PEER REVIEW JOURNAL
FOR
RESEARCH AND READINGS IN
APPLIED STATISTICS

ISSN 2321-0877

SANKHYA VIGNAN
सांख्य विज्ञान

NEW SERIES (NSV 14) DECEMBER 2018 No. 2

CONTENTS

Published by
GUJARAT STATISTICAL ASSOCIATION

SANKHYA VIGNAN

સંખ્યા વિજ્ઞાન

(e-mail ID : svgsa2015@gmail.com)

EDITORIAL

INSIDE EVERY SELF MADE MAN THERE IS A POOR KID WHO FOLLOWED HIS DREAM

We are extremly happy to present this issue (NSV 14, Dec. 2018, No.2) to our readers. We express our sincere thanks to all our contributors, evaluators, readers and well wishers for their continous and consistent support which has helped us to achieve our goal.

This issue contains **two articles, two research articles, one review articel, one note, one biography** along with **other items** as usual.

Under the caption of **Management and Statistics first article** is an interesting presentation on New Product Research. It is presented by **A. C. Brahmbhatt.**

Second Article is given by **Dinesh S. Dave** and **Ajay Aggarwal** which is pertaining to global supply chain managemnet.

First Research article presented by **C D Bhavsar** is an interesting econometric study on ARIMA model for air pollutants.

Second Research article is given by **U B Gothi** which represents a study based upon sampling method with mathematical statistics.

One **Review Article** is presented by **Sanjay G. Raval** and **Mahesh Vaghela.** It corresponds to study in social sector development pertaining to India.

A brief **note** on Monte Carlo simulation application is given by **H. M. Dixit.**

Biographical Sketch of eminent Statistician **Prof. J.K. Ghosh** is given by **H. D. Budhbhatti.**

S.V.News letter provides some useful information about seminars and conferences etc. It is presented by **K. Muralidharan.**

Readers Forum section provides some views expressed by readers. It is given by **A.M.Patel.**

We are highly indebted to our following referees who have done excellent job of evaluations for the articles/papers submitted for this issue.

(Their names are given one by one in order of their appearance in the journal.)

(1)	**Ajay Aggarwal**	**(2)**	**Jayesh R. Purohit**
(3)	**R. G. Bhatt**	**(4)**	**M.N. Patel**
(5)	**P. H. Thaker**	**(6)**	**H. S. Mediwala**
(7)	**Shailesh Teredesai.**		

Digital Copy of this issue will be sent to all our readers whose email ID are with us. Printed copy will follow soon. Our contributors will be given offprints of their published articles along with the printed copy and certificate.

Wishing you good health and seasons greetings.

WISH YOU ALL HAPPY AND PROSPEROUS NEW YEAR 2019.

Ahmedabad

Date : 31-12-2018

SANKHYA VIGNAN

PEER REVIEWED REFREED BI-ANNUAL JOURNAL

ISSN:2321-0877

(Journal of Research and Readings in Applied Statistics)

- **Listed at International ISSN Directory, PARIS**

Average Circulation Rate : 850
Average Journal Evaluation Score: 7.50 (As on 30th June, 2018)

PUBLISHED BY GSA
GUJARAT STATISTICAL ASSOCIATION

FROM EDITOR'S DESK

SANKHYA VIGNAN is a peer reviewed refereed Bi-Annually journal that publishes empirical, conceptual and review papers of exceptional quality that contribute to Statistics Theory and enriched Applications of Statistical Techniques in various fields. The objective of the Journal is to disseminate knowledge, which ensures good practice of professional management and its focal point is on research and reflections relevant to academicians and practitioners in the field of **Applied Statistics**.

PEER REVIEW PROCESS

(1) To maintain PEER REVIEW REGISTER giving full details about authors, Refrees, decisions and internal rankings, before printing the Issue

(2) All research articles published in the **SANKHYA VIGNAN** will undergo full peer review process. It should be clear for authors that the Editor In Chief is responsible for the final decision about the submitted papers; have the right to accept\reject any paper. The Editor In Chief will choose any option from the following to review the submitted papers:

a). send the paper to two reviewers, if the results were negative by one reviewer and positive by the other one; then the editor may send the paper for third reviewer or he take immediately the final decision by accepting\rejecting the paper. The Editor In Chief will ask the selected reviewers to present the results within 7 working days, if they were unable to complete the review within the agreed period then the editor have the right to resend the papers for new reviewers using the same procedure. If the Editor In Chief was not able to find suitable reviewers for certain papers then he has the right to accept\reject the paper.

b). sends the paper to a selected editorial board member(s) or reviewer.

c). the Editor In Chief himself evaluates the paper.

(3) The authors of the papers/ Articles accepted for publication need to submit a duly signed copyright form within maximum one week of their receiving intimation of acceptance of their paper/ article.

(4) To give each published issue for evaluation by two external Refrees other than editors) obtain evaluation reports and journal evaluation Score out of 10 points.

(5) To declare average Journal Evaluation Score on the basis of the scores given by these two Refrees for each published issue.

(6) To keep all records in perfect order.

(7) Invite and accept suggestions made by the Refrees for Peer Review.

PEER REVIEW COMMITTEE:

It consists of six members:

- **Editor In Chief**
- **Two Joint Editors**
- **Two External Referees**
- **Editorial Secretary**

PAPERS/ MANUSCRIPT EVALUATION CRITERIA

Each submitted manuscript is evaluated on the following basis:

- The originality of its contribution.
- The soundness of its theory and methodology.
- The coherence of its analysis.
- The ability to communicate to readers (grammar and style).
- All research papers are reviewed by at least two suitably qualified experts.
- All publication decisions are made by the journals' Editor on the basis of the reviews provided.
- Normal turn-around time for evaluation of manuscripts is 21 days from the date of recipient.

GUIDELINES FOR AUTHORS

1. The first page of the article should contain full name of authors with designation and correspondence address, contact Numbers and E-Mail Ids (Both Institution and residential)
2. The research must be original in nature and is neither published nor under consideration for publication anywhere else.
3. The review process may take 2-3 days and the status would be known within 24 hours of submission of the article. It is mandatory to mention a valid email and mobile number. We acknowledge the receipt of your article and subscription fee by email.
4. The abstract should not be more than 200-250 words and it should adequately describe the work and highlight its significance. Avoid the use of abbreviation and reference in the abstract. The abstract should be followed by relevant keywords.
5. The main paper (Manuscripts) should not exceed more than 3,000 to 4,000 words(including graphs&charts).
6. Article should be typed in 12 point - Times New Roman Fonts English with a one and half space and single column with 1 Margin on a Standard A4 size Paper. The Page numbers should be at the center of every page. All headings & sub headings must be in bold letters.
7. Table should be numbered consecutively,the title of the table should be placed above the table. The source should be indicated at the bottom.
8. All the tables, charts, graphs, diagrams should be in black and not in colors.
9. Footnotes, italics, and quotation marks should be kept to the minimum.
10. References should be mentioned in APA Referencing Format.

HOW TO SUBMIT

a) We will accept soft copies of article through online submissions at the E-Mail ID: (i) svgsa2015@gmail.com, (ii) drjayesh.purohit@gmail.com (iii)bbjani2012@gmail.com

b) Two hard copies should be sent to

Dr. B. B. Jani

Chief Editor, Sankhya Vignan

B/14, Bansidhar Appartments, Mirambica Road, Naranpura, Ahmedabad - 380 013

(R) 079-27476770 (M) 9824057902

NEW PRODUCT RESEARCH

A. C. Brahmbhatt[1]

ABSTRACT

In this modern world, manynew consumer products are generated day by day. This is attributed due to customer demands and new innovations. However the failure rate for new products is nearly about 70 to 80 percent. THis signifies the need for new product research. In this paper, an effort is made briefly to specify the new research product areas by means of Mathematical Programming and Simulation techniques.

KEY WORDS

New Product Development (NDP), NonLiner Programking (NLP), Monte Carlo Simulation

* * *

The heart and soul of any business is creativity and innovation. Peter Drucker, the father of modern management said," A business has two and only two basic functions—Marketing and Innovation, Marketing and Innovation produce results, all the rest are costs."In these days of cut –throat competition, the companies have to continually do the product innovations. There are companies that attach tremendous significance to New Product Development(NPD), the glaring example is 3M company in U.S.

The companies can develop new products in their own laboratories or it can contact with independent researchers or new product consulting firms. NPD is a challenging task as 'newness' is a relative term . The company may claim that it has created a new product , but similar product by its competitor may have already entered into the market – may be a day before!! Again, new product

(1) Research Mentor, PDPU, Gandhinagar. email : acbpramukh@Hotmail.com
(rcd. Oct'18 / rvd. Dec. 18)

have different connotations –it can be totally new to the world products, it could be a new product line, it can be a supplement to the established product lines by adding flavors, package sizes etc.,it could be a modified version of existing product etc. Developing a new product is daunting task because the failure rate of the new products in the world context is very high. According to Harvard Business School Professor Clayton Christensen , each year 30,000 new consumer products are launched and 80% of them failed! In Grocery sector, according to Professor Inez Blackburn of University of Toronto, the failure rate is 70% TO 80%!!

The New Product Development is a process that begins with the first stage of generating ideas followed by the stages like screening ideas, developing product concept and testing, business analysis, developing a prototype, test marketing and finally commercializing the product. In this paper the focus is on only the first stage of idea generation and a brief discussion as to how a few quantitative approaches could be useful.

The new ideas emanate from varieties of following sources , call them search areas.

i) Brainstorming sessions
ii) Surveying customers on their future needs
iii) Focus Group Discussions(FGDs) with customers, company engineers, designers.
iv) Monitoring the competitor's moves for developing new products
v) Search of existing patents
vi) Allow time off-scouting time –to technical people to come up with new breakthrough ideas
vii) lateral marketing- - that combines two product concepts or ideas to create a new offering.
e.g. cafeteria and internet , these two concepts combined, gave rise to Cyber Café's, audio and portable combined, gave rise to Sony Walkman etc.

The researchers have successfully employed quantitative approaches to find out the optimum allocation of search efforts.

(A) Mathematical Programming Approach:

Here the objective function is

To Maximize the Expected Return(ER)

$$ER = \sum_i R_i(X_i)$$

Subject to

$$\sum_i X_i \le B, \quad X_i \ge 0 \text{ for all } i$$

Where Xi :money spent on search efforts in area i(the above 7 are search areas) B is the total search budget available $R_i(X_i)$ is a general function representing the Expected Return from new ideas generated by searching in area i , with search budget X_i.

If we plot a curve with Xi on X-axis and Ri(Xi) on Y-axis, the shape of the curve would be, Increasing, Concave down, reflecting the diminishing returns to search efforts. Such a response to search efforts that the company is faced with is more akin to Non Linear Programming Problem(NLP) and could be solved by piecewise linear approximations.

(B) Simulation Approach

Pessemier [1] has developed a simulation model to aid product mangers in the generation of new product ideas. He has used the input data such as the areas to be considered for search, the cut off rate of return for acceptance of proposal, search strategies to be considered, response in each search area in terms of the rate of generation of new product proposals etc. Using the input data, a Monte Carlo simulation generates a distribution of return on investment to searching in each area i, with search policy j and cutoff return criterion k. Simulation is run a sufficient number of times to have a clear idea of the distribution. Then select the best combination of search area, strategy and cut off return on the basis of expected return on investment and risk measured by the standard deviation of return for the area-strategy-cut off combination.

RERFERENCE:

[1] **Pessemier E.A.**, New Product Decisions: An Analytical Approach (New York, McGraw-Hill, 1966, Chapter:2

ACKNOWLEDGEMENT:

I thank the referee for reviewing my paper.

MANAGEMENT AND STATISTICS
ARTICLE

A NOTE ON GLOBAL SUPPLY CHAIN MANAGEMENT AND SWOT

Dinesh S. Dave[(1)] and Ajay K. Aggarwal[(2)]

ABSTRACT

Supply chain management is becoming a vital business function in today's global competitive environment. In order to receive global business opportunities, it would be important for an organization to be a part of a global supply chain network. A supply chain network could involve global entities such as suppliers, manufacturers, warehouses, logistics companies, and retailers. It is important for an organization to conduct a thorough SWOT analysis in order to mitigate risk prior to entering in a global supply chain network. This study briefly describes the SWOT approach for supply chain variables.

KEYWORDS:

Supply chain management, global supply chain management, SWOT approach.

*****

Supply Chain Management has been gaining an enormous popularity in today's global competitive environment, particularly the world is continue to become flat. SCM includes the management of such activities as demand planning, sourcing, production, scheduling, pricing, logistics, materials management, warehousing, and customer satisfaction (see Chopra and Meindl, 2015). Additionally, successful organizations participate in a supply chain

(1) Head and Professor, Department of Computer Information Systems & Supply Chain Management, John A. Walker College of Business, Appalachian State University, Boone, North Carolina 28608 USA

(2) Professor of Management, School of Business, Henderson State University, Arkadelphia, Arkansas 71999 USA

(rcd. Sept.' 18 / rvd. Dec.'18)

network which involves suppliers, manufacturers, warehouses, retailers, transporters, and others. The main objective of a successful supply chain network is to create the value of each organization in the network, minimize the risk of uncertainty and volatility, minimize the cost, and continually enhance the customer satisfaction. Thus, an efficient supply chain network manages information flow, material flow, and financial flow while minimizing the risk and add the value to it all partner in the network.

In an ever-changing global environment, SCM integrates business processes from the end consumers via original suppliers, manufacturer, retail organizations, and third party logistics companies. This business to business collaboration could be a critical success factor, especially when almost all organizations are involved in today's dynamic global competitive environment. The results of this collaboration can improve value of each entity of the network and it can be accomplished by sharing resources, information, and knowledge, to improve cycle times, costs, inventory, and customer satisfaction. Additionally, an effective collaboration can provide a better production scheduling for all partners in the network as well as mutual beneficial product development, demand strategies, and logistics planning. Also, mutual cooperation in the SC network with increased information availability and knowledge sharing, supply chain partners can be more responsive to uncertain demand in a global environment. Further, to achieve a greater supply chain performance that include cost, quality, and flexibility, organizations should consider functional, geographical, and even information technology integration.

In today's competitive environment, an organization's global strategy becomes more important. A company's global strategy may lead toward lower cost as well as an access to items that are not available in the local region. Also, the global strategy can provide a competitive advantage through the production of products as per the needs of customers of different countries. A global supplier network of suppliers, manufacturers, distribution centers, and retailer organizations to acquire raw materials, perform manufacturing processes, and use the logistics systems to deliver products to customers in the right quantity at the right price, at the lowest possible lead-time. In a global supply chain network, partner companies of one country can achieve competitive advantages by developing collaboration and dependency with organizations in other country (see Srinivasan, et al., 2014). Some of the benefits of global supply chain management (GSCM) include accomplishing lower production costs of all partners in the GSC network while enhancing customer satisfaction. Additional benefits of GSCM include exploring new markets, enables business growth as well as learning new technologies and management practices from their international partners.

It is very critical for an organization to conduct a thorough SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis before becoming an active partner in a global supply chain. The SWOT methodology can address company's short-term, medium range, and long-term goals of global supply chain and logistics functions including lower cost, shorter lead-time, and high customer satisfaction (see Johnson, 2011). The SWOT approach may provide company's information on lower production cost, a better sourcing strategy, better negotiation strategy, consolidation of shipment for lower logistics cost, multiple vendor selection, and inventory turns. Weakness, for example, may include longer lead-times, inconsistency in quality specifications from multiple vendors, fluctuation of exchange rates, and the management of vendor selection. Opportunities may include lower cost of production facility and warehouse space, better economy of scale, expansion of product line, and knowledge of new management practices. Threats might include the environment risk, political situation in other countries, and arrival of defective products from the vendors. Some of the threats may include higher work-in process inventory that may lead into a bullwhip effect. An organization can mitigate the risk by performing a detail SWOT analysis since several variables change constantly in the global environment. This additional information regarding the environmental variables will potentially enhance the management decision in pursuing a global supply chain strategy.

ACKNOWLEDGEMENTS:

We appreciate insightful comments provided by the referees to enhance the quality of our paper.

REFERENCES:

[1] **Chopra, S. and Meindl,** P., "Supply Chain Management: Strategy, Planning, and Operation," Sixth Edition, Pearson, 2015

[2] **Johnson, I.**, "Assessing the Company's Supply Chain with a SWOT Analysis," Best Practices, Inventory Management & Supply Chain, Drive Your Success, https://www.driveyoursuccess.com/2011/03/index.html

[3] **Srinivasan, M., Stank, T., Dornier, P., and Petersen, K.**, "Global Supply Chains: Evaluating Regions on an EPIC Framework," McGraw Hill, 2014

RESEARCH ARTICLE

ARIMA : A TOOL FOR MODELLING AIR POLLUTANTS

C. D. BHAVSAR*

ABSTRACT

Developing technologies affect the social, economic, and environmental dimensions of the world, often in ways that are entirely unanticipated. The most important formidable challenges which we are facing in the 21st-22nd centuries are to maintain and improve soil, water and air quality. Pollutants from oil and chemical spills, Pesticide and fertilizer runoff, abandoned industrial and mining sites, air borne gaseous and particulate matter from automobile are responsible for exacerbation situation. Air pollution creates eco-toxic effects on terrestrial and aquatic flora and fauna and is sensitive to eco systems. A study is carried out to build up an appropriate model and to predict the future values of critical pollutants such as Sulphur (SO_2), nitrogen oxides (NO_2), and Suspended Particulate Matter (SPM) and Respirable Suspended Particulate Matter (RSPM) present in the ambient air of a Naroda area of Amdavad city using Box-Jenkins methodology (ARIMA).

KEY WORDS

Air Pollutants, NO_2, SO_2, SPM, RSPM, ARIMA

1. INTRODUCTION:

Air pollution is rapidly becoming a major environmental issue which is adversely affecting both the developed and developing countries. World Health Organization has estimated 4-8 per cent death can be directly attributed to air pollution. It is believed

* Professor, Dept. of Statistics, Gujarat University, Ahmedabad
email : chetna_bhavsar@yahoo.co.in.
(rc. Dec'18 / rv. Jan. '19)

that situation is rapidly becoming grave in south-east Asian cities including China & India and is primarily responsible for death of about 50,000 people. Despite increasing awareness and legislations for tapping air-pollution, the air quality is deteriorating in several cities across the globe at alarming rate. The situation is becoming more complicated as about half of the world population has migrated to urban areas and corresponding measures for pollution control are not adequate. Today, a considerable number of people breathe air having quality below WHO air quality guidelines.

Fresh air is essential for good health of people and good environment. Improvements have been made in many countries over last three to four decades to maintain the batter environment. Particulate and gaseous emission of pollutants from industries decreasing productivity and deteriorate that of artistic and cultural patrimony in urban cities.

