
50		3.2222	0.0191	5.2990	0.0191
100		3.0987	0.0097	5.1902	0.0097
500		3.0242	0.0019	5.0183	0.0019
BEs and PRs for DELF when {a, α, θ} ∈ {1, 3, 1}					
n	**Priors**	**b=3**		**b=5**	
		BE	**PR**	**BE**	**PR**
30	**Uniform**	3.3056	0.0313	5.5134	0.0313
50		3.1680	0.0192	5.3019	0.0192
100		3.0757	0.0098	5.1253	0.0098
500		3.0153	0.0019	5.0336	0.0019
30	**Jeffrey's**	3.2337	0.0323	5.3461	0.0323
50		3.1056	0.0196	5.2097	0.0196
100		3.0565	0.0099	5.1025	0.0099
500		3.0171	0.0019	5.0235	0.0019
30	**Exponential**	2.8326	0.0322	4.3686	0.0322
50		2.8984	0.0196	4.6058	0.0196
100		2.9354	0.0099	4.7683	0.0099
500		2.9857	0.0019	4.9621	0.0019
30	**Gamma**	3.3559	0.0309	5.5622	0.0309
50		3.3579	0.0191	5.3516	0.0191
100		3.1033	0.0097	5.1493	0.0097
500		3.0241	0.0019	5.0386	0.0019

We have used this simulation technique to check the behavior of the scale parameter of Kumaraswamy Exponentiated Inverse Raleigh Distribution. The above tables show the result of posterior distribution under Uniform, Jeffrey's, Exponential and Gamma priors by using different loss functions which are the Square error loss function, Precautionary loss function, and DeGroot loss function. After the simulation and comparing these loss functions the results show that as we increase sample size Bayes estimator approaches its true value of the parameter and posterior risk tends to decrease. From table we can see that DELF is performing better because its posterior risk is minimum as compared to other loss functions. Similarly, in the above table DELF is performing better. Whereas, the overall results show that the posterior distribution under Uniform and Gamma prior with DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors.

APPLICATION

In this study, we applied the proposed distribution on real data set of strength measured in GPA for single carbon fibers and impregnated 1000-carbon fiber tows. The data set consists of, 63 observations:

1.901, 2.132, 2.203, 2.228, 2.257, 2.350, 2.361, 2.396, 2.397, 2.445, 2.454, 2.474, 2.518, 2.522, 2.525, 2.532, 2.575, 2.614, 2.616, 2.618, 2.624, 2.659, 2.675, 2.738, 2.740, 2.856, 2.917, 2.928, 2.937, 2.937, 2.977, 2.996, 3.030, 3.125, 3.139, 3.145, 3.220, 3.223, 3.235, 3.243, 3.264, 3.272, 3.294, 3.332, 3.346, 3.377, 3.408, 3.435, 3.493, 3.501, 3.537, 3.554, 3.562, 3.628, 3.852, 3.871, 3.886, 3.971, 4.024, 4.027, 4.225, 4.395, 5.020.

Priors	**SELF**		**PLF**		**DELF**	
	BE	**PR**	**BE**	**PR**	**BE**	**PR**
Uniform	0.6267	0.00614	0.6315	0.0098	0.6365	0.00154
Jeffrey's	0.3866	0.00604	0.6218	0.0098	0.6267	0.00156
Exponential	0.6095	0.0058	0.6143	0.0096	0.6192	0.0156
Gamma	0.6121	0.0057	0.6168	0.0095	0.6398	0.0153

Whereas, the overall results show that the posterior distribution under Uniform and Gamma prior with DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors.

CONCLUSION

We have conducted this study to check the behavior of the scale parameter of Kumaraswamy Exponentiated Inverse Rayleigh Distribution. We have conducted a simulation study to complete Kumaraswamy Exponentiated Inverse Rayleigh Distribution to find out the Bayes Estimator and Posterior Risk of different priors which are Uniform, Jeffrey's, Gamma and Exponential priors. We have considered three different loss functions for the simple model, namely the Square error loss function (SELF), DeGroot loss function (DLF), and Precautionary loss function (PLF). The performance of the estimators has been compared based on their simulated risks, obtained under different loss functions. DeGroot loss function under Uniform and Gamma prior has been found better than the others because under this loss function posterior risks are minimum. Further the results indicate that the increase in sample size reduces the posterior risks of the estimates. In this study we have used only non-informative and informative priors other can extend this study by using priors. We have used Uniform, Jeffrey's, Gamma and Exponential priors other can use other prior distributions. We have used only three loss functions others can extend by using other loss functions.

REFERENCES

1. Diab, L. and Muhammed, H.Z. (2015). Statistical properties of kumaraswamy exponentiated gamma distribution. *International Journal of Reliability and Applications*, 16(2), 81-98.
2. Ahsan-ul-Haq, M. (2016a). Kumaraswamy exponentiated inverse rayleigh distribution. *Math. Theo. Model*, 6(3), 93-104.
3. Ahsan-ul-Haq, M. (2016b). Transmuted exponentiated inverse rayleigh distribution. *J. Stat. Appl. Prob*., 5(2), 337-343.
4. George, R. and Thobias, S. (2017). Marshall-olkin kumaraswamy distribution. In *International Mathematical Forum*, 12, 47-69.
5. Muhammad, M., Muhammad, I. and Yaya, A.M. (2018). The kumaraswamy exponentiated u-quadratic distribution: Properties and application. *Asian Journal of Probability and Statistics*, 1(3), 1-17.
6. Sule, B.O., Adegoke, T.M. and Uthman, K.T. (2021). Bayes estimators of exponentiated inverse rayleigh distribution using lindleys approximation. *Asian Research Journal of Mathematics*, 17(2), 60-71.
7. Jayalakshmi, S. and PK, N.K. (2022). Designing of Double Sampling Plan for Truncated Life Tests Based on Percentiles using Kumaraswamy Exponentiated Rayleigh Distribution. *International Journal of Mathematics Trends and Technology*, 68(1), 29-35.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 143-150

BAYESIAN ANALYSIS OF EXPONENTIATED INVERSE RAYLEIGH DISTRIBUTION BY USING NON-INFORMATIVE PRIORS

Maryam Hina[1], Wajiha Nasir[2], Sadia Munawar and **Zakia Kousar**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]maryamhina112@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

The Exponentiated inverse Rayleigh distribution is a generalized variant of the inverse Rayleigh distribution. Using Bayesian analysis, we investigated the exponentiated inverse Rayleigh distribution. The posterior distribution has been obtained using non-informative priors (Jeffrey's and uniform), and the Bayes estimates (BEs) and posterior risks (PRs) have also been determined. We also investigated the performance of the posterior distribution using the difference loss function, the square error loss function (SELF), the Degroot error loss function (DELF), and the Precautionary loss function (PLF). Using varying sample sizes, a simulation study was used to evaluate the performance of the Bayes estimator and its associated risks.

KEYWORDS

Bayesian analysis, Loss functions, Posterior Distribution under Non-Informative Prior.

INTRODUCTION

Inverse Rayleigh distribution was proposed by Trayer (1964). Many researchers consider the Inverse Rayleigh distribution to be a better and more versatile distribution than the earlier Rayleigh distribution named after Lord Rayleigh, and significant research has been done in this area to further improve Trayer's (1964) work, and many distributions have been introduced. Over the years, researchers have gone above and beyond to achieve superior and more effective methods of estimating the parameter of this distribution, and in statistical estimation, there are clearly two groups of people with opposing views on estimations; some prefer the classical approach, while others prefer the Bayesian approach.

Kumar and Garg (2014), investigation of the generalized inverted Rayleigh distribution with randomly censored data is done in this paper. The parameters have been estimated utilizing maximum likelihood and Bayesian methods. Based on the observed Fisher's information matrix, exponential confidence intervals for the parameters are calculated. The parameters' Bayes estimates are produced using Lindley's approximation approach and the importance sampling procedure. The significance sampling approach is used to generate maximum posterior density reliable ranges for the parameters. Sample size n rises, the average length of 95% asymptotic confidence and HPD credible ranges shortens. Furthermore, the average length of HPD credible intervals is always less than that of exponential confidence intervals.

Ali et al., (2022), this article is about the formation of the Alpha Power Generalized Inverse Rayleigh (APGIR) Distribution. The proposed model gives a more efficient fit of lifetime experience data. Moment-generating function, quantile, mode, order statistics, stress-strength parameter, and entropies are derived as significant aspects of the proposed model. The MLE approach is used to estimate parameters. The proposed model's performance is assessed using real-world data sets. The results of the simulation and real-world data sets indicate that the newly suggested model outperforms other present rival models. The suggested model's performance was assessed using two real data sets and several goodness-of-fit criteria. The findings clearly show that our suggested model outperforms other types of Rayleigh distributions available in the literature.

Ahmad et al., (2014), to create a transmuted inverse Rayleigh distribution, generalize the Inverse Rayleigh distribution using the quadratic rank transmutation map described. The features of this distribution are calculated, and the model parameters are estimated using the maximum likelihood technique. Statistical analysis processes' accuracy strongly depends on the assumed probability model or distribution. In this article, the transmutation map technique to create a novel model that generalizes the Inverse Rayleigh model. The generalized distribution will be referred to as the transmuted inverse Rayleigh distribution. Some mathematical features are examined, as well as estimating concerns. The reliability behavior and hazard rate function of the transmuted inverse Rayleigh distribution demonstrate that the subject distribution may be utilized to simulate dependability data.

Ogunsanya et al., (2021), a three-parameter Weibull Inverse Rayleigh (WIR) distribution is proposed in this paper. The new WIR distribution is a one-parameter Inverse Rayleigh distribution with a modification of the Weibull distribution and Log-logistic as the quantile function. The statistical features of the newly suggested distribution, such as the quantile function, order statistic, monotone likelihood ratio property, hazard, reverse hazard functions, moments, skewness, kurtosis, and linear representation, were theoretically investigated. Maximum likelihood estimators cannot be expressed explicitly. As a result, the standard errors of the parameters were found to be comparatively fewer than those offered by other competing existing distributions when compared, and these standard errors decreased as sample sizes increased for all parameter values indicated in the simulation research. Basically, the findings of this work demonstrated that Bayesian estimates of the novel WIR distribution are superior to those obtained by the standard frequentist approach of MLE

Banejee and Bhunia (2022), this article addresses an extension of the inverse Rayleigh distribution utilizing the DUS transformation, known as the Exponential Transformed Inverse Rayleigh (ETIR) distribution. Some of the statistical aspects of this newly suggested distribution, such as mode, quantiles, moment, moment generating function, survival, and hazard rate function, have been thoroughly investigated. To estimate the parameter of this distribution, four possible estimation approaches are briefly discussed: maximum likelihood estimation (MLE), maximum product spacing method (MPS), least square method (LSE), and weighted least square method (WLSE). The performance of these estimations is compared using large simulations. From an application standpoint, the model's superiority is demonstrated using two real datasets. Four estimating approaches are used to estimate the unknown model parameter. We did a thorough simulation analysis to compare different methodologies. We compared estimators based on mean square error (MSE).

Taniş and Saraçoğlu (2022), this article suggest a novel lifespan distribution termed Cubic Rank Transmuted Inverse Rayleigh as an alternative to the inverse Rayleigh distribution. Some distributional aspects of the proposed distribution are investigated, including moments, incomplete moments, Bonferroni and Lorenz curves, and moment generating function, quantile function, median, and mean residual life function. To estimate the parameters of the suggested distribution, we evaluate five methods: maximum likelihood, least squares, weighted least squares, Anderson Darling method, and the Cramer-von-Mises method. A complete Monte Carlo simulation analysis is also carried out to compare the performances of the evaluated estimators in terms of mean square errors and biases.

EXPONENTIATED INVERSE RAYLEIGH DISTRIBUTION (EIRGD)

The Exponentiated Inverse Rayleigh distribution has been developed by Rao & Mbwambo (2019). The P.d.f of Exponentiated Rayleigh distribution is as follows:

$$f(x)=\frac{2\alpha\sigma^2}{x^3}e^{-\left(\frac{\sigma}{x}\right)^2}\left(1-e^{-\left(\frac{\sigma}{x}\right)^2}\right)^{\alpha-1}\quad x\geq 0,\sigma,\alpha>0$$

Respectively, where the parameter σ is the scale parameter, and α is the shape parameter of the Exponentiated Rayleigh distribution respectively according to Rao & Mbwambo (2019).

MODEL AND LIKELIHOOD FUNCTION

The P.d.f of distribution is as follows:

$$f(x)=\frac{2\alpha\sigma^2}{x^3}e^{-\left(\frac{\sigma}{x}\right)^2}\left(1-e^{-\left(\frac{\sigma}{x}\right)^2}\right)^{\alpha-1}\quad x\geq 0,\sigma,\alpha>0$$

The log-likelihood function is as follows:

$$L(x,\alpha,\sigma)=\frac{2^n\alpha^n\sigma^{2n}}{\prod_{i=1}^{n}x_i^3}e^{-\sum\left(\frac{\sigma}{x}\right)^2}.e^{(\alpha+1)\ln\sum(z)}.$$

PRIOR DISTRIBUTION

The prior distribution is a method that includes non-informative prior.

NON-INFORMATIVE PRIORS

Sometimes, the deviation of the prior distribution based on information other than the current data is impossible. In that case, we use non-informative priors. They are used when we don't have any information about the parameters. A non-informative prior is one in which little new explanatory power about the unknown parameter is provided by intention. It is often improper. Some non-informative priors are given below:

USING UNIFORM PRIOR

The Uniform prior of parameter α is defined as:

$$p(\alpha \mid x) = 1 \quad 0 < \alpha < \infty,$$

The posterior distribution for the parameter is:

$$p(\alpha \mid x) = \frac{\left(\ln(t)\right)^{n+1} \alpha^{(n+1)-1} e^{-\alpha(\ln(t))}}{\Gamma(n+1)} \qquad \alpha \sim G(\alpha_1, \beta_1)$$

where $\alpha_1 = n+1$, $\beta = \ln\left(\sum_{i=1}^{n}\left(1-e^{-\left(\frac{\theta}{x_i}\right)}\right)\right)$ and $t = \sum_{i=1}^{n}\left(1-e^{-\left(\frac{\theta}{x_i}\right)}\right)$ which is the density function of Gamma Distribution with parameters α and β.

USING JEFF $Var(\alpha \mid x)$ REY'S PRIOR

The Jeffrey's prior of parameter θ is defined as:

$$p(\alpha \mid x) = \frac{1}{\alpha}, 0 < \alpha < \infty$$

The posterior distribution for parameter θ is

$$p(\alpha \mid x) = \frac{\left((\ln(t))\right)^{n} \alpha^{(n-1)} e^{-\alpha(\ln(t))}}{\Gamma(n)} \quad \alpha \sim G(\alpha_1, \beta_1)$$

where $\alpha_1 = n$, $\beta_1 = \ln \sum_{i=1}^{n}\left[1+e^{-\left(\frac{e}{x_i}\right)^2}\right]$ and $t = \sum_{i=1}^{n}\left(1-e^{-\left(\frac{\theta}{x_i}\right)}\right)$ which is the density function of Gamma Distribution with parameters α_1 and β_1.

BES AND PRS UNDER DIFFERENT LOSS FUNCTIONS

This section includes Bayes estimators and posterior risks for different loss functions, as well as their conclusions for non-informative priors. The following loss functions are defined:

Loss Function (LF)	Bayes Estimator (BE)	Posterior Risk (PR)
Square error (SELF)	$E(\alpha \mid x)$	
Precautionary (PLF)	$\sqrt{E(\alpha^2 \mid x)}$	$2*\left(\sqrt{E(\alpha^2 \mid x)} - E(\alpha \mid x)\right)$
DeGroot (DELF)	$\dfrac{E(\alpha^2 \mid x)}{E(\alpha \mid x)}$	$\dfrac{Var(\alpha \mid x)}{E(\alpha^2 \mid x)}$

EXPRESSION FOR BES AND PRS

This section represents BE and PR equations for different loss functions utilizing non-informative priors. They are obtained using the expressions given in the section below:

Expression For Bes and PRs Under SELF		
Prior	**BEs**	**PRs**
Uniform	$\dfrac{n+1}{\left[\sum\left(-\log(t)\right)\right]}$	$\dfrac{n+1}{\left[\sum\left(-\log(t)\right)\right]^2}$
Jeffrey's	$\dfrac{n}{\left[\sum\left(-\log(t)\right)\right]}$	$\dfrac{n}{\left[\sum\left(-\log(t)\right)\right]^2}$
Expression For Bes and PRs Under PLF		
Prior	**BEs**	**PRs**
Uniform	$\dfrac{\sqrt{(n+1)(n+2)}}{\left[\sum\left(-\log(t)\right)\right]}$	$2*\left(\dfrac{\sqrt{(n+1)(n+2)}-(n+1)}{\left[\sum\left(-\log(t)\right)\right]}\right)$
Jeffrey's	$\dfrac{\sqrt{n(n+1)}}{\left[\sum\left(-\log(t)\right)\right]}$	$2*\left(\dfrac{\sqrt{n(n+1)}-(n)}{\left[\sum\left(-\log(t)\right)\right]}\right)$
Expression For Bes and PRs Under DELF		
Priors	**BEs**	**PRs**
Uniform	$\dfrac{n+2}{\sum\left(-\log(t)\right)}$	$\dfrac{1}{n+2}$
Jeffrey's	$\dfrac{n}{\sum\left(-\log(t)\right)}$	$\dfrac{1}{n+1}$

SIMULATION STUDY

Simulation is the process of taking a model, often stated as a computer program, and executing it multiple times to generate results, then gathering the data and computing the estimated probability. Simulation is used to explore the performance of the Bayes estimator and posterior risk under prior distributions with different loss functions. We have simulated the random sample of sizes n=50, 70, 200, and 500 of Exponentiated Inverse Rayleigh Distribution with parameter α =1 and 2 and parameters σ =2, 3 for 10,000 times. The results of the simulation study are summarized in the following tables:

BEs and PRs for SELF					
n	**Priors**	**α =1**		**α =2**	
		BEs	**PRs**	**BEs**	**PRs**
50	**Uniform**	1.4884	0.0438	2.3013	0.1725
70		1.4756	0.0308	2.2451	0.0713
200		1.4585	0.0106	2.2197	0.0245
500		1.4526	0.0042	2.2105	0.0097
50	**Jeffrey's**	3.2225	0.3466	5.3563	0.9586
70		3.1047	0.1929	5.2236	0.5463
200		3.1405	0.0924	5.0830	0.2584
500		3.0305	0.0459	5.0445	0.1271
BEs and PRs for PLF					
n	**Priors**	**α =1**		**α =2**	
		BEs	**PRs**	**BEs**	**PRs**
50	**Uniform**	1.4976	0.0289	2.3403	0.0737
70		1.4901	0.0207	2.2619	0.0315
200		1.4609	0.0072	2.2241	0.0110
500		1.4534	0.0028	2.2119	0.0044
50	**Jeffrey's**	3.2540	0.1024	5.4253	0.1708
70		3.1495	0.0608	5.2600	0.1016
200		3.0680	0.0301	5.1303	0.0504
500		3.0377	0.0150	5.0659	0.0251
BEs and PRs for DELF					
n	**Priors**	**α=1**		**α =2**	
		BEs	**PRs**	**BEs**	**PRs**
50	**Uniform**	1.5125	0.0192	2.3761	0.0312
70		1.4994	0.0138	2.2794	0.0138
200		1.4651	0.0049	2.2310	0.0049
500		1.4540	0.0019	2.2153	0.0019
50	**Jeffrey's**	3.2912	0.0312	5.4979	0.0312
70		3.1832	0.0192	5.2911	0.0192
200		3.0908	0.0098	5.1534	0.0098
500		3.0410	0.0049	5.0648	0.0049

This simulation approach was used to examine the behavior of the shape parameter of the Exponentiated Inverse Rayleigh Distribution. The data tables indicate the results of the posterior distribution under the Uniform and Jeffery priors using multiple loss functions, including the Square error loss function, the Precautionary loss function, and the Degroot loss function. The simulation results and comparison of these loss functions reveal that as sample size increases Bayes estimator approaches its actual value of the parameter and the posterior risk decreases. We can conclude By Increasing the sample size PRs decreases. DELF gives the best estimators under all the priors. Uniform Prior with DELF is performing better among others.

25 / 27 APPLICATION

In this study, we analyzed our proposed distribution on a set of real-life data of coating weights (gm/m^2) from the ALAF industry, in Tanzania. The data set of 72 observations on coating weight by a chemical method on the top center side (TCS) is taken from the local Website Hindwai which is as follows:

{36.8, 47.2, 35.6, 36.7, 55.8, 58.7, 42.3, 37.8, 55.4, 45.2, 31.8, 48.3, 45.3, 48.5, 52.8, 45.4, 49.8, 48.2, 54.5, 50.1, 48.4, 44.2, 41.2, 47.2, 39.1, 40.7, 40.3, 41.2, 30.4, 42.8, 38.9, 34.0, 33.2, 56.8, 52.6, 40.5, 40.6, 45.8, 58.9, 28.7, 37.3, 36.8, 40.2, 58.2, 59.2, 42.8, 46.3, 61.2, 58.4, 38.5, 34.2, 41.3, 42.6, 43.1, 42.3, 54.2, 44.9, 42.8, 47.1, 38.9, 42.8, 29.4, 32.7, 40.1, 33.2, 31.6, 36.2, 33.6, 32.9, 34.5, 33.7, 39.9}

Priors	SELF	PLF	DELF
	BEs PRs	BEs PRs	BEs PRs
Uniform	1.4852 0.0320	1.4953 0.0202	1.5055 0.0135
Jeffrey's	1.4648 0.0298	1.4750 0.0202	1.4852 0.0134

From the above-provided table, it is concluded that DELF is performing better as compare to others. In comparison to other loss functions and priors, the overall result showed that the posterior distribution under Uniform and Jeffery prior to DELF performs better as it provides the smallest result of posterior risk. Real-world examples and simulated studies provide almost the same results.

CONCLUSION

This study was conducted out to investigate the behavior of the shape parameter of the Exponentiated Inverse Rayleigh Distribution. We did a simulation analysis for the whole Exponentiated Generalized Inverse Rayleigh Distribution to figure out the Bayes Estimator and Posterior Risk of two alternative priors, Uniform and Jeffery. For the basic model, we analyzed three possible loss functions: square error loss function, precautionary loss function, and DeGroot loss function. The estimators' performance was compared using their simulated risks calculated under various loss functions. Degroot loss function with Jeffrey prior has been determined to be better to the others because it has the lowest posterior risks. Furthermore, the results show that increasing the sample size decreases the posterior risks of the estimates. The graphical representation for the basic model shows that the posterior distribution is symmetric under non-informative prior.

REFERENCES

1. Banerjee, P. and Bhunia, S. (2022). Exponential Transformed Inverse Rayleigh Distribution: Statistical Properties and Different Methods of Estimation. 51, 60-75.
2. Ahmad, A., Ahmad, A.S. and Ahmed, A. (2014). Transmuted Inverse Rayleigh Distribution: A Generalization of the Inverse Rayleigh Distribution. 4, 90-98.
3. Almarashi, M.A., Badr, M.M., Elgrahy, M., Jamal, F. and Chesneau, C. (2020). Statistical Inference of the Half-Logistic Inverse Rayleigh Distribution. *Entropy*, 22(4), 449. https://doi.org/10.3390/e22040449
4. Ali, M., Khalil, A., Almaspoor, Z., Hussain, S., Khalil, U. and El-Morshedy, M. (2022). Alpha Power Generalized Inverse Rayleigh Distribution: Its Properties and Applications. Article ID 7847110, https://doi.org/10.1155/2022/7847110
5. Ogunsanya, A.S., Yahya, W.B., Adegoke, T.M., Iluno, C., Aderele, O.R. and Ekum, M.I. (2021). A New Three-Parameter Weibull Inverse Rayleigh Distribution: Theoretical Development and Applications. 9, 249-272.
6. Kumar, K. and Garg, R. (2014). Estimation of the parameters of randomly censored generalized inverted Rayleigh distribution. *International Journal of Agricultural and Statistical Sciences*, 10(1), 147-155.
7. Taniş, C. and Saraçoğlu, B. (2022). Cubic rank transmuted Inverse Rayleigh distribution: Properties and Applications. *Sigma Journal of Engineering and Natural Sciences*, 40(2), 421-432.
8. Rao, S.G. and Mbwanbo, S. (2022). Exponentiated Inverse Rayleigh Distribution and an Application to Coating Weights of Iron Sheets Data. *Journal of Probability and Statistics,* Article ID 7519429, https://doi.org/10.1155/2019/7519429

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 151-158

STATISTICAL ANALYSIS ON PROGONOSTIC FACTORS OF CARDIOVASCULAR DISEASE: A COMPARATIVE STUDY BETWEEN LAHORE AND SIALKOT

Minahil Shakeel[§], Alina Ramzan, Mashal Talib
and **Anam Asmatullah**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [§]alina.ramzan3@gmail.com

ABSTRACT

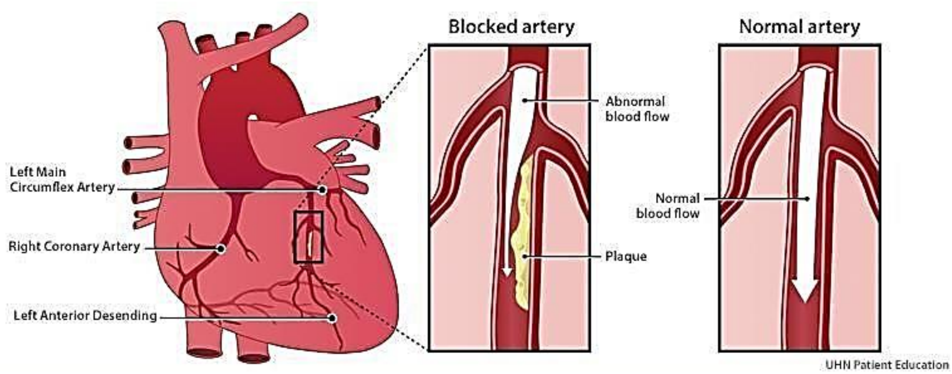
Cardiovascular disease (CVD) is a major public health concern, responsible for significant morbidity and mortality worldwide. The pathogenesis of CVD is complex and multifactorial, involving the interplay of genetic, environmental, and lifestyle factors. This analysis aims to provide a comprehensive overview of CVD, including its epidemiology, risk factors, and current management strategies in two major cities of Pakistan (Lahore and Sialkot). In addition, the current management strategies for CVD, including lifestyle modification, medicinal and pharmacological therapy are discussed. The sample size of 342 patients is collected by conducting a survey from Punjab Institute of Cardiology (Lahore) and Allama Iqbal Memorial Hospital (Sialkot). Descriptive statistics is used for screening the variables. The chi-square test is used for association between predictors (Cholesterol, Pain or discomfort, high blood pressure, diabetes, Age, Lack of exercise, obesity, smoking, family history smoking status, high blood pressure, and cholesterol level) and CVD patients to identify which patients are at higher risk for CVD and to provide prevention and treatment strategies.

KEYWORDS

CVD, chi square, phi, crammers v, hypertension, diabetes.

INTRODUCTION

In early 20th century, almost 10% deaths were caused by cardiovascular diseases, but this number was increased to 30% by the end of 2001. Maximum deaths were recorded in low and middle income countries and will be leading cause of death by 2020 in developing countries (Mathers et al., 2001, WHO, 2002a). 16.7 million deaths in 2000 and 17 million in 2008 were recorded due to CVD around the globe (Rosamond et al., 2008, WHO, 2011). Out of which 25% deaths in developed and 80% in middle and low income countries occur (Yarmohammadian et al., 2012). 85% global disabilities caused due to CVD (Alwan et al., 2011). Females are at a larger risk of CVD than males in developing countries (Mosca et al., 2011).



Cardiovascular diseases referred to situation in which heart is not properly pumping or receiving oxygen rich blood. This generally happens when solid mass made up of fats become deposited in blood vessels and blood clots formed. Sometimes damage to arteries of different organs (brain, heart, kidneys and eyes) is also related to CVD. The situation like this put strain on heart and different problems will arise like angina (severe chest pain produced due to lesser flow of blood to heart), heart attack (blood to heart muscle is totally blocked) or heart failure (heart not able to pump properly blood to different parts of body). Major risk factors include smoking, blood pressure, physio social stress, diabetes, lipid abnormalities (dyslipidemia), obesity, alcohol consumption and physical inactivity (Finucane et al., 2011; Mosca et al., 2011; Ueshima et al., 2008). Obesity increases risk factors of cardiovascular problems, 15% in men and 22% in women, observed from the data obtained from Framingham Heart Study. High cholesterol and systolic blood pressure give rise to CVD in Asian women (Barzi et al., 2007). Symptoms of CVD include sweating, dizziness, irregular heartbeat, shortness of breath, nausea and weakness. Various treatment methods are applied including physical examination, blood tests, echocardiography, computed tomography scan and angiography. Angioplasty and bypass can also be performed in severe cases depending upon patient conditions (WHO, 2008).

Coronary artery disease (CAD), congenital heart defects (CHD), cardiomyopathy (CMP) and myocardial infarction (MI) are different cases of cardiovascular problems. CMP is most common form of CVD in Pakistan (Ahmad et al., 2005; Khan et al., 2009). MI occurs 19/1000 and 192/1000 in males and females, respectively (Samad et al., 1996). Almost 250/100000 people get stroke and this number is increasing rapidly (Khatri and Wasay, 2011). In this paper data from 171 patients in Lahore (Punjab Institute of Cardiology) and 171 patients in Sialkot (Allama Iqbal Memorial Hospital) is gathered and compared. Firstly we use Descriptive statistics for screening the variables. Secondly the chi-square test is used for association between predictors (cholesterol, pain or discomfort, high blood pressure, diabetes, age, lack of exercise, obesity, smoking, family history smoking status, high blood pressure, and cholesterol level) and CVD (angina) patients to identify which patients are at higher risk for CVD and to provide prevention and

MATERIAL AND METHODS

Data Source

Treatment strategies. Risk factors are same in both cities having higher number of patients in Lahore as it is a big and more populated city compared to Sialkot.

Data Collection

A structured questionnaire is prepared, and the verbal consent was taking every CVD patient. The response variable in the presence of CVD (angina) (Yes, No). All the elicited information including the basic demographic details like social-economic status were collected. The sample consisted of 342 CVD patients from different hospitals. Being a CVD patient was taken an inclusion criteria. The independent variables were first selected using descriptive statistics and then variables associated with chi-square and phi crammer's v test. Odds ratio and 95% confidence interval for different risk factors were computed. The data was analyzed using Statistical Package for Social Sciences (SPSS) version 20. Cardiovascular disease (angina) cases were obtained from two major hospital namely (Punjab Institute of Cardiology, Allama Iqbal Memorial Hospital).

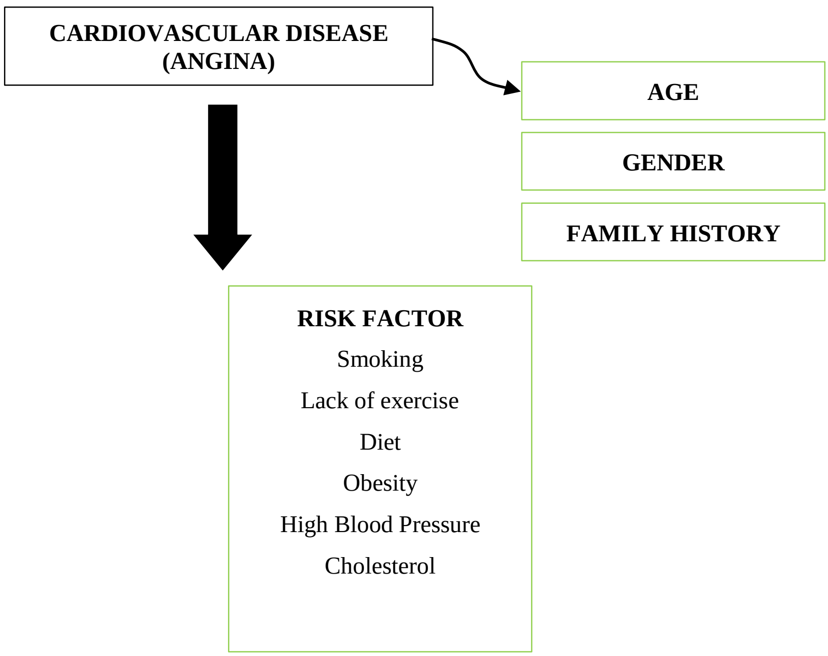


Figure 1:

Table 1
Sample Characteristics of Cardiovascular Disease

Variables		Sialkot	Sialkot	Lahore	Lahore
Age	30 to 50	26	15.2%	21	12.3%
	50 to 60	53	31.0%	77	45.0%
	60 to onwards	92	53.8%	73	42.7%
Gender	Male	133	77.8%	135	78.9%
	Female	38	22.2%	36	21.1%
Family History	Yes	126	73.7%	133	77.8%
	No	45	26.3%	38	22.2%
Chest Pain	Yes	23	13.5%	34	19.9%
	No	21	12.3%	25	14.6%
	Often	80	46.8%	48	28.1%
	Sometimes	47	27.5%	64	37.4%
Hypertension	Yes	139	81.3%	106	62.0%
	No	32	18.7%	65	38.0%
Diabetes	Yes	61	35.7%	61	35.7%
	No	110	64.3%	110	64.3%
Blockage of Vessels	Yes	146	85.4%	121	70.8%
	No	25	14.6%	50	29.2%
Obesity	Yes	123	71.9%	82	48.0%
	No	48	28.1%	87	50.9%
Lack of Exercise	Yes	107	62.6%	78	45.6%
	No	64	37.4%	93	54.4%
Angina	Yes	162	94.7%	163	95.3%
	No	9	5.3%	8	4.7%
Irregular Heartbeat	Yes	100	58.5%	108	63.2%
	No	71	41.5%	63	36.8%
Angiography	Yes	99	57.9%	129	75.4%
	No	72	42.1%	42	24.6%
Cholesterol	Yes	15	8.8%	36	21.1%
	No	16	9.4%	23	13.5%
	Often	63	36.8%	48	28.1%
	Sometimes	77	45.0%	64	37.4%
Balanced Diet	Yes	60	35.1%	92	53.8%
	No	111	64.9%	79	46.2%
Pain in Shoulder, Arms	Yes	114	66.7%	109	63.7%
	No	57	33.3%	62	36.3%
Shortness Of breath	Yes	110	64.3%	118	69.0%
	No	61	35.7%	53	31.0
Red Meat	Yes	20	11.7%	9	5.3%
	No	21	12.3%	16	9.4%
	Often	75	43.9%	65	38.0%
	Sometimes	55	32.2%	81	47.4%

Variables		Sialkot	Sialkot	Lahore	Lahore
Smoking	Yes No	36 135	21.1% 78.9%	53 118	31.0% 69.0%
Fatigue	Yes No Often Sometimes	23 21 74 53	13.5% 12.3% 43.3% 31.0%	34 25 46 66	19.9% 14.6% 26.9% 38.6%
Indigestion, Nausea	Yes No Often Sometimes	20 21 72 58	11.7% 12.3% 42.1% 33.9%	9 16 63 83	5.3% 9.4% 36.8% 48.5%

Table 2
Chi-Square and Phi, Crammer's V

Variables		Sialkot	Sialkot	Lahore	Lahore
		P-Value	Phi Crammer's V	P-Value	Phi Crammer's V
Age	30 to 50 50 to 60 60 to onwards	0.042	0.192	0.014	0.223
Gender	Male Female	1.000	0.000	0.779	0.021
Family History	Yes No	0.005	0.216	0.287	0.081
Chest Pain	Yes No Often Sometimes	0.017	0.244	0.013	0.250
Hypertension	Yes No	0.042	0.156	0.000	0.283
Diabetes	Yes No	0.46	-0.153 0.153	0.105	0.124
Blockage of Vessels	Yes No	0.009	0.199	0.000	0.284
Obesity	Yes No	0.001	0.261	0.906	0.236
Lack of Exercise	Yes No	0.333	-0.074 0.074	0.326	-0.075 0.075
Acquired Age	10-30 30-50 50-70 More than 70	0.467	0.122	0.016	0.246
Irregular Heartbeat	Yes No	0.609	-0.039 0.039	0.002	0.233
Angiography	Yes No	0.000	0.122	0.000	0.324

Variables		Sialkot	Sialkot	Lahore	Lahore
		P-Value	Phi Crammer's V	P-Value	Phi Crammer's V
Cholesterol	Yes No Often Sometimes	0.001	0.305	0.000	0.355
Balanced Diet	Yes No	0.041	-0.156 0.156	0.016	0.184
Pain in Shoulder, Arms	Yes No	0.029	0.167	0.000	0.294
Shortness of Breath	Yes No	0.046	0.153	0.000	0.271
Red Meat	Yes No Often Sometimes	0.015	0.247	0.000	0.332
Smoking	Yes No	0.353	-0.071 0.071	0.707	0.029
Fatigue	Yes No Often Sometimes	0.015	0.248	0.013	0.250
Indigestion, Nausea	Yes No Often Sometimes	0.014	0.249	0.000	0.331
ECG	Yes No	0.000	0.462	0.00	0.346
Marital Status	Married Unmarried	0.005	0.214	0.172	0.104

RESULTS

The results from the Table 1 shows that out of 342 CVD patients who suffer from cardiovascular disease (angina) 268 were male and 74 were female both the city having large number of probability of having this disease are male as compared with females, Aged (60 to onwards) 92 from Sialkot and 73 from Lahore people were more likely to suffer, those who have their family history with this disease are 126 from Sialkot and 133 from Lahore other risk factors that includes are chest pain 127 from Sialkot and 112 from Lahore, high blood pressure 139 from Sialkot and 106 from Lahore. Diabetes 61 from Sialkot and 61 from Lahore, blockage of vessels 146 and 121, obesity 123 and 82, lack of exercise 107 and 78, angina 162 and 163, irregular heartbeat 100 and 108, angiography 99 and 129, cholesterol 155 and 148, balanced diet 60 and 92, pain in shoulder and arms 114 and 109, shortness in breath 110 and 118, red meat 150 and 155, smoking 36 and 53, fatigue 150 and 146, indigestion and nausea 150 and 155 these are those patients who have CVD.

The results from Table 2 shows the association between predictors and CVD (angina) with p-value less than 0.05. The significant variables from Lahore are Age (0.014), ECG (0.000), Chest pain (0.013), High blood pressure (0.000), Cholesterol (0.000), Red meat

(0.000), Irregular heartbeat (0.002), Pain in shoulder and arms (0.000), Blockage of vessels (0.000), Acquired age (0.016), Angiography (0.000), Balanced diet (0.016), Shortness of breath (0.000), Fatigue (0.013), Nausea and Indigestion (0.000). And the significant variables from Sialkot are Marital status (0.005), Age (0.042), Family history (0.005), ECG (0.00), Chest pain (0.017), High blood pressure (0.042), Obesity (0.001), Cholesterol (0.001), Red meat (0.015), Pain in shoulder and arms (0.029), Blockage of vessels (0.009), Angiography (0.000), Balanced diet (0.041), Shortness of breath (0.046), Fatigue (0.015), Nausea and Indigestion (0.014). Similarly, the odds ratio of significant variables of Lahore are Gender (1.265), Marital status (3.063), Family history (2.194), ECG (1.143), Smoking (1.366), Irregular heartbeat (13.375), Blockage of vessels (19.5), Angiography (25.6), Balanced diet (8.847), Shortness of breath (17.804). As well as from Sialkot the odds ratio of significant variables is Gender (1.000), Marital status (6.021), Family history (6.308), ECG (1.286), High blood pressure (3.829), Obesity (10.329), Pain in shoulder and arms (4.353), Blockage of vessels (5.371), Shortness of breath (3.891).

CONCLUSION

Numerous researches have been conducted to identify the risk factors for CVD, and several significant variables have been identified. Therefore, a comprehensive assessment of an individual's risk of CVD should take account all relevant risk factors, along with other factors. Age, Gender, Smoking, High blood pressure, Cholesterol, Obesity, Diabetes, Physical Inactivity and Stress are all significant risk factors for CVD and Angina according to all performed statistical analysis. Treatment depends on patient's condition and vulnerability. Lahore is at greater risk for CVD as it is more populated and environment unhealthy city. However it is personal adaptation for a person to live a healthy life by adopting a healthy lifestyle through balanced diet, personal hygiene, and exercise.

REFERENCES

1. Samad, A., Sehibzada, W.A., Nazir, F. and Khan, A.A. (1996). Incidence of acute myocardial infarction. *Pak. J. Cardiol.*, 7, 13.
2. Ahmad M, Afzal S, Malik IA, Mushtaq S, Mubarik A (2005) An autopsy study of sudden cardiac death. *J-Pakistan Med Assoc*., 55, 149.
3. Finucane, M.M., Stevens, G.A., Cowan, M.J., Danaei, G., Lin, J.K., Paciorek, C.J., Singh, G.M., Gutierrez, H.R., Lu, Y. and Bahalim, A.N. (2011). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *The Lancet*, 377(9765), 557-567.
4. Khan, H., Jan, H. and Hafizullah, M. (2009). A hospital-based study on causes peculiar to heart failure. *J Tehran University Heart Center*, 4, 25-28.
5. Ueshima, H., Sekikawa, A., Miura, K., Turin, T.C., Takashima, N., Kita, Y., Watanabe, M., Kadota, A., Okuda, N. and Kadowaki, T. (2008). Cardiovascular disease and risk factors in Asia. *Circulation*, 118, 2702-2709.
6. WHO (2008). *The Global Burden of Disease*. World Health Organization.
7. WHO (2011). Global Status Report on Non communicable Diseases 2010. In: *Global Status Report on Non communicable Diseases* 2010, Geneva

8. Zubair, F., Nawaz, S.K., Nawaz, A., Nangyal, H., Amjad, N. and Khan, M.S. (2018). Prevalence of cardiovascular diseases in Punjab, Pakistan: A cross-sectional study. Journal of Public Health, 26, 523-529.
9. Sajid, M.R., Muhammad, N., Zakaria, R., Shahbaz, A. and Nauman, A. (2020). Associated factors of cardiovascular diseases in Pakistan: Assessment of path analyses using warp partial least squares estimation. *Pakistan Journal of Statistics and Operation Research*, 16(2), 265-277.
10. Nabel, E.G. (2003). Cardiovascular disease. *New England Journal of Medicine*, 349(1), 60-72.
11. Gaziano, T., Reddy, K.S., Paccaud, F., Horton, S. and Chaturvedi, V. (2006) Cardiovascular Disease. In: Jamison, D., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D.B., Jha, P., Mills, A. and Musgrove, P. (Eds.), *Disease Control Priorities in Developing Countries*, 2nd Edition, World Bank, Washington DC.
12. Verhaar, M.C., Stroes, E. and Rabelink, T.J. (2002). Folates and cardiovascular disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 22(1), 6-13.
13. Anderson, K.M., Wilson, P.W., Odell, P.M. and Kannel, W.B. (1991). An updated coronary risk profile. A statement for health professionals. *Circulation*, 83(1), 356-362.
14. Anderson, K.M., Odell, P.M., Wilson, P.W. and Kannel, W.B. (1991). Cardiovascular disease risk profiles. *American Heart Journal*, 121(1), 293-298.
15. Wald, N.J. and Law, M.R. (2003). A strategy to reduce cardiovascular disease by more than 80%. *BMJ*, 326(7404), 1419.
16. Cosselman, K.E., Navas-Acien, A. and Kaufman, J.D. (2015). Environmental factors in cardiovascular disease. *Nature Reviews Cardiology*, 12(11), 627-642.
17. Berry, J.D., Dyer, A., Cai, X., Garside, D.B., Ning, H., Thomas, A. and Lloyd-Jones, D.M. (2012). Lifetime risks of cardiovascular disease. *New England Journal of Medicine*, 366(4), 321-329.
18. Knutsson, A. and Bøggild, H. (2000). Shiftwork and cardiovascular disease: review of disease mechanisms. *Reviews on Environmental Health*, 15(4), 359-372.
19. Mbewu, A. and Mbanya, J.C. (2006). Cardiovascular disease. In: Jamison, D.T., Feachem, R.G., Makgoba, M.W., Bos, E.R., Baingana, F.K., Hofman, K.J. and Rogo K.O. (Eds.) *Disease and Mortality in Sub-Saharan Africa*. 2nd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2006. Chapter 21.
20. Chen, W.W., Gao, R.L., Liu, L.S., Zhu, M.L., Wang, W., Wang, Y.J., Wu, Z.S., Li, H.J., Gu, D.F., Yang, Y.J. and Zheng, Z. (2017). China cardiovascular diseases report 2015: A summary. *Journal of Geriatric Cardiology: JGC*, 14(1), 1-10.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 159-166

BAYESIAN ANALYSIS OF TOP-LEONE FRECHET DISTRIBUTION BY USING NON- INFORMATIVE PRIORS

Zakia Kousar[1], Wajiha Nasir[2], Sadia Munawar and **Maryam Hina**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]zakia.kousar55@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

This study focuses on Top-Leone Frechet distribution. It is the generalization of Frechet distribution and is used for modeling rainfall data, temperature, wind speed, etc. We have studied the proposed distribution by using Bayesian analysis. Posterior distribution has been derived by using informative priors (uniform and Jeffrey). The Bayes estimates (BEs) and posterior risks (PRs) have been estimated. The performance of the proposed distribution has been studied by using the loss function square error loss function (SELF), the simple precautionary loss function (SPLF), the DeGroot error loss function (DELF), and the precautionary loss function (PLF). The performance of the Bayes estimator and their corresponding risks has been checked by a simulation study. A real data set has also been used to check the performance of Bayes estimates and posterior risks.

KEYWORDS

Frechet, Bayesian estimation, Non-informative priors, Posterior risk, Loss functions, Simulation.

INTRODUCTION

Extreme value theory plays a significant role in statistical analysis. In probability theory and inferences, various lifetime distributions have been built to apply to extreme events. This paper has proposed an extreme value distribution known as the Topp-Leone Frechet distribution. We have extended one parameter Frechet distribution to more than one parameter and named it as Topp-Leone Frechet distribution. Real data sets from fields such as hydrology, engineering, and social sciences do not appropriately fit the lifetime distributions. To overcome this issue, we have introduced Topp-Leone Frechet distribution.

Shanker and Rahman (2021) used type II Topp-Leone Fréchet distribution. They have studied the properties of the proposed distribution including hazard rate function, reverse hazard rate function, Mills ratio, quantile function, and order statistics. Maximum likelihood estimation has been used for estimating the parameters of the distribution. They have discussed the applications of the distribution for modeling datasets relating to temperature. They have checked the goodness of fit. They have a practical implementation of this distribution on a real dataset i.e., the monthly mean temperature series of Silchar City.

Reyad et al. (2021) have studied the Fréchet Topp-Leone-G family of distributions. They have extended the distribution by including three parameters. All the statistical properties of the distribution had been calculated. This distribution can be used for fitting skewed data. For skewed data, other common distributions are not appropriate. They reported that the proposed distribution can be used in modeling problems in various fields such as financial, industrial reliability, and survival analysis. They have compared the distribution with other distributions to check its performance. Two applications on real-life data have been proven the efficiency of distribution. Different plots of the functions had been plotted, i.e., plots of skewness, kurtosis, hazard rate function, probability density function, etc. Moment generating function had been computed. A simulation study has been conducted.

Yousof et al. (2021) discussed a new extension of Fréchet distribution with regression models, residual analysis, and characterization. They have applied different models to voltage and Stanford heart transplant data sets. They proposed a new model that can be useful in analyzing, and modeling real-life data and provides more accurate fits than other regression models. The flexibility of the new model had been illustrated. Nadarajah (2000) had shown the applications of the Fréchet distribution on real data sets of earthquakes, rainfall, floods, horse racing, wind speed, etc. Graphical representation of the various functions has been shown. A residual analysis of the model had been done to check the adequacy of the model. A real univariate data set had been studied and the Fréchet distribution had been used for modeling this data. Real data sets had been used for this study.

Bantan et al. (2019) had introduced a new general family of the distribution obtained by the combination of two well-established families of distributions. They discussed some theoretical properties of the new family in generality. They constructed a statistical parametric model and used inverse Lomax distribution as a baseline distribution. Asymptotic confidence intervals of the parameters had been constructed. Different models had been used and the best model was selected based on error values for two data sets. Estimated Probability density functions and cumulative density functions had been plotted.

(Ramos et al., 2020) had studied Fréchet distribution. They used two approaches for estimating the parameters of the distribution, i.e., the Bayesian approach and the Frequentist approach. Frequentist approaches such as maximum likelihood estimation, method of moments, percentile estimator, and L-moments had been briefly described. They used the Bayesian approach by referencing priors. Both approaches had been compared. Five real data sets had used to illustrate the efficiency of both approaches. The results showed that the Fréchet distribution had fitted the data properly from all the distributions. Based on AIC and BIC values, the model had chosen. They concluded that the Bayesian approach performs better for estimating unknown parameters than the frequentist approach.

(Vasileva et al., 2021) had described new properties of the odd Weibull inverse Topp-Leone cumulative distribution. A new adaptive model with polynomial variable transfer had used. COVID-19 data had been used for the simulation study. Hausdorff approximation of the Heaviside step function had studied.

(Yahia and Mohammed, 2019) had introduced a new three-parameter distribution called type-II Topp-Leone generalized inverse Rayleigh distribution. They concluded that the proposed distribution can be uni-model, symmetric, and right-skewed. Reliability analysis, statistical properties, and linear representation of the pdf had been discussed. This distribution is very useful in Biostatistics.

TOPP-LEONE FRECHET DISTRIBUTION

Topp-Leone Frechet distribution is an extension of the one-parameter Frechet distribution that we have created using the Topp-Leone generating family of distributions. Consider X to be a non-negative random variable with a TLF distribution if its CDF is:

$$F(x) = [1-\{1-\exp(-x^{-\lambda})\}^2]^{\alpha} \;\; ; \;\; x>0, \alpha>0, \lambda>0$$

And Pdf is of the form as follows:

$$f(x) = 2\alpha\lambda x^{-(1-\lambda)} \exp(-x^{-\lambda})\left\{1-\exp(-x^{-\lambda})\right\}\left[1-\left\{1-\exp(-x^{-\lambda})\right\}^2\right]^{\alpha-1}, \quad x>0, \alpha>0, \lambda>0$$

where α is the scale parameter and λ is the shape parameter of the distribution.

MODEL AND LIKELIHOOD FUNCTION

The Pdf of distribution is as follows:

$$f(x) = 2\alpha\lambda x^{-(1-\lambda)} \exp(-x^{-\lambda})\left\{1-\exp(-x^{-\lambda})\right\}\left[1-\left\{1-\exp(-x^{-\lambda})\right\}^2\right]^{\alpha-1}, \quad x>0, \alpha>0, \lambda>0$$

The log-likelihood function is as follows:

$$L(\underline{x};\alpha,\lambda) = 2^n \alpha^n \lambda^n \prod_{i=1}^{n} x_i^{-(1+\lambda)} . \prod_{i=1}^{n} exp(-x_i^{-\lambda})\left\{1-exp(-x_i^{-\lambda})\right\} e^{-\alpha \log \sum_{i=1}^{n}(-z_i)}$$

where $z = 1-\left\{1-e^{-x^{(-\lambda)}}\right\}^2$.

PRIOR DISTRIBUTION

The prior distribution of a parameter is the probability distribution that indicates your uncertainty about the parameter prior to examining the current data. The posterior distribution of the parameter is obtained by multiplying the prior distribution and the likelihood function together.

NON-INFORMATIVE PRIORS

It is occasionally hard to calculate the prior distribution's deviation using information other than the available data. In that circumstance, we employ non-informative priors. When we don't know the parameters, we use them. A non-informative prior is one that has little new explanatory power about the unknown parameter offered by intention. It is frequently incorrect. Below are some non-informative priors:

USING UNIFORM PRIOR

The Uniform prior of parameter α is defined as:

$$p(\alpha \mid y) = 1 \;;\; 0 < \alpha < \infty,$$

The posterior distribution for parameter α is:

$$p(\alpha \mid x) = \frac{\left[\log \sum_{i=1}^{n} (-z_i)\right]^{n+1}}{\Gamma(n+1)} \; .\alpha^{(n+1)-1} . e^{-\alpha \log \sum (-z_i)}$$

where $\alpha_1 = n+1, \beta_1 = \log \sum (-z_i)$ and $z = 1 - \left\{1 - e^{-x^{(-\lambda)}}\right\}^2$ which is the density function of Gamma Distribution with parameters α_1 and β_1.

USING JEFFREY'S PRIOR

The Jeffrey's prior of parameter α is defined as:

$$p(\alpha \mid y) = \frac{1}{\alpha}; \quad 0 < \alpha < \infty$$

The posterior distribution for parameter θ is

$$p(\alpha \mid x) = \frac{\left[b + \log \sum_{i=1}^{n} (-z_i)\right]^{n+a}}{\Gamma(n+a)} \; .\alpha^{(n+a)-1} . e^{-\alpha[b + \log \sum_{i=1}^{n} (-z_i)]} \quad \alpha \sim G(\alpha_1, \beta_1)$$

where $\alpha_2 = n + a$, $\beta_2 = b + \log \sum_{i=1}^{n} (-z_i)$ and $z_i = 1 - \left\{1 - \exp\left(-x^{-\lambda}\right)\right\}^2$ which is the density function of Gamma Distribution with parameters α_2 and β_2.

BES AND PRS UNDER DIFFERENT LOSS FUNCTIONS

This section contains Bayes estimators and posterior risks of the different loss function and also contain their results for non-informative priors. Some loss functions are defined below:

Loss Function (LF)	Bayes Estimator (BE)	Posterior Risk (PR)
Square error (SELF)	$\alpha^* = E(\alpha)$	$\mathrm{P}(\alpha^*) = E_{\alpha\mid X}\left\{L(\alpha - \alpha^*)\right\} = Var(\alpha)$
Precautionary (PLF)	$\alpha^* = \sqrt{E\left(\alpha^2 \mid X\right)}$	$P(\alpha^*) = 2 * \left[\sqrt{E\left(\alpha^2 \mid X\right)} - E(\alpha \mid X)\right]$
DeGroot (DELF)	$\alpha^* = \frac{E\left(\alpha^2 \mid X\right)}{E(\alpha \mid X)}$	$P\left(\alpha^*\right) = \frac{Var(\alpha / X)}{E\left(\alpha^2 / X\right)}$

EXPRESSION FOR BES AND PRS

This section contains the expressions for BEs and PRs under different loss functions using non-informative priors. They are derived using the expression in the section which are given below:

Expression For Bes and PRs Under SELF		
Prior	**BEs**	**PRs**
Uniform	$\frac{n+1}{\left[\sum\left(-\log(z)\right)\right]}$	$\frac{n+1}{\left[\sum\left(-\log(z)\right)\right]^2}$
Jeffrey's	$\frac{n}{\left[\sum\left(-\log(z)\right)\right]}$	$\frac{n}{\left[\sum\left(-\log(z)\right)\right]^2}$
Expression For Bes and PRs Under PLF		
Prior	**BEs**	**PRs**
Uniform	$\frac{\sqrt{(n+1)(n+2)}}{\left[\sum\left(-\log(z)\right)\right]}$	$2*\left(\frac{\sqrt{(n+1)(n+2)}-(n+1)}{\left[\sum\left(-\log(z)\right)\right]}\right)$
Jeffrey's	$\frac{\sqrt{n(n+1)}}{\left[\sum\left(-\log(z)\right)\right]}$	$2*\left(\frac{\sqrt{n(n+1)}-(n)}{\left[\sum\left(-\log(z)\right)\right]}\right)$
Expression For Bes and PRs Under DELF		
Priors	**BEs**	**PRs**
Uniform	$\frac{n+2}{\sum\left(-\log(z)\right)}$	$\frac{1}{n+2}$
Jeffrey's	$\frac{n}{\sum\left(-\log(z)\right)}$	$\frac{1}{n+1}$

SIMULATION STUDY

Simulation is a method of modelling random events so that simulated outcomes closely resemble real-world outcomes. Researchers acquire insight into the real world by observing simulated outcomes. A simulation study is carried out to obtain the BEs and PRs under different loss functions using different priors. The simulation process is repeated 10, 000 times considering generation of random samples of sizes 50, 70, 200 and 500 from inverse method assuming $\alpha = 1, 2$, $\beta = 2, 3$ and the results have then been averaged. These results are shown in the following tables.

BEs and PRs for SELF					
n	Priors	α =1		α =2	
		BEs	PRs	BEs	PRs
30	**Uniform**	1.0380	0.0215	2.1344	0.1522
50		1.0295	0.0151	2.05992	0.0606
100		2.0170	0.0203	2.02061	0.0204
200		1.0046	0.0020	2.00764	0.0080
30	**Jeffrey's**	1.0196	0.0212	2.0473	0.0856
50		1.0140	0.0149	2.0262	0.0595
100		2.0015	0.0201	2.0093	0.0202
200		1.0019	0.0020	2.0032	0.0080
BEs and PRs for PLF					
n	Priors	α =1		α =2	
		BEs	PRs	BEs	PRs
30	**Uniform**	1.0458	0.0202	2.1693	0.0683
50		1.0374	0.0144	2.0757	0.0289
100		1.0119	0.0050	2.0223	0.0100
200		1.0054	0.0020	2.0089	0.0040
30	**Jeffrey's**	1.0358	0.020	2.0418	0.0851
50		1.0207	0.0144	2.0453	0.0289
100		2.0166	0.0100	2.0161	0.0100
200		1.0028	0.0020	2.0064	0.0040
BEs and PRs for DELF					
n	Priors	α =1		α =2	
		BEs	PRs	BEs	PRs
30	**Uniform**	1.0637	0.0192	1.0064	0.0019
50		1.0425	0.0138	2.2086	0.0312
100		1.0144	0.0049	2.0899	0.0138
200		1.0061	0.0019	2.0333	0.0049
30	**Jeffrey's**	1.0395	0.0196	2.0425	0.0851
50		1.0283	0.0140	2.0585	0.0140
100		2.0219	0.0049	2.0087	0.0202
200		1.0041	0.0019	2.0032	0.0080

This simulation method was employed to examine the behavior of the Topp-Leone Frechet Distribution's scale parameter. In the aforementioned tables, different loss functions—square error loss function, precautionary loss function, and DeGroot loss function—have been used to demonstrate the posterior distribution under uniform and Jeffery priors. The simulation's findings and comparison of these loss functions demonstrate that posterior risk tends to decline as sample size increases and Bayes estimator approaches its true value of the parameter. We can see that DELF is performing better because its posterior risk is minimum as compared to other loss functions. Similarly, DELF is performing better. Whereas, the overall results show that the posterior distribution under Jeffery prior to DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors.

APPLICATION

For the illustration, we have taken a real dataset that was reported by (Hinkley, 1977). The data consists of the amount of precipitation (inches) having 30 successive values in March at Minneapolis/St Paul.

2.05, 2.10, 3.00, 1.95, 0.90, 0.81, 2.81, 0.59, 1.43, 1.20, 0.52, 3.09, 1.74, 0.32, 1.51, 1.87, 0.47, 1.89, 2.20, 1.62, 3.37, 1.35, 0.77, 2.48, 4.75, 0.96, 1.31, 1.20, 1.18, 0.81

Priors	**SELF**		**PLF**		**DELF**	
	BEs	**PRs**	**BEs**	**PRs**	**BEs**	**PRs**
Uniform	2.3584	0.1794	2.3618	0.0754	2.2345	0.0312
Jeffrey's	2.2823	0.1738	2.3200	0.0754	2.3584	0.0322

From the above table, we can see that DELF is performing best. The DELF under gamma prior gives the best estimates as the values of posterior risks are lower than the other the other priors.

Whereas the overall result showed that the posterior distribution under Uniform and Jeffery prior with DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors. Real-life examples and simulation studies give the almost same result.

CONCLUSION

We have analyzed the scale parameter of Topp-Leone Frechet distribution and derived posterior distributions by using informative and non-informative priors. The performance of Bayes estimates and posterior risks have been checked under different loss functions, i.e., square root error loss function, precautionary loss function and DeGroot error loss function. From overall results, it can be concluded that by increasing sample size, the posterior risk decreases. DeGroot error loss function gives best estimates under all priors. Gamma prior gives best estimates as compared to Jeffery and uniform prior. Gamma prior with DELF is performing better among other priors.

RECOMMENDATIONS

From the overall results, it can be concluded that:

- We have used only non-informative priors' others can use informative priors.
- We have used only three loss functions others can others can use other loss functions.
- Two non-informative priors have been utilized other can extend this study by using other non-informative priors.
- We have utilized only complete sampling techniques other can used other sampling techniques.

REFERENCES

1. Bantan, R.A.R., Jamal, F., Chesneau, C. and Elgarhy, M. (2019). A new power topp-leone generated family of distributions with applications. *Entropy*, 21(12), 1177.
2. Hassan, A.S., Elgarhy, M. and Ragab, R. (2020). Statistical properties and estimation of inverted topp-leone distribution. *Journal of Statistics Applications and Probability*, 9(2), 319-331.
3. Pasha-Zanoosi, H. (2022). Estimation of Multicomponent Stress-strength Reliability under Inverse Topp-Leone Distribution. *Pakistan Journal of Statistics and Operation Research*, 18(4), 837-852.
4. Radwan, S. (2021). Topp-Leone Lomax Distribution. *Al-Azhar Bulletin of Science*, 32(1), 17-25.
5. Ramos, P.L., Louzada, F., Ramos, E. and Dey, S. (2020). The Fréchet distribution: Estimation and application - An Overview. *Journal of Statistics and Management Systems*, 23(3), 549-578.
6. Reyad, H., Korkmaz, M., Afify, A.Z., Hamedani, G.G. and Othman, S. (2021). The Fréchet Topp Leone-G Family of Distributions: Properties, Characterizations and Applications. *Annals of Data Science*, 8(2), 345-366.
7. Shanker, R. and Rahman, U.H. (2021). Type II Topp-Leone Frechet distribution: Properties and applications. *Statistics in Transition*, 22(4), 139-152.
8. Vasileva, M., Malinova, A., Rahneva, O. and Angelova, E. (2021). New properties of the odd weibull inverse topp-leone cumulative distribution function. *International Journal of Differential Equations and Applications*, 20(2), 263-272.
9. Yahia, N. and Mohammed, H.F. (2019). The type II Topp Leone generalized inverse Rayleigh distribution. *International Journal of Contemporary Mathematical Sciences*, 14(3), 113-122.
10. Yousof, H.M., Altun, E. and Hamedani, G.G. (2021). A New Extension of Fréchet Distribution with Regression Models, Residual Analysis and Characterizations. *Journal of Data Science*, 16(4), 743-770.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 167-174

REGRESSION MODELS FOR FORECASTING WHEAT PRODUCTION IN PAKISTAN

Maira Asghar[1], Wajiha Nasir[2], Qurat ul Ain and **Zainab Waris**
Zakia Kousar[1], Wajiha Nasir[2], Sadia Munawar and **Maryam Hina**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]mairaasghar61@outlook.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

In this study, using regression analysis, wheat production is estimated for the future based on past data. From 1960 to 2022, annual data on wheat production in Pakistan have been analyzed using simple linear and some nonlinear methodologies. The t and F tests are employed to determine the significance of the model and the coefficients individually. Using mean absolute percentage error (MAPE), the performance of the suggested models is also compared. The results demonstrate that a simple linear regression model can be utilized for wheat production forecasting. Wheat production is forecasted from 2023 to 2030.

INTRODUCTION

Wheat production is an essential part of modern food systems. Wheat is the most important cereal crop in Pakistan. Wheat is Pakistan's primary crop, and Pakistan is the world's eighth-largest producer of the wheat, with an annual production capacity of 26.5 million metric tons. In the marketing year 2021–2022, wheat production was projected to increase by 8% to 27 million tonnes from the 25 million tonnes produced the year before. In an effort to boost domestic reserves, the Pakistani government has permitted the import of up to three million tonnes of duty-free wheat during the marketing year 2021–2022.

Figure 1: Production of Wheat during 1960–2022

Figure 1 shows the graphical representation of the data during 1960-2022 as it is shown that the production of wheat showed an increasing trend.

2. LITERATURE REVIEW

Islam et al. (2021) stated that one of most essential food crop on world is wheat. His study's main goals were to find data sets that could give a model with improved prediction capabilities and to look into the important variables that affect wheat production. On the data, hierarchical regression analysis is used. From the first data set, three more data sets (clusters) were produced. These models have been subjected to model selection criteria, including R^2, adjusted R^2, MSE, SIC, AIC, ER (AIC) and Wi (AIC), In addition to this, they used the histogram with normal plots and Q-Q plots to check the normality of the response variable in the data set. They concluded the results of block 2 (model 2 with 15 predictors) in a hierarchical regression analysis are superior They also came to the conclusion that, based on their research, all fertilizer, advanced varieties, soil type, seed treatment, spry, irrigated land, and planting timing are important factors in the production of a better yield of wheat crop. Jamal and Rind (2007) conducted a study where he took the assumptions of OLS estimation into consideration conducting a study to create forecasting models for wheat crop acreage and production for the Chakwal district in the Rawalpindi region. Additionally, he said that manures, urea fertilizer, and DAP fertilizer all play important roles in boosting wheat crop output. To create the forecasting model, he considered all the major variables, including rainfall, temperature, chemical fertilizers, manure, and cultural practices that have an impact on the region and productivity. There was multicollinearity in the data for the wheat area model. To address the multicollinearity issue in data, ridge regression was applied. The forecasting models were created using data collected over a period of 15 years, from 1984–1985, to 1998–1999. Wheat area and production for the following five years, from 1999–2000 to 2003–2004, were then predicted using models and compared to the actual data. After assessing the models' correctness, final models were created using data spanning 20 years, from 1984–1985 to 2003–2004. He claimed that these linear models can be used to predict wheat production and area for the following five years. OLS estimate assumptions had also been tested, and the data did not demonstrate any violations of the assumptions under consideration. He concluded that the yield of wheat crops is considerably increased by manures, DAP fertilizer, and urea fertilizer. The number of ploughs in the wheat fields has a big impact on increased crop output. Wheat crop output is impacted heavily the rainfall during the months of October and November and the mean maximum temperature in the month of March severely reduces it. Ahmed et al. (2017) stated that sinusoidal model served purpose to detect the trends in production and productivity of wheat grown in Punjab and Pakistan.

3. DATA AND METHODOLOGY

The study's flowchart is depicted in Figure 2. This research compiles Index Mundi data on the annual production of wheat in Pakistan (1000 Mt) from 1960 to 2022. It is considered that the historical relationship between wheat production and years can be used to forecast future wheat production values. In the study, it is thus expected that future production would follow the established historical trend. In other words, the forecasts are derived from all available data.

