STOCHASTIC MODELS FOR POPULATION OF PAKISTAN

Muhammad Zakria
Roll No. T800583

Submitted in partial fulfillment of the requirements for the
Doctor of Philosophy degree in **Statistics**
at the Faculty of Sciences
Allama Iqbal Open University, Islamabad.

Supervisor
Professor Dr. Faqir Muhammad

December, 2009

To the memory of my sweet Mother & Father
To my Wife & Daughters

PAGES: 182

STUDENT: Muhammad Zakria

SUPERVISOR: Professor Dr. Faqir Muhammad

UNIVERSITY: Allama Iqbal Open University, Islamabad, Pakistan

YEAR: 2005-2009

SUBJECT: Statistics

DEGREE: Ph.D

Population of Pakistan is projected by different countries, scientists and bureaus using different methodologies. In this study, population projections, its age-sex distribution vision 2030 and inequality of the recorded and projected age-sex distribution is projected by different methods. Moreover, the reproductive cohort measure and fertility trends of the population during the last 20 years are measured. The said goals are achieved by using the population censuses data.

First of all, the quality of all censuses data is checked and found to be very poor especially of 1972 census. Different popular smoothing techniques are used to smooth the

(Acceptance by the Viva Voce Committee)

Title of Thesis: **Stochastic Models for Population of Pakistan**
Name of Student: **Muhammad Zakria**

Accepted by the Faculty of **Sciences**, Allama Iqbal Open University, in partial fulfillment of the requirements for the Doctor of Philosophy Degree in **Statistics**

Viva Voce Committee

Professor Dr. Faqir Muhammad
Dean, Faculty of sciences/Chairman, Department of Math. & Stat.

Prof. Dr. Muhammad Aslam
External Examiner

Professor Dr. Faqir Muhammad
Supervisor

Member: i) _____ Prof. Dr. Zafar Ilyas

ii) _____ Dr. Bahrawar Jan

iii) _____ Dr. Noor Muhammad Larik

iv) _____ Prof. Dr. Nowshad Khan

April 8, 2010

ACKNOWLEDGEMENTS

First of all my humble thanks to Almighty Allah- the most beneficent, gracious and merciful – without His infinite blessings, it was not possible for me to achieve this target.

The following few sentences encouraged me a lot:

كُنْ عَالِمًا أَوْ مُتَعَلِّمًا أَوْ مُسْتَمِعًا أو محبًّا (حضرت علی رضی الله تعالی عنه)

Syedina Hazrat Ali Murtaza said "Become a teacher or student or listener or lover to knowledge"

"Three sentences for getting success:"

"Know more than other, Work more than other, Expect less than other"

(William Shakespeare)

Words are too debilitating to express thanks to my Supervisor, **Professor Dr. Faqir Muhammad, Chairman**, Department of Mathematics and Statistics, Allama Iqbal Open University (AIOU), Islamabad, Pakistan. I am deeply indebted to his valuable, commendable, encouraging, inspiring, and cooperative supervision. I found his valuable feedback and unfailing enthusiastic guidance during every step of my research work. Thank you very much Sir! **Professor Dr. Faqir Muhammad**.

My special thanks are due to Professor Dr. Inayat Khan, Chairman, Department of Mathematics & Statistics, UAF and Dr. Irshad Ahmad Arshad, Associate professor & Chairman Department of Statistics, IIU, Islamabad for their friendly cooperation,

motivation as well as encouragement with their positive and inspiring attitude. Heartiest thanks are also extended to Mr. M. Ikram Janjua, Dy Director (Admin), Population Census Organization for providing the data for the research project and to Mr. Zahoor Ahmad, M. Tariq Ejaz, Dr M. Farid Khan, Professor Dr. Muhammad Aslam, Mr. Sultan Mahboob (Senior Demographer, NIPS), Mr. Tanveer, Mr. Fateh-ud-din, Mr. Zafar Zaheer, Miss Zainab Javed Dar, Mr. Islam-ud-din Shazad, M. Arif, Nadeem Saeed, M. Yaseen, Ahmed Nadeem and Kamran Abbas for their well wishes.

I am highly obliged to Professor Faisal Arif Sukhera and Ms. Shabnam Nasim for sparing their valuable time in proofreading the thesis. My thanks for the supportive staff, especially, Adeel Sheikh. Asad-ullah Qureshi, Asif Ali and Asif Mahmood.

I appreciate the encouragement which I received from my beloved wife and caring daughters who have always been a great source of inspiration, love and affection to accomplish this uphill task.

My heartiest and sincere sense of gratitude is extended to my affectionate brothers and sisters especially Baji Asia and Bhai Muhammad Yaqoob (United Kingdom) whose prayers, encouragement, moral support and well wishes enabled me to attain this gigantic goal.

(Muhammad Zakria)

TABLE OF CONTENTS

a student of Ph.D. at the Allama Iqbal Open University do hereby solemnly declare that the thesis entitled **Stochastic Models for Population of Pakistan** submitted by me in the partial fulfillment of Ph.D. degree in **Statistics** is my original work, except where otherwise acknowledged in the text, and has not been submitted or published earlier and shall not, in future, be submitted by me for obtaining any degree from this or any other University or institution.

Signature: _____

Name in Full: Muhammad Zakria

Date: 03/12/2009

FORWARDING SHEET

The thesis entitled

Stochastic Models for Population of Pakistan

submitted by **Muhammad Zakria** in the partial fulfillment of Ph.D. degree in **Statistics** has been completed under my guidance and supervision. I am fully satisfied with the quality of student's research work.

Dated: 03/12/2009

Signature of the Supervisor: _____

Name: Professor Dr. Faqir Muhammad

ACRONYMS

HEC	Higher Education Commission
NIPS	National Institute of Population Studies
PIDE	Pakistan Institute of Development Economics
ME	Mean Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
CRV	Coefficient of Relative Variation
NWFP	North West Frontier Province
FATA	Federally Administered Tribal Area
MSMR	Multi-Stage Mark Recapture
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
AR (1)	Autoregressive of order 1
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
AIC	Akaike Information Criteria
MSE	Mean Squared Error
SC	Schwarz Criterion
GNP	Gross National Product
GMD	Gini's Mean Difference
ID	Index of Dissimilarity
MLD	Mean Logarithmic Deviation
PAS	Population Analysis Spreadsheet
CBR	Crude Birth Rate
CWR	Child Women Ratio
TFR	Total Fertility Rate
GFR	General Fertility Rate
NRR	Net Reproduction Rates
GRR	Gross Reproduction Rate
ASFR	Age Specific Fertility Rate
IMR	Infant Mortality Rate
MCH-FP	Maternal and Child Health Family Planning
MAC	Mean Age of Childbearing
FHS	Fertility Health Survey
PDHS	Pakistan Demographic Health Survey
PCPS	Pakistan Contraceptive Prevalence Survey
PFS	Pakistan Fertility Survey
BBS	Bangladesh Bureau of Statistics
FBS	Federal Bureau of Statistics
CVPP	Cross Validity Prediction Power

PRB	Population Reference Bureau
WPP	World Population Prospects
NGO's	Non Government Organizations
IDB	International Data Base
SAARC	South Asian Association for Regional Cooperation

DECLARATION

I Prof. /Dr FAQIR MUHAMMAD

Supervisor of Mr. MUHAMMAD ZAKRIA

an AIOU research student, do hereby solemnly declare that the thesis

entitled **Stochastic Models for Population of Pakistan**

being submitted as partial fulfillment of **PhD** degree in the discipline

of **STATISTICS**

is an original work of the student except where otherwise acknowledged in the text, has not been submitted or published earlier for obtaining any degree from this or any other university or institution. The thesis is complete in all respects and ready to evaluate by foreign-external subject experts.

][

Signature

Date: 03-12-2009

Name in full Prof. Dr. Faqir Muhammad

Address: Chairman, Department of Mathematics & Statitics, AIOU, Islamabad. Pakistan.

CHAPTER 1

INTRODUCTION

In the most common sense, a population is the total number of persons inhabiting in a country, city, any district or area and ecologically, the population is the assemblage of a specific type of organism living in a given area (Population, 2009). Usually, the population censuses in all countries of the world have been conducted at decennial intervals and provide information about the size and structure of the population during the census years.

In 1950, Pakistan was ranked 14th populous country of the world after China, India, USSR, USA, Indonesia, Brazil, Japan, Bangladesh etc. The first population census of Pakistan was held in 1951 which recorded 34 million population while the 2nd census was held in 1961 and reported 43 million thus exhibiting an average annual growth rate of 2.45% (Anonymous 1967). The 3rd population census was delayed by one year and held in 1972 due to India-Pakistan war. In this census, population was reported 65 million with an average annual growth rate of 3.67% indicating 52.31% increase as compared to that of the 1961 population census. The growth rate remained highest during this census in the history of Pakistan and became the major cause of the drastic increase in the population (Anonymous, 1972). The 4th population census was held in 1981 and reported 84 million population with an average annual growth rate of 3.06% (Anonymous, 1984)

whereas, the latest and 5th Population census delayed by 7 years, was held in March 1998. This delay was owing to the volatile and disturbing political scenario in the province of Sind. The Population and growth rate of this census were reported 132 million and 2.69% respectively (Anonymous, 2001). It means that up-to March 1998; the population of Pakistan was quadrupled during the past 47 years (1951-1998).

National Institute of Population Studies [NIPS] (2006) reported that Pakistan was ranked 6th among the most populous countries of the world with 156.26 million population and growth rate 1.86% respectively in 2006 whereas Iqbal (2007) reported 158 million population with annual average growth rate 1.83% respectively. According to Nation Master (2008), Pakistan was the 7th populous country of the world with population 172.8 million whereas it would be the 5th populous country of the world in 2050 with 295 million population (Population Reference Bureau [PRB], 2007).

Taking into account the said statistics, the population of Pakistan drastically increased from 34 million to 172.8 million during 1951 to 2008, experiencing the average annual growth rates of Pakistan i.e. 3.66%, 3.05%, and 2.69% during the periods 1961-72, 1972-81 and 1981-98 respectively (Anonymous, 2001). Although, the growth rate of Pakistan has decreased from 2.69% to 1.86% during 1998 to 2006 respectively, even then the current population size of Pakistan is not so less as compared to the other countries of the world (NIPS, 2006).

According to the Population Growth Rate (2008), there are still 156 countries out of 229 in the world having less growth rate as compared to Pakistan. The logic behind this increase is that, Pakistan witnessed a very high growth rate in its early decades after gaining independence. Most of the developing countries of the world like Pakistan are

trying to limit their population size to maintain a balance between the population and the available resources of the country. It is indispensable during this transition period of Pakistan.

Different scientists studied the population from different aspects and projections were made using different methodologies. Stoto (1983) highlighted the importance of population projections of United States and used different statistics to measure the accuracy of population projections. Smith & Sincich (1988) discussed the stability over time in the distribution of population forecast errors. Smith & Rayer (2008) evaluated the accuracy and bias of total population projections as well as the population of different age groups of sub county areas in Florida. Keilman (1998) measured the data accuracy for the world and its major regions of the United Nation projections for the years 1950-1995 using mean error (ME) and mean absolute error (MAE) as an evaluation statistics. Carter (1996) forecasted the United States mortality using the structural time series models and MAPE as an evaluation criteria.

Jan et al. (2007) projected the population of province NWFP of Pakistan for the next 40 years by different growth models and different evaluation statistics were used for model selection. Loh & George (2007) forecasted the population of Canada up to 2056 with and without international migration. Cohen (1986) projected the population of Sweden by using the Model Based and Empirical approaches as well as constructed the confidence intervals.

Fujiwara & Casewell (2002) projected the population of North Atlantic right whale *(Eubalaena glacialis*) using the multi-stage mark recapture (MSMR) to convert transition

probabilities into a matrix population model. Keyfitz (1964) projected the Norwegian brown rats female population using the matrix operator technique.

The population projections are the estimates of total size or composition of populations in the future (United Nations, 1984) whereas the inequality is the disparity among the age distribution. Knowledge about past populations and assumptions about future populations are fundamental to planning decisions in every aspect of community life (Krueckeberg & Silvers, 1974).

The population Projections are based on the current and past trends of the population using different assumptions i.e. the current growth rate will increase, decrease or remain constant. The population is usually projected for short time as well as for long time period, but it should be kept in mind, the longer the period of projection, the greater the errors will be in the assumptions and lesser utilization of the population projections.

The increase in population ultimately puts a pressure on the available resources of the country and demands more food, residence, textile products, transport, educational and health facilities etc. No doubt, the government should be fully aware of the social, economic and health requirements of the communities. The planning and management of different spheres of life concerns with population is impossible without the complete knowledge of the projected population and its distribution. So the population projections must be updated and as much accurate as possible. Information about future population in the form of projections is required at the national as well as the regional level. It helps the policy makers, ministries, planners, and NGO's at the national and the regional level to develop an efficient infrastructure and essential social services especially education and health etc.

Keeping in view the importance of future population, change in age-sex distribution, inequality and fertility; the study of these different aspects of population of Pakistan has been carried out.

The main objectives of the present study are:

- Projection of the population of Pakistan vision 2030
- Projection of the age-sex distribution of population
- Estimation of actual and projected age-sex distribution inequality
- Estimation of the fertility trend of the population

The first objective is achieved by Time series and population growth models. Modified Markov chain modeling technique is also used to project the population.

The second objectives regarding the age-sex distribution of population is attained by Modified Markov chain model.

The third objective about the inequality of age sex distribution is predicted by the Gini Coefficients, confidence intervals and Lorenz curves.

The fourth objective is achieved by computing the cohort measures as well as by modeling the Age Specific Fertility Rates (ASFR).

Outlines of the thesis

The thesis comprises of seven chapters:

Chapter 1 provides a brief introduction of the thesis.

Chapter 2 consists of the related review of the literature.

Chapter 3 presents the findings regarding the accuracy of population census data including Whipple index, Myers Blended index, Bachi index, sex ratio, age ratio and age sex accuracy index to measure the accuracy of the population census. Population is also smoothed by applying different smoothing techniques.

Chapter 4 consists of the population projection of Pakistan using the Time series model Autoregressive integrated moving average (ARIMA) model.

Chapter 5 presents the population projection using traditional growth models and comparison with that of ARIMA model.

Chapter 6 includes the age sex distribution projections of Pakistan by Modified Markov chain modeling as well as the inequality of age sex distribution using the Gini Coefficient technique.

Chapter 7 contains fertility analysis using different useful demographic cohort measures and the modeling of Age Specific Fertility Rates (ASFRs) of women of Pakistan and its geographical regions i.e. rural and urban.

Finally, the thesis ends with a comprehensive summary of the thesis and a list of references for this study.