Within India, the qualities of air vary across the cities and within cities, across areas. Thus, there is spatially and temporal variation in quality of air. An in depth understanding of such distribution is required to put things in right perspective and initiate remedial measures. Keeping this in view, this paper attempts to build up an appropriate model and to predict the future values of critical pollutants present in the ambient air of a Naroda area of Amdavad city using Box-Jenkins methodology (ARIMA). The data of air pollutants is collected from CPCB (central Pollution Control Board) web site for the time period February 2006 to October 2009 of Naroda area of Amdavad city.

Time series is a fundamental technique of statistical analysis where data is arranged according to time. 'Box-Jenkins' forecasting models are based on statistical concepts and are able to model a wide spectrum of time series behaviour. It has a large class of models and a systematic approach for identifying the correct model. There are statistical tests for verifying model validity and statistical measures of forecast uncertainty.

2. METHODOLOGY:

In the initial stage of ARIMA Methodology, simple graph of the data points should be plotted to determining an appropriate model. The input data must be adjusted to a stationary series. Autoregressive moving average model is the general model introduced by Box and Jenkins (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. The first step is to make the input series of ARIMA, a stationary series that is the series under study must have a constant mean, variance, and autocorrelation through time. We usually start with taking the first difference to make it stationary.

Estimation and Forecasting:

At the next step, the parameters are estimated, so that the sum of squared residuals is minimized. The estimates of the parameters are used in the last stage of forecasting to calculate new values of the series and confidence intervals for those predicted values. The estimation process is performed on transformed (differenced) data while the forecasts are generated after integrating the series so that the forecasts are expressed in values compatible with the input data.

Evaluation of the Model:

Model is evaluated on the basis of approximate t values computed from the parameter standard errors. If these values are not significant, the respective parameter can be dropped from the model without affecting substantially the overall fit of the model. The analysis of ARIMA residuals constitutes an important test of the model. The estimation procedure assumes that the residual are not auto correlated and that they are normally distributed.

Limitations:

The ARIMA method is appropriate only for a time series that are stationary (i.e., its mean, variance, and autocorrelation should be approximately constant through time) and it is recommended that there are at least 50 observations in the input data. It is also assumed that the values of the estimated parameters are constant throughout the series.

3. INTERPRETATION OF MODEL FIT STATISTICS

Stationary R-squared is a measure that compares the stationary part of the model to a simple mean model. This measure is preferable to ordinary R-squared when there is a trend or seasonal pattern. R-squared is an estimate of the proportion of the total variation in the series that is explained by the model. This measure is most useful when the series is stationary.

RMSE (Root Mean Square Error) is a measure of how much a dependent series is varies from its model-predicated level, expressed in the same units as dependent series. MAPE (Mean Absolute Percentage Error) is a measure of how much a dependent series is varies from its model-predicated level. It is independent of the units used and therefore be used to compare series with different units.

MaxAPE (Maximum Absolute Percentage Error) is the largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for our forecasts. MAE (Mean Percentage Error) measures how much a series varies from its model-predicated level. MAE is reported in the original series units. MaxAE (Maximum Absolute Error) is the largest forecasted error, expressed in the same units as the dependent series. Maximum absolute error for a large series value is slightly larger than the absolute error for a small series value.

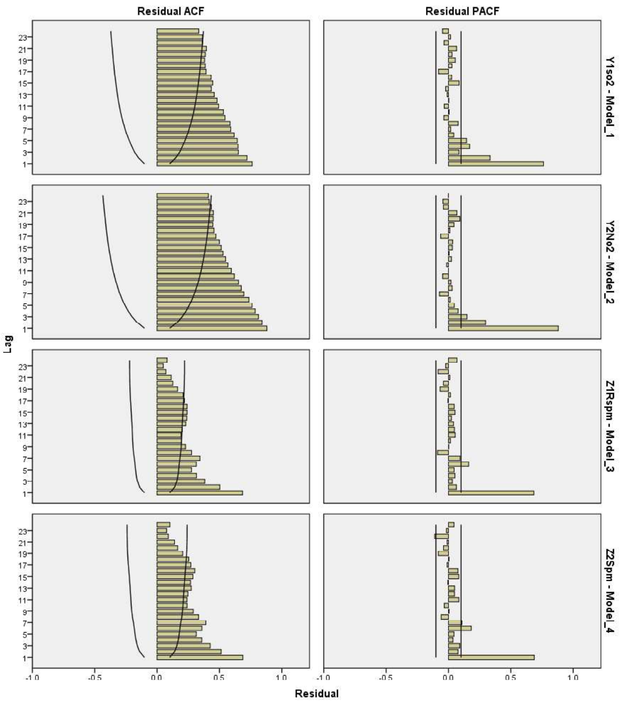


Fig-1: Chart of ACF and PACF for original series up to 23 lags

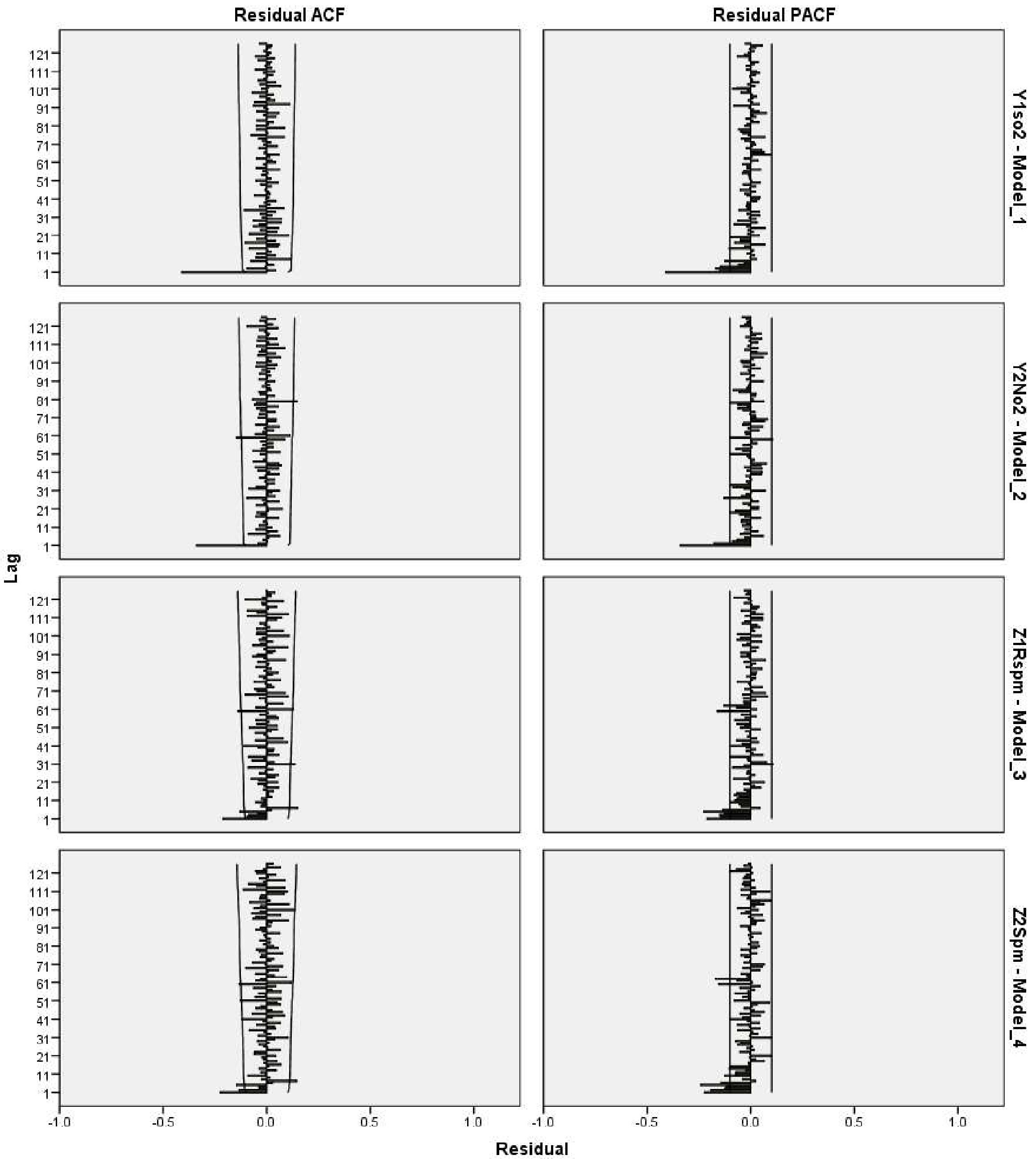


Fig-2: Chart of ACF and PACF after first difference up to 120 lags

Normalized BIC (Normalized Bayesian Information Criterion) is a general measure of the overall fit of a model that attempts to account for model complexity. It is a score based upon the mean square error and includes a penalty for the number of parameters in the model and the length of the series. The penalty removes advantage of models with more parameters, making the statistic easy to compare across different models for the same series. Ljung-Box Statistic Q(18) is a lack of fit statistic used to test the null hypothesis that the model is correctly specified. Also known as the modified Box-Pierce statistic.

4. DATA ANALYSIS

The data for air pollutants SO_2, NO_2, Rspm and Spm collected for Naroda area of Amdavad city for the period February 2006 to October 2009 is considered for the analysis. Firstly we check the stationary of the time series. Before perusing any formal test, we first plot the graph of the observed time series and following plots are observed. We can observe from the graph of observed values that the likely nature of all the original time series for air pollutants SO_2, NO_2, Rspm and Spm are not of stationary time series.

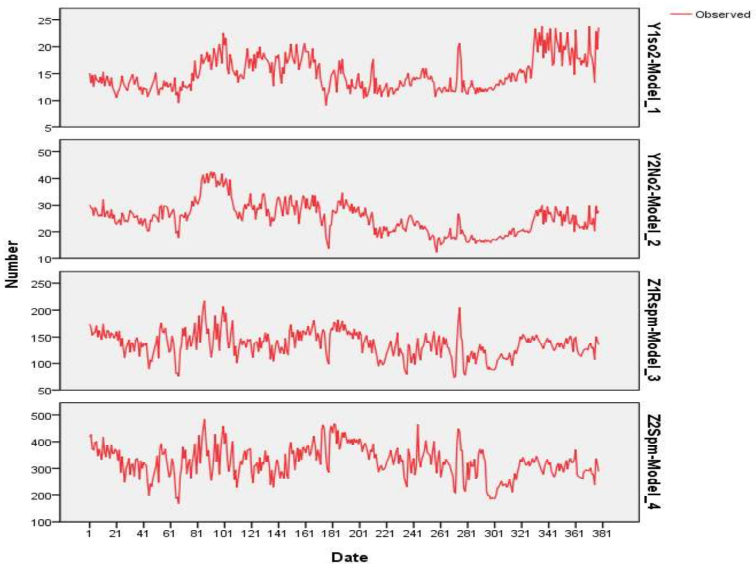


Fig-3: Chart of the original series for observed values

We plot the auto correlation function (ACF) and partial autocorrelation function (PACF) of the original time series and following graphs are observed. The ACF graph is decreasing slowly for SO_2 and NO_2 while it decreasing faster as compare to SO_2 &

NO_2 for Rspm and Spm. We can observe a spike at the first step of PACF graph for all the air pollutants – SO2, NO2, Rspm & Spm which indicate non stationary nature of the series. Generally such kind of the series becomes stationary at the first difference.

5. ARIMA MODEL FOR AIR POLLUTANTS

We considered data for air pollutants SO_2, NO_2, Rspm and Spm for Naroda area of Amdavad city for the period February 2006 to October 2009 to model ARIMA model and to predict the future values of air pollutants. Since all the series are seems to be non-stationary, we have to convert these non-stationary time series to stationary time series by differencing the series. We take first difference of the data points and again find ACF and PACF and draw the graph of it.

After identifying the stationary of the series, the next step is to identify appropriate value of auto regressive terms (p) and moving average terms (q). We have considered 378 data points for air pollutants SO_2, $NO_{2,}$ Rspm and Spm. According to the rule of thumb we should check at least 1/3rd to 1/4th lag of auto regressive terms to get proper value of auto regressive terms (p). Hence, we check '126' autoregressive lag terms for all air pollutants series one by one and try to obtain appropriate model for the series.

Analysis for SO_2:

We check '126' autoregressive lag terms for SO_2 series and observed that 1st ,2nd ,3rd and 4th lags has significant effect while other lags does not have significant effect which can be observed in the following table.

Table showing significant lags of AR for SO_2 :
ARIMA Model Parameters for ARIMA(126,1,0)

Y1so2-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	.013	.025	.521	.603
AR	**Lag 1**	**-.548**	**.065**	**-8.369**	**.000**
	Lag 2	**-.303**	**.074**	**-4.078**	**.000**
	Lag 3	**-.345**	**.077**	**-4.471**	**.000**
	Lag 4	**-.194**	**.081**	**-2.404**	**.017**
	...	**....**	**....**	**....**	**.....**
	Lag 125	.064	.106	.604	.546
	Lag 126	-.053	.095	-.558	.577
	Difference	1			

Thus after reducing the lag order we consider ARIMA model (4, 1, 0) for SO_2. In case of AR terms, we considered only the significant lag terms for modelling. Note that in case of MA no lag is found to be significant for ARIMA (0, 1, 126). But if we consider the ARIMA (0, 1, 30) model then we can observe that 1^{st}, 3^{rd}, 17^{th}, 21^{st} and 26^{th} lags has significant effect while other lags does not have significant effect which can be observed in the following table.

Table showing significant lags of MA for ARIMA(0, 1, 30) for SO_2

Y1so2-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	.017	.026	.646	.519
	Difference	1			
MA	**Lag 1**	**.523**	**.054**	**9.654**	**.000**
	Lag 3	**.162**	**.062**	**2.632**	**.009**
	Lag 17	**.180**	**.063**	**2.858**	**.005**
	Lag 21	**-.155**	**.065**	**-2.402**	**.017**
	Lag 26	**.140**	**.067**	**2.100**	**.036**
	Lag 30	-.106	.059	-1.782	.076

Similarly we check all combinations of significant lags of AR as well as significant lags of MA to obtain the most significant model. The various fit parameters, model fit statistics and LB-statistic for the significant models is observed and the details are given in the following table.

Table-1: Table of various Models fit Parameters and Model fit Statistics for SO_2

		Model Type			
		ARIMA(4,1,0)	ARIMA(0,1,30)	ARIMA(4,1,26)	**ARIMA(2,1,26)**
Model fit statistics	St. R-squared	.235	.312	.278	**.273**
	R-squared	.652	.687	.671	**.669**
	RMSE	1.839	1.809	1.800	**1.798**
	MAPE	8.778	8.597	8.705	**8.696**
	MAE	48.450	46.223	43.848	**46.400**
	MaxAPE	1.335	1.306	1.319	**1.318**
	MaxAE	7.131	6.771	6.854	**6.918**
	Normalized BIC	1.298	1.674	1.333	**1.284**
Ljung-Box Q(18)	Statistics	17.116	-	7.833	**9.130**
	DF	14	-	9	**12**
	Sig.	.250	-	.551	**.692**
AR	Sig. lag No.	1, 2, 3, 4	0	1, 2, 3, 4	**1, 2**
MA	Sig. lag No.	0	1, 3,17,21,26	1, 3,17,21,26	**3,17,21,26**

The ARIMA model parameters with significant lags and significance of the fitted model are presented in the following table.

ARIMA Model Parameters for SO_2

		Estimate	SE	t	Sig.
	Constant	.018	.026	.687	.493
AR	Lag1	-.562	.051	-10.914	.000
	Lag2	-.300	.059	-5.089	.000
MA	Lag3	.323	.058	5.576	.000
	Lag17	.141	.051	2.782	.006
	Lag21	-.106	.052	-2.042	.042
	Lag26	.137	.053	2.595	.010

Table showing parameters of the fitted model for SO_2

Model fit statistics								Ljung-Box Q(18)		
Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC	Statistics	DF	Sig.
.273	.669	1.798	8.696	46.400	1.318	6.918	1.284	9.130	12	.692

On the basis of RMSE, MAPE and Maximum absolute percentage error we conclude that the ARIMA model (2, 1, 26) with significant AR lag numbers 1st & 2nd and significant MA lags numbers 3rd, 17th, 21st & 26th is the best fitted model for the SO_2 time series. The best fitted model can be written as follows:

$$Y_t = 0.018 - 0.562Y_{t-1} - 0.300Y_{t-2} + 0.323u_{t-3} + 0.141u_{t-17} - 0.106u_{t-21} + 0.137u_{t-26}$$

20.60, 21.59, 19.48, 21.47, 21.26, 20.46, 20.90 and 21.60 are the predicated values for fitted model.

Analysis for NO_2:

We check '126' autoregressive lag terms for NO_2series and observed that 1st ,2nd ,3rd ,27th , 51st and 60th lags has significant effect while other lags does not have significant effect which can be observed in the following table.

Table showing significant lags of AR for NO_2:

ARIMA Model Parameters for ARIMA(126,1,0)

Y2No2-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	-.015	.037	-.396	.693
AR	**Lag 1**	**-.441**	**.065**	**-6.801**	**.000**
	Lag 2	**-.261**	**.071**	**-3.699**	**.000**
	Lag 3	**-.176**	**.073**	**-2.408**	**.017**
	Lag 27	**-.174**	**.077**	**-2.264**	**.024**
	Lag 51	**-.170**	**.083**	**-2.060**	**.040**
	Lag 60	**-.212**	**.083**	**-2.565**	**.011**
	....	**...**	**...**	**...**	**...**
	Lag 126	-.052	.078	-.667	.506
	Difference	1			

After reducing the lag order we consider ARIMA model (60, 1, 0) for NO_2. In case of AR terms not all 60 lags are considered but only lags with significant effect are taken into account for modelling the series. Note that in case of MA no lag is found to be significant.

Table showing lags of MA for NO_2 :

ARIMA Model Parameters for ARIMA(0,1,126)

Y2No2-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	-.010	.053	-.197	.844
	Difference	1			
MA	Lag 1	.359	76.447	.005	.996
	Lag 2	.054	35.593	.002	.999
	Lag 3	.033	65.670	.001	1.000
	Lag 4	-.017	123.876	.000	1.000
	...	...	...	...	...
	Lag 125	-.044	8.116	-.005	.996
	Lag 126	.046	4.764	.010	.992

Similar to SO_2 we check six significant lags of AR to obtain the most significant model for the NO_2 series. The various fit parameters; model fit statistics and LB-statistic for the significant models for the NO_2 series are observed and the details are given in the following table.