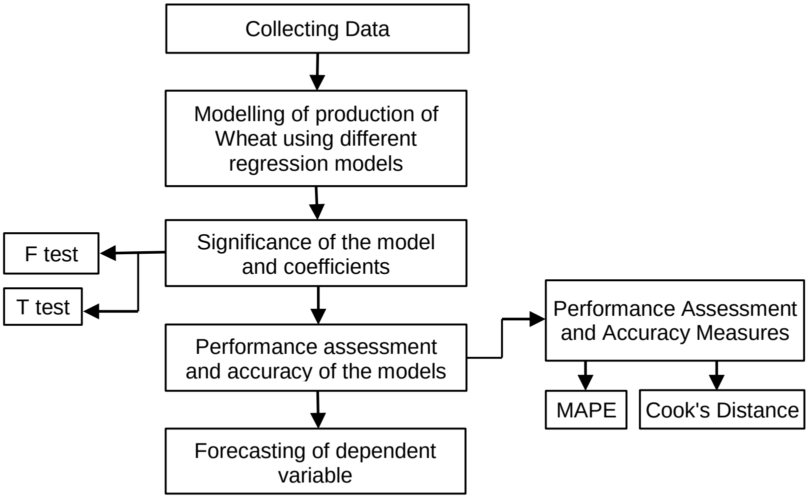


Figure 2: Flow Chart of the Study

For the development of models, linear and nonlinear regression analysis methodology, including simple linear regression, logarithmic regression, power regression, and exponential regression are utilized. R Studio is the program that is used for the analysis of the data. The information was separated into two categories: data to be used for training the model, and data to be used for testing of the model. The linear and nonlinear functions are given below, which are represented in the equations [1] to [4]:

Simple Linear regression model: $y = a + bt + \varepsilon_t$ (1)

Exponential regression model: $y = a\, e^{bt} + \varepsilon_t$ (2)

Power regression model: $y = a\, t^b + \varepsilon_t$ (3)

Logarithmic regression model: $y = a + b\, ln(t) + \varepsilon_t$ (4)

where y is the production of wheat measured in 1000 mt, t is the market year , b is the slope, a is the constant or the intercept, εt is the difference between the actual values and the predicted values. The developed models were checked for accuracy by taking into consideration the F Statistic the behavior of the determination coefficient (R^2), and the t value of the coefficients. The R^2 provides a measure against which the accuracy of the model may be evaluated. R^2 values that are high are indicative of a more reliable model. The F test is used to measure the significance of the results of the overall model then t values are used to determine the individual significance of the coefficients of the model. The validity of the model is validated if the computed t and F values (calculated by R Studio) are higher than the tabulated F and t values. This confirms the validity of the model. In order to choose the most effective model for forecasting, performance evaluations of the models that are constructed are carried out. For the purpose of performance evaluation, the MAPE Equation [5] is utilized, and the accuracy of the regression model is determined by obtaining cook's distance, which is validated using graph. MAPE values that are closer to zero suggest more accurate prediction.

$$\text{MAPE} = \frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{|e_i|}{y_i}\right) \tag{5}$$

where yi are the actual values, ei are the differences that exist between the actual values and the predicted values, and n is the total number of measurements.

Table 1
Results of the Regression Analysis

Model	Coefficients	Unstandardized Coefficients	R2	Adjusted R2
SIMPLE LINEAR	(Intercept) t	-791649.767 405.155	0.9787	0.9782
EXPONENTIAL	(Intercept) t	-65.102932 0.037485	0.9506	0.9495
POWER	(Intercept) t	-555.860 74.429	0.9517	0.9506
LOGARITHMIC	(Intercept) t	-6091331 803850	0.9783	0.9778

Table 2
Results of the Significance (F and t) Tests for the Developed Regression Models

MODEL	F	T (SLOPE)	T (INTERCEPT)
SIMPLE LINEAR	2201	46.91	-46.19
EXPONENTIAL	923.1	-26.59	30.38
POWER	944.9	30.74	-30.23
LOGARITHMIC	2163	46.50	-46.41

Table 3
MAPE Values of Developed Regression Models

MODEL	MAPE
SIMPLE LINEAR	0.03799452
EXPONENTIAL	0.9995846
POWER	0.9995848
LOGARITHMIC	0.03832779

Table 4
The Criteria of MAPE for Model Evaluation Based on Lewis (1982)

MAPE %	Evaluation
$\boldsymbol{MAPE \leq 10\%}$	High Accuracy Prediction
$\boldsymbol{10\% < MAPE \leq 20\%}$	Good Prediction
$\boldsymbol{20\% < MAPE \leq 50\%}$	Reasonable Prediction
$\boldsymbol{MAPE > 50\%}$	Inaccurate Prediction

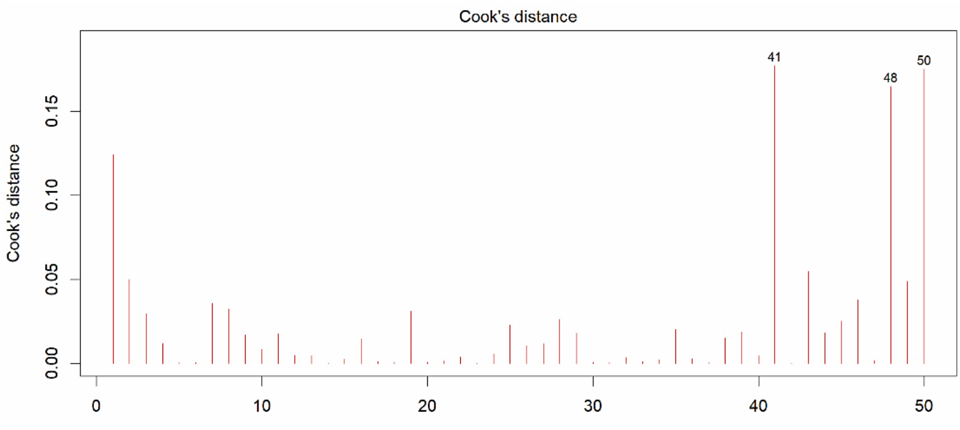


Figure 3: Cook's Distance of the Best Proposed Model

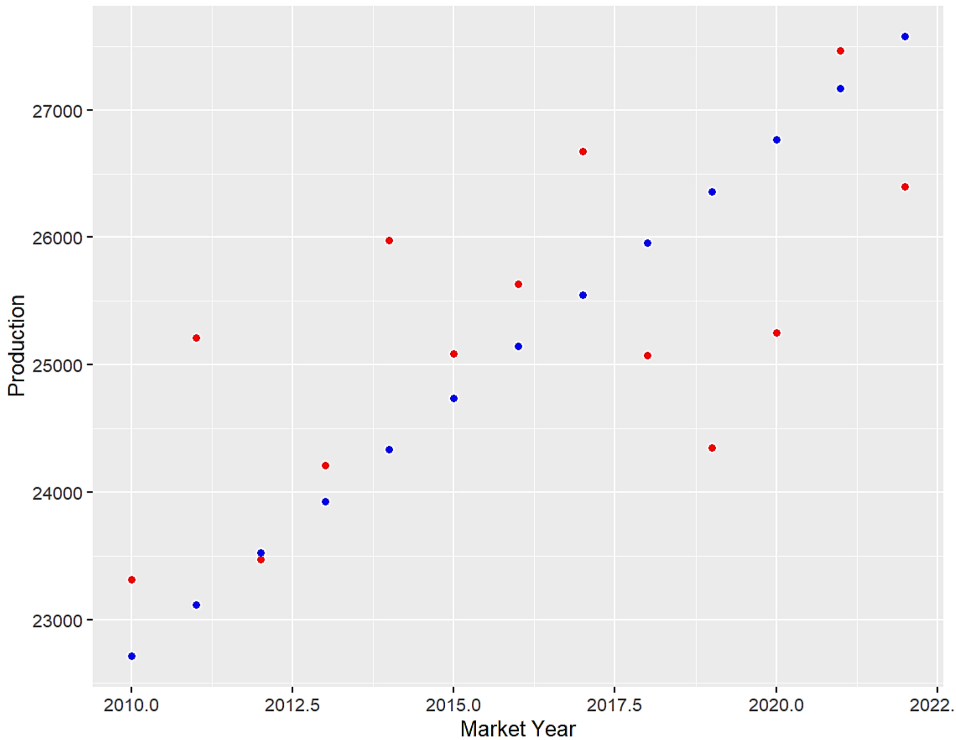


Figure 4: Illustration of the Actual and Predicted Production of Wheat of Simple Linear Regression

4. MODELING AND FORECASTING PRODUCTION OF WHEAT

Table 1 shows the statistical analysis of the regression models. As Shown in Table 1 values of intercept, slope, R^2, adjusted R^2 are shown we can see that value of intercept of simple linear regression is -791649.767 whereas slope is 405.155 and, market year can account for 97.87% of the variation in production of wheat. The value of intercept of exponential regression is -65.102932 whereas slope is 0.037485 and, market year can account for 95.06% of the variation in production of wheat. The value of intercept of power regression is -555.860 whereas slope is 74.429 and, market year can account for 95.17% of the variation in production of wheat. The value of intercept of logarithmic regression is -6091331 whereas slope is 803850 and, market year can account for 97.83% of the variation in production of wheat. Developed models (Eqs. [6] to [9]) are presented subsequently.

Simple Linear regression model: $y = 405.155t - 791649.767$ (6)

Exponential regression model: $y = (-65.102932)\, e^{0.037485\, t}$ (7)

Power regression model: $y = (-555.860)\, t^{74.429}$ (8)

Logarithmic regression model: $y = 803850 \ln(t) - -6091331$ (9)

where y is the Production of Wheat (1000 Mt) and t is the market year.

Table 5
Forecasted Production of Wheat (1000 Mt) Using the Proposed Models

Regression Type				
YEAR	**Simple Linear**	**Exponential**	**Power**	**Logarithmic**
2023	27979.68	10.72968	10.71859	27848.09
2024	28384.84	10.76717	10.75537	28245.34
2025	28790.00	10.80465	10.79214	28642.40
2026	29195.15	10.84214	10.82888	29039.27
2027	29600.31	10.87962	10.86561	29435.94
2028	30005.46	10.91711	10.90232	29832.41
2029	30410.62	10.95459	10.93901	30228.69
2030	30815.77	10.99208	10.97569	30624.77

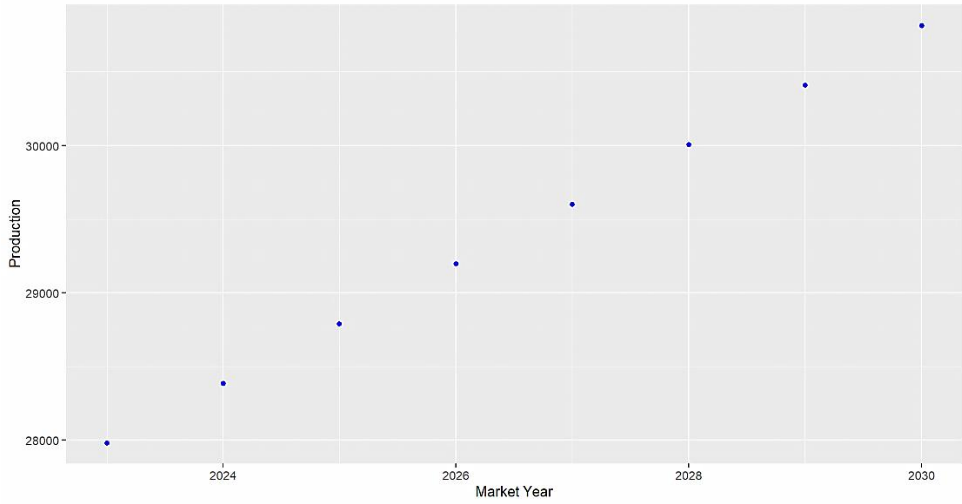


Figure 5: Forecasted Production of Wheat using the Best Proposed Model

Table 2 displays the results of the validation tests that were conducted. As can be seen, the calculated F and t values are higher than the values that have been tabulated, indicating that the model as a whole and the individual variables that are involved in the model are correct, respectively. The MAPE values that were determined for the forecasting models ranged between 0.03799452 and 0.9995848 (See Table 3). It can be seen that of all the regression models, the Simple Linear regression model has the lowest MAPE value. This suggests that the best model that is proposed is a high accuracy prediction model because the MAPE values for the model is less than 10% (See Table 4). Therefore, it can be utilized for the purpose of obtaining forecasts for wheat production that are more accurate hence it is the best fitted model. In addition, Figure 3 illustrates the cook's distance of the simple linear regression model. This figure demonstrates that all of the values are less than 1, indicating that our proposed model is functioning effectively. Furthermore, Figure 4 shows the predicted Production of Wheat based on the model that performed the best. These results indicate that the simple linear regression model can be utilized successfully for the purpose of forecasting the production of wheat. Table 5 contains forecasts of wheat production for the years 2023 to 2030 based on the models that have been developed, and Figure 5 illustrates the forecasts using the model that is considered to be the most accurate. According to the results of the modelling, the Production of Wheat in 2030 could range anywhere from 10.97569 (1000 Mt) to 30815.77 (1000 Mt) depending on the models that were proposed. The simple linear regression model estimated that the annual production of wheat would reach 30815.77 (1000 Mt) in the year 2030.

5. CONCLUSION

Recently, there has been a rise in the production of wheat, and this trend is expected to continue in the future. The high determination coefficients of the proposed models indicated that the models had a high reliability for forecasting the production of wheat. The significance of the whole model as well as coefficients of the model are validated by the results of the significance test that was performed on the model and the coefficients. Because the MAPE values for the proposed model is less than 10% so it is determined that the proposed regression model can be successfully utilized as a forecasting tool for the production of wheat. According to the findings of the research, the production of wheat in the year 2030 could range anywhere from 10.97569 (1000 Mt) to 30815.77 (1000 Mt), depending on the type of regression employed. It was discovered that the simple linear regression model provides the most accurate forecasting performance for wheat production. Using the simple linear regression model, the production of wheat is forecasted as 30815.77 (1000 Mt) in 2030. In addition, the following suggestions pertaining to the production of wheat are given below:

- There are a variety of resources, including seed ratio, soil, irrigation, fertilizer, pesticides, and regional factors, that can be utilized effectively to increase wheat production.
- The farmers should implement effective methods of irrigation, as this will help them cut their expenses and contribute to the production of abundant harvests.
- Education also provides the opportunity to adapt to emerging technological trends, which can increase both production and income. Literacy levels of farmers are a crucial factor in determining how open they are to new ideas and how effectively they allocate resources, and the adoption of modern education is widely acknowledged as a key factor in agricultural development.
- Farmers who have received formal education will find it much easier to take advantage of new opportunities, and they will be better able to adapt to shifting sociocultural norms. For this reason, the government should establish a number of small agricultural training institutions specifically for the farming community.

REFERENCES

1. Islam, M., Shehzad, F. and Omar, M. (2021). Modeling wheat productivity using hierarchical regression: a way to address food security concerns. *Ilköğretim Online*, 20(2), 1184-1195.
2. Jamal, N. and Rind, M.Q. (2007). Ridge Regression: A tool to forecast wheat area and production. *Pakistan Journal of Statistics and Operation Research*, 3(2), 125-134.
3. Ahmed, M., Khan, M.S. and Iqbal, N. (2017). Wheat Production Trend in Pakistan—*A Statistical Analysis. Journal of Agricultural Research*, 55, 115-124.
4. Lewis, C.D. (1982). *International and business forecasting methods*. London, England: Butterworths.
5. Gujarati, D.N. (2003). *Basic Econometrics*, Fourth Edition, McGraw-Hill. New York.
6. Gujarati, D.N. and Porter, D.C. (2004). Basic econometrics (ed.) McGraw-Hill. Irwin, a business.
7. Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models*, 5th Edition, McGraw-Hill, Inc., New York.
8. Hoerl, A.E. anb Kennard, R.W. (1970a), Ridge Regression: Biased Estimation for Non-orthogonal Problems. *Technometrics*, 12(1), 55-67.
9. Gruber, M.H.J. (1998). *Improving Efficiency by Shrinkage*, Marcel Dekker, Inc. New York.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 175-184

COMPARISON OF ARTIFICIAL NEURAL NETWORK AND AUTO REGRESSIVE INTEGRATED MOVING AVERAGE MODELING, A CASE STUDY OF FORECASTING OF RICE CROP OF PUNJAB, PAKISTAN

Ayat Mahfooz[1] and **Wajiha Nasir**[2]
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]ayyatmahfooz@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

The major goal of this research is to develop a reliable model to predict Pakistan's annual food crop. The information was gathered from the Crop Reporting Service Punjab website and spans the years 1990 through 2021. By contrasting the Box Jenkins Methodology and Artificial Neural Network, Time series analysis techniques with EViews-10 and Zaitun Time series software, the area, production, and yield of the Pakistani rice harvest are observed and examined. For analysis, estimate, and forecasting purposes, these models were employed. In the empirical investigation, the stationarity of the data in the EViews software was examined using the ADF (Augmented Dickey fuller) unit root. Zaitun Software employs MAE, while GMDHL software uses RMSE and MAPE to evaluate forecasting accuracy. We have projected the values of rice crops for the years 1990 through 2021. To forecast future values of rice crop area, production, and yield, we compared the results of Box Jenkins and artificial neural networks.

KEYWORDS

Rice, Area, Production, Yield, Time series, Autoregressive-Integrated-Moving Average model (ARIMA), Artificial Neural Networks (ANN).

1. INTRODUCTION

The second most widely farmed cereal grain, rice is the main food source for more than half of the world's population. Almost 3 billion people consume more than 100 kg of rice per year. In the past 30 years, the area under rice cultivation has increased by 155.5 million ha or 0.39% annually.

In terms of global rice production, Pakistan ranks tenth. Around 8% of all rice trade worldwide is represented by exports from Pakistan. The Pakistani agricultural economy depends heavily on this crop. One crucial Kharif crop is rice. Rice is Pakistan's second-most significant food crop after wheat and a key source of income. The areas of Pakistan where rice is grown range in latitude from 24° to 36°, in elevation from sea level in the south to 2500 meters in the northern mountain range and valleys and terraces, and climate from arid hot plains to coastal tropical damp districts in southern Sindh. Traditionally, forecasting has remained the most important part of structural econometric modeling because generally, it focuses on the one-dimensional time series model known as Autoregressive Integrated Moving Average (ARIMA) model. The foundation of the contemporary method of time series analysis is the Box and Jenkins (1970) procedure. This model is widely applicable for time series forecasting, simulation, and distributions. Environmental science has recently made use of some non-parametric and adaptive models for forecasting. When compared to mechanistic models, artificial neural network (ANN) models are comparatively simple while being a powerful empirical modeling tool.

The demand for rice has so far been met by the global supply. We have the opportunity to improve food security, lower poverty, and preserve the environment during the International Year of Rice, which benefits the billions of people for whom rice is essential to life (van Nguyen & Ferrero, 2006). In this work, three layers of feed-forward back propagation ANN and ARIMA hybrid models will be developed to predict the index value and trend of the Indian stock market in the future. The results demonstrate that for BSE Oil & Gas, ANN, and hybrid ANN ARIMA can manage the input data set and estimate the future closing price, but ARIMA and ARIMA ANN cannot forecast future values (Merh et al., 2010)**.** Rice is crucial for the nutrition of sizable populations in the Asia-Pacific region, some of Latin America and the Caribbean, and, more and more, in Africa. More than 200 million households in underdeveloped countries rely mostly on it for work and income (Muthayya et al., 2014). In this study, day-ahead electricity price forecasting for the Denmark-West region is done using a 24-hour forecasting window. An analysis of external regressors using the backward elimination method suggests that excluding temperature is beneficial to forecasting models (Karabiber & Xydis, 2019). The variations in rice production and area during the past 30 years in China show that changes in rice area had a substantial impact on the rise or fall of total production. The results show that changes in rice area significantly impacted changes in rice production.(Liu et al., 2013)**.** One of the objectives of agricultural production is to maximize crop yield at the lowest possible cost. Utilizing a variety of soil and atmospheric parameters, crop prediction technology predicts the best crop to plant. Parameters (SDahikar & VRode, 2014)**.** Data on the output of milled rice from the years 1960 to 2010 were gathered from secondary sources and used in this analysis. Although a ten-year forecast using the model indicates an upward trend in production, the forecast value for 2015 was insufficient to compare with Nigeria, the continent's top rice producer at the time, which had a production of 2700 thousand tonnes at the time (Suleman & Sarpong, 2012). In this study, rice yields were forecast using a range of algorithms, including the non-linear feed-forward back propagation neural networks (FFBN) tool and the linear model of partial least squares regression. Although climatic data could be utilized to predict rice yields, appropriate adjustments need to be performed to increase the modeling's accuracy and dependability (Guo et al., 2021). This article provided us with a general overview of how remote sensing data is used to map regions used for rice farming and predict how much will be produced. According to the research, new satellites like the MODIS constellation and the future MODIS will improve the temporal, spectral, and spatial resolutions that could be a good fit for mapping rice fields and predicting their production (Mosleh et al., 2015).

2. METHODOLOGY

2.1 Autocorrelation Function (ACF)

A time series' correlation with itself $(z_1, z_2, \dots, z_n)$ is known as the autocorrelation function Lag-k autocorrelation of z_t is the term for the correlation coefficient between z_t and z_{t-1}, which is indicated by t and is based on the weak stationary assumption.

The average correlation between data points in a time series and earlier values in the series calculated for various lag lengths is quantified and plotted using an ACF.

2.2 Partial Autocorrelation Function (PACF)

The correlation coefficient between z_t and z_{t-1} after removing the impact of the intervening $z_{t-1,t-2}, \dots, z_{t-k+1}$ is called partial autocorrelation function at lag-k, denoted by: ϕ_{kk}

The partial autocorrelation function (PACF) in time series analysis provides the partial correlation of a stationary time series with its own lagged values, regressing the time series values at all shorter lags.

2.3 Autoregressive Moving Average Model (ARMA)

ARIMA (p, d, q) is used to represent the autoregressive moving average model,

where

p: The order of autoregressive.

q: The order of moving average.

d: The order of non-seasonality difference.

The general form of an ARMA model for stationary time series can be expressed as follows:

$$(B)x_t = \delta + \theta(B)a_t$$

A non-stationary series must first be converted into a stationary one by taking into account pertinent differences:

$$\nabla^d x_t = (1 - B)^d x_t = x_t - x_{t-s}$$

$(1 - B)^d$: The d^{th} difference.

B: The backward shift operator.

For non-stationary time series, the general form of an ARMA model can be written as:

$$(B)(1 - B)^d x_t = \delta + \theta(B)a_t$$

where

$(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, is the autoregressive operator of order p.
$(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^1$, is the moving average operator of order q.
$a_t = NI(0, \sigma_a^2)$

2.4 Model Building

The identification, parameter estimation, diagnostic checking, and forecasting stages make up the Box-Jenkins iterative approach for building linear time series models. The most significant phase is identification. The proper ARIMA model is included in it. To find any areas where the model is deficient, diagnostic model checking entails putting the model's assumptions to the test.

To determine the suggested ARIMA model's goodness of fit, such that

$$H_0 = \rho_k(a_t) = 0$$
$$H_1 = \rho_k(a_t) = 0$$

If Q lies in the extreme 5% of the right side tail of the chi-square distribution, we reject thehypothesis where the residuals are random.

2.5 Neural Networks

2.5.1 Biological Neural

The biological neuron gets inputs from all parts of the body, combines the inputs, engages in nonlinear processing, and then provides the output result. The brain is a very sophisticated, nonlinear, parallel computer. Over 100 billion neurons make up the human brain. The average neuron is as complicated as a small computer and can physically link to up to 10,000 other cells. The cell body, dendrites, axons, and synapses are the four components of a neuron.

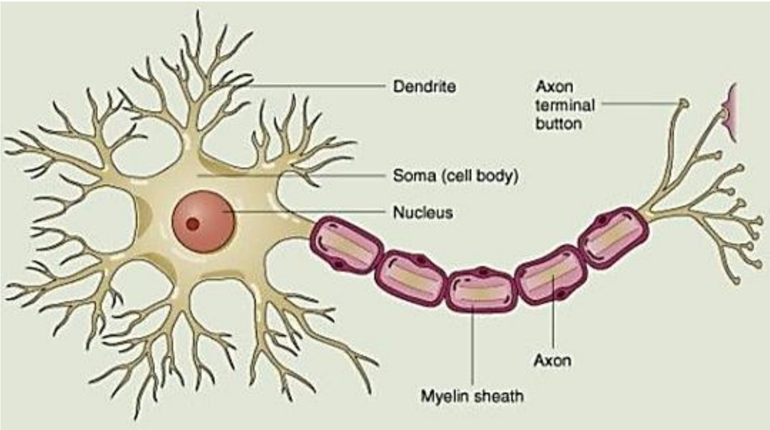


Figure 1: Biological Neural Network

2.5.2 Artificial Neural Network (ANN)

By the use of biological neurons, an artificial neural network simulates how the human brain works. Artificial neural networks imitate intricate brain functions and carry out useful calculations. McCulloch and Pitts proposed the initial model of an artificial neuron in 1943. The use of neural networks in biology, engineering, medicine, and finance is widespread. Either the feed-forward approach or the fed-back approach can be used to create an artificial neural network. The three types of layers in an artificial neural network are input, hidden, and output layers. As shown in Figure 2.

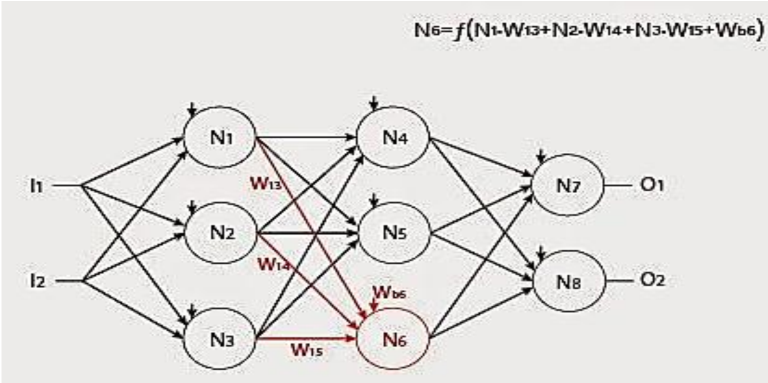


Figure 2: Artificial Neural Network

3. DATA ANALYSIS AND RESULTS

3.1 Data Description

In order to develop a suitable model to be used to forecast the yearly area, production, and yield of rice crops, the data for the current study were the annual area, production, and yield of rice crops. The data were obtained from the website "Crop Reporting Service Punjab" for the period (1990 to 2021). The statistical programs: EViews-10 and Zaitun Time Series were utilized.

3.2 Comparison of ARIMA and ANN

For ARIMA, the normality of area, production, and yield of the rice crop is determined by the probability value of the Jarque-Bera test. If the probability of Jarque-Bera is greater than 0.05 so we accept the null hypothesis, which means residuals are normally distributed. By taking the first log, in the case of the rice field, we make our data stationary. The model (0,1,0) for the annual rice area is then chosen based on ACF and PACF values since it has the lowest AIC and SIC values. In order to make our data stationary for yearly rice production, we additionally use the first log, and we choose the (0,1,0) model because it has the lowest AIC and SIC values. For the annual yield of the rice crop, the statistics become flat and steady. We choose the model (1,0,5) because it has the highest adjusted R-squared value and the lowest AIC and SIC values. Using the Augmented Dickey-Fuller (ADF) test, we demonstrate the stationary nature of the annual data on rice crop area, production, and yield.

In order to forecast the area, productivity, and yield of the rice crop for ANN, we placed twelve input layer neurons, ten hidden layer neurons, and one output layer neuron. With a learning rate of 0.05 and momentum of 0.5, we employ the bipolar sigmoid function and calculate mean absolute error (MAE) on Zaitun Time Series. Using GMDHL software, we calculate Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

The comparison of the mean absolute error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) of ARIMA and ANN is as:

Table 1
MAE, RMSE and MAPE of ARIMA

ARIMA	MAE	RMSE	MAPE
Rice Area	196.2352	263.4299	206.3611
Rice Production	174.9886	262.0721	248.4726
Rice Yield	0.547049	0.678643	3.041142

Table 2
MAE, RMSE and MAPE of ANN

ANN	MAE	RMSE	MAPE
Rice Area	37.66845	173.430	3.1710
Rice Production	50.59351	130.399	2.6353
Rice Yield	0.243813	0.48283	1.9494

The ANN forecasted values and graph of the area, production, and yield of rice crops are as:

- Artificial Neural Network ANN Forecasted Values

Table 3
ANN Forecasted Values of Rice Area

2022	6418.5261
2023	6454.7758
2024	6462.0176
2025	6473.4408
2026	6475.2119
2027	6473.9238
2028	6474.5444
2029	6475.3212
2030	6475.4823

Table 4
ANN Forecasted Values of Rice production

2022	5905.6170
2023	5959.8845
2024	5973.8917
2025	5988.9945
2026	5995.7007
2027	5994.0784
2028	5996.2576
2029	5998.3455
2030	5997.8308

Table 5
ANN Forecasted Values of Rice Yield

2022	22.6185
2023	22.5188
2024	22.7139
2025	22.5760
2026	22.6981
2027	22.8107
2028	22.8025
2029	22.8420
2030	22.8520

- Artificial Neural Network (ANN) Forecasted Graph

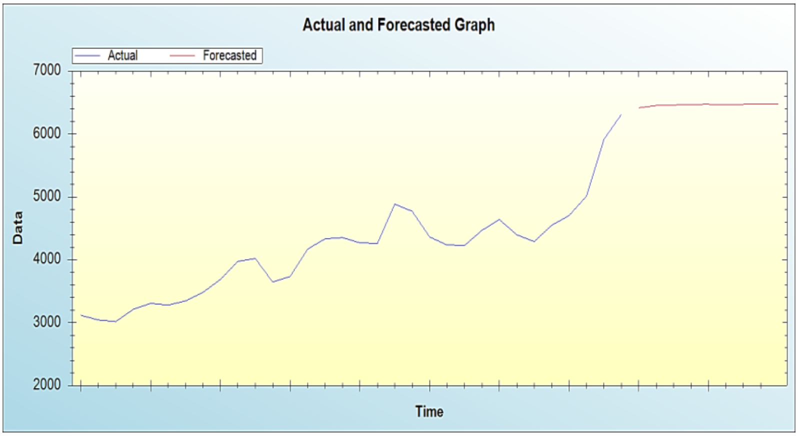


Figure 3: ANN Forecasted Graph of Rice Area

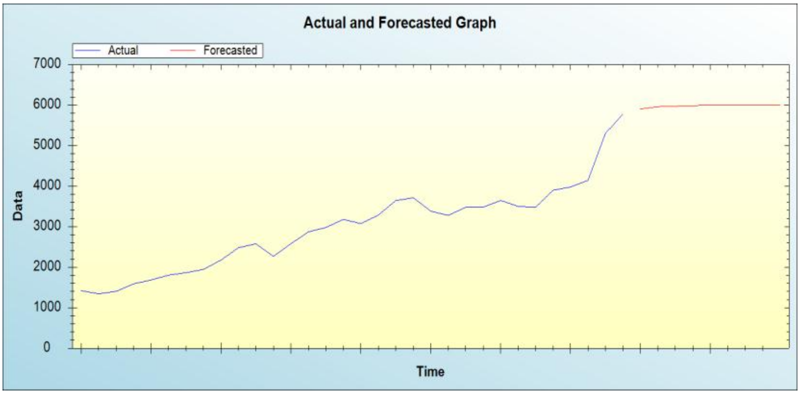


Figure 4: ANN Forecasted Graph of Rice Production

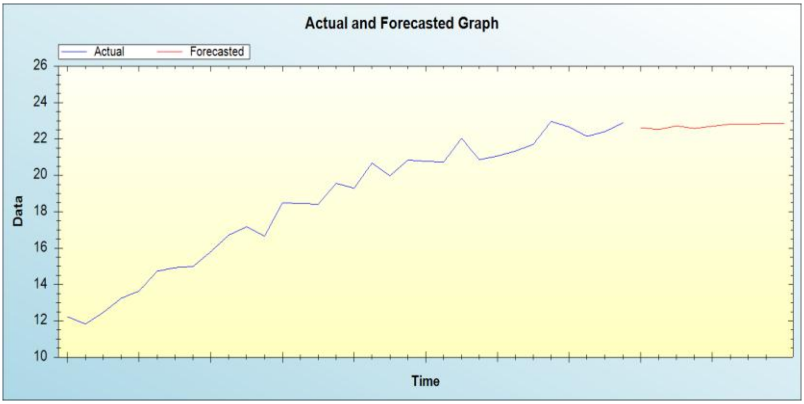


Figure 5: ANN Forecasted Graph of Rice Yield

4. CONCLUSIONS

Rice is the staple food for over half of the world's population. Almost a billion households depend on rice systems as their main source of employment and livelihood. Rice systems support a wide variety of plants and animals.

In the current research, an attempt was made to study whether an ARIMA model achieves better results than Artificial Neural Network. The study recommended that the most suitable and parsimonious model is Artificial Neural Network (ANN) as an appropriate forecasting tool to predict the area, production, and yield of Rice crops in Punjab, Pakistan for the next 9 years. When we use ANN instead of ARIMA, it gives us less valueof Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). This is an accurate sign of this model.

We recommend that government should use the ANN models forthe forecasting of area, production, and yield of rice crops in Punjab, Pakistan.

REFERENCES

1. Fernandes, P.O., Teixeira, J.P., Ferreira, J.J. and Azevedo, S.G. (2008). Modelling tourism demand: A comparative study between artificial neural networks and the Box-Jenkins methodology. *Journal for Economic Forecasting, Institute for Economic Forecasting*, 5(3), 30-50.
2. Guo, Y., Xiang, H., Li, Z., Ma, F. and Du, C. (2021). Prediction of rice yield in East China based on climate and agronomic traits data using artificial neural networks and partial least squares regression. *Agronomy*, 11(2). DOI:10.3390/agronomy11020282
3. Ji, B., Sun, Y., Yang, S. and Wan, J. (2007). Artificial neural networks for rice yield prediction in mountainous regions. *Journal of Agricultural Science*, 145(3), 249-261.
4. Karabiber, O.A. and Xydis, G. (2019). Electricity price forecasting in the Danish day-ahead market using the TBATS, ANN and ARIMA methods. *Energies*, 12(5). https://doi.org/10.3390/en12050928
5. Kihoro, J.M., Otieno, R.O. and Wafula, C. (2004). Seasonal Time Series Forecasting: A Comparative Study of Arima and Ann Models. *African Journal of Science and Technology (AJST) Science and Engineering Series*, 5(2). DOI:10.4314/AJST.V5I2.15330
6. Liu, Z., Li, Z., Tang, P., Li, Z., Wu, W., Yang, P., You, L. and Tang, H. (2013). Change analysis of rice area and production in China during the past three decades. *Journal of Geographical Sciences*, 23(6), 1005-1018. https://doi.org/10.1007/s11442-013-1059-x
7. Merh, N., Saxena, V.P. and Pardasani, K.R. (2010). A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting. *Business Intelligence Journal*, 3(2), 23-43.
8. Mosleh, M.K., Hassan, Q.K. and Chowdhury, E.H. (2015). Application of remote sensors in mapping rice area and forecasting its production: A review. *Sensors (Switzerland)*, 15(1), 769-791. MDPI AG. https://doi.org/10.3390/s150100769
9. Muthayya, S., Sugimoto, J.D., Montgomery, S. and Maberly, G.F. (2014). An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, 1324(1), 7-14.

10. Pandey, O.P., Somvanshi, V.K., Pandey, O.P., Agrawal, P.K., Kalanker, N. v, Prakash, M.R. and Chand, R. (2006). Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, 10(2), 141-151.
11. SDahikar, M. and VRode, D. (2014). Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. *International Journal of Innovative Research in Electrical*, 2(1), 683-686.
12. Suleman, N. and Sarpong, S. (2012). Forecasting milled rice production in Ghana using Box-Jenkins approach. *International Journal of Agricultural Management and Development*, 2(2), 79-84.
13. Tuong, T.P. and Bouman, B.A. (2003). Rice production in water-scarce environments. In J.W. Kijne, R. Barker and D. Molden (Eds.) *Water productivity in agriculture: Limits and opportunities for improvement*, CABI Publishing (pp 53-68).
14. Van Nguyen, N. and Ferrero, A. (2006). Meeting the challenges of global rice production. *Paddy and Water Environment*, 4, 1-9.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 185-194

FORECASTING OF SOME ECONOMIC VARIABLES OF PAKISTAN: A COMPARISON OF SOME TIME SERIES METHODOLOGIES

Fareeha Nosheen[1]**, Wajiha Nasir**[2] and **Zulaikha**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]fareehanosheen56@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

The purpose of this paper is to analyze and evaluate efficient forecasting model of some economic indicators of Pakistan. In this article, we consider the problem of multicollinearity between economic variables in the time series data. Some univariate and multivariate time series methodologies have been compared. Economic variables such as GDP, GNP, Exchange rate, and CPI has been considered. Annual data on these variables has been considered. Box Jenkins methodology and vector autoregressive methods has been compared by using different forecasting evaluations measures such as RMSE, MAPE, and MAE statistics has been used.

KEYWORDS

MAPE, ARIMA, GDP, GNP, VAR.

1. INTRODUCTION

Multicollinearity defined as the correlation between independent variables in regression is common problem while handling time series data. When multicollinearity occurs, the coefficient estimates for the individual predictor variables become unstable, making it difficult to interpret the effect of each variable on the dependent variable. If their presence a high correlation among the variables then we apply VAR. In this article, our objective is to compare the performance of different methods of forecasting in the presence of multicollinearity. We use economic variables such as GDP, GNP, exchange rate (ER) and CPI have a significant impact on the economy and can affect the country’s growth and development of economic performance. Gross Domestic product (GDP) and Gross National product these indicators reflect the value of all goods and services produced within the country (GDP) and by its citizens, including those produced abroad (GNP). Higher GDP and GNP indicate a strong and growing economy, while a decline in these indicators can signal an economic downturn. Exchange rate represent the value of one nation's currency is exchanged with respect to other currencies. Exchange rate are relative measures represent as a comparison of the currencies of two countries. Predicting exchange rates has become one of the most challenging applications of financial time series forecasting. Pakistan is a developing country facing a continuous decrease in dollar exchange rates. Compared to 1990, Pakistan's currency is declining in value gradually. It was 21.71 PKR on a dollar, while in 2000, it became equal to 53.65 PKR. Lately, in 2021, the US dollar got a record high exchange rate of 176 PKR in interbank rates. There was a rapid decline in the value of Pakistani rupees with respect to other currencies, particularly the US dollar and British pound. There were several reasons behind the depreciation of Pakistan's currency. A significant reason for the increase in dollar demand is still the trade imbalance, stagnant exports, and increased imports. The consumer price index (CPI) is a measure of inflation, which is the rate at which the prices of goods and services increase over time. Inflation can erode the purchasing power of a currency and can make it more expensive for consumers to purchase goods and services. Bagshaw (1986) has compared the performance of univariate ARIMA, multivariate ARIMA and VAR. Khan (2011) forecast the total import for Bangladesh using SRIMA and ANN. They have used monthly data of total imports. They have found VAR better than other methods.

2. LITERATURE REVIEW

Jilani et al. (2013) evaluated the impact of macroeconomic variables such as inflation, real exchange rate, and interest rate on the GDP of Pakistan from 1980 to 2013. Multivariate regression analysis was used to analyze the data consisting of GDP as a dependent variable while the independent variables were interest rate, exchange rate, and inflation rate. Shahini and Haderi (2013) study used ARIMA and VAR model to forecast the GDP of Albania. VAR model gives better results in GDP forecasting rather than the ARIMA model. Zhang Haonan (2013) examined the forecasting of per capita GDP for five regions of Sweden using ARIMA, VAR, and AR (1). Based on their study, autoregressive first-order models are the best for forecasting the per capita GDP of five Swedish regions.

Kwasi and Sharma (2015) have modeled and forecast the monthly price of groundnut in Bikaner district of India. They have used ARIMA and var for the purpose. Var has comparatively performed better as compared to ARIMA on comparing their forecasting accuracies which are measured by MAPE. Oluyemi and Isaac (2017) has studied the impact of exchange rate on exports and imports in Nigeria using a VAR model. VAR results show that the exchange rate positively affects imports but negatively affects exports.

Khan and Khan (2020) have compared the performance of forecasting techniques of economic data of Bangladesh. They have used annual data and found var more accurate then ARIMA modeling in the presence of correlated variables.

Ghazo (2021) has studied modeling and predicting GDP and CPI in Jordan. Box- Jenkins methodology was applied for the period 1976-to 2019. Based on the results, ARIMA (3,1,1) was found to be the best model for the GDP, and ARIMA (1,1,0) was the best model for forecasting the CPI. In 2020, the CPI is forecasted to rise, while the GDP is forecast to decline. After 2020, both GDP and CPI increased, which indicates that the Jordanian economy is trending toward cost-push inflation.

Mohanapriya and Ganapati (2021) has trading volume for cotton using ARIMA and VAR. they have used monthly data of trade volume for their study. They have selected their best model using AIC and SBC criteria. For comparing the performance of forecasting methods, RMSE, MAPE, MAE and Theil U statistics. They have found ARIMA model better than VAR. Katris (2021) has studied the unemployment of Greece by using auto regression methods. They have used monthly data of unemployment in the presence of COVID-19.

Hussain et al. (2022) has studied the monthly data of macroeconomic variable such as inflation, exchange rate and GDP. ARIMA and ANN model were used. The result conclude that ANN model better perform rather than ARIMA. Saleem et al. (2022) has studied the GDP data of Pakistan by utilized Box Jenkins methodology. Based on Box Jenkins methodology, he forecasted the GDP of Pakistan for the next 55 years from 2021 to 2075. As compared to 2025, He found that Pakistan's GDP in 2035 would be 2.51199, and it would decrease in 2075.

In this article, we want to compare univariate and multivariate forecasting methods for correlated data. For this Box-Jenkins will be served as univariate and vector autoregression will be served as multivariate technique. Several economic variables are available for study but we will utilize GDP, GNP, ER and CPI. We will apply both methods and will compare the performance of these methods by using forecasting evaluation techniques such as MAE, MAPE, RMSE and Theil u statistics.

3. DATA AND METHODOLOGY

The data of economic variables in Pakistan have been forecast using Box-Jenkins methodology and VAR model using EViews. GDP, GNP, ER and CPI are selected as economic variables because of their relationship. The data have been collected from the World Bank website ranging from 1962 to 2021.

Box Jenkins Methodology

It is used for forecasting. Box-Jenkins refers to a systematic method of identifying, fitting, checking, and using integrated auto-regressive, moving average (ARIMA) time series models. It has the following steps:

- Stationary checking
- Model identification
- Parameter estimation
- Diagnostic checking
- Forecasting

Vector Autoregressive model

The vector autoregression (VAR) model is widely used in multivariate time series analysis. It is an extension of the univariate autoregressive model to dynamic multivariate time series. Using the VAR model for multivariate analysis, it has been shown to be particularly useful for analyzing economic and financial time series dynamically and predicting future outcomes.

Forecasting Errors

Error is an essential part of forecasting, therefore to the maximum accuracy of error RMSE, MAE, and MAPE are used to compare the performance of the forecasting models. The formulas for calculating these errors are as follows,

Root means square error: $$\text{RMSE} = \sqrt{\frac{1}{T}\mathrm{e}^2_{t+h,t}} \qquad (01)$$

Mean absolute error: $MAE = \frac{1}{T}\sum_{t=1}^{T}\left|e_{t+h,t}\right|$ (02)

Mean absolute percent error: $MAPE = \frac{1}{T}\left|\sum_{t=1}^{T} p^{2}{}_{t+h,t}\right|$ (03)

4. RESULTS AND DISCUSSION

At first, we have studied the correlation between these variables and the results are shown in following table:

Table 1
Correlation Matrix

Variables	GNP	ER	CPI
GDP	0.99255	0.96419	-0.00568
GNP		0.96581	0.017143
ER			0.021704

From the above table, we can see that there is strong correlation between the GDP, GNP and ER but there is weak correlation of all variables with CPI.

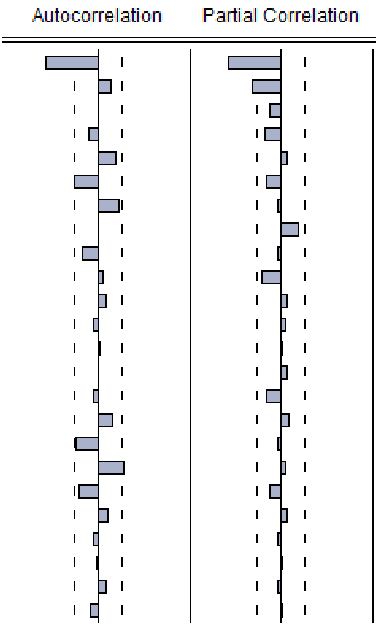
Box-Jenkins Methodology

For this we firstly apply unit root test to check the stationary of the data, the results of unit root tests are shown in the Table 2, we can conclude that the GDP, GNP and ER are stationary at second difference, CPI is stationary at 1st level. Now we will construct correlogram for these variables and identify the possible models. For modeling GDP, GNP and ER the term ARIMA (Auto regressive integrated moving average) and for CPI AR (Auto regressive) or MA (Moving average) terms will be used due to the stationary at 2nd difference and at level.

Table 2
P-Values of ADF Test

Variables	Level	1st difference	2nd difference
GDP	1.0000	0.9891	0.0000
GNP	1.0000	0.4931	0.0000
ER	1.000	0.9999	0.0000
CPI	0.0056		

Now, we will plot the correlogram of these variables. The correlogram of these variables are as follows.



	AC	PAC	Q-Stat	Prob
1	-0.596	-0.596	20.994	0.000
2	0.148	-0.322	22.308	0.000
3	0.005	-0.122	22.310	0.000
4	-0.100	-0.186	22.929	0.000
5	0.189	0.066	25.203	0.000
6	-0.262	-0.167	29.673	0.000
7	0.222	-0.031	32.950	0.000
8	0.004	0.190	32.951	0.000
9	-0.174	-0.026	35.034	0.000
10	0.062	-0.212	35.303	0.000
11	0.091	0.074	35.898	0.000
12	-0.065	0.050	36.211	0.000
13	0.014	0.010	36.226	0.001
14	-0.002	0.070	36.227	0.001
15	-0.065	-0.166	36.561	0.001
16	0.170	0.089	38.919	0.001
17	-0.261	-0.037	44.608	0.000
18	0.282	0.061	51.399	0.000
19	-0.210	-0.120	55.274	0.000
20	0.113	0.070	56.421	0.000
21	-0.049	-0.044	56.648	0.000
22	-0.022	0.021	56.695	0.000
23	0.089	-0.028	57.481	0.000
24	-0.088	0.009	58.259	0.000

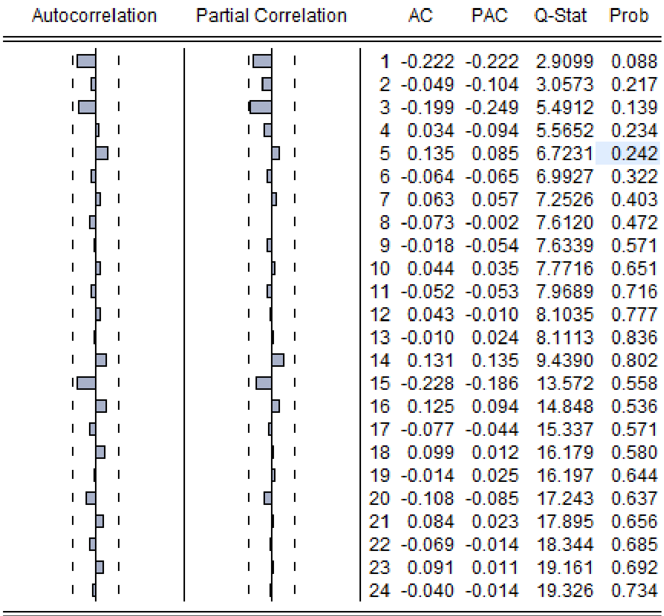
Figure 1: Correlogram of GDP

From Figure 1, the model for GDP identified are ARIMA (1,2,1), ARIMA (0,2,2) and ARIMA (17,2,0). Now will select the best model by comparing their AIC and SBC as given in the following table:

Table 3
AIC and SBC for GDP

Model	AIC	SBC
ARIMA (1,2,1)	49.34136	49.48346
ARIMA (0,2,2)	49.31952	49.42609
ARIMA (17,2,0)	49.68989	49.73140

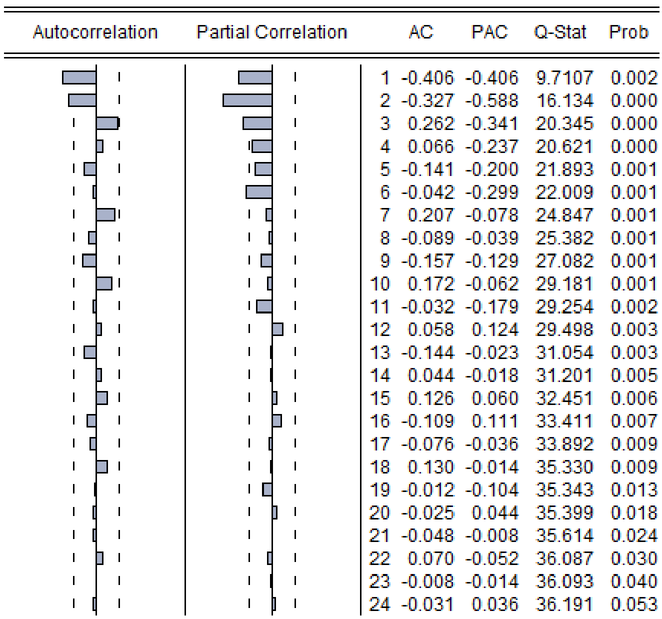
By comparing the AIC and SBC our best model is ARIMA (0,2,2) and our fitted model will be $GDP = 3.9\times10^{8} - 0.675274GDP_{(t)} + 1.36\times10^{20}GDP_{(t-2)}$. Now, we will plot the correlogram for GNP.



Lag	AC	PAC	Q-Stat	Prob
1	-0.222	-0.222	2.9099	0.088
2	-0.049	-0.104	3.0573	0.217
3	-0.199	-0.249	5.4912	0.139
4	0.034	-0.094	5.5652	0.234
5	0.135	0.085	6.7231	0.242
6	-0.064	-0.065	6.9927	0.322
7	0.063	0.057	7.2526	0.403
8	-0.073	-0.002	7.6120	0.472
9	-0.018	-0.054	7.6339	0.571
10	0.044	0.035	7.7716	0.651
11	-0.052	-0.053	7.9689	0.716
12	0.043	-0.010	8.1035	0.777
13	-0.010	0.024	8.1113	0.836
14	0.131	0.135	9.4390	0.802
15	-0.228	-0.186	13.572	0.558
16	0.125	0.094	14.848	0.536
17	-0.077	-0.044	15.337	0.571
18	0.099	0.012	16.179	0.580
19	-0.014	0.025	16.197	0.644
20	-0.108	-0.085	17.243	0.637
21	0.084	0.023	17.895	0.656
22	-0.069	-0.014	18.344	0.685
23	0.091	0.011	19.161	0.692
24	-0.040	-0.014	19.326	0.734

Figure 2: Correlogram of GNP

From the Figure 2, our model for GNP is ARIMA (0,2,3) and our model is $GNP = 1.42 - 0.81 GNP_{(t)} + 1738.42 GNP_{(t-3)}$. Now, we will plot correlogram for ER.



Lag	AC	PAC	Q-Stat	Prob
1	-0.406	-0.406	9.7107	0.002
2	-0.327	-0.588	16.134	0.000
3	0.262	-0.341	20.345	0.000
4	0.066	-0.237	20.621	0.000
5	-0.141	-0.200	21.893	0.001
6	-0.042	-0.299	22.009	0.001
7	0.207	-0.078	24.847	0.001
8	-0.089	-0.039	25.382	0.001
9	-0.157	-0.129	27.082	0.001
10	0.172	-0.062	29.181	0.001
11	-0.032	-0.179	29.254	0.002
12	0.058	0.124	29.498	0.003
13	-0.144	-0.023	31.054	0.003
14	0.044	-0.018	31.201	0.005
15	0.126	0.060	32.451	0.006
16	-0.109	0.111	33.411	0.007
17	-0.076	-0.036	33.892	0.009
18	0.130	-0.014	35.330	0.009
19	-0.012	-0.104	35.343	0.013
20	-0.025	0.044	35.399	0.018
21	-0.048	-0.008	35.614	0.024
22	0.070	-0.052	36.087	0.030
23	-0.008	-0.014	36.093	0.040
24	-0.031	0.036	36.191	0.053

Figure 3: Correlogram of ER

From Figure 3, the model for ER identified are ARIMA (1,2,1), ARIMA (2,2,2), ARIMA (0,2,3) and ARIMA (0,2,6). Now will select the best model by comparing their AIC and SBC as given in the following table:

Table 4
AIC and SBC for ER

Model	AIC	SBC
ARIMA (1,2,1)	5.744246	5.886345
ARIMA (2,2,2)	5.559127	5.701227
ARIMA (0,2,3)	5.992533	6.0099108
ARIMA (0,2,6)	5.999078	6.105653

By comparing the AIC and SBC our best model is ARIMA (2,2,2) and our fitted model will be $ER = 0.1876 - 0.336068ER_{(t-2)} - 0.462753ER_{(t-2)}$.

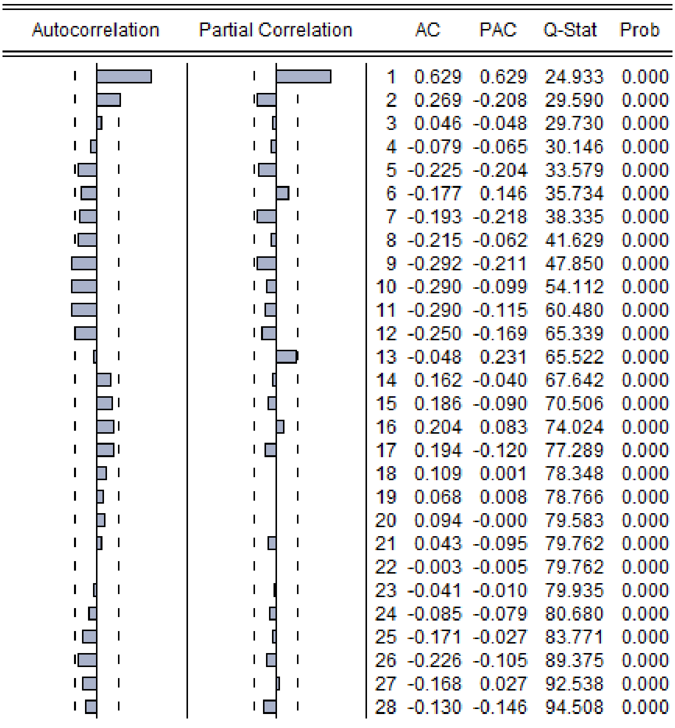


Figure 4: Correlogram of CPI

From Figure 4, the model for CPI identified are ARMA (1,1), ARMA (2,2), AR (5), AR (9), AR 10), AR (11), and AR (12). Now will select the best model by comparing their AIC and SBC as given in the following table:

Table 5
AIC and SBC for CPI

Model	AIC	SBC
ARMA (1,1)	5.668229	5.807852
ARMA (2,2)	6.139108	6.278731
AR (5)	6.145991	6.250708
AR (9)	6.100604	6.205322
AR (10)	6.102465	6.207182
AR (11)	6.103001	6.20778
AR (12)	6.117278	6.221995

By comparing the AIC and SBC our best model is ARMA (1,1) and our fitted model will be $CPI = 8.0126 + 0.510736 CPI_{t-1} + 0.253685 CPI_{t-1}$.

Vector Auto Regression Method:

It is a multivariate method used to study the relationship between variables which changes over time. This method has been applied by using EVIEWS and results are as follows.

$$\begin{aligned} CPI = {} & 0.7443 * CPI(-1) - 0.2045 * CPI(-2) + 0.14887 * ER(-1) \\ & - 0.1688 * ER(-2) + 0.0072 * GNP(-1) - 0.0322 * GNP(-2) \\ & - 8.51\times10^{-11} * GDP(-1) - 2.521\times10^{-11} * GDP(-2), + 3.9514 \end{aligned}$$

$$\begin{aligned} ER = {} & 0.0661 * CPI(-1) - 0.1281 * CPI(-2) - 0.1467 * ER(-1) \\ & - 0.5815 * ER(-2) + 0.0607 * GNP(-1) + 0.0217 * GNP(-2) \\ & - 8.69\times10^{-11} * GDP(-1) - 9.77\times10^{-11} * GDP(-2) + 0.9441 \end{aligned}$$

$$\begin{aligned} GNP = {} & - 0.1384 * CPI(-1) - 0.0964 * CPI(-2) - 6.769 * ER(-1) \\ & - 2.0786 * ER(-2) - 0.1395 * GNP(-1) - 0.7760 * GNP(-2) \\ & - 3.252\times10^{-10} * GDP(-1) + 1.9131\times10^{-10} * GDP(-2) + 5.1975 \end{aligned}$$

$$\begin{aligned} GDP = {} & - 139260491.275 * CPI(-1) + 134243457.935 * CPI(-2) \\ & - 2140638274.3 * ER(-1) - 484262976.492 * ER(-2) \\ & - 118570215.573 * GNP(-1) - 161330575.206 * GNP(-2) \\ & - 0.5727 * GDP(-1) + 0.0632 * GDP(-2) + 955444207.683 \end{aligned}$$

Table 6
AIC and SBC for VAR Model of GDP, GNP, ER and CPI

VAR Model	AIC	SBC	R square	Adjusted R^2
GDP	48.3449	48.6705	0.7978	0.7978
GNP	9.4148	9.7404	0.7673	0.7277
CPI	5.7897	6.1152	0.4508	0.3573
ER	5.5341	5.8596	0.5125	0.4295

Now, we will compare the performance of both methods. For this mean absolute error, root means square error, mean absolute percent error, and Theil coefficients has been used and their results for Box-Jenkins methodology and for vector auto regression are presented in Table 7.

Table 7
Forecast Evaluations of VAR & ARIMA

Model	Variables	MAE	RMSE	MAPE	Theil
VAR	GDP	4.47×10^9	6.49×10^9	7.7319	0.0225
	GNP	17.5499	22.8234	3.7388	0.0163
	ER	2.0983	3.2785	7.9688	0.0261
	CPI	2.5387	3.7255	30.7790	0.1948
ARIMA	GDP	7.01×10^9	1.17×10^{10}	10.1334	1.0721
	GNP	21.6045	41.7508	4.0216	0.6441
	ER	2.3353	3.6619	7.2628	0.9711
	CPI	2.4241	3.7743	100.272	0.3891

5. CONCLUSION

In this paper, we have studied the problem of multicollinearity among time series variables. Some univariate and multivariate methodologies have been used to study the problem. We have used Box-Jenkins Methodology as univariate and Vector-auto regression as multivariate technique. Economic variables such as gross domestic product, gross national product, exchange rate and consumer price index has been used for this study. Time series data of these variables has been obtained from the world bank website from 1962-2021. On studying the correlation between these variables, GDP, GNP and ER are highly correlated while CPI has weak correlation with all these variables. Both methodologies have been applied and best model is obtained for both. On comparing the forecast evaluation methodologies Vector auto regression method has performed well as its performance metrics, RMSE, MAPE, MAE and Theil coefficient are least as compared to Box-Jenkins methodology. So, in the presence of multicollinearity we can use var method to forecast these economic variables. We have used only few variables others can use different highly correlated variables for further studies and other methodologies.

REFERENCES

1. Bagshaw, M.L. (1986). *Comparison of univariate ARIMA, multivariate ARIMA and vector autoregression forecasting*. No. 86-02.
2. Katris, C. (2021). Unemployment and Covid-19 impact in Greece: A vector autoregression (VAR) data analysis. *Engineering Proceedings*, 5(1), 41.
3. Khan, M. and Khan, U. (2020). Comparison of forecasting performance with VAR vs. ARIMA models using economic variables of Bangladesh. *Asian Journal of Probability and Statistics*, 10(2), 33-47.
4. Khan, T. (2011). Identifying an appropriate forecasting model for forecasting total import of Bangladesh. *Statistics in Transition New Series*, 12(1), 179-192.
5. Kwasi, B.R. and Sharma, A. (2015). Comparing the forecasting power of multivariate VAR and univariate ARIMA models: A case of groundnut prices in Bikaner district of Rajasthan. *Asian Journal of Research in Business Economics and Management*, 5(3), 51-64.
6. Mohanapriya, M. and Ganapati, P.S. (2021). Forecasting Futures Trading Volume for Cotton using Vector Auto Regression and ARIMA Model. *Madras Agricultural Journal*, 108(special), 1. https://doi.org/10.29321/MAJ.10.000550
7. Parvin, S. and Khanam, M. (2018). Comparison between ARIMA and VAR Model Regarding the Forecasting of the Price of Jute Goods in Bangladesh. *Dhaka University Journal of Science*, 66(2), 91-94.
8. Tang, Z., De Almeida, C. and Fishwick, P.A. (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation*, 57(5), 303-310.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 195-204

A COMPARATIVE STUDY OF DIABETES CAUSES IN DIFFERENT COUNTRIES

Munazza Rafaqat and **Mehwish Asgher**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan

ABSTRACT

Background: Diabetes is a significant health problem in many countries. Absolute number of people with diabetes varied greatly from country to country. It is predicted that number of people with diabetes worldwide will be double in the next generation. This paper use the data about Diabetes causes from five countries which is collected recently. The aim of this review was to provide an overview of diabetes management of selected five countries and predict those diabetes causes which are at peak.

Method: We search many websites of health department and ageing, Agency of health care, research, quality and national institute of diabetes. We collect data relevant to some specific indicators of diabetes. GLM, Logistic regression and Hierarchical clustering techniques to compare the quality of diabetes management among indigenous and general population in selected countries.

Results: Data drawn from many websites, reports and many associated data sources**.** Countries had national data available including many indicators, in this review countries compare with each other to check the diabetes level which country at the peak. And which causes of diabetes most common.

Conclusion: substantial effort have been made in the four countries to collect the routine data in four countries to monitor performance of diabetes management. These countries can be learn from each other to improve their health care departments.

KEYWORDS

Diabetes, countries comparison, logistic regression, GLM, hierarchical regression.

1. INTRODUCTION

Diabetes is one of the most common human disease and become a significant public health concern worldwide. There were approximately 450 million people diagnosed with diabetes result in around 1.37 million deaths globally in 2017 (Joshi & Dhakal, 2021). In the 21st century, the global burden of diseases has shifted from communicable to non-communicable diseases, diabetes mellitus is one of the challenging global health issues of the 21st century. Presently, 500 million are living with diabetes mellitus worldwide (Meo et al., 2022). Diabetes mellitus, a long term and non-communicable disease is a major healthcare issue because it has touched alarming levels (Ganie & Malik, 2022). To better understand diabetes, we need to understand the process of metabolism (Factors, 2020). Our body is made up of millions of cells, when we eat food it contains a large amount of glucose which is oxidized to produce energy to these body cells. This produces glucose is then used for daily physical activities. Diabetes mellitus is a significant health problem in many countries, including India, Iraq, united state etc. an individual at high risk factors associated with it. Given the high prevalence and severity of diabetes, as it could be due to a combination of several reasons. Determining the risk factors and early prediction of diabetes have vital in reducing diabetes complications. It is predicted that the number of people worldwide with diabetes will double in the next generation. Due to rising living standard, diabetes has become more prevalence in people’s everyday lives. The only method of preventing diabetes complications is to identify and treat the disease early (Collins et al., 2011). The early detection of diabetes is important because its complications increase over time (WHO. Country and regional data on diabetes).

The worldwide obesity prevalence has increased to double since 1980 and tripled in developing countries last 20 years, a significant increase in obesity among adult’s females with prevalence reaches up to 55% several studies were conducted in Iraq recently to estimate the prevalence of DM among population in general and women particularly. Developed logistic regression based prediction models for type 2 diabetes occurrence these models are also screen individuals who are at a high risk of having diabetes. In contrast several others studies have shown that logistic regression perform at least as well as machine learning techniques for disease risk prediction. Risk factors for one ethnic group may or may not be generalized to others, for example: the prevalence of diabetes is reported to be higher among the pima Indian community. More than 100 million US adults with diabetes was the seventh leading cause of death in the US in 2020. In addition, there are substantial economic costs associated with the disease. The total estimated price of diagnosed diabetes in the US increased to USD 237 billion in 2017 from USD 188 billion in 2012. The excess medical cost per person USD 9601 from USD 8417 during the same period (Si et al., 2010) and (Laila et al., 2022).

2. DATA AND SUMMARY STATISTICS

There are four different country’s datasets which include several factors might be related to diabetes, including pregnancies, Glucose, Blood pressure, skin thickness, insulin, BMI, diabetes pedigree function, age etc. there are 394 diabetes patient in our dataset of India. As a component of diabetes management, it would be helpful to know which variables are related to diabetes. Several factors component of diabetes management several factors might be related to the United States diabetes patient are age, gender, sudden weight loss, weakness, genital thrush, visual blurring, itching, irritability delayed healing, obesity etc., there are 520 diabetes patient in our dataset of US. And the factors include in the dataset of Iraq that is age, sex, smoke, height, weight, active, lifestyle, bpsys, bpdias, ecgpatt, target, there are 333 diabetes patient include in our dataset of Iraq. The factors of France dataset are smoker, arterial hypertension, diabetes, dyslipidemia, immunossupressants, clinical activity, sex-female-male, arterial event. There are 90 patients of diabetes in our dataset of France.

Table 1
Descriptive statistics.

Countries	Factors	Highest Mean	Factors	St.dv
US	Age	48.03	Age	12.151
Iraq	Height	162.10	Bpsys	21.342
France	Sex	0.59	Smoker	0.497
India	insulin	155.32	Insulin	118.987

Table 1 explain that the factor age of US country's dataset has highest mean and standard deviation, the factors height from Iraq data has highest mean and bp has the highest standard deviation, sex factor has highest mean and smoker has highest standard deviation from France dataset, insulin has highest mean and highest standard deviation from India.