It is worth mentioning that

1) An article entitled "Forecasting the population of Pakistan using ARIMA models" from the contents of chapter 4 regarding the population of Pakistan using ARIMA model has been published in the HEC recognized journal *Pak. J. Agri. Sci., 46*(3), 2009.
2) An article entitled "Population projections of Pakistan using Traditional and Time series models" from the contents of chapter 5 regarding population projections has been accepted for publication in the HEC recognized journal of Humanities and Social Sciences, XVII(2), 2009.
3) An article entitled "Overtime changes and disparity in the Age-sex distribution of the population of Pakistan" from the contents of chapter 6 regarding population projection has been submitted for publication.
4) An article entitled "Modeling the fertility and reproductivity of Pakistan and its regions during 1984-2005" from the contents of chapter 7 regarding the modeling of fertility pattern of Pakistan has been submitted for publication.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Introduction

The work done in the area of accuracy of the population census data, population projection by different methods, population age-sex distribution projection, population inequality as well as the population fertility has been reviewed. Some comments have also been given wherever necessary.

2.2 Accuracy of Population Census Data

Kemal et al. (2003) pointed out the poor quality of population censuses data of Pakistan and emphasized that it should not be used for further population analysis without the smoothing of the data. Moreover, the strong smoothing techniques should be used instead of weaker one. In this study, the strong smoothing technique is used to smooth the 1998 population census data.

Pullum (2006) advocated the problem of misreporting of ages in many developing countries, although this problem exists in developed countries but with lesser extent since the majority of the people are aware about the importance of their date of birth. It is a fact that the problem of misreporting poses difficulties for the estimation of age sex distribution and birth death rates. The pervasiveness of these problems is illustrated with

that of Pakistan Demographic Survey data and the erroneous age reporting is due to non birthday celebrating societies. The age is usually reported and understood in terms of rounded years rather than completed years.

2.3 Population Projection

McDonald (1979) discussed the relationship among three types of models i.e. classical demographic deterministic models, Time series models and stochastic structural econometric models using the Australian total live births data. The autoregressive moving average (ARMA) model is fitted on the said data. The main interest was to analyze the transforming time series to stationarity and the properties of the forecasts as well as the comparison of actual forecast performance with that of the forecast made by various models. The deterministic cohort models were also used to forecast the total live births. Later on, these forecasts were compared with that of the long run future forecast performance.

Smith & Sincich (1988) evaluated the distribution of population forecast errors of different states in different time periods during the 20th century. To achieve the objectives of the study, the analysis is split into different section. In the 1st section, the degree of the extent of the distribution is determined by which it remains stable over time. Furthermore, the validity of the past forecast errors is measured to predict the distribution of future forecast errors. In the 2nd section, the data and population projection techniques are discussed whereas in the 3rd section, the characteristics and stability of forecast errors over time are studied. In the next section, empirical confidence limits are constructed for the past forecast errors and population projections. At the end, it is concluded that the distribution of absolute parentage errors remained relatively stable over time.

McNown & Rogers (1989) estimated the parameters of ARIMA model to develop the forecast of United States mortality to the year 2000. The data consisted of mortality by age and sex from 1900 to 1985. The forecasts provided the summary of entire mortality distribution and the estimated parameters, a useful instrument and convenient basis for comparing the mortality forecasts at different points in time.

Klosterman (1990) reported that twenty years forecasting is assumed to be long term and two to three years forecasting is a short term. Klosterman also gave the mathematical form of the mean absolute percentage error and coefficient of relative variation.

Smith & Shahidullah (1995) evaluated the accuracy and bias of the total and age group wise population projection for census tracts in three counties in Florida. Population data of 1970 and 1980 were used to project the population for the years 1990. Mean absolute percentage error is computed for the projected total population which range lies 17% to 20% and found no indication of overall bias. Mean absolute percentage error was also computed for each age group which range lies 20% to 29%.

Keilman (1998) projected the population of the world and its seven regions; Africa, Asia, Europe, Northern America, Former Soviet Union, Latin America and Oceania during the years 1950-1990 each with five years of interval. Mean error (ME) and Mean absolute percentage error (MAPE) is computed as an evaluation Statistics. It is also computed for crude death and crude birth rates. It is concluded that the quality of projected data was poor for Africa, Asia, good for Europe and Northern America and relatively good for Oceania regions respectively. Life expectancy at birth was also computed during the years 1965-1990. Keilman also reported that the errors in fertility were greater than the

average for Asia, Northern and Latin America as well as for Oceania whereas the errors in mortality for the world regions, Africa and Asia were relatively difficult to find.

Population Policy of Pakistan (2002) reported 34 million and 144 million population of Pakistan in 1951 and 2001 respectively. It indicated the increase of 108 million population during the last 50 years. Growth rate remained very high during the last 3 decades than the first two decades after independence. Ministry reported that population would be 220 million in 2020 with the current growth rate of the year 2002. Although the growth rate declined up to 2.1, even then, it was very high as compared to the other developing countries. Approximately 33 % of the population of Pakistan is living in poverty. On the other hand, if the growth rate decreased gradually up-to 1.9 during the year 2004, it will be in a replacement level of fertility during 2020. Vision of this policy was to stabilize the demographic transition i.e. declining both the mortality and fertility by the year 2020.

The main objectives of this population policy were to attain the balance between resources and pressure of population, awareness of high growth rate at all levels of the country, promotion of late marriages, spacing interval, small family size and family planning services, reduction in birth rate and population momentum.

The short term objectives of the policy 2002 were the reduction of growth rate and total fertility rate up to 1.9 and 4 respectively during the year 2004 whereas, long term objectives of the policy are; the reduction in growth rate and TFR up to 1.3 and 2.1 respectively during the year 2020, increase in safe family planning method's delivery, training and capacity building of the population, decentralization of responsibilities from federal to district level. Data analysis and research services would be provided by NIPS,

Islamabad. Approximately 49.5 billion rupees will be spent to achieve the long term objectives during the total eighteen years in different segments.

The achievement of such a low growth rate i.e. r =1.3 and TFR = 2.1 during the years 2020 may be an imagination only, not practically possible. Basically, 98 percent of Pakistani population is Muslim, majority of them have strong roots/links of their religion, especially the province NWFP (Tribal areas). Province Sind and NWFP are male dominated societies and prefer a larger family size. The residents of these areas have less acceptability about family planning practicing. They think that the baby birth is a God's will, so the birth should not be stopped in any form.

Maximum percentage of the population of Pakistan lives in the province Punjab but a very little percentage is accepting the family planning practices and small family size. The smaller family size might be the reason of busy schedule of modern era or practicing the demographic transcription theory of low fertility and low mortality. It is also mentioned that this percentage is positively correlated with the literacy rate. In spite of that, the above mentioned drastic decrease in growth rate seems impossible during next 10-20 years. But if the strong, solid and feasible initiatives are taken to increase the literacy rate then it will be possible within one or two generations. Otherwise, the decrease in fertility does not seem possible.

The government should concentrate on the health and education sector emphasizing the schooling education for both sexes, especially females along with psychological training regarding the benefits of small family sizes to promote the family planning services, late marriages and restructure the legislation of child labour etc. In this way, the government may achieve its targets smoothly. Otherwise, it would be the wastage of time and money.

Smith & Tayman (2003) investigated the precision and bias of projected population by age at national and state levels in the United States as well as for the counties of Florida. In this study, the population is projected by two methods i.e. the first one is full blown applications of the cohort component method and the other one is simpler, less data intensive version of method. Later on, the accuracy of State and county projects were compared and concluded that age group patterns were different for national as well as sub-national projections. It was also found that some age groups had substantial larger error than the others. Both the methodologies were equally good and had insignificant impact on the precision and bias of age group projections. On the other hand, when projection horizon became longer, then the difference in errors among age groups declined.

NIPS (2006) projected the population of Pakistan and compared it with developed and developing countries. Out of 6555 million population of world, approximately 5339 million populations belong to less developed and about 1216 million populations to more developed countries. Overall 74.7% population is Non Muslim and 25.3% is Muslim whereas 78.5% population is living in Non Muslim countries and the remaining 21.5% in the Muslim countries. NIPS reported that 3968 million population belong to Asia, out of which 1472 million population from SAARC countries. The growth rates during the years 1951-1961, 1961-1972, 1972-1989, 1989-1998 and 2005 were also reported 2.45, 3.66, 3.05, 2.69 and 1.9 respectively. Furthermore, NIPS also projected and reported 161.86 million, 175.65 million, 189.42 million and 202.11 million population of Pakistan during the years 2010, 2015, 2020, 2025 respectively. It was also mentioned that 7863, 239166, 2870000 persons were added in one day, one month and one year whereas

approximately 5 persons were added in one minute in the world population respectively. The Infant Mortality Rate (IMR) of Pakistan is about 77 which is less than Nigeria but greater than the other Muslim countries like Malaysia, Morocco, Sudan, Turkey and Uzbekistan whereas the life expectancy is about 63.2 and 63.6 years for male and female respectively. It is also mentioned that the Total Fertility Rate (TFR) would be 2.1 and the working population would be 66% in 2020.

Loh & George (2007) revealed the effects of net international migration on the population growth and age sex distribution of the Canadian population during the next 50 years. Canadian Population of 2005 is assumed to be the base population for forecasting purposes and the effects of net international migration are also examined on the provincial age and sex population distribution. Population projection is made up to 2056 with and without considering net international migration. Loh & George concluded that if net international migration continues, the size of the population would be increased but the age sex distribution of Canadian population would not be changed as desired. Percentage of the dependent population having age less than 15 and over 65 is also computed during 2005, 2031 and 2056 considering the net international migration which would be 44.3%, 61.2% and 68.7% respectively. Similarly without net international migration, it would be 68.0% and 80.9% in 2031 and 2056 respectively.

World Population Prospects (2008) reported that the population of Pakistan would be 185 million, 206 million, 226 million, 246 million and 266 million in 2010, 2015, 2020, 2025 and 2030 respectively.

International Data Base [IDB] U.S. Census Bureau (2008) also projected and reported the population of Pakistan which would be 177 million, 191 million, 204 million, 218 million and 231 million up to the years 2010, 2015, 2020, 2025 and 2030 respectively.

Total Population by Country (2009) projected the population of Pakistan for the next 40 years i.e. 204 million, 251 million and 344 million in 2015, 2025 and 2050 respectively and pointed out that Pakistan would be the 4th populous country in the world in 2050 with this population.

People Facts & Figures (2009) revealed that the population of Pakistan would be 268 million whereas Populous Pakistan (2009) reported 292 million population in 2050. Furthermore, according to the United Nations, Pakistan would be the 5th populous country of the world in 2050 with 300 million population.

2.4 Transition Probabilities and Population Inequality

Keyfitz (1964) projected the female population using the matrix approach including the probability of survivals. These probabilities are obtained from the female population of aged 0-45 with five years of interval during the years 1940-1955. Later on, the whole population was divided mainly into three groups i.e. 0-14, 15-29. 30-44 and projection was made for the next 60 years up-to the 2000 by taking one million (10,00,000) population of an age 0-14 as an initial vector.

Nichols et al. (1992) introduced the two approaches to estimate the transition probabilities for stage based projection matrices using capture recapture data. In the first approach, the maximum likelihood parameter estimates were computed from the transition probabilities of interest using SURVIV software. In the second approach, Pollack's robust design is modified. Both the approaches produced approximately similar

results when the ecological data were used. The advantages of these two approaches include the directness of parameter estimation, less restrictive assumptions regarding the independence of survival and growth as well as the testing of ecological interest related hypothesis.

Molenberghs et al. (2001) used the stochastic model on a set of population data of interest and concluded that if incompleteness exists in the data then uncertainty exists but not a full attention is devoted on this issue. Some proposals have been made for assessing the sensitivity to the fitted modeling assumptions; many are based on fitting several plausible but competing models e.g. it is assumed that data are missing at random in one model and then an additional model is fitted where non random missing is assumed. It is also indicated that such an ad hoc procedure may be misleading. One approach is proposed which identifies and incorporates both sources of uncertainty in inference i.e. imprecision due to finite sampling and ignorance to incompleteness. A simple sensitivity analysis considers a finite set of plausible models and this idea has one step further by considering more degrees. It produces a set of estimates and confidence regions.

Keilman (2001) examined the accuracy of United Nations population projection 1950-95 and analyzed the mortality and fertility data as well as the impact of accuracy of data on the projected results. The data about seven regions and ten largest countries of the world were used in the analysis. No doubt, there is considerable variation in the accuracy and data quality among large countries and regions of the world. The major problem was in region Asia as well as in countries like China, Pakistan and Bangladesh. The quality of African data was poor as a whole and Nigera which was the only African country that data had to be adjusted especially in the analysis. It is not surprising that poor data quality

for total fertility rate and life expectancy tend to go together with poor projection performance. It is also mentioned that the data quality is not only the main factor that accounts for the forecast accuracy. Some other variables to account for forecast accuracy are projection length, population size, unexpected development etc.

Fujiwara & Caswell (2002) improved the multistage mark recapture methodology in four major ways. In the first way, the Markov chain formation of the life cycle was used to express the likelihood function in matrix form which makes the numerical calculations simpler. In the second way, the procedure of capture histories incorporation with uncertain stage and sex identification is explained when the information was incomplete. In the second last way, the procedure of the writing the multinomial transition probabilities is introduced as a function of covariates. Lastly, it is described the conversion of estimated transition probabilities into a matrix population models using multistage mark recapture method. These methodologies were applied to the North Atlantic right whale (Eubalaena glacialis) data.

Goesling & Firebaugh (2004) measured the international global health inequality of the past 20 years (1980-2000). The global health inequality is the sum of health inequality among countries and within countries. In the 1st step, the level of health inequalities among the individuals is measured for each country where as in the 2nd step; the level of inequality among the countries is measured. The four well known measures of inequality are used for this purpose i.e. Gini coefficient, Theil index, mean logarithmic deviation (MLD) and squared coefficient of variation. The life expectancy of eight regions of the world except Japan and China is used for the years 1980, 1990 and 2000. The findings indicated that, those countries in which life expectancy is declined, the inequality among

the countries is increased as well as in the countries of Sub Saharan Africa that are faster-than-world-average population growth. On the other hand, the inequality is decreased in South Asian countries, especially in those where the life expectancy is rising.