Table-2: Table of various Models fit Parameters and Model fit Statistics for NO_2

		ARIMA (126, 1,0)	ARIMA (0,1,126)	ARIMA (60,1,0)	ARIMA (60,1,0)
Model fit statistics	St .R-squared	.855	.354	.181	**.174**
	R-squared	.855	.846	.805	**.804**
	RMSE	2.726	2.813	2.603	**2.607**
	MAPE	7.040	7.164	8.025	**8.033**
	MAE	28.824	33.103	35.320	**35.368**
	MaxAPE	1.738	1.783	2.005	**2.005**
	MaxAE	6.986	7.271	8.488	**8.400**
	Normalized BIC	3.999	4.067	2.024	**1.995**
Ljung-Box Q(18)	Statistics	-	-	14.286	**16.974**
	DF	0	0	12	**14**
	Sig.	-	-	.284	**.258**
AR	Significant lag no	1,2,3,27,51, 60,	0	1,2,3,27, 51,60	**1,2,27,60**
MA	Significant lag no	0	0	0	**0**

The ARIMA model parameters with significant lags and significance of the fitted model are presented in the following table.

ARIMA Model Parameters for NO_2:

		Estimate	SE	t	Sig.
	Constant	-.006	.075	-.085	.932
AR	Lag1	-.386	.050	-7.680	.000
	Lag2	-.170	.050	-3.391	.001
	Lag27	-.117	.051	-2.275	.024
	Lag60	-.151	.054	-2.798	.005

Table showing parameters of the fitted model for NO_2

Model fit statistics								Ljung-Box Q(18)		
Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC	Statistics	DF	Sig.
.174	.804	2.61	8.033	35.37	2.0	8.40	1.995	16.974	14	.258

On the basis of RMSE, MAPE and Maximum absolute percentage error we conclude that the ARIMA model (60,1,0) with significant AR lag numbers 1st , 2nd , 27th and 60th is the best fitted model for the NO_2 time series. The best fitted model can be written as follows:

$$Y_t = -0.006 - 0.386Y_{t-1} - 0.170Y_{t-2} - 0.117Y_{t-27} - 0.151Y_{t-60}$$

27.39, 27.64, 27.24, 26.98, 27.97, 27.69 ,27.32 and 27.43 are the predicated values for fitted model.

Analysis for Rspm:

We check '126' autoregressive lag terms for Rspm series and observed that Almost first 12 lags has significant effect while other lags does not have significant effect which can be observed in the following table.

Table showing significant lags of AR for Rspm :
ARIMA Model Parameters for ARIMA(126,1,0)

Z1Rspm-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	-.036	.098	-.372	.710
AR	**Lag 1**	**-.359**	**.065**	**-5.519**	**.000**
	Lag 2	**-.330**	**.069**	**-4.773**	**.000**
	Lag 3	**-.397**	**.073**	**-5.454**	**.000**
	Lag 4	**-.312**	**.077**	**-4.040**	**.000**
	Lag 5	**-.410**	**.079**	**-5.182**	**.000**
	Lag 6	**-.275**	**.083**	**-3.305**	**.001**
	Lag 7	**-.196**	**.085**	**-2.311**	**.022**
	Lag 8	**-.170**	**.086**	**-1.963**	**.051**
	Lag 9	**-.235**	**.087**	**-2.683**	**.008**
	Lag 10	**-.224**	**.089**	**-2.527**	**.012**
	Lag 11	**-.192**	**.090**	**-2.142**	**.033**
	Lag 12	**-.239**	**.091**	**-2.634**	**.009**
	...	**...**	**...**	**...**	**...**
	Lag 125	-.021	.083	-.251	.802
	Lag 126	-.029	.076	-.377	.706
	Difference	1			

We check all combinations of significant lags for ARIMA modelling with their various model fit parameters and observed the following detail for Rspm series.

Table-3: Table of various Models fit Parameters and Model fit Statistics for Rspm

		Model Type			
		ARIMA(12,1,0)	ARIMA(10,1,0)	ARIMA(10,1,0)	ARIMA(6,1,0)
Model fit statistics	St. R-squared	.187	.178	.168	**.164**
	R-squared	.489	.484	.478	**.475**
	RMSE	17.436	17.455	17.511	**17.538**
	MAPE	10.239	10.280	10.334	**10.410**
	MAE	53.314	55.527	55.607	**56.623**
	MaxAPE	13.403	13.473	13.533	**13.632**
	MaxAE	59.815	59.219	62.788	**65.117**
	Normalized BIC	5.922	5.877	5.852	**5.839**
Ljung-Box Q(18)	Statistics	6.524	10.439	14.768	**18.263**
	DF	6	9	11	**12**
	Sig.	.395	.316	.193	**.108**
AR	Significant lag No.	1,2,3,4,5,6,7,8,9,10,11,12	1,2,3,4,5,6,8,9,10	1,2,3,4,5,6,10	**1,2,3,4,5,6**
MA	Significant lag No.	0	0	0	**0**

The ARIMA model parameters with significant lags and significance of the fitted model are presented in the following table.

ARIMA Model Parameters for Rspm:

		Estimate	SE	t	Sig.
	Constant	-.085	.362	-.235	.814
AR	Lag 1	-.343	.052	-6.666	.000
	Lag 2	-.274	.053	-5.191	.000
	Lag 3	-.262	.054	-4.886	.000
	Lag 4	-.224	.054	-4.175	.000
	Lag 5	-.269	.053	-5.078	.000
	Lag 6	-.135	.052	-2.598	.010

Table showing parameters of the fitted model for Rspm

Model fit statistics								Ljung-Box Q(18)		
Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC	Statistics	DF	Sig.
.164	.475	17.538	10.41	56.623	13.632	65.117	5.839	18.263	12	.108

On the basis of RMSE, MAPE and Maximum absolute percentage error we conclude that the ARIMA model (6,1,0) with significant AR lag numbers 1st ,2nd ,3rd, 4th, 5th, and 6th is the best fitted model for the Rspm time series. The best fitted model can be written as follows:

$$Y_t = 0.085 - 0.343Y_{t-1} - 0.274Y_{t-2} - 0.262Y_{t-3} - 0.224Y_{t-4} - 0.269Y_{t-5} - 0.135Y_{t-6}$$

134.96, 135.32, 130.16, 129.67, 134.14, 135.02, 134.61 and 134.57 are the predicated values for the fitted model.

Analysis for Spm:

We check '126' autoregressive lag terms for Spm series and observed that 1-12 , 14th , 58th , 60-63, 114th , 118-120, and 122nd lags has significant effect while other lags does not have significant effect which can be observed in the following table.

Table showing significant lags of AR for ARIMA (126, 1,0) for Spm

Z2Spm-Model_1, No Transformation		Estimate	SE	t	Sig.
	Constant	-.148	.245	-.607	.545
AR	**Lag 1**	**-.406**	**.066**	**-6.109**	**.000**
	Lag 2	**-.449**	**.072**	**-6.254**	**.000**
	Lag 3	**-.472**	**.079**	**-5.944**	**.000**
	Lag 4	**-.346**	**.085**	**-4.058**	**.000**
	Lag 5	**-.490**	**.086**	**-5.673**	**.000**
	Lag 6	**-.342**	**.092**	**-3.732**	**.000**
	Lag 7	**-.231**	**.093**	**-2.489**	**.013**
	Lag 8	**-.205**	**.093**	**-2.197**	**.029**
	Lag 9	**-.236**	**.092**	**-2.558**	**.011**
	Lag 10	**-.289**	**.092**	**-3.151**	**.002**
	Lag 11	**-.211**	**.092**	**-2.300**	**.022**
	Lag 12	**-.225**	**.092**	**-2.444**	**.015**
	Lag 14	**-.192**	**.092**	**-2.096**	**.037**
	Lag 58	**-.248**	**.093**	**-2.660**	**.008**
	Lag 60	**-.331**	**.098**	**-3.379**	**.001**
	Lag 61	**-.241**	**.102**	**-2.367**	**.019**
	Lag 62	**-.289**	**.102**	**-2.835**	**.005**
	Lag 63	**-.333**	**.104**	**-3.205**	**.002**
	Lag 114	**-.216**	**.105**	**-2.060**	**.040**
	Lag 118	**-.223**	**.104**	**-2.154**	**.032**
	Lag 119	**-.230**	**.104**	**-2.204**	**.028**
	Lag 120	**-.232**	**.105**	**-2.208**	**.028**
	Lag 122	**-.253**	**.100**	**-2.535**	**.012**
	Difference	1			

We check all combinations of significant lags for ARIMA modelling with their various model fit parameters and observed the following detail for Spm series.

Table-4: Table of various Models fit Parameters and Model fit Statistics for Spm

		Model Type			
		ARIMA (122,1,0)	ARIMA (63,1,0)	ARIMA (63,1,0)	ARIMA (60,1,0)
Model fit statistics	St. R-squared	.256	.256	.213	.209
	R-squared	.538	.538	.511	.508
	RMSE	41.191	41.191	41.565	41.614
	MAPE	9.849	9.849	10.113	10.124
	MAE	50.940	50.940	52.263	52.194
	MaxAPE	30.565	30.565	31.411	31.451
	MaxAE	145.824	145.824	149.854	149.953
	Normalized BIC	7.814	7.814	7.612	7.599
Ljung-Box Q(18)	Statistics	-	15.148	18.581	17.089
	DF	0	4	9	10
	Sig.	-	.004	.029	.072
AR	Significant lag No.	1,2,3,4,5,6,7,8,9,10,11,12,14,58,60, 61,62,63,114,118, 119,120,122	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 58, 60, 62,63	1, 2, 3, 4, 5, 6, 10, 60, 63	1, 2, 3, 4 5, 6, 10, 60
MA	Significant lag No.	0	0	0	**0**

The ARIMA model parameters with significant lags and significance of the fitted model are presented in the following table.

ARIMA Model Parameters for Spm:

		Estimate	SE	t	Sig.
	Constant	-.282	.768	-.368	.713
AR	Lag 1	-.362	.051	-7.139	.000
	Lag 2	-.324	.052	-6.233	.000
	Lag 3	-.236	.054	-4.380	.000
	Lag 4	-.225	.053	-4.198	.000
	Lag 5	-.305	.053	-5.794	.000
	Lag 6	-.149	.051	-2.915	.004
	Lag 10	-.095	.048	-1.996	.047
	Lag 60	-.128	.048	-2.693	.007

Table showing parameters of the fitted model for Spm:

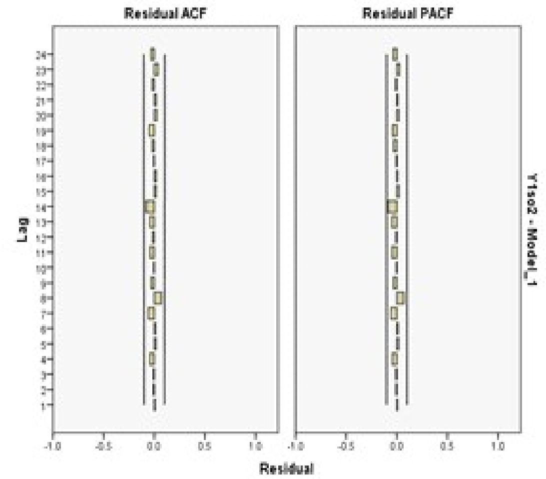
On the basis of RMSE, MAPE and Maximum absolute percentage error we conclude that the ARIMA model (60,1,0) with significant AR lag numbers 1st , 2nd ,3rd , 4th , 5th , 6th ,10th and 60th is the best fitted model for the Spm time series. The best fitted model can be written as follows:

$$Y_t = -0.282 - 0.362Y_{t-1} - 0.324Y_{t-2} + 0.236Y_{t-3} - 0.225Y_{t-4} - 0.305Y_{t-5} - 0.149Y_{t-6} - 0.095Y_{t-10} - 0.128Y_{t-60}$$

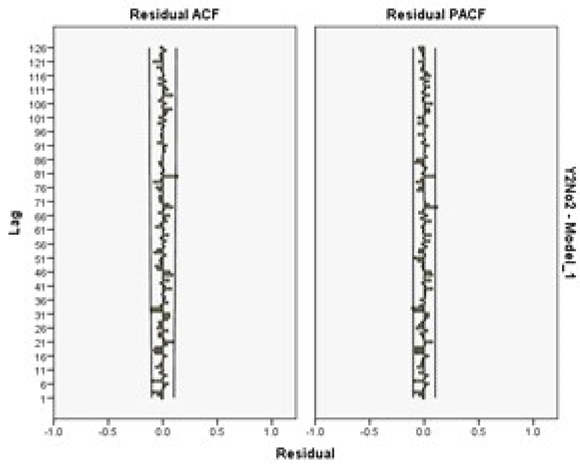
291, 291, 277, 275, 295, 296, 296 and 283 are the predicated values for fitted model.

Model fit statistics								Ljung-Box Q(18)		
Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	Normalized BIC	Statistics	DF	Sig.
.204	.508	41.614	10.124	52.194	31.451	149.953	7.599	17.089	10	.072

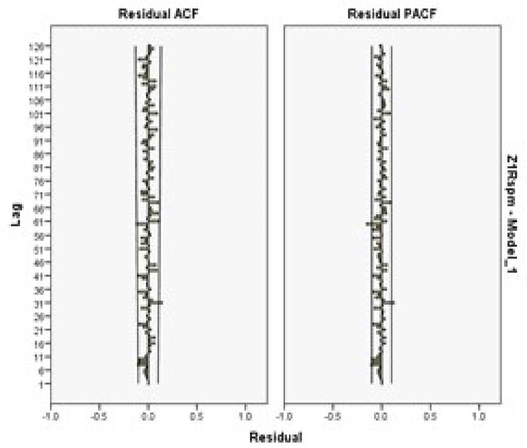
Note that we can observe that ACF and PACF of residual lies within 95% confidence interval and the histogram with normal curve for noise residuals indicating good fitting of the model for all the air pollutants.



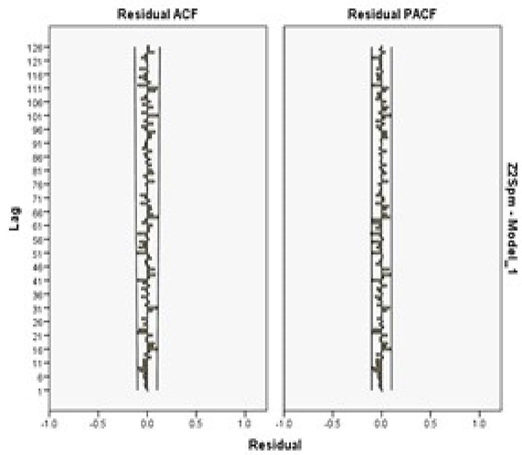
1 ACF and PACF of Residual for SO_2



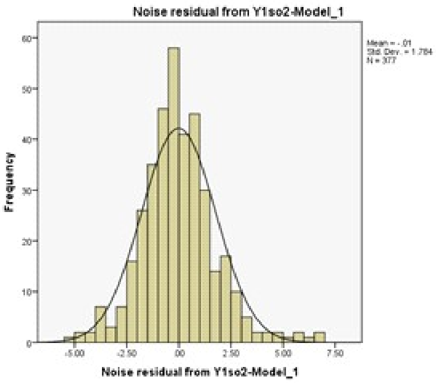
2 ACF and PACF of Residual for NO_2



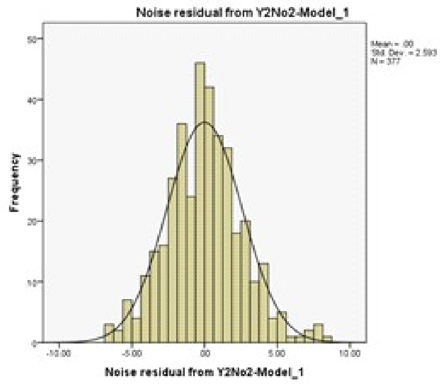
3 ACF and PACF of Residual for *Rspm*



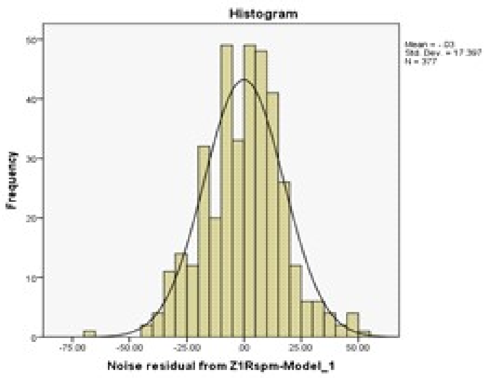
4 ACF and PACF of Residual for *Spm*



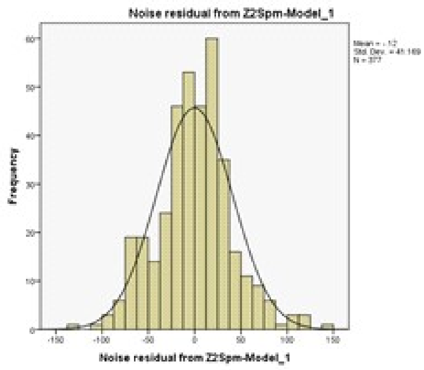
Normal curve for Noises Residual for SO_2



Normal curve for Noises Residual for NO_2



Normal curve for Noises Residual for Rspm



Normal curve for Noises Residual for Spm

6. RESULST AND DISCUSSION

In this paper we apply ARIMA model to predict the future values of air pollutants; SO_2, NO_2

Rspm and *Spm*. The model fit parameters and model fit statistics for significant models are given in table-1 to table-4 for SO_2 , NO_2 ,*Rspm* and *Spm* respectively. Different model fit statistics like stationary R-squared, R-squared, RMSE, MAPE, MAE, MaxAPE, MaxAE, Normalized BIC, Ljung-Box Q (18) statistics and specified significant lags of autoregressive terms as well as moving average terms are given in the respective table.

7. ACKNOWLEDGEMENT

The author is grateful to Shri. G. G. Shah, chief executive officer, NDDB, Anand for their help in collection of data for this research work.

I also thank the referee for reviewing my paper.

8. REFERENCES:

[1] **Box, G. E. P. And Jenkins,** G. M. (1970); Time series Analysis, Forecasting and Control, Holden-Day, Inc., San Francisco.

[2] **Douglas C. Montgomery and Lynwood A. Johnson** (1976); Forecasting and Time series Analysis, McGRAW-HILL Book Company, New York.

[3] **Gujarati, Damodar N.**(2006); Basic Econometrics, Tata Mc Grow hill.

[4] **T. Senthilnathan,** (2009) ; Study of Criteria Air Pollutants in the Industrial Background Ambient Air, IJEP, 29(4), 312-315.

[5] http://www.cpcb.nic.in/RealTimeAirQualityData.php#

RESEARCH ARTICLE

ESTIMATORS OF POPULATION MEAN AND VARIANCE IN THE PRESENCE OF EXTREME OBSERVATIONS

U. B. GOTHI[1]

ABSTRACT

In this paper, the estimators $\bar{y}_t$ and S_t^2 are defined to estimate population mean μ and variance σ^2 by replacing all the sample observations smaller than pre-determined cutoff point t_1 by t_1 and greater than the pre-determined cutoff point t_2 by t_2. The expressions of bias and mean square error (MSE) of the estimators $\bar{y}_t$ and S_t^2 are derived. The relative efficiencies (REF) of these estimators as compared to usual unbiased estimators are obtained. It is demonstrated by means of an example that the regions of t_1 and t_2 exist for which $\bar{y}_t$ and S_t^2 are better than $\bar{y}$ and S^2 respectively.