3. RESEARCH METHODOLOGY

Logistic regression model a relationship between the categorical response variable and covariance. Specifically, there is a linear log-odds of the probability of an event in a logistic model. Binary logistic regression estimate the likelihood that a characteristic of a binary variable is present, given the values of the covariates. Suppose Y is a binary response variable where if the character is present and $Yi = 1$ if the character is absent and the data $[Y_1, Y_2, \dots\dots Y_n]$ are independent. In order to investigate differences between countries, seconds and third model was stratified by type of data source and adjusted for country of origin. We used the largest sub- groups as the comparison groups. Data from each source were included in each analysis where information was available for more than 100 people in each group to reduce variability due to small numbers.

Estimation and Likelihood Ratio Test

Maximum likelihood is the preferred method to estimate ß since it include the good Statistical properties, we can use the least square approach. The logistic model with the single predictor variable x given by the logistic function of

$$\pi(X) = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}$$

Thus the likelihood ratio test tests whether this ratio is likelihood-ratio tests whether this ratio is significantly different from one, or equivalently whether its natural logarithm is significantly different from zero. We want to estimates the value of ß into the model $\pi(x)$ gives a number near to the all subjects who has diabetes and close to the zero. The selection to estimate ß for maximize likelihood function, we apply log on both sides to calculate the use of log –likelihood function for estimate our purpose. If subset of estimate β is zero we use likelihood ratio, suppose that p, r represent ß's full model and reduce model. The test statistics of likelihood ratio is as follow.

$$\Delta^* = -2\left[l(\widehat{ß}^{(0)}) - l(\hat{\beta})\right]$$

The full model log-likelihoods are $l(\hat{\beta})$ and $l(\widehat{ß}^{(0)})$ and the reduce model.

Model Selection Criteria

We used Alaike's information criteria (AIC), Schwarz's Bayesian Information Criteria (BIC) adjusted R^2 and PRESS to select the best predictive model. Information criteria are procedures that attempt to choose the model with the lowest sum of squared errors (SSE), with penalties for including too many parameters. AIC estimates the relative distance between the true and fitted likelihood functions of the data and model plus a constant. The AIC criteria are to choose the model which yields the smallest value of AIC, as defined by

$$AICp = n\,log(SSE) - n\,log(n) + 2p,$$

where n, and p number of observations and the number of parameters, respectively. The BIC gives a function of the posterior probability of a true model under a certain Bayesian setup. The BIC criteria are to choose the model which yields the smallest value of BIC. We define BIC as

$$BICp = n\,log(SSE) - n\,log(n) + p\,log(n)$$

where parameters are as defined earlier. Note that BIC incorporates a higher penalty for a higher n, and so it rewards more parsimonious models.

4. DISCUSSION

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood glucose. Hyperglycaemia, also called raised blood glucose or raised blood sugar, is a common effect of uncontrolled diabetes and over time leads to serious damage to many of the body's system, especially the nerves and blood vessels. In 2014, 8.5% of adults aged 18 years and older had diabetes. In 2019, diabetes was the direct cause of 1.5 million deaths and 48% of all adults due to diabetes occurred before the age of 70 years. Another 460,000 kidney disease deaths were caused by diabetes. Between 2000 and 2019, there was a 3% increase in age- standardized mortality rates from diabetes. In lower-middle-income countries, the mortality rate due to diabetes increased 13%. By contrast, the probability of dying from any one of the four main non communicable disease between the age of 30 and 70 decreased by 22% globally between 2000 and 2019. Risk factor for diabetes depend on the type of diabetes. Family history may play a part in all types. Environment factors and geography can add to the risk of type 1 diabetes.

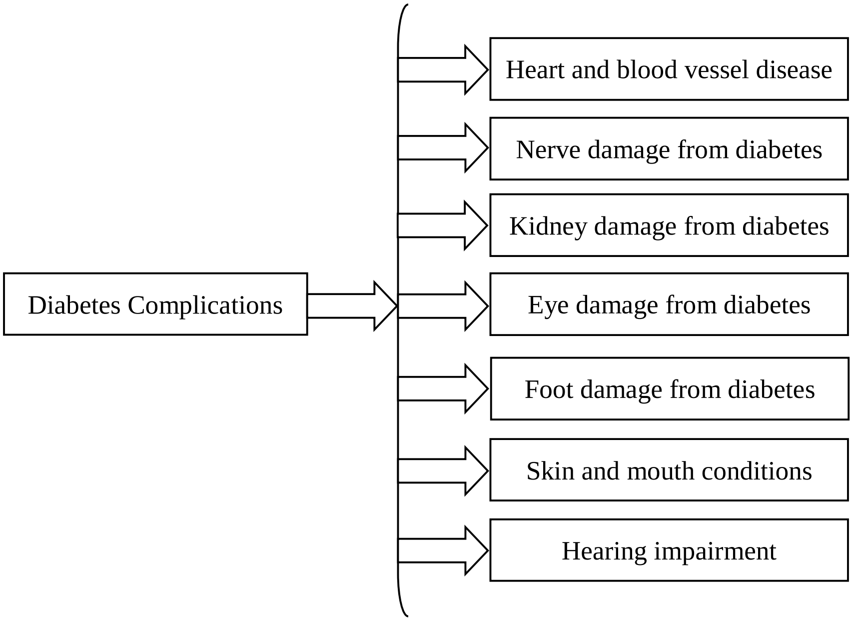


Table 2
Model Summary

Step	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Square
1	549.329[a]	.230	.314

Table 3
Hosmer and Lemeshow Test

Countries	Chi-square	Df	Sig.
US	9.005	8	.342
India	3.960	8	.861
Iraq	19.234	8	.014
France	5.174	8	.739

The value of model chi-square is 9.005 for us, 3.960 for India, 5.174 for France, and the 19.234 for Iraq this is highly significant. From the predict of datasets we say that the variables which including in these datasets are significant. Smoker, hypertension, sex, clinical activity, sudden weight loss, genital thrush, weakness, itching, obesity, smoke, Height, weight, age, etc.

Table 4
Contingency Table for Hosmer and Lemeshow Test (US)

	Gender = No		Gender = Yes		Total
	Observed	Expected	Observed	Expected	
Step 1	40	40.937	12	11.063	52
	34	34.752	17	16.248	51
	30	29.629	20	20.371	50
	25	25.445	27	26.555	52
	23	19.002	31	34.998	54
	16	14.590	36	37.410	52
	13	10.575	39	41.425	52
	2	8.164	50	43.836	52
	5	6.455	49	47.545	54
	5	3.450	45	46.550	50

Table 5
Contingency Table for Hosmer and Lemeshow Test (France)

	Diabetes = 0		Diabetes = 1		Total
	Observed	Expected	Observed	Expected	
Step 1	9	8.963	0	.037	9
	11	10.905	0	.095	11
	9	8.865	0	.135	9
	8	8.756	1	.244	9
	9	8.645	0	.355	9
	8	7.622	0	.378	8
	9	9.367	1	.633	10
	8	8.114	1	.886	9
	7	5.747	1	2.253	8
	3	4.016	5	3.984	8

Table 6
Contingency Table for Hosmer and Lemeshow Test (Iraq)

	Sex = 0		Sex = 1		Total
	Observed	Expected	Observed	Expected	
Step 1	33	32.935	0	.065	33
	33	32.440	0	.560	33
	32	31.114	1	1.886	33
	28	26.446	5	6.554	33
	16	17.385	17	15.615	33
	6	9.672	27	23.328	33
	2	3.308	31	29.692	33
	3	1.254	30	31.746	33
	1	.370	32	32.630	33
	1	.076	35	35.924	36

Table 7
Contingency Table for Hosmer and Lemeshow Test (India)

	Diabetes = -1		Diabetes = 1		Total
	Observed	Expected	Observed	Expected	
Step 1	38	37.908	1	1.092	39
	38	36.780	1	2.220	39
	36	35.475	3	3.525	39
	36	33.777	3	5.223	39
	30	31.510	9	7.490	39
	28	28.360	11	10.640	39
	22	24.644	17	14.356	39
	17	18.676	22	20.324	39
	13	10.818	26	28.182	39
	5	5.054	38	37.946	43

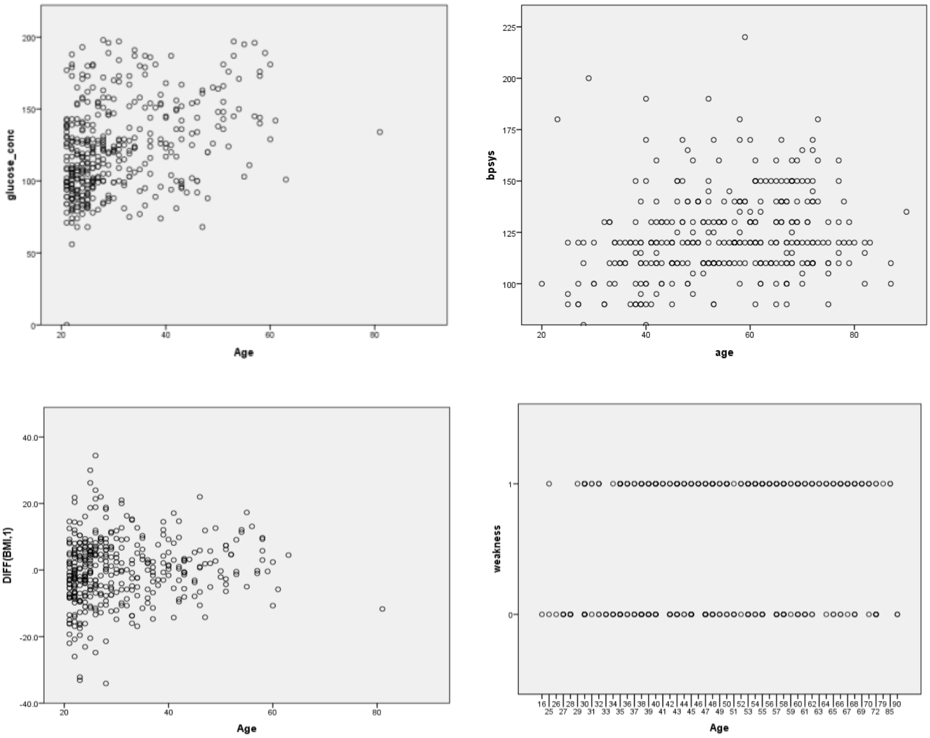


Figure 1: Correlation Plot of the Variables used in the Study
Note: BMI, Age, Time-Pregnant, BP, Glucose

5. RESULTS

Figure 1 is the correlation plot which give the strength of correlation between different pairs of the predictors. The pairwise correlation between time- weakness and age (0.680), p-value = 0.000 the result is significant therefore we can say that 95% level of confidence that there is 68% correlation between the predictors. And the correlation between the age and BMI is (0.106), and between age and blood pressure is (0.203), and between glucose and age (0.348) are ($R > 0.5$) high compared to other pairs, including these four pairs of predictors are significantly correlated.

Table 8
Internet Sources for Identification of Diabetes Related to Selected Countries

Countries	**Internet address**
US	
Department of health and human services	http://www.hhs.gov
Centres for diabetes control and prevention	http://www.cdc.gov/diabetes/
American diabetes association (ADA)	http://www.diabetes.org
Kaggle	http://www.kaggle.com
INDIA	
Kaggle	http://www.kaggle.com
Dataverse	http://www.dataverse.harvard.edu
ResearchGate	http://www.researchgate.net/figure/ pima-indian-diabetes-
Diabetes data and statistics	http://www.cdc.gov/diabetes/data/index.html
IRAQ	
Data world bank organization	http://data.worldbank.org/indicator
IDF diabetes atlas	http://diabetesatlas.org
Iraq-humanitarian data exchange	http://www.data.humdata.org
Kaggle	http://www.kaggle.com
International diabetes federation	http://www.idf.org
Data word	http://www.data.world
FRANCE	
The GHS program	http://dhsprogram.com
Kaggle	http://www.kaggle.com
Data hum data organization	http://www.humdata.org
Archive. ics.uci.education	http://www.archive.ics.uci.edu

REFERENCES

1. Collins, G.S., Mallett, S., Omar, O. and Yu, L.M. (2011). Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC medicine*, 9(1), 1-14.
2. Factor. R. (2020). *Diabetes Mellitus*.
3. Ganie, S.M. and Malik, M.B. (2022). Comparative analysis of various supervised machine learning algorithms for the early prediction of type-II diabetes mellitus. *International Journal of Medical Engineering and Informatics*, 14(6), 473-483.
4. Joshi, R.D. and Dhakal, C.K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health*, 18(14), 7346.
5. Laila, U.E., Mahboob, K., Khan, A.W., Khan, F. and Taekeun, W. (2022). An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. *Sensors*, 22(14), 5247.
6. Meo, S.A., Halepoto, D.M., Meo, A.S. and Klonoff, D.C. (2022). Impact of green space environment on the prevalence of diabetes mellitus in European countries. *Journal of King Saud University-Science*, 34(7), 102269.
7. Si, D., Bailie, R., Wang, Z. and Weeramanthri, T. (2010). Research article Comparison of diabetes management in five countries for general and indigenous populations: an internet-based review. *BMC Health Serv. Res*., doi: 10.1186/1472-6963-10-169.
8. WHO (n.d.). *Country and regional data on diabetes*.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 205-214

FORECASTING OF SUGARCANE PRODUCTION IN PUNJAB PAKISTAN USING ARIMA MODEL

Zakia Kousar[1] and **Hafiza Mamona Nazir**[2]
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]zakia.kousar55@gmail.com
[2]mamona.nazir@gcwus.edu.pk

ABSTRACT

Pakistan is an agricultural country and agriculture is very crucial for Pakistan's economy and employment. Sugarcane is a cash crop in Pakistan and Pakistan is a major contributor to World's sugarcane production. In Pakistan, Punjab province is the major producing sugarcane province. The objective of this study is to look at the stability of the production of sugarcane in Punjab. For this study, a secondary data set of the time period 1991-92 to 2021-22 has been obtained from Crop Reporting Service Punjab. ARIMA model has been used for forecasting. The ARIMA (10,1,2) has been considered the best model for this data set and it has been used for forecasting the sugarcane production for the time period 2022-23 to 2029-30. The results have shown that there is an increasing trend in sugarcane production in Punjab in the upcoming years. All this graphical and numerical work is performed by using R software.

KEYWORDS

Sugarcane, Stationarity, ARIMA, Forecast, AIC.

INTRODUCTION

Sugarcane is one of the most important crops in Pakistan due to its diverse variety of uses. As it is a multipurpose crop, we can use it for making sugar, jaggery, molasses, Khansari, and even paper. It is used in providing raw materials to the sugarcane industry and to the chemical and paper industries. It also reduces energy crises. It is one of the most important Kharif crops. It has a 0.71% share in gross domestic production (GDP) in the year 2017-18.

Pakistan is a major sugarcane-producing country. It grows about 1 million hectares of sugarcane from which 95000 hectares of sugarcane are grown by Punjab. Pakistan is the world's 6th largest sugarcane-producing country. In Punjab, 60-65% area is used for sugarcane cultivation. The plantation of sugarcane requires a very appropriate climate. Punjab's climate supports the plantation of sugarcane throughout the year. It grows in a hot and humid climate with a temperature of 210C to 270C and annual rainfall between 75cm.

Food security has become a major problem nowadays. Agriculture is not only a source of employment but also provides raw materials to the industries which export various commodities. The basic purpose of this study is to forecast the production of sugarcane.

Forecasting plays a vital role in managing our resources according to our requirements. There is a close association between crop price and production. There exists an effective relationship between productivity and the price of sugarcane. Price and productivity are linked with each other. A glut in the production of sugarcane can lead to a slump in prices and has an adverse effect on farmers' incomes. That's why forecasting production is very valuable.

Most of the studies have forecasted the production of cash crops in different areas of Pakistan. Ahmad et al., (2017) reported major crop forecasting areas, production, and yield from the agriculture sector of Pakistan. Wheat, rice, cotton, and sugarcane are cash crops of Pakistan as they fulfill our nutrition needs as well as we export these crops. Hence, their exportation plays important role in our economy. They used Arima for forecasting area, production, and yield. Ali et al. (2015) discussed when a number of diagnostic tests on the fitted model were successfully completed, it was determined that the ARIMA model is appropriate for predicting the yields of Pakistan's two main crops, sugarcane and cotton. Mehmood et al. (2019) forecasted the production of sugarcane in Pakistan for the year 2018-2030. They showed that the production of sugarcane will increase in the upcoming years. Boken (2000) used different forecasting techniques but he showed that quadratic smoothing is the best forecasting technique. Yaseen et al. (2011) used the Arima model and showed that forecasted values are very close to the actual values. Ali et al. (2015) presented a detailed study for forecasting the production and yield of sugarcane and cotton. Sharma et al., (2021) applied a combination approach using a rough set approach for forecasting. They suggested Arima (2,1,1) as an appropriate model. Meyler et al., (1998) had drawn a framework for Arima in time series. Suresh and Krishna Priya (2011) used the Box-Jenkins approach for forecasting. They used two models for area and production. Paswan et al. (2022) compared Arima models with Artificial Neural Networks and checked the efficiency of both methods. Both methods provided accurate results. When they employed secondary data and the ARIMA forecasting technique to predict the main food crop output in Khyber Pakhtunkhwa, they found that the results of the ARIMA model were adequate. The Box-Jenkins ARIMA model in Tamil Nadu to anticipate sugarcane production. Sugarcane is used in the production of over 70% of the world's sugar. Due to variations in the area under sugarcane cultivation, sugarcane production varies from year to year. Hossain and Abdulla (2015) reported that since sugar is primarily produced from sugarcane in Bangladesh, it is crucial to estimate production in order to meet domestic demand. This is why we conducted this study. Hussain et al. (2006) Finding the Auto-Regressive Integrated Moving Average (ARIMA) model that might be utilized to predict Bangladesh's sugarcane production is the major goal of this study. This analysis used yearly secondary data that had been published.

In the case of commercial crops like sugarcane, production level affects raw material costs. Over the last few years, various techniques have been applied to single-time series models. Low productivity is a big problem for developing countries. As Pakistan is an agricultural country. Most of its population lives in villages and they depend on agriculture.

STATIONARY OR NON-STATIONARY

A time series is a collection of observations that are recorded over time. Time series is classified as stationary and non-stationary. A stationary time series is one whose properties (mean, variance) do not vary with time. It means that mean and variance are independent of time. A non-stationary time series is one whose properties (mean, variance) depend on time i.e. it is time dependent. The first thing we do in modeling time series data is to check the stationarity of our data. Stationarity can be checked by plotting data. The approach which can check stationarity only by visualization is called the subjective approach. The objective approach is one in which we apply various tests like the unit root test, Dickey-Fuller test and Augmented Dickey-Fuller test for checking the stationarity of our data.

If the data is not stationary, we transform data and make it stationary by applying suitable transformations to our data. The following are the transformations that can be applied to make it stationary:

1. Differencing
2. Log transformation
3. Box cox transformation
4. Power transformation
5. Square root transformation
6. These transformations can stabilize the mean and variance of the data set.

DATA & METHODOLOGY

A secondary data set of sugarcane production in Punjab has been obtained from Crop Reporting Service Punjab for the time period 1991-92 to 2021-22.

Box-Jenkins Modeling Procedure

ARIMA methodology is called the Box-Jenkins methodology. This methodology is concerned with fitting a mixed ARIMA model to a given set of data. The main objective of fitting an ARIMA model is to identify the stochastic process of time series and predict future values accurately. The following steps are performed in the Box-Jenkins methodology:

1. Plot the data. The Histogram has to be plotted as it is the graph of the time series dataset.
2. Apply transformation if needed. If the series is non-stationary, make it stationary by applying different transformations.
3. Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the selection of lag values.
4. Select the model according to lag values.
5. Estimate parameters of the selected model.
6. Forecast the future values by using a model.

The Autoregressive integrated moving average model (ARIMA (p, d, q)) is as follows:

$$\varphi p(B)\Delta d\ ht = c + \theta q(B)gt$$

To develop the ARIMA model, the series has to be stationary. If the series is non-stationary, we make it stationary by applying different transformations according to the nature of our data set. ARIMA model mainly consists of model identification, parameter estimation, testing, model validation, and forecasting model. Identification involves finding out the order of AR, MA, parameters, and differentiation value of the model. First, we check the stationarity of our data by using the ADF test and by visual anticipation. As our series is non-stationary, then we have taken its first difference and made it stationary.

Table 1
Summary Statistics of Production of Sugarcane (1991-92 to 2021-2022)

Parameter	Production ('000 tonns)
Mean	35920
Standard Deviation	10877.12
Kurtosis	0.302
Skewness	0.684
Range	44612
Minimum	19633 (1991-92)
Maximum	64245 (2021-22)

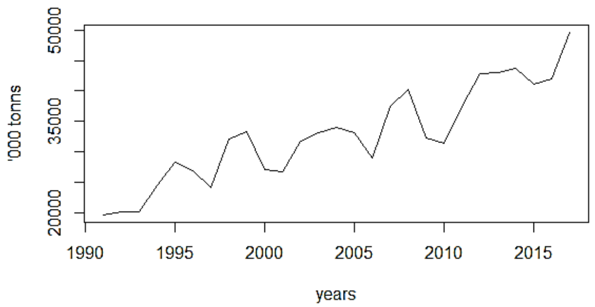


Figure 1: Sugarcane Production of Study Period

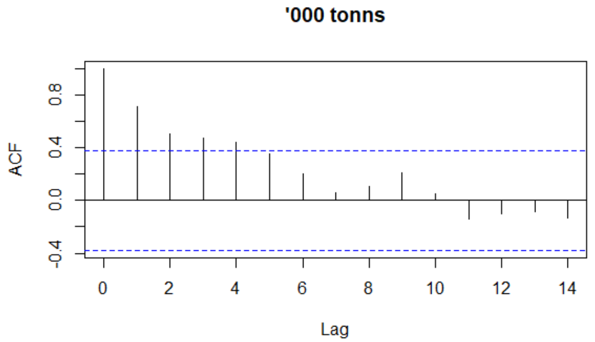


Figure 2: Sugarcane Production ACF Plot

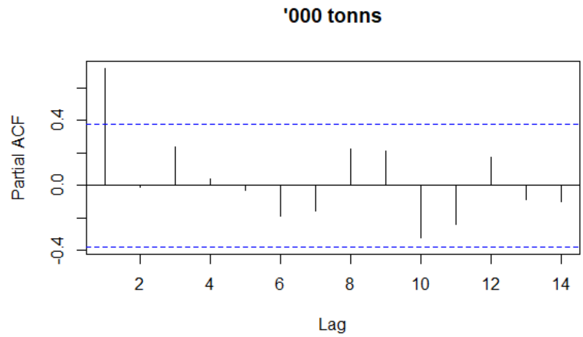


Figure 3: Sugarcane Production PACF Plot

The graphs of the sugarcane production historigram, Autocorrelation function (ACF) and Partial autocorrelation function (PACF) have been plotted (Figures 1, 2, 3). From Figure 1, we can see that there is a positive trend in the data set. It means that the data is non-stationary. The transformation applied here to make the series stationary is difference. By taking 1st difference, our series has become stationary. The ADF test has been used to check the stationarity of data.

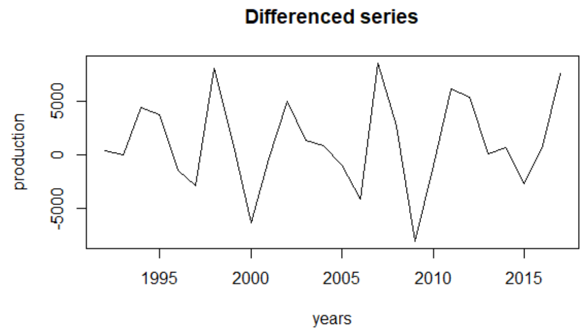


Figure 4: Time graph of 1st Difference in Production

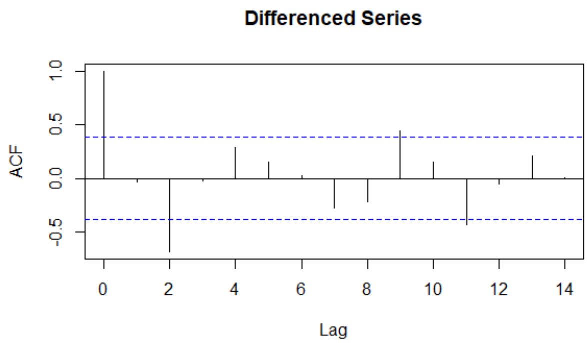


Figure 5: ACF Plot of 1st Difference in Production

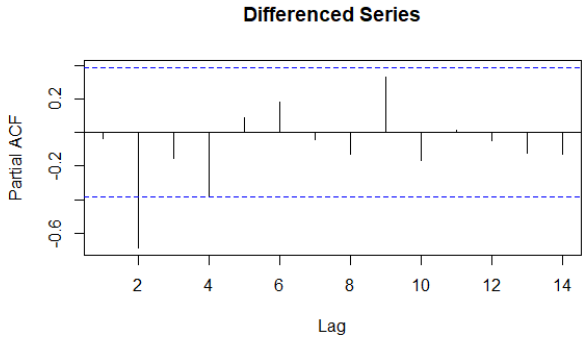


Figure 6: PACF Plot of 1st Difference in Production

The plots of 1st difference in production, Autocorrelation function (ACF), and Partial Autocorrelation function (PACF) have been plotted (Figure 4, 5, 6). From ACF and PACF, we can see moving average and autoregressive lag values. There are only 4 significant spikes in the ACF plot and only one significant spike in the PACF plot.

Table 2
Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value
-2.9626	2	0.205

Table 3
Augmented Dickey-Fuller Test

Dickey-Fuller	Lag order	p-value
-4.4079	2	0.01

RESULTS AND DISCUSSIONS

From the above results, Table 1 shows that the average production in Punjab is 35920 tons. The standard deviation is 10877.12, kurtosis is 0.302 and skewness is 0.684. The value of skewness shows that it is a positively skewed data set. The range of the data set is 44612. The minimum value is 19633 which is the production in the year 1991-92 and the maximum value is 64242 which is the production value in the year 2021-22. Table 2 shows the results of the Augmented Dickey-Fuller test. The value of the ADF test is 0.80 which is greater than 0.05. It means that we do not reject H0 and conclude that the series is non-stationary. After taking 1st difference of the series, the ADF test was again used and this time the value of the ADF test is 0.01 which is less than 0.05. So, the series is now stationary. ACF and PACF plots have been obtained for estimating the lag values of p and q in ARIMA. ARIMA (10,1,2) has considered being the best for forecasting future production in Punjab. The model provides the forecasted values which are close to the actual values. After estimating the model, diagnostic checking has been performed. From the ACF, PACF, and normal probability plot, it is clear that the residuals are normally distributed. The test for checking the normality of residuals is the Box L-Jung test is used and results are obtained which also showed that residuals are normally distributed.

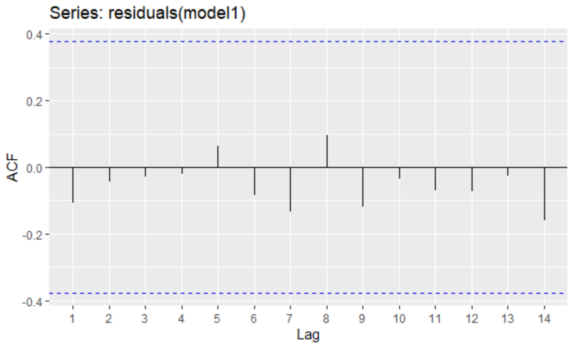


Figure 7: ACF Plot of Residuals

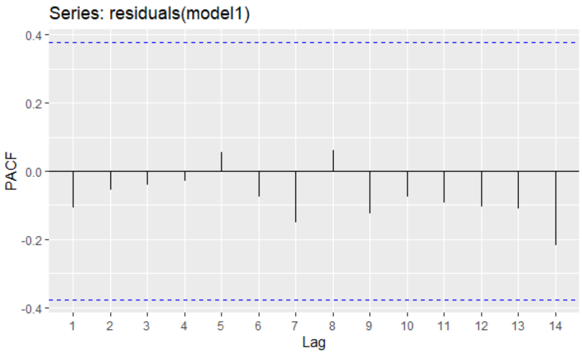


Figure 8: PACF Plot of Residuals

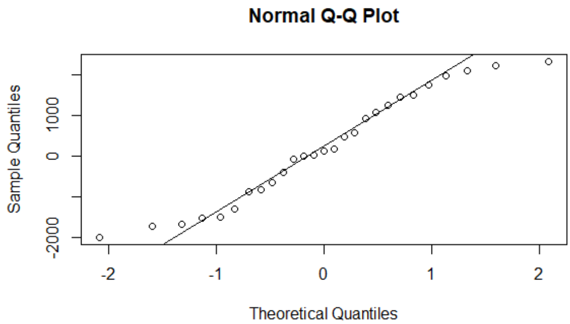


Figure 9: Normal Probability Plot

Table 3
AIC Values for Various Combinations of ARIMA

Arima Model	AIC
Arima (6, 1, 0)	499.96
Arima (9, 1, 0)	496.94
Arima (10, 1, 0)	497.58
Arima (9,1,1)	497.9
Arima (10, 1, 1)	499.44
Arima (10, 1, 2)	496.91

From Table 3, we can see that ARIMA (10,1,2) has the lowest value of AIC. So, the best model for our data set is ARIMA (10,1,2). AIC value is an error value and a model selection criteria. We select the model whose value of AIC is less than the others.

Testing for Residuals

Q*	DF	p-value
7.4178	3	0.05971

From the provided results, we can say that our p-value is 0.05971 so we reject H_0 and conclude that our residuals are normally distributed. It means that our selected model is appropriate for the data set.

As our data set consists of 32 observations, we have split the data set in the training and testing part. The time period 1990-91 to 2014-15 has been used as training and from 2015-16 to 2021-22 has been used as testing. The forecasted values by using ARIMA (10,1,2) are as follows:

Table 4
Forecasting of Data

Year	Forecast	Observed
2017	43385.02	41968
2018	43122.43	49613
2019	50840.27	55067
2020	53351.35	44906
2021	51648.88	57000
2022	49779.26	64245

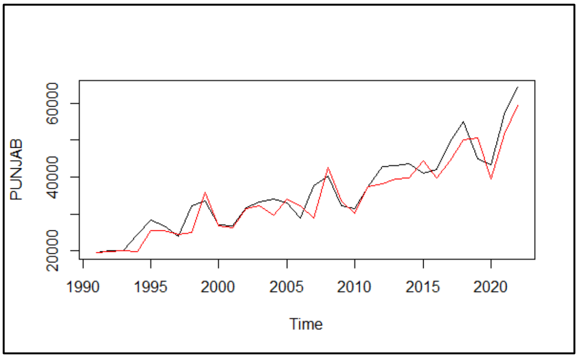


Figure 10: ARIMA for Observed and Predicted Sugarcane Production

Figure 10, shows that the observed values are very close to the predicted values. Figure shows that the production of sugarcane in Punjab will be increased in the upcoming years.

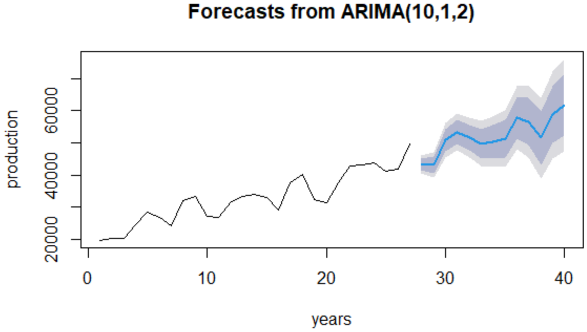


Figure 11: Forecasted Graph of Sugarcane Production

Year	Forecast	Lower Limit	Upper Limit
2023	49779.26	42833.72	56724.80
2024	50305.78	42477.68	58133.87
2025	51342.21	42445.99	60238.44
2026	57742.55	47917.17	67567.93
2027	56691.21	45501.71	67880.70
2028	51484.39	38829.34	64139.43
2029	58799.48	45233.44	72365.53
2030	61735.24	47294.16	76176.33

CONCLUSION

From various models, ARIMA (10,1,2) has been shown to be the best for modeling and forecasting. The forecasting values for 2015-16 to 2021-22 have shown that there is a significant increase in the production of sugarcane in Punjab which has been verified by the actual values. As Punjab is a major contributor to Pakistan's sugarcane production, Pakistan's production of sugarcane will also be high. The ARIMA (10,1,2) has been proven the best model for this data set as the forecasted values are very close to the actual ones. This proposed ARIMA model can be used by the Government to increase its resources for keeping stability or increasing the production of sugarcane for the coming years. It can also be useful for researchers, businesses, and farmers who want to make important decisions regarding the production of sugarcane.

REFERENCES

1. Ahmad, D., Chani, M.I. and Humayon, A. A. (2017). Major Crops Forecasting Area, Production and Yield Evidence from Agriculture Sector of Pakistan. *Sarhad Journal of Agriculture*, 33(3), 385-396.
2. Ali, S., Badar, N. and Fatima, H. (2015). Forecasting production and yield of sugar cane and cotton crops of Pakistan for 2013-2030. Sarhad Journal of Agriculture, 31(1), 1-10.
3. Boken, V.K. (2000). Forecasting spring wheat yield using time series analysis: A case study for the Canadian Prairies. *Agronomy Journal*, 92(6), 1047-1053.
4. Hossain, M.M. and Abdulla, F. (2015). Forecasting the Sugarcane Production in Bangladesh by ARIMA Model. *Journal of Statistics Applications & Probability an International Journal*, 4(2), 297.
5. Hussain, M.F., Anwar, S. and Hussain, Z. (2006). Economics of sugarcane production in Pakistan: A Price risk analysis. International Research Journal of Finance and Economics, 4(4), 33-39.
6. Mehmood, Q., Sial, M., Riaz, M. and Shaheen, N. (2019). Forecasting the production of sugarcane crop of Pakistan for the year 2018-2030, using Box-Jenkin's Methodology. *The J. Anim. Plant Sci*., 29(5), 1396-1401.
7. Meyler, A., Kenny, G. and Quinn, T. (1998). Forecasting Irish inflation using ARIMA models. Published in *Central Bank and Financial Services Authority of Ireland Technical Paper Series*, Vol. 1998, No. 3/RT/98, 1-48.
8. Paswan, S., Paul, A., Paul, A. and Noel, A.S. (2022). Time series prediction for sugarcane production in Bihar using ARIMA & ANN model. *The Pharma Innovation Journal*, 11(4), 1947-1956.
9. Sharma, H.K., Kumari, K. and Kar, S. (2021). Forecasting sugarcane yield of India based on rough set combination approach. *Decision Making: Applications in Management and Engineering*, 4(2), 163-177.
10. Suresh, K.K. and Krishna Priya, S.R. (2011). Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Tech.*, 13, 23-26.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 215-220

FORECASTING GDP OF PAKISTAN USING ARIMA MODEL

Alia Riaz[1], **Mamona Nazir**[2] and **Wajiha Nasir**[3]
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]aliariaz770@gmail.com
[2]mamona.nazir@gcwus.edu.pk
[3]wajiha.nasir@gcwus.edu.pk

ABSTRACT

Gross Domestic Product (GDP) is an important indicators of a country's economic growth. To assist in decision making process, this study intends to forecast the GDP for seven years and identify the factors affecting the GDP in context of Pakistan using the data from World Bank. The Autoregressive Integrated Moving Average (ARIMA) model is adopted to predict the GDP from 2013 to 2025 and multiple linear regression has been used to explore the factors affecting GDP. Different types of ARIMA (P, I, Q) tested and applied the ARIMA (0, 1, 0) model are found as best appropriate forecasting. Q-Q plot, residuals plot, PACF and ACF graphs of the residuals are drawn for checking model adequacy. The finding shows that the forecast values of Pakistan's GDP will be 23477 Billion rupees in 2013 and 103918 billion rupees in 2025.

1. INTRODUCTION

GDP is the total value of all products and services produced within a country's geographic boundaries during a certain period of time. For every economy, especially the emerging ones that frequently experience difficulties, sustainable economic growth is of utmost importance. The gross domestic product (GDP), which measures the market value of all finished products and services produced in a nation over a given period of time, is a key indicator of how well an economy is doing. Thus, economists have concentrated on examining the ways in which GDP may enhance economic growth. GDP can be described with three approaches; the income, the expenditure and the product approach. According to the rule of income approach, each manufacturer's income must match the value of their product, and the GDP is calculated by adding the incomes of all producer's. The expenditure approach rule is that every product must be purchase by someone (Kennedy, 2000).

GDP is intended to measure the market worth of production that flow throughout the economy, (a) Consist of only services and goods bought by their ultimate consumer, thus GDP dealswith final production. (b) Keep out transfer payments and financial transactions as they do not stand for present production. (c) Count up only the services and goods produced inside the country's borders for the duration of a year, whether through foreigners or citizens. (d) Measures both productivity and income, which are equivalent.

The economy of the Islamic Republic of Pakistan is the world's 27th largest economy found on its purchasing power. Pakistan economic development faced a serious hinder in fiscal year 2009 for the reason that the dejected consumer credit market, sluggish growth of public sector programs, inflation, cutback in subsidies, security menace, energy crisis and instability in the State. Furthermore, no concentration was given to the agriculture division. The exports and imports turn down by six percent and 10 percent respectively. The only credible thing was the increase in remittances by 22%. Agriculture division has revealed convincing outcomes for the reason of good weather. Major crops, rice, wheat and maize proof notable growth i.e., 7.7% against the goal of 4.5%. So, it is essential to forecast GDP for the prediction of the future economy as well as the advancement of a country based on the current data. On the other hand, a significant change in GDP, whether up or down, usually has an impact on the market economy. Therefore, it is necessary to realize the associated factors that affect GDP. There are many approaches that can be applied for macroeconomic estimating like linear regression, AR model, MA model, ARIMA model, VAR model, etc. The specific objectives to be investigated throughout this study are forecasting GDP of Pakistan and finding the significant factors affecting GDP in Pakistan.

2. LITERATURE REVIEW

GDP is one the main needle exercised to find out the health of a country's economy. GDP be a sign of the total dollar value of all services and goods produced over a particular time period, one can suppose of it as the size of the economy.Therefore, an extensive effort has been made to search the literature concerning with the time series modeling on GDP using ARIMA model. Lu (2009) attempt to construct a time series model which was utilized to forecast the gross domestic product to get its future estimations up to first quarter of 2009.The study found on the figures collected through secondary sources from 1962 to 2008.ARIMA models were seek on the collected data and to conclude ARIMA (4, 1, 0) is create to be an appropriate model, which is then apply for forecasting purpose. The outcome of the future forecast explains the significant improvement in the fourth quarter of 2009. Li, Liu and Zhao (2007) focused on Monetary policy, have an effect on the economy among long and variable lags, and meant for this explanation policy-makers have need of reliable forecasts of economic activity. Thus, forecasts of real GDP growth have turn into more and more needed. Projected a new modified ARIMA model and make use of it to forecast the GDP growth of China since 1978 to 2004. Within this paper, they propose a new genetic programming technique to forecast the GDP time series of United States, Japan and China from 1980 to 2006. In addition, they apply the proposed technique to forecast the prospect GDP growth of United States, Japan and China from 2007 to 2020, and they come across that the GDP of Japan vary periodically, though the GDP of United States and China raise steadily in the near future.

With the predicted data it is found that by 2020 the GDP of China will go above the GDP of Japan for the first time. The evaluation on the state of the Indian economy throughout recession by analyzing different macro-economic factors such as GDP, inflation, exchange rate, fiscal deficit and capital markets. This study forecast some of the major economic variables by ARIMA modeling and presents a depiction of the Indian economy in the coming years. The results indicate that Indian economy is stimulating after a slowdown in the phase of global recession. It is forecasted that GDP, fiscal deficit, capital markets and foreign investments will increase in 2010-11 (Sinha, 2010). The Study of

Mamun (2009) focused on ARIMA model to evaluate and forecast GDP; figures are extracted from Bangladesh Bureau of Statistics. This paper made an evaluation of two forecasting methods to observe which suits in forecasting the GDP and its divisions. ARIMA model plays quite an admirable part in uni-variate econometric time series data study of GDP of Bangladesh and its divisions. Ning (2010) made an effort to build a time series model which was then utilize to forecast the GDP of Shaanxi to achieve its future assessment up to the year 2013.This paper based on the figures collected from secondary sources as of the year 1952-2007. ARIMA Journal of Contemporary Issues in Business Research Volume 3, Issue No. 4, 2014 202 models were seeking on the collected figures and at last ARIMA (0,1,0) is found to be suitable model, which is then employ for forecasting purpose. The outcomes of the prospect forecasts showed the considerable progress in GDP of Shaanxi Province up to the year 2013.

3. RESEARCH METHODOLOGY

Forecast GDP of Pakistan. In this chapter ARIMA Model and diagnostic checked are described.

4. ANALYSIS AND RESULTS

Model Development

To stabilize the variability, the GDP is plotted for main datasets, log transformation and square root transformation of the data.

The plots are given in the following:

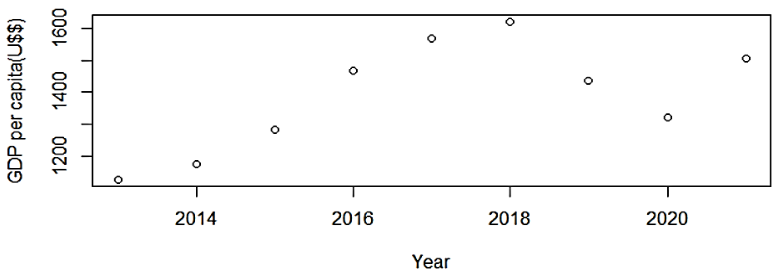


Figure 1: Actual GDP of Pakistan

Exploring d value: difference is taken to non-stationary data into a stationary data into a stationary time series. Here in this study, first, second and third differences are taken. From the Figure 2a, 2b and 2c, it can be seen that, the first and second graph looks more stationary than the third one. So, these plots suggest for $d = 1$ or $d = 2$.

Exploring p and q for $d = 1$: Here, to explore the value of p and q for $d = 1$, the example ACF and PACF of the first differenced data are plotted. It can be realized that the sample ACF and PACF do not cut off the blue boundary. So, the plots suggest, $p = 0$ and $q = 0$ for $d = 1$. So, the proposed model is ARIMA (0, 1, and 0).

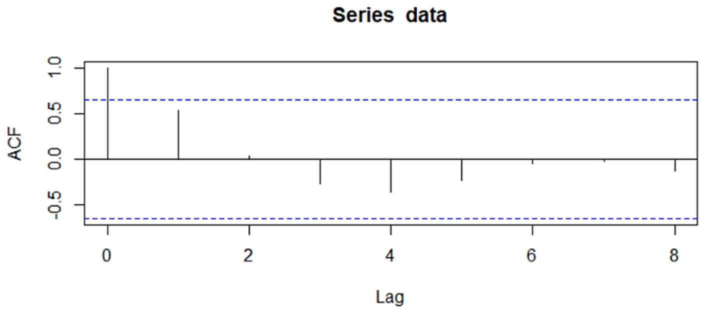


Figure 2: Exploring d Value

Model Adequacy Checking

For checking model adequacy, Q-Q plot, residuals plot, ACF and PACF graphs of the residuals are drawn.

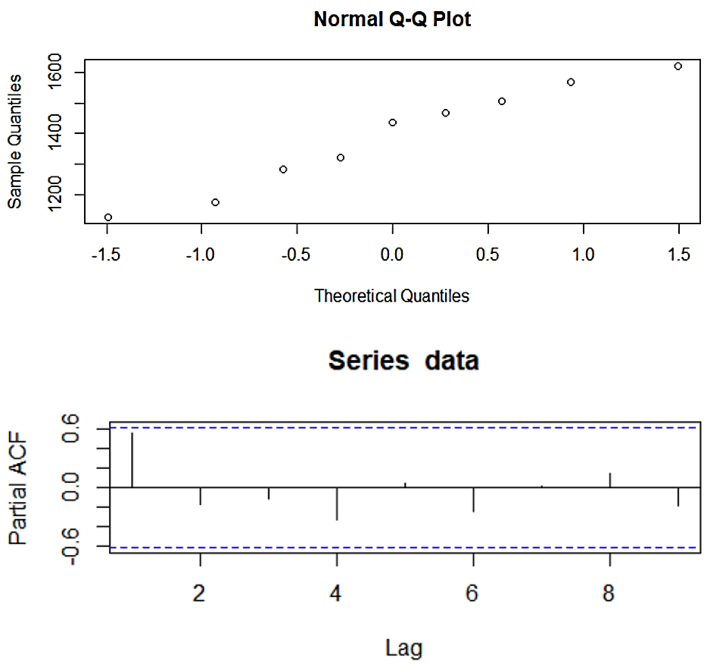


Figure 3: Q-Q, residual, ACF and PACF plots

The graph of the standardized residuals was also drawn to make sure the non-existence of the serial correlation among the residuals. The goodness of fit of the model was tested by constructing the QQ plot of the residuals. Approximately, 90% to 95% points are on the straight line which justifies the normality assumption of parsimonious model.

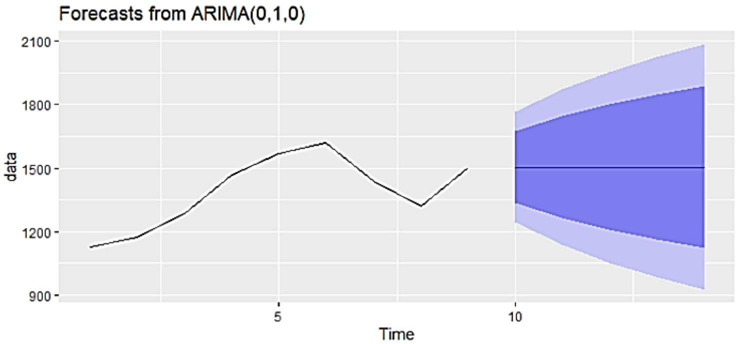


Figure 4: Forecasted Value of GDP

5. CONCLUSION

In this study, an ARIMA (0, 1, and 0) model has been fitted to forecast the Pakistan's annual GDP and its growth rates for a couple of years. The time plot of GDP data and ADF test connoted that GDP of Pakistan is stationary at I (2). The parameters of the ARIMA (0, 1, and 0) are estimated by MLE method and found to be significant. Diagnostics tests of residuals satisfied all the assumptions of parsimonious model i.e., ACF, PACF, QQ- plot and standardized residuals etc. The point base forecasts for the period 2013-25 anticipated a sustainable upward trend in Pakistan's GDP, whereas, a decreasing trend in point base growth rates respectively. It endorsed that precise forecasts offer useful information for devise sustainable development strategies which may helpful for both government and businessmen to ensure successful future policies.

REFERENCES

1. Box George, E.P., Jenkins Gwilym, M., Reinsel Gregory, C. and Ljung Greta, M. (1976). *Time series analysis: forecasting and control*, San Francisco: Holden Bay.
2. Dritsaki, C. (2015). Forecasting real GDP rate through econometric models: An empirical study from Greece. Journal of International Business and Economics, 3(1), 13-19.
3. Government of Pakistan (2004). *Economic Survey 2003-2004*. Finance Division, Economic Adviser's Wing, Islamabad.
4. Hamid, N.P., V. Thomas, Alberto and G. Suzanne (1987). The wheat economy of Pakistan setting and prospects. International Food Policy Research Institute, Ministry of Food and Agriculture, Government of Pakistan, Islamabad, Pakistan
5. Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T-C. (1988). *Introduction to the Theory and Practice of Econometrics*. 2nd ed., John Wiley and Son, INC. New York, Toronto, Singapore.
6. Muhammad, F., Siddique, M., Bashir, M. and Ahmad, S. (1992). Forecasting rice production in Pakistan using ARIMA models. J. Anim. Plant Sci, 2, 27-31.
7. Muhammad, K. (1989). Description of the historical background of wheat improvement in Baluchistan, Pakistan. *Agriculture Research Institute* (Sariab, Quetta, Baluchistan, Pakistan).

8. Pindyck, R.S. and Rubinfeld, D.L. (1991). Econometric Models and Economic Forecasts (3rd Edition), New York : McGraw-Hill
9. Qureshi, K.A., Buland Akhtar, M.A., Ullah, A. and Hussain, A. (1992). An Analysis of the relative contribution of area and yield to Total production of wheat and maize in Pakistan. *Pak. J. Agri. Sci*, 29(2), 166-9.
10. Rahman, N.M. and Baten, M. (2016). Forecasting area and production of black gram pulse in Bangladesh using ARIMA models. *Pakistan Journal of Agricultural Sciences*, 53(4) 759-765.
11. Sabur, S.A. and Haque, M.E. (1993). An analysis of rice price in Mymensing town market: Pattern and forecasting. *Bangladesh Journal of Agricultural Economics*, 16(2), 1-15.
12. Sohail, A., Sarwar, A. and Kamran, M. (1994). Forecasting total food grains in Pakistan. *J. Eng. Appl. Sci.*, Department of Mathematics and Statistics, University of Agriculture, Faisalabad, 13, 140-146.
13. World Bank. (2018). World Development Indicators.
14. Younis, M. (1995). Present Status and Future Prospects of Wheat in Pakistan–An Economic Review. *Agricultural Economic Section, Ayub Agricultural Research Institute*, Faisalabad. Research Study, (30).

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 221-228

FORECASTING OF BUFFALO POPULATION AND LIVESTOCK PRODUCTION USING BOX-JENKINS METHODOLOGY

Arooj Imtiaz[1], Mariam Shoukat, Anam Asmatullah,
Alina Ramzan[2] and **Hafiza Mamona Nazir[3]**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]alina.ramzan@gcwus.edu.pk
[2]aroojimtiaz0786@gmail.com
[3]mamona.nazir@gcwus.edu.pk

ABSTRACT

Buffalo is called the 'Black Gold' of Pakistan because of its importance in agriculture and milk production. In this study, taking a secondary data set of buffalo population form livestock and dairy development, Punjab. We attempt to build the Univariate time series model to forecast buffalo population and livestock production in Pakistan. The historical time series data of buffalo population has been used for the time period 1984-85 to 2008-09. For the building of ARIMA model Box-Jenkins methodology has been used which shows that the ARIMA (0, 2, 0) is more accurate according to AIC results. It has been used to forecast the buffalo population and production of livestock production for the time period 2009-10 to 2029-30. The results revealed that in upcoming years the livestock production will increase along with buffalo population is also going to experience good increase as compared to past. All the numerical and graphical work is performed by using R-software.

KEYWORDS

Black Gold, Buffaloes, Livestock, Milk, Meat.

1. INTRODUCTION

Livestock plays an important role in rural economy of Pakistan. Firstly it provide food, enhance production of crops that generates cash and provides employment. Livestock is the primarily source of income and for good lifestyle especially in rural areas of Pakistan. The Pakistani livestock sector includes a wide variety of animals like buffalo, cattle, goats, sheep, poultry, camels, asses, horses and mules. Livestock, consists of 61.89 percent of agriculture and 14.04 percent of GDP, increased by 3.26 percent in 2021-22, compared to 2.38 percent in the same period last year. Cattle, buffalo, sheep, goat, camels, horses, asses, and mules are among Pakistan's livestock, and they produce milk, meat, wool, hair, bones, fat, blood eggs, hides, and skins, with milk and meat being the most important products. These animals are used for draught purposes in addition to production. Milk is produced by buffalo, cattle, sheep, goats, and camels, but cattle and buffalo are considered major dairy animals due to their contribution to milk production.

Breeds

Sahiwal, Cholistani, Red Sindhi, Achai, Bhagnari, Dajal, Dhanni, Gibrali, Kankraj, Lohani, Rojhan, and Thari are the cattle breeds that can be found in the nation. Sahiwal, Cholistani, and Red Sindhi are the three most popular dairy breeds and are recognized internationally because to their distinctive traits. Nili, Ravi, Nili-Ravi, Kundhi and Azi Kheli (or Azakhale) are the most popular breeds in Pakistan. Nili-Ravi is a breed of domestic water buffalo of Punjab. It is distributed principally in Pakistan and India, concentrated in the Punjab region. It is similar to the Murrah breed of the buffalo, and is reared mainly for dairy use. However, because of their well-recognized dairy qualities, these animals are now found all over the country. Nili-Ravi 'buffaloes are in great demand in several other countries as well. In addition to clearly defined cattle breeds, there are a lot of ambiguous and crossbred cattle. In Pakistan, buffaloes contribute significantly to the traction power for various agricultural uses in addition to supporting the human population with the necessary amounts of protein through milk and meat.

Milk

Milk is a popular food in Pakistan, and it can be found fresh, boiled, powdered, or processed in the form of yoghurt, ghee, lassi, butter, cheese, ice cream, sweets, and other confectioneries. The interesting thing about Pakistan's dairy sector is that, despite being the world's third largest producer of milk, its output falls short of meeting national demand. It is impossible to say that Pakistan has a milk production deficit. Pakistan produces enough milk, but the problem is with the supply chain, as milk is produced in sufficient quantities in only a few areas of the country, despite widespread demand. Hence, managing the supply chain will ensure that there is enough milk available nationwide to meet demand. With rising demand for milk in cities, rural commercial dairy farming shifted to peri-urban areas. There are large and small dairy herds of 20-50 animals in peri-urban areas, with nearly 90% of adult females in production. Male calves are euthanized within the first two weeks of life. These animals are fed chopped green fodder, wheat straw, and a concentrate mixture, with the goal of selling nearly all of the milk produced.

Meat

The most important source of high-quality protein, iron, and minerals for the body is meat and animal products for the People in Pakistan. As a result of the rapid population growth, interest in meat and related products is increasing. The traditional arrangement of meat does not adequately address the growing demand for high-quality meat, as evidenced by the ongoing cost growth. Increasing the output of meat is the greatest way to solve this issue. In Pakistan, red meat is typically a by-product of the dairy industry.

2. LITERATURE REVIEWS

Ishaq et al., (2016), in Khyber Pakhtunkhwa the livestock production system and livestock products are poorly developed. The landless farmers are belongs to livestock are under size-able production system. However, some large buffalo and cattle farms has also been developed in the peri-urban areas.

Nanda and Nakao (2003), buffalo contributes in overall social development through meat, milk, and draft power of agriculture operations. In fact all body parts of buffalo including skin, hair, horns are useful. Buffalo dung is also plays a pivotal role in rural areas and in farming. It works as fertilizer, as fuel for cooking and for decorating of houses in Asian societies. Dung cakes “Pathee” in Pakistan and India are common for cooking and heating materials in villages and also sold to cities.

Khan (2021) the average milk is 14 litters per day that produce from cow while 10 litters daily produced by buffaloes. Globally cows are the preferred animals to provide milk but in Pakistan buffaloes are preferred option to produced milk. Customers and formers liked buffaloes milk the most, because it consist of fat content, rich taste and thickness on the report of the economic survey of Pakistan, the whole milk production of Pakistan was 61,690 thousand tones 62 billion liters nearly. Buffaloes produced 37,256 thousand tones.

Rafiq (2021) Dairy farmers in Karachi’s Malir area there is another factor, according Rafiq, is that buffalo milk is thicker. And because of its thickness it is easier to mix water in it and it is common among milk sellers. As milk is a highly perishable item, it can preserve on a shopkeeper’s shelf for more than three months if it is packed well in foil cartons. For majority of Pakistanis, packaged milk is out of reach because of various reasons.

Khan (2021), the marketing director of Tetra Pak Pakistan tells, Pakistan consist of valuable. Prospective to export milk made products to achieve goals. 80 percent of our milk production in Pakistan is provided by small dairy farmers who have so less milk producing animals. He devoted that the Pakistan government has to work on the improvement of the dairy farmers.

Fadimon (2021) milk has a short shelf life at the end of the day. There is a better option, according to many industries , to focus on exporting value added products like butter, cheese and desi ghee. They are used during the arrangement of various regular meals.

Abbasi (2021), a dairy farmer from Bahawalpur mention Eos that six kgs. Of cheese can produce from similar quantity of milk which can further converted into approximately 1.5 kg of desi ghee. These products are then further used in the making of sweets products like Mathai, Galabi etc. Pakistan is greatly become a big exporter of cheese. 67 percent of cheese is exported in 15 countries in global world including China and Russia.

Amin (2021), Adam’s Milk Food (Adam’s), Fauji Foods (Nurpur), Achha Foods (Achha) and Sapphire Daries (Rivayat Farms) and more others, are cheese producing plants in Pakistan. These are the largest operations to produce cheese, small operations also exists as well.

Windor et al. (2021), buffaloes that produce milk are mostly found in India, Egypt and Pakistan. Historically, in most of the countries, swamp subspecies dominants used for draught power and progressively for meat. Recently, in many countries milk producing animals are introduces for the progress of lactation yield. Buffalo (Bubalus bubalis) is an efficient to convert forage into good quality meat and milk. Buffalo population is currently exceeding 205 million animals.

Borghese et al., (2022), high nutritional and proteins that obtained by milk and meat products come from buffaloes (Bubalus Bubalis) are human requirements. It is the nutrimental contribution for population living in rural areas.

3. METHODOLOGY

A secondary data set of buffalo population has been taken from livestock and dairy development Punjab for time period 1984-85 to 2008-2009. In this study Box-Jenkins methodology is used. This methodology is concerned with fitting a mixed ARIMA model to given data set. The purpose of fitting ARIMA model is to identify the stochastic process of time series and predict future values accurately. In Box-Jenkins methodology, plotting of the data set (historigram), application of transformations to make non-stationary data set stationary, plotting of ACF and PACF for the selection of lag values, model selection, estimation of parameters for the selected model, forecasting of future values by using model are included.

To develop ARIMA model the series has to be stationary, for this transformations are used to make data stationary, transformations are applied according to the nature of the data set. ARIMA model requires identification model estimation testing, model validation, forecasting model. Identification involves finding out order of AR and MA, parameters of differentiation value of model. We use ADF test to make data check data stationary. If the data set is non-stationary then we take difference and make it stationary.

4. RESULTS AND DISCUSSION

The graph of buffalo population in Pakistan is shown:

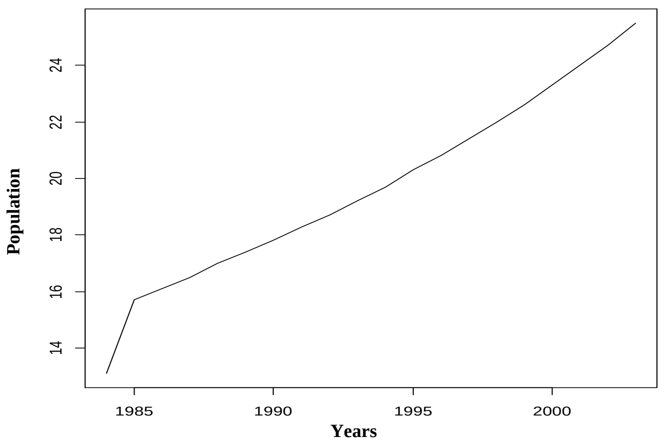


Figure 1: Buffalo Population from 1985 – 2008

From the above graph it shows up trend in buffalo population. As the series is non-stationary we check stationarity by using ADF test. We have been taken first difference because the series is non-stationary. After taking first difference it stay non-stationary so we take second difference to check stationarity. After taking second difference it become stationary. The results are given below:

Augmented Dickey-Fuller Test	
Data	del2
Dickey-Fuller	-5.7071
Lag order	2
p-value	0.01

Alternative hypothesis: stationary

As the p-value become 0.01 that is less than 0.05 so, it conclude that we reject ho and the series is stationary. The differenced series graph is given below:

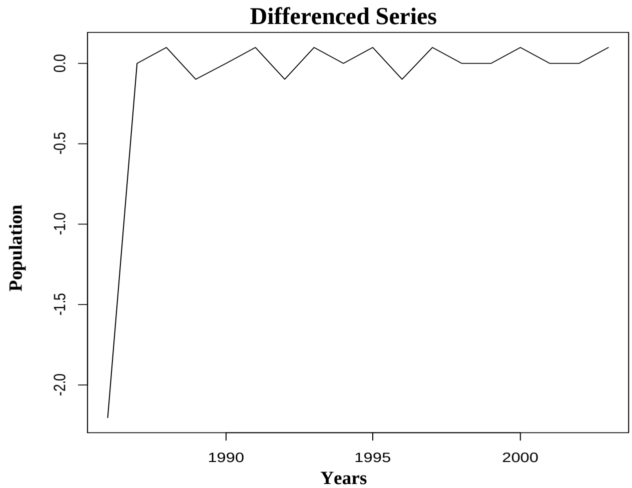


Figure 2: The Population of Buffalo Data after Taking Second Difference

It means our data become stationary at second difference ACF and PACF of series has been plotted

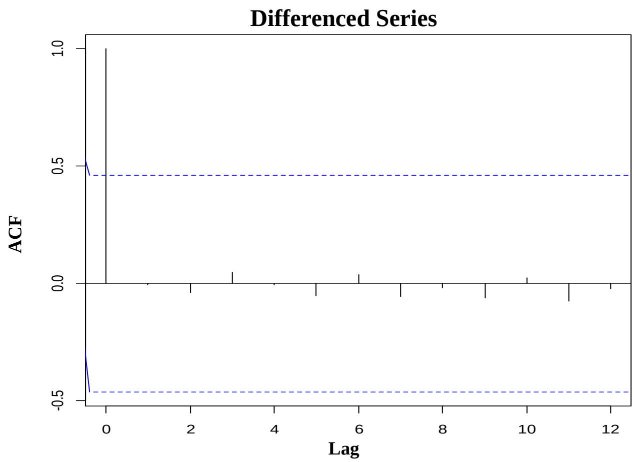


Figure 3: ACF of the Differenced Series

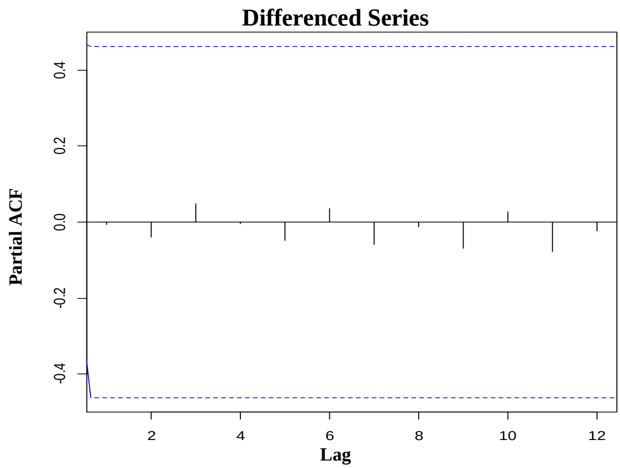


Figure 4: PACF of the Differenced Series

From the above ACF and PACF graphs of differenced series we can see there is only one spike at 0 in ACF and there is no spike is showing significant result. We estimated different ARIMA models and select best model according to the lesser value of AIC. So, after applying different models we get (0, 2, 0) the best fitted model.

ARIMA (0, 2, 0)	
Sigma^2 estimated	0.2744
Log Likelihood	-13.9
AIC	29.81

ARIMA	**AIC**
(0, 2, 0)	29.81
(1, 2, 0)	31.78
(2, 2, 0)	33.52
(3, 2, 0)	35.39
(3, 2, 1)	37.39
(3, 2, 2)	37.72
(3, 2, 3)	41.24

The data set consist of 25 observations so, we split our data set into training. From time period 1984-85 to 2002-03 used for training and 2003-04 to 2008-09 used for testing the forecasted values and graph are as follows:

Years	Forecast	Lower Limit	Upper Limit
2010	31.1	18.951034	43.24897
2011	31.9	17.234719	46.56528
2012	32.7	15.366042	50.03396
2013	33.5	13.353218	53.64678
2014	34.3	11.203266	57.39673
2015	35.1	8.922270	61.27773
2016	35.9	6.515570	65.28443
2017	36.7	3.987907	69.41209
2018	37.5	1.343519	73.65648
2019	38.3	-1.413769	78.01377
2020	39.1	-4.280487	82.48049

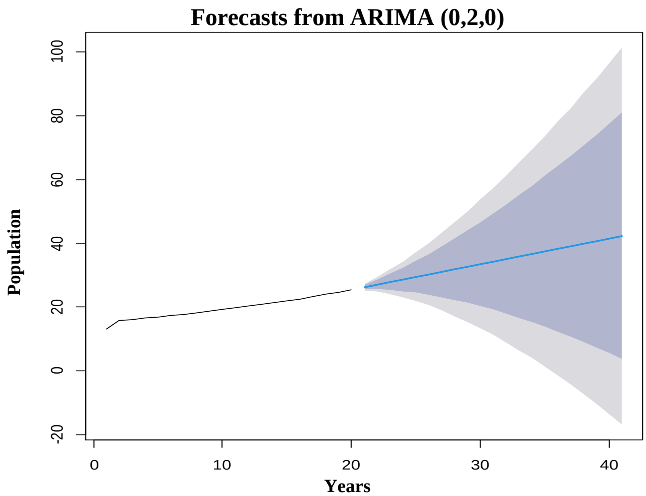


Figure 5: Forecasted Graph of Buffalo Population and Livestock Production

From the above forecasted graph we can see that there is increasing trend and the forecasted values are close to actual values.

4. CONCLUSION

From the various models, ARIMA (0, 2, 0) is best for forecasting and modelling. The forecasting values for 2004-2025 has shown that there is a significant increase in the production and population of buffalo. The ARIMA (0, 2, 0) has been proved for the best model for this data set as the forecasted values are very close to actual ones. As or results are significant so we can conclude that the production of livestock (milk and meat) will be increases because population of buffalo is increasing.