2.5 Fertility Analysis and Modeling

Kabir & Mosleh Uddin (1987) revealed the fertility transition of Bangladesh along with the trends and determinants of its fertility. During the late 1970s, in most of the developing countries, the fertility had declined except Bangladesh, Even some changes took place in fertility but the results were not so good owing to the poor quality and unreliability of the data. It is a fundamental fact that in using a cross sectional data, caution must be exercised in the use of methodology and interpretation of the estimates. Kabir & Mosleh Uddin also indicated that small changes in fertility may be attributed partly due to early marriages as well as the increased use of contraception. During 1961, the age at marriage of females in Bangladesh was 14 years whereas in 1981, it was 18 years. Similarly, the contraception use was 8% in 1975 whereas 25% in 1983. The main objective of this study was to know the occurrence of fertility transition, its degree and starting time in Bangladesh. Bangladesh retrospective survey 1974 and Bangladesh fertility survey 1975, National impact survey 1968-69 as well as contraceptive prevalence survey (CPS) 1979, 1981, 1983, and 1985 were used to assess the level of fertility. Age specific fertility rates (ASFR) of different periods and regions were also modeled.

Sathar & Kazi (1990) investigated the effects of education and employment of women of Karachi on their fertility pattern and status (autonomy). A survey was conducted and 1000 Karachi women were interviewed with one additional question regarding their outdoor/indoor employment. Analysis indicated that indoor employment/professions and

women education have less effect on the women status and fertility as compared to high paid jobs or professional employments. Mainly, it is concluded that the fertility of the women is affected by the nature of the women profession.

Warren et al. (1992) pointed out the most important fertility affecting factors in Swaziland i.e. nuptiality, contraception, Lactational amenorrhoea/sexual abstinence and abortion. These results are consistent with that of Bongaarts. On the other hand, some less important factors are frequency of intercourse, intrauterine mortality, natural sterility and involuntary infertility caused by sexually transmitted pelvic disease including gonorrhoea. Moreover, it is pointed out that if the prevalence of contraceptive changes, the fertility is surely changed provided that the other proximate determinants of fertility remained constant. This is not necessary the case for an indirect determinants e.g. income as well as education. The comprehension fertility health survey (FHS) in Swaziland was conducted first time in 1988. The main objective of this study was, to find the relative importance of different proximate determinants on fertility and its variation in different subgroups of the population. Study found that post partum sexual abstinence, separation of spouses, labour migration to South Africa and fecund ability seem to be the most important factors which are direct determinant of fertility in the sub Saharan Africa. Warren et al. reported that 20% of men having age 20-39 lived outside the country, out of which 95% men lived in South Africa and half of which worked in South African mines. No doubt, the contraception was less important than nuptiality but it seems to be the most likely method of fertility reducing. In this study, it is emphasized and recommended that Ministry of Health should focus on increasing the family planning programs in order to achieve fertility reduction.

Sathar (1993) studied the reality of fertility using the Pakistan Demographic Health Survey (PDHS) 1990-1991 data and reported the total fertility rate i.e. 5.5 where as it was 6.3 in 1975. On the other hand, according to different surveys, the total fertility rate ranged from 6.0 to 6.9 during 1980s which seems inconsistent with the fertility levels of Pakistan. The Pakistan Demographic Health Survey indicated that the TFR was 5.5 and 5.2 during the period 1985-91, 1990-91 respectively. Pakistan contraceptive prevalence survey (PCPS) also reported the total fertility rate with the decline of 0.5. Pakistan Demographic Health Survey (PDHS) illustrated that fertility rate declined due to the raised female age at marriage. Both the surveys, Pakistan Fertility Survey (PFS, 1985) and PDHS (1991) indicated that fertility declined among 15-19 years old females. Approximately 33 percent decrease in fertility has been examined during 1985-91 as compared to 1970-75 whereas the fertility in other age groups has fallen about 4-15% except the females of age group 45-49. Sathar criticized that in PDHS (1990-91), the use of contraceptive is under reported to report the fertility decline from 5-6 TFR to 4 TFR. Sathar is of the view that the means of fertility control should be more attractive, effective and easily available from the local market.

Paget & Timaeus (1994) revealed the fitting and assessment of relational Gompertz model to male fertility. This model is based on a standard pattern of fertility. Usually in fertility analysis, the attention is paid to female fertility but in this research article, an attempt is made to model the male fertility. The availability of male fertility data is too difficult than that of female fertility data. Fortunately, the United Nation Demographic yearbooks have compiled data from the whole world since 1948 e.g. two from the polygynous i.e. Cameroon and Central African Republic countries with total fertility rates

10.6 and 9.7 respectively, three from high fertility Population e.g. Libya, Kuwait and Pakistan with total fertility rates 9.9, 8.2, 8.6 respectively, two from medium fertility population i.e. Mexico and Trinidad with total fertility rate 3.7 and 3.9 respectively and two from low fertility populations i.e. Hong Kong and France with total fertility rates 2.12 and 2 respectively. Relational Gompertz model with two parameters was fitted to the observed fertility distributions against the male standard which seems to be sufficiently flexible.

Angeles et al. (1998) described that most of the times; the possibility of the distribution of services related to the fertility level in that area is ignored. The factors determining service placement might be related to the determinants of high or low fertility in that area. If this is the situation then it is difficult to account for the endogeneity of family planning services. Consequently, the estimated results are biased. In this study, new modeling approach is introduced in which model extends the simultaneous equation framework by integrating an individual level model of timing and spacing of children with the dynamic process of programme placement. Tanzania demographic health survey 1991-1992 data was used to demonstrate the said approach. It showed that the standard methods yield misleading results on the impact of different components of the family planning programme of the fertility due to the overstated effects of family planning hospitals on birth and understated the access to health centers offering the family planning services.

Sathar et al. (1988) reported that female education, workforce participation and women age at marriage are good indicators of women's status in Pakistan. According to the 1979-80 survey, the above mentioned three measures are significant determinants of fertility. Furthermore, in urban area, the female education, mean age at marriage and

workforce participation are positively associated/correlated variables. Sathar also mentioned that the education of next generation mothers depends on the education level of the parents and regions. Urban educated parents are more liberal to their daughters' education as that of sons' education. It may conclude that age at marriage; education and workforce participation have inverse relationship with that of the national fertility.

Sathar & Casterline (1998) discussed and reported the high fertility trend of Pakistan since 1960. If the current fertility rate continues, Pakistan would become the fourth most populous country of the world in 2050 (U.N. Projection). Pakistan has high total fertility rate than its populous neighbours of South Asia except Nepal. Some analysts criticized the poor management and low quality of the services regarding contraceptive (fertility controlling measures) offered to the majority of its clients. Sathar also pointed out that the whole responsibility of low fertility transition cannot be explicitly explained on the inadequacy of family planning services and country's population policies. Although some other factors also exist in Pakistan which affects the fertility.

Raab & Donnelly (1999) studied the information on sexual behaviour when some data are missing. A survey was conducted at the University of Edinburg in 1993 and the degree class student is considered as survey unit. Approximately 6110 questionnaires were registered to the students during their study period and the response of 3828 students were collected, giving an overall response rate of 62.7%. Out of these 73% respondents reported that they ever had sexual intercourse. The estimate derived from the response-saturated profile likelihood is 67 % with a 95 % confidence interval of 58-74 %. This is in line with other studies on response bias in the reports of young people's sexual behaviour which suggest that the respondents ever represent themselves sexually active.

Hussain & Bittles (1999) discussed the effects of consanguineous marriage on the mean age at marriage, contraceptive use as well as on the fertility in Pakistan. It is a fact that the age specific fertility rate of Pakistan remained high during the past three decades. Hussain & Bittles used different socio economic and demographic variables to measure the fertility pattern of Pakistan and pointed out that little attention had been given to consanguineous in the past. The Pakistan demographic and health survey (PDHS) 1990-91 and multi ethnic communities in Karachi data is used to compare the consanguineous and non consanguineous group of women. The results indicated that female age at first marriage is gradually increased whereas consanguineous marriages are held at younger ages, less likely the use of contraceptive, and have a higher mean number of pregnancies than non consanguineous union. The interesting thing is that the mean number of surviving children did not differ significantly in both groups of Karachi sample. On the other hand, the mean number of children and number of surviving children among women of consanguineous group of PDHS is lower. It is concluded that non consanguineous marriages should be encouraged for further fertility reduction in Pakistan.

Lee & Nelder (2000) studied the human sex ratio data which was collected by Geisster from the Saxony in the 19th century. Data were reanalyzed by joint modeling of the mean and dispersion. Un-normalized double-exponential family and extended quasi likelihood are used to lead to identical inference. This study described the relationship between multinomial and Poisson models using the over dispersed data.

Chen et al. (2000) suggested a Bayesian method for the analysis of toxicological multivariate mortality data when the discrete mortality rate for each family of subjects at

a given time depends on familial random effects as well as the toxicity level experienced by the family. The main aim is to model and analyze one set of such multivariate mortality data with large family sizes; the potassium thiocyanate (KSCN) contained fish tank data of O'Hara Hines. The used model is based on a discrete hazard with additional time varying familial random effects. A similar previous study (using sodium thiocyanate (NaSCN)) is used to construct posterior estimates of the model parameters of mortality rates and several other quantities of interest. Recent tools in Bayesian model diagnostics and variable subset selection have been incorporated to verify important modeling assumptions regarding the effects of time and heterogeneity among the families on the mortality rate. Furthermore, Bayesian methods using predictive distributions are used for comparing several plausible models.

Hinde & Mturi (2000) studied the recent trend of Tanzanian fertility and found the magnitude of Tanzania's fertility decline by using the different robust methods. Tanzanian's fertility declining pattern is similar to Zimbabwe and Kenya, but the urban fertility of Tanzania has more decline than its rural area which indicates the greater use of contraceptive and its prevalence. Before 1990, the contraceptive prevalence and its use were at very low level but currently, the use of contraceptive among the married women is increased and fertility declined. Although, the mean age at marriage is raised but it has a smaller contribution in fertility decline as compared to the contraceptive use. The Tanzanian fertility is still high as compared to the Zimbabwe and Kenya. It might be due to economic crisis which became the main cause of Tanzanian fertility disturbance. Such circumstances happened during the Tanzania-Uganda war in 1978 which cost about $500 million. Foreign trade was distributed badly and foreign aid was almost frozen in such

circumstances. It seems that couples attempted to postpone or prevent further childbearing, especially those couples who have already 3 or 4 children. It might be possible that such economic difficulties forced the community to think about delaying the marriage at a later age.

Bairagi & Datta (2001) revealed the demographic transition of Bangladesh and reported that Bangladesh's international centre for diarrheal disease started a research project. The purpose of this project is to test the hypothesis that fertility can be reduced by maternal and child health family planning (MCH-FP) without considering the socio economic status of the society. The total area was divided into two halves; one half of the area remained under the usual government programme i.e. the provision of contraception and the other half under the program (MCH-FP). MCH-FP programme includes the contraception and abortion. The study indicated that any one of the programs explicitly is insufficient to reduce the fertility in Bangladesh. Both the programs collectively may affect the fertility. Although, the socio-economic status, education, modernization, sex preferences determine the desired level of fertility but MCH-FP program helps to speed up the desired level of fertility. Finally, the family size, sex preference along with contraceptive measures and reproductive health services are the most important factors to determine the fertility level of Bangladesh.

Feeney & Iqbal (2003) reported that population growth rate of Pakistan was 2.6% per annum in the early 1960s which rose up to 3.5% during the late 1980s. During this period of rising growth rate, the total fertility rate was approximately seven 7 children per women. Again it was declined to an estimated growth rate 2.1% during the year 2003. The computed growth rate presented a very drastic picture which might be due to

differential accuracy of enumeration. These conclusions are based on the analysis of fertility and mortality data from all major sources, especially from Pakistan Demographic Survey conducted by the Federal bureau of statistics since 1984. The evidence for the rapid decline of fertility and the population growth from the late 1980s are strong and consistent.

McNay et al. (2003) concluded that the women's education is strongly associated with fertility in India but the interesting and important feature of India's current fertility transition is that the use of contraceptive among uneducated women is higher than educated women. That is why, uneducated women are playing major role in decreasing the fertility of the country. It might be due to the awareness of contraceptive through media. Different multilevel statistical methods were used to investigate the variation in the use of contraceptives among the uneducated women and to highlight the significant socio-economic variables. Analysis also showed that there was significant relationship between the use of contraceptives and education. Furthermore, mass media plays a vital role in the use of contraceptives which is also clear from the variables included in the model.

Islam & Ali (2004) studied the age specific fertility rates of rural Bangladesh during 1980-1998. The aim of this study was to estimate the demographic cohort measures i.e. total fertility rates (TFR), gross reproduction rates (GRR), net reproduction rates (NRR), mean age of childbearing (MAC) and mean length of generation (MLG) etc. The data on age specific fertility rates was taken from Bangladesh bureau of statistics (BSS). Total fertility rate during 1980 was 5.13 whereas in 1998, 3.29 respectively. Similarly the other measures were also decreased i.e. GRR, NRR, MAC etc. Polynomial models were fitted

taking age specific fertility rates as endogenous and age as an explanatory variable. The models were also fitted on forwarded cumulative ASFRs and on backward cumulative ASFRs. Goodness of fit of models was examined by cross validation production power and the coefficient of determination. Both the statistics were almost the same for different models.

Kostaki & Paraskevi (2007) found that age specific fertility pattern of human population was almost common. Different parametric models are recommended to describe the age specific fertility patterns. It is mentioned that fertility pattern of developed countries has changed from classical one but this distortion in age specific fertility pattern is greater and stronger in younger women as compared to the total fertility pattern. The intensity of the heterogeneity is greater among European countries and United States. Ethnic differences in number of births and timing are also related to fertility pattern in United States. It is directly related to marital status, religion, educational level as well as socio economic conditions of the population. Furthermore, it is pointed out that the new recommended models are more flexible than the existing ones to describe the old and new fertility patterns.

Lam & Marteleto (2008) discussed the stages of demographic transition from a child's perspective, family size, cohort size and children's resources of the developing countries over the last 50 years. The transition begins with declining mortality, especially the infant and child mortality. Consequently, the number of surviving children increases at family as well as national level. Mortality decline considered to be the second key element of the transition which becomes the cause of decline in fertility. Ultimately, it affects the family as well as the cohort size. The characterizations of the demographic transition stages are

also introduced and reported that a dramatic change may occur in the number of siblings and cohort size at family as well as national levels. These changes may move in different directions during the stages. In the 1st stage; the surviving children increase due to declining infant and child mortality. In the 2nd stage, the smaller family size persists due to declining fertility and mortality but the cohort size does not decline due to population momentum. In the 3rd stage, fertility decline overtakes population momentum and decreases the absolute birth cohort. The children born in different stages have to face different competitions at different levels. The micro census data was used to develop a simple model for the dynamics of cohort and family size of eight countries i.e. Brazil, Costa Rica, Kenya, Mexico, South Africa etc.