KEY WORDS:

Estimator, Mean square error, Relative efficiency.

1. INTRODUCTION

If the precise form of the population from which an independent random sample $y_1, y_2, \ldots\ldots, y_n$ is taken, is not known, then we know that sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ are used to estimate the population mean μ and variance σ^2. Suppose that the data have been collected to estimate μ

(1) Associate Professor, Department of Statistics, St. Xavier's College(Autonomous), Ahmedabad, Gujarat, India. (M) 9427020010
e-mail: ubgothi@gmail.com (rcd.Oct. '18 / rvd. Dec. '18)

and σ^2 from an unimodal population. Due to the sampling process, an occasional sample may contain one or more observations from the right tail or left tail of the distribution which are large or small sample observations and they cannot be termed outliers. When the sample size is small, $\bar{y}$ and σ^2 will differ from μ and σ^2 respectively by a considerable amount and therefore they are not proper estimators of μ and σ^2. So, if we are not interested in unbiased estimators but we prefer the estimators which are close to true values (i.e. having smaller mean square error), then we may even discard these offending observations.

If a pre-determined point t, which separates 'large' observations from the rest, is obtained by the external information, then Searls (1966) suggested a procedure for the estimation of mean. By the external information, what we need, is the rough idea about the general region in which the data may occur. He defined the estimator

$\bar{y}_t = \frac{1}{n}\left[\sum_{j=1}^{r} y_j + (n-r)t\right]$ (r = 1, 2,….,n, $y_j \le t$, j = 1,2,……r) of mean μ, by replacing

all the sample observations larger than the pre-determined cutoff point t by the value of t itself; where r is the number of observations less than or equal to t. He showed that a region for t exists for many distributions in which MSE($\bar{y}_t$) is smaller than Var($\bar{y}$).

Following the same technique, Ojha (1980) defined an estimator

$S_t^2 = \frac{1}{n-1}\left[\sum_{j=1}^{r} y_j + (n-r)t^2 - n\bar{y}_t^2\right]$ (r = 1, 2,….,n; y_j d" t, j = 1,2,……r) and he

showed by an example, that a region of t exists for which is better than .

In this paper, we define the other estimators $\bar{y}_t$ and S_t^2 to estimate population mean μ and variance σ^2. We demonstrate, by an example, that the regions of t_1 and t_2 exist for which $\bar{y}_t$ and S_t^2 are better than $\bar{y}$ and S^2; where t_1 and t_2 are pre-determined cutoff points.

2. The estimators $\bar{y}_t$ and S_t^2

Let y_1, y_2, …..,y_n be an independent random sample of size n from a distribution with p.d.f. f(y) and c.d.f. F(y). Let μ and σ^2 be the unknown population mean and variance which are to be estimated. Let us define the estimators $\bar{y}_t$ and S_t^2 to estimate population mean μ and variance σ^2 as

$$\bar{y}_t = \frac{1}{n}\left[\sum_{j=1}^{r_2} y_j + r_1 t_1 + r_3 t_2\right] \tag{2.1}$$

and

$$S_t^2 = \frac{1}{n-1}\left[\sum_{j=1}^{r_2} y_j^2 + r_1 t_1^2 + r_3 t_2^2 - n\bar{y}_t^2\right] \quad \left(t_1 \le y_j \le t_2, j = 1, 2, \ldots . r_2\right) \tag{2.2}$$

Where r_1 is the number of observations with values less than pre-determined cutoff point t_1, r_3 is the number of observations greater than the pre-determined cutoff point t_2 and r_2 is the number of observations which are greater than or equal to t_1 and less than or equal to t_2. Thus total number of observations are divided into three different groups and $r_1 + r_2 + r_3 = n$. We may note that when $r_2 = n$ i.e. $r_1 = 0 = r_3$, $\bar{y}_t$ and S_t^2 are same as the usual unbiased estimators and of the mean and variance respectively.

Let μ_1 and σ_1^2 be the mean and variance of first r_1 observations,

μ_3 and σ_3^2 be the mean and variance of last r_3 observations,

and $\mu_t, \sigma_t^2, \alpha_{3,t}, \alpha_{4,t}$ be the mean, variance, third and fourth moment about origin of the r_2 observations.

Note that the variable which denotes the number of observations in the different groups, follows a trinomial distribution with parameters n, p_1 and p_2 where

$$p_1 = P\left(y_j < t_1\right) = \int_{-\infty}^{t_1} f(y)\,dy \tag{2.3}$$

$$p_2 = P\left(t_1 \le y_j \le t_2\right) = \int_{t_1}^{t_2} f(y)\,dy \tag{2.4}$$

Let $p_3 = P\left(y_j > t_2\right) = \int_{t_2}^{\infty} f(y)\,dy = 1 - \left(p_1 + p_2\right)$ (2.5)

Here, we note that means μ_1, μ_3, μ_t, variances σ_1^2, σ_3^2, σ_t^2 and third and fourth moments about origin are defined as

$$\left.\begin{aligned}
\mu_1 &= E\left(y_j \mid y_j < t_1\right) = E\left(y_j \mid r_1\right) \\
\sigma_1^2 &= Var\left(y_j \mid y_j < t_1\right) = Var\left(y_j \mid r_1\right) \\
\mu_3 &= E\left(y_j \mid y_j > t_2\right) = E\left(y_j \mid r_3\right) \\
\sigma_3^2 &= Var\left(y_j \mid y_j > t_2\right) = Var\left(y_j \mid r_3\right) \\
\mu_t &= E\left(y_j \mid t_1 \le y_j \le t_2\right) = E\left(y_j \mid r_2\right) \\
\sigma_t^2 &= Var\left(y_j \mid t_1 \le y_j \le t_2\right) = Var\left(y_j \mid r_2\right) \\
\alpha_{3,t} &= E\left(y_j^3 \mid t_1 \le y_j \le t_2\right) = E\left(y_j^3 \mid r_2\right) \\
\alpha_{4,t} &= E\left(y_j^4 \mid t_1 \le y_j \le t_2\right) = E\left(y_j^4 \mid r_2\right)
\end{aligned}\right\} \qquad (2.6)$$

So that population mean is $\mu = p_1\mu_1 + p_2\mu_t + p_3\mu_3$ (2.7)

And second moment about origin is

$$\sigma^2 + \mu^2 = p_1\left(\sigma_1^2 + \mu_1^2\right) + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3\left(\sigma_3^2 + \mu_3^2\right) \qquad (2.8)$$

3. Bias and MSE of $\overline{y}_t$

It may be noted that

$$\left.\begin{aligned}
E\left(r_i\right) &= np_i \\
V\left(r_i\right) &= np_i q_i \\
Cov\left(r_i, r_j\right) &= -np_i p_j \;; \qquad q_i = 1 - p_i \qquad \left(i \ne j = 1,2,3\right)
\end{aligned}\right\} \qquad (3.1)$$

The expected value of the estimator $\overline{y}_t$ defined in (2.1) is

$$E\left(\overline{y}_t\right) = \frac{1}{n} E\left[\sum_{j=1}^{r_2} y_j + r_1 t_1 + r_3 t_2\right]$$

$$= E\left(\frac{r_2}{n} E\left[y_j \mid t_1 \le y_j \le t_2\right] + t_1 E\left(\frac{r_1}{n}\right) + t_2 E\left(\frac{r_2}{n}\right)\right)$$

Using the results of (2.6)

$$E\left(\bar{y}_t\right) = p_2\mu_t + t_1 p_1 + t_2 p_3$$

$$= z \quad (\text{say}) \tag{3.2}$$

Therefore bias of estimator $\bar{y}_t$ is

$$\text{Bias}\left(\bar{y}_t\right) = E\left(\bar{y}_t\right) - \mu = p_1\left(t_1 - \mu_1\right) + p_3\left(t_2 - \mu_3\right) \tag{3.3}$$

Now, variance of the estimator $\bar{y}_t$ defined in (2.1) is

$$V\left(\bar{y}_t\right) = V\left[\frac{1}{n}\left(r_1 t_1 + \sum_{j=1}^{r_2} y_j + r_3 t_2\right)\right]$$

$$= \frac{1}{n^2}\left[t_1^2 V\left(r_1\right) + V\left(\sum_{j=1}^{r_2} y_j\right) + t_2^2 V\left(r_3\right) + 2t_1 Cov\left(\sum_{j=1}^{r_2} y_j, r_1\right) + 2t_2 Cov\left(\sum_{j=1}^{r_2} y_j, r_3\right) + 2t_1 t_2 Cov\left(r_1, r_3\right)\right] \tag{3.4}$$

Using the results of (3.1), we get

$$V\left(\sum_{j=1}^{r_2} y_j\right) = V\left(E\left[\sum_{j=1}^{r_2} y_j \mid r_2\right]\right) + E\left(V\left[\sum_{j=1}^{r_2} y_j \mid r_2\right]\right) = np_2\left(\sigma_t^2 + q_2\mu_t^2\right) \tag{3.5}$$

$$Cov\left(\sum_{j=1}^{r_2} y_j, r_1\right) = E\left(r_1\sum_{j=1}^{r_2} y_j\right) - E\left(r_1\right)E\left(\sum_{j=1}^{r_2} y_j\right) = -np_1 p_2 \mu_t \tag{3.6}$$

$$Cov\left(\sum_{j=1}^{r_2} y_j, r_3\right) = -np_2 p_3 \mu_t \tag{3.7}$$

Using (3.5) to (3.7) in (3.4), we get

$$V\left(\bar{y}_t\right) = \frac{1}{n^2}\left[t_1^2 np_1 q_1 + np_2\left(\sigma_t^2 + q_2\mu_t^2\right) + t_2^2 np_3 q_3 - 2t_1 np_1 p_2 \mu_t - 2t_2 np_2 p_3 \mu_t - 2t_1 t_2 np_1 p_3\right] \tag{3.8}$$

$$= \frac{\sigma_*^2}{n} = \frac{\delta\,\sigma^2}{n}$$

Where $\sigma_*^2 = p_2\left(\sigma_t^2 + \mu_t^2\right) + t_1^2 p_1 + t_2^2 p_3 - z^2$ (3.9) and $\delta = \frac{\sigma_*^2}{\sigma^2}$ (3.10)

Now, mean square error of $\bar{y}_t$ is defined as

$$MSE\left(\bar{y}_t\right) = V\left(\bar{y}_t\right) + \left[Bias\left(\bar{y}_t\right)\right]^2$$

Using (3.3) and (3.8), we get

$$MSE\left(\bar{y}_t\right) = \begin{Bmatrix} \frac{1}{n}\left[p_2\sigma_t^2 + \left(p_2q_2\mu_t - 2t_1p_1p_2 - 2t_2p_2p_3\right)\mu_t + t_1^2p_1q_1 + t_2^2p_3q_3 - 2t_1t_2p_1p_3\right] \\ + p_1^2\left(t_1 - \mu_1\right)^2 + p_3^2\left(t_2 - \mu_3\right)^2 + 2p_1p_3\left(t_1 - \mu_1\right)\left(t_2 - \mu_3\right) \end{Bmatrix} \quad (3.11)$$

4. Relative Efficiency of $\bar{y}_t$

The relative efficiency of $\bar{y}_t$ with respect to $\bar{y}$ is given by

$$REF\left(\bar{y}_t\right) = \frac{Var\left(\bar{y}\right)}{MSE\left(\bar{y}_t\right)}$$

Using (3.8) and $Var\left(\bar{y}\right) = \frac{\sigma^2}{n}$, we get

$$REF\left(\bar{y}_t\right) = \frac{\sigma^2}{\sigma_*^2 + n\left\{p_1\left(t_1 - \mu_1\right) + p_3\left(t_2 - \mu_3\right)\right\}^2} \quad (4.1)$$

In this paper, we do not find the optimum values of t_1 and t_2 for which $REF\left(\bar{y}_t\right)$ is maximum but we can obtain them by solving $\frac{\partial}{\partial t_1} MSE\left(\bar{y}_t\right) = 0$ and $\frac{\partial}{\partial t_2} MSE\left(\bar{y}_t\right) = 0.$

5. An Example

The development presented in this paper is a general one. For any particular known underlying distribution further improvements on $\bar{y}_t$ could probably be found. Hence example presented in this section should be considered as providing guide lines for similar situations.

Let us take an example to demonstrate the gain, that can be achieved by using $\bar{y}_t = \frac{1}{n}\left[\sum_{j=1}^{r_2} y_j + r_1 t_1 + r_3 t_2\right]$ as an estimator of the mean over the usual unbiased estimator $\bar{y}$, by considering doubly truncated exponential distribution, truncated at t_1 on left and at t_2 on right whose pdf is given by

$$g(y) = \begin{cases} \frac{1}{p_2} \cdot \frac{1}{\theta} e^{-\frac{y}{\theta}} & ; t_1 \le y \le t_2 \\ = 0 & ; \text{otherwise} \end{cases} \tag{5.1}$$

Where $$p_2 = P(t_1 \le y \le t_2) = e^{-\frac{t_1}{\theta}} - e^{-\frac{t_2}{\theta}} \tag{5.2}$$

Now the mean square of the estimator $\bar{y}_t$ is defined as

$$MSE(\bar{y}_t) = V(\bar{y}_t) + \left[Bias(\bar{y}_t)\right]^2$$

Based on the results (A2.8) and (A2.9) of Appendix – 2 we find

$$MSE(\bar{y}_t) = \frac{\theta^2}{n}\left[(n-1)\left(p_2 + \frac{t_1}{\theta}\right)^2 - 2(n-1)p_2 - 2(n-1)\frac{t_1}{\theta} + \left(\frac{t_1}{\theta}\right)^2 - 2\left(\frac{t_2}{\theta}p_3 + \frac{t_1}{\theta}p_1\right) + n\right] \tag{5.3}$$

Relative efficiency of $\bar{y}_t$ compared to $\bar{y}$ is

$$REF(\bar{y}_t) = \frac{Var(\bar{y})}{MSE(\bar{y}_t)}$$

$$= \frac{1}{(n-1)\left(p_2 + \frac{t_1}{\theta}\right)\left(p_2 + \frac{t_1}{\theta} - 2\right) + \left(\frac{t_1}{\theta}\right)^2 - 2\left(\frac{t_2}{\theta}p_3 + \frac{t_1}{\theta}p_1\right) + n} \tag{5.4}$$

Note that, the efficiency of $\bar{y}_t$ relative to unbiased estimator $\bar{y}$ is a function of parameters n, $\frac{t_1}{\theta}$ & $\frac{t_2}{\theta}$

Table 1: Relative Efficiency for the estimator $\bar{y}_t$ of mean of two sided truncated exponential distribution (n = 5)

$\frac{t_1}{\theta}$ \ $\frac{t_2}{\theta}$	8	16	20	30	40	70	99
0.01	1.0055	1.0001	1.0001	1.0001	1.0001	1.0001	1.0001
0.05	1.0078	1.0024	1.0024	1.0024	1.0024	1.0024	1.0024
0.10	1.0145	1.0090	1.0090	1.0090	1.0090	1.0090	1.0090
0.50	1.1157	1.1087	1.1087	1.1087	1.1087	1.1087	1.1087
1.00	0.7869	0.7830	0.7830	0.7830	0.7830	0.7830	0.7830
3.00	0.0474	0.0474	0.0474	0.0474	0.0474	0.0474	0.0474

Table 2: Relative Efficiency for the estimator of mean of two sided truncated exponential distribution (n = 10)

$\frac{t_1}{\theta}$ \ $\frac{t_2}{\theta}$	8	16	20	30	40	70	99
0.01	1.0055	1.0001	1.0001	1.0001	1.0001	1.0001	1.0001
0.05	1.0078	1.0024	1.0024	1.0024	1.0024	1.0024	1.0024
0.10	1.0144	1.0090	1.0090	1.0090	1.0090	1.0090	1.0090
0.50	1.0500	1.0431	1.0431	1.0431	1.0431	1.0431	1.0431
1.00	0.5738	0.5118	0.5118	0.5118	0.5118	0.5118	0.5118
3.00	0.0238	0.0237	0.0237	0.0237	0.0237	0.0237	0.0237

Table 3: Relative Efficiency for the estimator of mean of two sided truncated exponential distribution (n = 15)

t_1/θ \ t_2/θ	8	16	20	30	40	70	99
0.01	1.0055	1.0001	1.0001	1.0001	1.0001	1.0001	1.0001
0.05	1.0078	1.0024	1.0024	1.0024	1.0024	1.0024	1.0024
0.10	1.0143	1.0088	1.0088	1.0088	1.0088	1.0088	1.0088
0.50	0.9910	0.9848	0.9848	0.9848	0.9848	0.9848	0.9848
1.00	0.3814	0.3802	0.3802	0.3802	0.3802	0.3802	0.3802
3.00	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158	0.0158

From tables 1 to 3, it can be observed that gains are achieved for extremely wide choices of t_1 and t_2. In fact, for a sample of size five or ten the gains are achieved for t_1 values anywhere between zero and half of the true mean and for t_2 values, between eight times the true mean and infinity. The gains become modest for large sample sizes and values of t_1 and t_2.