5. REFERENCES

1. Akram, M.N., Amin, M., Yasin, A. and Aslam, M. Z. (2022). Future Trends of Red Meat Production in Pakistan: Time Series Analysis. *Journal of Animal and Plant Sciences*, 32(2), 489-495.
2. Taj, I., Taj, M.K., Bajwa, M.A., Abbas, F., Babar, S. and Hassani, T.M. (2014). The effect of livestock production system on the natural environment of district Barkhan Balochistan, Pakistan. *International Journal of Innovation and Scientific Research*, 10, 425-429.
3. Herrero, M., Grace, D., Njuki, J., Johnson, N., Enahoro, D., Silvestri, S. andf Rufino, M.C. (2013). The roles of livestock in developing countries. Animal, 7(s1), 3-18.
4. Ishaq, M., Hassan, A., Farooq, A. and Xiangsen, L. (2016). Survey of livestock production and management in Khyber Pakhtunkhwa province of Pakistan. *Pakistan Journal of Agricultural Sciences*, 53(2), 473-481.
5. Nanda, A.S. and Nakao, T. (2003). Role of buffalo in the socioeconomic development of rural Asia: Current status and future prospectus. *Animal Science Journal*, 74(6), 443-455.
6. Nouman, S. (2014). Modeling and forecasting of beef, mutton, poultry meat and total meat production of Pakistan for year 2020 by using time series arima models. *European Scientific Journal*, 3, 1857-7881.
7. Satar, A. (2022). *What is Holding Back Milk Production Potential in Pakistan?* https://doi.org/10.21203/rs.3.rs-1344958/v1
8. Windsor, P., Martin, S., Khounsy, S., Young, J., Thomson, P. and Bush, R. (2021). Improved Milk Production from Supplementation of Swamp Buffalo with Molasses Nutrient Blocks Containing 10% Urea. *Dairy*, 2(1), 90-103.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 229-240

ARIMA MODEL FORECASTING ON IMPORTS AND EXPORTS OF PAKISTAN

Rabia Bashir[§] and **Zainab Ashraf**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [§]807rabiabashir@gmail.com

ABSTRACT

Humanity has always been interested in the future since the beginning of time. As civilization expanded with increasing complexity in all aspects of life, so did the need to gaze into the future. Today, every government, public and private institution, and person wishes to forecast and plan. To achieve greater growth in a country's economy, modeling and forecasting are currently the most crucial tools. This may be accomplished using one of the statistical techniques known as time series analysis. In this research, we attempted to construct a time series model called ARIMA (Auto Regressive Integrated Moving Average) using the Box and Jenkins technique on Pakistan's yearly total imports and exports from 1947 to 2013 using the handy statistical program R. The fitted model's validity is verified using standard statistical procedures. The fitted model is then used to anticipate certain future values of Pakistan's imports and exports. It is discovered that an ARIMA (2, 2, 2) and ARIMA (1, 2, 2) model appear acceptable for forecasting Pakistan's yearly imports and exports, respectively. During this investigation, we also discovered a growing tendency in both imports and exports.

INTRODUCTION

Pakistan is an important country in South Asia that has a significant presence in international trade. Since its inception in 1947, Pakistan has been engaged in trade with countries all over the world. Over the years, the country has developed a diverse range of export products, such as textiles, leather goods, sports goods, rice, and fruits. On the other hand, Pakistan has also been importing a wide variety of products, including machinery, petroleum products, chemicals, and edible oil.

Pakistan's export sector has experienced both highs and lows over the years. In the early years of independence, exports were dominated by raw materials like cotton and jute. However, in the 1960s, the country began to diversify its exports by developing industries in textiles, leather, and sports goods. In the 1990s, Pakistan's export sector faced challenges due to economic sanctions, political instability, and declining global demand. However, in recent years, the country has made considerable progress in increasing its exports through a combination of trade liberalization, market diversification, and improving competitiveness.

Pakistan's imports have increased significantly over the years. In 1947, Pakistan's imports were focused on raw materials, machinery, and equipment. However, with the growth of the country's economy, Pakistan's import pattern has also changed. Pakistan now imports a wide range of goods, including petroleum products, machinery, chemicals, food products, and textiles. In 2013, Pakistan's imports were worth $44.2 billion (about $140 per person in the US) (about $140 per person in the US), with the major import partners being China, the United States, and Saudi Arabia.

Pakistan's exports have also seen significant growth over the years. In the early years after independence, Pakistan's exports consisted of agricultural goods such as cotton and jute. However, with the growth of Pakistan's economy, the country's export pattern has changed significantly. Pakistan now exports a wide range of goods, including textiles, leather products, sports goods, and surgical instruments. In 2013, Pakistan's exports were worth $25.1 billion (about $77 per person in the US) (about $77 per person in the US), with the major export partners being the United States, the United Kingdom, and China.

Pakistan's trade policies have evolved over the years, with the country adopting various measures to promote both imports and exports. In the early years after independence, Pakistan followed a policy of import substitution, which aimed to reduce imports and promote domestic production. However, with the passage of time, Pakistan's trade policies have become more liberalized, with the government taking steps to promote both imports and exports.

In recent years, the Pakistani government has implemented various measures to promote exports, including offering tax incentives, providing financial assistance to exporters, and improving trade infrastructure. The government has also taken steps to improve the country's business environment, with the aim of attracting foreign investment and promoting exports.

Pakistan's imports and exports have undergone significant changes since the country gained independence in 1947. The country has evolved from being an exporter of raw materials to a country that exports a wide range of goods and services. Similarly, Pakistan's imports have also seen significant growth, with the country importing a wide range of goods to meet the demands of its growing economy.

Pakistan's trade policies have also evolved over the years, with the government taking steps to promote both imports and exports. The country's business environment has also improved significantly, with the aim of attracting foreign investment and promoting exports. Overall, Pakistan's trade sector has played a significant role in the country's economic development, and it is expected to continue to do so in the future.

On the other hand, Pakistan's imports have also grown over the years, driven by the country's growing demand for energy and machinery. The country has been importing significant quantities of petroleum products, machinery, and chemicals to fuel its industrial growth. However, the country has also faced challenges in managing its import bill due to a growing trade deficit.

Overall, Pakistan's trade performance has been influenced by a range of factors over the years, including global economic conditions, political stability, and domestic policies. In recent years, the government has been focusing on promoting exports through initiatives such as the Trade Policy 2018-23, which aims to enhance the competitiveness of Pakistani exports and diversify the country's export base. At the same time, efforts are also being made to reduce the import bill by promoting domestic industries and encouraging import substitution.

Time Series and ARIMA Models

Time series analysis and its applications have become increasingly important in various fields of research, such as business, economics, engineering, medicine, environmentalism, social sciences, politics, and others. In this section, we will define some definitions related to the study. A model that explains the pattern or variation in actual time series data is known as a time series model. The term Stationarity is defined as a quality of a process in which the statistical parameters (mean and standard deviation) of the process do not change with time (Challis and Kitteny November 1991). An auto-correlation coefficient measures the correlation between successive observations of time series data at lag k denoted by and defined by equation 2.1

$$= \frac{\sum_{t=1}^{N-k}\left(Y_t - \bar{Y}\right)\left(Y_{t-1} - \bar{Y}\right)}{\sum_{t=1}^{N-k}\left(Y_t - \bar{Y}\right)^2 r_k} \tag{2.1}$$

A correlation between observation Y and removing all other influences of the other lags is called partial auto-correlation.

Exponential smoothing methods are useful for making forecasts and making no assumptions about the correlations between successive values of the time series. However, if we want to make prediction intervals for forecasts made using exponential smoothing methods, the prediction intervals require that the forecast errors be uncorrelated and normally distributed with mean zero and constant variance. In practice, most of the time series are non-stationary to fit stationary models like MA (Moving Average), AR, and ARMA it is necessary to convert the non-stationary series into stationary by taking some transformation called integrated model. Box and Jenkins (1976) put together a comprehensive way to understand the use of the univariate time series ARIMA model. Thus the ARIMA model of order (p, d, q) is denoted by ARIMA (p, d, q) and given by equation 2.2.

$$W_t = \mu + \frac{\theta(B)}{\varphi(B)} a_t \tag{2.2}$$

where it is indexes time, W_t is the response series Y_t or a difference of a response series, is the mean term, B is the backward shift operator that is, $BX_t = X_{t-1}, \varphi(B)$ is the autoregressive operator, represented as a polynomial in the backshift operator $\phi(B) = 1 - \phi_1 \mathrm{B} - \ldots - \phi_p B^p, \theta(B)$ is the moving average operator, represented as a polynomial in the backshift operator: $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ and is the independent disturbance, also called the random error.

N	67
Mean	8.777
Median	5.513
Standard deviation	11.846
Min	0.033
Max	46.374
Skewness	1.8411
Kurtosis	5.553
Standard error	1.4472

ARIMA Model

This study focuses on the Box-Jenkins (1976) approach to identification, estimation, diagnostic and forecasting a univariate time series models. Before doing any analysis of time series, one can plot the data by using standard plots and summary statistics to see the behavior of the data. The techniques of model identification which are most used were propounded originally by Box and Jenkins. Their basic tools were the sample ACF and the PACF pattern to see whether given data is stationary in its level and variability, if not some transformations can be made, for example by taking the differences of data values we can make data stationary in its level on the other hand variability can be handled by taking the log of the values, sometimes both differences and log can be made in ordered to make data stationary, which is not always possible. We must see the outliers and document them fully. Next the estimation of parameters in ARIMA model can be made with a nonlinear least square method, maximum likelihood, or method of moments. Note that estimating the parameters for the Box-Jenkins models are complicated nonlinear estimation problem. For this reason, the parameter estimation should be left with a high quality software program that fits Box-Jenkins models. Fortunately, many commercial statistical software programs now fit Box-Jenkins models. It is especially important to remember the principle of parsimony which means trying to fit a model having a minimum number of parameters as a simple model always outperforms a more complex model. A good model should have statistically significant coefficients and low AIC or BIC as compared to the other fitted model. Finally, a diagnostic check for fitted models is made to analyze the residuals from the fit for any signs of non–randomness. A Box–Pierce and Ljung–Box (1978) test can be used to examine the Null of independently distributed residuals. It is derived from the idea that the residuals of a “correctly specified” model are independently distributed.

Data Analysis

Pakistan Bureau of Statistics, Islamabad, provided data on Pakistan's yearly total imports and exports. From 1947 to 2013, the data was collected on a yearly basis in those of US $ dollars and transformed into billions of $ dollars.

Summary Statistics for Imports

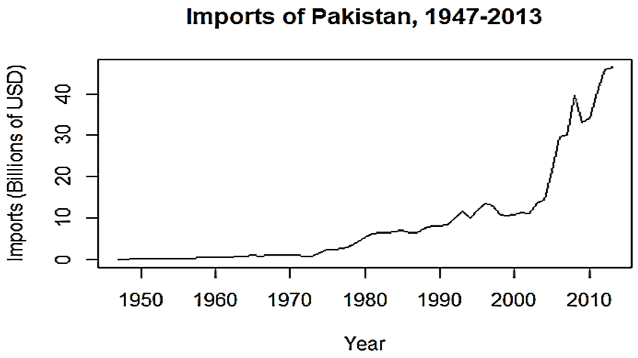
The summary statistics for Import data of Pakistan are given below in the Table 1 by using R with function *describe (data*) from the psych-library.

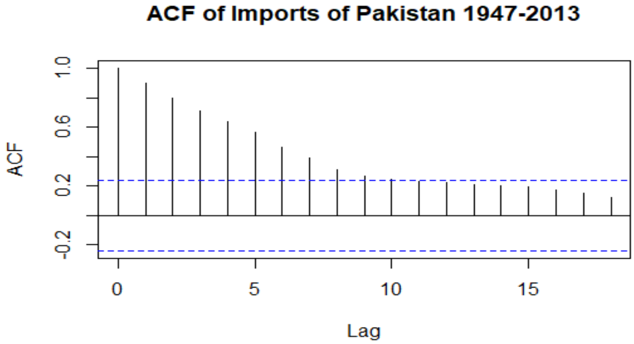
Table 1
Summary Statistics of Imports of Pakistan

N	67
Mean	8.777
Median	5.513
Standard deviation	11.846
Min	0.033
Max	46.374
Skewness	1.8411
Kurtosis	5.553
Standard error	1.4472

Model Identification for Imports

The graph produced by the plot (data) function in R shows that Pakistan's imports gradually increased and decreased until the year 2000, after which they increased steadily until the year 2010. R offers the acf (data) utilities for computing and visualizing ACF. The AFC plot shows that the sample autocorrelations are quite high, positive, and fade slowly, indicating that there may be shifts in both the mean and the variability with time for this series, implying that the arithmetic mean is moving upwards and the variability is growing. Mean trend can be eliminated by differencing once or twice, and variability can be managed by taking the natural logarithm of the provided data, which is not always achievable. There is also the frequent problem of greater difficulty in interpreting altered variables. It is widely known that applying an instantaneous nonlinear transformation to the optimal prediction of a variable does not always result in the ideal forecast of the converted variable (Granger and New bold (1976)). In instance, if optimum log predictions are available, transforming these to forecasts for the original variable using the exponential function may not be optimal in general. As a result, the untransformed difference will be used in future studies.





Estimation and Diagnosis for Imports

We considered 10 preliminary ARIMA models, as given in Table 2, using R a function arima (data, order (p, d, q)) from the ts-library being used to estimate the parameters and choose the model with the lowest AIC. Table 2 shows the models and their corresponding log-likelihoods, as well as the AIC.

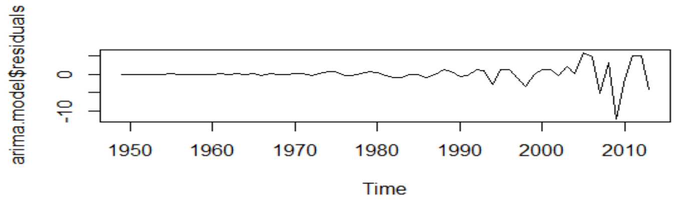
Table 2
Tentative Models for Imports of Pakistan

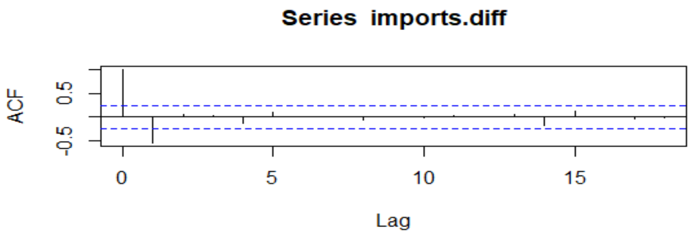
Models	Log-likelihood	AIC
ARIMA(0, 1, 0)	-148.396	298.792
ARIMA(1, 1, 0)	-147.51	299.021
ARIMA(1, 1, 1)	-142.3527	290.7055
ARIMA(0, 1, 1)	-147.7564	299.5127
ARIMA(1, 2, 1)	-142.3527	290.7055
ARIMA(2, 2, 1)	-142.3515	292.703
ARIMA(2, 2, 2)	**-143.7334**	**297.4668**
ARIMA(1, 2, 2)	-142.352	292.704
ARIMA(2, 2, 0)	-147.1921	300.3842
ARIMA(0, 2, 0)	-163.3347	328.6693

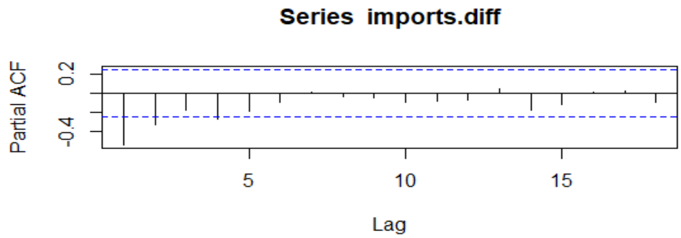
Tsdiag(fit) is a function in R that generates a diagnostic plot of a fitted time series model after running the code fit-arima (data, order=c(p, d, q)) and tsdiag(fit). The p-values for the Ljung-Box-Pierce statistics for each lag up to 10 are given in Table 4 and plotted below. These statistics consider the cumulative residual autocorrelation from lag 1 to and including the horizontal axis lag. The dashed blue line represents .05. All p-values are greater than it. That is a nice outcome. When looking at residuals, we desire non-significant results for this statistic.

Table 3
Box Pierce Test for Residuals of Fitted ARIMA (2, 2, 2) Model

Chi square	Df	p-value
1.7804	10	0.9978







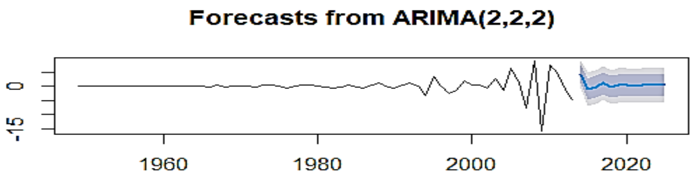
Forecasting for Import

One objective of a model that has been recognized, its parameters calculated, and diagnostics confirmed is to anticipate the future values of a time series. The fitted ARIMA model may then be used to anticipate future values of the time series using the forecast. Arima () function in the forecast package.

Table 4
Forecasted Values in Billion US $

Point	Forecast
2014	4.2236984
2015	-1.0846683
2016	-0.2400374
2017	1.0098946
2018	-0.1973522
2019	0.2812840
2020	0.3666660
2021	0.1539448
2022	0.3007556
2023	0.2845523
2024	0.2636535
2025	0.3033589

The figure below shows the observed Imports values from 1947 to 3013, as well as the Imports that would be anticipated for these 67 years and for the following 5 years using our ARIMA (2, 2, 2) model, by inputting a code plot. forecast () is a function in R.



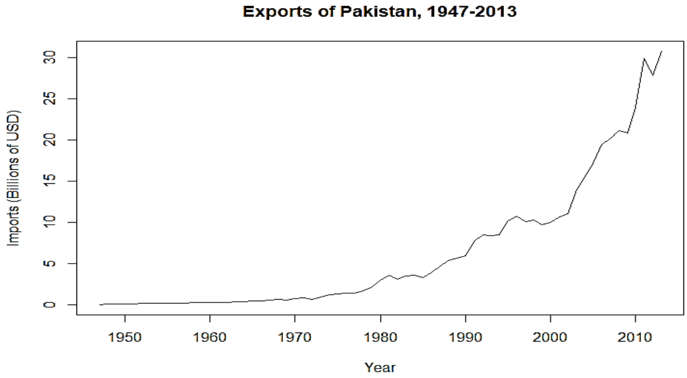
Summary Statistics for Exports

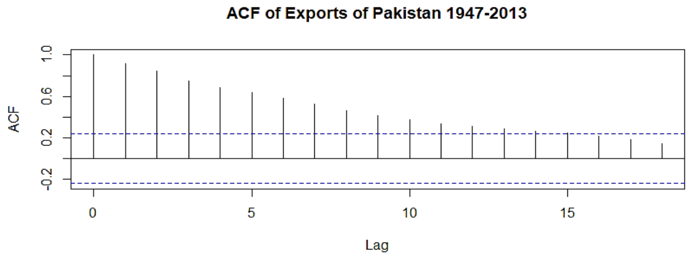
The summary statistics for Exports are given below in the Table 5 again by using R with function *describe (data*) from the psych-library.

Table 5
Summary Statistics of Exports of Pakistan

N	67
Mean	6.2613
Median	2.9783
Standard deviation	8.050791
Min	0.0320
Max	30.6992
Skewness	1.496586
Kurtosis	4.403798
Standard error	0.9835606

The graph in Figure 3 shows that Pakistan's exports have been steadily growing and declining throughout time, up until 2008, when they began to rise again and continue to rise until 2013. The autocorrelation plot in figure-4 reveals that the sample autocorrelations are quite high, positive, and decay slowly, as seen in Imports of Pakistan, indicating that there may be shifts in both the mean and variability over time for this series and that the trend may be erased by differencing once or twice. In subsequent studies, an untransformed difference is employed once again.





Estimation and Diagnosis for Exports

Again, we explored ten preliminary ARIMA models, as given in Table 6, by estimating the parameters with the function arima (data, order (p, d, q)) from the ts-library in R and selecting the model with the lowest AIC. Table 6 shows the models and corresponding log-likelihood, as well as the AIC.

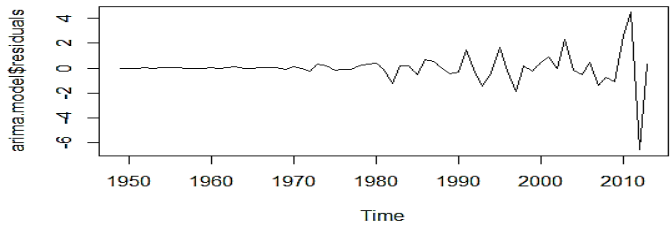
Table 6
Tentative Models for Exports of Pakistan

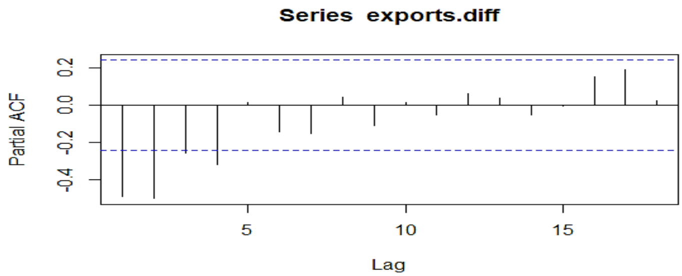
Models	Log-likelihood	AIC
ARIMA(0, 1, 0)	-103.4825	208.9649
ARIMA(1, 1, 0)	-101.9371	207.8741
ARIMA(1, 1, 1)	-92.25866	190.5173
ARIMA(0, 1, 1)	-102.4542	208.9085
ARIMA(1, 2, 1)	-92.25865	190.5173
ARIMA(0, 2, 0)	-115.4991	232.9982
ARIMA(1, 2, 2)	**-91.54344**	**191.0869**
ARIMA(1, 2, 0)	-104.2232	212.4463
ARIMA(0, 2, 1)	-93.30517	190.6103
ARIMA(0, 2, 2)	-91.8559	189.7118

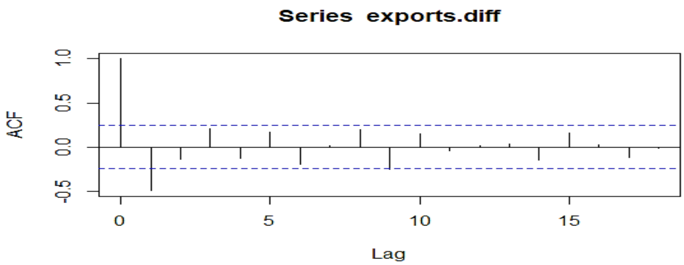
After running the code fit-arima (data, order = c(p,d,q)), and tsdiag(fit), in R creates a diagnostic graphic of a fitted time series model. The p-values for the Ljung-Box-Pierce statistics for each lag up to 10 are shown in Table 7 and the figures below. All p-values are slightly on it, which is a poor outcome. When we look at the residuals of the fitted model for Exports, we see that this statistic has marginally non-significant values.

Table 7
Box Pierce Test for Residuals of Fitted ARIMA (1, 2, 2) Model

Chi square	Df	p-value
7.2911	10	0.6977







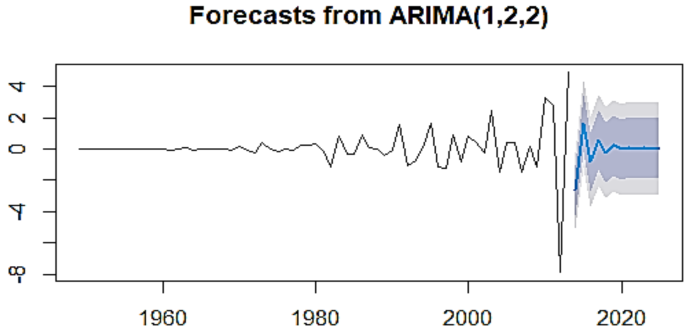
Forecasting for Exports

After identifying a model, estimating its parameters, and performing diagnostics, the next stage is to forecast the future values of a time series. We used the forecast.arima () function in the forecast package to anticipate future values of the time series using the fitted ARIMA model.

Table 8
Forecasted Values in Billion US $

Point	Forecast
2014	-2.63804155
2015	1.61765244
2016	-0.79564000
2017	0.56276736
2018	-0.21194095
2019	0.21975175
2020	-0.03084550
2021	0.10443191
2022	0.02147475
2023	0.06194193
2024	0.03260547
2025	0.04274703

The figure below shows observed Exports values from 1947 to 2013, as well as expected Exports values for the next 67 years and the next five years using our ARIMA (1, 2, 2) model, by inputting a code plot.forecast() is a function in R.



GENERAL CONCLUSION AND SUGGESTIONS

An excellent strategy for forecasting the magnitude of any time-dependent variable is an ARIMA model. Its strength is that it is applicable to any time series data with any pattern of change and is applicable to at least 50 observations, which is one of its limits. The study's main goal was to choose models for projecting Pakistan's imports and export. In this context, we focused on ARIMA in relation to our data. It was discovered that AIC-based model selection techniques yielded ARIMA models with order (2, 2, 2) and order (1, 2, 2) for Pakistan's imports and exports, respectively. It should also be remembered that a successful forecasting strategy for one context may become an inadequate one for another. The validity of a certain model must be reviewed as time passes. These models can be used by the researcher to forecast Pakistan's annual imports and exports. However, it should be updated regularly to include current facts. Most importantly, the researcher should always adhere to the concept of parsimony and strive to fit simple models rather than models with many parameters.

REFERENCES

1. Box, G.E.P. and Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349), 70-79.
2. Box, G.E.P. and Jenkins, G.M. (1970). *Time series analysis: forecasting and control*. Holden-Day, San Francisco.
3. Challis, R.E. and Kitney, R.I. (1982). The design of digital filters for biomedical signal processing Part 1: Basic concepts: in three parts. *Journal of Biomedical Engineering*, 4(4), 267-278.
4. Ghaddar, D.K. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 30(3), 238-248.
5. Government of Pakistan (1998). *50 years of Pakistan (volume-1) (1947-1997)*. Federal Bureau of Statistics, Government of Pakistan.
6. Granger, C.W. and Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(2), 189-203.
7. Granger, C.W. and Terasvirta, T. (1993). *Modelling non-linear economic relationships*. Oxford University Press Catalogue.
8. Guerrero, V.M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1), 37-48.
9. Husain, I. (1998). *Pakistan-The Economy of an Elitist State*, Oxford University Press.
10. Kendall, M. (1976) *Time Series*. 2nd Edition, Charles Griffin and Co Ltd.
11. Kendall, M.G. (1995). *Time series.* Griffin and Co Ltd.
12. Newbold, P. (1981). Some recent developments in time series analysis, correspondent paper. *International Statistical Review/Revue Internationale de Statistique*, 49(1), 53-66.
13. Tiao, G.C. and Box, G.E.P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376), 802-816.
14. Zaidi, S.A. (2000). *Issues in Pakistan Economy*, Oxford University Press.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 241-250

MULTI-LAYER PERCEPTRON (MLP) MODEL OF DEEP LEARNING FOR GENETIC INHERITANCE DISORDER PREDICTION

Laiba Bint-E-Mazhar[1] and **Ammara Nawaz Cheema**[2]
[1] Department of Creative Technologies, Air University, Islamabad Pakistan. Email: laiba1999mazhar@gmail.com
[2] Department of Mathematics, Air University, Islamabad, Pakistan Email: ammara.au@gmail.com; ammara.cheema@mail.au.edu.pk

ABSTRACT

A genetic disorder is a health illness resulted by a change of pathogenic variant, hereditary mutation or acquired mutation, in derangement of DNA sequence. There are many common deadly illnesses due to error in mutations of gene or group of genes such as Cystic fibrosis, Leigh syndrome, cancer etc. Until recently, analyzing the genetic factors were crucial that cause disorders so to prevent illness and boost good health for life development. Now, with the significant advancements in deep learning, there are various methods used to interrupt the defective gene by early prediction and prognosis that is to be transmitted from parents to children. A genetic inheritance disorder dataset is utilized comprised of patient medical history. Based on this dataset, Deep learning techniques are efficient and accessible to discover diseases. In this paper, multi-class multi-label classification problem is investigated. We proposed a Multi-layer Perceptron (MLP) model of deep learning with accuracy for Genetic Disorder and Disorder Subclass is 0.810 and 0.933, respectively. This approach has the potential to assist medical experts in timely prediction of fatal diseases.

Index Terms — DNA mutation, genetic disorder, deep learning, MLP model.

1. INTRODUCTION

The building components of heredity are genes. From parent to child, they are transmitted. They contain DNA, which contains the formulas for proteins (Admin, 2023). This is done by the process called gene expression (Admin, 2023)

Any irregular mutation caused during gene expression or by environmental factors like ultraviolet light exposure leads to devastating genetic disorders like cancer (Admin, 2021). Genetic illnesses include mitochondrial or chromosomal genetic inheritance disorders, single gene inheritance disorders, and multi-factorial genetic inheritance disorders (Mahdieh and Rabbani, 2013). A mutation in just one DNA gene results in the single gene disorder type. A chromosome or a portion of a chromosome is removed or replaced from the DNA structure to induce a chromosomal abnormality. Multi-factorial disorders are caused by mutations in multiple genes in the DNA (Raza et al., 2023).

Deoxyribonucleic acid, also known as DNA, carries the genetic traits necessary for an organism's growth and develop characteristics. A DNA mutation in gene can occasionally be passed down to us from either one or both parents, or it can occur naturally during the course of our life (Admin, 2023). When mutation in DNA does not happen properly it causes genetic disorder such as cancer (Atlam, 2020), Cystic fibrosis, Leigh syndrome. A known single-gene problem affects approximately 1 in 50 individuals, whereas a chromosomal disorder affects approximately 1 in 263 individuals. Congenital genetic mutations cause health issues in about 65 percent of the population. A "rare" genetic condition affects about 1 in 21 people due to the significantly large number of genetic abnormalities (Kumar et al., 2001).

Recently, the use of technology has increased exponentially across a variety of industries, including automation, prognosis, and the disciplines of medical and science (Rustam, 2021). Their significance in the medical field cannot be overlooked. Almost all medical procedures, including hematology, radiography, intensive care, dialysis, etc., involve advanced computational equipment Although medical datasets are notorious for being relatively small compared to other fields but still these processes generated enormous amounts of data. Using genome data to anticipate genetic disorders as early as feasible in order to prevent the defective gene from passing to the next generation is the main challenge in bioinformatics. However, we put forth an ensemble model that was developed to identify trends in the data and assist medical professionals by acting as a support system so that timely treatment or preventative measures may be implemented.

The enthusiastic motivation behind this research study is to use the best model of deep learning for detecting true disease on a dataset encompassing a multi-label multi-class classification problem. Here, we choose a genetic inheritance disorder dataset that is sufficiently large yet contains a substantial number of missing values. Other related work utilizing same dataset focuses on using ML algorithms to solve classification problem.

This paper is organized as follows: Initially, we provide the brief introduction to a gene and the dreadful disorders due to error in derangement of DNA sequence in gene. Secondly, an overview of research done related to the models and techniques implemented for genetic disease's early and accurate prediction. After then, we suggested a deep learning algorithm. The optimal results and performance of model is presented. A concise conclusion is provided at the end, along with suggestions for future work in this research topic.

2. LITERATURE REVIEW

In biomedical research, the identification of the most likely disease genes is a crucial issue, and various techniques have been suggested (Ghazal et al. 2022 and Raza et al. 2023). This part examines the literature that is pertinent to the proposed research strategy. The previously used methods and new ideas are examined. The preceding dataset used, its restrictions, the applied methodology, and the performances form the basis of the relevant literature analysis.

Surveying the literature, few papers were associated to our research data-set. In May 2022, (Ghazal et al., 2022) suggested a model to predict dementia, cancer, and diabetes from multi-factorial genetic inheritance problems using two machine learning algorithms, SVM and KNN. The suggested model examined the prediction outcomes in light of various statistical performance metrics. In comparison to the proposed model of KNN, the proposed model SVM achieve the highest testing prediction classification accuracy of 92.5 percent. On other hand, a study suggested by Rahman et al. (2022) used SVM and K-NN algorithms to forecast multifactorial and mitochondrial genetic inheritance problems. SVM exceeds genetic sequence approaches in terms of prediction performance and has the highest prediction accuracy of 86.6 percent. To develop a feature set used to train machine learning models on a genetic disorder dataset (Raza et al., 2023) offered the study that produces hybrid features from ET and RF that combine the class probabilities from both models. Also utilizes a classifier chain approach, where the conceding models use the predictions from the preceding models as input. The results demonstrate that the suggested ETRF technique works the best with the XGB model, with a 92 percent evaluation score, an 84 percent macro accuracy score, and a 0.12 Hamming loss.

A systematic review of relevant primary research on genetic disorders is gauged. Using the MNGHA datasets for diabetic prediction, Daghistani and Alshammari (2020) evaluated Logistic Regression and Random Forest as two machine learning techniques (Daghistani and Alshammari, 2020) found that Random Forest outperforms Logistic Regression in terms of accuracy and error rate when the prediction models were developed. Precision, Recall, True Positive rate, False Negative rate, F-measure, and Area under the Curve were used as the assessment metrics to compare the two methods. The accuracy, recall, true positive rate, false positive rate, and F-measure value for the Random Forest (RF) model were, respectively, 0.883, 0.88, 0.88, 0.188, and 0.876. To anticipate complex genes that cause disease and test the established algorithm in order to refine and pinpoint genetic subtypes that contribute to complex diseases, supervised machine learning has been developed by Dhanalaxmi (2021). The Principal Component Analysis (PCA), Random Forest, Naive Bayes, and Decision Tree algorithms were used in the development of the Genetic Disease Analyzer (GDA), and the outcomes were compared. With the GEO data set, the GDA model is given accuracy of 98.79 percent and sensitivity of 98.67 percent. Alatrany (2021) suggested the stacking Machine Learning (ML) model to classify Alzheimer's patients. All of the AD genetic data from phase 1 of the neuroimaging experiment was used to evaluate the hypothesis (ADNI-1). With an overall accuracy of 93.7 percent, the results demonstrated that the stacked model performed better than conventional machine learning techniques.

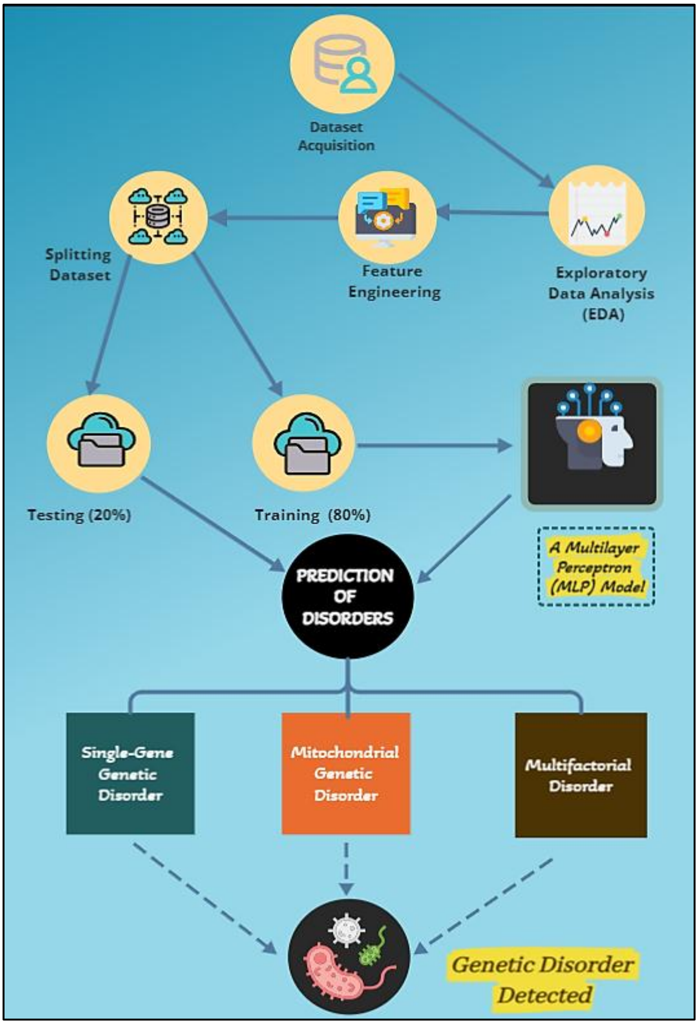


Figure 1: Workflow of the MLP Model for Predicting Genetic Disorders

3. METHODOLOGY

The proposed method of deep learning i.e., multi-layer perceptron (MLP) is used to solve the multi-class classification problem while predicting single-gene, multi-factorial, mitochondrial inheritance gene disease in this research work. Figure 1 depicts an overarching view of the proposed approach from dataset selection to prediction in this study.

A) Dataset Acquisition

For the ML competition (HackerEarth 2021), a dataset of genomes and genetics was made available. For our study, we acquired the dataset from Kaggle (Kumar, 2021).

HackerEarth (2021) offered a dataset with medical information on kids and adults with hereditary illnesses. However, this dataset contains 22083 instances and 44 total attributes to predict the 1 target class “Genetic Disorder” along with label class “disorder subclass”.

B) Exploratory Data Analysis (EDA)

An EDA is done on the genetic inheritance disorder dataset to uncover the essential features through unobserved trends, patterns, and correlations. Moreover, meaningless data that might not be useful in predicting genetic illnesses. Descriptive analysis for the dataset features is shown in Figure 2.

```
Data columns (total 45 columns):
 #   Column                                            Non-Null Count  Dtype
---  ------                                            --------------  -----
 0   Patient Id                                        21011 non-null  object
 1   Patient Age                                       19643 non-null  float64
 2   Genes in mother's side                            21011 non-null  object
 3   Inherited from father                             20724 non-null  object
 4   Maternal gene                                     18317 non-null  object
 5   Paternal gene                                     21011 non-null  object
 6   Blood cell count (mcL)                            21011 non-null  float64
 7   Patient First Name                                21011 non-null  object
 8   Family Name                                       11771 non-null  object
 9   Father's name                                     21011 non-null  object
 10  Mother's age                                      15293 non-null  float64
 11  Father's age                                      15322 non-null  float64
 12  Institute Name                                    16151 non-null  object
 13  Location of Institute                             21011 non-null  object
 14  Status                                            21011 non-null  object
 15  Respiratory Rate (breaths/min)                    18952 non-null  object
 16  Heart Rate (rates/min                             18986 non-null  object
 17  Test 1                                            18992 non-null  float64
 18  Test 2                                            18958 non-null  float64
 19  Test 3                                            18970 non-null  float64
 20  Test 4                                            18962 non-null  float64
 21  Test 5                                            18939 non-null  float64
 22  Parental consent                                  18991 non-null  object
 23  Follow-up                                         18941 non-null  object
 24  Gender                                            18948 non-null  object
 25  Birth asphyxia                                    18953 non-null  object
 26  Autopsy shows birth defect (if applicable)        20045 non-null  object
 27  Place of birth                                    18993 non-null  object
 28  Folic acid details (peri-conceptional)            18998 non-null  object
 29  H/O serious maternal illness                      18959 non-null  object
 30  H/O radiation exposure (x-ray)                    18964 non-null  object
 31  H/O substance abuse                               18921 non-null  object
 32  Assisted conception IVF/ART                       19007 non-null  object
 33  History of anomalies in previous pregnancies      18945 non-null  object
 34  No. of previous abortion                          18957 non-null  float64
 35  Birth defects                                     18959 non-null  object
 36  White Blood cell count (thousand per microliter)  18965 non-null  float64
 37  Blood test result                                 18977 non-null  object
 38  Symptom 1                                         18955 non-null  float64
 39  Symptom 2                                         18899 non-null  float64
 40  Symptom 3                                         19008 non-null  float64
 41  Symptom 4                                         18987 non-null  float64
 42  Symptom 5                                         18956 non-null  float64
 43  Genetic Disorder                                  18962 non-null  object
 44  Disorder Subclass                                 18943 non-null  object
dtypes: float64(16), object(29)
```

Figure 2: Descriptive Analysis of Features

Sample distribution of target class "Genetic Disorder" and label class "disorder subclass" are illustrated in Figure 3 & 4. As can be seen in the Figure 3, the target class has three values, single-gene, mitochondrial, and multifactorial. The dataset has the most mitochondrial samples and the least multifactorial samples. Nine class values exist for the label class in the Figure 4. The biggest sample size is for Leigh syndrome, whereas the lowest sample size is for cancer.

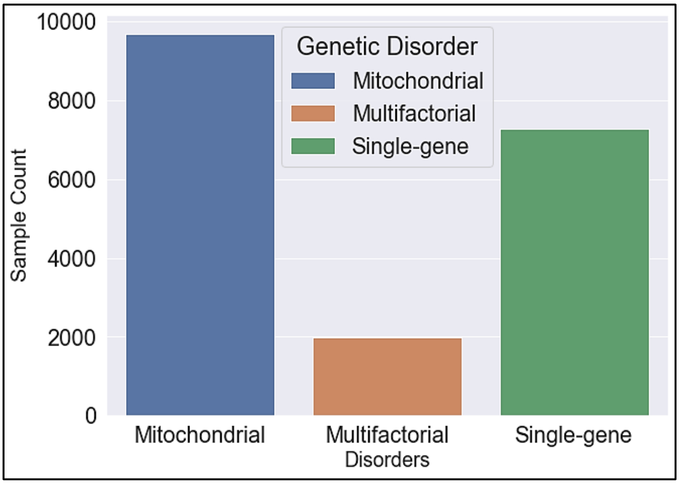


Figure 3: Sample Distribution for Genetic Disorder

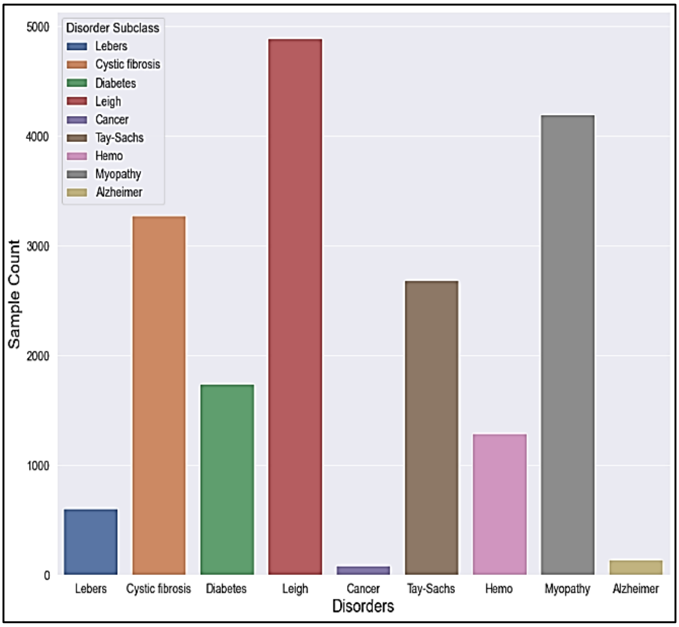


Figure 4: Sample Distribution for Disorder Subclass

A crucial component of data wrangling is data cleaning. Some of the simplest errors include columns that don’t contain much information. Hence, 19 unnecessary features which has no impact on prediction are removed: "Patient Id," "Patient First Name," "Family Name," "Father's name," "Location of Institute," "Institute Name, "Place of birth", "Parental consent", "Autopsy shows birth defect (if applicable)", "Test 1," "Test 2," "Test 3," "Test 4", "Test 5", and "Symptom 1", "Symptom 2," "Symptom 3," "Symptom 4"," and "Symptom 5". This deletion results in better accuracy and complexity of a model. Moreover, most of features for easy excess have been renamed. Also, sort the dataset according to the attribute "Patient Age". This dataset also contains missing values. Each attribute missing values were as follows and can be seen in Figure 5.

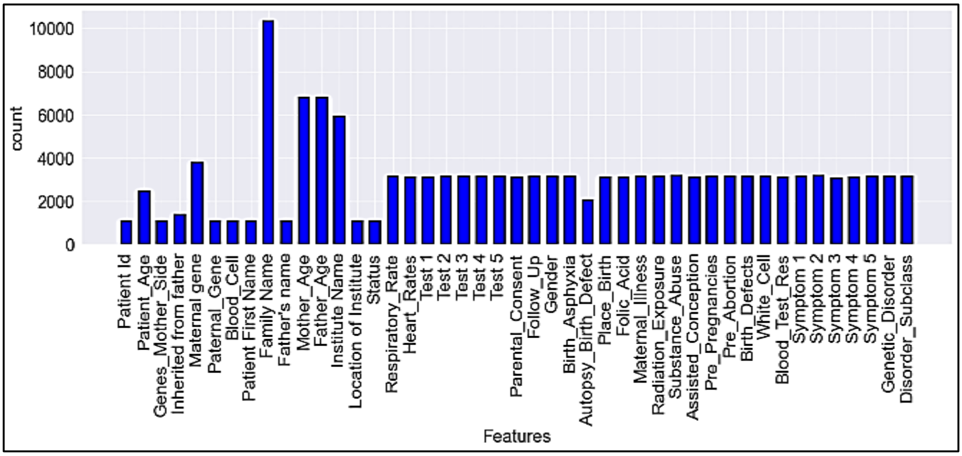


Figure 5: Null Values in a dataset

C) Feature Engineering

Feature engineering is useful for minimizing the complexity of prediction models, dealing with noisy and corrupted data, and reducing the size of the data set (Fan, 2019). Data is pre-processed by replacing the unclear values with NAN. The attribute “Birth asphyxia” holds values of “No record” and “Not available”, whereas “H/O substance abuse” and “H/O radiation exposure (x-ray)”, which have the values “Not applicable” and “-“, respectively are mapped with NAN. The handling of null values then involves categorical imputation for the following categories: "Genes Mother Side", "Inherited Father", "Maternal Gene", "Paternal Gene", "Status", "Respiratory Rate Breaths Min", "Heart Rates Min", "Follow Up", "Gender", "Birth Asphyxia", "Folic Acid", "Maternal Illness", "Radiation Exposure", "Substance Abuse", “Assisted Conception”, "History Previous Pregnancies", "Birth Defects", "Blood Test Result" and numerical imputation for the following : “Mother Age”, "Father Age", "Previous Abortion", "White Blood Cell", "Genetic Disorder" and "Disorder Subclass" . The dataset now contains 16146 and 9290 instances for training and testing, respectively, after the remaining null values are discarded.

Label Encoder normalizes the labels by transforming them into the machine-readable form. Figure 6 shows the mapping of the categorical values of genetic dataset to numerical form.

Feature Label	Feature value	Mapped Value	Feature Value	Mapped Value
Genes Mother Side	Yes	1	No	0
Maternal Gene	yes	1	No	0
Inherited Father	Yes	1	No	0
Paternal Gene	Yes	1	No	0
Status	Alive	1	Deceased	0
Respiratory Rate Breaths Min	Tachypnea	1	Normal (30-60)	0
Heart Rates Min	Tachycardia	1	Normal	0
Follow Up	High	1	Low	0
Gender	Female	1	Male	0
	Ambiguous	2	-	-
Birth Asphyxia	Yes	1	No	0
Folic Acid	Yes	1	No	0
Maternal Illness	Yes	1	No	0
Radiation Exposure	Yes	1	No	0
Substance Abuse	Yes	1	No	0
Assisted Conception	Yes	1	No	0
History Previous Pregnancies	Yes	1	No	0
Birth Defects	Singular	1	Multiple	0
Blood Test Result	Abnormal	0	Inconclusive	2
	Normal	2	Slightly abnormal	3
Genetic Disorder	Mitochondrial	0	Multifactorial	1
	Single-gene	2		
Disorder	cancer	0	Alzheimer's	1
	Cystic fibrosis	2	Diabetes	3
	Hemochromatosis	4	Leber's hereditary optic neuropathy	5
	Leigh syndrome	6	Mitochondrial myopathy	7
	Tay-Sachs	8		

Figure 6: Mapping of Categorical Values

D) ***Data Splitting***

Usually, data splitting is done to prevent overfitting of a model. The data is divided into training and test sets using the data splitting technique. For this study, the genetic inheritance disorder dataset is split into a ratio of 0.8:0.2.

E) ***Multi-layer Perceptron (MLP) Approach***

On utilized dataset of genetic inheritance disorder, multi-class multi-label classification problem is examined. Thus, we used MLP approach to solve the multi-class classification in our dataset. The Multi-layer Perceptron (MLP) is a form of feed-forward neural network used to solve multi-class classification problems (TensorFlow, 2022). A MLP classifier is used for deep learning model. The accuracy metrics is the performance metrics used for evaluation in this study.

4. RESULTS

The MLP deep learning classifier is evaluated on imbalanced dataset with a split of 0.8:0.2. The target class "Genetic Disease" obtains an accuracy of 0.810 and 0.781 for training and testing, respectively, according to the model's performance analysis. Whereas, label class “Disorder Subclass” obtained an accuracy of 0.933 and 0.917 for training and testing respectively. This analysis shows that our proposed approach helps us to achieve higher accuracy scores.

5. CONCLUSION

Deep learning algorithms can be helpful to interpret medical information to perform diagnosis and to assist experts timely. The occurrence of hereditary diseases is rising; thus, prediction is crucial to minimize the chance of deadly consequences. For this research work, a deep learning technique: Multilayer Perceptron (MLP) is utilized. Results show that the proposed technique produces best results. 0.810 and 0.933 training accuracy for Genetic Disorder and Disorder Subclass respectively. 0.781 and 0.917 testing accuracy for Genetic Disorder and Disorder Subclass respectively. This paper focuses on sequential model of MLP in deep learning. Also, in order to improve performance for genetic inheritance disorder prediction, we plan to use transfer learning approaches.

REFERENCES

1. Admin (2021). *Rescue mutations' that suppress harmful DNA changes could shed light on genetic disorders*, (Accessed: Feb. 26, 2023). Online available at http://resp.llas.ac.cn/C666/handle/2XK7JSWQ/327337
2. Admin (2023). *Genetic Diseases Review Articles| OMICS International| Journal of Hereditary Genetics: Current Research*, (Accessed: Feb. 26, 2023). Online available at https://www.omicsonline.org/genetics/genetic-diseases-review-articles.php
3. Admin (2023). *What is gene expression?*, (Accessed: Feb. 27, 2023). Online available https://www.yourgenome.org/facts/what-is-gene-expression/
4. Alatrany, A.S., Hussain, A., Jamila, M. and Al-Jumeiy, D. (2021). Stacked machine learning model for predicting alzheimer's disease based on genetic data. In 2021 *14th International Conference on Developments in eSystems Engineering (DeSE)*, Sharjah, United Arab Emirates, (pp 594-598), doi: 10.1109/DeSE54285.2021.9719449.

5. Atlam, M., Torkey, H., Salem, H. and El-Fishawy, N. (2020). A new feature selection method for enhancing cancer diagnosis based on DNA microarray. In 2020 *37th National Radio Science Conference (NRSC)* (pp. 285-295). IEEE.
6. Daghistani, T. and Alshammari, R. (2020). Comparison of statistical logistic regression and random forest machine learning techniques in predicting diabetes. *Journal of Advances in Information Technology*, 11(2), 78-83.
7. Dhanalaxmi, B., Anirudh, K., Nikhitha, G. and Jyothi, R. (2021). A Survey on Analysis of Genetic Diseases Using Machine Learning Techniques. In 2021 *Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 496-501). IEEE. doi: 10.1109/I-SMAC52330.2021.9640723.
8. Fan, C., Sun, Y., Zhao, Y., Song, M. and Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy*, 240, 35-45.
9. Ghazal, T.M., Al Hamadi, H., Umar Nasir, M., Gollapalli, M., Zubair, M., Adnan Khan, M. and Yeob Yeun, C. (2022). Supervised machine learning empowered multifactorial genetic inheritance disorder prediction. *Computational Intelligence and Neuroscience*, Article ID 1051388, 10 pages, https://doi.org/10.1155/2022/1051388
10. HackerEarth (2021). "*Of Genomes and Genetics: HackerEarth Machine Learning Challenge*," https://www.hackerearth.com/challenges/new/competitive/hackerearth-machine-learning-challenge-genetic-testing/ (accessed Feb. 26, 2023).
11. Kumar, A. (2021). Predict the genetic disorders dataset-of genomes. Online available at https://www.kaggle.com/datasets/aibuzz/predict-the-genetic-disorders-datasetof-genomes
12. Kumar, P., Radhakrishnan, J., Chowdhary, M.A. and Giampietro, P.F. (2001). Prevalence and patterns of presentation of genetic disorders in a pediatric emergency department. In *Mayo Clinic Proceedings*, 76(8), 777-783. Elsevier.
13. Mahdieh, N. and Rabbani, B. (2013). An overview of mutation detection methods in genetic disorders. Iranian Journal of Pediatrics, 23(4), 375-88.
14. Rahman, A.U., Nasir, M.U., Gollapalli, M., Alsaif, S.A., Almadhor, A.S., Mehmood, S., Khan, M.A. and Mosavi, A. (2022). IoMT-based mitochondrial and multifactorial genetic inheritance disorder prediction using machine learning. *Computational Intelligence and Neuroscience*, Article ID 2650742, 8 pages, https://doi.org/10.1155/2022/2650742
15. Raza, A., Rustam, F., Siddiqui, H.U.R., Diez, I.D.L.T., Garcia-Zapirain, B., Lee, E., and Ashraf, I. (2023). Predicting Genetic Disorder and Types of Disorder Using Chain Classifier Approach. *Genes*, 14(1), 71.
16. Rustam, F., Siddique, M.A., Siddiqui, H.U.R., Ullah, S., Mehmood, A., Ashraf, I. and Choi, G.S. (2021). Wireless capsule endoscopy bleeding images classification using CNN based model. *IEEE Access*, 9, 33675-33688. doi: 10.1109/ACCESS.2021.3061592.
17. TensorFlow (2022). *Multilayer perceptrons for digit recognition with Core APIs*, Online available at https://www.tensorflow.org/guide/core/mlp_core (accessed Feb. 26, 2023).

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 251-256

OPTIMIZATION TECHNIQUES TO ANALYZE INFLATION UNDER STATISTICAL METHODS AND MACHINE LEARNING APPROACH

Muhammad Faraz Ahsan[1] and **Ammara Nawaz Cheema**[2]
[1] Department of Creative Technologies, Air University, Islamabad Pakistan. Email: mf237862@gmail.com
[2] Department of Mathematics, Air University, Islamabad, Pakistan Email: ammara.au@gmail.com; ammara.cheema@mail.au.edu.pk

ABSTRACT

Highly dimensioned data becoming continuously more progressive and representable with the help of statistical tools. Financial data analytics (Fin Tech) and econometrics emerges the data science applications. Inflation data treated as an important piece of economic information. The major challenge in econometrics is economic data analysis which concern basically different configuration of data. Heuristic data analytic process is proposed using machine learning techniques and statistical methods for inflation data time stamp and financial modelling in economic system. World Bank inflation data is used for the modelling in economic applications. As cross-section and panel data econometrics and time series analysis are the two major branches of econometrics. The findings indicate the directions which followed the advancement of using multiple learning algorithm of machine learning for econometric applications. These algorithms made a progress in analysis of largely unknown, complex and dependent pattern of data.

KEYWORDS

Econometrics, Inflation, Statistical methods, Machine Learning, Time series data, Economic System Analysis.

1. INTRODUCTION

Different dataset has multidimensional and complex structure in modern applied sciences and is represented in multidimensional arrays but machine learning and statistics provides an effective and efficient background for modeling big data problems. In this article we proposed to apply machine learning and data science techniques to analyze inflation data. We took World Bank inflation data from 1970 to 2021. Econometrics is the study of economic relationships using mathematical models. It can be used to analyze data, to forecast future trends, and to make economic decisions. Econometrics is a relatively new field, and there is still much to learn. Inflation is a general increase in the prices of goods and services in an economy over a period of time (Abadir and Paruolo, 1997). It can be caused by many factors, including increases in the cost of goods, increases in the amount of money in circulation, and increases in the demand for goods and services. Inflation can have a negative impact on the economy by making it more difficult for people to afford goods and services, and by leading to a decline in the value of money. Statistical methods are used to analyze economic data. This can include looking at trends, measuring the impact of different variables on economic outcomes, and estimating the effects of policy changes (Aldasoro and Alves, 2018). Statistical methods can be used to analyze data from a variety of sources, including surveys, administrative records, and economic datasets. Machine learning techniques can be used to improve the accuracy of economic predictions. By using machine learning algorithms, economists can improve the accuracy of predictions by identifying patterns in economic data that they would not be able to see without the help of a machine. By identifying these patterns, economists can make more accurate predictions about future economic trends. Data science is a rapidly growing field that is constantly evolving (Anacleto and Queen, 2017). There are many different data science trends that are constantly changing, which can make it difficult to keep up with all the new developments. However, some of the most popular data science trends include big data, machine learning, and artificial intelligence. The tensor valued data is available increasingly of long temporal sequence like multidimensional tables but there are limitations in multivariate time series models [5]. Many research in this field more focused on analysis of cross-sectional data which apply on neuroimaging problems like FMRI (Functional Magnetic Resonance Image) and also signal processing, but the time series analysis literature is unique on its own way. Often tensor value covariate normally used to predict scalar outcomes and few research analyze tensor regression models. Parameter regularization are required in estimation of tensor regression. The number of entries of coefficient are larger than sample size. Guhaniyogi, Qamar and Dunson (2017) design the predictive model in cross sectional stream in order to investigate the relation between matrix valued brain images and scalar medical index, whereas Li et al. proposes low-rank approximation to tensor coefficient by using tucker decomposition. There is an alternate approach Tensor on-vector which is tensor-on-vector and is use for accessing the impact of vector factors on tensor valued observations. Behera, Nandi and Sahoo, (2020) introduces an autoregressive tensor model that generalizes existing tensor framework. This model provides dynamics in linear tensor with tools to analyze the shock propagation in dynamic system. High dimensional models' sparse estimation and parameter regularization can also be achieved with alternative approaches. The alternative approach assumes coefficient tensors reduced rank. Data modelling literature increased rapidly due to financial crisis on both side theoretical and also on empirical analysis (Dawid and Lauritzen, 1993). Natural framework for financial multilayer data analysis uses dynamic tensor models. International trade networks data consist of yearly snapshot collection is an example of time series data network. Although dynamic models may be more suitable for studying network data collected over time, most statistical models for network data have remained static to date. Several attempts have been made to model time-varying networks, and most existing approaches focus on providing a representation and description of time-evolving graphs. We contribute to this literature by providing an original study of time-varying economic and financial networks and show that our dynamic tensor model can be successfully used to perform impulse response analysis in a multivariate setting.

2. METHODOLOGY

Econometrics is a well-established field that involves the measurement and statistical inference of economic phenomena. So in starting we have the raw data we start some preprocessing on it, In data preprocessing we deal with the outliers and missing values, then we remove string columns and finally we have only two things including country code alongwith the yearly data from 1970s to 2021. We did data visualization by plotting different graphs. We split the data into training and testing in which we set 33 percent data for testing and remaining for training after this we applied machine learning algorithm naive bayes algorithm. For the prediction we use a function using a command. In the whole scenario we have dataset constraints that if we have more data then we can achieve more better accuracy with improved performance.

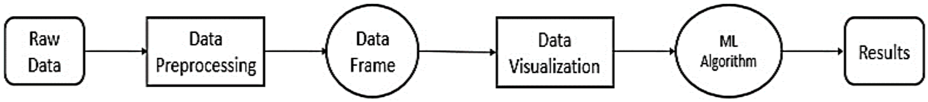


Figure 1: Block Diagram shows the Flow of Data

Financial econometrics, in turn, refers to the application of statistical techniques to problems in finance. While the definition of econometrics is widely accepted, the formal definition of financial data science is still emerging. Financial data science can be defined as an interdisciplinary process of scientific inquiry that uses statistical techniques to explore and explain financial data sets in order to advance financial decision making. Financial data science differs from econometrics in several ways. While econometrics' intellectual point of departure is statistical inference, financial data scientists focus on exploring and explaining data sets to advance financial decision making. Financial data science also emphasizes the importance of interdisciplinary collaboration, as it typically requires the expertise of financial econometricians, computer scientists, and individuals with deep knowledge about financial markets. Despite the differences between the two fields, econometrics and financial data science share many similarities. Both fields use econometric concepts and techniques, and both are likely to make use of big data resulting from digitalization and an increased willingness of commercial organizations to share proprietary data sets with academics. Furthermore, both fields are likely to experience an increased practical relevance due to their analysis of bigger and more often proprietary datasets. The emergence of financial data science as a distinct field of inquiry reflects the growing importance of big data in finance. Financial data scientists seek to leverage the insights gained from the exploration and explanation of large, complex datasets to improve financial decision making. This requires interdisciplinary collaboration and the use of advanced statistical techniques, including machine learning and artificial intelligence.

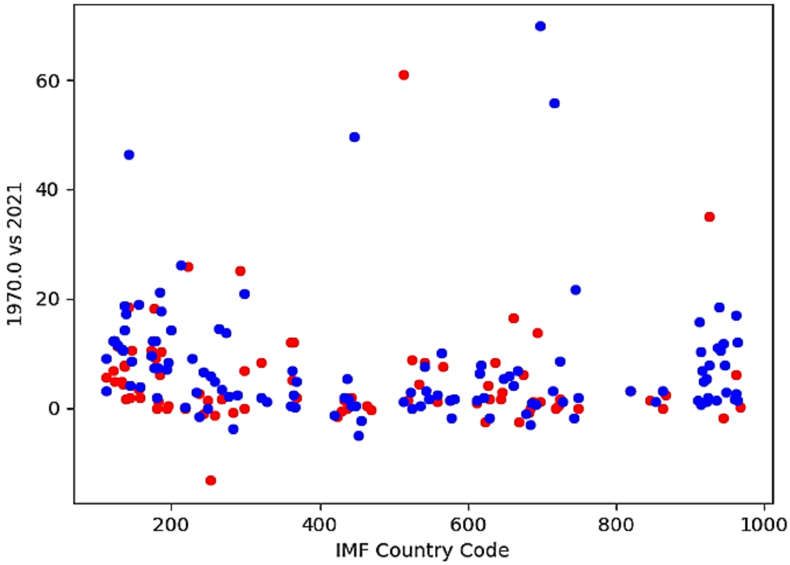


Figure 2: Scatter Plot Shows the Frequency of Inflation Occurrence with IMF Country Code

Financial data science has many practical applications in finance, including risk management, portfolio optimization, and algorithmic trading. The ability to process and analyze vast amounts of financial data has enabled financial institutions to make more informed decisions and to identify new opportunities for growth and innovation. One of the key challenges facing financial data scientists is the availability of high-quality data sets. While the move to big data is unstoppable, the quality of available data sets remains a key practical constraint for empirical researchers. Financial data scientists must be able to extract meaningful insights from unstructured data, process large data sets efficiently, and design and execute effective statistical analyses. To be effective, financial data science requires a strong interdisciplinary team spirit. No single member of the team should insist on the idiosyncratic attributes of their discipline being more worthy or truthful than another discipline's idiosyncratic attributes. Financial data scientists must be able to work collaboratively and share knowledge and expertise across disciplinary boundaries. Financial data science also has broader societal ambitions. Financial data scientists aim to leverage financial and computer science for the broader good, with a focus on advancing financial decision making and enlightening society as a whole. This requires a commitment to scientific rigor and a willingness to engage with the broader societal implications of their work. Econometrics and financial data science represent complementary perspectives on data-driven financial decision making. While econometrics has a greater emphasis on statistical inference, financial data science focuses on big data processing. However, both fields share many concepts, and both are likely to make use of increasingly large and complex data sets. The emergence of big data in finance has created many new opportunities for econometricians and financial data scientists. These opportunities include new analytical measurement techniques and new financial products that leverage the insights gained from the exploration and explanation of large data sets. In conclusion, econometrics and financial data science are two complementary fields that share a common goal of using statistical techniques to advance financial decision making. While they differ in their intellectual point of departure, process, and ambitions, they both use econometric concepts and techniques and are likely to make use of the increasing amount of human data resulting from digitalization. The emergence of financial data science as a distinct interdisciplinary field reflects the need for a specialized approach to studying the unique distributions that are produced by many financial markets. However, financial data science is not just an application of data science methods to finance but a discipline in its own right. As both fields continue to evolve and expand, it is important to maintain a collaborative and interdisciplinary approach in order to extract insights and maximize their impact on society. The increasing availability of high-quality data sets and the growing interest in leveraging financial and computer science for the broader good present's new analytical measurement opportunities and financial products that have the potential to transform the way we think about finance and decision-making.

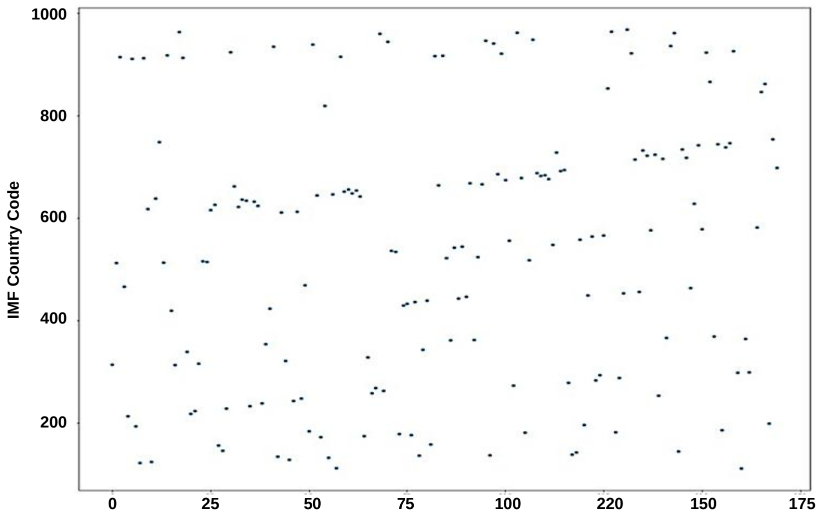


Figure 3: Visualized Data Dots Show Inflation Rate Value Range

3. RESULTS AND DISCUSSION

We provide a framework of inflation dataset by applying data preprocessing techniques on it. We plot different graph using countries ranges point. By analyzing the data using machine learning techniques, it is realized that before we provide our data for training and testing its distribution should be proper and in structured form. Now as we split data into training and testing when we apply machine learning algorithm, we achieve accuracy 55%.