Ruggles & Heggeness (2008) pointed out that co-residence between elderly people and their children have been declined in most of the developing countries. These findings are based on 42nd census data of 15 developing countries. The first objective of the study was to measure the relationship between the younger and older generation. The second objective was to measure the trends in inter-generational co-residence of the house hold headship patterns. A multivariate analysis was applied to assess the relationship between inter-generational co-residence and economic development. Some countries showed increasing trend in co-residence, while the others showed decreasing trend. A clear cut trend had not been examined in inter-generational co-residence over the past several decades. It is also concluded that traditional family norms have greater association in the developing world. Multivariate analysis indicated that economic development had positive relationship with that of the inter-generational families headed by the elders. Furthermore, it is found that life expectancy is strongly associated with that of the inter-

generational residence families headed by the elders and negatively associated with residence in a family headed by a younger person. It is also reported that the percentage of urban areas is positively associated to older headed co-residence among the younger generation. Only two less important Asian countries of the world are used, so the results should be generalized carefully. Almost similar situation exists for other regions of the country.

Nasir et al. (2009) estimated the demographic cohort measures i.e. Total fertility rate, gross reproduction rate, net reproduction rate, mean age of childbearing etc. The age specific fertility rates of Pakistan data was used during the years 1984-2005. Same polynomial models were fitted on the age specific fertility rate, forward and backward cumulative age specific fertility rates data. Goodness of fit of the models is also tested by the same model validation criteria as that of Islam & Ali (2004).

Feldman et al. (2009) studied the use of contraceptive, spacing between pregnancies and the autonomy of the women of the rural Mexico during the year 1997. The main theme of this study was to know the attitude of the women towards contraceptive, birth spacing and the women's autonomy with and without considering the health and education. To achieve the objectives of the study, cash incentive to mothers was given to invest on the health as well as on the education of the family members. Experiment was conducted during 1998, 2000 and 2003. Three controlled groups were also incorporated with that of the experimental group. In 2000, it was seen that the percentage to use contraceptive in experimental group was greater than that of the control group. But in 2003, the percentage to use contraceptive in experimental group and control group was insignificant. It implied that the change in autonomy was not a mediator; the baseline

autonomy modified the program's influence on the contraceptive use. Cox proportional hazard model parameters were estimated using the birth spacing data of both groups. These findings may be helpful for the planners of the family planning and it is indispensible to get the desired results of fertility decline as well as the birth spacing.

Ishida et al. (2009) reported that the total fertility rate of Paraguay was 4.3 during 1955-98 and it was 2.9 during 2001-2004 (Reproductive Health Surveys). It indicates approximately 33 percentage decrease in total fertility rate. A relationship is also established between the two surveys 1998 and 2004 as well as the use of Bongaarts framework of proximate determinants of fertility. The decline in fertility was accounted for from the prevalence of contraceptive. Study also reported that the young mother feed the ideal fertility and is likely to continue to decrease. Ishida et al. highlighted that the most effective and significant agents to decline the fertility rate are education, urbanization, region and mother language (language spoken at home). Guarani women like a larger family size imply the lower contraception rate where as Spanish speaking women like a smaller family size as compared to Guarani women imply the higher rate of contraceptive in Paraguay.

Eltigani (2009) revealed that the two countries i.e. Egypt and Tunisia have almost the same fertility levels and began their fertility transition approximately the same time period. It is also reported that Tunisia reached its replacement level by the year 2007 whereas, the Egypt remained above three live births per women (TFR = 3). These findings are based on several national representative surveys of the two countries during 1978-2005. The impact of contraception was also considered to determine the fertility decline differences of two countries which became an important factor in mid 1980.

Mean age at marriage is not less important than any other factors of fertility decline. In this regard, Tunisia significantly improved itself as compared to the Egypt. Furthermore, the consistency of fertility decline leads to the fertility transition. The influence of political parties in promoting the family planning programmes and their integration within the socio economic development has its own significance. Finally, it is concluded that without the expansion of family planning programmes, encouragement, improvement in family planning services, delivery and the use of these methods, the desired level of fertility decline cannot be achieved.

Steele & Curtis (2003) pointed out that the selection of contraceptive method is endogenous and an important determinant of contraceptive discontinuation. Indonesian Demographic and Health Survey (1997) data are used and the multilevel model was applied to examine the impact of method choice on three types of contraceptive discontinuation. Ignoring the endogeneity of contraceptive choice leads to various biases in the magnitude of estimated effects of method choice on abandonment and method switching, but the general conclusions are robust to these biases.

Elliott & Little (2005) studied the demographic analysis using births, deaths and migration data together with coverage measurement surveys that use capture-recapture methods. United States 1990 Census data are used in this research project and proposed a hierarchical Bayesian model. The presence of bias in the census data is also considered in the refined hierarchical model design. Bayes factors are used in the model selection which was then applied to the entire 2000 United States census data for comparison purposes.

CHAPTER 3

ACCURACY OF POPULATION CENSUS DATA

3.1 Introduction

Population census may be defined as the enumeration of people, houses, firms, or other important items in a country or region at a particular time. The modern periodic census and thorough statistical review began in the 17th century. The first U.S. population census was taken in 1790. Censuses of England, France, and Canada were taken in 1801, 1836, and 1871 respectively. Census information is obtained by using a fixed questionnaire covering such topics as the place of residence, sex, age, marital status, occupation, citizenship, language, ethnicity, religious affiliation and education etc. (Census, 2009).

In Pakistan, five population censuses have been conducted till now. After its independence dated 14th August, 1947, the first population census of Pakistan was held in 1951 and recorded 34 million populations whereas the latest and 5th population census was conducted in 1998. According to this census, the population of Pakistan was 132 million and the growth rate 2.69% respectively (Anonymous, 2001).

It is a proved and established fact that age misreporting is more common problem in developing countries. Although this problem exists in developed countries but with lesser extent as compare to the developed countries. The majority of the people report their ages ending at digits 0, 5 and less percentage on other digits. Usually, the larger age misreporting is recorded in the ages 0-4, 5-9 and on the digits 0 & 5 than the other ages and digits.

Actually, the people do not know the importance of their date of birth and population data. The reason might be the poor registration system of birth and death, home birth deliveries especially in remote areas and less literacy rate as well as we have not established the tradition of celebrating birthday in our country. It is a universal truth that poor information about the problem under study provides poor results.

The population census data is used to achieve the research objectives especially the census data of 1972, 1981 and 1998 of Pakistan. Before using the population census data, it is indispensable to measure the accuracy of the census data. If the data seems accurate then it can be used as such, otherwise it has to be smoothed by some suitable smoothing techniques for further analysis. The accuracy of population census data of Pakistan was examined using the same methodology (Kemal et al., 2003).

3.2 Objectives

The objectives of this chapter are:

- To estimate the extent of errors in age sex distribution of population census data
- To provide the smoothed age and sex distribution of population using various smoothing techniques.

3.3 Methodology

Some common measures of digit preference of census data are available in the literature i.e. Whipple's index, Myers blended index and Bache index etc. Whipple index is used to measure the digit preference at ages ending '0' and '5'. A data set is said to be highly inaccurate if Whipple's index is over 175, inaccurate if its value lies 125-175, fairly acceptable if its value lies 110-125 or less and highly accurate if its value less than 105.

The mathematical form of the Whipple index is

$$\frac{\sum(P_{25}+P_{30}+P_{35}+\cdots\cdots+P_{55}+P_{60})}{1/5\sum(P_{23}+P_{24}+P_{25}+\cdots\cdots\cdots P_{60}+P_{61}+P_{62})} X\ 100 \qquad \text{(Shryock, 1973)}$$

On the other hand, the Myers blended index and Bachi index are used to measure the digit preference at ages ending 0 to 9. The range of Myers blended index is 0 to 90, the index 0 means no digit preference and 90 means a digit preference at a single digit. The computation procedure of Myers blended index is given (Shryock, 1973). If these measures indicate that the population census data is inaccurate, then there is need to apply some suitable smoothing techniques to smooth the population of different age groups. Same measures have been used to measure the accuracy of population census data of Pakistan (Kemal et al., 2003). United States Population Analysis Spreadsheet (PAS) software was used to compute the accuracy indexes of digit preference of population census data (U.S. Census Bureau, 2008).

3.4 Results and discussion

Table 3.1 reveals the Whipple index, Myers blended index and Bachi index for the population census 1972, 1981 and 1998 of Pakistan. Whipple index for 1972

population census is 347 and 346 for male and female respectively whereas the Whipple index for both sexes is 347. Index indicates that 1972 population census data is highly inaccurate and the greater preference is at ages ending 0 or 5. Another measure i.e. Myers blended index for male and female is 76.8 and 78.2 respectively whereas, 77 for both sexes. The Myers blended index is positive and highest at the digits 0 and 5 which indicates the preference at digits 0 and 5.

Similarly, the Whipple index for 1981 and 1998 population census are 332(male), 327(female), 330(both sexes) and 172(male), 201(female) and 186(both sexes) respectively. Although, both these indexes are less than the 1972 indexes, even then the values of Whipple index are very high and greater than the acceptable range. No doubt the values of the Whipple index of 1998 are less than the 1981 but not in the acceptable range. It is a good indication about improvement in age reporting during 1998 as compared to 1981 and 1972. It might be due to the awareness of respondent about age reporting or effort of trained enumerator to reach the truth. Overall the quality of age distribution in the last three censuses remained poor. There is need to redesign the age based questions and trainings for the interviewers before to conduct the census. In this way, the trainers can teach the tactics to reach the truth regarding the age of the respondents during population information collection.

Table 3.2 consists of the United Nations accuracy index, sex ratio score, age ratio score for male and female of the reported and smoothed population of the census 1972, 1981, and 1998. The sex ratio scores for the population census 1972, 1981 and 1998 are 9.76, 8.39 and 4.77 respectively. These scores indicate that 1972 population census data is highly inaccurate as compared to 1981 and 1998. It seems that awareness regarding age

has been increased and people are caring to report their ages. Another measure of age misreporting is known as the age ratio score which is computed for male and female separately. Male and female age ratio scores of the population census 1972 are 24.61 and 19.10 respectively. Similarly, male and female age ratio scores for the population census 1981 are 19.62 and 14.50 respectively. Usually it is assumed that age misreporting in females is higher than males. But in both these census, the male age ratio scores are greater than females which is contradict to the usual expectation/practice. On the other hand, the male and female age ratio scores of 1998 census are 8.71 and 9.68 respectively. It indicates that age misreporting in females is higher than males. According to these measures of accuracy, the quality of population censuses 1972, 1981 and 1998 data is highly poor.

The third measure of census data is known as the United Nation accuracy index. The United Nation accuracy indexes for the population census 1972, 1981 and 1998 are 73.0, 59.3 and 32.7 respectively. Although, the accuracy index is decreased in 1998 as compared to 1981 and 1972 population census, even then the data are highly inaccurate. On the basis of all these measures, it is concluded that the population censuses data are highly inaccurate and it is indispensible to smooth the data before further use. The reported and smoothed data of 1972, 1981, 1998 for male and females separately are given in Tables 3.3, 3.4, 3.5, 3.6, 3.7 and 3.8 respectively.

Table 3.3, 3.5 and 3.7 reveal the reported and smoothed male populations of the census years 1972, 1981 and 1998 respectively. Different smoothing techniques i.e. Carrier Farrag, K-King Newton, Arriaga, United Nations and Strong are used to smooth the population. Except the strong smoothing technique, the other smoothing techniques are

known as light smoothing techniques. Similarly, Table 3.4, 3.6, and 3.8 present the reported and smoothed female population for the same census years respectively.

Kemal et al. (2003) reported the Arriaga and strong smoothed population for the years 1972, 1981 and 1998. Since the population census data of Pakistan is highly inaccurate, the strong smoothing technique might be preferred over the others. For onward analysis in other chapters, the strong smoothed population data will be used.

Table 3.9, 3.10 and 3.11 reveal the sex ratio of reported and smoothed population of 1972, 1981 and 1998 population census. The sex ratio increases gradually up to the age group 10-14 and then decrease and increase at different age groups. The reason might be the understating of age of young girls of ages 10-14 and overstating the age by girls of ages 15-19 (especially who become mothers). The sex ratio of other age groups indicates the less life expectancy of females as compared to the male life expectancy. The reason might be the high rate of fertility and deficiency of quality food nutrition of mothers. The last columns of these Tables are the sex ratios of the strong smoothed population of 1972, 1981 and 1998 census respectively. The sex ratio pattern of strong smoothed population of 1972 and 1981 are approximately same but different from 1998 smoothed sex ratio. The sex ratio in later ages during 1998 is decreased as compared to the 1972 and 1981 sex ratio. It indicates the improvement in female mortality, better health facilities and look after of females of older ages.

Table 3.12, 3.14, and 3.16 present the age ratios of males of the reported and smoothed population of census 1972, 1981 and 1998 respectively whereas the Table 3.13, 3.15, and 3.17 present the female age ratios of the reported and smoothed population for the same census respectively. Age ratios indicate the high distortion in age reporting for males as

Figure 3.8 and 3.9 show the trend of digit preference at ages ending 0 to 9 for the population census 1998. The Myers blended index value at digit preference 0 and 5 are less as compared to 1981 and 1972 population census index. The trend of preference on other digits is almost similar to that of 1972 and 1981 population census. It indicates that people are more caring to report their ages and understand the importance of age, although misreporting of ages still exists but with lesser extent. The improvement in male age reporting is more than the female age reporting.

3.5 Conclusion and Recommendations

It is concluded that the population census 1972, 1981 and 1998 are inaccurate and it cannot be used as such for onward use. The value of Whipple index for 1972 census is 347 for both sexes whereas the acceptable range is 110-125 which shows the preference at ages ending 0 and 5. The Whipple indexes are also computed for 1981 and 1998 census which is 330 and 186 respectively? It indicated that 1972 census data is highly inaccurate as compared to 1981 and 1998. Although both these indexes are less than 1972 but are outside the acceptable range. Another measures i.e. Myers blended index and Bachi index are also computed and almost same conclusions are drawn. The tendency at the digit 0 and 5 is reduced in 1998 as compared to 1981 and 1972 even then the people have a tendency at ages ending 0 and 5 to report their ages.

Age misreporting is also measured by using the sex ratio, age ratio and United Nations accuracy index. The same conclusions are obtained about the quality of data. An interesting thing is that greater misreporting is recorded in male than female during the census 1972 and 1998 whereas in 1998 census the female misreporting is greater than male. No doubt the departure from exact age also exists in developed society but with

less severity. Since the quality of Pakistan census data are very poor, it indicates that data should be smoothed before further use. Resultantly, different smoothing techniques are used to smooth the data especially the strong smoothing techniques is preferred for Pakistan census data.

In 1998 census, people showed more awareness in reporting their date of birth. It is a good indication which might be due to better training of the enumerators and increased literacy rate. In the light of these results, it is suggested that the age based question should be improved and repeat with some alteration for cross checking, better and tactful trainings be arranged for enumerators, the importance of census data should be publicized by the governments. Government should motivate and improve the confidence of the people so that they can give correct information during population census.