6. Bias and MSE of S_t^2.

Here we derive bias and MSE of $S_t^2 = \frac{1}{n-1}\left[\sum_{j=1}^{r_2} y_j^2 + r_1 t_1^2 + r_3 t_2^2 - n\bar{y}_t^2\right]$

Bias of S_t^2 is defined as

$$\text{Bias}\left(S_t^2\right) = E\left(S_t^2\right) - \sigma^2 \tag{6.1}$$

Now, $E\left[(n-1)S_t^2\right] = E\left(r_1 t_1^2 + \sum_{j=1}^{r_2} y_j^2 + r_3 t_2^2 - n\bar{y}_t^2\right)$

$$= t_1^2 E\left(r_1\right) + E\left(\sum_{j=1}^{r_2} y_j^2\right) + t_2^2 E\left(r_3\right) - nE\left(\bar{y}_t^2\right) \tag{6.2}$$

$$E\left(\sum_{j=1}^{r_2} y_j^2\right) = E\left[E\left(\sum_{j=1}^{r_2} y_j^2 \mid t_1 \le y_j \le t_2\right)\right] = E\left[\sum_{j=1}^{r_2} E\left(y_j^2 \mid r_2\right)\right] = np_2\left(\sigma_t^2 + \mu_t^2\right) \tag{6.3}$$

$$E\left(\bar{y}_t^2\right) = V\left(\bar{y}_t\right) + \left[E\left(\bar{y}_t\right)\right]^2$$

$$= \frac{1}{n}\left[p_2\left(\sigma_t^2 + \mu_t^2\right) + t_1^2 p_1 + t_2^2 p_3 + (n-1)z^2\right], \tag{6.4}$$

Using (6.3), (6.4), (3.1) in (6.2), we get

$$E\left[(n-1)S_t^2\right] = (n-1)\sigma_*^2 \text{ or } E\left(s_t^2\right) = \sigma_*^2 \tag{6.5}$$

$\therefore$ Bias of S_t^2 is

$$\text{Bias}\left(S_t^2\right) = E\left(S_t^2\right) - \sigma^2 = \sigma_*^2 - \sigma^2 = (\delta - 1)\sigma^2 \tag{6.6}$$

From (2.8), (3.10) and (6.6), we get

$$\text{Bias}\left(S_t^2\right) = \mu^2 - p_1\left(\sigma_1^2 + \mu_1^2\right) - p_3\left(\sigma_3^2 + \mu_3^2\right) + t_1^2 p_1 + t_2^2 p_3 - z^2 \tag{6.7}$$

Mean square error of S_t^2 is defined as

$$\text{MSE}\left(S_t^2\right) = E\left(S_t^2 - \sigma^2\right)^2$$

$$= E\left(S_t^4\right) - 2\sigma^2 E\left(S_t^2\right) + \sigma^4 \tag{6.8}$$

$$\text{Now, } E\left(S_t^4\right) = \frac{1}{(n-1)^2} E\left(r_1 t_1^2 + \sum_{j=1}^{r_2} y_j^2 + r_3 t_2^2 - n\bar{y}_t^2\right)^2$$

$$= \frac{1}{(n-1)^2}\left[\begin{array}{l} E\left(\sum_{j=1}^{r_2} y_j^2\right)^2 + n^2 E\left(\bar{y}_t^2\right)^2 + t_1^4 E\left(r_1^2\right) + t_2^4 E\left(r_3^2\right) + 2t_1^2 t_2^2 E\left(r_1 r_3\right) \\ +2t_1^2 E\left(r_1 \sum_{j=1}^{r_2} y_j^2\right) + 2t_2^2 E\left(r_3 \sum_{j=1}^{r_2} y_j^2\right) - 2nt_1^2 E\left(r_1 \bar{y}_t^2\right) - 2nt_2^2 E\left(r_3 \bar{y}_t^2\right) - 2nE\left(\bar{y}_t^2 \sum_{j=1}^{r_2} y_j^2\right) \end{array}\right] \tag{6.9}$$

Using the results (A1.1) to (A1.10) of Appendix - 1 in (6.9), we get

$$E\left(S_t^4\right) = \left\{\begin{array}{l} \frac{1}{n}\left[\begin{array}{l} \left(p_1 t_1^4 + p_2 \alpha_{4,t} + p_3 t_2^4\right) - 4\left(p_1 t_1^3 + p_2 \alpha_{3,t} + p_3 t_2^3\right)z \\ +6\left\{p_1 t_1^2 + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3 t_2^2\right\}z^2 - 3z^4 \end{array}\right] \\ +\frac{n^2 - 2n + 3}{n(n-1)}\left[p_1 t_1^2 + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3 t_2^2 - z^2\right]^2 \end{array}\right\} \tag{6.10}$$

If μ_4^* and σ_*^2 are the fourth central moment and variance respectively of the population truncated at the point t_1 on the left and at the point t_2 on the right, then

$$\left.\begin{aligned}\mu_4^* &= \left(p_1t_1^4+p_2\alpha_{4,t}+p_3t_2^4\right)-4\left(p_1t_1^3+p_2\alpha_{3,t}+p_3t_2^3\right)z+6\left\{p_1t_1^2+p_2\left(\sigma_t^2+\mu_t^2\right)+p_3t_2^2\right\}z^2-3z^4\\ \text{and } \sigma_*^2 &= p_1t_1^2+p_2\left(\sigma_t^2+\mu_t^2\right)+p_3t_2^2-z^2\end{aligned}\right\} \quad (6.11)$$

From (6.10) and (6.11), we get

$$E\left(s_t^4\right)=\frac{\mu_4^*}{n}+\frac{n^2-2n+3}{n(n-1)}\sigma_*^4 \quad (6.12)$$

Using (6.5) and (6.11) in (6.8), we get

$$MSE\left(S_t^2\right)=\left[\frac{\mu_4^*}{n}+\frac{3-n}{n(n-1)}\sigma_*^4\right]+\left(\sigma_*^2-\sigma^2\right)^2 \quad (6.13)$$

It may be noted that when t_1 approaches to the lower limit and t_2 approaches to the upper limit of the distribution,

$p_1 \to 0,\ p_3 \to 0,\ p_2 \to 1,\ \mu_4^* \to \mu_4,\ \sigma_*^2 \to \sigma^2$ and

$$MSE\left(S_t^2\right) \to Var\left(S^2\right)=\frac{\mu_4}{n}+\frac{3-n}{n(n-1)}\sigma^4$$

7. Relative Efficiency of s_t^2

The relative efficiency of the proposed estimator S_t^2 with respect to usual unbiased estimator S^2 is defined as

$$REF\left(S_t^2\right)=\frac{Var\left(S^2\right)}{MSE\left(S_t^2\right)}=\frac{\dfrac{\beta_2}{n}+\dfrac{3-n}{n(n-1)}}{\left[\dfrac{\beta_2^*}{n}+\dfrac{3-n}{n(n-1)}\right]\delta^2+(\delta-1)^2} \quad (7.1)$$

where $\beta_2=\dfrac{\mu_4}{\sigma^4}$ and $\beta_2^*=\dfrac{\mu_4^*}{\sigma_*^4}$

The estimator S_t^2 is better than the estimator S^2 of ó² if $REF\left(S_t^2\right)\geq 1$ which gives

$$\delta^2\ \beta_2^* \le \beta_2 - n(1-\delta)^2 - \frac{n-3}{n-1}(1-\delta^2) \tag{7.2}$$

It may be noted that the optimum values of t_1 and t_2, for which $REF(S_t^2)$ is maximum, can be obtained by solving the equations $\frac{\partial}{\partial t_1} MSE(S_t^2) = 0$ and $\frac{\partial}{\partial t_2} MSE(S_t^2) = 0$

8. An Example

We want to study the range of t_1 and t_2 for which $REF(S_t^2) \ge 1$.

To observe the gain which is obtained by using the estimator S_t^2 over the usual unbiased estimator S^2 of σ^2 here we consider a doubly truncated normal distribution at cutoff point t_1 on the left and at t_2 on the right, with pdf

$$g(y) = \frac{1}{p_2}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} \quad ; t_1 < y < t_2,\ \ \sigma>0 \tag{8.1}$$

Where $p_2 = \int_{t_1}^{t_2} f(y,\sigma)\,dy = \Phi(t_2') - \Phi(t_1')$ and $\Phi(\cdot)$ is the distribution function of a standard normal variate.

We know that for the usual unbiased estimator S^2 of variance ó^2 of a normal distribution,

$$V(S^2) = \frac{2\sigma^4}{n-1} \tag{8.2}$$

Using the results (8.2) and (A3.13) of Appendix -3 in (7.1) we get

$$REF(S_t^2) = \frac{\frac{2}{(n-1)}}{\frac{1}{n}\left[\frac{D'-4C'A'+6B'A'^2-3A'^4}{(B'-A'^2)^2} + \frac{3-n}{n-1}\right](B'-A')^2 + (B'-A'^2-1)^2} \tag{8.3}$$

A', B', C' , and D' of (A3.7) are the functions of t_1' and t_2' and so the efficiency of the estimator S_t^2 relative to usual unbiased estimator S^2 is a function of parameter n, $\frac{t_1}{\sigma}$ and $\frac{t_2}{\sigma}$.

Table 4. Relative Efficiency for Variance of two sided truncated Normal Distribution (n = 5)

t_1' \ t_2'	0.1	0.4	0.7	1.0	1.3	1.6	1.9	2.2	2.5	2.8	3.1
- 0.1	.5002	.4937	.4962	.5289	.5979	.6989	.8171	.9310	1.0236	1.0897	1.1232
- 0.4	.4937	.5165	.5421	.5964	.6902	.8193	.9645	1.0983	1.2018	1.2707	1.3072
- 0.7	.4962	.5421	.5881	.6634	.7818	.9377	1.1047	1.2491	1.3527	1.4137	1.4492
- 1.0	.5289	.5964	.6634	.7630	.9103	1.0940	1.2755	1.4149	1.5017	1.5398	1.5694
- 1.3	.5979	.6902	.7818	.9103	1.0887	1.2908	1.4597	1.5592	1.6008	1.5970	1.6145
- 1.6	.6989	.8193	.9377	1.0940	1.2908	1.4760	1.5826	1.6036	1.5822	1.5339	1.5370
- 1.9	.8171	.9645	1.1047	1.2755	1.4597	1.5826	1.5957	1.5373	1.4679	1.3935	1.3871
- 2.2	.9310	1.0983	1.2491	1.4149	1.5592	1.6036	1.5373	1.4299	1.3378	1.2553	1.2447
- 2.5	1.0236	1.2018	1.3527	1.5017	1.6008	1.5822	1.4679	1.3378	1.2382	1.1554	1.1433
- 2.8	1.0897	1.2707	1.4137	1.5398	1.5970	1.5339	1.3935	1.2553	1.1554	1.0755	1.0631
- 2.1	1.1232	1.3072	1.4492	1.5694	1.6145	1.5371	1.3871	1.2446	1.1433	1.0631	1.0505

Table 5. Relative Efficiency for Variance of two sided truncated Normal Distribution (n = 10)

t_1' \ t_2'	0.1	0.4	0.7	1.0	1.3	1.6	1.9	2.2	2.5	2.8	3.1
- 0.1	.2223	.2194	.2205	.2351	.2663	.3132	.3704	.4278	.4761	.5122	.5292
- 0.4	.2194	.2296	.2410	.2655	.3090	.3717	.4471	.5224	.5850	.6312	.6528
- 0.7	.2205	.2410	.2617	.2962	.3525	.4321	.5268	.6200	.6958	.7501	.7757
- 1.0	.2351	.2655	.2962	.3431	.4170	.5200	.6404	.7555	.8455	.9057	.9352
- 1.3	.2663	.3090	.3525	.4170	.5164	.6523	.8052	.9409	1.0375	1.0925	1.1229
- 1.6	.3132	.3717	.4321	.5200	.6523	.8252	1.0015	1.1343	1.2100	1.2357	1.2588
- 1.9	.3704	.4471	.5268	.6404	.8052	1.0015	1.1675	1.2562	1.2818	1.2643	1.2742
- 2.2	.4278	.5224	.6200	.7555	.9409	1.1343	1.2562	1.2825	1.2590	1.2094	1.2087
- 2.5	.4761	.5850	.6958	.8455	1.0325	1.2100	1.2818	1.2590	1.2051	1.1398	1.1335
- 2.8	.5122	.6312	.7501	.9057	1.0925	1.2357	1.2643	1.2094	1.1398	1.0689	1.0597
- 2.1	.5292	.6528	.7757	.9352	1.1229	1.2588	1.2742	1.2087	1.1335	1.0597	1.0497

Table 6. Relative Efficiency for Variance of two sided truncated Normal Distribution (n = 15)

t_1' \ t_2'	0.1	0.4	0.7	1.0	1.3	1.6	1.9	2.2	2.5	2.8	3.1
- 0.1	.1429	.1411	.1418	.1511	.1713	.2018	.2394	.2776	.3099	.3344	.3456
- 0.4	.1411	.1476	.1549	.1708	.1990	.2403	.2909	.3424	.3861	.4192	.4342
- 0.7	.1418	.1549	.1683	.1906	.2275	.2806	.3456	.4118	.4676	.5095	.5284
- 1.0	.1511	.1708	.1906	.2213	.2703	.3407	.4270	.5144	.5870	.6400	.6641
- 1.3	.1713	.1990	.2275	.2703	.3382	.4359	.5548	.6721	.7652	.8278	.8579
- 1.6	.2018	.2403	.2806	.3407	.4359	.5716	.7307	.8748	.9762	1.0316	1.0620
- 1.9	.2394	.2909	.3456	.4270	.5548	.7307	.9175	1.0582	1.1337	1.1542	1.1746
- 2.2	.2776	.3424	.4118	.5144	.6721	.8748	1.0582	1.1588	1.1856	1.1650	1.1723
- 2.5	.3099	.3801	.4676	.5870	.7652	.9762	1.1337	1.1856	1.1713	1.1239	1.1223
- 2.8	.3344	.4192	.5095	.6400	.8278	1.0316	1.1542	1.1650	1.1239	1.0632	1.0566
- 2.1	.3356	.4342	.5284	.6641	.8579	1.0620	1.1746	1.1723	1.1223	1.0566	1.0486

From the Table 4, we observe that, when n = 5, $t_1' \leq 1.3$ and $t_2' \geq 1.3$, the estimator S_t^2 is better than S^2 Similarly, from the Table 5 and Table 6, we can get the range of t_1' and t_2' for which S_t^2 is better than S^2 From Table 4 to Table 6, we observe that if other parameters remain constant, the relative efficiency REF(S_t^2) increase as n decreases.

9. Concluding Remarks

The estimators $\overline{y}_t$ and S_t^2 of mean and variance proposed in this paper are defined to reduce the effect of large true observations. It has been shown that under certain circumstances, these estimators have a smaller mean square error than the usual estimators $\overline{y}$ and S^2 of the population mean and variance respectively. The estimators proposed in this paper differ from others in the information external to the sample is utilized to predetermined cutoff points t_1 and t_2 which separate 'large' sample observations from the rest. The only external information required is a rough idea of the general range in which the data lie. The examples considered and the tables prepared reveal that this information need not be very exact for the gain to be achieved.

10. Acknowledgement

I thank the referee for reviewing this paper which has helped in revising the earlier draft of the paper.

11.1 Appendix – 1

Let $y_1, y_2, \ldots, y_n$ be an independent random sample of size n from a distribution with p.d.f. f(y) and c.d.f. F(y). Let μ and ó² be the unknown population mean and variance which are to be estimated. Let r_1 be the number of observations with values less than pre-determined cutoff point t_1, r_3 be the number of observations greater than the pre-determined cutoff point t_2 and r_2 be the number of observations which are greater than or equal to t_1 and less than or equal to t_2. Thus total number of observations are divided into three different groups and $r_1 + r_2 + r_3 = n$. In this regard the following results are obvious:

$$E\left(\sum_{j=1}^{r_2} y_j^2\right)^2 = E\left[E\left\{\left(\sum_{j=1}^{r_2} y_j^4 + \sum\sum_{j\neq k} y_j^2 y_k^2\right) | r_2\right\}\right]$$

$$= np_2\alpha_{4,t} + n(n-1)p_2^2\left(\sigma_t^2 + \mu_t^2\right)^2 \qquad \text{(A1.1)}$$

$$E\left(\bar{y}_t^2\right)^2 = \frac{1}{n^4}E\left[\left(r_1t_1 + r_3t_2\right) + \sum_{j=1}^{r_2} y_j\right]^4$$

$$= \frac{1}{n^4}\left\{\begin{array}{l} E\left(\sum_{j=1}^{r_2} y_j\right)^4 + 4E\left[\left(r_1t_1 + r_3t_2\right)\left(\sum_{j=1}^{r_2} y_j\right)^3\right] + 6E\left[\left(r_1t_1 + r_3t_2\right)^2\left(\sum_{j=1}^{r_2} y_j\right)^2\right] \\ +4E\left[\left(r_1t_1 + r_3t_2\right)^2\left(\sum_{j=1}^{r_2} y_j\right)\right] + E\left(r_1t_1 + r_3t_2\right)^4 \end{array}\right\}$$

$$= \frac{1}{n^4}\left[\begin{array}{l} n\left(p_1t_1^4 + p_2\alpha_{4,t} + p_3t_2^4\right) + 4n(n-1)\left(p_1t_1^3 + p_2\alpha_{3,t} + p_3t_2^3\right)\cdot Z \\ +6n(n-1)(n-2)\left\{p_1t_1^2 + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3t_2^2\right\}z^2 \\ +n(n-1)(n-2)(n-3)z^4 + 3n(n-1)\left\{p_1t_1^2 + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3t_2^2\right\}^2 \end{array}\right] \qquad \text{(A1.2)}$$

$$E\left(r_1^2\right) = np_1\left[1 + (n-1)p_1\right] \qquad \text{(A1.3)}$$

$$E\left(r_3^2\right) = np_3\left[1 + (n-1)p_3\right] \qquad \text{(A1.4)}$$

$$E\left(r_1 r_3\right) = n\left(n-1\right)p_1 p_3 \tag{A1.5}$$

$$E\left(r_1 \sum_{j=1}^{r_2} y_j^2\right) = n\left(n-1\right)p_1 p_2\left(\sigma_t^2+\mu_t^2\right) \tag{A1.6}$$

$$E\left(r_3 \sum_{j=1}^{r_2} y_j^2\right) = n\left(n-1\right)p_2 p_3\left(\sigma_t^2+\mu_t^2\right) \tag{A1.7}$$

$$E\left(r_1 \bar{y}_t^{2}\right) = \frac{1}{n^2}\begin{bmatrix} n\left(n-1\right)p_1\left\{p_1 t_1^2 + p_2\left(\sigma_t^2+\mu_t^2\right)+p_3 t_2^2\right\}+2n\left(n-1\right)p_1 t_1 z \\ +n\left(n-1\right)\left(n-2\right)p_1 z^2 + n_1 p_1 t_1^2 \end{bmatrix} \tag{A1.8}$$

$$E\left(r_3 \bar{y}_t^{2}\right) = \frac{1}{n^2}\begin{bmatrix} n\left(n-1\right)p_3\left\{p_1 t_1^2 + p_2\left(\sigma_t^2+\mu_t^2\right)+p_3 t_2^2\right\}+2n\left(n-1\right)p_3 t_2 z \\ +n\left(n-1\right)\left(n-2\right)p_3 z^2 + n\ p_3 t_2^2 \end{bmatrix} \tag{A1.9}$$

$$E\left(\bar{y}_t^{2}\sum_{j=1}^{r_2} y_j^2\right) = \frac{1}{n^2}E\left[\left(r_1 t_1 + \sum_{j=1}^{r_2} y_j + r_3 t_2\right)^2\left(\sum_{j=1}^{r_2} y_j^2\right)\right]$$

$$= \frac{1}{n^2}\begin{bmatrix} t_1^2 E\left(r_1^2 \sum_{j=1}^{r_2} y_j^2\right) + E\left\{\left(\sum_{j=1}^{r_2} y_j\right)^2\left(\sum_{j=1}^{r_2} y_j^2\right)\right\} + t_2^2 E\left(r_3^2\sum_{j=1}^{r_2} y_j^2\right) \\ +2t_1 E\left\{r_1\left(\sum_{j=1}^{r_2} y_j\right)\left(\sum_{j=1}^{r_2} y_j^2\right)\right\} + 2t_2 E\left\{r_3\left(\sum_{j=1}^{r_2} y_j\right)\left(\sum_{j=1}^{r_2} y_j^2\right)\right\} + 2t_1 t_2 E\left\{r_1 r_3\left(\sum_{j=1}^{r_2} y_j^2\right)\right\} \end{bmatrix}$$

$$= \frac{1}{n^2}\begin{bmatrix} np_2\alpha_{4,t} + 2n\left(n-1\right)p_2\alpha_{3,t} z + n\left(n-1\right)p_2\left(\sigma_t^2+\mu_t^2\right)\left\{p_1 t_1^2 + p_2\left(\sigma_t^2+\mu_t^2\right)+p_3 t_2^2\right\} \\ +n\left(n-1\right)\left(n-2\right)p_2\left(\sigma_t^2+\mu_t^2\right)z^2 \end{bmatrix} \tag{A1.10}$$