4. CONCLUSION

Applying machine learning techniques on financial dataset improve the analytical process. As we have the dataset limitation, the more data can increase the accuracy and also improve the performance of algorithm on financial data. Ensemble learning can also apply on financial dataset in order to achieve more better accuracy

REFERENCES

1. Abadir, K.M. and Paruolo, P. (1997). Two mixed normal densities from cointegration analysis. *Econometrica: Journal of the Econometric Society*, 65, 671-680.
2. Abraham, R., Marsden, J.E. and Ratiu, T. (2012). *Manifolds, tensor analysis, and applications* (Vol. 75). Springer Science & Business Media.
3. Aldasoro, I. and Alves, I. (2018). Multiplex Interbank Networks and Systemic Importance: An Application to European Data. *Journal of Financial Stability*, 35, 17-37.
4. Anacleto, O. and Queen, C. (2017). Dynamic Chain Graph Models for Time Series Network Data. *Bayesian Analysis*, 12, 491-509.
5. Arnott, R.D., Harvey, C.R. and Markowitz, H. (2019). A Backtesting Protocol in the Era of Machine Learning. *The Journal of Financial Data Science*, 1, 64-74.
6. Asimakopoulos, P., Asimakopoulos, S. and Fernandes, F.D.S. (2019). Cash holdings of listed and unlisted firms: New evidence from the euro area. *The European Journal of Finance*, 25(17), 1708-1729.
7. Bajgrowicz, P. and Scaillet, O. (2012). Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics*, 106(3), 473-491.
8. Balazsi, L., Matyas, L. and Wansbeek, T. (2018). The estimation of multidimensional fixed effects panel data models. *Econometric Reviews*, 37(3), 212-227.
9. Behera, R., Nandi, A.K. and Sahoo, J.K. (2020). Further Results on the Drazin Inverse of Even Order Tensors. *Numerical Linear Algebra with Applications*, 27(5), e2317.
10. Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.
11. Dawid, A.P. and Lauritzen, S.L. (1993). Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models. *The Annals of Statistics*, 21, 1272-1317.
12. De Paula, A. (2017). Econometrics of Network Models. In B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson (Eds.) *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, pp. 268-323, Cambridge University Press.
13. Del Brio, E.B., Mora-Valencia, A. and Perote, J. (2019). Expected shortfall assessment in commodity (L) ETF portfolios with semi-nonparametric specifications. *The European Journal of Finance*, 25(17), 1746-1764.
14. Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70, 1741-1779.
15. Guhaniyogi, R. (2020). *Bayesian Methods for Tensor Regression*. Wiley StatsRef: Statistics Reference Online, 1-18. https://tr.soe.ucsc.edu/sites/default/files/technical-reports/UCSC-SOE-20-09.pdf
16. Guhaniyogi, R., Qamar, S. and Dunson, D.B. (2017). Bayesian tensor regression. *The Journal of Machine Learning Research*, 18(1), 2733-2763.
17. Jackson, R.H. (2019). Sub-sequence incidence analysis within series of Bernoulli trials: application in characterisation of time series dynamics. *The European Journal of Finance*, 25(17), 1730-1745.
18. Kumar, G., Muckley, C.B., Pham, L. and Ryan, D. (2019). Can alert models for fraud protect the elderly clients of a financial institution? *The European Journal of Finance*, 25(17), 1683-1707.
19. Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228-250.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 257-262

ANALYSIS OF FOOD CROPS OF PAKISTAN TO AID AGRICULTURE AND FOOD SECURITY

Mahnoor Athar[1] and **Ammara Nawaz Cheema**[2]
[1] Department of Creative Technologies, Air University, Islamabad Pakistan. Email: m23athar@gmail.com
[2] Department of Mathematics, Air University, Islamabad, Pakistan Email: ammara.au@gmail.com; ammara.cheema@mail.au.edu.pk

ABSTRACT

In a world where our food production cannot sustain the growing population, now is a time as ever that we need to strategize to bring policymakers together to make decisions that will ensure food security for the coming years. In this paper, I have conducted a basic analysis of the various crops grown in Pakistan; the production versus cultivated area. The analysis is primarily based on stationary graphs of both food and cash crops of Pakistan. This will provide a base for further analysis of the same data.

1. INTRODUCTION

Global food insecurity is at an all-time high right now. With the rise in inflation all around the world due to climate change and the Russia-Ukraine War, cost of food is sky-rocketing. Up to 828 million people don't know where they will get their next meal. Providing food for these people has become a task of the utmost priority. Many African countries and war torn regions have little to no access to food and are on the brink of famine. Even those who have access to food are severely malnourished which can be attributed to scarce and poor food quality.

Another reason for high food prices is the COVID-19 pandemic which caused a worldwide economic deficit. There is a long history of fatal viruses emerging on our planet but not as deadly and contagious as COVID-19. The pandemic alone had devastated the global food supply chain, from the primary supply to the final demand. As the world becomes more interconnected to each other as a global village, it is more likely that all countries would be vulnerable to any disaster. Therefore, failure to prevent such a situation in one country means the whole world is exposed as well. For example, the Russia-Ukraine war has caused fuel and food shortages all across the Europe.

The situation is dire and exacerbated in the developing countries where meeting the basic livelihood needs of a majority of the population who are underprivileged is impossible. This calls for making use of the limited land resources intelligently to produce grains to feed the entire global population. Using smart agriculture can help overcome this problem. With the use of Artificial Intelligence (AI), Internet of Things (IoT) and Machine Learning (ML), a holistic approach can be determined to increase yield of crops. The rise of wireless and remote sensors has made it incredibly easy to acquire data regarding the agricultural land, giving rise to precision agriculture. The exponential growth of digital data is the primary factor behind the recent set of advances in the AI sector. With this vast amount of data and trainable AI startups, the fourth industrial revolution (4.0 IR) is on its way through creative solutions to challenges that are currently being faced in many different industries. The food industry is one of many that has recently had a significant impact from AI on its practises, equipment, and machines. The advent of AI-driven processes and machines into agriculture and the food business has transformed how crops are farmed, cultivated, produced, and processed.

Vast amounts of data are required to train AI and ML models and this has been made possible with the advent of hi-tech sensors and imaging tools and techniques have produced exabytes of data to date. Big data from agriculture has been used to estimate the financial effects of natural disasters on food production and evaluate crop performance across regions in a given field season.

All this discussion leads to one goal; to be able to provide food for the growing population of the world. Pakistan is one the developing nation where devastation due to climate change induced rainfall and consequent flooding has left many with food or shelter. A huge chunk of the total cultivatable area is underwater causing further crippling the supply chain. The motivation behind this study is to carry out a basic analysis on the different crops grown in Pakistan. Understating what is the trend of production of a particular crop compared to the area it is cultivated on and establishing what the future course of action will be.

2. METHODOLOGY

The main goal is to provide a bird’s eye view of the data to assist further analysis like model fitting and forecasting. The study is a simple one where not a lot of steps are involved. The main framework is the data acquisition, data analysis, interpretation and results. The data I am working on was acquired from the website of Pakistan Bureau of Statistics. On their website, a record of 28 years, from 1981-2009 is accessible. The data is divided into food crops and cash crops; food crops being ones which are necessary for survival and cash crops being a source of export and revenue. The data is provided in different forms for example, area and production of food crops in Pakistan, province-wise distribution of crops etc. This can help if different types of analysis.

The basic or the descriptive analysis of the data for this paper includes the following:

- Basic statistics; this includes mean, standard deviation and inter interquartile range (contains the second and third quartiles, or the middle half of your data set.)
- Skewness; who symmetric the data is
- Kurtosis; how flat the data is

All analysis mentioned in this study were carried out using the help of the statistics, numpy, pandas and matplotlib packages and the Python programming language in Anaconda environment using Jupyter Notebook.

3. ANALYSES

I began my analyses by determining whether the data contained null values or not. After this, basic statistics where calculated for both area and production. This is shown in Table 1. Similarly, descriptive statistics for production is calculated as show in Table 2.

Table 1
Descriptive Statistics of Area Cultivated in Food Crops in Pakistan

Crop	Mean	Standard Deviation	25%	50%	75%
Wheat	8015.336	468.3113	7723.75	8083.3	8355.45
Rice	2233.261	260.1539	2030.9	2143.2	2432.85
Maize	896.8036	84.52045	846.925	874.8	950.75
Jowar	367.4893	50.74302	349.725	382.3	395.9
Bajra	448.5893	91.508	379.525	466.15	510.4
Barley	147.071429	43.34701	109.875	151.35	166.65

Table 2
Descriptive Statistics of Production of Food Crops in Pakistan

Crop	Mean	Standard Deviation	25%	50%	75%
Wheat	16807.41	3739.119	14217.38	16778.95	19262.43
Rice	4129.457	996.4211	3301.6	3924.25	4813.85
Maize	1640.775	809.6701	1166.2	1255.3	1682.95
Jowar	218.125	28.51262	209.875	222	236.15
Bajra	213.625	47.90002	192.275	212.1	242.55
Barley	128.0821	29.70335	99.425	132.6	146.775

Next in the basics we have skewness and kurtosis plots for individual plots.

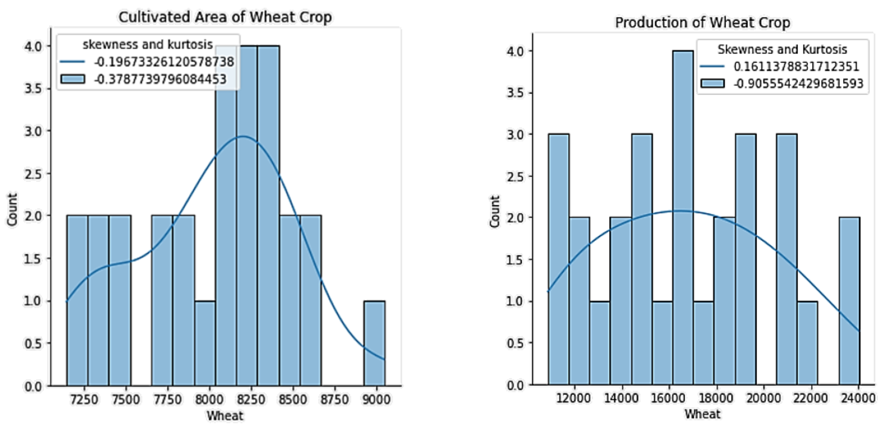


Figure 1: Skewness and Kurtosis Plot for Wheat Crop

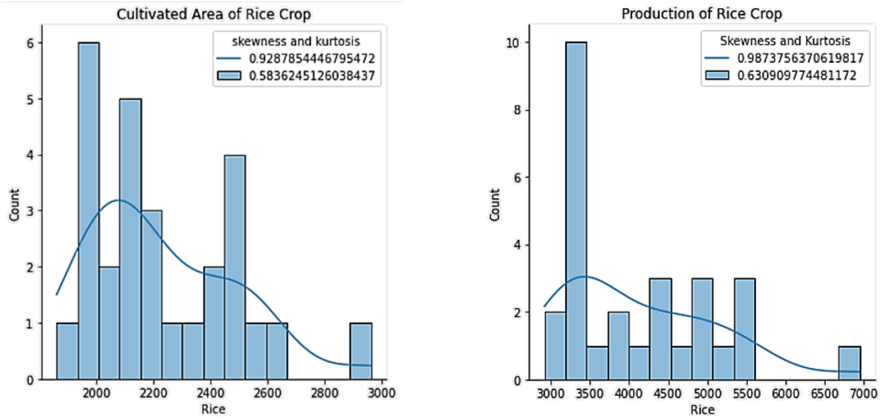


Figure 2: Skewness and Kurtosis Plot for Rice Crop

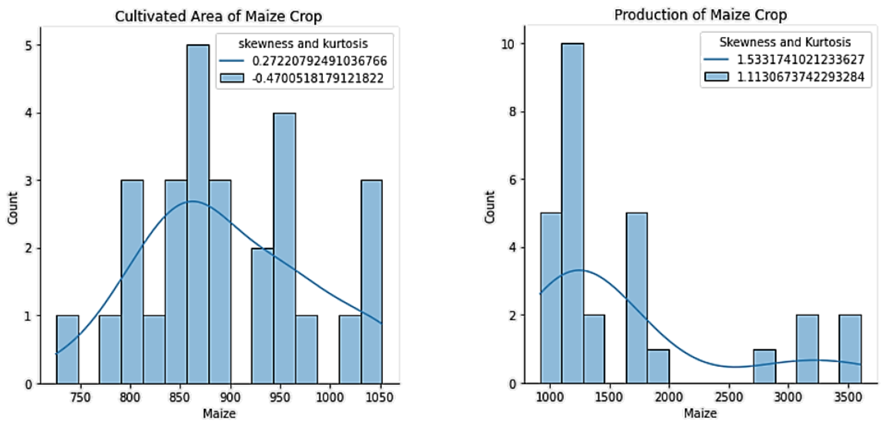


Figure 3: Skewness and Kurtosis Plot for Maize Crop

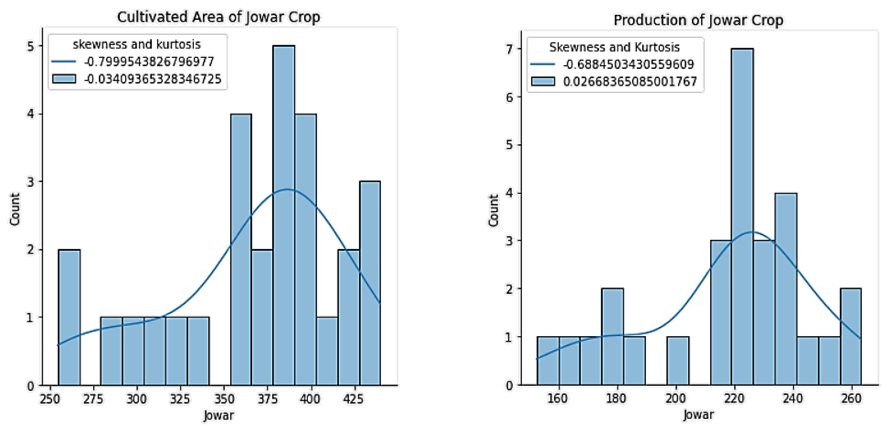


Figure 4: Skewness and Kurtosis Plot for Jowar Crop

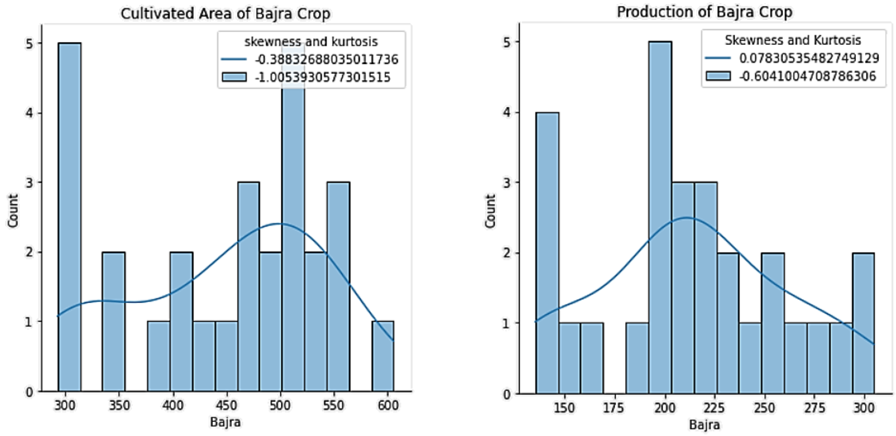


Figure 5: Skewness and Kurtosis Plot for Bajra Crop

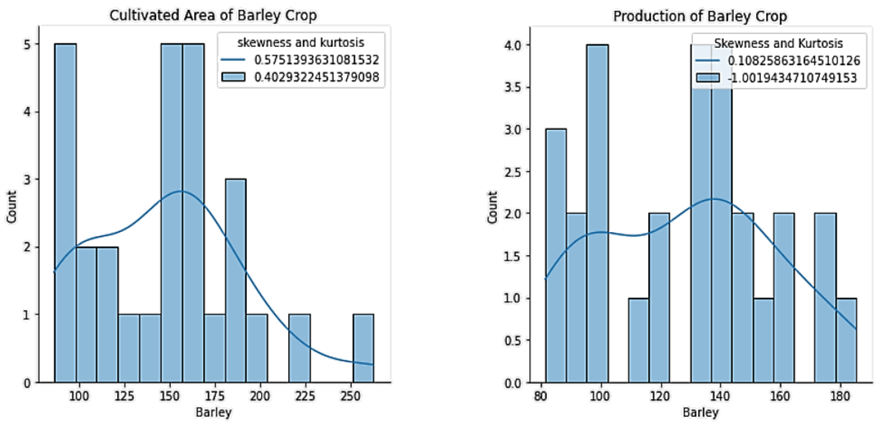


Figure 6: Skewness and Kurtosis Plot for Barley Crop

Figures 1-6 show the skewness and kurtosis plots for production and cultivatable area of the crops. Most of these plots are shown be not normal and different transformations can be to normalize the data which can help with further analysis.

4. COMMENTS AND CONCLUSION

Through the course of this paper, we saw a basic analysis of the production and area dataset. This work can be extended to model fitting and forecasting. Since the original data is time series data, ARMA and ARIMA models can be used for predictions and forecasting. Further ANOVA and chi-square tests can be used to determine goodness of the fitted model.

5. ACKNOWLEDGEMENT

I am extremely thankful to, Dr. Ammara Nawaz Cheema for her support and guidance during my research work. Her motivation, patience, ideas, constant support and immense knowledge considerably helped me to complete this paper. And my family, whose prayers and support have helped me to achieve this milestone.

REFERENCES

1. Goel, R.K., Yadav, C.S., Vishnoi, S. and Rastogi, R. (2021). Smart agriculture – Urgent need of the day in developing countries. *Sustainable Computing: Informatics and Systems*, 30, 100512.
2. Kakani, V., Nguyen, V.H., Kumar, B.P., Kim, H. and Pasupuleti, V.R. (2020). A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, 2, 100033.
3. Khan, S.B., Ahmad, F., Sadaf, S. and Kashif, R.H. (2010). *Crops area and production (by districts) (1981-82 to 2008-09)*. Volume I, Food and cash crops, Government of Pakistan, Statistics Division, Federal Bureau of Statistics (EconomicWing), Islamabad.
4. Khanna, A. and Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Computers and Electronics in Agriculture*, 157, 218-231.
5. Ramirez-Asis, E., Bhanot, A., Jagota, V., Chandra, B., Hossain, M.S., Pant, K. and Almashaqbeh, H.A. (2022). Smart Logistic System for Enhancing the Farmer-Customer Corridor in Smart Agriculture Sector using Artificial Intelligence. *Journal of Food Quality*, Article ID 7486974, https://doi.org/10.1155/2022/7486974
6. Shrivastava, S. and Marshall-Colon, A. (2018). Big data in agriculture and their analyses. In *Encyclopedia of Food Security and Sustainability* (pp. 233-237). Elsevier.
7. Workie, E., Mackolil, J., Nyika, J. and Ramadas, S. (2020). Deciphering the impact of COVID-19 pandemic on food security, agriculture, and livelihoods: A review of the evidence from developing countries. *Current Research in Environmental Sustainability*, 2, 100014.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 263-270

CONSEQUENCES OF GREEN HRM PRACTICES IN THE BANKING SECTOR OF PAKISTAN

Nadia Hanif[1], **Ali Junaid Khan**[2]**, Muhammad Adnan Sial**[1]**,**
Noman Arshed[1] and **Tanzila Abdul Karim**[3]

[1] Division of Management and Administrative Science
University of Education, Lahore, Pakistan

[2] Institute of Business, Management and Administrative Science
The Islamia University of Bahawalpur, Bahawalpur, Pakistan

[3] Federal Urdu University Science and Technology,
Islamabad, Pakistan

ABSTRACT

The degree to which a company successfully implements environmentally friendly programs indicates environmental performance. The authors investigated green human resource management (GHRM) practices, including green rewards and training for employee satisfaction, commitment, and performance. The GHRM practices were measured with green training and green rewards for the employees. As part of the study's quantitative methodology, a survey questionnaire was distributed to Lahore Bank customers, and SPSS was applied to evaluate the results of the hypotheses. Findings indicated that the data supported the study’s hypotheses. HR managers can use green training programs to teach staff members the value of sustainability and their actions’ effects on the environment. Employees can apply sustainable practices using the information and abilities acquired through green training programs.

1. INTRODUCTION

According to Marhatta & Adhikari, 2013, GHRM is the relevance to human resource management doctrines to advance resource sustainability within organizations while also advancing environmental sustainability. Several definitions of "green management" for long term sustainable development can be found in the environmental literature; however, they all aim to explain the need for stability between business development for prosperity and environmental preservation for the benefit of existing and forthcoming generations (Daily and Huang, 2001). According to Lober (1996), a few indicators, such as low environmental issues, pollution management, minimization of waste, and reusing activity, can be used to gauge an environment's success. Environmental protection can also refer to how a company's actions and products affect the environment, including resource use, waste production, and emissions. A product's Environmental Performance (EP) shows a corporation’s dedication to environmental preservation (hereinafter called "environment"). GHRM is often defined as "the use of HRM practices to encourage the sustainable utilization of resources within the organizations and, more broadly, supports environmental sustainability," according to many authors"(Marhatta & Adhikari, 2013). Organizational

Performance (OP) can be improved and increased by implementing GHRM, such as green recruiting and selection, green development and training of employees, and green learning programs. Although recent studies (e.g. Jackson and Seo 2010; Renwick et al. 2013) emphasized on the substantial role of HRM practices in improving higher levels of environmental outcomes, this field of study is still in its early stages. Paillé (2014) discovered a significant linkage among OP, EP, and HRM practices. The GHRM may improve employee commitment, performance, and work satisfaction. Although much work has been done on GHRM and its antecedents, more studies are encouraged (Dumont et al., 2017; Pham et al. 2019). The lack of research and analysis on how green HRM affects employees and organizations is another challenge. Due to the necessity to create an efficient system, it is vital to comprehend the effects of GHRM on the workforce (Ren et al., 2018). The results of applied green HRM. Because it advances understanding of green HRM, our study emphasizes effects at the individual and organizational level (i.e., Employee productivity, job Satisfaction (JS), and job commitment (JC).

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

2.1 Rewards and Green HRM

Organizations today are taking action to improve environmental performance by implementing reward systems (Renwick, Redman and Maguire, 2013). While it inspires workers and fosters a collaborative environment for solving problems, a reward system promotes self-esteem, respect and acknowledgement for the workforce (Jabbour and Santos, 2008). Financial incentives are the main motivator for encouraging staff to participate in eco-initiative projects. According to research, financial incentives significantly impact individuals’ satisfaction and motivation at work.

GHRM enhances an organization's environmental performance while impacting employee productivity and work satisfaction (Chan and Hawkins, 2010). According to a study, an important factor influencing an employee's conduct at work is how they rate the quality of their job (Yusoff et al., 2018). Pride, recognition, involvement, advancement, fairness, self-actualization, the work environment and the task itself is just a few employment factors that can affect how people view their jobs and, as a result, their dedication (Arnett et al., 2002). By fostering green employees who are concerned about environmental issues and green workplace culture, GHRM, as a strategy, helps firms achieve their environmental goals (Kim et al., 2019). Thus, GHRM initiatives like green awards can boost employee commitment to the job and job happiness.

H1: Green HRM Practices (e.g., green rewards) positively impact employee job satisfaction.

GHRM procedures support or uphold the organization's dedication (Jyoti, 2019). This connection might be understood as an individual’s commitment to environmental care. Employees who believe that a company's entire operation is devoted to environmental preservation do their part by displaying their loyalty to the business. Hence, one could argue that using GHRM techniques increases and strengthens company and employee commitment (Yusliza et al., 2019).

H2: Green HRM Practices (e.g., green rewards) positively impact employee commitment to the Job.

2.2 Training and Green HRM

The second crucial component of GHRM is green training, which promotes employees' knowledge of environmental issues and helps them connect business decisions to environmentally friendly activities. It enables people to use their professional skills to conform to social norms (Bansal and Roth 2000). Waste reduction was one of these skills named by May and Flannery (1995), along with knowledge regarding environmental protection and environmental education (Roy and Therin 2008). Workers deal with data management daily, and this training equips them with the managerial abilities they need to manage waste as part of their regular duties. There is a clear and noteworthy link between training of employees and development and organizational performance because it equips workers with the information and skills needed to achieve organizational objectives and provide sustainable performance. Innovations occur in an organization when skill development and inquisitive learning are coupled (Zakaria, 2013). Implementing the organization’s environmental management system includes environmental training (Delmas and Pekovic, 2013). Hence, environmental training is a methodical procedure that equips staff with the knowledge, attitudes, and skills necessary to achieve senior management's environmental management objectives (Daily et al., 2012). According to Renwick et al. (2013), ecological training gives staff members environmental knowledge and attitudes that enhance their environmental performance. Employees who receive environmental education and training at their place of employment are more likely to feel liable for environmental outcomes (Ramus, 2002). Training makes employees more aware of the company's environmental objectives (Jackson et al., 2011). As a result, environmental training improves staff attitudes towards environmental management practices, including waste management and pollution prevention (Hart and Ahuja, 1996). Offering environmental training to organizational employees and members (managers and non-managerial staff) to develop the necessary knowledge and skills is one of GHRM's key goals. This will make it easier to put business environmental management initiatives into action. Training to promote waste management practices that support telecommuting and flexible scheduling.

H3: Green training positively influences employee job satisfaction.

Employees feel more independent in carrying out environmental measures when given the freedom and encouragement to discuss their environmental goals. By evaluating, praising, and satisfying their environment-related performance, they may also be able to comprehend the genuine impact of their actions. Four stages are required to implement GHRM practices: creating an environmental vision to serve as a guide; teaching employees to communicate their environmental objectives, mission, and visions; evaluating employee environmental performance; and rewarding employees for their environmental efforts (Clair et al., 1996). The essential elements of the job characteristics model are linked with the phases for implementing GHRM. By establishing a common environmental vision, mission and goals, and providing environmental awareness training, GHRM increases skill variety, task identity, and work relevance.

H4: Green training positively influences employee job commitment.

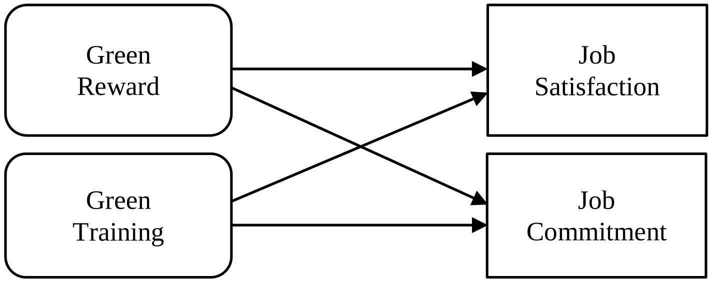


Figure 1

3. DATA, METHODOLOGY AND RESULTS

Data were composed from employees of the Lahore banking sector in Pakistan to test the theoretical model. The questionnaire was distributed to the target groups from May to July 2022.

3.1 Research Design

This study aimed to find a way to reconcile the interests of businesses and society without compromising business objectives. This research investigates the importance of green rewards and training as GHRM practices for employees' job-related satisfaction and commitment.

3.2 Population and sample

The studied population was the employees of the banking sector of Pakistan operating in Lahore. The construct measures of the study have been verified by earlier studies (Khan et al., 2019; Kraus et al., 2020; Rehman et al., 2019b). We used a straightforward random sample method to gather the data (Sekaran and Bougie, 2016). To the staff at each bank, we gave out 300 surveys. Our request for survey participants from the staff in a number of banks was not heeded. Two weeks later, we returned to these banks after giving reminders. 250 questionnaires were ultimately returned because of this. In total, 252 survey questions were examined.

3.3 Methods

To do the analysis of the study to test the connection amid dependent and independent variables, we used SPSS, and we did descriptive analysis and inferential statistics.

4. RESULTS AND DISCUSSION

Table 4.1
Descriptive Statistics

	Mean	Median	S. D
Job Satisfaction	2.219	2.000	0.895
Job Commitment	2.619	2.667	0.914
Green Reward	2.632	2.667	1.109
Green Training	2.056	2.00	0.957

Referring to Table 4.1, most respondents either strongly disagree or disagree with the statements measuring JS, job commitment, green reward, and green training.

Table 4.2
Correlations Analysis

	Job Satisfaction	Job Commitment	Green Reward	Green Training
Job Satisfaction	1	0.613**	0.456***	0.512***
Job Commitment		1	0.432***	0.233***
Green Reward			1	0.387***
Green Training				1

$N = 250$, $p < 0.05$**, $p < 0.01$***

Table 4.2 shows that JS, job commitment, green reward, and green training have positive and significant associations at $p < .01$. Whereas job commitment and JS have a positive and significant association with each other at $p < .05$.

Table 4.3
Regression Results of GHRM (Reward and Training) Practices and Job Satisfaction and Job Commitment

	Model 1 Job Satisfaction	Model 2 Job Commitment
Green Reward	0.304*** (0.045)	0.402*** (0.051)
Green Training	0.394*** (0.052)	0.077 (0.059)
Constant	0.815*** (0.134)	1.595*** (0.152)
F-S	63.714***	29.249***
R2	0.340	0.43

$N = 250, p < 0.10*, p < 0.05**, p < 0.01***$

Model 1 of Table 4.3 shows that green reward positively influences JS. Hence, H1 in this study is accepted. Model 1 of Table 4.3 illustrates that training is positively significant at $P < 0.001$; this shows that green training increases the JS of banking employees. Therefore, H2 of the study is accepted. Model 2 of Table 4.3 shows that green reward is significant at $P < 0.001$, which demonstrates that H3 is accepted, and green reward increases the commitment of banking employees. Whereas H4 is not accepted as green training does not impact employee commitment, as $P > 1.0$ level of significance.

5. COMMENTS AND CONCLUSION

This study's primary objective was to investigate the effects of GHRM methods, such as green training and rewards, from several angles, including JS and job commitment in the banking industry. A survey-based approach was chosen to accomplish the research goals, and statistical analysis was conducted using SPSS software. The results demonstrate that GHRM approaches, such as green training and rewards, benefit JS and corporate commitment.

6. ACKNOWLEDGEMENT

All the acknowledgement goes to our parents and family, who are pillars in our life and achievements. They are a source of encouragement, motivation, direction and support.

REFERENCES

1. Arnett, D.B., Laverie, D.A. and McLane, C. (2002). Using job satisfaction and pride as internal-marketing tools. *Cornell Hotel and Restaurant Administration Quarterly*, 43(2), 87-96.
2. Bansal, P. and Roth, K. (2000). Why companies go green: a model of ecological responsiveness. *Academy of Management Journal*, 43, 717-736.

3. Chan, E.S. and Hawkins, R. (2010). Attitude towards EMSs in an international hotel: an exploratory case study. *International Journal of Hospitality Management*, 29(4), 641-651.
4. Clair, J.A., Milliman, J. and Whelan, K.S. (1996).Toward an environmentally sensitive ecophilosophy for business management. *Industrial and Environmental Crisis Quarterly*, 9(3), 289-326.
5. Daily, B.F. and Huang, S. (2001). Achieving sustainability through attention to human resource factors in environmental management. *International Journal of Operations & Production Management*, 21(12), 1539-1552.
6. Delmas, M.A. and Pekovic, S. (2013). Environmental standards and labor productivity: Understanding the mechanisms that sustain sustainability. *Journal of Organizational Behavior,* 34(2), 230-252.
7. Dumont, J., Shen, J. and Deng, X. (2017). Effects of green HRM practices on employee workplace green behavior: the role of psychological green climate and employee green values. *Human Resource Management*, 56(4), 613-627.
8. Hart, S.L. and Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business strategy and the Environment*, 5(1), 30-37.
9. Jabbour, C.J.C. (2013). Environmental training in organisations: From a literature review to a framework for future research. *Resources, Conservation and Recycling*, 74, 144-155.
10. Jackson, S.E. and Seo, J. (2010). The greening of strategic HRM scholarship. *Organization Management Journal*, 7(4), 278-290.
11. Jackson, S.E., Renwick, D.W., Jabbour, C.J. and Muller-Camen, M. (2011). State-of-the-art and future directions for green human resource management: Introduction to the special issue. *German Journal of Human Resource Management*, 25(2), 99-116.
12. Jyoti, Koshish (2019). Green HRM – People Management Commitment to Environmental Sustainability. *Proceedings of 10th International Conference on Digital Strategies for Organizational Success*, Available at SSRN: https://ssrn.com/abstract=3323800 or http://dx.doi.org/10.2139/ssrn.3323800
13. Kim, Y.J., Kim, W.G., Choi, H.M. and Phetvaroon, K. (2019). The effect of green human resource management on hotel employees' eco-friendly behavior and environmental performance. *International Journal of Hospitality Managemen*t, 76, 83-93.
14. Lober, D.J. (1996). Evaluating the environmental performance of corporations. *Journal of Managerial Issues*, 8(2), 184-205.
15. Marhatta, S. and Adhikari, S. (2013). Green HRM and sustainability. *International e-Journal of Ongoing Research in Management & IT*. Retrieved from www.asmgroup.edu.in/incon/publication/incon13-hr-006pdf
16. May, D.R. and Flannery, B.L. (1995). Cutting waste with employee involvement teams. *Business Horizons,* 38(5), 28-39.
17. Paillé P. Chen, Y., Boiral O. and Jin, J. (2014). The Impact of Human Resource Management on Environmental Performance: An Employee-Level Study. *Journal of Business Ethics*, 121, 451-466.
18. Pham, N.T., Tuckova, Z. and Jabbour, C.J.C. (2019). Greening the hospitality industry: how do green human resource management practices influence organizational

citizenship behavior in hotels? A mixed-methods study. *Tourism Management*, 72, 386-399.

19. Ramus, C.A. (2001). Organizational support for employees: Encouraging creative ideas for environmental sustainability. *California Management Review*, 43(3), 85-105.
20. Ren, S., Tang, G. and E Jackson, S. (2018). Green human resource management research in emergence: A review and future directions. *Asia Pacific Journal of Management*, 35, 769-803.
21. Renwick, D., Redman, T. and Maguire, S. (2013), Green human resource management a Research and review agenda. *International Journal of Managemen*t, 5, 1-14.
22. Roy, M.J. and Thérin, F. (2008). Knowledge acquisition and environmental commitment in SMEs. *Corporate Social Responsibility and Environmental Management*, 15(5), 249-259.
23. Yusliza, M.Y., Norazmi, N.A., Jabbour, C.J.C., Fernando, Y., Fawehinmi, O. and Seles, B.M.R.P. (2019). Top management commitment, corporate social responsibility and green human resource management: *A Malaysian study. Benchmarking: An International Journal*, 26(6), 2051-2078.
24. Yusoff, Y.M., Nejati, M., Kee, D.M.H. and Amran, A. (2018). Linking green human resource management practices to environmental performance in hotel industry. *Global Business Review*, 21(3), 1-18.
25. Zakaria, N. (2013). Enhancing organizational performance of Malaysian SMEs through human resource management (HRM) practices and organizational innovative capability: A proposed framework. In: *3rd International Conference on Management (3rd ICM 2013)*, 10-11 June 2013, Penang, Malaysia.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 271-292

ON ESTIMATION OF POPULATION MEAN USING TRIMMED RANKED SET SAMPLING

Shakeel Ahmed[1] and **Javid Shabbir**[2]
[1] School of Natural Sciences, National University of Science and Technology, Islamabad, Pakistan
[2] Department of Statistics, University of Wah, Wah, Pakistan.
[§] Corresponding author Email: shakeel.ahmed@sns.nust.edu.pk

ABSTRACT

Statisticians always seek robust estimates for drawing more reliable statistical inferences about population parameters. Efforts are made to obtain a more robust estimate at the sample selection stage. In this article, we provide a new sample selection mechanism, namely, Trimmed Ranked Set Sampling (TrRSSS) for obtaining a relatively robust and efficient estimate of the trimmed mean of population. The method is based on the selection of ranked set samples after removing influential observations. The optimum proportion of observations to be removed from the ends has also been obtained for the newly constructed sampling scheme using tail weight measure approach. The properties of the proposed measure are studied using real-world data as well as artificially generated population.

KEYWORDS

Influential observation, Robust, Symmetric distribution.

1. INTRODUCTION

Survey practitioners have been working for improvement of sample selection mechanism to reach on a valid statistical inference. A wide range of sampling methods have been developed in last couple of decades. The methods are proposed keeping in view the structure of the sampled population, the resources available for sampling, the objective of the study, and the research question to be answered. The notable probability sampling methods include simple random sampling, stratified random sampling, multi-stage or cluster sampling, probability proportional to size (PPS) sampling etc. In continuation to the probability sampling, McIntyre (1952) introduced a new data collection mechanism, named, ranked set sampling for the situations where taking the actual measurements of a larger number of sampled observations is difficult (e.g., expensive, destructive, and time-consuming), but ranking a small set of sampled units is relatively cheap, easy and reliable. The basic purpose was to improve the precision in estimation of the average herbage yield from large plots without increasing the number of fields from which detailed expensive and tedious measurements required to be taken. The unbiasedness and efficiency of the mean estimator under ranked set sampling was studied by Takahasi and Wakimoto (1968), Dell and Clutter (1972), Stokes (1980) and Stokes and Sager (1988) etc. In last two decades of the 20th century, a significant contribution to ranked set sampling was made by several researchers. Among them Cobby et al. (1985), Ridout and Cobby (1987), Samawi and

Ahmed (1996), Muttlak (1997) and Muttlak (2003) are noteworthy. Most of the contributions to the literature on ranked set sampling are related to the efficiency improvement, truncation and extension of stages, etc. Such work can be found in Balci et al. (2013), Mandowara and Mehta (2014), Biradar and Santosha (2015), Linder et al. (2015) and references there in.

Although ranked set sampling provides at least an efficient result as simple random sampling with replacement for estimation of population mean it is highly affected by extreme values in ranked sets. On the other hand, median ranked set sampling is robust against outliers but does not provide the efficient result for fat-tailed distribution, for instance, normal. In such situations, the idea of estimating central tendency is trimming the extreme observations from the ends of the ranked sets. For past few decades, trimmed mean has been used as an estimator of location parameters when the data is extremely skewed (see: Tukey and McLaughlin (1963), Bickel (1965), Stigler (1973), Ruppert and Carroll (1980), Kim (1992) Kim (1992), Fraiman and Muniz (2001) Fraiman and Muniz (2001), Wilcox (2003) Wilcox (2003) and Wang and Cui (2015) for detail about its properties and history.).

On the lines of research contributions cited above, we extend the idea of trimming to a ranked set sampling setting. We suggest a new data collection mechanism, called trimmed ranked set sampling, where we take measurements from a proportion of units from the middle after removing a certain number of units from both ends of ranked samples. The newly suggested data collection mechanism is then used for estimating the mean of the population. Efficiency and unbiasedness properties are studied mathematically, through numerical example, and simulation studies. Section 2 covers the revision of the trimmed mean under simple random sampling with replacement. Section 3 presents the proposed data collection mechanism along with its properties, while Sections 4 and 5 provide efficiency comparison using Monte Carlo simulation and real-life data sets respectively. Section 6 is devoted to the concluding remarks of the study.

2. TRIMMED MEAN UNDER SIMPLE RANDOM SAMPLING

Let $Y_1, Y_2, \ldots\ldots Y_n$, be a simple random sample of size n taken with replacement from a population of size N having mean μ and variance σ^2. The sample mean along its variance is given by

$$\bar{y} = \frac{\sum y_i}{n}. \tag{1}$$

$$\mathrm{Var}(\bar{y}) = \frac{\sigma^2}{n} \tag{2}$$

Although simple mean estimator has many attractive features like unbiasedness, consistency and sufficiency etc., it has a distinct demerit which is its sensitivity with respect to outlying observations in the sample. To overcome this deficiency the idea of trimming is used which includes first sorting data in ascending order and a specific percentage of observations from the upper and/or lower tails are removed or trimmed. Let $Y_{(1)}, Y_{(2)}, \ldots\ldots Y_{(n)}$ be an ordered sample of size n from a density $f(y)$ having cumulative distribution function (cdf) $F(y)$. The marginal density of the ith order statistic and the join density of the ith and the jth order statistics are given respectively

$$f(y_{(i)}) = \frac{n!}{(i-1)!\,(n-i)!}\left[F(y_{(i)})\right]^{i-1}\left[1-F(y_{(i)})\right]^{n-i} f(y_{(i)}),$$
$$-\infty < y_{(i)} < \infty \text{ for } i = 1,2,\dots n$$

$$f(y_{(i)}, y_{(j)}) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\left[F(y_{(i)})\right]^{i-1}\left[F(y_{(j)}) - F(y_{(i)})\right]^{j-i-1}\left[1-F(y_{(j)})\right]^{n-j} \times f(y_{(i)}) f(y_{(j)}),\ -\infty < y_{(i)} < y_{(j)} < \infty, \forall i < j.$$

The sample mean after trimming $\gamma\%$ data is written as follow

$$\hat{\phi}_{t(k_1,k_2)} = \frac{\sum_{i=k_1+1}^{n-k_2} y_{(i)}}{n^*} \tag{3}$$

where $n^* = n - k_1 - k_2, k_1 = \gamma_1 n$ and $k_2 = \gamma_2 n$ be the number of smallest and largest units to be removed from the ordered samples respectively, where γ is the overall proportion of units to be trimmed from the sample such that $\gamma_1 + \gamma_2 = \gamma$. For symmetric distribution, we choose $k_1 = k_2 = k$. Expected value of $\hat{\phi}_{t(k_1,k_2)}$ is given by

$$E(\hat{\phi}_{t(k_1,k_2)}) = \frac{\sum_{i=k_1+1}^{n-k_2} E(Y_{(i)})}{n^{*2}} = \frac{\sum_{i=k_1+1}^{n-k_2} \mu_{(i)}}{n^*} = \phi_{t(k_1,k_2)} \text{ (say)} \tag{4}$$

which is unbiased only when underlying distribution is symmetric about μ. Many researchers attempted to find the exact expression for the variance of trimmed mean among them Balakrishnan and David (2001) and Rivest et al. (1995) are famous.
By definition, we have

$$\mathrm{Var}(\hat{\phi}_{t(k_1,k_2)}) = \frac{\mathrm{Var}\left(\sum_{i=k_1+1}^{n-k_2} Y_{(i)}\right)}{n^{*2}} = \frac{1}{n^{*2}}\left(\sum_{i=k_1+1}^{n-k_2} \sigma_{(i)}^2 + \sum_{i=k_1+1}^{n-k_2} \sum_{j\neq i; j=k_1+1}^{n-k_2} \sigma_{(i)(j)}\right) \tag{5}$$

Assuming $k_1 = k_2 = k$, we have

$$\mathrm{Var}(\hat{\phi}_{t(k,k)}) = \frac{\sum_{i=k+1}^{n-k} \sigma_{(i)}^2}{(n-2k)^2} + \frac{\sum_{i=k+1}^{n-k} \sum_{j\neq i; j=k+1}^{n-k} \sigma_{(i)(j)}}{(n-2k)^2} \tag{6}$$

where $\sigma_{(i)}^2 = E\left(Y_{(i)} - \mu_{(i)}\right)^2$ and $\sigma_{(i)(j)} = E\left(Y_{(i)} - \mu_{(i)}\right)\left(Y_{(j)} - \mu_{(j)}\right)$ are respectively the variance of the ith order statistic and the covariance between the ith and the jth order statistics.

3. TRIMMED RANKED SET SAMPLING

Trimming observations in simple random sampling requires observations of units for ordering which might be difficult when observation of units is costly or time taking.

Ranked set sampling, on the other hand, allows ranking of small sets on the basis of some concomitant variables or through visual inspection. In this way trimming of units with influential values of target variable is possible before the observation. To this extend, we propose a modified version of ranked set sampling is proposed following Wilcox (2003) and named as trimmed ranked set sampling. The newly developed ranked set sampling provides efficient estimator for a trimmed mean rather than a complete mean. The algorithm for selecting a trimmed ranked set sample is as follow

1. Select $m^* = m - l_1 - l_2$ random samples each of size m from a population of interest using simple random sampling (SRSWR).
2. Rank each set in itself according to certain ranking mechanism then discard l_1 smallest ranked units and l_2 largest ranked units from each set.
3. Observe the value of study variable from m^* sets each with size m^*.
4. Repeat Steps 1-3, r times to obtain a ranked set sample of size $r(m - l_1 - l_2) = n - k_1 - k_2$, where $k_1 = rl_1$ and $k_2 = rl_2$.

Note that in ranked set sampling $l_1 = \gamma_1 m$, and $l_2 = \gamma_2 m$ are assumed to be the number of the smallest and the largest units to be removed from each ranked sample respectively. For symmetric distribution, we choose $l_1 = l_2 = l$ such that $l_1 + l_2 = 2l$. But, in practice, the observed source often follows asymmetric distribution so choosing l_1 and l_2 equally may results a misleading sample. In such situations, we choose l_1 and l_2 on the line of Alkhazaleh and Razali (2010). Alkhazaleh and Razali (2010) introduced γ_1 based on numerator and denominator of tail weight measure (TWM) suggested by Brys et al. (2006) in simple random sampling based on quantile values of the data set as follow

$$\gamma_1 = \frac{1}{2}\left[\frac{Y_{\left(\left(1-\frac{p}{2}\right)n\right)} + Y_{\left(n\frac{p}{2}\right)} - 2Y_{(0.25n)}}{Y_{\left(\left(1-\frac{p}{2}\right)n\right)} - Y_{(0.25n)}}\right]\gamma \tag{7}$$

and $\gamma_2 = \gamma - \gamma_1$ and $Y_{(p)}$ denotes pth percentile of the sample of size n. To see statistical properties of TWM readers can read Alkhazaleh and Razali (2010). Under a ranked set sample of size rm one may estimate the TWM with the percentiles $Y_{\left(\frac{p}{2}rm\right)}, Y_{\left(\left(1-\frac{p}{2}\right)rm\right)}$ and $Y_{(0.25rm)}$, where p is assumed to be fixed in advance. However, we already assumed that the measurement of the units is performed after removing the influential units on the basis of some concomitant variable or through visual inspection. A proxy TWM can be obtained through some highly correlated variable used for ranking.

Suppose that $\{Y_{1(l_1+1)j}, Y_{1(l_1+2)j}, \dots, Y_{1(m-l_2)j}\}$ for $(j = 1,2,\dots,r)$ be a trimmed ranked set sample for the jth cycle obtained after discarding l_1 and l_2 smallest and largest units respectively from the ends. The mean estimator under the proposed ranked set sampling scheme after ignoring $k_1 = rl_1$ and $k_2 = rl_2$ from the ends, is given by

$$\hat{\phi}_{tr(k_1,k_2)} = \frac{\sum_{j=1}^{r}\sum_{i=1}^{m^*} Y_{i(l_1+i)j}}{rm^*} \tag{8}$$

Taking expectation, we have

$$E\left(\hat{\phi}_{tr(k_1,k_2)}\right) = \frac{\sum_{i=1}^{m^*}\mu_{(l_1+i)}}{m^*} = \phi_{(k_1,k_2)} \tag{9}$$

where $\mu_{(l_1+i)} = E(Y_{(l_1+i)})$ for $i = 1,2, \dots, m^*$ be the population mean of the m^* middle order statistics. If population is symmetric $\hat{\phi}_{tr(k_1,k_2)}$ will be unbiased. Assuming unbiased, we obtain following variance.

$$\mathrm{Var}(\hat{\phi}_{tr(k_1,k_2)}) = \frac{\sum_{i=1}^{m^*} \sigma^2_{(l_1+i)}}{rm^{*2}} \tag{10}$$

where $\sigma^2_{(l_1+i)} = E(Y_{(l_1+i)} - \mu_{(l_1+i)})^2 = E(Y_{(l_1+i)} - \mu + \mu - \mu_{(l_1+i)})^2$ for $i = 1,2, \dots, m^*$. After some algebraic operations, we get $\sigma^2_{(l_1+i)} = \sigma^2 - \Delta^2_{(l_1+i)}$ for $i = 1,2, \dots, m^*$. The mean estimator under trimmed ranked set sampling is at least as efficient as the mean estimator under a trimmed simple random sample with replacement i.e.

$$\mathrm{Var}(\hat{\phi}_{t(k_1,k_2)}) - \mathrm{Var}(\hat{\phi}_{tr(k_1,k_2)}) = \frac{1}{n^*} \sum_{i=k_1+1}^{n-k_2} \sum_{j \neq i; j=k_1+1}^{n-k_2} \sigma_{(i)(j)} \tag{11}$$

The trimmed ranked set sampling provides better efficiency for mean estimation than a trimmed simple random sample when the correlation between the values of ith and jth order statistics is positive for all i and j i.e. $\sigma_{(i)(j)} > 0$. For a negligible correlation the two approaches provide same efficiency. To compare the mean estimator under TrRSS with the mean estimator under un-trimmed simple random sample with replacement of size n^*, we write Equation (10) as follow

$$\mathrm{Var}(\hat{\phi}_{tr(k_1,k_2)}) = \frac{\sigma^2}{r(m - l_1 - l_2)} - \frac{\sum_{i=1}^{m-l_2} \Delta^2_{(l_1+i)}}{r(m - l_1 - l_2)^2} \tag{12}$$

The first term on the right side of Equation (12) is the variance of mean estimator under un-trimmed SRSWR with $n^* = r\,(m - l_1 - l_2)$ units which shows superiority of trimmed ranked set sampling estimator to the ean estimator under un-trimmed SRSWR for same number of sample units measured. For $l_1 = l_2 = l = k/r$, we have

$$\mathrm{Var}(\hat{\phi}_{t(k,k)}) = \frac{\sigma^2}{r(m - 2l)} - \frac{\sum_{i=1}^{m-l} \Delta^2_{(l+i)}}{r(m - 2l)^2} \tag{13}$$

where $\Delta_{(l+i)} = \mu_{(l+i)} - \mu$. We obtain special cases of the trimmed mean estimator given in Equation (8) for different choices of l_1 and l_2 in Table 1.

Table 1
Special Cases of $\hat{\phi}_{tr(k_1,k_2)}$ for Different Choices of l_1 and l_2

		Reduced forms of Proposed Estimators $\hat{\phi}$	Description
l_1	l_2	$\hat{\phi}_{tr(k_1,k_2)}$	
0	l_2	$\hat{\phi}_{tr(0,k_2)} = \frac{\sum_{j=1}^{r}\sum_{i=1}^{m-1}]_2 Y_{i(i)j}}{r(m-k_2)}$	(Right trimmed mean)
l_1	0	$\hat{\phi}_{tr(k_1,0)} = \frac{\sum_{j=1}^{r}\sum_{i=l_1+1}^{m} Y_{i(l_1+i)j}}{r(m-l_1)}$	(Left trimmed mean)
0	0	$\hat{\phi}_{tr(0,0)} = \frac{\sum_{j=1}^{\prime}\sum_{i=1}^{m} Y_{i(i)j}}{rm}$	(Usual mean under rss)
$\frac{m}{2}-1$ $\frac{(m-1)}{2}$	$\frac{m}{2}-1$ $\frac{(m-1)}{2}$	$\hat{\phi}_{tr\left(r\left(\frac{m}{2}-1\right),r\left(\frac{m}{2}-1\right)\right)} = \frac{1}{r}\sum_{j=1}^{r}\left[\frac{Y_{1\left(\frac{m}{2}\right)j} + Y_{2\left(\frac{m}{2}+1\right)j}}{2}\right]$ $\hat{\phi}_{\mathrm{tr}\left(\frac{r(m-1)}{2},\frac{r(m-1)}{2}\right)} = \frac{1}{r}\sum_{j=1}^{r} Y_{1\left(\frac{m+1}{2}\right)j}$	(Sample median with even m) (Sample median with odd m)

3.1 Theoretical Comparison

To conduct a theoretical comparison of the proposed data collection mechanism we consider that the sample is selected from a uniform population ih the range of$(0,1)$. The uniform distribution is considered due to less complex structure of mathematical derivation of its statistical properties. The density function of the variable interest is given by

$$f(y) = 1, 0 < y < 1$$

Let $Y_{(i)}$ be the ith order statistics of a random sample of size n drawn from a population having uniform density as given in Equation (15). The marginal distribution of $Y_{(i)}$ is given by

$$f(y_{(i)}) = \frac{n!}{(i-1)!\,(n-i)!}\left(1-y_{(i)}\right)^{i-1}\left(y_{(i)}\right)^{n-i+1}, 0 < y_{(i)} < 1, \text{ for } i = 1,2,..,n$$

The joint density of $Y_{(i)}$ and $Y_{(j)}$ is obtained as follow:

$$f(y_{(i)}, y_{(j)}) = \frac{n!}{(i-1)(j-i-1)!\,(n-j)!}\left(y_{(i)}\right)^{i-1}\left(y_{(j)} - y_{(i)}\right)^{j-i-1}\left(1-y_{(j)}\right)^{n-j},\ 0 < y_{(i)} < y_{(j)} < 1.$$

Using Equation (16) and (17), we obtain following identities

$$E\left(Y_{(i)}\right) = \mu_{(i)} = \frac{i}{n+1}, \mathrm{Var}\left(Y_{(i)}\right) = \sigma_{(i)}^2 = \frac{i(n-i+1)}{(n+1)^2(n+2)}$$

and

$$\mathrm{Cov}\left(Y_{(i)}, Y_{(j)}\right) = \sigma_{(i)(j)}^2 = \frac{i(n-j+1)}{(n+1)^2(n+2)}$$

Under uniform distribution, expected values of $\left(\hat{\phi}_{t(k_1,k_2)}\right)$ and $\left(\hat{\phi}_{tr(k_1,k_2)}\right)$ are expressed as follow

$$E\left(\hat{\phi}_{t(k_1,k_2)}\right) = \sum_{i=k+1}^{n-k} \frac{i}{(n-2k)(n+1)}$$
$$= \sum_{i=k+1}^{n-k} \frac{i}{(n-2k)(n+1)} = \frac{(n-2k)(n+1)}{2(n-2k)(n+1)} = \frac{1}{2}$$

and

$$E\left(\hat{\phi}_{tr(k_1,k_2)}\right) = \sum_{l=1}^{r} \sum_{i=k+1}^{m-k} \frac{i}{(m-2k)(n+1)}$$
$$= \sum_{l=1}^{r} \sum_{i=k+1}^{m-k} \frac{i}{r(m-2k))(m+1)} = \frac{(m-2k)(m+1)}{2(m-2k)(m+1)} = \frac{1}{2}.$$

Equations (18) and (19) provide that both $\hat{\phi}_{tr(k_1,k_2)}$ and $\hat{\phi}_{t(k_1,k_2)}$ both are unbiased estimators of population mean $\left(\mu = \frac{1}{2}\right)$. Similarly, the respective variances are given by

$$\mathrm{Var}\left(\hat{\phi}_{t(k_1,k_2)}\right) = \frac{1}{(n-2k)^2}\left(\sum_{i=k+1}^{n-k} \frac{i(n-i+1)}{(n+1)^2(n+2)} + \sum_{i=k+1}^{n-k} \sum_{j\neq i; j=k+1}^{n-k} \frac{i(n-j+1)}{(n+1)^2(n+2)}\right)$$

and

$$\mathrm{Var}\left(\hat{\phi}_{tr(k_1,k_2)}\right) = \frac{1}{(m-2k)^2} \sum_{i=k+1}^{m-k} \frac{i(m-i+1)}{(m+1)^2(m+2)}$$

Here, we assumed $r = 1$ for easy of computation. From (20) and (21) we can notice that for $m = n$ the variance of trimmed mean under ranked set sampling is smaller than that of trimmed mean estimator under simple random sampling. This is due to the fact that the covariance between the order statistics are positive i.e. $n + 1 - j > 0$ for $j = k + 1, k + 2, \ldots, n - k$. It is to be noted that the proposed data collection mechanism is applied only for relatively large m i.e. $m > 2$. The derivation of expectation and variance of the special cases of the proposed estimator is straightforward. To support theoretical findings, we provide a numerical study in next section.

4. NUMERICAL STUDY

Real life application of the proposed data collection mechanism along with efficiency comparison is incorporated in this section.

4.1 Illustration of Procedure

For illustration of the proposed ranked set sampling procedure a leaf data set collected by Pedro et al. (2013) is used, where aspect ratio of the leaves is taken as the study variable. The data set consists of $N = 340$ leaves which is taken as the study population. Figures 1-3 present the illustration of selection scheme (sampling layout) of a trimmed ranked set sample with different proportions units to be trimmed the ends for $m = 6$. For this purpose, 36 units are drawn using sampling with replacement from the population and divided into 6 groups each of size $m = 6$. Rank each set of leaves according to their sizes i.e. placing leaves in order of the smallest to the largest in each set.

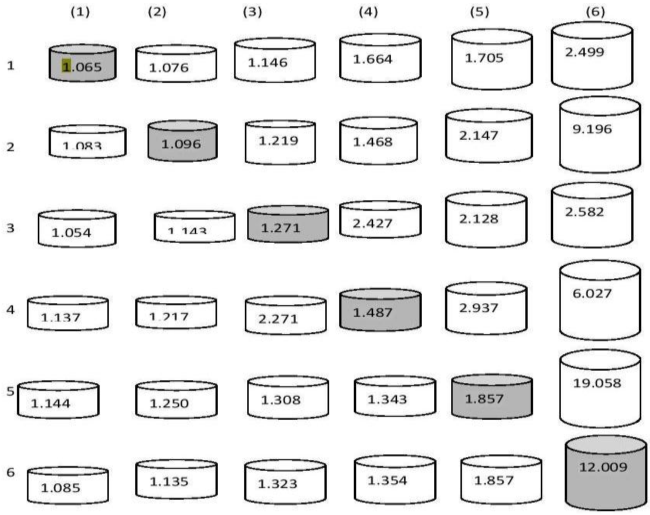


Figure 1: Sampling Layout with $k_1 = 0$ and $k_2 = 0$ (a) $\mathbf{K1 = 0, k2 = 3}$

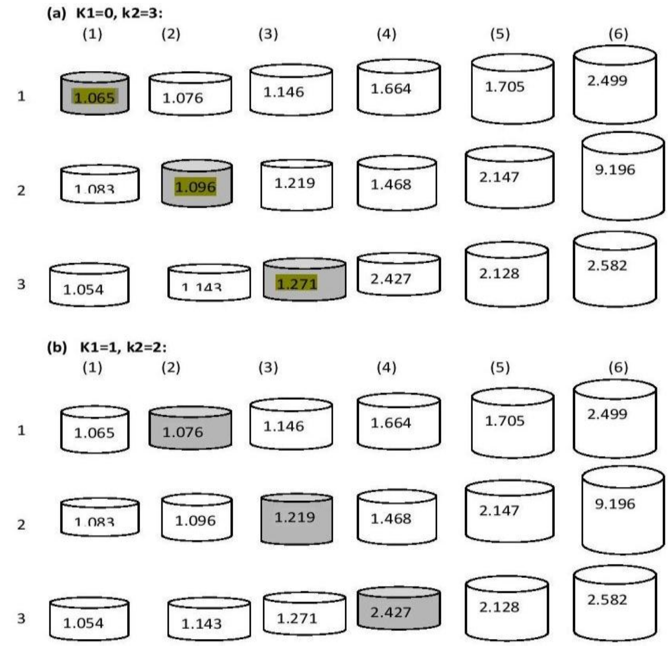


Figure 2: Sampling Layout with $k_1 = 0, 1$ and $k_2 = 3, 2$ (a) for $k1 = 2, k2 = 0$

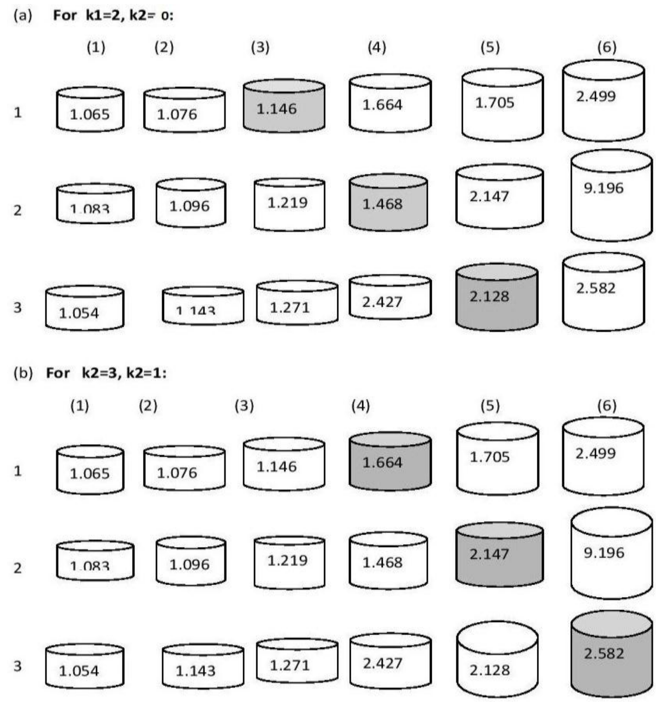


Figure 3: Sampling layout with $k_1 = 2, 3$ and $k_2 = 1$

In Figures 1-3, selected units are shown physically with respect to the study variable. The units might be the volume of timber of trees in forestry research or the area of plot in agriculture experiment where one might remove the unit with extremely large or extremely small values of the study variable from the data sets to be selected. In our example, these shapes are structured to just display the physical appearance of units with different sizes in a sample. Figure 1 displays the layout of an untrimmed ranked set sample of size $m = 6$ without trimming data. While Figures 2 and 3 provide a snapshot of the scheme with a 50% trimmed ranked set sample with different proportions of units to be trimmed from the ends. In the figures, the units selected for actual measurement are shown with shaded shapes. The sampling layouts present the units selected and ranked according to some ranking mechanism that is highly correlated with the study variable. For selecting a 50% trimmed ranked set sample, three sets each of size $m = 6$ are required to select for ranking final one unit is selected from each sample according to the values of k_1 and k_2.

4.2 Efficiency Comparison

For repeated sampling from the data set described above, the following steps are taken to obtain the variances and the relative efficiency (RE) of the mean estimators.

- **Step 1:** Select 20,000 each of sample of size $m = n$, for avoiding complexity setting $r = 1$, from the leaf data
- **Step 2:** Rank the samples from smallest to largest with respect to the leaf size to get $\mu_{(i)}, \sigma^2_{(i)}$ and $\sigma_{(i)(j)}$ for $i, j = 1,2, \dots, m$ and $i \neq j$.
- **Step 3:** Compute the variance of mean estimators $\hat{\phi}_{t(k_1,k_2)}$ and $\hat{\phi}_{tr(k_1,k_2)}$ for different combinations of k_1, k_2 and m.

We use the following expression to obtain the relative efficiency (RE) of mean estimator under proposed ranked set sampling is then obtained as follow:

$$RE = \frac{\mathrm{Var}\left(\hat{\phi}_{t(k_1,k_2)}\right)}{\mathrm{Var}\left(\hat{\phi}_{tr(k_1,k_2)}\right)}. \tag{14}$$

For efficiency comparison on the basis of leaf data set we take $m = n = 10{,}15$ by setting $r = 1$ for simplicity. The variances of $\hat{\phi}_{t(k_1,k_2)}$ and $\hat{\phi}_{\mathrm{tr}(k_1,k_2)}$ along with their ratios i.e. relative efficiency are given in Table 2. It can be inferred from Table 2 that efficiency of trimmed ranked set sample relative to simple random sample increases with increase in γ. It is also noticed that relative efficiency increases only with increase in k_2 while changing k_1 does not affect efficiency of trimmed mean under ranked set sampling. Relative efficiency also increases with increase in set size m for $r = 1$.

Table 2
Variances and Relative Efficiency for Different Values of n and γ

		$n=m=10$				$n=m=15$			
$k1$	$k2$	γ	$Var(\hat{\phi}_{t(k_1,k_2)})$	$Var(\hat{\phi}_{tr(k_1,k_2)})$	RE	γ	$Var(\hat{\phi}_{t(k_1,k_2)})$	$Var(\hat{\phi}_{tr(k_1,k_2)})$	RE
	0	0	0.6566	0.3073	2.1362	0	0.45	0.1654	2.7208
	1	0.1	0.3064	0.135	2.2694	0.067	0.2499	0.0903	2.7658
0	2	0.2	0.1121	0.0448	2.5026	0.133	0.1157	0.0391	2.9541
	3	0.3	0.0509	0.0163	3.12	0.200	0.056	0.0154	3.6582
	4	0.4	0.0315	0.0097	3.2442	0.267	0.0331	0.0072	4.6312
	5	0.5	0.0227	0.0074	3.0614	0.333	0.0233	0.0045	5.1154
	0	0.1	0.8256	0.3834	2.1532	0.067	0.5068	0.1872	2.7076
	1	0.2	0.3627	0.1653	2.1947	0.133	0.2905	0.1042	2.7872
1	2	0.3	0.1468	0.0594	2.4734	0.200	0.1424	0.0478	2.9776
	3	0.4	0.0673	0.0228	2.9561	0.267	0.0631	0.0176	3.5885
	4	0.5	0.0435	0.0136	3.1869	0.333	0.0395	0.0084	4.6990
	0	0.2	1.0237	0.4816	2.1257	0.133	0.5947	0.2198	2.7052
	1	0.3	0.468	0.2169	2.158	0.2	0.3426	0.1241	2.7603
2	2	0.4	0.188	0.0792	2.3736	0.267	0.1542	0.0529	2.9110
	3	0.5	0.0877	0.0316	2.774	0.333	0.0763	0.0218	3.4877
	0	0.3	1.3256	0.6241	2.124	0.133	0.6908	0.2557	2.7007
3	1	0.4	0.6223	0.2985	2.0849	0.200	0.4074	0.1472	2.7673
	2	0.5	0.2614	0.1154	2.2655	0.267	0.1843	0.0651	2.8321
	0	0.4	1.7751	0.8583	2.0682	0.133	0.8073	0.3031	2.6630
4	1	0.5	0.9037	0.4412	2.0484	0.200	0.4836	0.1784	2.7094
5	0	0.5	1.8439	0.8647	2.1324	0.200	0.8345	0.311	2.6829

5. SIMULATION STUDY

The effect of trimming observations from the end of a ranked set sample and a simple random sample are compared for different choices of k_1, k_2 and m with symmetric, moderately skewed and highly skewed distributions. The steps involved in simulation are as follow.

- **Step 1:** Generate a data set from a gamma distribution with different combinations of parameters a and b, where values of a and b are taken 1,2,3,4 and 2,3,4,5 respectively.
- **Step 2:** Select 20,000 samples each of size m, for avoiding complexity setting r=1, from the population generate $\sigma_{(i)}^2$ and $\sigma_{(i)(j)}$ for $i, j = 1,2,..,m$ and $i \neq j$
- **Step 3:** Obtain the variances of $\hat{\phi}_{t(k_1,k_2)}$ and $\hat{\phi}_{\mathrm{tr}(k_1,k_2)}$ for different combinations of k_1, k_2 and m.

The relative efficiency of the proposed mean estimator is then obtained using Equation (22). Tables 4-10, given in the Appendix, provide variances of mean estimators along with relative efficiency of the trimmed ranked set sampling estimator relative to trimmed mean under simple random sampling with replacement. Four panels of Figure 4 show different behaviors of gamma distribution i.e. moves from extreme skewed to symmetry within a panel and becomes fat tailed as b increases moving toward Panels (a) to (d). (a)

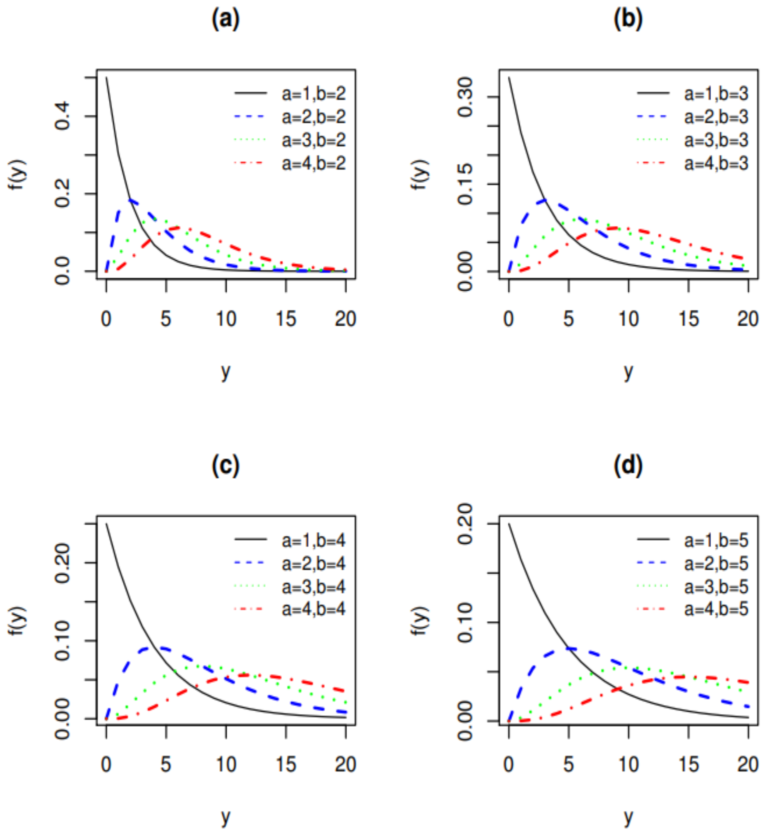


Figure 4: Gamma Distribution for Different Choice of (a, b)

6. DISCUSSION ON RESULTS

Tables 3-10 provide an idea about the behavior of the mean estimator under the trimmed ranked set sample with respect to its counterpart in SRSWR. It is noticed that relative efficiency increases with an increase in m for fixed r, i.e. $r = 1$. For different choices of scale parameter a and the shape parameter b different types of distributions are generated and shown in four different panels of Figure 4, like fat tail, thin tail, long tail, skewed and symmetric distributions, etc.