Table 3.1 Measuring the Digit Preference of Population Censuses 1972-1998

	1972			1981			1998		
Method of digit preference	Male	Female	Both Sexes	Male	Female	Both Sexes	Male	Female	Both Sexes
WHIPPLE INDEX	347	346	347	332	327	330	172	201	186
MYERS INDEX	76.8	78.2	77.1	70.1	71.9	70.9	27.2	36.7	31.8
0	22.5	23.1	22.7	19.8	20.3	20.0	6.9	9.8	8.2
1	-7.6	-7.7	-7.6	-6.9	-7.4	-7.1	-3.3	-4.3	-3.8
2	-0.9	-1.4	-1.1	-0.7	-0.5	-0.6	0.8	0.4	0.6
3	-6.2	-6.3	-6.2	-5.2	-5.8	-5.5	-1.8	-2.4	-2.1
4	-5.7	-5.6	-5.6	-5.2	-5.2	-5.2	-1.7	-2.4	-2.0
5	15.9	15.4	15.7	14.7	14.5	14.6	3.8	5.3	4.5
6	-3.8	-4.1	-3.9	-3.6	-3.8	-3.7	-1.8	-2.6	-2.2
7	-6.4	-6.5	-6.4	-6.2	-6.0	-6.1	-2.5	-3.3	-2.9
8	-0.3	0.6	0.1	0.6	1.1	-0.8	2.2	2.9	2.5
9	-7.5	-7.5	-7.5	-7.2	-7.3	-7.3	-2.6	-3.3	-2.9
BACHI INDEX	51.4	51	51.2	48.3	47.1	47.7	16	22.3	19.0
0	31.2	31.1	31.1	28.5	27.6	28.1	9.4	12.9	11.1
1	-8.7	-8.8	-8.8	-8.1	-8.5	-8.3	-3.5	-4.7	-4.1
2	-3.6	-4.0	-3.8	-3.7	-2.7	-3.2	-0.1	-0.7	-0.4
3	-7.4	-7.4	-7.4	-6.1	-6.7	-6.4	-1.7	-2.5	-2.1
4	-7.6	-7.4	-7.5	-7.1	-7.1	-7.1	-2.2	-3.1	-2.6
5	20.1	19.9	20.0	19.8	19.6	19.7	5.7	7.9	6.8
6	-5.8	-6.1	-5.9	-5.7	-6.0	-5.8	-2.2	-3.1	-2.6
7	-7.4	-7.5	-7.5	-7.2	-6.8	-7.0	-2.5	-3.4	-2.9
8	-2.6	-1.3	-2.0	-2.1	-1.0	-1.6	0.8	1.5	1.1
9	-8.4	-8.4	-8.4	-8.3	-8.2	-8.3	-3.8	-4.9	-4.4

Myers blended index is the sum of the absolute values of the deviations. Bachi index is the sum of the positive of the deviations (one half the sums of the absolute deviations).

Table 3.2 Accuracy Measures of the Population Censuses 1972-1998

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
1972						
Sex ratio score	9.76	4.56	5.32	4.57	3.95	2.80
Male age ratio score	24.61	2.40	2.41	2.47	4.86	1.41
Female age ratio score	19.10	2.01	2.27	2.11	4.35	1.74
Accuracy index	73.00	18.09	20.66	18.29	21.07	11.55
1981						
Sex ratio score	8.39	5.07	5.41	5.02	4.89	3.45
Male age ratio score	19.62	2.79	2.93	2.92	3.98	1.57
Female age ratio score	14.50	3.29	3.36	3.24	3.59	1.95
Accuracy index	59.30	21.28	22.51	21.23	22.25	13.86
1998						
Sex ratio score	4.77	3.47	3.75	3.46	3.53	1.20
Male age ratio score	8.71	2.10	2.01	2.27	1.82	1.48
Female age ratio score	9.68	2.57	2.54	2.63	2.10	1.54
Accuracy index	32.70	15.09	15.80	15.27	14.50	6.64

Note: The accuracy index is the sum of the male and female age ratio scores plus three times the sex ratio score, all calculated using data for ages 0-14 through 65-69.

Table 3.3 Reported and Smoothed Male Population Census 1972

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	33,393,646			33,393,646		33,393,646
Total, 10-69	22,274,590	22,274,590	22,274,590	22,274,590	22,389,554	22,274,590
0-4	4,725,325			5,375,266		5,412,270
5-9	5,316,861			4,666,920		4,629,916
10-14	4,384,059	3,982,431	3,974,545	3,969,221	4,354,467	3,879,251
15-19	2,909,927	3,311,555	3,319,441	3,324,765	3,017,001	3,287,037
20-24	2,350,945	2,592,227	2,616,095	2,584,041	2,406,884	2,721,549
25-29	2,450,404	2,209,122	2,185,255	2,217,308	2,339,593	2,329,403
30-34	2,056,573	2,042,344	2,040,671	2,039,930	2,095,870	2,008,129
35-39	1,790,693	1,804,922	1,806,596	1,807,336	1,811,272	1,747,620
40-44	1,645,256	1,587,516	1,582,317	1,583,360	1,585,882	1,526,732
45-49	1,283,493	1,341,233	1,346,432	1,345,389	1,391,134	1,313,571
50-54	1,318,614	1,066,583	1,072,949	1,063,544	1,137,475	1,091,820
55-59	641,572	893,603	887,237	896,642	885,710	927,245
60-64	1,041,546	775,449	776,734	773,590	796,569	775,598
65-69	401,508	667,605	666,320	669,465	567,697	666,635
70-74	524,067	3,982,431		577,919		575,294
75+	552,803	3,311,555		498,952		501,576

Table 3.4 Reported and Smoothed Female Population Census 1972

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	29,068,237			29,068,237		29,068,237
Total, 10-69	18,765,873	18,765,873	18,765,873	18,765,873	18,911,998	18,765,873
0-4	4,688,162			5,295,025		5,237,172
5-9	4,814,625			4,207,762		4,265,615
10-14	3,451,121	3,218,356	3,255,608	3,210,572	3,535,174	3,359,233
15-19	2,423,195	2,655,960	2,618,708	2,663,744	2,491,996	2,778,732
20-24	2,211,540	2,350,794	2,355,787	2,345,339	2,202,370	2,330,980
25-29	2,196,040	2,056,786	2,051,793	2,062,241	2,153,595	2,006,389
30-34	1,903,303	1,846,282	1,842,801	1,843,124	1,896,533	1,765,469
35-39	1,539,054	1,596,075	1,599,556	1,599,233	1,589,547	1,524,646
40-44	1,417,332	1,354,462	1,349,906	1,348,740	1,350,577	1,289,049
45-49	1,044,292	1,107,162	1,111,718	1,112,884	1,125,451	1,079,748
50-54	994,174	850,583	857,083	848,103	883,849	862,933
55-59	542,682	686,273	679,773	688,753	685,605	710,724
60-64	730,718	564,075	567,650	562,438	584,611	572,005
65-69	312,422	479,065	475,490	480,702	412,691	485,965
70-74	379,646			419,811		421,367
75+	419,931			379,766		378,210

Table 3.5 Reported and Smoothed Male Population Census 1981

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	43,089,811			43,089,811		43,089,811
Total, 10-69	28,618,507	28,618,507	28,618,507	28,618,507	28,727,354	28,618,507
0-4	6,200,434			6,837,736		6,947,949
5-9	6,811,487			6,174,185		6,063,972
10-14	5,856,744	5,492,814	5,452,798	5,472,077	5,819,577	5,195,810
15-19	4,192,513	4,556,443	4,596,459	4,577,180	4,295,519	4,406,716
20-24	3,269,776	3,388,207	3,426,886	3,380,312	3,299,291	3,578,730
25-29	2,891,427	2,772,996	2,734,317	2,780,891	2,827,049	2,993,937
30-34	2,388,124	2,409,659	2,417,705	2,403,302	2,420,140	2,466,736
35-39	2,120,580	2,099,045	2,090,999	2,105,402	2,125,349	2,112,104
40-44	1,937,256	1,904,528	1,899,487	1,901,342	1,891,880	1,846,585
45-49	1,610,303	1,643,031	1,648,072	1,646,217	1,713,972	1,601,378
50-54	1,637,892	1,349,721	1,354,512	1,346,029	1,438,859	1,364,472
55-59	859,488	1,147,659	1,142,868	1,151,351	1,136,074	1,177,688
60-64	1,299,090	989,373	992,077	986,911	1,020,897	999,581
65-69	555,314	865,031	862,327	867,493	738,748	874,770
70-74	677,869			768,738		771,027
75+	781,514			690,645		688,356

Table 3.6 Reported and Smoothed Female Population Census 1981

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	38,965,286			38,965,286		38,965,286
Total, 10-69	25,189,890	28,618,507	28,618,507	28,618,507	28,727,354	28,618,507
0-4	6,373,470			6,926,268		6,940,294
5-9	6,330,850			5,778,052		5,764,026
10-14	4,946,304	5,492,814	4,705,852	4,680,513	4,983,580	4,649,284
15-19	3,570,574	4,556,443	3,811,026	3,836,365	3,650,268	3,842,173
20-24	2,957,980	3,388,207	3,036,044	2,999,975	2,939,845	3,119,236
25-29	2,587,731	2,772,996	2,509,667	2,545,736	2,561,176	2,635,987
30-34	2,229,204	2,409,659	2,287,441	2,280,617	2,253,990	2,263,552
35-39	2,076,657	2,099,045	2,018,420	2,025,244	2,083,809	1,960,297
40-44	1,927,768	1,904,528	1,835,946	1,844,324	1,868,156	1,714,580
45-49	1,465,779	1,643,031	1,557,601	1,549,223	1,553,234	1,455,428
50-54	1,327,725	1,349,721	1,167,344	1,155,174	1,206,298	1,175,066
55-59	751,369	1,147,659	911,750	923,920	912,282	966,952
60-64	917,301	989,373	737,401	727,972	755,825	762,300
65-69	431,498	865,031	611,398	620,827	536,220	645,034
70-74	483,556			551,396		560,969
75+	587,520			519,680		510,107

Table 3.7 Reported and Smoothed Male Population Census 1998

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	67,221,639			67,221,639		67,221,639
Total, 10-69	45,263,089	45,263,089	45,263,089	45,263,089	45,418,375	45,263,089
0-4	9,761,275			10,704,639		10,815,059
5-9	10,570,613			9,627,249		9,516,829
10-14	8,909,137	8,543,139	8,511,636	8,524,377	8,964,683	8,243,424
15-19	6,909,333	7,275,331	7,306,834	7,294,093	7,033,786	7,094,142
20-24	5,814,957	5,845,260	5,867,489	5,827,531	5,759,974	5,932,940
25-29	4,878,521	4,848,218	4,825,989	4,865,947	4,925,662	5,006,092
30-34	4,232,271	4,071,657	4,080,918	4,059,853	4,131,759	4,115,994
35-39	3,254,204	3,414,818	3,405,557	3,426,622	3,372,160	3,461,214
40-44	2,930,509	2,877,201	2,881,575	2,868,839	2,833,082	2,897,138
45-49	2,360,081	2,413,389	2,409,015	2,421,751	2,460,370	2,432,300
50-54	2,200,655	2,048,455	2,041,907	2,038,851	2,069,974	2,005,594
55-59	1,505,344	1,657,544	1,664,092	1,667,148	1,644,918	1,658,707
60-64	1,418,158	1,250,352	1,263,997	1,247,403	1,289,024	1,325,544
65-69	849,919	1,017,725	1,004,080	1,020,674	932,985	1,090,001
70-74	777,588			860,320		892,780
75+	849,074			766,342		733,882

Table 3.8 Reported and Smoothed Female Population Census 1998

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
Total, 0-79	61,954,309			61,954,309		61,954,309
Total, 10-69	41,608,644	41,608,644	41,608,644	41,608,644	41,788,889	41,608,644
0-4	9,356,857			10,119,665		10,142,892
5-9	9,644,403			8,881,595		8,858,368
10-14	7,822,462	7,693,170	7,692,980	7,676,495	7,977,105	7,611,060
15-19	6,490,279	6,619,571	6,619,761	6,636,246	6,562,299	6,549,825
20-24	5,772,812	5,698,495	5,682,319	5,678,122	5,664,391	5,535,411
25-29	4,642,699	4,717,016	4,733,192	4,737,389	4,709,049	4,678,253
30-34	3,807,310	3,665,704	3,697,201	3,655,655	3,731,612	3,835,119
35-39	2,912,417	3,054,023	3,022,526	3,064,072	3,047,963	3,209,204
40-44	2,814,964	2,739,193	2,727,826	2,729,848	2,678,563	2,661,295
45-49	2,203,152	2,278,923	2,290,290	2,288,268	2,306,095	2,212,212
50-54	1,947,600	1,801,788	1,803,158	1,792,692	1,833,874	1,800,301
55-59	1,271,840	1,417,652	1,416,282	1,426,748	1,404,852	1,464,116
60-64	1,219,020	1,066,100	1,078,744	1,063,794	1,094,733	1,135,402
65-69	704,089	857,009	844,365	859,315	778,353	916,447
70-74	630,583			714,639		739,586
75+	713,822			629,766		604,819

Table 3.9 Sex Ratios of the Reported and Smoothed Population Census 1972

Age Group	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
0-4	100.8			101.5		103.3
5-9	110.4			110.9		108.5
10-14	127.0	123.7	122.1	123.6	123.2	115.5
15-19	120.1	124.7	126.8	124.8	121.1	118.3
20-24	106.3	110.3	111.0	110.2	109.3	116.8
25-29	111.6	107.4	106.5	107.5	108.6	116.1
30-34	108.1	110.6	110.7	110.7	110.5	113.7
35-39	116.4	113.1	112.9	113.0	113.9	114.6
40-44	116.1	117.2	117.2	117.4	117.4	118.4
45-49	122.9	121.1	121.1	120.9	123.6	121.7
50-54	132.6	125.4	125.2	125.4	128.7	126.5
55-59	118.2	130.2	130.5	130.2	129.2	130.5
60-64	142.5	137.5	136.8	137.5	136.3	135.6
65-69	128.5	139.4	140.1	139.3	137.6	137.2
70-74	138.0			137.7		136.5
75+	131.6			131.4		132.6