11.2 Appendix – 2

The p.d.f. of exponential distribution is

$$f\left(y\,|\,\theta\right) = \frac{1}{\theta}e^{-\frac{y}{\theta}} \quad ; \quad y>0,\ \ \theta>0 \tag{A2.1}$$

Hence, in this case using (2.3) to (2.6), we get

$$p_1 = P(y < t_1) = 1 - e^{-\frac{t_1}{\theta}} \quad \text{(A2.2)}$$

$$p_2 = P(t_1 \le y \le t_2) = e^{-\frac{t_1}{\theta}} - e^{-\frac{t_2}{\theta}} \quad \text{(A2.3)}$$

$$p_3 = P(y > t_2) = e^{-\frac{t_2}{\theta}} \quad \text{(A2.4)}$$

$$p_2\mu_t = \frac{1}{\theta}\int_{t_1}^{t_2} y e^{-\frac{y}{\theta}} dy = \left(t_1 e^{-\frac{t_1}{\theta}} - t_2 e^{-\frac{t_2}{\theta}} \right) + p_2\theta \quad \text{(A2.5)}$$

$$z = \theta\left(p_2 + \frac{t_1}{\theta} \right) \quad \text{(A2.6)}$$

$$p_2\left(\sigma_t^2 + \mu_t^2\right) + p_1 t_1^2 + p_3 t_2^2 = \theta^2\left[2p_2 - \frac{2p_3 t_2}{\theta} + \frac{2(1-p_1)t_1}{\theta} + \left(\frac{t_1}{\theta}\right)^2 \right] \quad \text{(A2.7)}$$

Using the results (A2.2) to (A2.7) in (3.2) and (3.8) we get bias of $\bar{y}_t$ and $V(\bar{y}_t)$ as

$$\text{Bias}(\bar{y}_t) = E(\bar{y}_t) - \theta = (p_2\mu_t + t_1 p_1 + t_2 p_3) - \theta = \theta\left(t_2 + \frac{t_1}{\theta} - 1 \right) \quad \text{(A2.8)}$$

$$V(\bar{y}_t) = \frac{\theta^2}{n}\left[2p_2 - 2p_3\frac{t_2}{\theta} + 2(1-p_1)\frac{t_1}{\theta} + \left(\frac{t_1}{\theta}\right)^2 - \left(p_2 + \frac{t_1}{\theta}\right)^2 \right] \quad \text{(A2.9)}$$

11.3 Appendix – 3

Consider a normal distribution with mean zero and variance σ^2, as an underlying distribution, to observe the gain which is obtained by using the estimator s_t^2 over the usual unbiased estimator S^2 of σ^2. The probability density function is

$$f(y,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} \quad ; -\infty < y < \infty, \ \sigma > 0 \quad \text{(A3.1)}$$

Here $p_1 = \int_{-\infty}^{t_1} f(y,\sigma)\, dy = \Phi\left(t_1'\right)$

$$p_2 = \int_{t_1}^{t_2} f(y,\sigma)\,dy = \Phi\left(t_2^{'}\right) - \Phi\left(t_1^{'}\right)$$

and $$p_3 = \int_{t_2}^{\infty} f(y,\sigma)\,dy = \Phi\left(-t_2^{'}\right) \qquad \text{(A3.2)}$$

where $t_1^{'} = \frac{t_1}{\sigma}$, $t_2^{'} = \frac{t_2}{\sigma}$ and $\Phi(\cdot)$ is the distribution function of a standard normal variate.

Then the p.d.f. of the normal distribution truncated is given by

$$g(y) = \frac{1}{p_2}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} \quad ; t_1 < y < t_2, \ \sigma > 0$$

$$\therefore p_1 t_1 + p_2 \mu_t + t_3 t_2 = \sigma A' \qquad \text{(A3.3)}$$

$$p_1 t_1^2 + p_2\left(\sigma_t^2 + \mu_t^2\right) + p_3 t_2^2 = \sigma^2 B' \qquad \text{(A3.4)}$$

$$p_1 t_1^3 + p_2 \alpha_{3,t} + p_3 t_2^3 = \sigma^3 C' \qquad \text{(A3.5)}$$

and $$p_1 t_1^4 + p_2 \alpha_{4,t} + p_3 t_2^4 = \sigma^4 D' \qquad \text{(A3.6)}$$

where $$A' = \left\{\phi\left(t_1^{'}\right) - \phi\left(t_2^{'}\right)\right\} + p_1 t_1^{'} + p_3 t_2^{'}$$

$$B' = \left\{t_1^{'}\phi\left(t_1^{'}\right) - t_2^{'}\phi\left(t_2^{'}\right)\right\} + p_2 + p_1 t_1^{'2} + p_3 t_2^{'2}$$

$$C' = \left\{t_1^{'2}\,\phi\left(t_1^{'}\right) - t_2^{'}\phi\left(t_2^{'}\right)\right\} + 2\left\{\phi\left(t_1^{'}\right) - \phi\left(t_2^{'}\right)\right\} + p_1 t_1^{'3} + p_3 t_2^{'3}$$

$$D' = \left\{t_1^{'3}\,\phi\left(t_1^{'}\right) - t_2^{'3}\,\phi\left(t_2^{'}\right)\right\} + 3\left\{t_1^{'}\phi\left(t_1^{'}\right) - t_2^{'}\phi\left(t_2^{'}\right)\right\} + 3p_2 + p_1 t_1^{'4} + p_3 t_2^{'4} \qquad \text{(A3.7)}$$

$\text{Bias}\left(S_t^2\right) = (\delta - 1)\sigma^2$ where (A3.8)

$$\delta = \left\{t_1^{'}\phi\left(t_1^{'}\right) - t_2^{'}\phi\left(t_2^{'}\right)\right\} + p_2 + \left(p_1 t_1^{'2} + p_3 t_2^{'2}\right) - \left[\left\{\phi\left(t_1^{'}\right) - \phi\left(t_2^{'}\right)\right\} + p_1 t_1^{'} + p_3 t_2^{'}\right]^2$$

$$= B' - A'^2 \qquad \text{(A3.9)}$$

$$\sigma_*^2 = \sigma^2\left(B' - A'^{\,2}\right) \tag{A3.10}$$

$$\mu_4^* = \sigma^4\left(D' - 4C'A' + 6B'A'^{\,2} - 3A'^{\,4}\right) \tag{A3.11}$$

$$\beta_2^* = \frac{\mu_4^*}{\sigma_*^4} = \frac{\left(D' - 4C'A' + 6B'A'^{\,2} - 3A'^{\,4}\right)}{\left(B' - A'^{\,2}\right)^2} \tag{A3.12}$$

Therefore, MSE $\left(S_t^2\right)$ can be obtained using the results (8.10), (8.13) and from (7.1) as

$$\text{MSE} = \frac{1}{n}\left[\frac{D' - 4C'A' + 6B'A'^{\,2} - 3A'^{\,4}}{\left(B' - A'^{\,2}\right)^2} + \frac{3-n}{n-1}\right]\left(B' - A'\right)^2 + \left(B' - A'^{\,2} - 1\right)^2 \tag{A3.13}$$

12. REFERENCES

[1] **Ojha, V.P. (1980),** 'Some Contribution to the Statistical Inference under conditional Specifications', *Unpublished Ph. D. Thesis*, *Uni. Of Gorakhpur, Gorakhpur (India)*.

[2] **Searls, D. T. (1966),** 'An estimator for a population mean which reduces the effect of large true observations', *Journal Amer. Stat. Asso*, Vol. 61, No. 316, pp. 1200-1204.

REVIEW ARTICLE

A BRIEF REVIEW FOR THE ROLE OF GOVERNMENT IN SOCIAL SECTOR DEVELOPMENT

Sanjay G Raval[1] & Mahesh Vaghela[2]

ABSTRACT

Social Sector is very useful and important for the growth of the economic system. The development aspects pertaining to social sector can enhance the economic growth of the country.

In this paper it is intended to visualize governmental role for economic development in which some specific aspects pertaining to study are reviewed in brief.

KEYWORDS

HDI, GEM, UNDP

1. INTRODUTION

The importance of social development is now a well recognized fact. However the problem is complex in nature not only from the point of view of resource availability in developing countries like India, but also because of the fact that it involves radical changes in social, political, cultural and other institutional factors which act as blocks to development. Increasingly there is academic interest in studying the problem of social development. So also even traditionally the State has been committing bulk of its resources in the social sector. The size, the pattern and its accomplishments were also subject to review

(1) Head of Statistics Department, Som–lalit College of Commerce, Gujarat University. E-mail: drsgraval@gamail.com

(2) Head of Statistics Department, N.C. Bodiwala Commerce College, Ahmedabad Gujarat University. E-mail: Mahesh.vaghela72@yahoo.com

(Rcd.Oct-18/rvd Dec-18)

by the scholars. Some of such works are reviewed below. In this study the available literature is grouped into two categories namely, (1) Studies made for social development, (2) Role of Government in social development.

This brief review intends two specify different areas in which research work has been carried out pertaining to our study on social sector development.

2. Studay on Social sector development

Amartya Sen (1989) discussed the nature and implications of the task of identification of ends, in terms of which the means of development can be systematically assessed. He argued for the necessity of seeing human beings as ends in themselves, rather than as means to other ends of development. According to him the expansion of real income and economic growth should be just an intermediate goal, the importance of which is contingent on what it ultimately contributes to human lives. For him the foundational importance of human capabilities provides a firm basis for evaluating living standards and quality of life and also points to a general format in terms of which problems of efficiency and equality can be discussed. It is concluded that the challenge of human development cannot be fully grasped without paying deliberate attention to the enhancement of freedoms and capabilities that matter most in the lives that we can lead.

Newman and Thomson (1989) investigated the controversy surrounding the interrelation between economic growth and social development using a lagged-dependent variable model for 46 developing countries for the years 1960, 1970 and 1980. They provided a test of the two competing hypotheses; 1) Economic growth is a cause of subsequent social development 2) Social development is the cause of subsequent economic development. The findings of the study suggest that the level of social development in a country is substantially a product of its earlier social indicators. Economic indicators do not significantly predict subsequent social development. On the other hand, economic growth appears to be a product of earlier economic and social indicators. Hence social development policies may be justified on the basis of both expected social and economic benefits, at least in the long run, while economic growth policies should be justified on the basis of economic benefits alone.

UNDP (1990) defined human development as “a process of enlarging people’s choice”. The three essential choices are, to lead a long and healthy life, to acquire knowledge and to have access to the resources needed for a decent standard of living. In this report UNDP has attempted to divert attention from the exclusive use of income measures to assess the development levels and gave ‘a human face’ to development. The most important contribution of the report is the construction of a composite indicator, the human development index (HDI) which assesses achieved development levels of countries. The countries are classified as: low human development (HDI<0.5); medium human development (HDI 0.5 to 0.79) and high human development (HDI 0.8 to 1). HDI is increasingly used to monitor the progress of nations and of global society.

Mark Gillivray (1991) examined both the composition and usefulness of HDI as a composite development indicator using simple and rank correlation coefficients. On the basis of the correlation analysis it is argued that the composition of the index is flawed as it is significantly and positively correlated with each of its component variables individually. As a consequence, assessing inter country development levels on any of these variables yields similar results to those that the index itself yields. According to him the index provides us with little more information regarding inter country development levels than the traditional indicator GNP per capita alone provides. It is concluded that the UNDP index is yet another redundant composite inter country development indicator.

Michael Hopkins (1991) made a critical appraisal of the UNDP’s first human development report and the composite index (HDI). He posed the question whether HDI can help understand development better and thus provide guidance for international and national development efforts. He examined the case of Sri Lanka and concluded that performing well on the HDI is not a necessary and sufficient condition for rapid economic growth or poverty elimination. Another important criticism leveled against HDI is that skating over the experiences of many countries; the report is of no use in addressing practical issues of implementing human resource development programmes. It is concluded that every country is different and that packages can not be borrowed from one country to fit another country's needs.

Das Gupta and Martin Weale (1992) made an attempt to extend the measures of general well-being in current use by including ordinal indices of political and civil liberties and provide a ranking of the world's poorest countries. They used Borda rules as the aggregator of a set of six constituents of well-being namely, per capita income, life expectancy, infant mortality rate, adult literacy rate, index of political rights and index of civil rights, to rank forty eight countries which were among the poorest in terms of income per capita. This new ranking was compared with rankings based on each of the six chosen elements of well-being. It is pointed out that there are countries which are very poor but enjoy relatively high levels of civil and political liberties.

Srinivasan and Verma (1993) discussed a conceptual framework within which the term 'development' can be defined for the purpose of identifying measurable indicators of development. They also made a brief review of the major approaches to the development of indicators and described human development as a multi-dimensional concept comprising economic, social, political, psychological and spiritual dimensions. It is argued that development should begin with the fulfillment of basic needs of an individual in terms of food clothing and shelter and gradually reach the highest level of self reliance. The authors concluded by expressing the hope that the review of the past major approaches will stimulate further discussion on the issue of development indicators.

S.P Pal and D.K Pant (1993) made an attempt to critically assess the methodology employed in UNDP's Human Development Reports for measuring progress and suggested modifications in the methodology and formulation. The conceptual framework used by UNDP is attacked on the following grounds. (1) The HDI is sensitive to the choice of maximum and minimum values of an indicator, (2) the HDI ignores the distributional aspects of an indicator across population groups within a country, (3) the introduction of the income variable in enlarging people's choices is not fully justified and (4) the three indicators chosen do not fully reflect the development goals of a country. To do away with the above said limitations of HDI, they introduced a few modifications and examined how meaningful and realistic are the numerical values generated by the two alternative measures. It is claimed that the values generated by modified HDI are more realistic than those by UNDP's HDL

Mahbub ul Haq (1995) explained how the focus of Development Economics shifted from national income accounting to people-centered policies. He also analysed the evolution of the human development index as a far more comprehensive measure of socio-economic progress than the traditional measure of GNP, and introduced a political freedom index in measuring overall social development, with political participation, rule of law, freedom of expression and non discrimination as indicators. Being one of the chief architects of the new paradigm, Haq discussed the development policies and strategies that link economic growth with human lives. It is opined that most development plans would look different if their pre occupation were with people rather than with production. It is concluded that the human dimension of development is not just another addition to the development dialogue. Instead of being the residual of development, human beings could finally become its principal object and subject.

UNDP (1995) focused on the pervasive and persistent inequality between men and women. It is pointed out that the focus on gender inequality is an important starting point since a widespread gender bias severely affects the social, economic and political situation of many countries. The innovative features of the report were the design of two new composite indices: the gender related development index (GDI) and the gender empowerment measure (GEM) and ranking of countries according to GDI and GEM. GDI measures achievements in the same basic capabilities measured by HDI, but takes note of inequality in achievements between men and women. The GEM examines whether women are able to participate actively in economic and political life and take part in decision making. It is argued that continuing exclusion of women from many opportunities of life totally warps the process of development. Thus the simple but far reaching message of Human Development Report (1995) is: "Human development, if not engendered is endangered".

Inderjeet Singh and Reena Singh (2001) have attempted a modification in UNDP's Human development Index formulation and have constructed an income disparity adjusted HDI and a gender disparity- adjusted HDI for the Indian States. The UNDP methodology is criticised on the ground that it assigned equal weight to all the three indicators of human development. In the wake of this the authors attempted to modify HDI by using weighted average instead of simple average used by UNDP. HDI is further criticised as it hides differences in the distribution of income, gender etc. So income is adjusted by a factor indicating distributional inequality (1- Gini coefficient). The adjusted income W(Y)(1-G) is used as the third component along with life expectancy and educational attainment. To highlight gender disparity in human development the per capita income of each state is discounted on the basis of relative ratio of work participation and wage differentials. It is concluded that income and gender disparities in a State have a strong bearing on its level of human development.

Achin Chakraborthy (2002) discussed the methodological issues behind construction of human development index (HDI) and its use in measuring and comparing the level of human development. It is pointed out that a composite index fulfils the need for a measure of the achievements of development, but it pre-supposes a deliberate conceptual aggregation of separable variables. For him, a composite index contains no more information than the individual indicators out of which it is made. It only presents that information in a form convenient and more amenable to some forms of analysis. But in the process, much useful information may be lost. This loss should be weighed against the gain from avoidance of the trouble of handling a large set of data. It is argued that indices can capture only the state of development and not the process aspect of it. Therefore it is suggested that a properly designed indicator capable of illuminating even some of the process aspects of development may be evolved.

UNDP (2002) examined political participation as a dimension of human development. The challenges faced by democracies in broadening people's participation and strengthening accountability of democratic institutions are briefly discussed. It also analysed the main tools - political participation, rule of law and freedom of expression - for measuring political and civil freedoms and their relationship to the human development index. The central message of this report is thateffective governance is central to human development and lasting solutions should be firmly grounded in democratic policies.

Paul Streeten (2003) made a brief survey of the progress of the development thinking and the place of people in it. He showed that our thinking has undergone an evolution and uneven progress. Both internal logic and new evidence have led to the continual revision of our views. He is of the view that previous and partly discarded approaches have taught us much that is still valuable and our current approach will surely be subject to criticisms and be overtaken. After giving a brief history of the development thinking up to Human development index, he concluded that human development and human development index are not ultimate insights and other ideas will take their place.

3. LITERATURE ON ROLE OF GOVERNMENT IN SOCIAL DEVELOPMENT

John Toye (1981) analysed empirical data on public expenditure in India between 1960 and 1970. The real growth rate of public expenditure and its functional and economic composition at the all India level are presented and the strong contrast between the patterns of the first and last five year periods is elucidated. He was of the view that one would not expect that States with different levels and growth rates of income and population density, as well as differences in political preferences would exhibit the same level or growth rates of expenditure. Regarding the problem of expenditure control he opined that it is a problem of intelligent selection based on detailed information about the economic effects of existing and potential items of public expenditure. Itis argued that the formal adoption of the administrative technique of programme and performance budgeting is no substitute for the creation of a series of official and political committees to decide priorities over the entire field of public expenditure.