Further relative efficiencies for various choices of l_1 and l_2 can be interpreted as follow:

1. With the increase in a, i.e. moving from extremely skewed to the symmetry left truncation becomes useless as it does not improve efficiency this can be noticed from Tables 3 and 4.
2. Trimming improves efficiency when the underlying distribution has a fat tail which can be noticed from the rigid line on the graph of different panels of Figure 4 and the results can be found in Tables 3-4.

3. Right trimming improves efficiency when the distribution is positively skewed (see different combinations in Panel (a) of Figure 4). The effect of trimming decreases when skewness in data decreases. The results can be seen from Tables 3-4.
4. The relative efficiency of the mean estimator under trimmed ranked set sampling increases with an increase in k_2.
5. The relative efficiency of the mean estimator under trimmed ranked set sampling also increases with an increase in set size m for $r = 1$.

7. CONCLUSION

A robust scheme of data collection is suggested based on ranked set sampling for situations in which extreme observations disturb the estimation of measure of central tendency of a population. Focus was to obtain a ranked set sample after excluding units containing influential observations before taking actual measurement. It is also justified that for asymmetric population one can obtain amounts of trimming from the ends on the line of Alkhazaleh and Razali (2010). The proposed data collection method gives greater efficiency than its counterpart in ranked set sample as well as to the situations where no truncation is performed. Efficiency comparison through the simulation and through leaf data set support the superiority of the proposed scheme.

8. REFERENCES

1. Alkhazaleh, A.M.H. and Razali, A.M. (2010). New technique to estimate the asymmetric trimming mean. *Journal of Probability and Statistics*, Article ID 739154, https://doi.org/10.1155/2010/739154
2. Balakrishnan, N. and David, H.A. (2001). A note on the variance of a lightly trimmed mean when multiple outliers are present in the sample. *Statistics Probability Letters*, 55(4), 339-343.
3. Balci, S., Akkaya, A. and Ulgen, B. (2013). Modified maximum likelihood estimators using ranked set sampling. *Journal of Computational and Applied Mathematics*, 238, 171-179.
4. Bickel, P.J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics*, 36(3), 847-858.
5. Biradar, B.S. and Santosha, C.D. (2015). Estimation of the population mean using paired ranked set sampling. *Open Journal of Statistics*, 5, 97-103.
6. Brys, G., Hubert, M. and Struyf, A. (2006). Robust measures of tail weight. *Computational Statistics Data Analysis*, 50(3), 733-759.
7. Cobby, J., Ridout, M., Bassett, P. and Large, R. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science*, 40, 257-263.
8. Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
9. Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2), 419-440.
10. Kim, S.J. (1992). The metrically trimmed mean as a robust estimator of location. *The Annals of Statistics*, 20(3), 1534-1547.

11. Linder, D., Samawi, H., Yu, L., Chatterjee, A., Huang, Y., and Vogel, R. (2015). On stratified bivariate ranked set sampling for regression estimators. *Journal of Applied Statistics*, 42(12), 2571-2583.
12. Mandowara, V. and Mehta, N. (2014). Modified ratio estimators using stratified ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, 43(3), 461-471.
13. McIntyre, G. (1952). A method of unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3, 385-390.
14. Muttlak, H.A. (1997). Median ranked set sampling. *Journal of Applied Statistical Sciences*, 6, 245-255.
15. Muttlak, H.A. (2003). Investigating the use of quartile ranked set samples for estimating the population mean. *Applied Mathematics and Computation*, 146, 437-443.
16. Pedro, F., Silva, Andre, R., Marcal, R.M. and Almeida, D. (2013). Leaf data set. *Springer Lecture Notes in Computer Science*, 7950, 197-204.
17. Ridout, M. and Cobby, J. (1987). Ranked set sampling with non-random selection of sets and errors in ranking. *Applied Statistics*, 36, 145-152.
18. Rivest, L., Abu-Dayyeh, and Caprra, P. (1995). On the variance of the trimmed mean. *Statistics Probability Letters*, 22, 79-85.
19. Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, 75(372), 828-838.
20. Samawi, H., Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling. *Biometrical Journal*, 30, 577-586
21. Stigler, S.M. (1973). The asymptotic distribution of the trimmed mean. *The Annals of Statistics*, 472-477.
22. Stokes, S. (1980). Estimation of variance using judgment ordered ranked set samples. *Biometrics*, 36, 35-42.
23. Stokes, S. and Sager, T. (1988). Characterization of a ranked set sample with application to estimating distribution functions. *Journal of American Statistical Association*, 83, 374-381.
24. Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20, 1-31.
25. Tukey, J.W. and McLaughlin, D.H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization 1. *Sankhȳa: The Indian Journal of Statistics*, Series A, 331-352.
26. Wang, S.-Y. and Cui, H.-J. (2015). Trimmed and winsorized transformed means based on a scaled deviation. *Acta Mathematicae Applicatae Sinica*, English Series, 31(2), 475-492.
27. Wilcox, R.R. (2003). Trimmed means. *Encyclopedia of Statistics in Behavioral Science*, 4, 2066-2067.

APPENDIX

Table 3

Gamma (a = 1, b = 2)										
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0233	0.007	3.3302	0.0223	0.0051	4.3493	0.0178	0.0044	4.0059
	1	0.0161	0.0039	4.1538	0.0175	0.0037	4.7913	0.0128	0.0024	5.3548
	2	0.0125	0.003	4.1702	0.0141	0.0028	4.9685	0.0106	0.0018	5.7803
0	3	0.0102	0.0025	4.0512	0.0111	0.0023	4.7368	0.0086	0.0015	5.7745
	4	0.0081	0.0022	3.6927	0.0094	0.0021	4.5417	0.0075	0.0013	5.6495
	5	0.0063	0.0019	3.2561	0.0072	0.0017	4.1201	0.0066	0.0012	5.3768
1	0	0.0309	0.0089	3.4528	0.0232	0.0066	3.5196	0.0207	0.0045	4.5752
	1	0.0228	0.0055	4.1639	0.0163	0.0035	4.6498	0.0167	0.003	5.4945
	2	0.0187	0.0045	4.1306	0.0131	0.0027	4.8868	0.014	0.0025	5.6763
	3	0.0151	0.0039	3.8282	0.0109	0.0023	4.7924	0.0116	0.0021	5.5728
	4	0.0128	0.0037	3.4475	0.0093	0.0021	4.4871	0.0099	0.0018	5.42
	0	0.0411	0.0129	3.1763	0.0325	0.0091	3.5746	0.0251	0.006	4.221
2	1	0.0281	0.0075	3.7262	0.0222	0.0051	4.3155	0.0187	0.0036	5.1933
	2	0.0221	0.0058	3.8285	0.0181	0.004	4.557	0.0153	0.0028	5.5192
	3	0.0183	0.0052	3.5326	0.0142	0.0033	4.376	0.0131	0.0024	5.4391
3	0	0.0413	0.0142	2.9183	0.0314	0.0093	3.3829	0.0251	0.0066	3.8209
	1	0.0283	0.0079	3.5832	0.0219	0.0051	4.3329	0.0184	0.0036	5.1024
	2	0.0228	0.0066	3.4619	0.0181	0.0041	4.3991	0.015	0.0029	5.2656
	0	0.0508	0.0167	3.0351	0.0368	0.0099	3.7174	0.0284	0.0067	4.2253
4	1	0.0394	0.0118	3.3362	0.0291	0.0069	4.2131	0.0235	0.0046	5.0743
5	0	0.0706	0.0265	2.6635	0.0487	0.0147	3.2997	0.0355	0.0093	3.8255

Gamma(a = 2, b = 2)										
k_1	k_2	m=10			m=12			m=14		
	0	0.0525	0.0139	3.7723	0.044	0.0103	4.2639	0.0374	0.0079	4.7533
	1	0.0407	0.0089	4.5587	0.0351	0.0066	5.3019	0.0306	0.0051	5.9814
	2	0.0346	0.0078	4.4698	0.0306	0.0057	5.3532	0.027	0.0044	6.2057
0	3	0.0301	0.0071	4.22	0.0274	0.0053	5.2016	0.0241	0.004	6.0964
	4	0.0268	0.0069	3.8584	0.0238	0.0049	4.8714	0.0218	0.0037	5.8433
	5	0.0238	0.0071	3.367	0.0218	0.0048	4.4987	0.0195	0.0035	5.5128
1	0	0.0632	0.0168	3.7699	0.0509	0.0119	4.2673	0.043	0.0088	4.8814
	1	0.0516	0.0116	4.459	0.0417	0.008	5.1798	0.0351	0.0059	5.9346
	2	0.044	0.0101	4.3558	0.0366	0.0069	5.2878	0.0315	0.0051	6.1334
	3	0.0399	0.0098	4.0698	0.0334	0.0065	5.1105	0.0289	0.0048	6.0398
	4	0.0365	0.0101	3.6206	0.0304	0.0065	4.7106	0.027	0.0046	5.8069
	0	0.0915	0.0268	3.4125	0.0703	0.0173	4.0698	0.0567	0.0124	4.5693
2	1	0.0734	0.0195	3.7748	0.0588	0.0125	4.7076	0.0477	0.0087	5.4889
	2	0.0639	0.0178	3.5882	0.0513	0.0111	4.6206	0.0423	0.0076	5.5603
	3	0.0579	0.0184	3.1418	0.0456	0.0106	4.2983	0.0383	0.0071	5.3749
3	0	0.0736	0.0214	3.4356	0.0596	0.0142	4.1845	0.0468	0.0099	4.7454
	1	0.0629	0.0162	3.8808	0.0484	0.0101	4.7806	0.0396	0.007	5.6307
	2	0.055	0.0153	3.6083	0.0448	0.0095	4.7059	0.0365	0.0064	5.7169
	0	0.0913	0.0275	3.3168	0.067	0.0166	4.0241	0.0539	0.0113	4.7665
4	1	0.0793	0.0232	3.4179	0.0588	0.0134	4.3762	0.0471	0.0089	5.2706
5	0	0.1454	0.0526	2.7659	0.1002	0.0292	3.4307	0.0776	0.0187	4.1457

Table 4

Gamma(a = 3, b=2)										
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0729	0.0175	4.1746	0.0611	0.0127	4.8110	0.0528	0.0097	5.4507
	1	0.0615	0.0129	4.7697	0.0522	0.0093	5.5837	0.0454	0.0071	6.3480
	2	0.0553	0.0118	4.6757	0.0484	0.0086	5.6548	0.0412	0.0063	6.4947
0	3	0.0541	0.0123	4.3774	0.0448	0.0084	5.3436	0.0395	0.0062	6.3914
	4	0.0503	0.0128	3.9352	0.0429	0.0086	4.9836	0.0380	0.0063	6.0256
	5	0.0473	0.0138	3.4232	0.0403	0.0089	4.5130	0.0364	0.0065	5.6084
1	0	0.0830	0.0205	4.0449	0.0672	0.0145	4.6481	0.0572	0.0107	5.3537
	1	0.0700	0.0157	4.4520	0.0579	0.0112	5.1892	0.0499	0.0082	6.0481
	2	0.0620	0.0142	4.3665	0.0527	0.0100	5.2940	0.0444	0.0073	6.1137
	3	0.0563	0.0140	4.0314	0.0486	0.0096	5.0846	0.0411	0.0069	5.9813
	4	0.0518	0.0144	3.5919	0.0434	0.0093	4.6643	0.0382	0.0067	5.6786
	0	0.1029	0.0269	3.8297	0.0822	0.0185	4.4368	0.0676	0.0133	5.0717
2	1	0.0858	0.0202	4.2551	0.0695	0.0136	5.1009	0.0676	0.0133	5.0717
	2	0.0769	0.0189	4.0769	0.0629	0.0123	5.1097	0.0676	0.0133	5.0717
	3	0.0731	0.0198	3.7003	0.0598	0.0124	4.8166	0.0676	0.0133	5.0717
3	0	0.1321	0.0374	3.5290	0.1017	0.0239	4.2484	0.0676	0.0133	5.0717
	1	0.1077	0.0282	3.8187	0.0854	0.0181	4.7164	0.0676	0.0133	5.0717
	2	0.0956	0.0264	3.6276	0.0771	0.0164	4.6982	0.0676	0.0133	5.0717
	0	0.1545	0.0464	3.3330	0.1173	0.0283	4.1373	0.0924	0.0190	4.8685
4	1	0.1357	0.0386	3.5109	0.1016	0.0226	4.4976	0.0852	0.0155	5.4808
5	0	0.1847	0.0625	2.9558	0.1346	0.0353	3.8078	0.1039	0.0228	4.5647
Gamma(a=4, b=2)										
k_1	k_2	m=10			m=12			m=14		
	0	0.0952	0.0217	4.3880	0.0786	0.0156	5.0425	0.0683	0.0119	5.7410
	1	0.0827	0.0173	4.7744	0.0703	0.0125	5.6304	0.0613	0.0095	6.4571
	2	0.0767	0.0165	4.6577	0.0669	0.0118	5.6911	0.0571	0.0087	6.5663
0	3	0.0772	0.0174	4.4405	0.0635	0.0118	5.3795	0.0563	0.0087	6.4616
	4	0.0729	0.0185	3.9463	0.0619	0.0123	5.0293	0.0537	0.0089	6.0620
	5	0.0721	0.0209	3.4501	0.0607	0.0133	4.5594	0.0529	0.0093	5.6641
1	0	0.1220	0.0312	3.9054	0.0991	0.0222	4.4603	0.0839	0.0167	5.0108
	1	0.0991	0.0215	4.6110	0.0822	0.0150	5.4831	0.0694	0.0111	6.2496
	2	0.0920	0.0205	4.4926	0.0767	0.0140	5.4977	0.0655	0.0101	6.4706
	3	0.0907	0.0220	4.1309	0.0752	0.0144	5.2116	0.0643	0.0102	6.2819
	4	0.0868	0.0239	3.6356	0.0728	0.0152	4.7981	0.0607	0.0104	5.8584
	0	0.1324	0.0331	3.9943	0.1051	0.0225	4.6790	0.0859	0.0161	5.3233
2	1	0.1150	0.0271	4.2361	0.0938	0.0183	5.1338	0.0793	0.0131	6.0363
	2	0.1059	0.0259	4.0938	0.0879	0.0171	5.1263	0.0729	0.0120	6.0746
	3	0.0977	0.0265	3.6919	0.0805	0.0167	4.8090	0.0686	0.0118	5.8404
3	0	0.1511	0.0418	3.6181	0.1159	0.0268	4.3243	0.0957	0.0188	5.0994
	1	0.1284	0.0329	3.9030	0.1008	0.0209	4.8260	0.0848	0.0147	5.7545
	2	0.1172	0.0320	3.6608	0.0941	0.0198	4.7484	0.0780	0.0136	5.7396
	0	0.1710	0.0534	3.2007	0.1306	0.0328	3.9875	0.1027	0.0223	4.6054
4	1	0.1426	0.0410	3.4820	0.1085	0.0244	4.4463	0.0897	0.0166	5.4097
5	0	0.2030	0.0673	3.0146	0.1442	0.0378	3.8167	0.1147	0.0247	4.6527

Table 5

		m=10			m=12			m=14		
Gamma(a=1, b=3)										
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0119	0.0039	3.0591	0.0096	0.0028	3.4139	0.0083	0.0022	3.7523
	1	0.0075	0.0018	4.1396	0.0064	0.0013	4.7570	0.0057	0.0011	5.2514
	2	0.0058	0.0014	4.2751	0.0051	0.0010	5.0565	0.0045	0.0008	5.8261
0	3	0.0047	0.0012	4.0463	0.0042	0.0009	4.9163	0.0039	0.0007	5.8074
	4	0.0038	0.0010	3.6757	0.0036	0.0008	4.6459	0.0034	0.0006	5.5684
	5	0.0031	0.0009	3.2498	0.0030	0.0007	4.2666	0.0028	0.0005	5.2175
1	0	0.0142	0.0041	3.4959	0.0112	0.0029	3.9154	0.0095	0.0022	4.3730
	1	0.0101	0.0025	4.0726	0.0085	0.0018	4.7822	0.0072	0.0013	5.4415
	2	0.0081	0.0020	4.1119	0.0071	0.0014	4.9439	0.0059	0.0010	5.6740
	3	0.0066	0.0017	3.8515	0.0058	0.0012	4.7958	0.0052	0.0009	5.6789
	4	0.0055	0.0016	3.4514	0.0050	0.0011	4.4955	0.0045	0.0008	5.4602
	0	0.0137	0.0039	3.4922	0.0106	0.0026	3.9896	0.0089	0.0020	4.5285
2	1	0.0104	0.0026	4.0098	0.0085	0.0017	4.8391	0.0071	0.0013	5.5331
	2	0.0089	0.0023	3.9318	0.0073	0.0015	4.8860	0.0061	0.0011	5.7245
	3	0.0077	0.0022	3.5505	0.0064	0.0014	4.5968	0.0055	0.0010	5.5541
3	0	0.0263	0.0090	2.9048	0.0196	0.0059	3.3260	0.0157	0.0042	3.7707
	1	0.0176	0.0052	3.3937	0.0136	0.0034	3.9809	0.0114	0.0024	4.6755
	2	0.0131	0.0038	3.3942	0.0103	0.0024	4.2866	0.0085	0.0017	5.0622
	0	0.0263	0.0087	3.0303	0.0188	0.0052	3.5886	0.0145	0.0035	4.1120
4	1	0.0194	0.0060	3.2290	0.0145	0.0036	4.0517	0.0117	0.0024	4.8302
5	0	0.0277	0.0102	2.7103	0.0189	0.0056	3.3866	0.0141	0.0036	3.9128
Gamma(a=2, b=3)										
k_1	k_2	m=10			m=12			m=14		
	0	0.0260	0.0067	3.8599	0.0214	0.0049	4.3952	0.0181	0.0037	4.9157
	1	0.0204	0.0044	4.5986	0.0175	0.0032	5.3772	0.0149	0.0024	6.1042
	2	0.0181	0.0040	4.5798	0.0153	0.0028	5.4220	0.0135	0.0022	6.2546
0	3	0.0156	0.0037	4.2428	0.0139	0.0027	5.2179	0.0123	0.0020	6.1392
	4	0.0140	0.0036	3.8541	0.0126	0.0026	4.9044	0.0112	0.0019	5.8882
	5	0.0121	0.0036	3.3739	0.0111	0.0025	4.4973	0.0099	0.0018	5.5110
1	0	0.0265	0.0065	4.0547	0.0216	0.0046	4.6935	0.0181	0.0034	5.3407
	1	0.0222	0.0051	4.3822	0.0188	0.0036	5.2175	0.0159	0.0027	5.9288
	2	0.0191	0.0044	4.3064	0.0164	0.0031	5.2641	0.0141	0.0023	6.1371
	3	0.0177	0.0044	4.0137	0.0146	0.0029	4.9950	0.0128	0.0021	5.9795
	4	0.0155	0.0044	3.5441	0.0133	0.0029	4.6298	0.0117	0.0021	5.6451
	0	0.0332	0.0094	3.5318	0.0257	0.0062	4.1218	0.0213	0.0046	4.5878
2	1	0.0256	0.0063	4.0527	0.0205	0.0042	4.8509	0.0172	0.0031	5.6110
	2	0.0214	0.0054	3.9601	0.0175	0.0036	4.9393	0.0153	0.0026	5.8855
	3	0.0194	0.0053	3.6332	0.0161	0.0034	4.7005	0.0138	0.0024	5.7632
3	0	0.0346	0.0097	3.5734	0.0265	0.0062	4.2520	0.0215	0.0044	4.9095
	1	0.0287	0.0075	3.8175	0.0229	0.0048	4.7633	0.0186	0.0033	5.5864
	2	0.0260	0.0072	3.6068	0.0200	0.0043	4.6355	0.0169	0.0030	5.6371
	0	0.0469	0.0157	2.9887	0.0336	0.0094	3.5912	0.0272	0.0065	4.1639
4	1	0.0351	0.0105	3.3333	0.0262	0.0062	4.2096	0.0213	0.0043	4.9928
5	0	0.0612	0.0222	2.7545	0.0429	0.0126	3.4166	0.0334	0.0082	4.0694

Table 6

		Gamma(a=3, b=3)								
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0368	0.0089	4.1285	0.0305	0.0064	4.7723	0.0260	0.0049	5.3339
	1	0.0304	0.0065	4.6542	0.0262	0.0048	5.4168	0.0223	0.0036	6.1394
	2	0.0267	0.0058	4.6349	0.0229	0.0041	5.5288	0.0204	0.0032	6.4292
0	3	0.0248	0.0057	4.3487	0.0212	0.0040	5.3086	0.0191	0.0030	6.3443
	4	0.0222	0.0057	3.8772	0.0197	0.0040	4.9588	0.0178	0.0030	6.0180
	5	0.0213	0.0062	3.4309	0.0187	0.0041	4.5339	0.0167	0.0030	5.6265
1	0	0.0357	0.0085	4.2005	0.0295	0.0060	4.9499	0.0247	0.0044	5.6381
	1	0.0319	0.0069	4.6486	0.0262	0.0048	5.4936	0.0224	0.0035	6.3345
	2	0.0296	0.0066	4.5116	0.0248	0.0045	5.5343	0.0214	0.0033	6.5110
	3	0.0283	0.0069	4.1285	0.0231	0.0044	5.1895	0.0203	0.0032	6.2639
	4	0.0271	0.0074	3.6513	0.0224	0.0047	4.7687	0.0194	0.0033	5.8785
	0	0.0412	0.0104	3.9573	0.0332	0.0070	4.7376	0.0278	0.0051	5.4564
2	1	0.0364	0.0085	4.2827	0.0295	0.0057	5.1948	0.0246	0.0041	6.0145
	2	0.0331	0.0080	4.1091	0.0274	0.0053	5.1606	0.0237	0.0038	6.2064
	3	0.0314	0.0085	3.6964	0.0260	0.0054	4.8392	0.0217	0.0037	5.8969
3	0	0.0515	0.0138	3.7276	0.0397	0.0089	4.4458	0.0328	0.0062	5.2516
	1	0.0448	0.0116	3.8721	0.0355	0.0073	4.8257	0.0296	0.0052	5.6940
	2	0.0407	0.0112	3.6214	0.0325	0.0069	4.6840	0.0272	0.0048	5.7139
	0	0.0673	0.0218	3.0822	0.0507	0.0134	3.7878	0.0403	0.0090	4.4647
4	1	0.0527	0.0158	3.3400	0.0409	0.0096	4.2732	0.0320	0.0063	5.0407
5	0	0.0747	0.0242	3.0914	0.0542	0.0136	3.9921	0.0419	0.0087	4.8301

		Gamma(a=4, b=3)								
k_1	k_2	m=10			m=12			m=14		
	0	0.0483	0.0111	4.3519	0.0403	0.0081	4.9879	0.0346	0.0061	5.6371
	1	0.0414	0.0086	4.8006	0.0349	0.0062	5.6298	0.0310	0.0048	6.5143
	2	0.0386	0.0082	4.7161	0.0321	0.0057	5.6168	0.0283	0.0043	6.5444
0	3	0.0364	0.0082	4.4150	0.0309	0.0057	5.3848	0.0271	0.0042	6.3792
	4	0.0342	0.0087	3.9417	0.0296	0.0059	5.0276	0.0258	0.0043	6.0568
	5	0.0332	0.0096	3.4525	0.0278	0.0061	4.5604	0.0249	0.0044	5.6630
1	0	0.0556	0.0133	4.1748	0.0456	0.0094	4.8536	0.0377	0.0069	5.4624
	1	0.0474	0.0104	4.5437	0.0396	0.0073	5.4161	0.0340	0.0055	6.1803
	2	0.0430	0.0096	4.4591	0.0362	0.0066	5.4466	0.0310	0.0049	6.3414
	3	0.0403	0.0098	4.1169	0.0340	0.0066	5.1634	0.0296	0.0048	6.1746
	4	0.0384	0.0106	3.6242	0.0322	0.0068	4.7615	0.0282	0.0048	5.8812
	0	0.0496	0.0133	3.7320	0.0391	0.0089	4.3825	0.0325	0.0064	5.0610
2	1	0.0408	0.0098	4.1733	0.0328	0.0065	5.0464	0.0283	0.0048	5.9243
	2	0.0366	0.0091	4.0055	0.0301	0.0060	5.0300	0.0255	0.0043	5.9407
	3	0.0329	0.0090	3.6460	0.0265	0.0057	4.6713	0.0240	0.0042	5.7643
3	0	0.0725	0.0199	3.6479	0.0547	0.0126	4.3255	0.0446	0.0090	4.9729
	1	0.0622	0.0159	3.9089	0.0470	0.0098	4.7797	0.0394	0.0069	5.7146
	2	0.0559	0.0153	3.6599	0.0445	0.0094	4.7384	0.0366	0.0064	5.7205
	0	0.0847	0.0255	3.3255	0.0648	0.0157	4.1283	0.0508	0.0104	4.8669
4	1	0.0721	0.0210	3.4397	0.0561	0.0126	4.4608	0.0453	0.0084	5.3668
5	0	0.0809	0.0258	3.1321	0.0590	0.0146	4.0358	0.0468	0.0095	4.9484

Table 7

		Gamma(a=1, b=4)								
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0071	0.0023	3.0505	0.0060	0.0018	3.4185	0.0050	0.0013	3.7136
	1	0.0046	0.0011	4.0518	0.0039	0.0008	4.6397	0.0034	0.0007	5.2705
	2	0.0035	0.0008	4.2188	0.0031	0.0006	4.9705	0.0028	0.0005	5.7063
0	3	0.0028	0.0007	4.0664	0.0025	0.0005	4.9708	0.0022	0.0004	5.7134
	4	0.0022	0.0006	3.7288	0.0021	0.0004	4.7110	0.0020	0.0003	5.6612
	5	0.0019	0.0006	3.2558	0.0018	0.0004	4.2854	0.0017	0.0003	5.3278
1	0	0.0072	0.0023	3.1198	0.0059	0.0016	3.6601	0.0050	0.0013	3.9408
	1	0.0052	0.0012	4.1688	0.0041	0.0009	4.8107	0.0036	0.0007	5.4889
	2	0.0042	0.0010	4.1721	0.0036	0.0007	5.0323	0.0031	0.0005	5.8621
	3	0.0035	0.0009	3.8514	0.0031	0.0006	4.8679	0.0027	0.0005	5.8057
	4	0.0030	0.0009	3.4415	0.0027	0.0006	4.4972	0.0023	0.0004	5.4764
	0	0.0082	0.0024	3.4295	0.0065	0.0016	3.9294	0.0053	0.0012	4.4863
2	1	0.0061	0.0016	3.8771	0.0050	0.0011	4.6131	0.0042	0.0008	5.2753
	2	0.0050	0.0013	3.8346	0.0042	0.0009	4.7483	0.0035	0.0006	5.5552
	3	0.0041	0.0012	3.5439	0.0034	0.0008	4.5514	0.0030	0.0005	5.5003
3	0	0.0107	0.0035	3.0345	0.0082	0.0023	3.6227	0.0067	0.0016	4.1338
	1	0.0079	0.0022	3.6591	0.0062	0.0014	4.4436	0.0050	0.0010	5.1528
	2	0.0065	0.0019	3.5001	0.0052	0.0012	4.4848	0.0043	0.0008	5.3380
	0	0.0142	0.0048	2.9561	0.0103	0.0030	3.4894	0.0079	0.0020	3.9913
4	1	0.0106	0.0032	3.3303	0.0078	0.0019	4.1731	0.0063	0.0013	5.0387
5	0	0.0200	0.0079	2.5450	0.0135	0.0043	3.1217	0.0102	0.0028	3.6375
		Gamma(a=2, b=4)								
k_1	k_2	m=10			m=12			m=14		
	0	0.0120	0.0029	4.0951	0.0101	0.0021	4.7272	0.0088	0.0016	5.3743
	1	0.0101	0.0022	4.6315	0.0084	0.0016	5.4354	0.0075	0.0012	6.2079
	2	0.0089	0.0019	4.5844	0.0075	0.0014	5.4501	0.0067	0.0011	6.3588
0	3	0.0079	0.0019	4.2799	0.0069	0.0013	5.2626	0.0062	0.0010	6.1950
	4	0.0071	0.0018	3.8814	0.0062	0.0013	4.9054	0.0056	0.0009	5.9413
	5	0.0065	0.0019	3.4117	0.0056	0.0013	4.4785	0.0053	0.0009	5.5811
1	0	0.0154	0.0042	3.6246	0.0127	0.0030	4.2058	0.0103	0.0022	4.6596
	1	0.0121	0.0028	4.4084	0.0100	0.0019	5.1274	0.0086	0.0015	5.8949
	2	0.0104	0.0024	4.3382	0.0087	0.0016	5.2670	0.0074	0.0012	6.1125
	3	0.0091	0.0023	4.0262	0.0078	0.0015	5.0314	0.0068	0.0011	6.0146
	4	0.0081	0.0023	3.5581	0.0071	0.0015	4.6539	0.0062	0.0011	5.7295
	0	0.0175	0.0049	3.5550	0.0139	0.0033	4.1509	0.0115	0.0024	4.7309
2	1	0.0136	0.0034	4.0590	0.0111	0.0023	4.9016	0.0094	0.0017	5.6394
	2	0.0118	0.0029	4.0216	0.0096	0.0019	4.9991	0.0081	0.0014	5.9251
	3	0.0107	0.0029	3.6575	0.0086	0.0018	4.7318	0.0073	0.0013	5.7667
3	0	0.0220	0.0064	3.4403	0.0162	0.0040	4.0547	0.0133	0.0028	4.6603
	1	0.0174	0.0047	3.7014	0.0138	0.0030	4.6086	0.0115	0.0021	5.3965
	2	0.0150	0.0043	3.5086	0.0122	0.0027	4.5595	0.0101	0.0018	5.4518
	0	0.0303	0.0096	3.1611	0.0221	0.0058	3.8113	0.0176	0.0039	4.4838
4	1	0.0245	0.0073	3.3714	0.0184	0.0043	4.2990	0.0147	0.0029	5.1281
5	0	0.0295	0.0106	2.7744	0.0208	0.0059	3.5302	0.0164	0.004	4.1429

Table 8

Gamma(a=3, b=4)										
k_1	k_2	m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0196	0.0046	4.2691	0.0162	0.0033	4.9448	0.0139	0.0025	5.5551
	1	0.0171	0.0036	4.7176	0.0145	0.0026	5.5146	0.0128	0.0020	6.3191
	2	0.0150	0.0033	4.6033	0.0130	0.0024	5.4985	0.0114	0.0018	6.3504
0	3	0.0138	0.0032	4.3332	0.0119	0.0023	5.2885	0.0105	0.0017	6.2779
	4	0.0123	0.0032	3.8864	0.0111	0.0022	4.9846	0.0097	0.0016	5.9929
	5	0.0113	0.0033	3.4140	0.0101	0.0022	4.5182	0.0092	0.0016	5.6256
1	0	0.0228	0.0061	3.7304	0.0187	0.0043	4.3195	0.0160	0.0033	4.8999
	1	0.0183	0.0041	4.4636	0.0152	0.0029	5.2224	0.0130	0.0022	6.0162
	2	0.0164	0.0037	4.4371	0.0135	0.0025	5.3533	0.0114	0.0019	6.1621
	3	0.0148	0.0036	4.1004	0.0125	0.0024	5.1740	0.0108	0.0018	6.1582
	4	0.0138	0.0038	3.6280	0.0118	0.0025	4.7686	0.0101	0.0017	5.8596
	0	0.0258	0.0070	3.6968	0.0207	0.0047	4.4076	0.0170	0.0034	4.9661
2	1	0.0209	0.0051	4.1256	0.0172	0.0035	4.9617	0.0142	0.0025	5.7276
	2	0.0184	0.0046	4.0200	0.0151	0.0030	4.9718	0.0129	0.0022	5.9034
	3	0.0168	0.0046	3.6433	0.0138	0.0029	4.7507	0.0119	0.0020	5.8045
3	0	0.0276	0.0073	3.7991	0.0212	0.0047	4.5460	0.0178	0.0033	5.3996
	1	0.0246	0.0063	3.9138	0.0193	0.0040	4.8752	0.0161	0.0028	5.7808
	2	0.0227	0.0063	3.6321	0.0181	0.0038	4.7285	0.0153	0.0026	5.7903
	0	0.0315	0.0096	3.2746	0.0242	0.0060	4.0384	0.0191	0.0040	4.7584
4	1	0.0272	0.0078	3.4742	0.0205	0.0046	4.4797	0.0167	0.0031	5.4278
5	0	0.0456	0.0154	2.9666	0.0322	0.0086	3.7335	0.0255	0.0057	4.5069

Gamma(a=4, b=4)										
k_1	k_2	m=10			m=12			m=14		
	0	0.0256	0.0062	4.1426	0.0213	0.0045	4.7247	0.0181	0.0034	5.2556
	1	0.0214	0.0045	4.7527	0.0184	0.0033	5.5758	0.0160	0.0025	6.3738
	2	0.0201	0.0043	4.7178	0.0171	0.0030	5.6736	0.0148	0.0022	6.5704
0	3	0.0196	0.0045	4.3885	0.0164	0.0030	5.3945	0.0142	0.0022	6.4320
	4	0.0190	0.0048	3.9336	0.0163	0.0032	5.0201	0.0142	0.0023	6.0801
	5	0.0182	0.0054	3.4017	0.0159	0.0035	4.5574	0.0138	0.0024	5.6326
1	0	0.0402	0.0116	3.4737	0.0317	0.0073	4.3558	0.0251	0.0049	5.1392
	1	0.0378	0.0106	3.5753	0.0290	0.0063	4.6325	0.0237	0.0042	5.6479
	2	0.0360	0.0114	3.1487	0.0280	0.0064	4.3692	0.0227	0.0041	5.4964
	3	0.0359	0.0141	2.5497	0.0277	0.0071	3.9048	0.0225	0.0044	5.1138
	4	0.0365	0.0199	1.8338	0.0275	0.0083	3.3156	0.0224	0.0048	4.6465
	0	0.0313	0.0074	4.2059	0.0251	0.0051	4.9607	0.0208	0.0037	5.7052
2	1	0.0289	0.0066	4.3892	0.0233	0.0044	5.3284	0.0196	0.0031	6.2251
	2	0.0270	0.0065	4.1299	0.0218	0.0042	5.1664	0.0183	0.0030	6.1789
	3	0.0257	0.0069	3.6987	0.0213	0.0044	4.8729	0.0178	0.0030	5.9351
3	0	0.0461	0.0124	3.7055	0.0358	0.0080	4.4681	0.0296	0.0057	5.2295
	1	0.0405	0.0103	3.9205	0.0318	0.0065	4.9023	0.0267	0.0046	5.8233
	2	0.0374	0.0102	3.6588	0.0295	0.0062	4.7323	0.0245	0.0042	5.7823
	0	0.0452	0.0133	3.3856	0.0332	0.0081	4.1163	0.0270	0.0055	4.9400
4	1	0.0390	0.0113	3.4654	0.0296	0.0065	4.5184	0.0244	0.0044	5.4996
5	0	0.0560	0.0197	2.8458	0.0409	0.0113	3.6236	0.0324	0.0074	4.3974

Table 9

Gamma(a=1, b=5)										
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0035	0.0010	3.5471	0.0029	0.0007	3.9892	0.0025	0.0006	4.4120
	1	0.0026	0.0006	4.2502	0.0021	0.0004	4.8994	0.0019	0.0003	5.6445
	2	0.0021	0.0005	4.2799	0.0018	0.0004	5.0433	0.0016	0.0003	5.8231
0	3	0.0016	0.0004	4.0168	0.0015	0.0003	4.9093	0.0014	0.0002	5.7540
	4	0.0013	0.0004	3.6987	0.0012	0.0003	4.6415	0.0012	0.0002	5.5348
	5	0.0011	0.0003	3.2902	0.0010	0.0002	4.2596	0.0010	0.0002	5.2468
1	0	0.0053	0.0015	3.4879	0.0042	0.0011	3.9730	0.0036	0.0008	4.4510
	1	0.0038	0.0010	3.9534	0.0032	0.0007	4.5554	0.0028	0.0005	5.2762
	2	0.0029	0.0007	3.9836	0.0025	0.0005	4.8055	0.0022	0.0004	5.4897
	3	0.0023	0.0006	3.7758	0.0021	0.0004	4.6815	0.0018	0.0003	5.5119
	4	0.0019	0.0005	3.4209	0.0017	0.0004	4.4312	0.0016	0.0003	5.3407
	0	0.0061	0.0019	3.2882	0.0048	0.0013	3.7252	0.0039	0.0009	4.1965
2	1	0.0043	0.0011	3.8999	0.0036	0.0008	4.6423	0.0030	0.0006	5.3484
	2	0.0037	0.0010	3.8561	0.0029	0.0006	4.7098	0.0025	0.0005	5.5220
	3	0.0031	0.0009	3.5264	0.0025	0.0006	4.4796	0.0022	0.0004	5.3930
3	0	0.0084	0.0027	3.1386	0.0062	0.0017	3.6595	0.0050	0.0012	4.1866
	1	0.0059	0.0017	3.4622	0.0048	0.0011	4.1916	0.0039	0.0008	4.8396
	2	0.0045	0.0013	3.3796	0.0037	0.0008	4.3120	0.0031	0.0006	5.0777
	0	0.0080	0.0028	2.8920	0.0057	0.0016	3.4474	0.0044	0.0011	3.9837
4	1	0.0057	0.0017	3.2946	0.0043	0.0010	4.1848	0.0034	0.0007	4.9543
5	0	0.0114	0.0043	2.6747	0.0078	0.0024	3.2579	0.0059	0.0016	3.8201

Gamma(a=2, b=5)										
k_1	k_2	m=10			m=12			m=14		
	0	0.0094	0.0024	3.8599	0.0077	0.0018	4.3952	0.0065	0.0013	4.9157
	1	0.0073	0.0016	4.5986	0.0063	0.0012	5.3772	0.0054	0.0009	6.1042
	2	0.0065	0.0014	4.5798	0.0055	0.0010	5.4220	0.0048	0.0008	6.2546
0	3	0.0056	0.0013	4.2428	0.0050	0.0010	5.2179	0.0044	0.0007	6.1392
	4	0.0050	0.0013	3.8541	0.0045	0.0009	4.9044	0.0040	0.0007	5.8882
	5	0.0044	0.0013	3.3739	0.0040	0.0009	4.4973	0.0036	0.0006	5.5110
1	0	0.0111	0.0030	3.7577	0.0090	0.0021	4.3173	0.0075	0.0016	4.8507
	1	0.0088	0.0020	4.4474	0.0073	0.0014	5.2664	0.0063	0.0010	6.0578
	2	0.0079	0.0018	4.3789	0.0064	0.0012	5.2446	0.0055	0.0009	6.1145
	3	0.0069	0.0017	4.0087	0.0059	0.0012	5.0629	0.0051	0.0008	6.0126
	4	0.0063	0.0018	3.5750	0.0054	0.0012	4.6620	0.0046	0.0008	5.6544
	0	0.0132	0.0037	3.5871	0.0103	0.0025	4.1580	0.0084	0.0018	4.7435
2	1	0.0104	0.0025	4.1917	0.0083	0.0017	4.9644	0.0072	0.0012	5.8836
	2	0.0094	0.0023	4.0513	0.0077	0.0015	5.0364	0.0064	0.0011	5.9463
	3	0.0083	0.0023	3.6295	0.0068	0.0014	4.7244	0.0058	0.0010	5.7669
3	0	0.0140	0.0040	3.5128	0.0107	0.0025	4.2268	0.0087	0.0018	4.8572
	1	0.0114	0.0031	3.7017	0.0089	0.0020	4.5386	0.0074	0.0014	5.3479
	2	0.0098	0.0028	3.5086	0.0080	0.0018	4.5337	0.0066	0.0012	5.4141
	0	0.0167	0.0051	3.2720	0.0122	0.0031	3.9825	0.0096	0.0021	4.6521
4	1	0.0140	0.0041	3.3967	0.0104	0.0024	4.3409	0.0083	0.0016	5.1838
5	0	0.0189	0.0066	2.8681	0.0135	0.0037	3.6323	0.0103	0.0024	4.3056

Table 10

		Gamma(a=3, b=5)								
		m=10			m=12			m=14		
k_1	k_2	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE	$Var(\hat{\varphi}_t)$	$Var(\hat{\varphi}_{tr})$	RE
	0	0.0112	0.0026	4.3798	0.0092	0.0018	5.0465	0.0081	0.0014	5.6893
	1	0.0100	0.0021	4.8530	0.0084	0.0015	5.6820	0.0071	0.0011	6.4440
	2	0.0088	0.0019	4.6644	0.0076	0.0014	5.6349	0.0067	0.0010	6.5532
0	3	0.0082	0.0019	4.3749	0.0072	0.0013	5.3637	0.0062	0.0010	6.3228
	4	0.0079	0.0020	3.9656	0.0068	0.0014	5.0006	0.0061	0.0010	6.0803
	5	0.0074	0.0021	3.4234	0.0064	0.0014	4.5558	0.0056	0.0010	5.6446
1	0	0.0139	0.0035	3.9673	0.0112	0.0024	4.6054	0.0097	0.0019	5.1780
	1	0.0120	0.0026	4.6393	0.0098	0.0018	5.4637	0.0083	0.0013	6.2797
	2	0.0110	0.0024	4.4762	0.0091	0.0017	5.4656	0.0079	0.0012	6.4433
	3	0.0107	0.0026	4.1732	0.0086	0.0017	5.2005	0.0074	0.0012	6.2446
	4	0.0100	0.0027	3.6459	0.0085	0.0018	4.8338	0.0071	0.0012	5.8905
	0	0.0175	0.0046	3.7918	0.0141	0.0032	4.4590	0.0114	0.0022	5.1038
2	1	0.0143	0.0034	4.1951	0.0119	0.0023	5.0863	0.0101	0.0017	5.9581
	2	0.0131	0.0032	4.0595	0.0108	0.0021	5.0347	0.0095	0.0016	6.0606
	3	0.0121	0.0033	3.6790	0.0098	0.0021	4.7681	0.0084	0.0015	5.8087
3	0	0.0180	0.0048	3.7407	0.0139	0.0031	4.4965	0.0113	0.0022	5.2530
	1	0.0157	0.0040	3.8816	0.0123	0.0026	4.7901	0.0099	0.0018	5.6545
	2	0.0141	0.0039	3.6032	0.0114	0.0024	4.6788	0.0093	0.0016	5.6454
	0	0.0227	0.0071	3.1917	0.0165	0.0042	3.8816	0.0131	0.0029	4.5219
4	1	0.0186	0.0054	3.4402	0.0141	0.0032	4.4181	0.0117	0.0022	5.3836
5	0	0.0289	0.0102	2.8303	0.0204	0.0057	3.5579	0.0160	0.0037	4.3215
		Gamma(a=4, b=5)								
k1	k2	m=10			m=12			m=14		
	0	0.0160	0.0036	4.3874	0.0138	0.0027	5.1922	0.0115	0.0020	5.8119
	1	0.0144	0.0030	4.7469	0.0121	0.0022	5.5488	0.0107	0.0017	6.3823
	2	0.0130	0.0028	4.6796	0.0111	0.0020	5.5889	0.0097	0.0015	6.4906
0	3	0.0124	0.0028	4.4175	0.0103	0.0019	5.3654	0.0090	0.0014	6.3454
	4	0.0116	0.0029	3.9565	0.0099	0.0020	5.0353	0.0087	0.0014	6.0662
	5	0.0113	0.0033	3.4414	0.0096	0.0021	4.5500	0.0084	0.0015	5.6561
1	0	0.0182	0.0045	4.0702	0.0151	0.0032	4.7137	0.0122	0.0023	5.2513
	1	0.0155	0.0033	4.6664	0.0129	0.0023	5.5925	0.0108	0.0017	6.4377
	2	0.0147	0.0033	4.5141	0.0120	0.0022	5.5201	0.0105	0.0016	6.5461
	3	0.0145	0.0035	4.1383	0.0118	0.0023	5.2314	0.0102	0.0016	6.2706
	4	0.0140	0.0038	3.6681	0.0114	0.0024	4.7883	0.0100	0.0017	5.8991
	0	0.0220	0.0055	4.0005	0.0177	0.0037	4.7525	0.0147	0.0027	5.4482
2	1	0.0196	0.0045	4.3490	0.0159	0.0030	5.2591	0.0133	0.0022	6.1597
	2	0.0183	0.0044	4.1458	0.0150	0.0029	5.1950	0.0125	0.0020	6.1914
	3	0.0171	0.0046	3.6795	0.0140	0.0029	4.8407	0.0119	0.0020	5.9262
3	0	0.0253	0.0068	3.7007	0.0198	0.0044	4.4644	0.0164	0.0031	5.2159
	1	0.0224	0.0056	3.9579	0.0178	0.0036	4.9401	0.0144	0.0025	5.7707
	2	0.0208	0.0057	3.6788	0.0166	0.0035	4.7850	0.0138	0.0024	5.8128
	0	0.0293	0.0088	3.3482	0.0223	0.0054	4.1639	0.0179	0.0036	4.9041
4	1	0.0257	0.0074	3.4969	0.0200	0.0044	4.5270	0.0163	0.0030	5.4960
5	0	0.0347	0.0117	2.9620	0.0252	0.0066	3.8004	0.0199	0.0043	4.6140

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 293-304

ENVIRONMENTAL SITUATION OF RAVI CHEMICAL COMPLEX

Mohammad Rafiq Khan[§], Hoor Shmail Rana, Taha Atique,
Sehar Salim, Urooj Fatima, Khadija Shahid,
Mohammad Nawaz Waseem and **Ameer Mustafa**
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
[§]Corresponding Author Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

The goal of this study was the analysis of the entire system of Ravi Medical Complex and understanding of the intoxication of environment that takes place due to the formation of unlimited chemicals produced by it. The expected outcome of this study is not only important, but also lethal in determining the lethality of Ravi Chemical Complex, considering the fact that it is located on the hub of Pakistan, and all its pollutants can spread far and wide. Another concern is the fact that close by passes Ravi River, which is the waste way for all chemical intoxicants produced here. A variety of research approaches were followed to gather the information that was pertinent to the subject in order to carry out this study, which focused on the harmful effects that the chemical industry has on the environment and the general public. Both primary and secondary data were used in the study; they were combined to get results that were pertinent to the subject and to draw a conclusion based on the research findings. On the onset of the study multiple survey questionnaires were designed to target the general public, the huge chunk of employees working at Ravi Chemical Complex and for the administration that is running Ravi Chemical Complex for responses. Along with the survey questionnaires for employees and public, an interview with the Operations Manager at Ravi Chemical Complex was conducted. In-depth research and analysis were conducted on the information found in sources, including the data and discoveries. After the collection of data were complete, the data were sorted and computed and analysed applying Google. Form Software. The results were recorded in the form of bar charts and pi diagrams. The results were discussed and compared to make recommendations for control of pollutants.

1. INTRODUCTION

Chemical pollution is the contamination of our environment with chemicals that are not naturally present. But how widespread is chemical pollution, and how does it affect our health and the environment is a very serious question. Chemical intoxication is caused by chemical pollutant exposure and can have immediate or delayed effects that appear weeks or even months after the exposure. Severe chemical intoxication may result in the death of the person who inhales an excessive amount of such substances. Chemical compounds are organic or inorganic chemicals that are major contributors to chemical pollution. The most common chemical pollutants are those that are used over large areas and are persistent, meaning they do not degrade easily in nature. All these are important raw materials used to make essential products that boost the economy, and also decrease the budget for their import. Acetic acid is used in the manufacture of acetic anhydride, cellulose acetate, vinyl acetate monomer, acetic esters, chloroacetic acid, plastics, dyes, insecticides, photographic chemicals, rubber and many others. The most essential use of acetic acid is in food preservation as preservative. As food industry is enormous in a country like Pakistan, and our production of food related products is much more in quantity as compared to import, we have to keep up with the requirement and demand of food preservatives that is used in many products such as all jar related products, like jars, sauces etc. Huge quantities of acetic acid are also used to make ink, dyes, pesticides, pharmaceuticals rubber and plastics. On the other hand, other major produce of Ravi chemical complex is Rayon, which is majorly used to produce cloth and fibre. Pakistan has the second-best quality of cotton in the world and is a master on clothes, with 57 percent of country’s total exports coming from cloth alone. Thus, large quantity of Rayon is produced in Pakistan, and it is essential to keep up with the economy of the country. Revenue from this segment amounts to almost US $6.01 billion and the largest segment is Women’s apparel. Lastly, alcohol is the most common ground organic compound, used in perfumes, medicines and is the most abundant produced organic chemical in industry. The reason behind its abundant production is that it is used for the synthesis of other compounds, which leads to the formation of other chemical products (Aljamali et al., 2020).

All these chemicals are essential for the economy of Pakistan, and to fulfil the dire needs of the enormous population that exists within Pakistan, thus the option of reducing the production of these chemicals is out of the equation. The main aim should be to have a proper disposal system of these chemicals, so that they may not interfere with the cycle of life, and environmental pollution can be reduced as much as it can be (Ketin et al., 2014).

That is the status of pollution caused by one chemical complex located in Lahore. If the whole country is focused upon, the picture will be alarming because consumer goods produced for consumption by 220 million population will be very high and large scale production will also translate into big quantity of pollution from the chemical industries, The most common chemical pollutants are those that are used over large areas and are persistent, Most pesticides, herbicides, and insecticides used in agriculture and gardening are examples. Similarly, chlorinated solvents used in many industrial processes and dry-cleaning operations may be cited in this context (Golbabaie et al., 2012). Chemical contaminants are classified as naturally occurring or man-made based on their chemical structure. Harmful pollutants such as arsenic (As), lead (Pb), cadmium (Cd), and mercury (Hg), as well as smog and air-borne particulate pollution in large cities, have been documented since ancient Rome and Athens, whose citizens suffered from contaminated water supplies, air, cooking and eating utensils, and food.

While chemical toxicity is not new, the phenomenal 40-fold increase in chemical production and resource extraction over the last 100 years poses a serious risk to humanity.

Emissions of pollutants can be continuous but they are often under-reported and there is great variability in reported values. The emission, dispersal, and exposure of hazardous chemical pollutants and their mixtures are frequently sporadic and not time or space confined. This is the primary reason for humans' increasing chronic exposure to them. Evidence of their global migration can be found in the form of airborne particles, gases and aerosols, waterborne suspended particles, and dissolved pollutants (Rovira et al., 2021).

There is a large variety of pollutants that could poison the soil. Examples of the most common and problematic soil pollutants can be Lead, Mercury, arsenic, copper, zinc and nickel. Soil pollution affects plants, animals and humans alike. If the soil is contaminated, home-grown vegetables and fruits may become polluted too. This happens because most of the soil pollutants present in the soil are extracted by the plants along with water every time they feed.

The chemical industrial activity also causes water pollution. The most common water pollutants are crude oil and petroleum products, chlorinated solvents such as TCE, PCE, 1,1,1-TCA, carbon tetrachloride, and Freons, petroleum solvents such as benzene, toluene, xylenes, ethylbenzene, organic solvents and chemicals such as acetone, methyl ethyl ketone, alcohols such as ethanol, isopropanol; or oxygenate compounds such as MTBE, antibiotics and other pharmaceuticals, perchlorate salts used in rocket fuels and many others (Aljamali et al., 2020).

More than 10,000 chemicals are used or find their way into the modern food supply. Food chain pollution endangers humans directly through the consumption of contaminated food contamination can occur at several stages before consumption, including crop or forage or animal production and harvesting, as well as post-harvest storage, processing, transport, and processing. Heavy metal(loid)s, pesticides, dioxin, PCBs, antibiotics, growth-promoting substances, packaging residues, preservatives, and excess nutrients (e.g., nitrate) have all been found to contaminate food at levels that are above acceptable (Dutkiewicz et al., 1992). Through soil, surface water, groundwater, or aerial deposition, this affects vegetables, grains, fish, and livestock. Healthy soils are essential for safe, healthy food, ecosystem service delivery, climate change mitigation, abundant food and fiber production, pollutant attenuation, and freshwater storage, all of which are critical to the global food supply's sustainability. Nearly one-third of the Earth's total arable land has been lost in the last 40 years due to soil erosion, desertification, urban expansion, and contamination (Rovira et al., 2021). Nearly one-third of the Earth's total arable land has been lost in the last 40 years due to soil erosion, desertification, urban expansion, and contamination. Contaminated soils with heavy metals and pesticides result in the loss of productive agricultural land and jeopardize food production and quality.

Goal and Objectives

The goal of the study was to highlight the impact Ravi Chemical Complex on its employees and general society in Pakistan. The objectives were as under

1. Collection of secondary data
2. Collection of primary data
3. Sorting and computation of data.
4. Analysis of data
5. Interpretation of results
6. Conclusion and recommendations

2. RESEARCH METHODOLOGY

The research design and process involved the stages as given in the form of objectives. The procedures are explained below.

Sampling

A sample is a section or a single item that serves as an example of a larger whole or group, especially when it is presented for analysis or displayed as evidence of the overall product's quality. This is especially true when the sample is used to demonstrate the whole product's quality. An interview as part of the investigation was held with an executive which turned out to be the primary cornerstone for the information we obtained. The executive from the Ravi Chemical Complex, largely regarded as Pakistan's most renowned Chemical Industry, was the person who was interviewed for this article. The executive was directly questioned for his opinions and ideas on the issue of pollution brought on by the expanding chemical sector. These inquiries cantered on the executive's and the executive's perspectives on the problem. He was also questioned on the steps his business is taking to avoid pollution from spreading further and to build a better, more sustainable, and pollution-free future. The future was the subject of this query.

The second step of the process involved conducting a questionnaire-based interview with a former employee of the highly regarded Ravi Chemical Complex. A number of queries were directed toward the employee.

Additionally, a questionnaire for the general public was created and distributed to a group of individuals for feedback. The sample's participants were asked about the difficulties they are currently experiencing as a direct result of the rise in pollution levels, as well as the ways in which ongoing exposure to contaminated air has harmed their health. In addition to being questioned about their views on pollution, participants were also asked what they expected their future to hold.

Survey Questionnaires

Three survey questionnaires were designed: for the general public, the employees of our chosen chemical complex and interview for executive. The questions ranged from simple yes or no to short answer-based ones.

1) **Questionnaire for General Public:**
   Questions that were asked:
   i. Name
   ii. What sort of Pollution according to you is affecting the environment of Pakistan the most?
   iii. Explain your answer to the previous question. How is it affecting you?
   iv. To what extent do you think this sort of pollution has impacted you?
   v. Who do you think is responsible for the pollution?
   vi. Do you think shifting the industries to rural areas has made any progress in terms of pollution?
   vii. Has the pollution affected your health in any way?
   viii. Do you think chemical industries contribute most towards pollution?
   ix. Do you believe the industries should keep working the way they do or they should adopt green chemistry methods (using safer chemical products that are less harmful for the environment)?
2) **Questionnaire for the Administration of Ravi Chemical Complex**:
   Questions?

3) **Questionnaire for the Employees of Ravi Chemical Complex:**
   Questions that were asked:
   i. Age
   ii. How long have you been working at Ravi Chemical Complex?
   iii. Has the pollution from the industry affected your health in any way?
   iv. To what extent chemical pollution from the industry has impacted you?
   v. In your chemical career have you ever experienced a chemical-related emergency such as a spill or injured co-worker due to exposure?
   vi. Do you think working in the chemical industry increases fatal diseases?
   vii. Tell us about any safety-related training you have received when it comes to the handling of chemicals and other toxic substances.
   viii. In your opinion, what do you think has been the biggest change in the chemical industry in the past 10 years?
   ix. Why is this job with Ravi Chemical Complex right for you at this time in your career?

Collection of the Secondary Data

To collect secondary data a number of research papers and reports to acquire access to previously published work on the issue. In addition to that the websites of the concerned establishments were visited. The secondary thus collected formed a good baseline to clear about materials used to make different plastic products, their production lines, technological progress in the focused field.

Collection of the Primary Data

The primary data was collected by taking responses against appropriately designed questionnaires depending upon the status of the respondents in the current study: Executive/Managers, of the establishment targeted, Workforce/Employees and the Locals affected by the pollutants of the polymer industry. Thus, three types of questionnaires designed were Interview Questionnaire, Employee Response Questionnaire and Public Response Questionnaire. The information collected from the executives was used to ease up serving the questionnaire to the employees of Pakistan Polymers to know their perspective on the ways in which they believe the mills was contributing to the degradation of the environment and looking after the employee welfare. The third kind of questionnaire was served to the general public to know their perspective on the ways in which they believe the mills are contributing to degradation of the environment and the extent to which people were being affected by pollutants.

Processing of the Data

The data collected in the form of responses was sorted, computed and analyzed applying Google. Form Software

Interpretation of Results

The results of the analysis were recorded in the form of pi bar diagrams, and bar line graphs. The results of interview of the executives are reported as descriptive research.

3. RESULTS AND DISCUSSION

Results of the Employees Survey

What is your Age?

4 responses

22

35

27

50

The age group is around 20-50.

How long have you been working at Ravi Chemical Complex?

4 responses

2 years

10 years

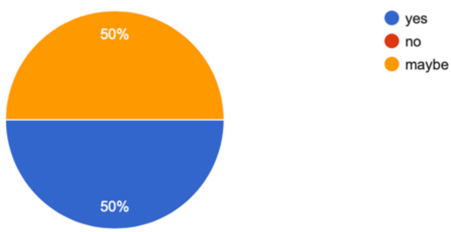
5 years

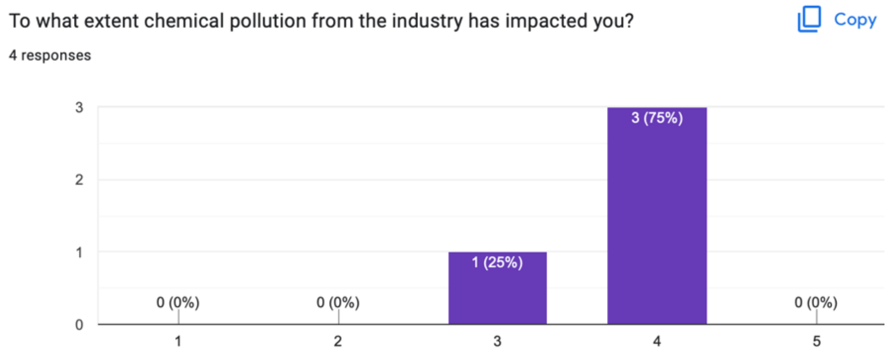
15 years

Employees are working there for almost 9-10 years.

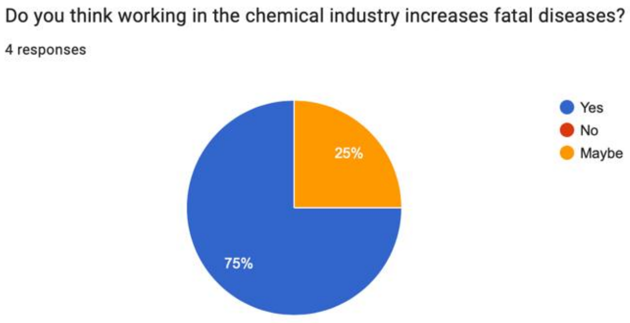
Has the pollution from the industry affected your health in any way?

4 responses





50% of the people believe that the pollution from industry has affected them and 50% say it may have some way or the other. Moreover, almost 75% of the people say that the pollution from the factory has impacted them a lot.



100% of the employees admit that they have experienced a chemical-related emergency such as spills and injured coworkers. Furthermore, 75% of the employees agree to the fact that chemical industry increases the number of fatal diseases for the people working there.

Tell us about any safety-related training you have received when it comes to the handling of chemicals and other toxic substances.

4 responses

Wearing gloves and lab coats to protect ourselves and wearing masks

Mask and gloves

Gloves and mask

Wearing gloves and tackling the situation properly with care .

In your opinion, what do you think has been the biggest change in the chemical industry in the past 10 years?

4 responses

More machines

More toxic now Bcz of the chemicals used

Technology

When asked about the safety measures and the training provided most of them mentioned the use of masks, gloves, lab coats etc. to protect themselves. And according to the employees, the biggest change in the industry in the past 10 years was the new technology, machinery, and ideas.

Why is this job with Ravi Chemical Complex right for you at this time in your career?

4 responses

Income

Less education

Because of the survival

Because I have to meet the needs of my family

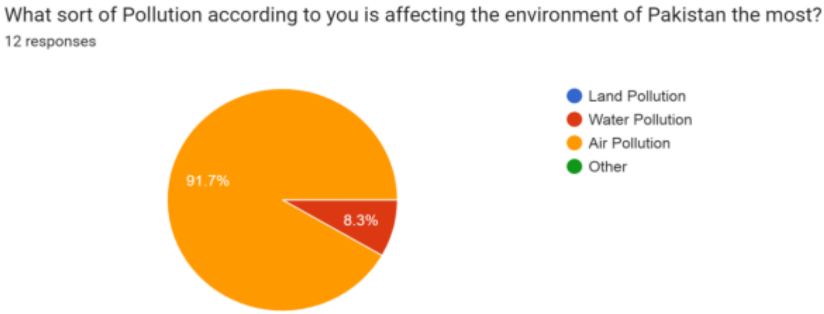
A most important reason for the employees to keep working there is the income for survival and to meet the needs of today.

Description:

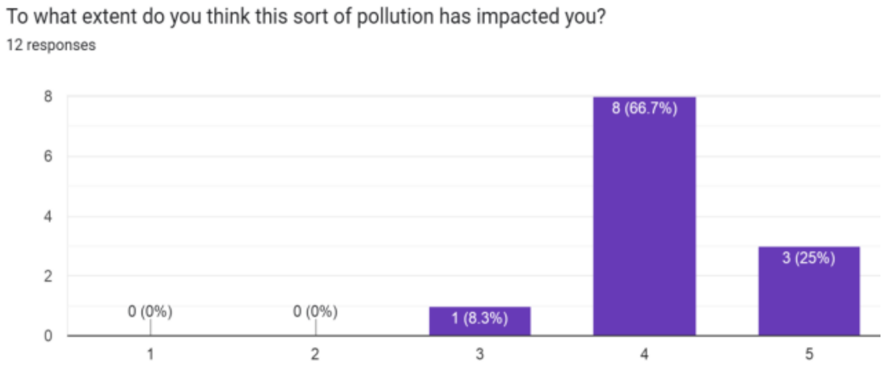
Hence, after looking at the results of the survey we got to know that the employees are facing a lot of problems too. They agree that the chemicals from the industry are affecting their health and increasing the number of fatal diseases in them which is not only hurting them but their family.

They do use safety precautions to keep themselves safe from disasters, but the chemicals do impact them in one way or the other. One of the people who filled out the questionnaire said that once he experienced a chemical disaster himself and was burnt and hospitalized. Moreover, they agree that they are doing this job because of survival in society and meeting the needs and fulfilling the basic necessities of their family.

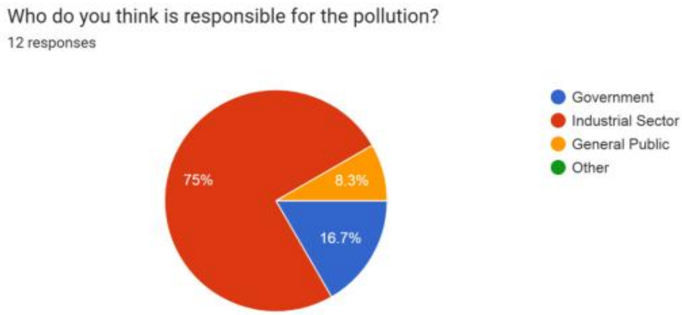
Results of the General Public Survey:



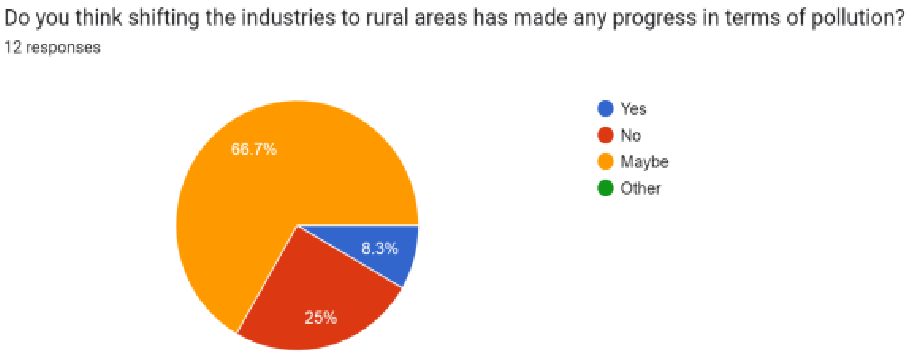
91.7% of the general public believes that most pollution comes from the air.



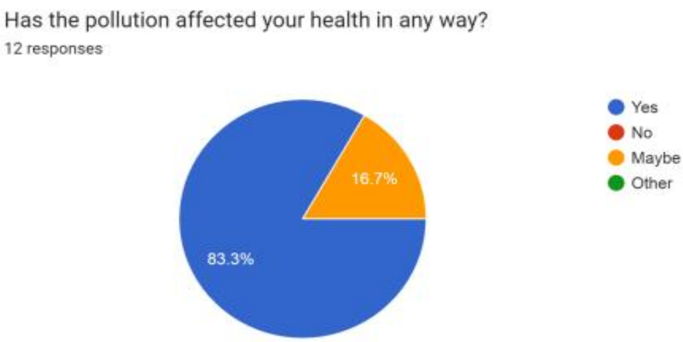
Almost 80% of the people believe that air pollution is having a great impact on them and their health.