Table 3.10 Sex Ratios of the Reported and Smoothed Population Census 1981

Age Group	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
0-4	97.3			98.7		100.1
5-9	107.6			106.9		105.2
10-14	118.4	116.9	115.9	116.9	116.8	111.8
15-19	117.4	119.3	120.6	119.3	117.7	114.7
20-24	110.5	112.6	112.9	112.7	112.2	114.7
25-29	111.7	109.3	109.0	109.2	110.4	113.6
30-34	107.1	105.5	105.7	105.4	107.4	109.0
35-39	102.1	103.9	103.6	104.0	102.0	107.7
40-44	100.5	102.9	103.5	103.1	101.3	107.7
45-49	109.9	106.5	105.8	106.3	110.3	110.0
50-54	123.4	116.5	116.0	116.5	119.3	116.1
55-59	114.4	124.7	125.3	124.6	124.5	121.8
60-64	141.6	135.5	134.5	135.6	135.1	131.1
65-69	128.7	139.8	141.0	139.7	137.8	135.6
70-74	140.2			139.4		137.4
75+	133.0			132.9		134.9

Table 3.11 Sex Ratios of the Reported and Smoothed Population Census 1998

Age Group	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
0-4	104.3			105.8		106.6
5-9	109.6			108.4		107.4
10-14	113.9	111.0	110.6	111.0	112.4	108.3
15-19	106.5	109.9	110.4	109.9	107.2	108.3
20-24	100.7	102.6	103.3	102.6	101.7	107.2
25-29	105.1	102.8	102.0	102.7	104.6	107.0
30-34	111.2	111.1	110.4	111.1	110.7	107.3
35-39	111.7	111.8	112.7	111.8	110.6	107.9
40-44	104.1	105.0	105.6	105.1	105.8	108.9
45-49	107.1	105.9	105.2	105.8	106.7	109.9
50-54	113.0	113.7	113.2	113.7	112.9	111.4
55-59	118.4	116.9	117.5	116.8	117.1	113.3
60-64	116.3	117.3	117.2	117.3	117.7	116.7
65-69	120.7	118.8	118.9	118.8	119.9	118.9
70-74	123.3			120.4		120.7
75+	118.9			121.7		121.3

Table 3.13 Female Age Ratios of the Reported and Smoothed Population Census 1972

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
5-9	118.3			98.9		99.2
10-14	95.4			93.4		95.4
15-19	85.6	95.4	93.3	95.9	86.9	97.7
20-24	95.8	99.8	100.9	99.3	94.8	97.4
25-29	106.7	98.0	97.7	98.5	105.1	98.0
30-34	101.9	101.1	100.9	100.7	101.3	100.0
35-39	92.7	99.7	100.2	100.2	97.9	99.8
40-44	109.7	100.2	99.6	99.5	99.5	99.0
45-49	86.6	100.4	100.7	101.3	100.7	100.3
50-54	125.3	94.9	95.7	94.1	97.6	96.4
55-59	62.9	97.0	95.4	97.7	93.4	99.1
60-64	170.9	96.8	98.3	96.2	106.5	95.6
65-69	56.3			97.9		97.8
70-74	103.7			97.6		97.5

Table 3.14 Male Age Ratios of the Reported and Smoothed Population Census 1981

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
5-9	113.0			100.3		99.9
10-14	106.4			101.8		99.2
15-19	91.9	102.6	103.5	103.4	94.2	100.4
20-24	92.3	92.5	93.5	91.9	92.6	96.7
25-29	102.2	95.7	93.6	96.2	98.9	99.0
30-34	95.3	98.9	100.2	98.4	97.7	96.6
35-39	98.1	97.3	96.9	97.8	98.6	97.9
40-44	103.8	101.8	101.6	101.4	98.6	99.5
45-49	90.1	101.0	101.3	101.4	102.9	99.7
50-54	132.6	96.7	97.1	96.2	101.0	98.2
55-59	58.5	98.1	97.4	98.7	92.4	99.6
60-64	183.6	98.3	99.0	97.8	108.9	97.4
65-69	56.2			98.8		98.8
70-74	101.4			98.7		98.7

Table 3.15 Female Age Ratios of the Reported and Smoothed Population Census 1981

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
5-9	111.9			99.6		99.5
10-14	99.9			97.4		96.8
15-19	90.3	99.1	98.5	99.9	92.1	98.9
20-24	96.1	94.7	96.1	94.0	94.7	96.3
25-29	99.8	95.9	94.3	96.4	98.6	97.9
30-34	95.6	100.3	101.0	99.8	97.1	98.5
35-39	99.9	97.7	97.9	98.2	101.1	98.6
40-44	108.8	103.9	102.7	103.2	102.7	100.4
45-49	90.0	102.5	103.7	103.3	101.0	100.7
50-54	119.8	94.1	94.5	93.4	97.9	97.0
55-59	66.9	97.4	95.7	98.1	93.0	99.8
60-64	155.1	94.9	96.8	94.3	104.4	94.6
65-69	61.6			97.1		97.5
70-74	94.9			96.7		97.1

Table 3.16 Male Age Ratios of the Reported and Smoothed Population Census 1998

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
5-9	113.2			100.1		99.9
10-14	101.9			100.8		99.3
15-19	93.9	101.1	101.6	101.6	95.5	100.1
20-24	98.7	96.4	96.7	95.8	96.3	98.1
25-29	97.1	97.8	97.0	98.4	99.6	99.6
30-34	104.1	98.6	99.2	97.9	99.6	97.2
35-39	90.9	98.3	97.8	98.9	96.8	98.7
40-44	104.4	98.7	99.1	98.1	97.1	98.3
45-49	92.0	98.0	97.9	98.7	100.4	99.2
50-54	113.9	100.6	100.3	99.7	100.8	98.0
55-59	83.2	100.5	100.7	101.5	97.9	99.6
60-64	120.4	93.5	94.7	92.8	100.0	96.4
65-69	77.4			96.9		98.3
70-74	91.5			96.3		97.9

Table 3.17 Female Age Ratios of the Reported and Smoothed Population Census 1998

Age and Sex	Reported	Smoothed				
		Carrier Farrag	k-king Newton	Arriaga	United Nations	Strong
5-9	112.3			99.8		99.8
10-14	97.0			98.9		98.8
15-19	95.5	98.9	99.0	99.4	96.2	99.6
20-24	103.7	100.5	100.1	99.8	100.5	98.6
25-29	96.9	100.7	100.9	101.5	100.2	99.9
30-34	100.8	94.3	95.3	93.7	96.2	97.2
35-39	88.0	95.4	94.1	96.0	95.1	98.8
40-44	110.1	102.7	102.7	102.0	100.1	98.2
45-49	92.5	100.4	101.1	101.2	102.2	99.2
50-54	112.1	97.5	97.3	96.5	98.8	97.9
55-59	80.3	98.9	98.3	99.9	95.9	99.7
60-64	123.4	93.7	95.4	93.1	100.3	95.4
65-69	76.1			96.6		97.8
70-74	88.9			96.0		97.2

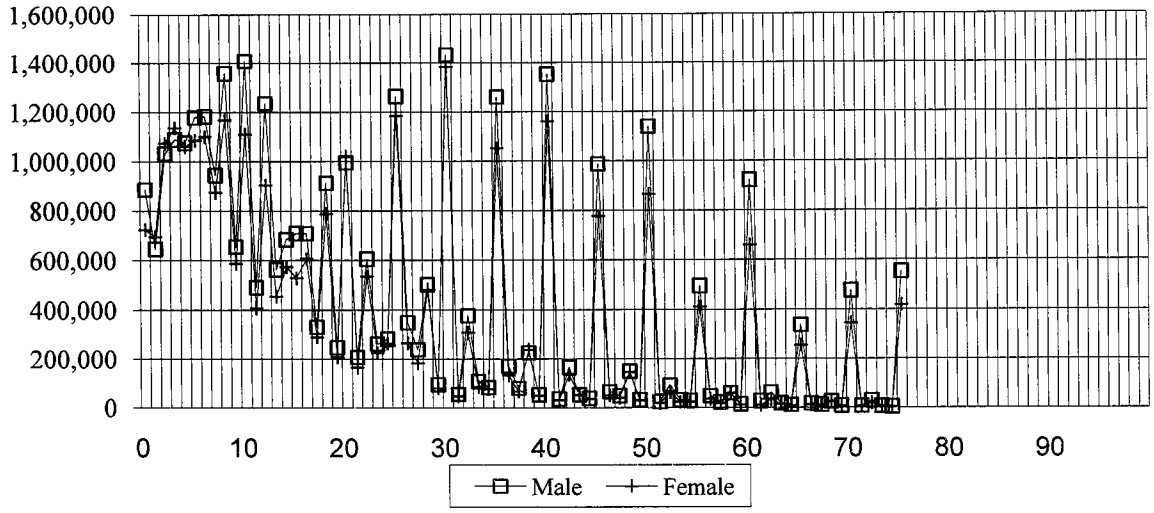


Figure 3.1 Age and sex distribution of population of Pakistan 1972

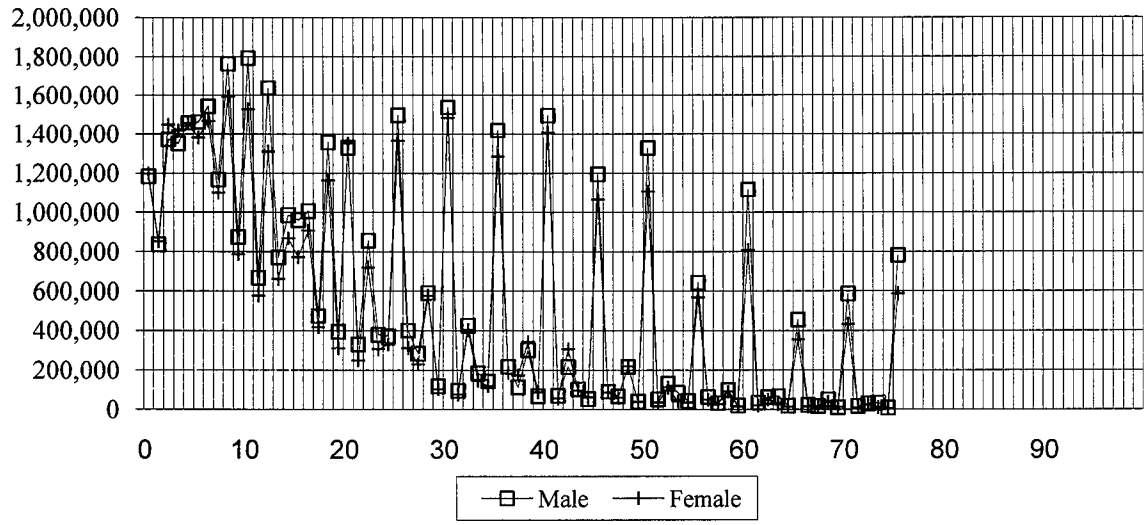


Figure 3.2 Age and sex distribution of population of Pakistan 1981

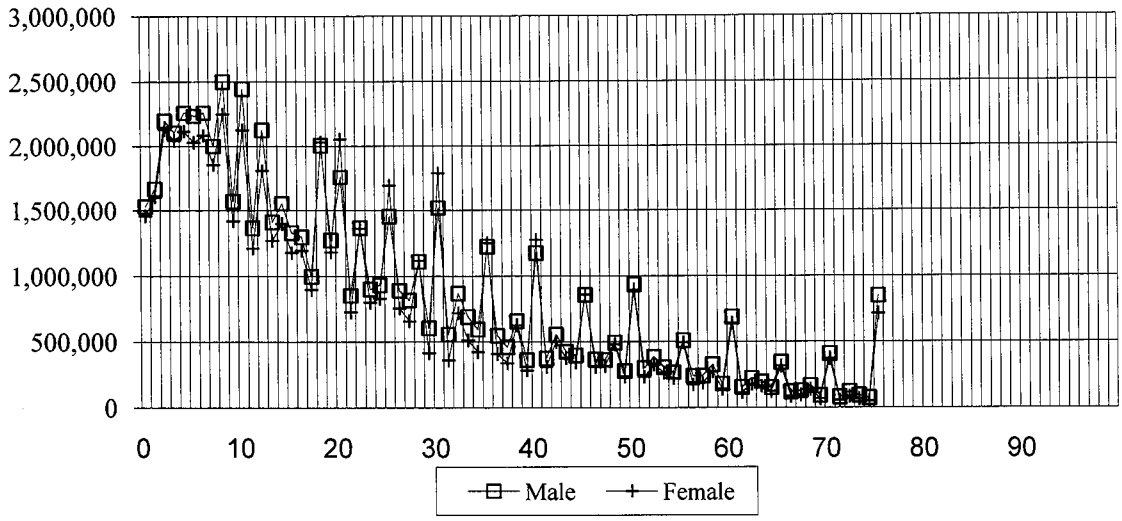


Figure 3.3 Age and sex distribution of population of Pakistan 1998

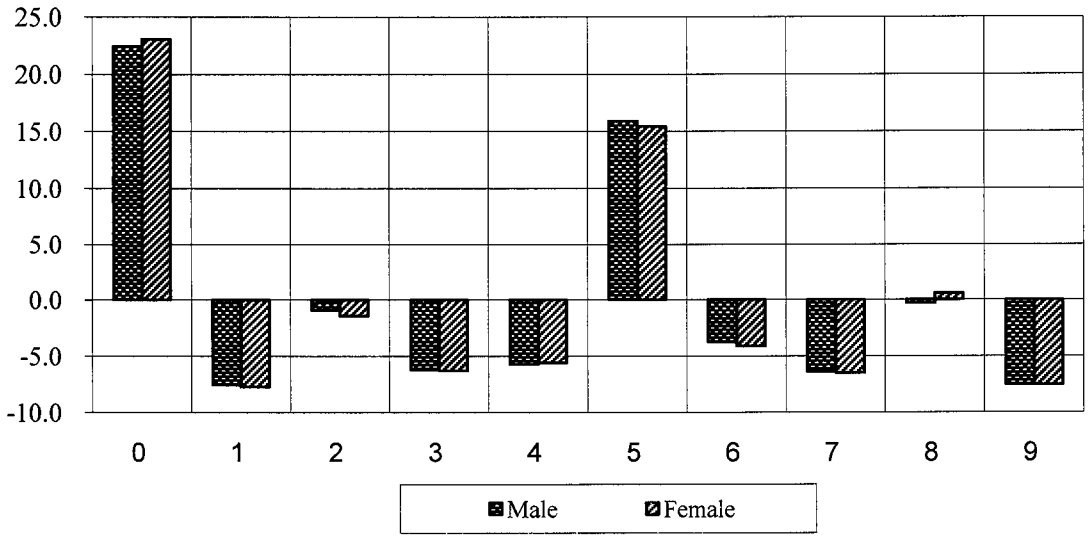


Figure 3.4 Myers Digit Preferences in the Population Census 1972

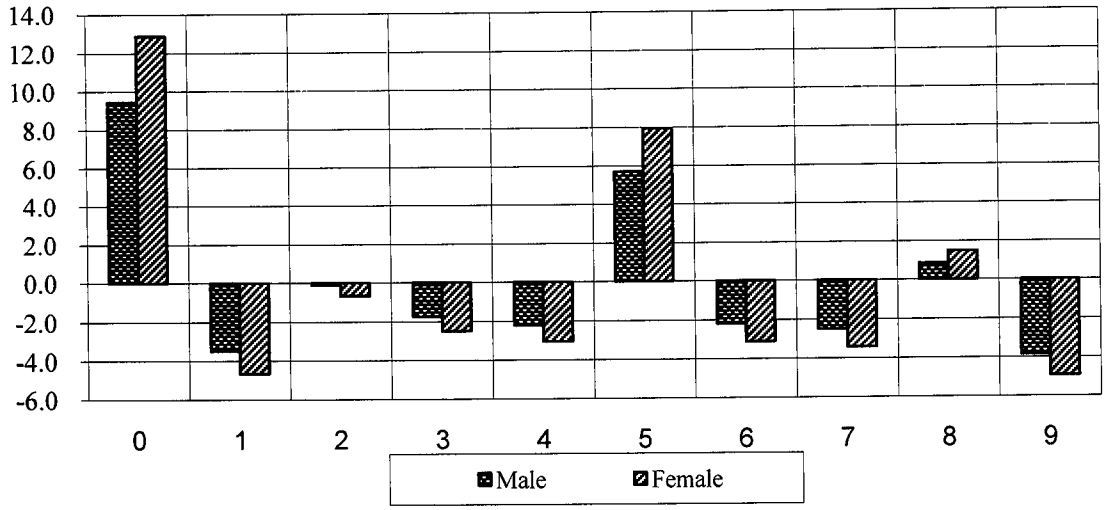


Figure 3.9 Bachi Digit Preference of Population of Pakistan 1998

CHAPTER 4

FORECASTING USING TIME SERIES MODELS

4.1 Introduction

It is unanimously accepted that with the increase of population, the number of associated problems are generated like food and accommodation, education, medical, traffic etc. So much so, the criminal offence rate also raise among such societies due to heavy pressure of the population. Different measures and strategies are being adopted by multicultural societies of the world to limit the size of population according to their feasibility and circumstances. The logic behind is that the current and future economy as well as planning of the country depends on the population.