B.N.P Singh (1983) made an analytical presentation of the growth and trends of public expenditure in the Indian institutional framework, with special reference to Bihar. He has presented a critical examination of the performance of the Government of Bihar on fiscal front from its inception in 1947 to the end of the Fifth Five Year Plan. It is pointed out that there has been a deceleration in the rate of growth of tax revenues of the state due to various concessions and exemptions granted. On the other hand the expenditure of the government has tremendously increased leading to budgetary deficits. The Government has failed in exercising effective and powerful control. It is opined that the successful operation of fiscal policy ultimately depends upon stable governments and efficient and honest administrative officials.

R.K Sinha (1983) examined the experiences of developing countries in their effort to achieve the goal of economic and social development. The experiences of developing nations indicate that they are far away from the goals to be realized. An attempt is made to find out the reasons for the failure of the developing nations and the following suggestions are made, (i) In neglecting agriculture, most of the countries have committed a great mistake. (ii)After agriculture the other important choice relates to the area of human development.

In selecting an appropriate structure of educational levels for accelerating economic growth, the developing countries should learn lesson from and make use of the experience of countries that have proceeded well ahead in economic development, (iii) Like expenditure on education, expenditure on housing, health and social services have important effect on redistributive justice and help economic growth.

Sharda (1986) provided an in depth analysis of the expenditure, sources of revenue and the impact of state budgetary operations on the social and economic development of the people with special reference to Himachal Pradesh. The impact of budgetary operations is traced in terms of the changes in the behaviour of capital formation in public sector and trend of Government's expenditure. Effectiveness of public expenditure has been evaluated by examining the changing importance of expenditure categories that are supposed to benefit the poor. It is pointed out that a vast majority of the population in Himachal Pradesh has for long remained deprived of the social services. Since these services will help in improving the productive capacity of the masses, it is essential for the Government to extend services like education and health care to the deprived groups.

Dholakia (1990) made an endeavour in the direction of developing a methodology for empirically estimating the socio-economic impact of the government expenditures. She has developed an alternative approach called Welfare Indicator Approach for measuring impact of Government expenditure. This approach attempted to measure welfare in real terms through combination of various socio-economic indicators, improvement in which would indicate improvement in the well-being of the poorest section of the society and expressed changes in the welfare variables as a function of government expenditure. It is evident from the analysis that in terms of the composite welfare index, government effort on human capital had increasing returns whereas; efforts on physical capital had constant returns. It is inferred that a policy in favour of expenditure on human capital would help achieving the targeted welfare level more efficiently than expenditure on physical capital.

Keith Hinchliffe (1993) discussed some analytical issues involved in a number of alternative ways of increasing total resources for human development programmes. The focus was on the mobilisation of resources for basic schooling, health services and drinking water. It is pointed out that mobilising additional resources for human development programmes is only one of the methods to expand their scope and coverage. The other one is to use existing resources efficiently. It is a central position of this paper that Governments are essentially responsible for providing basic services to their populations and that with few exceptions they are financially capable of this. It is concluded that Governments can fulfill the responsibility of providing basic services by increasing total revenue, by redirecting resources from some activities to others and by targeting of programmes.

Prabhu and Chatterjee (1993) examined the size and composition of State Government expenditures on health, education and nutrition and analysed the trends therein. They also assessed the impact of such expenditures on the levels of health, education and nutrition attainment. In addition, the study made an attempt in arriving at the social priority ratio and human expenditure ratio which are considered by UNDP to be the most telling indicators of the Government's commitment to the cause of human development. In order to assess the impact of these expenditures on levels of human development, indices of educational and health attainment have been constructed. These attainment indices were related to the index of physical infrastructure and revenue expenditure. It is concluded that there is an urgent need to increase the range and scope of Government interventions both directly and indirectly. It is also suggested that the increase in allocations are to be accompanied by a drastic real location of resources in favour of primary level facilities, particularly in health sector.

Shariff (1997) made an analysis of Central budget allocations for social sector and poverty alleviation programmes since 1990-91. The most relevant fact which emerges from a comparison of budget allocations and actual expenditures since 1990-91 is that there has been significant increase in social sector allocations. The study is concluded with two important suggestions. The first is that the corporate sector may be made a dominant partner in the development of social services in the country, allowing them to invest in social services in rural and backward areas. Secondly, to find out resources for social service sector, a 3-5 percent levy, which may be called as 'social service tax', may be imposed on the profits of large and medium business houses. The 'social fund' so generated may be transferred entirely to the states for investing exclusively in social sector.

Sunando Roy (1997) made an attempt to review the developments in the provision of basic needs by the state and tried to locate the shortcomings and weaknesses of the schemes and also to draw lessons for the future role of government in providing for the basic needs in the Indian context. According to the author restructuring the public policy towards basic needs involve tackling three basic issues namely targeting the resources, improving quality of services provided and reducing the existing gap between provision and utilisation of welfare activities. It is pointed out that as the nation stands at a crucial phase seeking to integrate with the rest of the world; it faces the grim reality that the number of people deprived of even the basic necessities of life remains substantially large. Attempts must therefore be made to promote basic needs with a conviction that this will help to improve productivity.

World Bank (1997) discussed the role and effectiveness of the State: what the State should do, how it should do and how it can do it in a rapidly changing world. It explored why and how some states have been more effective than others in playing a catalytic and sustainable role in economic development. It has provided a broad historical and conceptual introduction to the issues and examined the empirical evidence of the impact of State policies and institutions on development. The principal message conveyed is that sustainable economic and social development is impossible without an effective State. An effective State is central to social and economic development more as a partner and facilitator than as director. It is concluded that state should work to compliment markets, not to replace them.

Allen Roy et.al (2000) estimated the normative levels of expenditure on primary, secondary and higher education, utilising the pooled data for 15 States over the period 1992-93 to 1997-98. On the basis of normative expenditures, a comparative analysis of the normative and actual expenditure levels is made with the objective of classifying States on the basis of the relative emphasis laid on the provision of education. Findings of the study reveal that the actual spending on educational services in low income States is lower than their requirement. This implies that the existing fiscal equalisation mechanism has not been effective in offsetting the revenue and cost disabilities of the poor States in India.

Shariff and Ghosh (2000) analysed State level and national level patterns of public expenditure on various heads of account in education. Patterns of intra sectoral allocation of resources also are analysed. The priority to education in the national development framework is studied in terms of share of education in gross national product and in government expenditure. The unit cost of education is studied in terms of per pupil government expenditure at different levels of education. It was found that the annual rate of growth as well as the share of education expenditure in GDP has declined considerably during the post reform period. Per pupil expenditure on education, especially by the lessdeveloped states also has declined. As 90 percent of the expenditure on education is currently met from state funds, there are limitations on the availability of resources for education. Therefore it is suggested that the central government may expand its role in contributing resources, especially in respect of elementary education. Lack of resources cannot be a convincing argument for failing to discharge this national duty.

Kuldip Kaur and Bawa (2000) examined the impact of social services on poverty alleviation through lagged regression equations separately for each sector for rural and urban areas in major States in India. The State-wise estimates of poverty index showed that poverty alleviation varied significantly across the states. It is pointed out that much of these interstate differences can be attributed to State specific growth promoting or inequality reducing policies. Results of the analysis do not lead to any optimistic turn about the effectiveness of social sector expenditure on incidence of poverty, despite a rise in government expenditure on these sectors. Therefore it is concluded that the need of the hour is to ensure that social expenditures are cost effective and well targeted.

Panchamukhi (2000) made an attempt to focus on the impact of economic reforms on social sector in India, by comparing the data of the pre-reform and post-reform periods. The important questions addressed are (1) Are allocations to social sector, especially to education and health being affected during the period of economic reforms? (2) What are the effects of such changes in social sector outlays? (3)Are all the components of the social sector uniformly affected? It is clear from the data that during the reform period, expenditures on social sector have relatively declined. It is shown that with regard to individual components of social sector also, the initiative of the government presents a contrasting picture for the pre-reform and reform period. It is shown that during the period of reforms, the relative allocations to the sub-sectors of the social sector have been disturbed even showing declining trends. Therefore it is concluded that the effects of economic reforms appear to be unfavourable to countries like India, particularly when their social impacts are considered.

Sarker and Prabhu (2001) estimated the rate of growth of government expenditure on social services, on education and health in 15 major Indian states over the period 1974-75 to 1995-96, using kinked exponential growth model. The main objective of this exercise was to investigate whether there has been a break or shift in the trend path of social sector expenditure series and if there has been a shift, what was the direction of the shift. The results of the trend analysis for real per capita social expenditures revealed a clear shift in 14 out of 15 States in the latter part of 1980s. They have estimated the levels of expenditure that would have been incurred had the past trends continued. They also attempted to indicate the impact of these expenditures on social indicators. It is concluded that the deceleration of social sector expenditures in low income and low human development States like Bihar, Orissa and Madhya Pradesh demands special attention.

Shariff et.al (2002) :-have analysed trends in expenditures on social sector and poverty alleviation programmes from 1990-91. The trends in State expenditure, expenditure by the Central Government and Central and State adjusted combined expenditures are examined in detail. It is shown that overall expenditure on social schemes was increasing in real terms but mainly through increased expenditure of the Central Government. The State Governments seem to be easing out of their constitutional commitment to sustain programme in social sector. It is also seen that there are large inter- sectoral real location of funds in the poverty alleviation sectors. One major development has been that large funds that were allocated to employment generation have now been diverted to the rural road construction. It is cautioned that this reallocation may have serious implications on employment generation.

Dev and Mooij (2002):- examined the trends in social sector expenditure in the Centre and State budgets for the period from 1990-91 to 2000-01 and compared expenditure levels in the 1980s with those in the 1990s. They analysed the trends in social sector expenditures as a proportion of state domestic product, as a percentage of the aggregate budget expenditure and also as real per capita expenditure. The trends are examined at three levels - Centre, States and combined Centre and State. The most significant change visible both at the Centre and State is a shift away from the traditional way of addressing rural poverty to basic needs interventions. Another finding is that the social sector expenditure in India is low compared with East Asian countries and compared with the UNDP recommended ratios. The study is concluded with two important observations First; there is an urgent need for stepping up social sector expenditure. Second, there is an obvious need for better utilisation of the allocated money.

Kaur and Misra (2003) have made an attempt to analyse the present state of social sector development across States and to examine the effectiveness of public spending on social sectors namely education and health in terms of select human development indicators of various States. Empirical analysis to examine whether increased public outlays have been reflected in improved social indicators is done for 15 States. The variables chosen are gross enrolment ratio-both primary and secondary-in education sector and infant mortality as health attainment indicator. State-wise revenue expenditure on education and health as percentage of GSDP is taken as the policy variable. The study clearly brings out that the relationship between public spending on education and primary enrolment is stronger for poor States. However it is shown that the association between public spending and health outcome is weak. Income turns out to be a more significant determinant of health outcome than public spending. The authors have recommended future research exploring relative role of different factors for different states by examining the slope coefficients.

Seenuvasan (2004) discussed the welfare programmes and issues relating to the expenditure on welfare programmes in India. Ananalysis of welfare expenditure in the pre reform and reform periods is made using state expenditure on welfare programmes. The study revealed that among the welfare programme components, the expenditure share of labour and labour welfare, employment and nutrition showed broadly declining trend in the reform period when compared to pre- reform period. In contrast the expenditure share of rural development and social security has improved in the reform period in almost all States. A brief analysis of the other components of social service expenditure namely health and education revealed that during the reform period the expenditure on health declined, while most of the States witnessed increasing trend in education expenditure. It is cautioned that unless people in the rural areas, SC/ST/OBC, Labourers and other welfare dependents are incorporated, supported and empowered through sound welfare programme policies, efforts for removing poverty and attainment of human welfare would be ineffective.

Mukta.S.Adi (2004) made an attempt to examine as to how the expenditure on social sector has changed in the post reform period vis-a-vis the pre-reform period. It is seen that over the years, public expenditure on social sector has increased substantially in absolute terms but the rate of increase during the 1990s has been at declining rate until 1997. It is pointed out that multiplicity of programmes for universalizing education, providing primary health care, ensuring employment, subsidizing food distribution and a number of social welfare programmes have proved to be counterproductive. Many of them are deeply criticised for being unsuccessful in generating intended outputs due to poor implementation. It is pointed out that the per capita public expenditure on social sector and poverty alleviation programmes is too low and inadequate and a major part of the allocated money is disbursed as salaries and spent on establishment. The author has suggested further studies to explore reasons for low spending and to explore corrective measures that National and State Governments can undertake on an annual basis.

We have made a humble effort for identifying some important studies pertaining to the theme for social development indicators. There are many facts that can be viewed at length and such studies can be fruitful for the development in social sector of the economy.

4. ACKNOWLEDGEMENT:

We thank the referee for his valuable comments and suggestions to revise this paper.

5. REFERENCES:

[1] **Abusaleh Shariff(1995)** Health Transition in India, NCAER,Oct 1995

[2] **Dasgupta MK & Dasgupta T. (2008)** Human development in India

Indexing across Selected States Economics of education and Health in India (Ed).

[3] **D. Pant (1997)** Human development,demographic transition and economic growth linkage, July 1997.

[4] **Dholakia (2003)** Regional Disparity in economics and Human development in india Economic and political weekly 38 (39).

[5] **G.chakrabarty and S.Pal (1995)** Human development profile of Indian state,Jan. 1995

[6] **Ghosh M. (2007)** Economic growth and Human development reforms and Indian economic development Bookwell publications New Delhi.

[7] **Mc Gilivray (1999)** The Human development Index –yet another redundant composite development indictor? World development – Vol-19, No-[10],1999

[8] **Michael Hopkins (1991)** World development Vol-19, 1999 Human development revisited – A new UNDP report.

[9] **Newman and Thomson (1989)** Economic growth and social development Vol-17, Issue-4 1989

[10] **Seenivasan R (2014)** Analysis of Relationship between FDI and Economic growth Vol-7, Jan. 2014

Note

MONTE CARLO SIMULATION APPROACH TO ESTIMATE CLRM

H. M. DIXIT*

ABSTRACT

An essential feature of the parameters of classical liner regression model is that they are unbiased, satisfying BLUE property. In this article it is shown how to use simulation approach to estimate the parameters establishing BLUE property.

KEY WORDS

Simulation, Repeated sampling, BLUE, CLRM

1. INTRODUCTION

For the classical linear regression model (CLRM), we know how to obtain least sauares estimators of the regression coefficients (parameters). An essential feature of these estimators (as provided by Gauss Marov theorem) is that they are BestLiner Unbiased Estimators (BLUE).

No doubt we can obtain these estimators using standard formulae based upon least sauares principle - that is how they are called OLSE (Ordinary Least Sauares Estimators)

How one knows empirically whether these estimators justify BLUE properties? How one knows if the OLS estimators are unbiased ?

This can be accomplished by what is known as Monte Carlo Simulation aproach for estimating the parameters. In this article, it is illustrated using standard two variables linear model under usual assuptions.

* Head, Dept. of Statistics, Arts & Commerce College, PILVAI (N.G.)
dixithm13@gmail.com
(rcd. Dec. '18 / rvd. Jan.'19)

2. Monte Carlo Simulation Method for estimating Bivariate Linear regression Model.

Let us consider the following two variables linear regression model. Under usual assumptions it is also termed as classical linear regression model (CLRM)

The PRF of the model is expressed as

$Yi = \alpha + \beta Xi + Ui$ (1)

$E(Yi/Xi) = \alpha + \beta Xi$ (2)

and $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ (3)

our aim is to obtain estimaters α and β by using Monte Carlo Simulation approach.

Procedure :

(1) Suppose that the true values of the parameters are $\alpha = 40$, $\beta = 0.8$

(2) Now choose the sample size (say n = 30)

(3) Fix some values for each of Xi, Thus we have 30 values of Xi.

(4) Use random number table and select 30 values and call them Ui values (Now a days most of statistical packages have already built in random number generators)

(5) Now we have 30 values of Xi and Ui with assumed values of α and β using eauation (1), obtain 30 values of Yi.

(6) Now using 30 values of Yi thus generated regress them on the 30 Xi values and compute $\hat{\alpha}$ and $\hat{\beta}$ the least sauares estimaters.

(7) Suppose you repeat this experiment 100 times, each time using the same α, β and Xi values. Each time Ui values will be different as specified by random numbers.

Thus, we may have 100 experiments generating 100 values of α and β.

(8) Find the average value of α and β, so obtaind, call them α^* and β^*.

(9) Compare α^* and β^* with values α and β. If these values are nearly the same, we arrive at our estimators of the parameters.

This also establishes that the least sauares estimators are indeed unbiased as $E(\alpha^*) = \alpha$ and $E(\beta^*) = \beta$.

It may be noted that this expriement needs to be repeated many times (sometimes 1000 to 2000) to arrive at the final estimators.

3. Concluding Remarks

This simulation approach can be done by means of computer. Such an approach is particularly useful to study the behaviour of estimators in small or finite sample. These experiements also explain the concept of repeated sampling.

4. Acknowledgements

I thank the referee for reviewing this article.

5. Selected Reference

[1] **Demodar Gujarati** (1995), Basic Econometrics (Mc Graw Hill Co, N.Y.)

[2] **Damodar N. Gujarati, Dawn C. Porter and Sangeeth Gunasekar,** Tata Mcgraw Hill Education Pvt. Ltd. (New Delhi)

[3] **H. A. TAHA** (1989)
Operations Research - An introduction (MC Millan Publising Co., N.Y.)

BIOGRAPHY

PROFESSOR J. K. GHOSH*

**H. D. BUDHBHATTI

In this article, it is my very humble attempt to present briefly the biographical sketch of an eminent internationally reputed Indian statistician 'Prof. J. K. Ghosh'. He was a thinker, an academician, and visionary personality who spent his whole life for giving his versatile deliberations in many areas of statistics.

Professor Jayantkumar Ghosh (or J. K. Ghosh) as he is commonly known was born in Calcutta on May 23, 1937. He lived for 80 years and passed away on september 30, 2017. He was graduated from the dept. of Statistics at Calcutta University. He carried out his doctaral and post doctaral research work in India and Abroad. He has been a prominent contributor to statistics discipline for about five decades.

Ghosh's earliest work was influenced by Abraham Wald's invention of the SPRT. As a twenty-something, in 1960, he studied the ASN and the efficiency of the sequential t-test, as a part of his Calcutta University PhD dissertation under the supervision of H.K. Nandi. A few years later, simultaneously and independently, with Jack Hall and Bob Wijsman, he wrote a fundamental paper formalizing the results of Charles Stein and Don Burkholder on commutativity of reduction by invariance and sufficiency. This work was later followed up

* Adapted from wikipedia (the free encyclopedia) and other related resources. (We express our sincere thanks and gratitude for this assitance)

** Ex. CSO, Head, Statistics Dept., GSRTC, Ahmedabad

(Thanks to the referee for reviewing this article.)
(rcd. Oct.'18 / rvd. Dec.'18)

by others, and was cited in Ferguson's classic text.