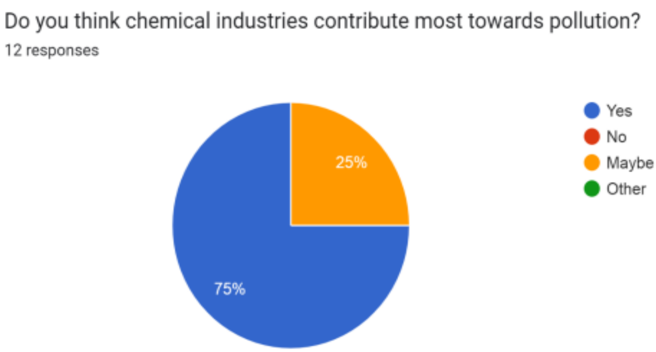
Industrial sector is the reason of pollution according to 75% of the people.



66.7% people believe that shifting industries to the rural areas has made progress in terms of pollution.

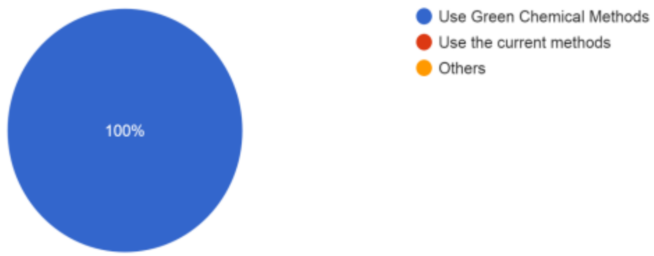


The general public also believes that pollution from the chemical industries affects their health in many ways almost 83.3% response proves this.



75% of people agrees that chemical industries contribute towards air pollution as well as land pollution.

Do you believe the industries should keep working the way they do or they should adopt green chemistry methods(using safer chemical products that are less harmful for the environment)?
12 responses



100% of the responses are in favor of the use of green chemical methods to protect the environment

Description:

Hence, after gathering the results of the general public's survey, we got to know kay people think that the most common type of pollution affecting the environment of Pakistan is Air Pollution, as the quality of the air is getting worse day by day, it has impacted their health greatly. According to them, the government is not taking adequate measures to tackle these issues concerning the health of the public. They were of the opinion that by shifting the factories to the rural areas, there has been a positive impact, as a significant reduction can be seen in the pollution. They also believed that chemical industries played a huge part in causing pollution as they emit harmful particles that negatively affect the human health. Lastly, they thought that to bring less harm to the environment, the industries should use new green chemical methods, that entail using safer products that are less harmful for the environment instead of just using the current methods.

Interview:

Along with the survey questionnaires for the primary research, we also conducted an interview with the Operations Manager at Ravi Chemical Complex. In the interview he talked about how a Chemical Complex contributes to pollution, in which the most eminent is "Air Pollution". As the chemicals are produced, the emissions from the manifesting of these chemicals tends to contains numerous harmful and toxic substances which can have negative effects on the human health and on the environment as well. The production of these chemicals is also important because these chemicals are used in our daily lives through different means such as, in our soaps, table salt, baking soda, sugar, soaps and etc. to make our lives more convenient and comfortable so, there production can't be stopped all together due to the release of the chemical eminent but, adoption of different ways can be used to minimize their harmful effects on the humans and the environment. New and improved methods for the production of these chemicals are being used. The Government of Pakistan is also contributing towards reducing the pollution in the country and is guiding the Chemical Complexes to use more sustainable and environment-friendly methods to produce the chemical. Hopefully in the near future by the use of such environmentally friendly methods, the environment will become healthier with minimum pollution of all sorts.

5. ACKNOWLEDGEMENT

We are extremely grateful to Dr. Shahid Amjad Chaudhry, Rector, Lahore School Economics for patronization of our research and sanction of expenditure involved and providing us requisite facilities to participate in 20th ISOSS Conference held in Lahore.

REFERENCES

1. Aljamali, N.M., Alfatlawi, I.O., Husien, N.S.M., Hussein, A.A.A. and Almosawy, M.G.A.A. (2020). Review on Effect of Chemical Waste on Environment. *International Journal of Green Chemistry*, 6(1), 20-29.
2. Dutkiewicz, T., Rolecki, R., Konczalik, J. and Swiatczak, J. (1992). The impact of the chemical industry on the human environment. *Polish Journal of Occupational Medicine and Environmental Health*, 5(1), 13-26.
3. Golbabaie, F., Eskandari, D., Azari, M., Jahangiri, M., Rahimi, A. and Shahtaheri, J. (2012). Health risk assessment of chemical pollutants in a petrochemical complex. *Iran Occupational Health*, 9(3), 11-21.
4. Ketin, S., Sacirovic, S., Plojovic, S., Skrijelj, R. and Biocanin, R. (2014). Method for comparison of the maximum range of chemical pollutants. *Russian Journal of General Chemistry*, 84, 2677-2680.
5. Rovira, J., Nadal, M., Schuhmacher, M. and Domingo, J.L. (2021). Environmental impact and human health risks of air pollutants near a large chemical/petrochemical complex: Case study in Tarragona, Spain. *Science of the Total Environment*, 787, 147550.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 305-328

ROLE OF STATISTICS IN FRAMING APPROPRIATE STRATEGY FOR RECOVERY OF PAKISTAN ECONOMY

Mohammad Rafiq Khan
Department of Environmental Science and Policy
Lahore School of Economics, Lahore, Pakistan
Email: drrafiq@lahoreschool.edu.pk

ABSTRACT

This keynote develops an interlink between cause, effect and measure that led to the destabilization of Pakistan economy and suggests some measures to stabilize it by discovering the missing link between Microeconomic and Macroeconomic models in the light of the local conditions and environment of Pakistan and thrashing out where the leadership has remained trapped. The primary question in this context may be who are to be blamed for this economic downfall? The author sincerely feels that not only politicians, bureaucrats and financially defined upper class are to be blamed but also the experts such as economists, statisticians and others. The statistics has equipped the economic experts with statistical models that identify cause and effect but fail to offer any measure. Unfortunately, this track has translated into transformation of discipline of Economics into Econometrics. Most of the research is aimed at mass award of higher degrees without any significance in applied context. The keynote concludes that unless the leadership doesn't realize its faults, it is not possible to correct our position to recover Pakistan economy. It suggests need for active multidisciplinary search of solutions at both macro and micro level that may be helpful in pulling Pakistan economy back to normal. To do so the experts of statistics can play a potential role. How it can do so? This article answers this question effectively.

1. INTRODUCTION

Before talking about something, it is imperative to know that thing significantly. Thus, it is imperative to know salient features of the historical landmarks to know clearly what has been the shape of Pakistan economy in the past and what is its current status to design an appropriate strategy for its future development?

When India was partitioned in 1947, the lion's share of resources went to its twin brother Bharat. Pakistan was left to solely depend on traditional agriculture for its survival. There was no large scale industry and it had a very small industrial sector wit, insignificant industrial base. Thus Pakistan had to take start from a scratch to build up its economy. If we compare our techno-economic status at that time with that we have as nuclear power now, the pace of progress is worth appreciating. The analytical view of this march from nothing to something has been presented by different experts both at home and abroad have divided the economic life of Pakistan from 1947 to 2019 into the following periods:

Period 1 (1947-1958)

Period 1 is characterized by initial buildup of mercantile capital with highly significant state role and neglect of the agriculture sector. Two factors that helped in transformation of agriculture based mercantile capital into industrial capital were devaluation of Indian currency in 1949 that led to the suspension of trade between India and Pakistan which finally translated into blocking of major imports from India to Pakistan and Korean Boom (Amjad, 2008). The Korean Boom that spread over 1950 to 1952 gave s fairly good chance to traders to make high profits for subsequent investment in installing new industries to produce consumer goods later under Import Substitution Policy that followed the Korean Boom. This gesture was further supported by the state protection of the local industrial products and high duties on imports.

The state created Pakistan Industrial Development Corporation that played an extensive role to partially cover the consumer goods home needs and partially helping the entrepreneur in the private sector through the mechanics of installation, operation and sale to interested parties for private or corporate ownership.

The reason for the neglect of agriculture sector by state was on the assumption that Pakistan was self-sufficient in agricultural produce for the survival of its residents (Hasan, 2007) and will be able to meet its food needs comfortably in near future because the rate of population increase was low (1.4%). This dream was shattered by wheat crop failures of 1952 and 1953 and the unexpected rise in increase of population resulted into increased food grain demand that had to be met through supplies of food grain from USA under PL-480 Scheme that initiated in mid-fifties and lasted till mid-sixties and thus continued for almost a decade.

The import of food grain also reduced the investment in industry as the capital was to be essentially invested in import of food stuffs. The second setback to the industry was that there was a slump in the home market of food products (Hamid, 1974). Another setback was the tighter controls on foreign exchange which discouraged exports of the manufactured goods by industrialists.

Period 2 (1958-1971)

Period 2 was triggered with the first military takeover accompanied by pro-western policies due to the message of significantly high increase in foreign aid. It provided some attractions for investment by entrepreneur from both home and abroad. Ayyub regime also banned the trade unions to make business environment more conducive for the investor. This resulted into mass industrialization and thus Pakistan encountered an industrial boom with a high growth rate.

The regime did not neglect the agriculture sector and tried to encourage the farmer by giving subsidy on fertilizer and plant protection (Rashid, 2008). The agriculture sector was further supported by the Green Revolution and growth rate of agriculture sector increased from 3.5% in 1960-65 to 4.1% in 1965-70 (Hasan, 2007). The overall growth rate of economy in this period was 6.7% which was comparable with some front line Asian countries. So much so, Pakistan started being referred as a development model for many developing countries.

The major setbacks to the Pakistan economy in this period were the emergence of 1965 War and nationwide political turmoil by civil society against Ayyub Regime which ended into Martial Law of 1969 and take over by Yahiya Khan whose short period had to face 1971 War. The drop scene of this event was the separation of East Pakistan.

In spite of a fairly good progress in agriculture sector, the Country was pulled into another severe drought in 1966-67 which translated into grain shortage. Moreover, the defense budget increased from Rs. 5.5 billion (2.8% of GDP) in 1960-65 to 6.89 billion (3.5% of GDP) in 1965-70 which in actual terms was 4.0% of GDP.

In spite of all these negative forces, with 6.7% overall growth rate of economy, this period is considered as “successful period of economic management” (Hasan, 2007).

Period 3 (1971-1977)

Period 3 started with Bhutto’s taking over as first civilian martial law administrator, perhaps never heard before in the history of the world. This start was with the nation in extreme shock due to separation of East Pakistan from West Pakistan to create an independent nation Bangladesh. The public at large was mentally prepared to accept him in any form. He came into the field with the philosophy of socialism that guaranteed on paper or verbally, ‘Roti, Kapra and Makan’ to every one which fortified his acceptability.

Under the said philosophy, Bhutto attempted a change from private to public ownership and thus tried to nationalize major sectors particularly industry and education. He rather tried to have full control of educational institutions as students those days were the major political force in the Country.

The consequences of this unplanned change were disastrous. Extraordinary encouragement of workers against the owners resulted into almost no discipline in industry. The industrial units started falling sick. This sickness ultimately translated into closure of many units. Thus, the growth of industry stopped and the attempted change set the stage for a major bias against export strategy.

The economic consequences included increase in the defense establishments, expansion of government departments, and additional burden of education expenditure due to nationalization and elimination of private enterprise. The worst of all were the dents that could be clearly seen on the structure of the public institutions, particularly in education sector, the major result of which was the lawlessness. The investors drained their capital out of the country and there was a drastic reduction in foreign exchange reserves and foreign transmittances.

To reverse the effects, the succeeding governments tried to privatize the nationalized hunts but some major segments like banking, insurance and industry remained with the state till late 1990s (Hasan, 2007)

Period 4 (1977-1988)

Period 4 started with military takeover under the leadership of General Zia-ul-Haq. This period is marked by successful governance with the help of centrally controlled democratic forces. It didn't encounter any significant labor unrest or student agitation. There was no novelty and no innovation, yet it was helpful in getting Pakistan out of the economic recession that it received from the previous regime.

The economy grew at a reasonable growth rate of 6.6 for the following major reasons:

1. Long term Tarbela Dam Project was completed as a result of which the cultivated land increased by 10 million acres (Hasan, 2007).
2. Fertilizer and cement installations started earlier maturity and thus became functional to add value to the performance of the manufacturing sector and also to the overall economy.
3. The state support to Afghan Mujahidin at War against the Soviet Union being basically in interest of USA, led to transfer of huge aid from America to Pakistan and thus Pakistan was helped indirectly. As the economic situation was favorable, there were extensive remittances from abroad and foreign investment.

There were many structural problems that remained unattended by the government in this period. The economic situation started deteriorating in later half of the decade (1985 to 1990). The Pakistan economy faced severe challenges in terms of declining economic growth, accelerating inflation, increasing poverty and widening income inequality. These also included the poor climate for private sector investment, heavy dependence on exports on cotton based goods and the inelasticity of the tax system (Hasan, 2007).

Some of the problems highlighted above were intensified towards the end of 1980s. Had President Zia survived to live a few years more, the consequences of these overlooks would have enriched his accountability.

Period 5 (1988-1999)

Period 5 is characterized by the rule of the rotational democratic political governments of Benazir and Nawaz Sharif. There is hardly any doubt that this period had to absorb the shocks that previous regime that had to absorb if it had survived.

Over 1990s, weak macroeconomic management under the new democratic governments of the time, lack of commitment to execute competent structural reforms and corruption in public spending resulted in poor economic outcomes.

The structural problems of public debt, weakness in exports, poor revenue collection, inadequate investments in human capital and physical infrastructure, and low quality of public institutions and governance further added to the pressures facing the economy.

Economic growth slowed down to an average of 4% by the end of the 1990s from 6.7% per annum in the 1980s with more and more people falling below the poverty line while investment rate decelerated from an average of over 19% of GDP to 15.6% by 1999 (MCB, 2007). The economy passed through crisis after crises during the decade.

Inappropriate sequencing of financial reforms in the early 1990s particularly introduction of foreign currency accounts and the use of short-term commercial borrowings translated into rapidly increasing total indebtedness of the economy while the repayment of both internal and external liabilities created excess pressure on government expenditure.

The fiscal and the current account deficit reached as high as 7% and 5% of GDP respectively and the associated buildup of public debt and external debt was recorded at over 100% of GDP and 33.5% of Foreign Exchange Earnings respectively. This emerged as a major source of macroeconomic imbalances during this time (MCB, 2007).

Period 6 (1999-2007)

Pakistan economy of period 7 headed by General Pervez Musharraf with Shaukat Aziz as major economic advisor, has been widely hailed by a large number of expert organizations and institutions (Gulf Research Center, 2008, MCB, 2007, Daily Times of Pakistan, 2007, Pakistan Times, 2007, IMF, 2007). Its major features are as under:

1. It sustained growth rate remained 7% from 2004 to 2007(Pakistan Times, 2007)
2. There was great improvement in the foreign exchange situation and there was recorded a rapid growth in hard currency reserves. The 2005 estimate of foreign debt was around US$40 billion, which decreased with the assistance of IMF and debt-relief from the United States.
3. On the contrary, it has been widely admitted that inflationary pressures and low saving rate along with other economic factors, would have rendered it very difficult to sustain a high growth rate. The structure of the Pakistan economy has been said to change from agriculture base to a strong service base. The agriculture accounts for roughly 20% of GDP. The service sector, on the other hand, accounts for 53% of GDP (Wikipedia). Significant foreign investments can be seen in several areas such as telecommunications, real estate, energy and many others (Governor SBP, 2007).
4. Pakistan also signed a Free Trade Agreement with China with the hope to triple bilateral trade from $4.2 billion to $15 billion within in the next five years and Pakistan's exports in 2007 amounted to $20.58 billion (World Fact Book, 2008)
5. There are many for and against opinions on the performance of Pakistan economy in this period. Some experts are of the view that the current success of the economy was due to economic growth largely due to the comprehensive structural reforms, macroeconomic policies and financial discipline of the Musharraf and Shaukat Aziz rule while others indebt it to 9/11 events after which foreign direct investment and remittances encountered high growth globally.

6. Before political unrest and judicial crises in Pakistan, Pakistan economy started being considered among the fastest growing economies of the region as it was growing at an average rate of 7% for the years 2004- 2007. Compared with other emerging economies in Asia, this put Pakistan as one of the fastest growing economies in the region, along with China and India (MCB, 2007). The economic growth prospects were considered better than ever before as foreign direct investment, remittances and foreign exchange reserves exhibited new records. Tax revenue collection increased substantially. The agricultural performance had been strong while that of the service sector had been remarkable.

Period 7 (2007-2018)

The new government of Pakistan Muslim League Nawaz in Period 7 believed that it was vital to restore macroeconomic stability for creating employment opportunities and preventing people from falling below the poverty line while extensive structural reforms in almost all the key sectors of the economy were needed to enhance economic incentives, improve resource allocation, and remove hurdles in the way of private sector development. It is with this view that a series of structural reform measures were initiated in areas of privatization and deregulation, trade liberalization, banking sector, capital markets, tax system and tax administration and the agriculture sector that ultimately translated into results as under:

The most remarkable development was seen on the external front where for the first time in the history of Pakistan, economy was able to exhibit a current account surplus of USD 331 million and thereby substantial improvement could be seen in Pakistan's balance of payments. When the foreign investors started showing interest in rapidly growing economies of Asia, Pakistan also attracted the attention of foreign investors as economic growth continued to gain pace. The foreign direct investment and inflow of remittances accelerated to accelerate the pace of economy, even further. The important economic indicators of this period are as under:

- Population: 207.68 million (5th (2017 National Census)
- GDP: $299 billion (nominal; Jun 2021) $1.36 trillion (PPP; Jun 2021)
- GDP Rank: 46th (nominal; 2020) 22nd (PPP; Jun 2021)
- GDP Growth: 5.5% (17/18) 1.9% (18/19) −0.4% (19/20) 3.9% (20/21e)

The sustainability of this period due to the strong support of China that nullified the hostility of India and up and down treatment of Pakistan by USA. In 2017, China invested billions in energy sector and betterment of infrastructure in Pakistan and the World Bank also praised Islamabad for accelerating GDP growth to 5.5% in fiscal year 2017 that which was the highest in the decade (Pakistan Economy - The Heritage Foundation Online). After the country's Supreme Court convicted Prime Minister Nawaz Sharif in July 6, 2018, Pakistan's military openly criticized his governance that headed Pakistan Muslim League Nawaz (PML-N) party, for its mishandling of the economy and with this historical event, governance of Pakistan went to Tehreek-e-Insaf headed by Prime Minister Imran Khan.

The economy under the administration of Prime Minister Imran Khan faced the slowest growth rate in South Asia because about 220 million Pakistanis were struggling with food inflation and rising unemployment. The country recorded negative growth for the first time since 1952 in 2019. Beijing's enthusiasm for betterment of Pakistan's economic fortunes did cool down, also. Different authorities and military high ups gave talks that oscillated between despair and optimism. The IMF conditions that required Pakistan increasing revenue via raising of taxes and adopting severity by cutting subsidies that contribute to lower down inflation. For example, in 2019, Pakistan started following a $6 billion IMF loan program requiring increased tax collection via reforms while taking measures in bridging the budget deficit. In spite of his hardline foreign policy due to which he lost the sympathies of some big shots, and earned the appreciation by common man, Imran Khan, could not survive due big inflationary pressure and other economic pressures and was ousted in the parliament with a vote of no confidence and the rule went back to PMLN with Shahbaz Sharif as Prime Minister and economy further destabilized due to unaffordable political leadership with vested interest. The prices are almost doubled even of common goods and rupee is facing worst devaluation (1US $= 275) and skyrocketing prices of petrol and all routine needs.

After taking into consideration the facts and figures computed above, it can be concluded that in spite of economic recessions encountered in Period 3 and 5, the Pakistan economy performed well but in terms of macroeconomic indicators. The fruits of this performance were reaped by selected few. The benefits never adequately reached the common man in any period.

2. CURRENT SITUATION

The current situation of Pakistan Economy is well known to all who are equipped with average literacy level, even to known a common man claiming to be a Pakistani. Still it seems logical to present opinions of some experts in the field.

Noor (2019) reviewed the situation and attempted to prove it an eventful year for Pakistan starting from February Pulwama/Balakot crisis with India to diplomatic battles to avoid blacklisting by the Financial Action Task Force (FATF) in October. The year encountered internal political and economic challenges, troubled regional security, and unstable global dynamics, which collectively had a direct impact on the country. The Government of Tehreek-e-Insaf encountered many domestic challenges such as facing a stunted economic growth due to high inflation and low employment rates. The author advocated that "In spite of massive borrowing from friendly countries and the bailout package provided by the IMF, Pakistan's economic outlook remains grim." She concluded that "2019 was a tough year for the relatively inexperienced PTI government. Notably, Imran Khan's comparatively strong performance on the foreign policy front stands in sharp contrast to his handling of domestic affairs, where he has received criticism for failing to deliver on promised reforms. While patterns from 2019, such as hostile relations with India and oscillating relations with the United States, are likely to remain a factor in Pakistan's foreign policy for the coming year, the PTI's real opportunity in 2020 will be improving its performance on the domestic front."

The Heritage Foundation (2021) answered the question," How bad is the Pakistan Economy?" saying that "Pakistan's economic freedom score is 51.7, making its economy the 152nd freest in the 2021 Index. Pakistan is ranked 34th among 40 countries in the Asia–Pacific region, and its overall score is below the regional and world averages". It continues that "Pakistan is one of the poorest and least developed countries in Asia. Pakistan has a growing semi-industrialized economy that relies on manufacturing, agriculture and remittances. Although since 2005 the GDP has been growing at an average of 5 percent a year, it is not enough to keep up with fast population growth."

Nadir while communicating to Gandhara (Siddique, 2021) argued that "the government's accountability drive, which largely targeted opposition politicians for alleged corruption, only served to scare the business community and send the country's vast civil bureaucracy into shutdown mode. The administration's lack of a clear economic vision beyond joining the IMF program in 2019 only made matters worse. He also commented on World Bank projection and IMF estimates said, "The structural adjustment policy followed under the IMF was contractionary and low growth and put the economy on a path to very slow growth, high interest rates, and increasing unemployment," Nadir told Gandhara. The World Bank has projected 1.3 percent GDP growth for 2020-2021 while the IMF's estimate is slightly higher at 1.5 percent. He also highlighted a better area saying, "The only area where things are better is the external account, where the reduction in imports and the increase in remittances [have] improved the situation somewhat," Nadir said. "But the fiscal position is very poor with an unmanageable increase in debts." Government and international statistics indicate poverty among Pakistan's rapidly expanding estimated 220 million populations is on the rise. International lenders expect poverty to rise to more than 40 percent from 24 percent in 2015, and the country could be home to as many as 80 million poor compared to some 55 million in 2015.

"With shrinking GDP, employment has also shrunk considerably," Ismail noted. "In a country where 1.8 million people join labor market every year, you need growth of at least 5.5 percent to 6 percent to absorb all the growth."

Talking in comparative context, let us take the example of Bangladesh. It celebrated 50 years of separation from Pakistan in March 2021. Based on the key economic indicators, it can be said that it is well ahead of Pakistan. Its $1,855 GDP per capita in 2019 was a third more than Pakistan's $1,284 because of following vastly different ways of growth and human development.

Younus (2021) also viewed potential in Pakistan's industrial and agricultural base, as well as in its young population in a way that "There are good green shoots in the ground, particularly in the technology and startup sector, which is thriving, and the government has so far played a role in ensuring that," he said.

He warns that the political instability is overshadowing the quest for economic prosperity and thus. "So long as the political game of thrones continues, governments will find it difficult to pursue these reforms," he said. "So long as those in power [do] not lose out because of successive IMF programs [and] are not impacted by sky-high food inflation, there is not a lot of incentive to pursue reforms."

The COVID-19 pandemic had recently the most serious impact on public health and economic crises faced by Asian countries including India, Bangladesh, and Pakistan (Yunous, 2021). This economic impact on these nations has some future implications with reference to their relations with the United States. The author identifies key areas of focus for ensuring the subcontinent's recovery is equitable—which, in the context of an erosion of democratic norms, growing authoritarianism, and severe crackdown on dissent, could help avoid economic and social instability.

3. CAUSES AND EFFECTS

The cause and effect inter-linkages with reference to the factors responsible for the current economic crises in Pakistan are highlighted below.

International Political Scenario

If interpreted in political terms, the current economic crises may be having many reasons behind it, but the major one that every one of us knows is the 9/11. The post-script of this story is that it has pushed out the war of terror from the home of many global players to our doorsteps. The transaction has been very smoothly done taking the advantage of our innocence and our strategic situation in 2001. The investment, the big lords had to make can be coined in terms of some grants, promises made to us for our future development and help in boosting of our foreign trade through quotas in international trade, etc. Many out of us who had their interests beyond the borders helped to conclude these transactions. These apparent favors have now taken the shape of a coercive force as if we don't stand by our benefactors in the war of terror, these may be withdrawn and this action may pull us to the position where we will be more backward even than our economic position in 2001 even. If this happens, our projects will be blocked where they are.

The roads will remain unfinished with dust and debris in their environment and the luxurious salaries being enjoyed by the people even from the educational institutions may be withdrawn or may suffer severe cuts. The question arises that what will happen to the common man who is already under the intense pressure of inflation and sky-rocketing prices of stuffs urgently needed to keep body and soul together. The promises made by great economists imported from abroad and joiners from home to make us join the front line of Asian Tigers have already left him helpless in depression and sorrows as the benefits of what was planned have not reached the lower strata of society

Lack of Confidence in Collective Self

The major wrong that we are doing and are still continuing to do is our sole reliance on imported ideas that translate here into imported economic models that ultimately lead to lack of national self-reliance and development of indigenous innovation systems. Unfortunately, the ideas that crop up in the developed countries are in tune with their own conditions and environment don't suit when applied in the conditions and environment of the developing countries. The end result is that there is a lack of self-sustained innovation systems that are urgently required to solve indigenous problems. Moreover, local talent and expertise is in a better position to solve these problems as they can view the national scenarios better than those imported from abroad.

The results of our sole reliance on foreign advice and consultancy are disastrous. The experts and think tanks in the developed countries mostly depend on macroeconomic indicators for the control of their economies and they are successful in achieving the desired results. Thus, the same model was imported from abroad and applied for the control of economy of Pakistan.

There is a sanction of worldwide recognized experts behind the fact that the macroeconomic model has failed to deliver the goods in the developing countries including Pakistan. One of these experts relevant to quote here may be Mehboob-ul-Haq. He, in his article entitled, "Employment and Income distribution in the 1970s: A New Perspective" published in 'Pakistan

Economic and Social Review' in 1971 gave his impartial opinion given below.

> "The problem of development must be defined as selective attack on the worst forms of poverty. Development goals must be defined in terms of progressive reduction and eventual elimination of malnutrition, disease illiteracy, squalor, unemployment and inequalities. We were taught to take care of the GNP because it will take care of poverty. Let us reverse this disorder and take of poverty because it will take care of GNP. In other words, let us worry about contents of GNP even more than its rate of increase."

The statement was made in 1971 and even after forty years after that the macroeconomic policies have not been able to correct the condition of lower classes in poor countries. That is why the major issue before the developing states currently is how to overcome the problem of unemployment, inequality and poverty in these countries. The irony is that these victims of macroeconomic model yet look towards the West for the solutions of their socio-economic problems without realizing that if the same trend continues, this problem will magnify year by year along with the creation of many others.

Budget Allocation and Utilization

The next aspect for serious consideration may be the over allocation of non-development budgets to different public departments, underutilization of annual budgets and subsequent surrender to the circles, which are already getting more than their due. The author recalls an episode of his career when he was the Director Colleges, Gujranwala Division. During his short tenure, he encountered many colleges of his division where in spite of budget allocation against proportionate number of teaching posts, they were deplorably deficient in teaching staff. In one of the girl colleges, there were only 3 staff members against 14 sanctioned posts. In another, there were 7 against 35. There were many subjects in which not a single teacher was available to meet the classes and thus the teachers of other disciplines were asked to teach the students of the disciplines where there was the deficiency. When the people from the Education Department,

Government of the Punjab were asked to tell, "where goes the unutilized budget?", the answer was that it is surrendered to the Government of the Punjab for utilization otherwise.

How the surplus would have been spent otherwise, everyone knows. This is just one example from an education department; there could be a large number that one can come across during the survey of all government departments. It may be fruitful to cast a glance on different establishments for having a rough idea about the wastage situation over there.

Better not narrate the names; there are many institutes which were opened due to the reason that a single person who came from abroad after getting a degree or training in a small field and thus a special establishment was erected to accommodate him. In our universities, there is a large number of departments which are teaching if not the same, very similar courses but they are functioning independently. Examples may be public administration, business administration, business education, etc. These can be integrated under one roof in a department which may be called an institute/department of management sciences. A recent wise step taken by Government of Pakistan is the integration of three institutions i.e. Administrative Staff College, NIPA and Civil Service Academy into one institution. This may translate into a drastic reduction in the expenditure. Similarly, there are many other subjects that can be combined to function under the same roof. The instances may be social work and sociology, philosophy and psychology, Urdu, Persian and Arabic (Oriental Languages) and so on.

Similar is the case of many government departments. For example, the major activity of the Department of Manpower Planning and Social Welfare, mostly, encountered is the preaching of the family planning to keep the family size limited to a few members. The same can be accomplished by a small wing opened in the Health Department.

Food and Energy

Currently, the food and energy are the biggest references while taking decisions at the macroeconomic level. We have been and we are yet in the list of the agriculturist countries expected to be self-sufficient in our food supplies. We are facing an acute shortage of wheat flour is really an irony. It is our misfortune that we are suffering from water shortages, occasional famines and desertification.

After paying due regards to these factors, there are yet many things to be thought about in context of food. One of these additional factors is the rapid urbanization that is taking from us the agriculture land. Have we ever put this question to our collective self, “If mushroom growth of rapidly expanding urban colonies continues at the current pace, from where will come our food in future?” The answer is not difficult but none talks in these terms due to the benefit of the people involved in the real estate business. Once the author put this question to the chief executive of a high class lending bank and the answer he received was that that it was none of his business as he was concerned with loans and not with its impact on our food problem.

Where we are facing an acute shortage of energy, we are also wasting it unconsciously. The first question that comes in mind in the shortage context is, “Are we managing an efficient and effective distribution of the amount of electricity that we are producing in the Country?”. The answer will be “certainly not”. The next question in this context will be that have we ever rightly identified the causes of line losses and theft and if done so, have we ever developed some measures and implemented them to eradicate this menace from our society. We have huge literature on the renewable resources of energy but we are not in a mood to consider that burning of solid waste openly in homes, on roads and in the fields for its disposal not only creates air pollutants but is also synonymous with the burning of the national wealth? Have the managers of hospital waste disposal ever imagined that the incineration of hospital waste translates into the creation of secondary air pollutants but also is wastage of huge quantity of energy, which could be otherwise harnessed to produce thermoelectric power (Khan and Sana, 2020). Similarly, dumping of large quantities of solid waste underground without reaping the benefits associated with techno-economic disposal is a wasteful exercise. This situation exists when it is now globally resolved that the solid waste resource can be better exploited through the proper management of constructing landfills in which the naturally distributed services of anaerobic bacteria can be engaged to split its high molecular weight complex and compact structured components to simple low components that can be easily taken up by plants as nutrients to promote their growth, along with the production of large quantities of biogas that is mostly methane. Methane is also the major component of natural gas which finds extensive use both in home and industry. The gas we burn in our stoves and gas heaters as fuel for cooking, warming and sometime lighting is the natural gas subsequently refined for making it suitable for home use. Similarly, it is used as fuel in industries and a raw material for commercial products. An important example is the manufacture of urea. Natural gas is extensively substituting gasoline as a fuel for vehicles as it produces relatively less air pollutants. In many countries, it is also being used as a fuel for the production of thermoelectricity in power plants. In some, even, the generators have been installed near the landfills to avoid expensive network of pipelines for supplying gas elsewhere.

The Missing Link

It will not be fair to advise that we should let aside macroeconomic model and switch over to the microeconomic alternative. The point to be made here is that there is a missing link between these two models that must be highlighted and thrashed out in detail to assess what is the best from both sides that may be combined and subsequently applied not only to solve the problem of unemployment, inequality and poverty but also bring a package that guarantees prosperity for the people of Pakistan as a whole. This is essential with reference to the gloomy socioeconomic future ahead and the problem multiplicity that is due as a gift of ensuing globalization, for whose combating we are not prepared from any angle.

4. SOLUTIONS AT DIFFERENT LEVELS

What can be the solution to the mega problems narrated above? It is the question which is extremely difficult to answer. Yet there is no harm in making an attempt to dig out the answer and present it to the suffering humanity. The same has been done in the following sections basing everything on extreme sincerity on behalf of author and expectation of similar sincerity from others. The situation cannot be changed until the vested interests are put aside by the implementing forces. An important example of this thought from an Indian expert, currently working with IMF are given below as an illustration of the mindset of the author here.

Rajan (2009) attempted to identify some causes of persistence of underdevelopment and stated that of these were the constituencies and rent preservation. He argued that each “constituency prefers reforms that preserve only its rents and expand its opportunities”. That is the basis of the fact that no comprehensive reform path can command broad support. The conclusion leads to the overall reality that the vested interests don’t give way to any reform to be through. As our socio-economic environment is not different from India, we can apply the theme to illustrate the hurdles in the way of the reforms such as first generation and second generation land reforms and nationalization of industries.

It may be safely conceived that there are two levels where the solutions can be offered: Macroeconomic Level and Microeconomic Level. It sounds appropriate to discuss them separately. The same exercise has been carried out below.

Macroeconomic Level

At the macroeconomic level, following the current mechanics of problem solving, are of the opinion that the focus has to be on enhancing the agriculture sector and formalizing the country’s mining sector assuming that these are the two real wealth generating sectors in Pakistan with subsequent encouragement of industries, starting from the low skilled ones, to develop (Malik, 2021).

Taking into consideration the key challenges and solutions at this level, the solution that can be transparently viewed as a solution in the prevailing kaleidoscopic socioeconomic scenario on the screen of twenty-first century is ‘Fair Distribution of Income’. This solution has been floated many times in the history of Pakistan in context of national economy. The irony is that this has been publicly stated and heard by all concerned but silently turned down by the decision makers. The reason behind this turning down is that a set of society which owns most of the wealth is not in a mood to accept it as a problem solver due to their vested interests. This does not mean that no attempt has been made on the goal of redistribution of income. The first attempt was through different generations of land reforms. Although the reforms were done but these reforms translated into no requisite results as these were done with no intention to see their practical manifestation. The big landlords did not forego even an acre of land in favor of tenancy and the have-nots continued to remain have-nots.

On the industrial side, nationalization was done under the cover of the slogan of “Roti, Kapra and Makan”. As it was done in a hatch patch manner with untimely and irresponsible encouragement of the workforce, the results were highly undesirable as they ultimately translated into drastic reduction of production efficiency and net outflow of capital abroad by the entrepreneur. A significant percentage of industrial units fell sick and many were also closed.

The overall victim was the national economy, which had to bear the strongest economic blow hardly encountered in the history of the nation. Since then many crests and troughs are seen on the graph of national economic efficiency measured in terms of macroeconomic indicators.

These ups and downs have not yet stopped out of which the current economic crises, has appeared on the scene. It is so severe that every Pakistani feels concerned about it.

The situation analysis gives us a warning that we have to be very careful while reorganizing our economic model based on redistribution of income. The major question to be put to those who claim to be the saviors of the nation will be whether they are sincerely in a mood to take up this exercise with strong determination and good intentions. If the answer is real yes, Allah will positively show us the ways to accomplish it smoothly.

The next aspect for serious consideration may be imposition of cuts in non-development budgets of different public departments through the realistic allocations in the national annual budget, which is a major tool in the hand of the state to control national economics. The data about all wasteful exercises done in the past should be gathered and analyzed to know the cause and effect, which will form the basis of thrashing out where the scissors can be applied to cut short the expenditures. This does not imply that cuts are imposed even if the budget allocation is adequate. It will also be better to identify who have been the victims of budget underutilization of annual budgets and subsequent surrender to the circles, which are already getting more than their due.

There is a solution to the problem of rapid urbanization and mushroom growth of colonies on agriculture land. The state should decide, in the first instance, that none will be permitted to own plot for building his residence more than the permitted dimensions. The next solution will be that we should strongly decide on option of vertical movement instead of horizontal movement.

This will economize in land both in general and agricultural contexts. This will certainly lead to economy in use of agriculture land and ultimately to a secure future for our following generations.

Another serious aspect to be highlighted is the luxurious architecture that advises to make every room air conditioned without keeping in view from where the required amount of electricity will be supplied to run them. State level decisions are requisite to mentally prepare the people for future shortages of electric power and revert to the traditional methods of keeping the rooms cold or semi-classical techniques that offer substitution of air conditioners by room air coolers, etc.

Microeconomic Level

A number of solutions can emerge from the microeconomic model. The examples may be techno-economic disposal of wastes, micro-financing and many others. Let us recall that Gandhi asked the Indian leadership to put spinner wheal as an insignia on the Indian flag. Its significance was chaining of technology from the lowest to the highest level in such a way that it is in tune with the national conditions and environment. Moreover the labor intensity during the transitions from lower to upper level would not affect the labor intensity of the nation as a whole. The proposal did not materialize but its impact can be viewed in the pattern of Indian technology that exhibits diversity from lowest to highest scale of production. All the players are competing very well with each other in the market. This does not mean that India has done very well in the sense that it has been able to eradicate poverty and its common man is very prosperous. The economy as a whole is quite stable.

Techno-economic Disposal of Solid Waste (A Case Study):

Techno-economic disposal of solid waste can effectively address the problem of electrical shortage. This is because the techno-economic disposal of waste is based on the concept of 'Law of Conservation of Resources' which states that the total amount of resources in the universe remains constant although may change from one form to another. The transformation from one form to another is just like of mass and energy. This law provides a firm basis for sustainable development, the concept floated as a cure of all economic evils to stabilize the shaky economies. According to the law, a waste is no more a waste as every waste can be assigned an economic value. Sometimes, a waste, after proper evaluation may even turn out to be more valuable than the stuff from which it is produced.

The production of thermoelectric power by incineration of solid waste is more than a half century old story. The author was first time introduced to this important technology in late 1960s or early 1970s by a young man who was doing a job in a New York Council power plant that was based on the incineration of municipal waste to produce thermoelectric power. He suddenly stepped in author's room in Government College Lahore where he was a Lecturer and told him how the solid waste of New York was being techno-economically disposed to overcome the problem of solid pollution along with the production of electricity for the City of New York. He disappeared and reappeared next day with a shield in his hand. He told the author that he delivered a talk on production of thermoelectric power by incineration of biomass of municipal solid waste to the councilors of Lahore Municipal Corporation and they have honored him with the shield as a token to the guest speaker. He was back to the job and was almost forgotten for a long period. After a long time, it was disclosed to the author (participant) in a PTV program on environment that LMC got a study done on production of thermoelectric power by incineration of municipal solid waste of Lahore City but its recommendations were not implemented.

The next two episodes that can be quoted in present context are visits of the post-graduate students of Government College Lahore to Rahwali Sugar Mill and Pattoki Sugar Mill. Former utilized its sugarcane solid waste popularly known as 'Bagasse' by incinerating it only as a source of heat for the multiple effect evaporators for concentrating the clarified sugarcane juice.

The excess energy dissipated in the atmosphere and was not harnessed to produce electricity.

Latter was reaping both benefits; it was concentrating the juice as well as harnessing energy to produce electricity by the use of steam boilers and turbines to meet their domestic demand.

These days, almost all mills of Pakistan are using this technology to dispose of bagasse and produce electricity side by side.

The fourth episode started its constitutional make up when author became actively involved in studies in the field of Environmental Economics at Lahore School of Economics and carried out some projects on techno-economic disposal of liquid and gaseous pollutants. Suddenly, he felt that he should do something on the similar disposal of solid pollutants. At this juncture he recalled all his memories about the techno-economic disposal of solid waste by incineration of its biomass to produce electricity. All the events were integrated which finally translated into a pilot plan to carry out a series of studies in a systematic manner. The central theme was techno economic disposal of solid wastes produced from an institution to city level.

In the first phase, the work on economic viability of the production and disposal of solid waste was undertaken on four educational institutions: LSE, LUMS, KC and GCUL. The primary data concerning the amount and composition of solid waste was collected through successive visits to the institutions and subsequent interviews of the concerned personnel. The primary data concerning the production of thermoelectric power from the solid waste such as bagasse was collected by visiting Pattoki Sugar Mill and the prices of machinery and equipment were supplied by the Mill authorities which were later verified by visiting the websites of different local machinery manufacturing companies. The data were used to design the project that was subsequently appraised to determine its B/C Ratio, NPV by applying discounted cash flow techniques and their payback period (PBP) was determined by conventional method.

The results indicated that the installation of solid waste based thermoelectric power plant was not feasible in LSE, LUMS and KC but it was feasible for GCUL. The data, of course, guided that the projects could be pulled towards feasibility through well planned strategies.

In the second phase, the extension of work narrated above was undertaken by one of the research groups being supervised by the author. The studies on village level, urban colony level and on city of Lahore as a whole have been completed. The group is likely to give promising messages to the residents of Lahore and to the nation at large.

The major constraint on the installation of solid waste thermoelectric power plants seems to be low heating value of municipal solid waste due to its being mixed with a high proportion of unwanted material such as construction material that cannot be burnt to produce heat energy and presence of toxic material which produces toxicity in the form of toxic pollutants. An action plan may be framed to handle these problems. For example, the municipal authorities can be advised to introduce the concept of separate collection of waste components at source. The source may be a residential house, a site in an institution, a spot in a manufacturing factory and so on. The concerned authorities are expected to provide plastic bags of different colors for each spot for collecting different waste components separately. The public may be counseled for awareness of the value of the process with the help of media and personal contact which will guarantee a significant saving in terms of recyclables and thermoelectric power production, etc. Moreover, it will provide a clean way of separating recyclables from non-recyclables, infectious from non-infectious, toxic from non-toxic and incinerable from non-incinerable wastes. Thus, the people carrying out this exercise will have multifold benefits through resale of recyclables at the registered and authorized points and production of thermoelectric power from the burnable biomass. Chemists and mechanical engineers can help in installing scrubbers to trap air polluting gases coming out of incinerators to avoid their exhaustion into the atmosphere. The biomass that will be free from construction material, dust, etc. will definitely have higher heating value than that of biomass which will be mixed with other non-burnable materials.

The solid wastes can be disposed of through the alternative technique of landfills. The major requisite is that they should be properly designed. These are to be equipped with porous pipes for internal collection of biogas. The gas thus produced may be integrated and supplied as fuel for the production of electricity, cooking food at homes and other purposes. The residue left after the natural breakdown of biomass by anaerobic bacteria is rich in nutrients and that may be supplied to the farmers as natural manure.

In continuation of the above context, the liquid pollutants can be disposed of techno-economically producing methane, natural manure and clean water that can be supplied to agricultural farms for irrigation purpose. Waste water treatment can be studied and installed at different scales of processing such as lowest for a locality or village and the highest for big town like Karachi and Lahore. The designs can be better worked out by the combined effort of chemists, engineers and economists.

To cut short, there can be hundreds of the projects like the ones narrated above. These can be identified and enlisted to study them to add significant value to the national production which is the basic thing to target alleviation of poverty and minimization of inequality from Pakistan and also from the developing countries at large.

5. ROLE OF STATISTICS IN RECOVERY OF ECONOMY

To define the role of statistics, let us start from the first principles and start with the definition of Statistics and then switch over to its applications in different fields with special focus on what economic recovery.

The statistics as defined by Jim Frost (Online) is “the science of learning from data. Statistical knowledge helps you use the proper methods to collect data, employ the correct analyses, and effectively present the results. Statistics is a crucial process behind how we make discoveries in science, make decisions based on data, and make predictions”. The field of statistics is concerned with collecting, analyzing, interpreting, and presenting data. The most important concepts covered in statistics include mean, median, mode, range, and standard deviation. There are two types of statistics as narrated below:

Descriptive Statistics – Through graphs or tables, or numerical calculations, descriptive statistics uses the data to provide descriptions of the population. Economists often calculate the following descriptive statistics for a given region (Byjus Online):

- The mean household income.
- The standard deviation of household incomes.
- The sum of gross domestic product.
- The percentage change in total new jobs.

Inferential Statistics – Based on the data sample taken from the population, inferential statistics makes the predictions and inferences.

Both types of statistics are equally employed in the field of statistical analysis.

Its multiple prong importance in different fields with focus on economics is highlighted below by Jim (Online), Britannica (Online), and Satology (Online) as under:

1. It helps in designing the effective and proper planning of the statistical inquiry in any field. It gives valid inferences with the reliability measures about the population parameters from the sample data. Statistics help Economists to understand problems on production, distribution and consumption. Statistics of consumption describes the way in which people from different income group spend their income. It helps in understanding standard of living and taxable capacity of people.

2. Statistics allows economists to understand the state of the economy using descriptive statistics. It allows economists to spot trends in the economy using data visualizations. It allows economists to quantify the relationship between variables using regression models.

3. It allows economists to forecast trends in the economy. Using these metrics, economists can gain a better understanding of the state of the economy in a particular region. They can then use these metrics to inform politicians or law makers on the best methods to use to ensure that the economy remains healthy and grows. These types of descriptive statistics are used at every level of economics. For example, economists at the national level, state level, city level, county level, etc. all use descriptive statistics to gain a better understanding of the state of the economy in their area.

4. Another common way that statistics is used in economics is through data visualizations such as line charts, histograms, boxplots, pie charts and other charts. These types of charts are often used to help economist’s spot trends that can help them see if the economy is improving or declining.

5. It helps to **quantify relationship between variables using Regression Models which** allow economists to quantify the relationship between one or more predictor variables and a response model.

6. It is applied in economics is in the form of forecasting trends.

The Merospark.com (Online) has reported the following important applications of statistics in the field of economics:

1. Consumption statistics reveals the pattern of the consumption of the various commodities by different sections of society and also enable us to have some ideas about their purchasing capacity and their standard of living.
2. **To solve production related economic problems** and is very useful in solving basic economics problems like what to produce, how to produce, for whom to produce, etc. which arise due to the scarcity of resources. It also helps in adjusting the market supply to market demand.
3. **The e**xchange statistics reflects upon the process of pricing regarded to maximize profit of business firms or to promote consumer's welfare.
4. **It helps in the study of the nature and process of income distribution by advising** how the national income is calculated in the distribution of income and wealth, statistics.
5. **The statistical method** is very useful to formulate various economic policies like trade policy, taxation policy, industrial policy, etc., and evaluate the effectiveness of the framed policies. For example, by evaluating the absolute figure of poverty government formulates policies related to poverty alleviation and also checks the effectiveness of these policies to alleviate poverty on the basis of related figures.
6. **The statistical methods** are methods of economic planning. Statistics helps to evaluate the achievements and problems of the previous plans. Similarly, it helps to fix targets and revenue and expenditure statements of the current plan.
7. **The statistics** is also useful to build economic models and functions.

Importance and Application of Statistics on Economics (A Consolidated Statement)

Emathzone.com/tutorials (Online) has given a utilizable consolidated statement with reference to the applications of statistic in economics:

> "Statistics play an important role in economics. Economics largely depends upon statistics. National income accounts are multipurpose indicators for the economists and administrators. Statistical methods are used for preparation of these accounts. In economics research statistical methods are used for collecting and analysis the data and testing hypothesis. The relationship between supply and demands is studies by statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics."

After taking into consideration the basics and the applications of statistics in different fields and economics, it can be concluded that this discipline can advise its benefactors but lacks its active involvement in planning and development, policy framework, production and supply, consumption and assessment of standards of living, fair distribution of wealth, choice of appropriate technology, sustainable development, forecasting the future supply for demand and securing and many others. Unfortunately, it is not equipped with the planning and implementation power to recover the sinking economy of Pakistan. Still it can be said with confidence that it can play a significant active role in achieving this objective. How it can be done? Here are some suggestions that may be helpful for statisticians to play the desired role in recovery of Pakistan Economy.

Correctness of Statistical Data

There is no doubt about it that incorrect data always translates into incorrect decisions at all levels and the major outcome of incorrect decisions is incorrect planning and incorrect policies that ultimately lead to the destabilization of the economies. Thus, it seems imperative to identify the causes of incorrectness in different set ups. Because the author has been in long contact with collection, computation and analysis of the data, he encountered a number of problems in this exercise and feels the following major causes of the incorrectness of data:

a. The major cause seems to be that the respondents, if executives or managers, don't disclose information due to their secrecy policy for hiding their income particularly to avoid taxes and other liabilities, if employees, due to lack of interest or of ability to answer questions, fear of action from the executives that may even lead to the firing of the employees and finally the public from whom a mixed response is often encountered; some are not interested to answer the question or lack of the consciousness about the sensitivity of the issue being investigated.

b. A big source of incorrectness in data is the government directed reporting of the statistical data by the statistical organizations. The objective is to show that the economic situation is well under control of the government in power. How far dangerous is such an incorrectness can be well imagined average by a common man with an average intelligence? One should not forget that a model erected on a weak foundation always remains exposed to the environmental shocks.

c. Due to difficulties involved the researchers find easy ways to get their questionnaires filled and contact only the easily available respondents. An important example quotable in this context is a project, title, "The university teachers/employees are satisfied with their salaries and emoluments or not?" The author received questionnaire of this project continuously from different universities for responses in his 18 year service in this field and filled the forms. The central theme of such a project was only the award of a higher degree without any applied significance. Another example for filling in the blank can be given from a welfare center opened under the Ministry of Manpower Planning in his Colony. The Principal of the institution called a meeting of residents all highly educated for preaching of family planning. After his address, the author asked him, "What was the sense calling the meeting of the educated people who are already aware: why don't you call instead the meeting of the illiterate folk having rich number of children from villages for raising the consciousness? " He was stunned at this question because he could not have in his mind any answer.

d. According to a recent report of Transparency International (2021), Pakistan ranks 124 out of 180 countries on the Corruption Perceptions Index 2020. Realistically speaking, education and research cannot be kept non-aligned from this malpractice.

There is always a chance that those research students working for higher degrees may produce concocted data for submission of their theses and the research supervisor may overlook and excuse them due to his own efficiency being at stake. Thus due to false data, the degrees awarded will also be fake and of no use in applied context.

Suggestions to the Statisticians in Recovery of Pakistan:

Following suggestions be made for the recovery:

1. The researchers mostly express that they are refused the data by the executives of different establishments and ask for engaging survey firms for collection and supply of required data. The author of this keynote claims that he was never refused the technical data during his 18 year research career. As for corporate sensitive data related to taxes is concerned, that in the first instance was not need, it was also made available for cost analysis. Collection of data, of course, depends upon how to approach the respondents? Out of different approaches the personal contact turned out to be the best. Here the most appropriate example may be that of Pattoki Sugar Mill for the supply of data required in the project, "Production of thermoelectric from solid waste." The boiler house staff prepared some of the alternative projects depending upon our requirement and their mill experience that were processed by the members of the research group working at different levels. The only condition is that the research supervisors should accompany the research workers which is usually not the case. Once the author visited Small and Medium Enterprises Development Authority (SMEDA) for collection of the primary data. He was well received and the officials said "You are the first research supervisor who has visited our place in context of data collection." The story proves that officials are willing to supply data to the responsible authorities from the research institutions.

2. To elaborate on strategic approach to gather the data from different establishments, a recent example is the data collection for author's pilot project "Environmental Situation of Industrial Sector of Pakistan". He is currently teaching BBA undergraduates in Lahore School of Economics .where the students have to submit a term thesis on a project. There were more than 100 students in three sections he is teaching. He made around 30 groups having 3 or 4 students in each group. The pilot project was framed having in mind two bases: Students work very hard under the pressure of the degree and can be made to collect data due to their personal contacts. They were questioned for their contacts and as a result it was observed that some of them were running their own industrial units. Many of them said that they will be able to get the data through some employees working there. The group was successful in gathering the data even if the supervisor could not accompany them due to his other involvements in LSE. Thus the author was able to bag primary data for 30 projects out of which 30 papers can be written. The group was lucky enough to present 6 research papers in 20th ISOSS Conference with author of keynote as Corresponding Author.

3. The next major point to be taken into account is that the statistics presents only the statistical methodology with no power of decision making and thus it remains in advisory capacity. Let us not forget offer of these methods in the form of models have not able to solve the problems because the models identify the cause and effect where the experts responsible for control have no measure to offer the solution. It may be elaborated by taking the case of poverty eradication. It is a fact that billions of dollars are being spent on calculation of poverty at different levels with statistical models but the graph of the poverty curve is rising day by day. The solution of the problem is the creation of an interdisciplinary culture in the country. The author is aggrieved to say that experts attending the ISOSS conferences are all statisticians barring a few sometime himself feeling alone among the statisticians. The irony is that even the experts of economics who are considered to be allies of statistics are not seen in ISOSS conferences. The experts of this discipline after interdisciplinary grooming will be able to claim an equal share as decision makers.

4. There is an urgent need to change the direction of research from academic research to applied research. Here author's question for all experts claiming the champions of control of economy: How many of you are targeting the problem of the recovery of sinking Pakistan economy? Either there will be no answer or if it is there, it will be extremely depressing because the target will be irrelevant. Here research leading to the diversified production is required not the one that academic leading to the award of the higher degrees. Here also the research articles with application to solve national problems and the problems of ailing humanity must be assigned more weightage compared to the academic articles with even highest impact factor.

5. There is no end to the roles the statistics and other discipline can play thus it is better to stop here wait or more elaborated suggestions.

CONCLUSION AND COMMENTS

From the integration of the facts, figures, experiment, observations and inferences, it can be concluded that the destabilized Pakistan economy can be putt back to normal by developing linkages between causes, effects and measures and resuming linkage between macroeconomic and microeconomic economic models. The former will help by reorganization of whole economic model through development of a mechanism for fair distribution of wealth and genuine allocations in the budgets particularly through the elimination of the set ups that produce little and consume more. Development of linkages of economic problems with environmental problems will add significant value.

A final tip may be;

> "Let the macroeconomic players play their game as they cannot escape the international pressures to use macroeconomic indicators as tools. They are the need of the day for sorting out short time solutions. A group should be created to identify the production oriented projects, enlist them, and study them and implement them. To this end.

Economists should work in collaboration with Technologists, Engineers and Scientists to learn about technologies in depth and make them learn practical economics. This exercise will produce interdisciplinary minds with versatile approach. After achieving this target, the nation will own real personnel to handle both sides of the picture. A natural consequence will be that the Country will also economize in use of its talented workforce.

The plan framed to carry out this exercise should be long term; spread over at least two decades."

REFERENCES

1. Ahmed, V. and Rashid, A. (1982). The Management of Pakistan's Economy, Oxford University Press Karachi.
2. American Institute of Pakistan Studies (2006). Pakistan Studies Economy (Retrieved on 2006-11-20).
3. Amjad, R. (2008). *An Employment and Decent Work Strategy for the Punjab:* Background Paper for Speech at Government of the Punjab Lecture Series (Available with Punjab Planning and Development Website)
4. Burki, S.J. (2007). *Punjab Bounty and its Economic Prospects.* Background Paper for Speech at Government of the Punjab Lecture Series (Available with Punjab Planning and Development Website).
5. BYJU'S. *Types of Statistics in Mathematics*. Online available at https://byjus.com/maths/types-of-statistics/
6. Chaudhry, T.T. and Chaudhry, A.A. (2008). The effects of rising food and fuel costs on poverty in Pakistan. *Fourth Annual Conference on the Management of the Pakistan Economy* (April 24-25, 2008). 117-138.
7. Daily Times (2006). *7% Growth Achieved in FY 2005-06*. Jan 12, 2006.
8. Frost, J. (Online). *The Importance of Statics*. Online available at https://statisticsbyjim.com/basics/importance-statistics/
9. Government of Pakistan. Overview of the Economy, In *Pakistan Economic Survey 2021-22*, https://www.finance.gov.pk/survey/chapter_22/Overview.pdf
10. Governor, State Bank of Pakistan (2007). Foreign Investment to Reach $7 Billion during Current Fiscal Year, *Pak Tribune*, April 1, 2007.
11. Gulf Research Center (2008). *GCC Investments in Pakistan and Future Trends*.
12. Hamid, N. (1974). The Burden of Capitalist Growth—A Study of Real Wages and Consumption in Pakistan. *Pakistan Economic and Social Review*, 12(1), 47-70.
13. Haq, M. (1971). Employment and income distribution in the 1970's: A new perspective. *Pakistan Economic and Social Review*, 9(1/2), 1-9.
14. Hasan, P. (2007). *State and Pakistan Economy: Where Have We Come From? Where Do We Go*? https://www.brecorder.com/news/3355442/state-and-pakistan-economy-where-have-we-come-from-where-do-we-go---v-20070116517676
15. International Monetary Fund (2007). *Report for Selected Countries and Subjects*.

16. Khan, M.R. and Qaiser, S. (2020). Comparative Techno-Economic Study of Incinerators and Thermal Power Plants for Disposal of Hospital Waste, Proceedings of the *ISERD Conference: on Basic Sciences and Environment* at Sien Diego USA.
17. Lewis, S.R. Jr. (1970). *Pakistan's Industrialization and Trade Policies*, London Oxford University Press.
18. MeroSpark. *Application of Statistics in Economics*. Online available at https://www.merospark.com/content/173/application-of-statistics-in-economics/
19. Meyer, S. (2008). *Index on Afghanistan April 2008: Famine*. http://www.indexresearch.blogspot.com/2008_04_01_archive.html
20. Muslim Commercial Bank Ltd (2007). *Tales of Success Quid Pro Quo 45*.
21. Noor, S. (2019). *Year in Review: Pakistan in 2019*, (December 30, 2019) https://southasianvoices.org/year-in-review-pakistan-in-2019/
22. Papanek, G.F. (1967). *Pakistan's development*, *social goals and private incentives*. Harvard University Press.
23. Rajan, R.G. (2009). Rent Preservation and the Persistence of Underdevelopment. *American Economic Journal: Macroeconomics*, 1 (1): 178-218.
24. Siddique, A. (2021). *How Pakistan's Economy is Failing its People*. https://gandhara.rferl.org/a/pakistan-failing-economy/31272373.html
25. The Heritage Foundation (2021). 2021 Index of Economic Freedom-Pakistan. Paper presented in the 22nd Annual General Meeting and Conference at Lahore https://www.heritage.org/index/pdf/2021/countries/2021_IndexofEconomicFreedom-Pakistan.pdf
26. The Pakistan Times. Feb 2, 2007. Pakistan Economy Registers 7% Growth Rate for 4th Consecutive Year (Retrieved on 2008-02-12).
27. The World Fact Book. 2008. Pakistan: Economy
28. Trading Economics. *Pakistan GDP Growth Rate* | 2021 Data | 2022 Forecast | Chart. https://tradingeconomics.com/pakistan/gdp-growth
29. Transparency International (2021). *Pakistan Slides in Transparency International's Corruption Perception Rankings*. https://www.geo.tv/latest/332167-pakistan-is-more-corrupt-than-last-year-transparency-international
30. Wikipedia. *Economy of Pakistan*. https://en.wikipedia.org › wiki › Economy of Pakistan
31. Wikipedia. *Pakistan*. http://en.wikipedia.org/wiki/Pakistan
32. Williams, T.A., Anderson, D.R. and Sweeney, D.J. (2023). "Statistics". *Encyclopedia Britannica*, https://www.britannica.com/science/statistics.
33. YesPakistan.com (2004). *Overview of Pakistan's Economy*. Online available at http://www.yespakistan.com/economy/overview-pak.asp
34. Yunous, U. (2021). *The Impact of COVID-19 on South Asian Economies*, Special Report No. 500, United States Institute of Peace.
35. Zach (2022). *The Importance of Statistics in Economics (with Examples).* https://www.statology.org/importance-of-statistics-in-economics/

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 329-332

LOG LINEAR MODELING ON WOMEN'S REPRODUCTIVE HEALTH OF PAKISTAN

Zahra Malik[1] and **Zulaikha Mashkoor**[2]
School of Statistics, Minhaj University, Lahore, Pakistan
Email: [1]malikzahra323@gmail.com
[2]zulaikhamashkoor@gmail.com

ABSTRACT

The women are cornerstones of the family so, the health of women should be good it is essential for whole family. The women's reproductive health with reference to the sustainable development Goal is carried out by utilizing Multiple Indicator Cluster Survey (MICS) for Punjab, secondary data analysis is performed.

Different variables are under studied, Factor analysis technique is applied to identify those factors that affects the women's reproductive health badly, log linear predictive modeling is applied on that identified factors of women's reproductive health.

1. INTRODUCTION

The reproductive health of a women can be described as, entire physical &social development. These elements are related to the reproductive health system, its activity, and its development.

In Pakistan, the population growth rate is 1.9 (World Bank, 2019) while the maternal mortality death rate is very high at 14.47% (world health organization, 2000-2017) the maternal mortality ratio is 186 deaths per 100000 live births, Infant mortality rate is 74 per 1000 live births. There are multiple reasons that impact women's health badly e.g. domestic violence, air pollution, tobacco usage, gender disadvantage, mental disorders in women, Antenatal care, women's autonomy on contraception, and the environment.

Sustainable development goals indicators framework developed by inter-Agency on SDG indicators 48th session of United Nations statistical commission held in the month of March (2017).The indicators and list of targets of 17 SDGs goals were published in the month of July (2017).

In 17 goals of sustainable development goal 3 Good health and well-being is one of the most important target. In united nation these goals were established. To ensure the health lives and promote well- being. The SDGs goal three focus on various aspects of healthy lifestyles.

To predict the salient factors affecting the reproductive health of women.

To purpose a statistical model on factors that affect women's reproductive health.

To examine the association between factors regarding women's reproductive health.

2. LITERATURE REVIEW

Agha (2011) studied the impact of a maternal health scheme on institutional delivery among low income women in Pakistan. Only 39% of deliveries in Pakistan are attended by skilled birth attendants, while Pakistan's target for skilled birth attendance by 2015 is 90%. Primary data was collected through interview in 2008. Regression Analysis was conducted to determine the impact of the voucher scheme on ANC, PNC and institutional delivery. Marginal effect estimated from logistic regression Analysis was used to assess the magnitude of the impact of intervention. 22% increase in ANC, 22% increase in institutional delivery, & 35% increase in PNC.

Aslam et al. (2020) assessed the service utilization of maternal health in Pakistan. To check the factor association with the maternal health regression model was used $n = 16314$ sample size for the selection of covariates, poisson, negative binomial Anderson and Newman's behavioural model were fitted and compared to identify the good model Hypothesis testing was applied the test statistic value $= 6113$, critical value $= 2.7$ $P < 0.001$.

Akyüz et al. (2011) studied the effects of domestic violence on female reproductive health. The domestic violence might be sexual, physical, or emotional. Force women for Abortion to take contraception pills and their husbands having illegal relationships with other women were the indicators of bad reproductive health. The WHO on women's health and domestic violence for a sample of 24000, 23-49% of women had physical violence from their partner, 10-50% had sexual violence. Hypothesis testing showed domestic violence have significant effect on the reproductive health of women.

Asif and Pervaiz (2019) studied the married women for family planning of unmet need. Binary Multinomial logistic Regressions on Demographic health survey (2012-13) data was used. About 72.3% women have fear of side effect of contraceptive use. 21.7% women are employed and 78.3% are unemployed.

Azmat et al. (2015) evaluated the predictors of demand of family planning services and contraceptive use in extreme area of Punjab Pakistan, $n = 3998$ sample of married women were selected from three various districts, Mianwali, Chakwal and Bhakkar. The statistical technique descriptive and logistic regression were applied. Currently used contraceptive ranged from 17% to 21% in these districts. Unsatisfiable need for contraception was 36.6%, 40.6% and 31.9% in three districts.

3. RESEARCH DESIGN AND METHODOLOGY

- The multivariate analysis of women's reproductive health with reference to the sustainable development Goal carried out utilizing Multiple Indicator cluster survey (MICS) for Punjab, secondary data analysis performed.
- To build the multivariate model regarding awareness about the factors that affect the women reproductive health factor analysis technique is applied.
- The log linear predictive model is also used as the method of multidimensional data analysis.

Factor Analysis

Factor analysis is used for factor reduction to identify a small no of factors that explain most of the variance that is observe in much larger number of variables. This technique is applied in this research work to identify the factors because a large amount of data of Multiple Indicator Cluster Survey Punjab (2017-2018) is used.

Scree plot indicates that the 5 factors that are generated by the analysis. These factors are lies on the Eigen value 1.0. These five factors are significant that most affect the women's reproductive health. In component matrix, ever tried cigarette smoking, in the past 12 months, felt discriminated and able to get pregnant can be extracted. The Received prenatal care will be extracted.

From the above analysis it suggests that five factors can be retained which are highly correlated and affect the women's reproductive health badly. However if the variable have the lower value these are not well represented a communality less than 0.4 does not contribute much to the underlying factor. The communality is the proportion of each variable variance that gets explain by the retained factors so we have value against each of the variable here. In the past 12 months, felts discriminated gender, ever tried cigarette smoking, Able to get pregnant and family planning methods will be extracted

The goodness of fit measures can be useful for model and for evaluating the presence of over dispersion and under dispersion. Departures from equi-dispersion can be assessed by computing the ratio of the deviation to its degrees of freedom. Values greater than 1, indicate the presence of over dispersion, whereas values less than 1 signal the presence of under dispersion the values/df column in the output.

The over-dispersion is most likely to be problematic when the ratio of the chi square to its degrees of freedom is greater than 2. Pearson chi-square value 4117.888 highest value. Married or lived with a man once or more than once. This is a likelihood chi square test aimed at testing whether the model containing the full set of predictors fits significantly better than a null model. If the test result is significant, then our model is a significant improvement in fit over a null model.

In this model parameter estimates physically attacked was negative and significant predictor.

$(b = -19.523, S.E = 0.6427, P = 0.000)$

The regression coefficient for start of last menstrual period unit are significant.

$(b = .141, S.E = 0.0436, P = 0.001)$

Something taken by using force or someone threatening to use force are positive but nearly significant.

$(b = .438, S.E = .3418, P = .200)$

Family planning methods slope is negative but significant.

$(b = .498, S.E = 0.1815, P = .006)$

Own a mobile phone slope is negative and insignificant.