The governments and industrialists of the world emphasize on the accuracy of population statistics that is inevitable for future planning. Each and every product of the industry, even the wastage of the industry is used by the population in one or the other way. Furthermore all products are designed for national and international population. It might be medical items, food items, gas, petrol, gold etc. In addition, the employments of population in different departments like educational institutions, industrial units, stock market and infrastructure depends upon the age and sex distribution of that country. It seems that age and sex distribution of the population is directly related to the

consumption of industrial products and plays a significant role in improving a GNP of the country. So the population data is of great importance for national and international governments, Non Governments Organizations and industrialists etc.

Developed countries are much more aware about their past, current and future population trends, sizes as well as needs. On the other hands, developing countries are least aware of their population and needs. Only those countries are in good economic and social status which is more aware about the current and future population distribution. It is fact that without the optimum knowledge of the population, no country can be on the right track of scientific and technological development in today's world. Keeping in view the importance of population distribution, population census is conducted after every 10 years in most of the countries but some surveys are also conducted after every five years to update population data.

Population data being a most important asset in the past and present era, the scientists focus on determining the significant trends of the population distribution. That is why; mostly publications are on population trends, size, and age and sex distribution. To forecast the population, different linear, nonlinear, first and higher degree regression models, simple and double exponential, logistic regression, simple decay and growth models are being used (Shryock et al., 1973: Jan et al., 2007 & Agrawal, 2000). Component method of population projection is most widely used method in which data on fertility, mortality and migration is used (Srinivasan, 1998). Autoregressive integrated moving average (ARIMA) model is also used to model the mortality and fertility data.

4.2 Objectives

The objectives of this chapter are:

- Testing the stationary of the time series data
- Modeling the time series data of population
- Projection of population for vision 2030
- Goodness of fit of the models

4.3 About the data

The data used in this chapter consist of the population of Pakistan for both sexes spread over the past 57 years (1951-2007) on yearly basis. Most of the data are taken from "an analysis of 1998 population and housing census" (Kamal et al., 2003) and some of the data is taken from (Iqbal, 2007). Eveiw-5 and Minitab-14 Statistical Softwares are used for model fitting and forecasting.

4.4 Methodology

The ARIMA model technique is not common to forecast the population. Verbeek (2005) gives the following general form of ARMA (p, q) model

$$Y_t = \delta + \phi_1 Y_{t-1} + \ldots\ldots\ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots\ldots\ldots + \theta_q \varepsilon_{t-q}$$

The major steps involved are: model identification, fitting, validation and forecasting. The procedure adopted for model selection in this study is: to check the stationarity of the series as well as for model identification, different time series plots i.e. ACF and PACF are constructed using actual, differenced and transformed data. Box Cox transformation is used to transform the population data (Box & Jenkins, 1976). After the identification of

the model, different stochastic ARIMA models are fitted on the 2nd differenced logarithmic population series. For the selection of the parsimonious model, different model validation statistics are recommended i.e. Mean squared error (MSE), AIC and P-values etc. These statistics are computed for each candidate model and the model having smallest AIC is recommended as a parsimonious model assuming that it is to be closest to the unknown reality by which the series is generated (Burnham et al., 2002). Similarly graphical validation approaches are also applied e.g. normal probability plot, histogram, residual plots, and PACF, ACF plots of the residuals for the confirmation of parsimonious model and its comprehensive graph is given in 4.13. After selecting a parsimonious model, population of Pakistan is forecasted for the next 20 years along with confidence interval. Both the forecasted and fitted population is plotted on the same graph to examine the model adequacy and is given in Figure 4.15.

4.5 Results and Discussion

Figure 4.1 presents the time series plot of original population of Pakistan from 1951 to 2007. It exhibits upward increasing trend and suggests that the given time series is non stationary. Figures 4.2 and 4.3 depict the trend of the population after taking the 1st and 2nd difference of the original population from 1951 to 2007 respectively. Figure 4.2 is unusual whereas the Figure 4.3 indicates that the second differenced logarithmic series is approximately stationary.

Table 4.1 consists of the values of sample ACF, sample PACF, student's t statistics, modified Box-Pierce (Ljung Box) Chi-square statistics, and P-values corresponding to different lags ranging from 1 to 14. ACF of the actual population given in Table 4.1 declines very slowly from high correlation to low correlation during the years 1951-2003.

ACF value at lag 1 is 0.9518 which is very high as compare to the value of ACF at lag 14 i.e. 0.2939 which is not too low. The correlogram of the sample ACF of the original series given in Figure 4.4 indicates high positive correlations at lower lags and low positive correlations at higher lags; Moreover, it has an exponential decay which is an indication that the given population series is non stationary series. The most striking feature of this correlogram is that the autocorrelation coefficients at different lags are very high and out of the confidence limits.

One of the major advantages of the correlogram is that it helps in determining the p, q values of the ARIMA model. Figure 4.4 is the autocorrelation function (ACF) whereas Figure 4.5 is the sample partial autocorrelation function (PACF). The correlogram given in Figure 4.5 has only one spike out of the limits at lag 1 which clearly suggests that an Autoregressive Stochastic model of parameter one seems to be suitable for said series. Box Cox transformation gave the values of $\lambda = 0.22$ whose interval (-0.41, 0.85) contains the value zero (Chatfield, 1996). It suggests that the log transformation is appropriate choice to make the series stationary. It will use before to take the difference of the series. Figure 4.6 shows the trend of the natural logarithm population of Pakistan from 1951 to 2007. The behaviour of the logarithmic population slightly differs from the Figure 4.1. Figure 4.7 and 4.8 present the trend of the population after taking the first and second difference of the natural logarithm of the population series during the same period. No doubt, Figure 4.7 is unusual as that of Figure 4.2 but Figure 4.8 is approximately stationary and behaves more sophisticated way than that of Figure 4.3. Therefore, the 2nd differenced natural logarithmic population series is being used for onward analysis.

Figure 4.9 is the sample autocorrelation function (ACF) of the 2nd difference of natural logarithm of the population of Pakistan. Almost all the spikes at different lags are within the 95% confidence limits; this is an indication that the selected parsimonious model might be without moving average components. Figure 4.10 is the sample partial autocorrelation function (PACF) of the same series used in Figure 4.9. All spikes at different lags of the Figure 4.10 are within the 95% confidence limits except two spikes, one at lag 1 and second at lag 5. The spike at lag 5 is clearly out of the positive limit whereas spike at lag 1 is close to the negative limit; other spikes at different lags in PACFs are clearly within the 95 % limits. The values of sample ACF; sample PACF, Students t statistics, Ljung Box statistics, and P-values corresponding to the different lags from 1 to 14 of 2nd difference of natural logarithmic population series are given in Table 4.2. The characteristics of this correlogram are totally different as that of the characteristics of correlogram of the actual population given in Figure 4.4 and 4.5. It is concluded that all the autocorrelation coefficients (ACF) and partial autocorrelation coefficients (PACF) given in Table 4.2 do not differ significantly from zero and consequently the 2nd difference of the natural logarithm of the population series seems to be stationary. This indicates that different stochastic stationary models can be studied on this series.

Table 4.3 consists of the parameter estimates and goodness of fit measures of fourteen ARIMA models. Last three rows of Table 4.3 present three different models but they provide almost the same information. According to the scientific approach, if more than one model provides same information, the researcher should recommend the insignificant model which has minimum number of parameters as it makes easy the estimation and

interpretation of the model parameters. If the P value corresponding to an estimate of model parameter is less than 0.05, the hypothesis that the parameter equal to zero is rejected. Similarly, if the P value corresponding to an estimate of model parameter is greater than 0.05, the hypothesis that the parameter equal to zero is not rejected which suggests that the explanatory variable should not be included in the model. In time series studies, the selection of the parsimonious model is an art not a mathematical science. Actually, in real phenomena, it is tried to model the real situation by relaxing the level of significance. The model ARIMA (1, 2, 0).W is declared as parsimonious model which is listed at the end of the Table 4.3.

Figure 4.11 and 4.12 are the autocorrelation and sample partial autocorrelation functions of the residuals. These residuals are computed after fitting the model ARIMA (1, 2, 0) W on the 2nd differenced natural logarithm population of Pakistan. All the autocorrelation and partial autocorrelations of the residuals at different lags are within the 95% confidence limits which strengths the recommended parsimonious stochastic time series model.

Figure 4.13 presents the different graphical measures for the adequacy of the model. The first measure is the normal probability plot of the residuals which is not as good as required for an adequate model. Although some of the residuals in this plot are in scattered positions but most of the residuals are on the straight line. Second measure of model adequacy is the histogram of the residuals which does not show exact normality of the residuals but luckily the majority of the residuals lie at centre. Some of the residuals are very large which complicate the situation or the main cause of difficulty. Third and fourth measures are the plot of residuals Vs fitted values and order of the data

respectively. Almost all of the residuals are within acceptable limits which indicate the adequacy of the recommended model.

Table 4.4 consists of the forecasted population for the next 20 years including the years 2010, 2015, 2020, 2025, and 2027 using different ARIMA models. On the basis of goodness of fit criteria, again the same three models provide the consistent results and out of which ARIMA (1, 2, 0) W is the simplest one which strengths the recommended parsimonious model.

Table 4.5 consists of the forecasted population using ARIMA (1, 2, 0) W model for the next 20 years. The second column is the forecasted population in the form of natural logarithm whereas third column is the forecasted population (in million) from 2008 to 2027. Fourth and fifth columns are the lower limits and upper limits of the forecasted population during the years 2008-2027. If the current growth rate remains continue, population of Pakistan would be approximately 230.7 million with 95 % confidence limits 193.3 million & 275.24 million in 2027.

Figure 4.14 presents the comparative trend of last three models of Table 4.4 which are ARIMA, State Space ARIMA and GARCH models. Overall the forecast with these three time series models are almost same during the next 20 years but after 10 years in 2017, these models slightly differ from each other. The forecast by GARCH model is slightly less than State Space model whereas forecast with State Space model is little bit less than the recommended ARIMA model. Figure 4.15 indicates the trend of the actual and forecasted population along with the confidence interval for the next 20 years.

4.6 Conclusion and Recommendations

In this chapter, population of Pakistan of 57 seven years (1951-2007) is modeled using Box and Jenkins ARIMA methodology. The Akaike information criteria's value for the model ARIMA (1, 2, 0) W is -9.46. Although AIC value is approximately same as that of the other fitted models given in Table 4.3 but this model has minimum number of parameters. The model P-value is 0.022 which indicates the independency and randomness of the residuals. If the current growth rate continues, the population of Pakistan would be approximately 230.7 million with 95% confidence limits (193.33 million, 275.25 million) in 2027 using the said parsimonious model. The ACF and PACF of the residuals also strengthened the recommendation of parsimonious model.

The forecasted population using different models is given in Table 4.4 for next specific years i.e. 2010, 2015, 2020, 2025, and 2027. The forecasted population would be 228.73 million in 2027 using ARIMA (1, 2, 0) W with State Space Kalman filter approach. Forecasted population using the State Space Kalman filter estimation approach is almost same as that of the model ARIMA (1, 2, 0) W above. According to the parsimonious model, there will be 74.29% increase in the Population till 2027 with respect to the population census 1998 whereas 45.74% increase in population as compared to the estimates (Iqbal, 2007).

The forecasted population by parsimonious model is close to the projected population by different following bureaus. WPP (2006) estimates are slightly higher than the estimates of ARIMA (1, 2, 0) W. Approximately 7 million people are more in each of the first three estimates and about 3 million people are less in the last estimates. Our estimates are close to the estimates (PRB 2007; U. S. Census Bureau 2008; Pakistan Reality 2008 &

Population of Pakistan 2008). These bureaus reported approximately 229 million population of Pakistan for the year 2025. PRC (2007) reported that population of Pakistan would be 228 million by 2025 which is almost same as our projected population. Rauf Textile & Printing Mills (2008) estimated 213 million population of Pakistan for the year 2025. Rauf's estimates are greater than the estimates of NIPS (2006) but less than all other above mentioned estimates. The ARIMA methodology is simple, easy to apply and requires minimum information in population projection.

In short, the estimates provided in Table 4.4 using ARIMA (1, 2, 0) W are consistent and equally important for the government of Pakistan as well as Non Government Organizations for future planning and projects.