Around this time, C.R. Rao was publishing his path-breaking results on second order efficiency of the MLE among asymptotically normal first order efficient estimates. Around 1970, Ghosh, jointly with his student Kasala Subrahmanyam, began his long journey into higher order asymptotics, and in well known later work with many others, established second order risk optimality of the MLE under essentially all bounded convex loss functions. It was evident that Ghosh admired MLEs (and perhaps LRT). He saw MLEs as more-or-less Bayesian objects. In his discussion of Brad Efron's 1975 curvature paper, Ghosh shows open disdain for frequentist statistics, likening it to Lewis Carroll's Cheshire cat, all but dead save its ironic grin. It is not clear if he was already a committed Bayesian at that point, although he was obviously supportive of Basu's epic articles on likelihood and information, and edited a Springer monograph compiling Basu's main work at that point.

Professor Ghosh's most cited work is his 1978 joint paper with Rabi Bhattacharya on the validity of Edgeworth expansions for smooth functionals in the iid or independent situations. The work is widely regarded as a masterpiece in controlling the error in the central limit theorem. Earlier, Professor Ghosh had given a very sweet alternative proof of Bahadur's classic stochastic representation of quantiles as asymptotically linear statistics, under one less derivative than Jack Kiefer would need, but with an in probability remainder, unlike Kiefer. For many first order results, this suffices.

In the mid-80s, Ghosh became the Director of the Indian Statistical Institute (ISI). His tenure was marked by prudence and negotiation. As the ISI directorship was coming to an end, he was hired as a visiting professor at Purdue University in 1990, and tenured in 1997. On his arrival there, he converted his NSF-CBMS lectures into the well known IMS monograph on higher order asymptotics. At this time, he became the President of the International Statistical Institute.

Professor Ghosh also showed his first interests in Bayesian model selection at that time by revisiting Schwarz's work. The interest later blossomed into full fledged work with Jim Berger and numerous other collaborators. Undoubtedly, he helped popularize and sometimes clarify fractional Bayes factors, and objective priors custom-made for model selection. He provided clarifications and publicity for Bernardo priors, and in a long series of papers with many coauthors established Bayesian (hierarchical) resolutions of various types of Neyman–Scott quagmires. Many of these things, along with modern Bayesian computing, found their way into his Springer text on Bayesian analysis written with Mohan Delampady and Tapas Samanta.

Although for a large part of his career he concentrated on parametric models, one of his prime contributions to Bayesian statistics is the fundamental Springer book with R.V. Ramamoorthi on Bayesian nonparametrics. Later, jointly with Aad van der Vaart and his students Subhashis Ghoshal and R.V. Ramamoorthi, he wrote two major papers on consistency of Ferguson mixtures. Ghosh presented a part of this research at the International Congress of Mathematicians.

Although a theoretician, Ghosh had an interest in modeling and looking at data. He used to regularly collaborate with Bijoy Singha Mazumder and Supriya Sengupta on sediment transportation and grain size distributions. At the inspiration of Vijay Nair, jointly with Pulakesh Maiti, T.J. Rao and Bikas Sinha, he wrote a historical article on statistics in India. He was visibly proud of the heritage of the ISI, and often mentioned in conversation students and colleagues there.

In the post Rao–Basu era, Professor Ghosh acted as a unique coalescing force for Indian statistics within India, and served as a bridge between Indian statistics and statistics in the west. Two important examples are his organizing roles in the 1974 conference at the ISI in Calcutta on Recent Trends in Statistics, and the 1980 conference at the ISI in Delhi in honor of C.R. Rao's sixtieth birthday.

Over a period of 35 years, Professor Ghosh received numerous awards. He was awarded the Bhatnagar prize by the CSIR of India in 1981, the Presidency of the statistics section of the Indian Science Congress in 1991, the Mahalanobis gold medal of the Indian Science Congress in 1998, the P.V. Sukhatme prize in statistics of the Government of India in 2000, an honorary D.Sc. degree by the B.C. Roy Agricultural University in West Bengal, India in 2006, the Lifetime Achievement award of the International Indian Statistical Association in 2010, an honorary D.Sc. degree by the ISI in 2012, and the Padma Shree award of the Government of India in 2014. He was also a Fellow of the National Academy of Science of India and the IMS.

Ghosh was a private person, but maintained committed friendships with many people. He was well read in both Indian and western literature and was fond of Tagore's poetic songs. He was spectacularly devoted to his family and was an exemplary father and husband (his wife, Ira, passed away a few weeks before him). In a clearly distinguished career, Professor Ghosh influenced the work of various spectra of statisticians at different times, was an undisputed leader of Indian statistics, and produced an enviable number of successful students. It is an extraordinary legacy.

RESEARCH SPECTRUM

The spectrum of his contributions encompasses sequential analysis, the foundations of statistics, finite populations Edgeworth expensions, second order efficiency, Bartlett corrections, noninformative, and especially matching, priors, semiparametric inference, posterior limit theorems, Byesian nonparametrics, model selection Bayesian hypothesis testing and high dimensional data analysis, as well as some applied work in reliability theory, statistical quality control, modeling hydrocarbon discoveries, geological mapping and DNA fingerprinting. By itself, covering suchdiverse topics in depth is a major career achievement. He has authored over 130 publications including three monographs and several edited volumes. His books, one entitled Higher Order Asymptotics and published as an IMS monogrpah and another entitled Byaesian Nonparametrics, co-authored by R.V.Ramamoorthi and published by Springer - Verlag, continue to hold respected positions for researchers in these areas. His recently published third book is a fine graduate text on Bayesian inference. In addition, his service to the profession, especially as the editor of Sankhya, has been invaluable.

1. SEQUENTIAL ANALYSIS

J. K. Ghosh started his research career in Sequential Analysis in the early sixties as a Graduate Student in the Department of Statistics at Calcutta University. Wald had recently introduced his sequential probability ratio test (SPRT), but its properties were not well understood in the composite case. This was the first topic to which Ghosh turned his attention. Through his work, many of the properties of SPRT and related procedures were established and better understood. For instance, in the testing context, double minimaxity essentially means simultaneous minimization of average type I and II error probabilities. In his first published work, Ghosh clarified a result of Wald on the double minimaxity of the SPRT for normal two sided alternative hypothesis (with unknown scale) separated from the null by δ.

It is well-known that the power function is monotonic in many common families for fixed sample sizes. Ghosh established an analog of this result in , namely that the operating characteristic function of the (generalized) SPRT continues to be monotonic. Also in the sequential context, considered the admissibility of sequential tests based on a simple identity which later became known as the Ghosh–Pratt identity. Ghosh compared the SPRT not just with the class of all tests with finite expected sample size but also within other classes, for instance, the class which requires at least one observation or which requires no more than a predetermined number of observations to reach a conclusion. Following this, Ghosh continued to elucidate more properties of the SPRT, and its variants, which could be seen as analogs of the corresponding properties Neyman–Pearson or Bayes tests for fixed sample size. He proved that for J. K. Ghosh's contribution to statistics: A brief outline 3 exponential families, truncated or untruncated Bayesian sequential decision rules' terminal decisions describe regions in terms of sufficient statistics, and also showed that for testing problems, truncated generalized SPRT's form a complete class. About two decades later, Ghosh returned to sequential problems, along with various co-authors. He studied an invariant SPRT to identify two normal populations with equal variance and obtained bounds for error probabilities. Most recently, similar bounds for an invariant SPRT with respect to an improper prior have also been obtained in. Two-stage procedures are closely related to sequential procedures. Recall Stein's famous problem of finding a bounded length confidence interval for the normal mean with unknown variance. Stein proposed a two-stage procedure for doing this: In the first stage, the sample variance determines how many samples are to be taken in the second stage. An obvious shortcoming of the procedure is that the second stage sample variance is not used in the construction of the interval. So, it is natural to ask whether one can improve Stein's procedure by using the second stage sample variance. Surprisingly, it is impossible to better Stein's procedure, as shown in. However, the procedure can be improved in a different, and perhaps more appropriate sense. The confidence coefficient does not in general properly reflect the true sense of confidence about a parameter after observing data. For instance, if two observations are obtained from a $U(\theta, \theta + 1)$ family, then the assessment of θ is very precise when the two observations differ in magnitude by nearly 1, while the assessment is much less precise if the two observations are close to each other. This means that classical confidence intervals fail to indicate the true difference in the level of confidence after observing the sample.

Motivated by this, Kiefer suggested letting the confidence coefficient depend on the data. After all, in reality, for a given random interval I, we often want to predict the indicator function $1\{\theta \in I\}$. Since this object is unknown, it is traditionally estimated by a constant, the best constant being the expectation $P\ \theta\ (\theta \in I)$, which becomes fixed (or asymptotically fixed) for many classical intervals. However, from a prediction theory point of view, it makes more sense to let the predictor of $1\{\theta \in I\}$ depend on the observed data. The predictor considered in this way is called the random confidence coefficient associated with the confidence interval I. It is shown in that the second stage sample variance can be used to boost the random confidence coefficient of a bounded length confidence interval.

2. FOUNDATIONS OF STATISTICS

From the examination of individual data points as they relate to the testing problem, Ghosh shifted his attention to data summarization, focussing on the relationship between sufficiency and invariance. Sufficiency isolates features of the collection of observations from those of the individual ones which are independent of the features of the collection. Invariance, on the other hand, summarizes data by imposing symmetry constraints. In practice, both sufficiency and invariance restrictions are applied, but their order of application is an issue of interest.

3. ASYMPTOTICS

The asymptotic point of view undergirded Ghosh's thinking, even in problems that were not primarily focussed on asymptotic properties. In a sense, much of his work on sequential analysis, Bayesian analysis and Bayesian nonparametrics are also, at least implicitly, work on asymptotics. In fact, many of the most important asymptotic ideas, such as higher order asymptotics and Edgeworth expansions, were pioneered by him. Moreover, in terms of how his thinking progressed, asymptotics can be regarded as the next natural conceptual step after thinking about data points in sequential analysis, and sufficiency or invariance as a data summarization technique. That is, once we have gathered and summarized our data, we want to see where it seems to be leading us. Ghosh's work on asymptotics can be broadly grouped into seven categories. He worked on the Bahadur–Ghosh–Kiefer representation for a quantile. He made foundational contributions to establishing the existence of Edgeworth expansions. In higher order asymptotics, Ghosh examined second order efficiency, Bartlett correction and contributed to our understanding of how Wald, Rao and likelihood ratio tests compare. Then he turned his attention to Bahadur efficiency and the vexing Neyman–Scott problem.

Among his best-known discoveries are the Bahadur-Ghosh-Kiefer representation (with R. R. Bahadur and Jack Kiefer) and the Ghosh-Pratt identity along with John W. Pratt.

* His research contributions fall within the fields of :
* Bayesian inference
* Asymptotics
* Modeling and Mdoel selection
* High dimensional data analysis
* Nonparametric regressiona and density estimation
* Survival analysis
* Statistical genetics

AWARDS/HONORS

* Elected Member of the International Statistical Institute
* Advisory Editor, Journal of Statistical Planning and Inference
* Fellow of the Institute of Mathematical Statistics
* Fellow of the Indian National Science Academy
* Life Member and Director of the Calcutta Statistical Association
* Fellow of the Indian Academy of Sciences
* Japanese Society for Promotion of Sciences Fellowship, 1978
* Shanti Swarup Bhatnagar Prize for Science and Technology, 1981
* President, Statistics Section of the Indian Science Congress Association, 1991
* President, International Statistical Institute, 1993
* Mahalanobis Gold Medal of Indian Science Congress Association, 1998
* P. V. Sukhatme Prize for Statistics, 2000
* Mahalanobis Memorial Lecture, State Science and Technology Congress, W. Bengal, 2003
* D.Sc. (h.c.), B.C. Roy Agricultural University, W. Bengal, India, 2006

* International Indian Statistical Association (IISA) Lifetime Achievement*Award, 2010 [5]
- Padma Shree (2014) by the Government of India

BIBLIOGRAPHY

He has published over 50 research papers. He has also published four books, which are:

* Invariance in Testing and Estimation (Lecture Notes), 1967, published by Indian Statistical Institute, Calcutta.
* Higher Order Asymptotics (based on CBMS-NSF lecture), published jointly by Institute of Mathematical Statistics and American Statistical Association, 1994.
* (with R.V. Ramamoorthi) Bayesian Nonparametrics (Springer 2003).
* (with Mohan Delampady and Tapas Samanta) An Introduction to Bayesian Analysis - Theory and Methods, Springer 2006.

REFERENCES

[1] **Dasgupta, Anirban** (16 November 2017). "Obituary: Jayanta Kumar Ghosh, 1937–2017". IMS Bulletin. Retrieved 2018-02-21.

[2] **Clarke, Bertrand; Ghosal, Subhashis** (2008). J. K. Ghosh's contribution to statistics. projecteuclid.org. ISBN 9780940600751. Retrieved 18 August 2018.

[3] **Lahiri, S. N** (1992). "On the Bahadur–Ghosh–Kiefer representation of sample quantiles". Statistics & Probability Letters. 15 (2): 163–168. doi:10.1016/0167-7152(92)90130-w.

[4] **Casella, George** (1996). "The Ghosh–Pratt Identity". Wiley StatsRef: Statistics Reference Online. doi:10.1002/9781118445112.stat01501.

[5] "International Indian Statistical Association".

EXTERNAL LINKS

* Indian Statistical Institute Statistics Department homepage
* Dr. Ghosh's profile at Purdue University
* Dr. Ghosh's webpage at the Statistics Department of Purdue University
* A biography of Dr. Ghosh written by Professor Anirban Dasgupta

SV NEWS LETTER

K. Muralidharan*

Out of numerous activities in the country and abroad only some of them are indicated here under this caption.

* EMRC, Gujarat University prepared MOOC course which is expected to run from this semester. It is on Mathematics and Statistics for business economics and it will be displayed on SWAYAM platform. For registration and further details contact http://swayam.gov.in/courses/5695-mathematics and statistics for business economics.

* Dept. of statistics, Gujarat University had organised one conference on clinical research and Life science on 8th September 2018. It was financially and academically supported by PHARMA STAT by the courtesy of Nirali Mehta. It was all India course in which practitioners cum academicians like A. V. Prabhakar, Arjun ROy, Shubha Rani, Ronak Patel, Hitesh Chauhan etc. had given their deliberations along with the faculty members of Dept. of Statistics, G.U.

* Dept. of Statistics, G.U. had organised one day programme on Data Analysis on 24th Dec. 2018, at Ahmedabad University. It was under the auspicies of Association for computer management, Ahmedabad Chapter.

* Arts and commerce college, Naroda, Ahmedabad will organise National Level Seminar on Multidisciplinary subjects on 4th January 2019.

* Dept. of Statistics, Gujarat University may start certificate course from the next semester on (1) Biostatistics (2) Data Science Analysis. One can contact the dept. for further details.

* 106th Indian Science Congress conference was held during Jan. 3 to 7, 2019 at Lovely University, Delhi G.T. Road, Phagwara, Punjab. PM had inaugurated this conference. The focal theme was Future of Indian Science and Technology.

* Professor, Dept. of Statistics, M. S. University, VADODARA
email : muralikustat@gmail.com

* ORSI conference was held during Dec. 16 to 18, 2018 at IIT Bombay.

* 55th annual conference of The Econometrics society of India (TIES) will be held during January 8-10, 2019. once can contact tiessecreta@gmail.com

* ICSO 19 (International Conference on Softwares and optimisation Techniquesh will take place during February 15, 16, 2019 at Agartala.

* Interdisciplinary Research on Automation, computation, Technology and Management will take place at Amit University, London Campus during April 24-26, 2019. Contact : icactm.london@gmail.com

READERS FORUM

A. M. PATEL*

(Readers are requested to feel free for sending their views and constructive suggestions for this section.)

* **C. D. Bhavsar (Ahmedabad)**

SV Journal has its own style and status. It is particularly useful for students, teachers and researchers in the subject. My best wishes to the team.

* **Bhushan J. Bhatt (Bhavnagar)**

This journal is beneficial for the research workers. It gives some useful informations also. Its frequency should be increased. My congrats and blessings for the team work.

* **P. Prajapati (Ahmedabad)**

It is my request to the editorial team of Sankhya Vignan Journal to arrange for the forthcoming issues in the following manner.

(1) First Issue (April) - Theoretical Articles

(2) Second Issue (July) - Applications

(3) Third Issue (October) - Data analysis research

(4) Forth Issue (Dec./January) - Research Articles pertaining to econometries, biometrics, sociometry etc.

This highlights the works in much lucid manner. Best wishes to the editorial team.

* Rtd. Principal, H. K. Commerce College, Ahmedabad
and Ex. Secretary, Gujarat Vidyasabha and Brahmachariwadi Trust, Ahmedabad.

Gujarat Statistical Association

Established : 1969

[Registered under Public Trust Act of 1950 (Bombay)]

R. No. E2502 A'bad-1974

The objective of the association is primarily to promote statistical ideas in pure and applied fields in the form of study, teaching and research in statistics.

The membership of GSA consists of Life / Institutional / ordinary members.

Membership	Fees	
	In India	Abroad
Insitutional Member (for 3 years)	₹ 2,000/-	US $ 500
Life Member	₹ 1,500/-	US $ 300
Ordinary Member (p.a.)	₹ 500/-	US $ 50

Annual Subscription for Gujarat Stastistical Review / Sankhya Vignan for individuals, Institutions and library is ₹ 500 (in India), US $ 50 (Abroad)

J. K. GHOSH*

Professor Jayantkumar Ghosh was born in Calcutta On May 23, 1937. He lived for 80 years and passed away on September 30, 2017. He was graduated from Dept. of Statistics at Calcutta University. He carried out his doctoral and post doctoral research works in India and abroad. Over a period of nearly 60 years, Professor Ghosh made timely and insightful contributions to a wide swath of theoretical statistics. He is an Indian Statistician who is very famous for his research work internationally. He was a very humble and polite person who was well read in both Indian and Western Literature and was also fond of Tagore's poetic songs.

His research contributions are in the fields of (1) **Bayesian Inference** (2) **Asymptotics** (3) **Modelling and model selection** (4) **High dimensional data analysis** (5) **Non parametric regression and density estimation** (6) **Survival analysis** (7) **Statistical Genetics** etc.

Prof. Ghosh has received more than 16 awards /Honors. He was president of International Statistical Institute. He also edited **Sankhya Journal.** Among very notable honours, he received **S. Bhatnagar Prize** in 1981. **Mahalanobis Gold Medal** of Indian Science Congressin 1998, International **IISA Lifetime Achievement award** in 2010 and **PADMASHREE** by Govt. of India in 2014.

***(Brief Biographical sketch is given inside the journal)**

This space is specially donated by Prof. Shailesh Teredesai (Ex. Head, Statistics Dept., S. M. Patel Insitute of Commerce, GLS, Ahmedabad-380 009.

Printed Matter **BOOK-POST**

(Journal of GSA, Ahmedabad)

To,

From :

Gujarat Statistical Association,

C/o. Statistics Dept. Gujarat University, Ahmedabad-380 009. (INDIA)