$(b = -217, S.E = 5029, P = .666)$

Ever use internet slope positive and significant.

$(b = 20.728, S.E = .1017, P = .000)$

Received prenatal care slope was positive and significant.
$(b = .116, S.E = .0311, P = .000)$

Able to get pregnant was positive and significant.
$(b = .050, S.E = .0375, P = .185)$

Married or lived with a man once or more than once was negative and in significant.
$(b = -.044, S.E = .0809, P = .589)$.

4. COMMENTS AND CONCLUSION

Factor analysis is the factor reduction technique this is used to identify the factors that most affect the women's reproductive health badly.

With the help of factor analysis we conclude physically attacked, start of last menstrual period unit, Family planning methods and ever tried cigarette smoking and received prenatal care are those factors that most affect the women's reproductive health.

Log linear predictive modeling is apply on these identified factors and perform model this model display the factors which have p-values and standard error and regression slope that shows the values are positive or negative and p- values shows the results are significant and model is good.

These techniques help us to find results and conclude that various factors that damage the women's reproductive health.

5. ACKNOWLEDGEMENT

All Praise and thanks to Allah Almighty, who bestowed me with knowledge and skills, and enabled me to complete my thesis work, I would like to express my gratitude and love to my teachers, beloved family and colleagues specially; Madam Zulaikha Mashkoor, my supervisor, for her supervision, encouragement and patience to guide my thesis, a helping hand and all time cooperation.

REFERENCES

1. Agha, S. (2011). Impact of a maternal health voucher scheme on institutional delivery among low income women in Pakistan. *Reproductive Health*, 8(1), 1-14.
2. Aslam, M., Sadiq, M. and Mehmood, T. (2020). Assessment of maternal health services utilization in Pakistan: the role of socio-demographic characteristics. *Asian Biomedicine*, 14(1), 3-7.
3. Akyüz, A., Yavan, T., Şahiner, G. and Kılıç, A. (2012). Domestic violence and woman's reproductive health: a review of the literature. *Aggression and Violent Behavior*, 17(6), 514-518.
4. Asif, M.F. and Pervaiz, Z. (2019). Socio-demographic determinants of unmet need for family planning among married women in Pakistan. *BMC Public Health*, 19(1), 1-8.
5. Azmat, S.K., Ali, M., Ishaque, M., Mustafa, G., Hameed, W., Khan, O.F. and Munroe, E. (2015). Assessing predictors of contraceptive use and demand for family planning services in underserved areas of Punjab province in Pakistan: results of a cross-sectional baseline survey. *Reproductive Health*, 12(1), 1-10.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 333-342

ROBUST ESTIMATORS FOR THE MEDIAN ESTIMATION USING BIVARIATE AUXILIARY INFORMATION

Muhammad Zohaib
Shifa Tameer-e-Millat University, Islamabad, Pakistan
Email: da.qec@stmu.edu.pk

ABSTRACT

This study developed improved family of robust estimators for estimating the population median using information on two auxiliary variables in the presence of extreme observations with the help of robust measures of location under simple random sampling without replacement SRSWOR by extending Irfan et al. (2021). Estimation of Population Median under Robust Measures of an Auxiliary Variable [*Mathematical Problems in Engineering*, *2021*, 1-14]. For these purpose robust measures such as tri-mean, deciles mean, and Hodges-Lehmann estimator are used for the first time for progressive estimation of population median using bivariate auxiliary information. Mathematical properties associated with the robust estimators for median estimation are evaluated using mean square error or minimum mean square error up to first order of approximation. Robustness is also examined using "Mahalanobis Distance". Based on graphical and empirical study, it is found that proposed family of estimators perform more efficiently as compare to existing estimators.

KEYWORDS

Auxiliary variable; Median; Mean squared error; Robustness, Simple random sampling.

1. INTRODUCTION

A wide range of literature is available for estimating the finite population means and totals in sample surveys, relatively less effort has been devoted to the development of efficient methods for estimating finite population median using bivariate auxiliary information e.g. Afifa et al. (2019) firstly, introduced the difference type estimator for population median using two auxiliary variables. Recently, Shabbir et al. (2020) introduced the improved difference type estimator for population median using two auxiliary variables under simple random sampling without replacement (SRSWOR).

In our present study, the problem under consideration is to develop the improved robust family of estimators for median estimation in the presence of extreme observations using two auxiliary variables.

Objectives of this work is as follows:

1) For the first time, to investigate the progressive estimation of population median we used robust measures (i.e., deciles mean, Hodges-Lehman estimator and tri-mean) with the help bivariate information.

2) Mahalanobis distance is used for the first time to check the presence of outliers in real life applications for the median estimation.

Notations:

$$\text{Population Medians} = \begin{cases} M_y \\ M_x, \\ M_z \end{cases}$$

$$\text{Sample Medians} = \begin{cases} \hat{M}_y \\ \hat{M}_x, \\ \hat{M}_z \end{cases}$$

$$\text{Probability density Functions} = \begin{cases} f_y(M_y) \\ f_x(M_x), \\ f_z(M_z) \end{cases}$$

$$\text{Correlation Coefficients} = \begin{cases} \rho_{(\hat{M}_y,\hat{M}_x)} = 4P_{11}(y,x) - 1, \\ \rho_{(\hat{M}_y,\hat{M}_Z)} = 4P_{11}(y,z) - 1, \\ \rho_{(\hat{M}_x,\hat{M}_z)} = 4P_{11}(x,z) - 1, \end{cases}$$

$$\text{Population Tri-Means} = \begin{cases} T_M^x \\ T_M^z, \end{cases}$$

$$\text{Population Hodges Lehmann Estimators} = \begin{cases} H_L^x \\ H_L^z, \end{cases}$$

$$\text{Population Deciles Means} = \begin{cases} D_M^x \\ D_M^z \end{cases}$$

Following error terms are used to obtain the mathematical properties such as bias, mean squared error (MSE) and minimum MSE of the various estimators:

$\xi_0 = \frac{\hat{M}_y - M_y}{M_y}$ = Error term for study variable y,

$\xi_1 = \frac{\hat{M}_x - M_x}{M_x}$ = Error term for auxiliary variable x,

$\xi_2 = \frac{\hat{M}_z - M_z}{M_z}$ = Error term for auxiliary variable z

Such that $E(\xi_0) = E(\xi_1) = E(\xi_2) = 0$.To the first degree of approximation,

$$E(\xi_0^2) = \theta C_{My}^2, \quad E(\xi_1^2) = \theta C_{Mx}^2, \quad E(\xi_2^2) = \theta C_{Mz}^2,$$

$$E(\xi_o \xi_1) = \theta C_{Myx}, \quad E(\xi_o \xi_2) = \theta C_{Myz}, \quad E(\xi_1 \xi_2) = \theta C_{Mxz},$$

$$C_{Myx} = \rho_{yx} C_{My} C_{Mx}, \quad C_{Myz} = \rho_{yz} C_{My} C_{Mz}, \quad C_{Mxz} = \rho_{xz} C_{Mx} C_{Mz}$$

$$C_{My} = \frac{1}{M_y f_y(M_y)}, \quad C_{Mx} = \frac{1}{M_x f_x(M_x)}, \quad C_{Mz} = \frac{1}{M_z f_z(M_z)}$$

$\theta = \frac{1}{4}\left(\frac{1}{n} - \frac{1}{N}\right)$ be the finite population correction (f.p.c) factor.

2. REVIEWING WELL KNOWN MEDIAN ESTIMATORS

This study considered some well-known existing estimators for estimating population median under (SRSWOR) suggested by different researchers.

Traditional difference type estimator using dual auxiliary variables for median estimation

$$\hat{M}_{td0} = \hat{M}_y + h_1\left(M_x - \hat{M}_x\right) + h_2\left(M_z - \hat{M}_z\right) \tag{1}$$

where h_1 and h_2 are unknown constants

$$h_{1(opt)} = \frac{M_y(A_2A_3 - A_4A_5)}{M_x(A_1A_2 - A_5^2)} \text{ and } h_{2(opt)} = \frac{M_y(A_1A_4 - A_3A_5)}{M_z(A_1A_2 - A_5^2)}$$

$$MSE\left(\hat{M}_{td0}\right)_{min} \cong M_y^2\left\{\theta C_{My}^2 - \frac{A_1A_4^2 + A_2A_3^2 - 2A_3A_4A_5}{A_1A_2 - A_5^2}\right\} \tag{2}$$

where $A_1 = \theta C_{Mx}^2, A_2 = \theta C_{Mz}^2, A_3 = \theta C_{Myx}, A_4 = \theta C_{Myz}, A_5 = \theta C_{Mxz}$.

Baig et al. (2019) suggested the difference type estimator using dual auxiliary variable for median estimation. Their estimator consists of two parts, first part consists of usual difference estimator with weight h_3 and second part consist of the sum of exponential ratio and exponential product estimators with weight h_4.

$$\hat{M}_{bag} = \left\{\hat{M}_y + h_3\left(M_x - \hat{M}_x\right)\right\}\left\{h_4 exp\left(\frac{M_z - \hat{M}_z}{M_z + \hat{M}_z}\right) + (1 - h_4) exp\left(\frac{\hat{M}_z - M_z}{\hat{M}_z + M_z}\right)\right\} \tag{3}$$

where h_3 and h_4 are unknown constants

$$h_{3(opt)} = \frac{M_y(B_1B_4 - B_2B_5)}{M_x(B_1B_3 - B_5^2)} \text{ and } h_{2(opt)} = \left(\frac{B_2B_3 - B_4B_5}{B_1B_3 - B_4B_5^2}\right)$$

is given by

$$MSE\left(\hat{M}_{bag}\right)_{min} \cong M_y^2\left\{\theta C_{My}^2 - \frac{B_3B_2^2 + B_1B_4^2 - 2B_2B_4B_5}{B_1B_3 - B_5^2}\right\} \tag{4}$$

where $B_1 = \theta C_{Mz}^2, B_2 = \theta C_{Myz}, B_3 = \theta C_{Mx}^2, B_4 = \theta C_{Myx}, B_5 = \theta C_{Mxz}$.

Shabbir et al. (2020) suggested the improved difference type estimator under simple random sampling without replacement (SRSWOR).

$$\hat{M}_{shb} = \left\{h_5\hat{M}_y + h_6\left(M_x - \hat{M}_x\right) + h_7\left(M_z - \hat{M}_z\right)\right\} exp\left(\frac{M_x - \hat{M}_x}{M_x + \hat{M}_x}\right) \tag{5}$$

where h_5, h_6 and h_7 are unknown constants

$$h_5 = \frac{\lambda_5}{\lambda_4}, h_6 = \frac{\lambda_6}{\lambda_4}, h_7 = \frac{\lambda_7}{\lambda_4} \text{ and } MSE(\hat{M}_{shb})_{min} \cong M_y^2\left(\frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_4}\right) \tag{6}$$

where

$$\lambda_1 = -C_1C_2C_6^2 - C_1C_3C_5^2 + 2C_1C_5C_6C_9 - C_2C_3C_4^2 + 2C_2C_4C_6C_8 + 2C_3C_4C_5C_7 + C_4^2C_9^2 - 2C_4C_5C_8C_9 - 2C_4C_6C_7C_9 + C_5^2C_8^2 - 2C_5C_6C_7C_8 + C_6^2C_7^2,$$

$$\lambda_2 = C_1C_2C_3 - C_1C_9^2 - \; -C_2C_6^2 + 2C_2C_6C_8$$

$$\lambda_3 = -C_2C_8^2 - C_3C_5^2 + 2C_3C_5C_7 - C_3C_7^2 + 2C_4C_9^2 + 2C_5C_6C_9 - 2C_5C_8C_9 - 2C_6C_7C_9 + 2C_7C_8C_9,$$

$$\lambda_4 = C_1C_2C_3 - C_1C_9^2 - C_2C_8^2 - C_3C_7^2 + 2C_7C_8C_9 + C_2C_3 - C_9^2,$$

$$\lambda_5 = C_2C_3C_4 - C_2C_6C_8 - C_3C_5C_7 - C_4C_9^2 + C_5C_8C_9 + C_6C_7C_9 + C_2C_3 - C_9^2,$$

$$\lambda_6 = C_1C_3C_5 - C_1C_6C_9 - C_3C_4C_7 + C_4C_8C_9 - C_5C_8^2 + C_6C_7C_8 + C_3C_5 - C_3C_7 - C_6C_9 + C_8C_9,$$

$$\lambda_7 = C_1C_2C_6 - C_1C_5C_9 - C_2C_4C_8 + C_4C_7C_9 + C_5C_7C_8 - C_6C_7^2 + C_2C_6 - C_2C_8 - C_5C_9 + C_7C_9$$

where

$$C_1 = \theta\left(C_{My}^2 + C_{Mx}^2 - 2C_{Myx}\right), C_2 = \theta C_{Mx}^2, C_3 = \theta C_{Mz}^2,$$
$$C_4 = \theta\left(\frac{3}{8}C_{Mx}^2 - \frac{1}{2}C_{Myx}\right) C_5 = \frac{\theta C_{Mx}^2}{2}, C_6 = \frac{\theta C_{Mxz}}{2},$$
$$C_7 = \theta\left(C_{Mx}^2 - C_{Myx}\right), C_8 = \theta\left(C_{Mxz} - C_{Myz}\right) \text{ and } C_9 = \theta C_{Mxz}.$$

3. IMPROVED ROBUST FAMILY OF ESTIMATORS

One prominent drawback of well-known existing estimators is that they performed inefficient in the presence of extreme/wild observations. In this section, we define a proposed family of robust estimators for median estimation with the linear combination of deciles mean, tri-mean and Hodges Lehmann estimators.

Improved general robust estimator for population median is given as,

$$T_{i(z)} = \widehat{M}_y\left[\left\{\Delta_x d_1\left(\frac{\vartheta_x\widehat{M}_x + \tau_x}{\vartheta_x M_x + \tau_x}\right)^{\lambda_x}\right\} + \left\{\Delta_z d_2\left(\frac{\vartheta_z M_z + \tau_z}{\vartheta_z\widehat{M}_z + \tau_z}\right)^{\lambda_z}\right\}\right] \tag{7}$$

where d_1 and d_2 Are Unknown constants, $\lambda_x = \lambda_z = -1$, for designing new estimators, $\vartheta_x, \vartheta_z, \tau_x$ and τ_z may be any constant values or functions of the known robust measures of location associated with x and z variables.

Suppose $\Delta_x = exp\left(\frac{M_x - \widehat{M}_x}{M_x + \widehat{M}_x}\right)$ and $\Delta_z = exp\left(\frac{M_z - \widehat{M}_z}{M_z + \widehat{M}_z}\right)$.

Table 1
Combination of Robust Measures which are used as Constants for Making Robust Family of Estimators

S#	ϑ_x	ϑ_z	τ_x	τ_z	Estimator
$T_{1(z)}$	D_M^x	D_M^z	1	1	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{D_M^x M_x+1}{D_M^x\widehat{M}_x+1}\right)\right\}+\left\{\Delta_z d_2\left(\frac{D_M^z\widehat{M}_z+1}{D_M^z M_z+1}\right)\right\}\right]$
$T_{2(z)}$	H_L^x	H_L^z	1	1	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{H_L^x M_x+1}{H_L^x\widehat{M}_x+1}\right)\right\}+\left\{\Delta_z d_2\left(\frac{H_L^z\widehat{M}_z+1}{H_L^z M_z+1}\right)\right\}\right]$
$T_{3(z)}$	T_M^x	T_M^z	1	1	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{T_M^x M_x+1}{T_M^x\widehat{M}_x+1}\right)\right\}+\left\{\Delta_z d_2\left(\frac{T_M^z\widehat{M}_z+1}{T_M^z M_z+1}\right)\right\}\right]$
$T_{4(z)}$	T_M^x	T_M^z	H_L^x	H_L^z	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{T_M^x M_x+H_L^x}{T_M^x\widehat{M}_x+H_L^x}\right)\right\}+\left\{\Delta_z d_2\left(\frac{T_M^z\widehat{M}_z+H_L^z}{T_M^z M_z+H_L^z}\right)\right\}\right]$
$T_{5(z)}$	D_M^x	D_M^z	T_M^x	T_M^z	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{D_M^x M_x+T_M^x}{D_M^x\widehat{M}_x+T_M^x}\right)\right\}+\left\{\Delta_z d_2\left(\frac{D_M^z\widehat{M}_z+T_M^z}{D_M^z M_z+T_M^z}\right)\right\}\right]$
$T_{6(z)}$	H_L^x	H_L^z	T_M^x	T_M^z	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{H_L^x M_x+T_M^x}{H_L^x\widehat{M}_x+T_M^x}\right)\right\}+\left\{\Delta_z d_2\left(\frac{H_L^z\widehat{M}_z+T_M^z}{H_L^z M_z+T_M^z}\right)\right\}\right]$
$T_{7(z)}$	T_M^x	T_M^z	D_M^x	D_M^z	$\widehat{M}_Y\left[\left\{\Delta_x d_1\left(\frac{T_M^x M_x+D_M^x}{T_M^x\widehat{M}_x+D_M^x}\right)\right\}+\left\{\Delta_z d_2\left(\frac{T_M^z\widehat{M}_z+D_M^z}{T_M^z M_z+D_M^z}\right)\right\}\right]$

3.1 Bias, MSE and Minimum MSE of $T_{i(z)}$

Expressing the suggested generalized class of estimators $T_{i(z)}$ in terms of ξ_0 and ξ_1 as

$$T_{i(z)}=M_Y(1+\xi_0)\left[\left\{d_1\left(1+\frac{\vartheta_x M_x\xi_1}{\vartheta_x M_x+\tau_x}\right)^{\lambda_x} exp\left(\frac{M_x-\widehat{M}_x}{M_x+\widehat{M}_x}\right)\right\}\right.$$
$$\left.+\left\{d_2\left(1+\frac{\vartheta_z M_z\xi_2}{\vartheta_z M_z+\tau_z}\right)^{-\lambda_z} exp\left(\frac{M_z-\widehat{M}_z}{M_z+\widehat{M}_z}\right)\right\}\right] \tag{8}$$

After some simplification of Eq. (8), we have

$$T_{i(z)}=M_Y(1+\xi_0)$$
$$\left[\begin{array}{l}\left\{d_1\left(1+\varpi_x\xi_1\lambda_x+\frac{1}{2}\lambda_x(\lambda_x-1)(\varpi_x\xi_1)^2\right)exp\left(-\frac{\xi_1}{2}+\frac{\xi_1^2}{4}\right)\right\}+\\ \left\{d_2\left(1-\varpi_z\xi_2\lambda_z+\frac{1}{2}\lambda_z(\lambda_z+1)(\varpi_z\xi_2)^2\right)exp\left(-\frac{\xi_2}{2}+\frac{\xi_2^2}{4}\right)\right\}\end{array}\right] \tag{9}$$

where $\varpi_x = \dfrac{\vartheta_x M_x}{\vartheta_x M_x + \tau_x}$ and $\varpi_z = \dfrac{\vartheta_z M_z}{\vartheta_z M_z + \tau_z}$.

The bias of the proposed estimators, $T_{i(d)}$ are defined as

$$Bias\left(T_{i(z)}\right) \cong E\left(T_{i(z)} - M_Y\right)$$

The bias of generalized class of estimators is as

$$Bias\left(T_{i(z)}\right) = M_y\left[d_1\left\{\begin{array}{c}1+\dfrac{\theta C_{Mx}^2}{2}\left(\dfrac{3}{4}-\lambda_x\varpi_x+\lambda_x\left(\lambda_x-1\right)\varpi_x^2\right)\\ +\theta\rho_{yx}C_{My}C_{Mx}\left(\lambda_x\varpi_x-\dfrac{1}{2}\right)\end{array}\right\} + d_2\left\{\begin{array}{c}1+\dfrac{\theta C_{Mz}^2}{2}\left(\dfrac{3}{4}+\lambda_z\varpi_z+\lambda_z\left(\lambda_z+1\right)\varpi_z^2\right)\\ -\theta\rho_{yz}C_{My}C_{Mz}\left(\lambda_z\varpi_z+\dfrac{1}{2}\right)\end{array}\right\} - \{1\}\right] \qquad (10)$$

The MSE of the proposed estimators, $T_{i(z)}$ are defined as

$$MSE\left(T_{i(z)}\right) = M_Y^2\left[1 + d_1^2 Q_1 + d_2^2 Q_2 + 2d_1 d_2 Q_3 - 2d_1 Q_4 - 2d_2 Q_5\right] \qquad (11)$$

where

$$Q_1 = \Big[1 + \theta C_{My}^2 + \theta C_{Mx}^2(1 + 2\lambda_x^2\varpi_x^2 - \lambda_x\varpi_x^2 - 2\lambda_x\varpi_x) - 2\rho_{yx}\theta C_{My}C_{Mx}(1 - 2\lambda_x\varpi_x)\Big]$$

$$Q_2 = \Big[1 + \theta C_{My}^2 + \theta C_{Mz}^2(1 + 2\lambda_z^2\varpi_z^2 + \lambda_z\varpi_z^2 + 2\lambda_z\varpi_z) - 2\rho_{yz}\theta C_{My}C_{Mz}(1 + 2\lambda_z\varpi_z)\Big]$$

$$\begin{aligned}Q_3 = \Bigg[1 + \theta C_{My}^2 &+ \frac{\theta C_{Mx}^2}{2}\left(\frac{3}{4} + \lambda_x\varpi_x + \lambda_x(\lambda_x - 1)\varpi_x^2\right)\\ &+ \frac{\theta C_{Mz}^2}{2}\left(\frac{3}{4} + \lambda_z\varpi_z + \lambda_z(\lambda_z + 1)\varpi_z^2\right)\\ &+ \theta\rho_{yx}C_{My}C_{Mx}(2\lambda_x\varpi_x - 1)\\ &- \theta\rho_{yz}C_{My}C_{Mz}(1 + 2\lambda_z\varpi_z) + \theta\rho_{xz}C_{Mx}C_{Mz}\left(\frac{1}{4} + \frac{\lambda_x\varpi_x}{2}\right.\\ &\left.- \frac{\lambda_z\varpi_z}{2} - \lambda_x\varpi_x\lambda_z\varpi_z\right)\Bigg]\end{aligned}$$

$$Q_4 = \left\{1 + \theta\rho_{yx}C_{My}C_{Mx}\left(\lambda_x\varpi_x - \frac{1}{2}\right) + \frac{\theta C_{Mx}^2}{2}\left(\frac{3}{4} - \lambda_x\varpi_x + \lambda_x(\lambda_x - 1)\varpi_x^2\right)\right\}$$

$$Q_5 = \left\{1 - \theta\rho_{yz}C_{My}C_{Mz}\left(\lambda_z\varpi_z + \frac{1}{2}\right) + \frac{\theta C_{Mz}^2}{2}\left(\frac{3}{4} + \lambda_z\varpi_z + \lambda_z(\lambda_z + 1)\varpi_z^2\right)\right\}$$

Partially differentiating Eq. (12) with respect to d_1 and d_2 and equating them to zero, we get the optimal values of d_1 and d_2 as follows.

$$d_{1(opt)} = \frac{(Q_2Q_4 - Q_3Q_5)}{(Q_1Q_2 - Q_3^2)}$$

$$d_{2(opt)} = \frac{(Q_1Q_5 - Q_3Q_4)}{(Q_1Q_2 - Q_3^2)}$$

Placing these optimal values in Eq. (12), we obtained the minimum MSE as given by

$$MSE\left(T_{i(z)}\right)_{min} \cong M_Y^2 \left[\frac{Q_1Q_2 + 2Q_3Q_4Q_5 - Q_3^2 - Q_2Q_4^2 - Q_1Q_5^2}{(Q_1Q_2 - Q_3^2)}\right]. \tag{12}$$

4. REAL LIFE APPLICATIONS

Comparison of the $T_{i(z)}$ estimators with the other existing estimators under study is given by using two real-life applications. The population I is taken from Singh [9] and The Population II is taken from Gujrati [7]. Performance of current study is evaluated in terms of percentage relative efficiency (PRE) and scatter plot is considered with the help of Mahalanobis distance which shows the presence of extreme/wild observations.

4.1 Percentage Relative Efficiency

The performance of proposed class of estimators as compared to other competing estimators in terms of percentage relative efficiency (PRE). For this purpose, we selected two real-life populations. For deep study we take different sample size ($n =$ 12, 15 and 17).

$$PRE = \frac{MSE\left(\widehat{M}_{td0}\right)}{MSE\left(\widehat{M}_i\right)}$$

where $\widehat{M}_i = \widehat{M}_{bag}, \widehat{M}_{shb}, T_{i(z)}$.

4.2 Mahalanobis Distance

Mahalanobis distance (MD) is a classical approach for checking the potential outliers in real life applications. For each observation " i " Mahalanobis distances are calculated as follows.

$$MD = \sqrt{(x_i - M)C^{-1}(x_i - M)^T}$$

where "M" indicate the location matrix and "C" shows the covariance matrix of independent variables. The cut of value is $\sqrt{\chi^2_{q,\alpha}}$ where q is the number of variables. The term $\alpha = 97.5\%$ shows level of error. The observations are considered outlier/wild, when they exceed the cut-off value. In the current study, we calculated Mahalanobis distance for population I which have 9 values which are greater than $\chi^2_{3,0.975} = 3.06$ and population II have 14 values which are greater than $\chi^2_{3,0.975} = 3.06$ whose are considered outliers/wild observations. We used scatter plot graphs with the help of Mahalanobis distance to identify the outliers in two real life populations. The cut-off value is given with a straight line in diagrams. Clearly, we see that both data sets have potential outliers/wild observations and our estimator performs efficiently in the presence of extreme observations.

Descriptive measures related to population I is given as follows:

$N = 69, M_y = 2068, M_x = 2307, M_z = 20,$
$\rho_{yx} = 0.1505, \rho_{yz} = 0.3166, \rho_{xz} = 0.1431,$
$f_y(M_y) = 0.00014, f_x(M_x) = 0.00014, f_z(M_z) = 0.00014,$
$T_M^x = 4043,\ T_M^z = 4043, D_M^x = 3853.4,\ D_M^Z = 3615.2,$
$H_L^x = 3004,\ H_L^z = 2935$

Descriptive measures related to population II is given as follows;

$N = 81, M_y = 32.6, M_x = 109, M_z = 101,$
$\rho_{yx} = -0.2098, \rho_{yz} = -0.3087, \rho_{xz} = 0.5308,$
$f_y(M_y) = 0.0396, f_x(M_x) = 0.0282, f_z(M_z) = 0.0178,$
$T_M^x = 109.5,\ T_M^z = 101,\ D_M^x = 110.44,\ D_M^Z = 97.67,$
$H_L^x = 108.75,\ H_L^z = 98.$

Table 2
PRE-of Existing and Proposed Estimators

Estimator	**Population I**			**Population II**		
	$\boldsymbol{n}$ **= 12**	$\boldsymbol{n}$ **= 15**	$\boldsymbol{n}$ **= 17**	$\boldsymbol{n}$ **= 12**	$\boldsymbol{n}$ **= 15**	$\boldsymbol{n}$ **= 17**
$\widehat{M}_{td0}$	100.000	100.000	100.000	100.000	100.000	100.000
$\widehat{M}_{bag}$	100.000	100.000	100.000	100.000	100.000	100.000
$\widehat{M}_{shb}$	125.564	119.089	116.107	101.008	100.771	100.660
$T_{1(z)}$	169.265	148.893	140.184	103.576	103.384	103.294
$T_{2(z)}$	169.265	148.893	140.183	103.576	103.384	103.294
$T_{3(z)}$	169.265	148.893	140.183	103.576	103.184	103.294
$T_{4(z)}$	169.247	148.888	140.183	103.467	103.275	103.185
$T_{5(z)}$	169.237	148.883	140.181	103.465	103.273	103.182
$T_{6(z)}$	169.226	148,878	140.177	103.464	103.272	103.181
$T_{7(z)}$	169.240	148.885	140.181	103.466	103.274	103.184

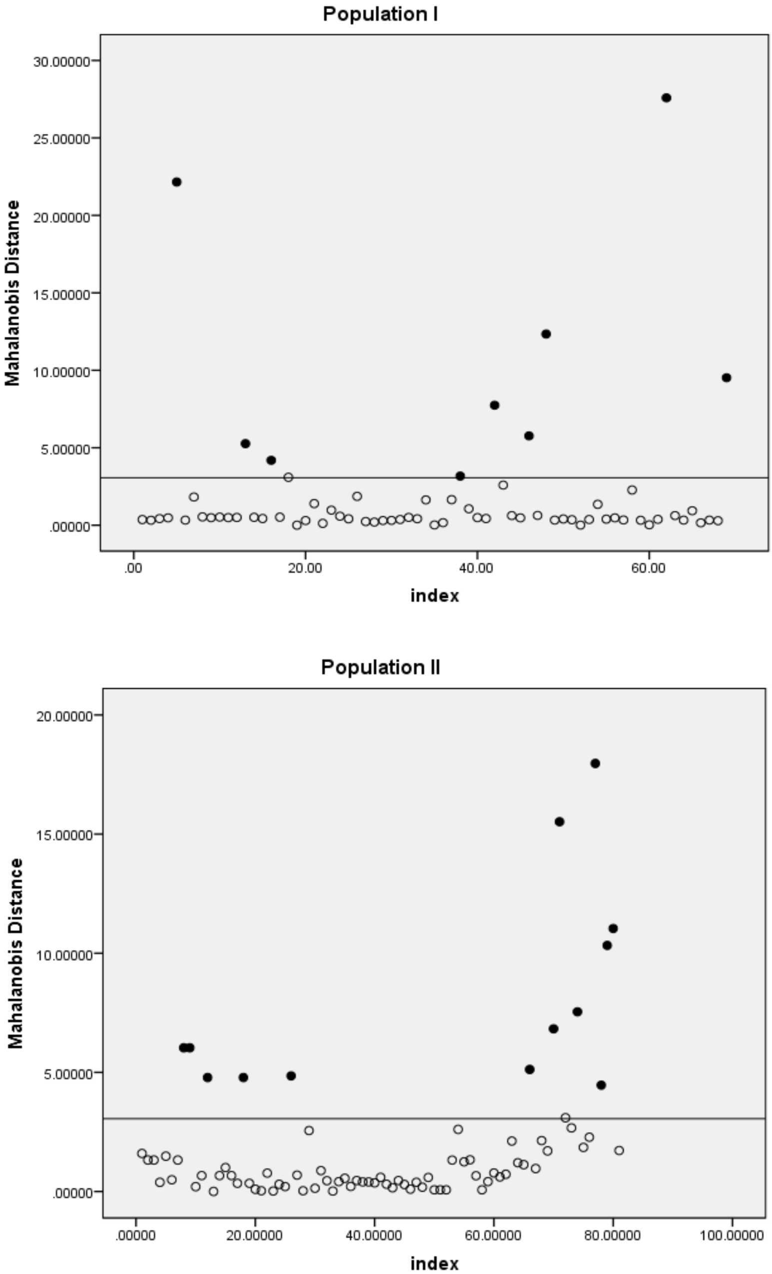


Figure 1: Scatter Plot for Population I and Population II

Important Findings:

The following is revealed from Table 2 and Figure 1:

1) $T_{i(z)}$ Performs efficiently than all existing estimators, that is $\widehat{M}_{td0}, \widehat{M}_{bag}$ and $\widehat{M}_{shb}$.
2) As sample size increase percentage relative efficiency (PRE) increase as compare to existing estimators.
3) Figure 1 shows the presence of extreme values in two real life applications which are used in current study.

5. CONCLUDING REMARKS AND RECOMMENDATIONS

This study proposed an exponential type estimator for estimating population median under simple random sampling using bivariate auxiliary information. Bias, mean squared error and minimum mean squared error of the suggested class of estimators is derived up to first degree of approximation. Two real-life data sets are used to check the numerical performance of suggested estimator in the presence of extreme values. On the basis of numerical findings, it is concluded that suggested estimator can generate efficient results as we increase sample size. Therefore, use of suggested robust family of estimators is recommended for future applications.

REFERENCES

1. Baig, A., Masood, S. and Ahmed Tarray, T. (2019). Improved class of difference-type estimators for population median in survey sampling. *Communications in Statistics-Theoryand Methods*, 49(23), 5778-5793.
2. Government of Pakistan (2004). *Crops Area Production*, Ministry of Food and Agriculture, Islamabad, Pakistan.
3. Government of the Punjab (2012). *Punjab Development Statistics 2012,* Bureau of Statistics, Government of the Punjab, Lahore, Pakistan.
4. Gujrati, D.N. (2004). *Basic Econometrics*, The McGraw-Hill Companies, New York.
5. Hettmansperger, T. and McKean, J.W. (2011). *Robust Nonparametric Statistical Methods*. 2nd Edition Boca Raton: Chapman & Hall / CRC Press.
6. Irfan, M., Javed, M., Shongwe, S.C., Zohaib, M. and Haider Bhatti, S. (2021). Estimation of Population Median under Robust Measures of an Auxiliary Variable. *Mathematical Problems in Engineering*, 2021, 1-14.
7. Murthy, M.N. (1967). Sampling theory and methods, *Statistical Publishing Society*, Calcutta, India.
8. Rousseeuw, P.J. and Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633-639.
9. Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael"" Selected"" Amy* (Vol. 2). Springer Science & Business Media.

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 343-352

ON ESTIMATION OF EXPONENTIATED GENERALIZED INVERTED GOMPERTZ DISTRIBUTION USING BAYESIAN ANALYSIS UNDER NON-INFORMATIVE PRIOR

Zulaikha[1], Wajiha Nasir[2], **Aima** and **Fareeha**
Department of Statistics, Govt. College Women University
Sialkot, Pakistan
Email: [1]zoyamehar464@gmail.com
[2]wajiha.nasir@gcwus.edu.pk

ABSTRACT

In this paper, Exponentiated Generalized Inverted Gompertz Distribution is studied using Bayesian Analysis under non-informative prior. Exponentiated Generalized Inverted Gompertz Distribution is a modified form of Exponential distribution. The four parameters Exponentiated Generalized Inverted Gompertz Distribution $\alpha, \beta, \Upsilon, \theta$ is a generalization of the Exponentiated and the Gompertz distributions. Posterior distributions are derived by using, Uniform and Jeffrey's prior. Bayes estimators and Posterior risks are conducted using different loss functions under a complete sample. Simulation results have been implemented to obtain the accuracy of the estimators. Graphs of posterior distributions for different values of shape parameters are also drawn.

KEYWORDS

Bayesian analysis, MLE, Loss functions, Posterior Distribution under Non-Informative Prior.

INTRODUCTION

Ahmed et al., (2021), have introduced a new three-parameter model called the Extended Inverse-Gompertz (EIGo) distribution. They have studied parameters estimated by the maximum likelihood and Bayesian approaches under Type-II censoring. They have provided an adequate and improved fit with respect to its competing inverted model. They have proposed that the extended inverse-Gompertz distribution provided all important shapes for the hazard rate including increasing, bathtub, decreasing, and unimodal shapes.

Aminzadeh (2019), has studied Inverse Gamma-Pareto composite distribution is considered a model for heavy-tailed data. The maximum likelihood (ML), smoothed empirical percentile (SM), and Bayes estimators (informative and non-informative) for the parameter theta, which is the boundary point for the supports of the two distributions are derived. A Bayesian predictive density was derived via a gamma prior for theta and the density is used to estimate risk measures. The accuracy of estimators of theta and the risk measures are assessed via simulation studies.

Elbatal et al., (2014) said that the inverse Weibull distribution could be readily applied to a wide range of situations including applications in medicine, reliability, and ecology. They have introduced a new model of Generalized inverse Weibull distribution referred to as the Exponentiated generalized inverse distribution. They have derived the moment-generating functions and the rth moment. They also obtained the expressions of a density function, moment generating function, and rth moment of the order statistics. Further, they discussed the estimation of parameters by maximum likelihood and provided the information matrix.

Singh et al., (2013) have proposed Bayes estimators of the parameter and reliability function of inverted exponential distribution under the general entropy loss function for complete type I and type II censored samples. They have compared proposed estimators with the corresponding maximum-likelihood estimators for their simulated risks. From their study they have observed that the risk of the Bayes estimators with an informative prior is the least. Therefore, they have concluded that the Bayes estimator with an informative prior can be used particularly when some a priori information about the parameter θ is known. However, if the prior information about the parameter is not available, one can use the MLE (maximum likelihood estimation).

Guure and Ibrahim (2012), have Determined the Best method by Comparing the Classical maximum likelihood against the Bayesian estimators. They have used an Informative prior and proposed a data-dependent prior known as a Generalized non-informative prior. Then they considered Bayesian estimation under three loss functions and employed Lindley's approximation procedure to reduce the ratio of the integrals. They have obtained the Mean square error and the absolute bias for the purpose of comparison. They conducted this study via simulation by utilizing different sample sizes. From this study, they observed that the generalized prior performed better than the other under linear exponential loss function with respect to mean square error and under general entropy loss function with respect to absolute bias for both the survival function and the failure rate.

Barrigac et al., (2011) have proposed a new two-parameter lifetime Poisson exponential distribution with an increasing failure rate. The new distribution arises on a latent complementary risk scenario. The properties of have proposed distribution are discussed, including a formal proof of its density function and explicit algebraic formulae for its quantiles and survival and hazard functions. Also, we have discussed inference aspects of the model proposed via Bayesian inference by using Markov chain Monte Carlo simulation. A simulation study investigates the frequentist properties of the proposed estimators obtained under the assumptions of non-informative priors. Further, some discussions on model selection criteria are given. The developed methodology is illustrated on a real data set.

Francisco (2010), has proposed a new two-parameter lifetime distribution with an increasing failure rate. He has studied new distribution arising on a latent complementary risk scenario. He has also discussed inference aspects of the model proposed via Bayesian inference by using Markov chain Monte Carlo simulation. A simulation study investigates the frequentist properties of the proposed estimator so obtained under the assumption of non-informative priors. He has developed a methodology illustrated on a real data set.

EXPONENTIATED GENERALIZED INVERTED GOMPERTZ DISTRIBUTION (EGIGD)

The Exponentiated Generalized Inverted Gompertz distribution has been developed by Mahmoud et al. (2021). The P.d.f of Exponentiated Generalized Inverted Gompertz distribution is as follows:

$$f(x)=\frac{\alpha\gamma\theta}{x^2}e^{\frac{\beta}{x}}e^{-\frac{\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\left(1-e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\right)^{\gamma-1}\left[1-\left(1-e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\right)^{\gamma}\right]^{\theta-1};$$

$$x>0;\alpha,\beta,\gamma,\theta>0$$

Respectively, where the parameter β is the scale parameter, and the three parameters α, Υ, θ are the shape parameters of the Exponentiated Generalized Inverted Gompertz distribution respectively according to Mahmoud et al. (2021).

Model and Likelihood Function

The P.d.f of distribution is as follows:

$$f(x)=\frac{\alpha\gamma\theta}{x^2}e^{\frac{\beta}{x}}e^{-\frac{\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\left(1-e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\right)^{\gamma-1}\left[1-\left(1-e^{\frac{-\alpha}{\beta}\left(e^{\frac{\beta}{x}}-1\right)}\right)^{\gamma}\right]^{\theta-1};$$

$$x>0;\alpha,\beta,\gamma,\theta>0$$

The log-likelihood function is as follows:

$$L(x;\alpha,\beta,\Upsilon,\theta)=\frac{\left(\alpha\Upsilon\theta\right)^n}{\prod_{i=1}^{n}x_i^{\,2}}e^{\sum_{i=1}^{n}\frac{\beta}{x_i}}e^{-\frac{\alpha}{\beta}\sum_{i=1}^{n}\left(e^{\frac{\beta}{x_i}}-1\right)}$$

$$\sum_{i=1}^{n}\left(1-e^{-\frac{\alpha}{\beta}\left(e^{\frac{\beta}{x_i}}-1\right)}\right)^{\Upsilon-1}\sum_{1=1}^{n}\left[1-\left(1-e^{-\frac{\alpha}{\beta}\left(e^{\frac{\beta}{x_i}}-1\right)}\right)^{\Upsilon}\right]^{\theta-1}$$

Prior Distribution

The prior distribution is a method that include non-informative prior.

Non-Informative Priors

Sometimes, the deviation of the prior distribution based on information other than the current data is impossible. In that case, we use non-informative priors. They are used when we don't have any information about the parameters. A non-informative prior is one in which little new explanatory power about the unknown parameter is 4 provided by intention. It is often improper. Some non-informative priors are given below:

Using Uniform Prior

The Uniform prior of parameter θ is defined as:

$$p(\theta \mid y) = 1, \ 0 < \theta < \infty,$$

The posterior distribution for parameter θ is:

$$p(\theta \mid y) = \frac{\left[-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right]^{n+1} \theta^{(n+1)-1} \exp\left\{-\theta\left(-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right)\right\}}{\Gamma(n+1)}$$

where $\alpha 1 = n + 1$ and $\beta_1 = -\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)$ which is the density function of Gamma Distribution with parameters $\alpha 1$ and $\beta 1$.

Using Jeffrey's Prior

The Jeffrey's prior of parameter θ is defined as:

$$p(\theta \mid y) = \frac{1}{\theta}, 0 < \theta < \infty$$

The posterior distribution for parameter θ is

$$p(\theta \mid y) = \frac{\left[-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right]^{n} \theta^{(n)-1} \exp\left\{-\theta\left(-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right)\right\}}{\Gamma(n)}$$

where $\alpha 2 = n$ and $\beta_2 = -\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)$ which is the density function of Gamma Distribution with parameters $\alpha 2$ and $\beta 2$.

BES and PRS under Different Loss Functions

This section contains Bayes estimators and posterior risks of the different loss function and also contain their results for non-informative priors. Some loss functions are defined below:

Loss Function (LF)	Bayes Estimator (BE)	Posterior Risk (PR)
Square error (SELF)	$\theta^* = E(\theta)$	$P(\theta^*) = E_{\theta\|X}\left\{L(\theta-\theta^*)\right\} = Var(\theta)$
Precautionary (PLF)	$\theta^* = \sqrt{E(\theta^2 \| X)}$	$P(\theta^*) = 2*\left[\sqrt{E(\theta^2 \| X)} - E(\theta \| X)\right]$
DeGroot (DELF)	$\theta^* = \dfrac{E(\theta^2 \| X)}{E(\theta \| X)}$	$P(\theta^*) = \dfrac{Var(\theta/X)}{E(\theta^2/X)}$

Expression for BES and PRS

This section contains the expressions for BEs and PRs under different loss functions using non-informative priors. They are derived using the expression in the section which are given below:

Expression For Bes and PRs Under SELF		
Prior	**BEs**	**PRs**
Uniform	$\dfrac{n+1}{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$\dfrac{n+1}{\left\{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right\}^2}$
Jeffrey's	$\dfrac{n}{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$\dfrac{n}{\left\{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)\right\}^2}$
Expression For Bes and PRs Under PLF		
Prior	**BEs**	**PRs**
Uniform	$\dfrac{\sqrt{(n+1)(n+2)}}{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$2*\left\{\dfrac{\sqrt{(n+1)(n+2)}-(n+1)}{-\sum_{i=1}^{n}\left(\frac{\lambda}{xi}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}\right\}$
Jeffrey's	$\dfrac{\sqrt{(n)(n+1)}}{-\sum_{i=1}^{n}\left(\frac{\lambda}{xi}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$2*\left\{\dfrac{\sqrt{(n)(n+1)}-(n)}{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}\right\}$
Expression For Bes and PRs Under DELF		
Priors	**BEs**	**PRs**
Uniform	$\dfrac{(n+2)}{-\sum_{i=1}^{n}\left(\frac{\lambda}{x_i}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$\dfrac{1}{n+2}$
Jeffrey's	$\dfrac{(n+1)}{-\sum_{i=1}^{n}\left(\frac{\lambda}{xi}\right)\left(1-e^{-\frac{\lambda}{x_i}}\right)}$	$\dfrac{1}{n+1}$

SIMULATION STUDY

Simulation is the act of taking a model and generally expressed as a computer program and running it several times to obtain results, then collecting the data and calculating the estimated probability. Simulation is conducted to investigate the performance of Bayes estimator and posterior risk under prior distributions using different loss functions. We have simulated the random sample of sizes $n = 30$, 50, 100, and 200 of Exponentiated Generalized Inverted Gompertz Distribution with parameter α =1,1 and parameters β =2,3 Υ =1,1 and θ =3,5 for 10,000 times. The results of the simulation study are summarized in the following tables:

BEs and PRs for SELF when {α, β, ϒ} ∈ {1, 2, 1}					
n	**Priors**	**θ =3**		**θ =5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.1710	0.3556	5.3285	0.9474
50		3.1131	0.1937	5.2456	0.5515
100		3.0674	0.0940	5.1343	0.2636
200		3.0303	0.0449	5.0424	0.1271
30	**Jeffrey's**	3.0811	0.3285	5.1788	0.9250
50		3.0946	0.1956	5.0847	0.5277
100		3.0285	0.0926	5.0702	0.2596
200		3.0065	0.0456	5.0172	0.1265
BEs and PRs for SELF when {α, β, ϒ} ∈ {1, 3, 1}					
n	**Priors**	**θ =3**		**θ =5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.2225	0.3466	5.3563	0.9586
50		3.1047	0.1929	5.2236	0.5463
100		3.1405	0.0924	5.0830	0.2584
200		3.0305	0.0459	5.0445	0.1271
30	**Jeffrey's**	3.0992	0.3306	5.1370	0.9100
50		3.0657	0.1929	5.1131	0.5343
100		3.0385	0.0932	5.0587	0.2583
200		3.0157	0.0457	5.0293	0.1270
BEs and PRs for PLF when {α, β, ϒ} ∈ {1, 2, 1}					
n	**Priors**	**θ =3**		**θ =5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.2595	0.1026	5.3995	0.1700
50		3.1686	0.0612	5.3020	0.1024
100		3.0545	0.0300	5.1193	0.0503
200		3.0568	0.0151	5.0693	0.0251
30	**Jeffrey's**	3.1645	0.1029	5.2423	0.1704
50		3.0894	0.0608	5.1320	0.1011
100		3.0495	0.0302	5.8092	0.0505
200		3.0223	0.0150	5.0922	0.0250

BEs and PRs for PLF when {α, β, ϒ } ∈ {1, 3, 1}					
n	**Priors**	θ =**3**		θ =**5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.2540	0.1024	5.4253	0.1708
50		3.1495	0.0608	5.2600	0.1016
100		3.0680	0.0301	5.1303	0.0504
200		3.0377	0.0150	5.0659	0.0251
30	**Jeffrey's**	3.1232	0.1015	5.2383	0.1703
50		3.0633	0.0603	5.1352	0.1011
100		3.0493	0.0302	5.0781	0.0504
200		3.0196	0.0150	5.0403	0.0251
BEs and PRs for DELF when {α, β, ϒ } ∈ {1, 2, 1}					
n	**Priors**	θ =**3**		θ =**5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.3200	0.0312	5.5186	0.0312
50		3.1967	0.0192	5.3164	0.0192
100		3.0823	0.0098	5.1479	0.0098
200		3.0425	0.0049	5.0727	0.0049
30	**Jeffrey's**	3.1992	0.0322	5.4032	0.0322
50		3.1486	0.0196	5.1999	0.0196
100		3.0540	0.0099	5.1311	0.0099
200		3.0331	0.0049	5.0530	0.0049
BEs and PRs for DELF when {α, β, ϒ } ∈ {1, 3, 1}					
n	**Priors**	θ =**3**		θ =**5**	
		BEs	**PRs**	**BEs**	**PRs**
30	**Uniform**	3.2912	0.0312	5.4979	0.0312
50		3.1832	0.0192	5.2911	0.0192
100		3.0908	0.0098	5.1534	0.0098
200		3.0410	0.0049	5.0648	0.0049
30	**Jeffrey's**	3.1924	0.0322	5.3719	0.0322
50		3.0946	0.0196	5.1888	0.0196
100		3.0602	0.0099	5.1137	0.0099
200		3.0343	0.0049	5.0462	0.0049

We have used this simulation technique to check the behavior of shape parameter of Exponentiated Generalized Inverted Gompertz Distribution. The above tables show the result of posterior distribution under Uniform and Jeffery prior by using different loss functions which are Square error loss function, Precautionary loss function and DeGroot loss function. After the simulation and comparing these loss functions the results show that as we increase sample size Bayes estimator approaches its true value of the parameter and posterior risk tends to decrease. We can see that DELF is performing better because its posterior risk is minimum as compared to other loss functions. Similarly, DELF is performing better. Whereas, the overall results show that the posterior distribution under Jeffery prior to DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors.

APPLICATION

In this study, we applied the proposed distribution on the real data set of the strength of glass for a sample of thirty-one aircraft windows. The data set consists of these observations:

{18.83, 20.8, 21.657, 23.03, 23.23, 24.05, 24.321, 25.5, 25.52, 25.8, 26.69, 26.77, 26.78, 27.05, 27.67, 29.9, 31.11, 33.2, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381}

Priors	**SELF**	**PLF**	**DELF**
	BEs PRs	**BEs PRs**	**BEs PRs**
Uniform	29.0835 26.431	29.0825 0.9018	24.9803 0.0303
Jeffrey's	28.1737 25.605	29.1737 0.9016	28.17 0.0303

DELF is performing better. Whereas the overall result showed that the posterior distribution under Uniform and Jeffery prior with DELF is performing better because it is providing minimum posterior risk as compared to other loss functions and priors. Real-life examples and simulation studies give the almost same result.

CONCLUSION

We have conducted this study to check the behavior of shape parameter of Exponentiated Generalized Inverted Gompertz Distribution. We have conducted simulation study for complete of Exponentiated Generalized Inverted Gompertz Distribution to find out the Bayes Estimator and Posterior Risk of two different priors which are Uniform prior and Jeffery prior. We have considered three different loss functions for simple model, namely square error loss function, Precautionary loss function, DeGroot loss function. The performance of the estimators has been compared on the basis of their simulated risks, obtained under different loss functions. DeGroot loss function under Jeffrey prior has been found better than the rest, because under this loss function posterior risks are least. Further the results indicate that the increase in sample size reduces the posterior risks of the estimates. Also, the graphical representation for simple model shows that the posterior distribution under non-informative prior is symmetrical.

REFERENCES

1. Aminzadeh, M.S. and Deng, M. (2019). Bayesian predictive modeling for Inverse Gamma-Pareto composite distribution. *Communications in Statistics-Theory and Methods*, 48(8), 1938-1954.
2. Elbatal, I. and Muhammed, H.Z. (2014). Exponentiated generalized inverse Weibull distribution. *Applied Mathematical Sciences*, 8(81), 3997-4012.
3. El-Morshedy, M., El-Faheem, A. A., Al-Bossly, A. and El-Dawoody, M. (2021). Exponentiated Generalized Inverted Gompertz Distribution: Properties and Estimation Methods with Applications to Symmetric and Asymmetric Data. *Symmetry*, 13(10), 1868.

4. Elshahhat, A., Aljohani, H.M. and Afify, A.Z. (2021). Bayesian and classical inference under type-II censored samples of the extended inverse Gompertz distribution with engineering applications. *Entropy*, 23(12), 1578.
5. Francisco, L.M., Sage, P.T. and Sharpe, A.H. (2010). The PD-1 pathway in tolerance and autoimmunity. *Immunological Reviews*, 236(1), 219-242.
6. Guure, C.B. and Ibrahim, N.A. (2012). Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data. *Mathematical Problems in Engineering*, https://doi.org/10.1155/2012/329489.
7. Louzada-Neto, F., Cancho, V.G. and Barriga, G.D. (2011). The Poisson–exponential distribution: a Bayesian approach. *Journal of Applied Statistics*, 38(6), 1239-1248.
8. Singh, S.K., Singh, U. and Kumar, M. (2013). Estimation of parameters of generalized inverted exponential distribution for progressive type-II censored sample with binomial removals. *Journal of Probability and Statistics*, https://doi.org/10.1155/2013/183652

Proc. 20th International Conference on Statistical Sciences
Lahore, Pakistan – March 9-11, 2023, Vol. 37, pp. 353-358

CLOUD COMPUTING IN HEALTH SECTOR

Imran Anwar Ujan and **Assadullah Shah**

[1] Faculty of Engineering & Technology, University of Sindh, Jamshoro, Pakistan. Email: iujaniium@gmail.com

[2] Kulliyyah of Information & Communication Technology, International Islamic University, Malaysia.

ABSTRACT

Cloud computing is recognised as an extremely beneficial application for businesses due to benefits including long-term cost savings, simple access to data at any time, and economics. In fact, today's customers can choose from a plethora of free cloud storage services that make it simple to store and distribute data across an entire organisation. The private and governmental sectors are both interested in bettering cloud data storage. Some of these initiatives, however, are just getting off the ground, and there are still technological and human factors concerns that need to be solved before they can be fully implemented. Research pertaining to health is the focus of this analysis. The most significant factors in cloud computing adoption were discovered to be technological readiness, human readiness, organisational support, environment, and security and privacy. We anticipate that this research will aid in expanding our understanding of and preparedness for adopting cloud computing.

INTRODUCTION

Many businesses now consider Cloud Computing (CC) a viable choice, especially as they seek to mitigate the effects of the current economic climate and make decisions that will have little negative impact on the business. While the physical data centre was once a need and a source of IT pride, with the advent of cloud computing, the physical data centre is becoming obsolete due to the lower cost and greater efficiency of its replacement, cloud storage. Easy and ubiquitous access to data and applications, greater cost effectiveness, and a competitive advantage are just a few of the many benefits a business may reap from a well-adopted cloud Hamid, 2015. There will be significant cost reductions in both capital and operating expenses as a result of the company's adoption of CC.

One of the most significant shifts that has affected many industries, not just computing, is the introduction of cloud computing and the business models that have arisen as a result. Nearly 80% of today's global firms are expected to make the switch to cloud computing by 2020 (Kuttikrishnan, 2011). Organizations who lack the capital to construct their own application deployment infrastructure and platforms can now take advantage of cloud services. Pay-as-you-go pricing allows customers to only pay for the services they really utilise. The cloud provides the necessary infrastructure for users to launch and operate their own software. Clients can use these various platforms running a selection of OSes to create, test, and release software for use on virtual servers. On top of that, the cloud is a highly scalable platform that can easily accommodate increased traffic. Since the necessary infrastructure and platforms already exist, businesses and individuals can save a tonne of money by not having to create or manage them themselves. Larger companies are investing in the development of private clouds, which they will operate using in-house resources.

Computing in the cloud has the ability to cut expenses associated with integrating healthcare systems, maximise the use of existing resources, and usher in a new era of technological progress. Because of the way things are now going, people want to be able to access information whenever they want and from wherever they happen to be. One strategy for achieving this objective is to upload patient records and other types of data related to healthcare to the cloud. This novel approach to the delivery of healthcare has the potential to enhance both the efficacy and efficiency of medical care, while also lowering the financial load that is imposed on technology budgets (Horowitz, 2011). On the other hand, it does come with a few challenges, the most obvious of which are issues around the confidentiality of patient information and the adherence of vital legislation such as HIPAA. On the whole, however, it is a viable alternative to the current system. Utilizing this technology can be beneficial to healthcare organisations, despite the fact that cloud computing poses some threats to the safety and privacy of consumers. It has the potential to contribute to an improvement in the overall quality of care that is offered to patients while simultaneously contributing to a reduction in the total costs of providing that care (Muir, 2011).

The capacity to interchange data between various systems will be one of the most significant advantages. This expertise is badly needed in healthcare IT. For instance, cloud computing can facilitate the sharing of EHRs, doctor's references, prescriptions, insurance information, and test results that are maintained in several information systems by healthcare institutions. This is already occurring in the radiological field, where several institutions have migrated to the cloud to save storage costs and facilitate picture exchange (Terry, 2012).

Clinics, hospitals, insurance agencies, pharmacies, and other healthcare organisations may now pool their resources and exchange patient records thanks to cloud computing, leading to higher standards of care and lower costs for the consumer. In light of current market developments, it seems that once all the hurdles associated with cloud-based solutions in healthcare are cleared, they will become the norm.

Migration to the cloud could improve inter-health organisation communication, application development, and collaboration depending on whether the health organization's IT infrastructure is centrally located or scattered among its various medical locations. This method helps cut down on the costly and unnecessary IT personnel needed at each satellite location. It's no secret that most healthcare networks, including those connecting hospitals and their many outposts, were designed and implemented decades ago.

CLOUD COMPUTING IN HEALTHCARE

The ecology that supports the healthcare approach is enormous, rich in variety, and exceptionally convoluted. Companies that provide health insurance, hospital and doctor networks, laboratories, pharmacies, patients, and other types of groupings are included in this category (Wan, Greenway, Harris, & Alter, 2010). All of these are required to behave in accordance with a number of regulations imposed by the government (Raut, 2011). It is imperative that several key pieces of information are communicated between the various components of this ecosystem in a timely manner that is accurate and protected from unauthorised access if this ecosystem is to function effectively and efficiently. In the field of healthcare, maintaining the confidentiality of a patient's medical history is a very delicate and personal topic. This is most likely one of the factors that has slowed down the process of moving healthcare operations to the cloud. When it comes to the cloud, the process of exchanging information needs to make use of modern technology and tools.

There are, however, a great deal of other sorts of data, information, and services, which might be dispersed throughout a number of towns, states, or even nations, and which might gain from cooperating in the cloud. According to Wan, Greenway, Harris, and Alter's (2010) research: As things stand right now, it appears that private clouds will be employed first due to worries about security, and then they will be used in public infrastructure after that. To determine how cloud computing might be of assistance to the healthcare industry, it might be a good idea to begin by making a list of the most important goals currently being pursued in this sector (Cloud Security Alliance, 2010; Wan, Greenway, Harris, and Alter, 2010). It would appear that the things that are most important right now are the increasing expenses of health care, the quality of services provided to patients and customers, privacy, the security and integrity of data, and the ability to recover from disasters. Some of these priorities can be accomplished with the assistance of built-in features such as scalable infrastructure, data centres that store data for an extended period of time, security models, easy access to information, and other similar characteristics.

AVAILABILITY

The health care sector can benefit greatly from cloud services that have a high availability rate, since this will allow them to provide their services with fewer interruptions. The cloud allows for programmatic control, which allows for near real-time scalability. High availability can be achieved by forming a cluster with multiple nodes (Kupferman, Silverman, Jara, & Browne, 2009). Then applications can be scaled up or down as needed to accommodate the workload, even though the resources are fixed at the outset of the computation. Cloud-based administration of healthcare software is expected to increase their accessibility and make them available around the clock, according to some forecasters. Potentially, this might greatly lessen the expense of routine maintenance. The first step in migrating healthcare apps to the cloud could be a thorough grasp of security and privacy problems (Rui & Ling, 2010).

CLOUD MONITORING TOOLS

As cloud computing grows in popularity, many companies are creating systems to keep tabs on the services it delivers. Healthcare firms collaborate with these suppliers to tailor security settings on these technologies to their operations. Cloudkick (Rackspace, 2011), LogicMonitor (LogicMonitor, 2012), Pandora FMS (Pandora, 2011), and others are just a few examples of the growing number of third-party providers offering monitoring and cloud management technologies. Additionally, many companies sell software that does nothing but track whether or not a server or service is online (Barry, 2011). If you use EC2 instances on Amazon Web Services, you can take advantage of AWS's built-in monitoring features.

CLOUD COMPUTING BENEFITS IN HEALTHCARE

There are several ways in which the healthcare industry as a whole could benefit from the new business model made possible by cloud computing. The adoption of the cloud in medical services will greatly benefit patients and healthcare organisations alike through improvements in service quality for patients, more collaboration between healthcare organisations, and lower IT costs for healthcare businesses.

When switching to the cloud, healthcare firms also gain a significant advantage in terms of IT expenses. The adoption of the cloud model will result in the transfer of all information technology activities to a remote cloud-computing infrastructure, where these activities will be stored and executed. Because the new "pay-as-you-go" model enables businesses to pay only for what they use, there is no reason for businesses to invest in expensive hardware infrastructure or software licences. Additionally, there is no need for businesses to keep or train on-site personnel for maintenance, security, or replications because cloud computing providers take care of these tasks.

The value of a human life much exceeds the capacity of the healthcare system (Wang, 2010) Therefore, healthcare services adopted in cloud providers are in line with a cost-effective concept where patients and health organisations benefit from this new technology by enhancing the quality of service patients receive through a distributed high-integrated platform (Wang, 2010), coordinating medical processes, and decreasing the need for costly IT infrastructure investments or upkeep.

CONCLUSION

Cloud computing is becoming more and more popular in the medical field. This can improve and solve a wide range of collaborative information problems in healthcare firms, as well as make them more cost-effective. Costs could also go down if you do your computing in the cloud. When information is shared between various medical institutions, utilising cloud-based applications that are standardised will be beneficial for everyone involved. Patients, physicians, insurance companies, pharmacies, imaging facilities, and other types of businesses fall under this category. The concept of cloud computing will give rise to new issues, such as those pertaining to interoperability and security concerns, which are only a couple examples. The shift toward using cloud computing is progressing at a glacial pace as a direct result of this factor. In spite of all of the obstacles, a future expansion in the adoption of cloud-based systems can hopefully be generated by the implementation of best practises in the design, deployment, and utilisation of it.

REFERENCES

1. Ahuja, S.P., Mani, S. and Zambrano, J. (2012). A survey of the state of cloud computing in healthcare. *Network and Communication Technologies*, 1(2), 12-19.
2. Allen, S. (2011). Cloud Computing and Health Care Security. *Cloud Computing Journal*. Retrieved from http://cloudcomputing.sys-con.com/node/1796151
3. Barry, J. (2011). *Testing the Cloud: Assuring Availability*. Retrieved from https://www.hpcwire.com/2011/08/16/testing_the_cloud:_assuring_availability/

4. Blaisdell, R. (2012). *Cloud benefits in the health Industry*. Retrieved from http://www.cloudtweaks.com/2012/02/cloud-benefits-in-the-health-industry/
5. Cloud Security Alliance. (2010). *Top Threats to Cloud Computing V1.0*. Retrieved from https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf
6. Gullo, C. (2011). *Half of doctors to use medical apps in 2012*. Retrieved from http://mobihealthnews.com/14703/half-of-doctors-to-use-medical-apps-by-2012/
7. Hamid, H.A. and Yusof, M.M. (2015). State-of-the-art of cloud computing adoption in Malaysia: A review. *Jurnal Teknologi*. DOI: 10.11113/jt.v77.6499
8. HealthImaging (2010). *Acumen Nabs ONC Cloud Computing Contract*. Retrieved from http://www.healthimaging.com/index.php?option=com_articles&view=article&id=20648:acumen-nabs-onc-cloud-computing-contract&division=hiit
9. Horowitz, B. (2011). *Cloud Computing Brings Challenges for Health Care Data Storage, Privacy*. Retrieved from http://www.eweek.com/c/a/Health-Care-IT/Cloud-Computing-Brings-Challenges-for-Health-Care-Data-Storage-Privacy-851608/
10. Kabachinski, J. (2011). What's the forecast for cloud computing in healthcare? Biomedical Instrumentation & Technology, 45(2), 146-150. http://dx.doi.org/10.2345/0899-8205-45.2.146
11. Korea IT Times (2010). *Telstra Plans Launch of E-Health Cloud Services, Tip of the Iceberg for Opportunity*. Retrieved from http://www.koreaittimes.com/story/9826/telstra-plans-launch-e-health-cloud-services-tip-iceberg-opportunit
12. Kuo, A.M. (2011). Opportunities and Challenges of Cloud Computing to Improve Health Care Services. *Journal of Medical Internet Research*, 13(3), e67. http://dx.doi.org/10.2196/jmir.1867
13. Kupferman, J., Silverman, J., Jara, P. and Browne, J. (2009). Scaling Into the Cloud. Retrieved from http://cs.ucsb.edu/~jkupferman/docs/ScalingIntoTheClouds.pdf
14. Kuttikrishnan, D. (2011). *Cloud Computing: The road ahead*. Retrieved from https://www.datamation.com/cloud/cloud-computing-the-road-ahead/
15. Logic Monitor (2012). *LogicMonitor: ARCHITECTURE WHITE PAPER*. Retrieved from http://www.logicmonitor.com /downloads/Architecture.pdf?84cd58
16. Microsoft (2012). *Microsoft HealthVault*. Retrieved from http://www.microsoft.com/en-us/healthvault/organize/medical-records.aspx
17. Muir, E. (2011). *Challenges of cloud computing in healthcare integration*. Retrieved from http://www.zdnet.com/news/challenges-of-cloud-computing-in-healthcare-integration/6266971
18. Myers, J.E. (2012). *Data Modeling for Healthcare Systems Integration: Use of the MetaModel*. Retrieved from http://www.metadata.com/whitepapers/myers1.pdf
19. Pandora (2011). *FMS-Virtualization and cloud computing monitoring*. Retrieved from http://pandorafms.com/downloads/PandoraFMS_Virtual_Enviroment_Monitoring.pdf
20. Rackspace Hosting (2011). *High Availability Cloud Environments*. Retrieved from http://www.codeproject.com/Articles/157992/High-Availability-Cloud-Environments
21. Raut, V. (2011). Cloud Computing and Health Care. *Cloud Computing Journal*. Retrieved from http://cloudcomputing.sys-con.com/node/2026409
22. Shimrat, O. (2009). *Cloud Computing and Healthcare. Technology Matters*. Retrieved from http://www.himss.org/content/files/Code%2093_Shimrat_CloudComputingandHealthcare_2009.pdf

23. Terry, K. (2012). *Cloud computing in healthcare: the question is not if, but when*. Retrieved from http://www.fiercehealthit.com/story/ cloud-computing-healthcare-question-not-if-when/2012-01-
24. Violino, B. (2011). *Accenture, AT&T Offer Cloud-Based Medical Imaging*. Retrieved from http://www.information-management.com/news/Accenture-ATT-cloud-computing-medical-imaging-solutio
25. Wan, D., Greenway, A., Harris, J.G. and Alter, A.E. (2010). *Six questions every health industry executive should ask about cloud computing*. Retrieved from http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Cloud_Healthcare_PoV.pdf
26. Wang, X. (2010). *Application of Cloud Computing in the Health Information System*. Computer Application and System Modeling (ICCASM). Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5619051
27. Webb, G. (2012). *Making the cloud work for healthcare, Health Management Technology*. Retrieved from http://connection.ebscohost.com/c/articles/71285402/making-cloud-work-healthcare
28. Whitemore, J. (2012). *Five key considerations for healthcare facilities before moving to the cloud.* Retrieved from http://www.mhimss.org/news/five-key-considerations-healthcare-facilities-moving-cloud
29. Zhang, R. and Liu, L. (2010, July). Security models and requirements for healthcare application clouds. In *2010 IEEE 3rd International Conference on cloud Computing* (pp. 268-275). IEEE.