Table 4.1 ACF and PACF of the Actual Population of Pakistan

LAG	ACF	PACF	T-STAT	LB-STAT	P-VALUE
1	0.9518	0.9518	7.1863	54.4090	0.0000
2	0.9029	-0.0327	4.0653	104.2620	0.0000
3	0.8534	-0.0325	3.0568	149.6180	0.0000
4	0.8025	-0.0409	2.4945	190.4820	0.0000
5	0.7512	-0.0326	2.1155	226.9750	0.0000
6	0.6996	-0.0324	1.8315	259.2450	0.0000
7	0.6477	-0.0324	1.6040	287.4620	0.0000
8	0.5957	-0.0327	1.4129	311.8180	0.0000
9	0.5437	-0.0325	1.2467	332.5330	0.0000
10	0.4929	-0.0216	1.1004	349.9150	0.0000
11	0.4424	-0.0306	0.9673	364.2210	0.0000
12	0.3923	-0.0303	0.8440	375.7200	0.0000
13	0.3428	-0.0303	0.7285	384.7000	0.0000
14	0.2939	-0.0298	0.6189	391.4560	0.0000

Table 4.2 ACF and PACF of 2nd Difference of Logarithmic Population

LAG	ACF	PACF	T-STAT	LB-STAT	P-VALUE
1	-0.2666	-0.2666	-1.9774	4.1273	0.9947
2	0.1292	0.0626	0.8967	5.1149	0.9841
3	0.1180	0.1805	0.8071	5.9543	0.9676
4	0.1201	0.2068	0.8118	6.8407	0.9407
5	0.2993	0.4084	1.9995	12.4569	0.5697
6	-0.3167	-0.2213	-1.9772	18.8754	0.1698
7	0.2599	-0.0295	1.5183	23.2880	0.0557
8	0.0475	0.0230	0.2667	23.4386	0.0535
9	0.0481	0.0303	0.2696	23.5964	0.0512
10	0.0401	0.0416	0.2246	23.7086	0.0497
11	0.0307	0.1747	0.1717	23.7758	0.0488
12	0.0249	-0.1839	0.1393	23.8210	0.0482
13	0.0156	-0.0585	0.0872	23.8392	0.0479
14	0.0104	-0.0330	0.0583	4.1273	0.0478

Table 4.3 Parameter Estimates and Goodness of Fit of Different ARIMA (p, d, q) Models

ARIMA (p,d,q)	$Y_t = \delta + \phi_1 Y_{t-1} + \ldots\ldots\ldots\ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots\ldots\ldots + \theta_q \varepsilon_{t-q}$							
ARIMA(5,2,1),W	$\emptyset_1$	$\emptyset_2$	$\emptyset_3$	$\emptyset_4$	$\emptyset_5$	θ_1	θ_2	Intercept
Coefficients (Coeff.s P-value)	-0.633 (0.001)	-0.138 (0.324)	0.331 (0.018)	0.667 (0.000)	0.801 (0.000)	-0.243 (0.291)		-6.87558E-05 (0.919)
MSE 0.0000028, AIC -9.908, BIC -9.641, D.W. Stat. 1.951, P-value 0.391								
ARIMA(5,2,1) (Coeff.s P-value)	-0.732 (0.001)	-0.181 (0.250)	0.192 (0.214)	0.471 (0.006)	0.654 (0.000)	-0.421 (0.104)		
MSE 0.0000032, AIC -9.898, BIC -9.669, D.W. Stat. 1.940, P-value 0.844								
ARIMA(2,2,2),W (Coeff.s P-value)	1.174 (0.000)	-0.166 (0.328)				1.731 (0.000)	-0.866 (0.000)	-2.28224E-05 (0.764)
MSE 0.0000034, AIC-10.164,BIC -9.978, D.W. Stat. 2.208, P-value 0.077								
ARIMA (2,2,2) (Coeff.s P-value)	0.699 (0.004)	-0.317 (0.160)				-0.317 (0.000)	-0.812 (0.000)	
MSE 0.0000039, AIC -9.793, BIC -9.64, D.W. Stat. 2.034 P-value 0.331								
ARIMA(2,2,0),W (Coeff.s P-value)	-0.253 (0.0740)	0.068 (0.653)						0.000 (0.834)
MSE 0.0000045, AIC -9.458, BIC -9.347, D.W. Stat. 2.058, P-value 0.025								
ARIMA(2,2,0) (Coeff.s P-value)	-0.253 (0.0710)	0.065 (0.6620)						
MSE 0.0000044, AIC -9.495, BIC -9.421, D.W. Stat. 2.058, P-value 0.039								
ARIMA(1,2,1),W (Coeff.s P-value)	1.002 (0.000)					1.047 (0.000)		-7.45839E-07 (0.977)
MSE 0.0000047, AIC -9.531, BIC 8.669, D.W. Stat. 2.866, P-value 0.001								
ARIMA(1,2,1) (Coeff.s P-value)	-0.360 (0.487)					-0.093 (0.865)		
MSE 0.0000044, AIC -9.522, BIC -9.448, D.W. Stat. 2.698, P-value 0.028								
ARIMA(1,2,0) (Coeff.s P-value)	-0.273 (0.042)							
MSE 0.0000044, AIC -9.501, BIC -9.464, D.W. Stat.1.982, P-value 0.034								
ARIMA(0,2,1),W (Coeff. P-value)						0.213 (0.120)		0.000 (0.882)
MSE 0.0000044, AIC -9.442, BIC -9.369, D.W. Stat. 2.032, P-value 0.016								
ARIMA(0,2,1) (Coeff. P-value)						0.213 (0.115)		
MSE 0.0000045, AIC -9.478, BIC -9.441, D.W. Stat. 2.029, P-value 0.026								
ARIMA(1,2,0),W (Coeff.s P-value)	-0.270 (0.046)	$Y_t^* = \emptyset_\circ + \emptyset_1 Y_{t-1}^* + \varepsilon_t$						0.000 (0.954)
MSE 0.0000044, AIC -9.464, SC -9.390, D.W. Stat.1.981, P-value 0.022								

ARIMA(1,2,0) State Space (Coeff.s P-value)	$Y_t^* = \varphi + \emptyset_1 Y_{t-1}^* + \varepsilon_t$	
	φ	$\emptyset_1$
	-12.367 (000)	-0.272 (0.000)
	MSE 0.0000043, AIC -9.455, SC -9.382, HQ -9.427	

ARIMA (1,2,0) W + GARCH (0,2) (Coeff.s P-value)	$Y_t^* = \emptyset_\circ + \emptyset_1 Y_{t-1}^* + \gamma_\circ + \omega_1 \sigma_{t-1}^2 + \omega_1 \sigma_{t-2}^2$				
	$\emptyset_\circ$	$\emptyset_1$	$\gamma_\circ$	ω_1	ω_1
	-2.19E-05 (0.907)	-0.389 (0.000)	1.92E-07 (0.000)	1.868 (0.000)	-0.923 (0.000)
	AIC -9.567, SC -9.383, D. W. Stat. 1.718				

Table 4.4 Projected Populations for the Years 2010, 2015, 2020, 2025, 2027

Model	2010 (million)	2015 (million)	2020 (million)	2025 (million)	2027 (million)
ARIMA 5,2,1),W	165.49	177.78	188.22	196.29	198.90
ARIMA (5,2,1)	166.09	180.57	195.76	211.91	218.69
ARIMA (2,2,2),W	160.72	165.40	171.44	171.98	234.31
ARIMA (2,2,2)	167.57	184.98	204.13	225.26	234.31
ARIMA(2,2,0),W	167.48	184.20	184.20	223.66	232.66
ARIMA(2,2,0)	167.42	183.84	201.86	221.65	230.10
ARIMA(1,2,1),W	165.99	175.26	179.29	177.65	175.40
ARIMA(1,2,1)	167.33	183.56	201.37	220.91	229.24
ARIMA(0,2,1),W	167.20	183.33	201.19	22097	229.47
ARIMA(0,2,1)	167.16	183.10	200.56	219.67	227.81
ARIMA(1,2,0)	167.27	183.39	201.07	220.46	228.72
ARIMA(1,2,0),W	167.31	183.67	201.82	221.99	230.68
STATE SPACE ARIMA (1,2,0) W	167.27	183.40	201.08	220.46	228.73
ARIMA(1,2,0),W + GARCH (0,2)	167.24	183.23	200.64	219.59	227.62

*W indicates the inclusion of intercept in the model.

Table 4.5 Projected Population for the Years 2008-2027 by ARIMA (1, 2, 0) W

YEAR	NATURAL LOG OF PROJECTED POPULATION	PROJECTED POPULATION (million)	LOWER LIMIT OF PROJECTED POPULATION (million)	UPPER LIMIT OF PROJECTED POPULATION (million)
2008	18.90	161.23	160.56	161.90
2009	18.92	164.24	162.89	165.60
2010	18.94	167.31	165.10	169.55
2011	18.95	170.45	167.23	173.73
2012	18.97	173.65	169.28	178.36
2013	18.99	176.92	171.26	182.77
2014	19.01	180.26	173.18	187.63
2015	19.03	183.67	175.04	192.72
2016	19.05	187.15	176.84	198.10
2017	19.07	190.70	178.58	203.64
2018	19.09	194.33	180.28	209.48
2019	19.10	198.04	181.92	215.60
2020	19.12	201.82	183.51	221.96
2021	19.14	205.69	185.05	228.63
2022	19.16	209.64	186.55	235.58
2023	19.18	213.67	188.10	242.85
2024	19.20	217.79	189.40	250.43
2025	19.22	222.00	190.75	258.35
2026	19.24	226.29	192.06	266.62
2027	19.26	230.68	193.33	275.25

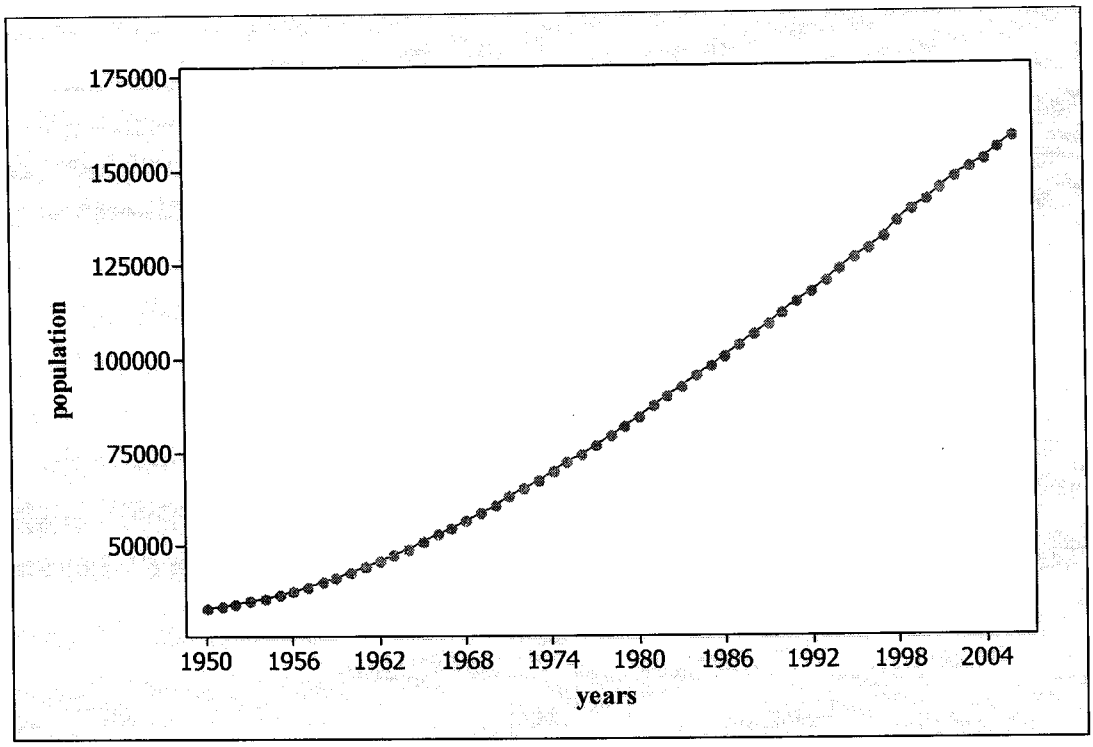


Figure 4.1 Trend of Population of Pakistan During 1951-2007

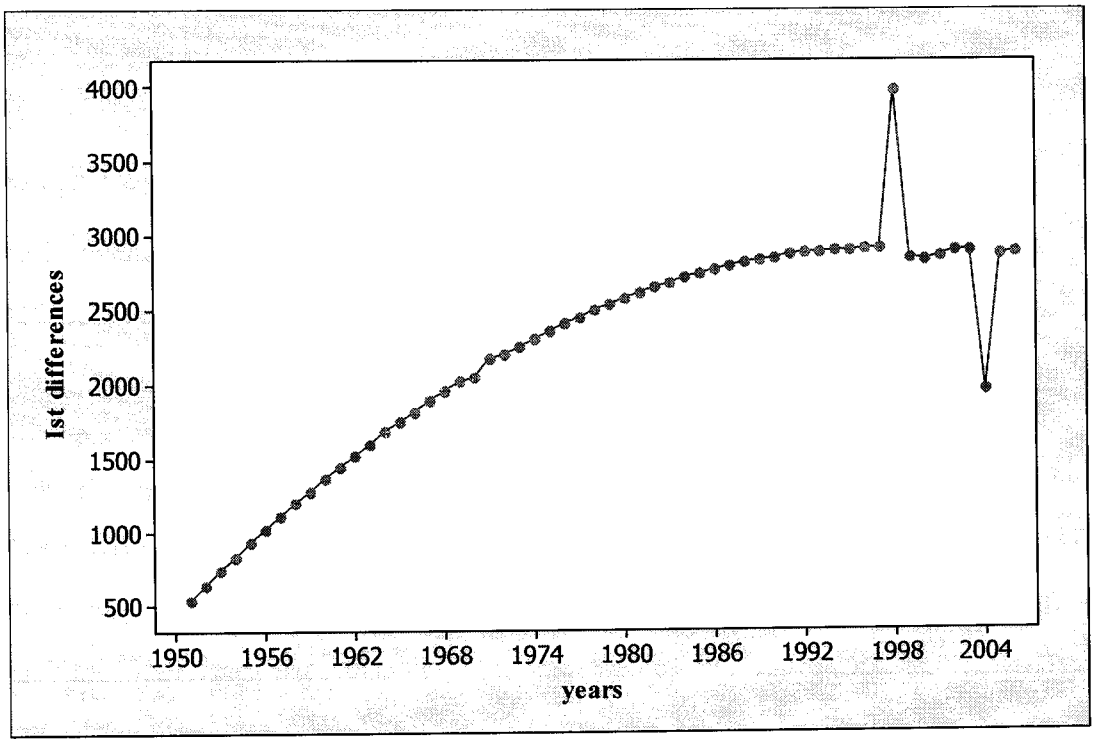


Figure 4.2 Trend of Population of Pakistan after 1st Differencing

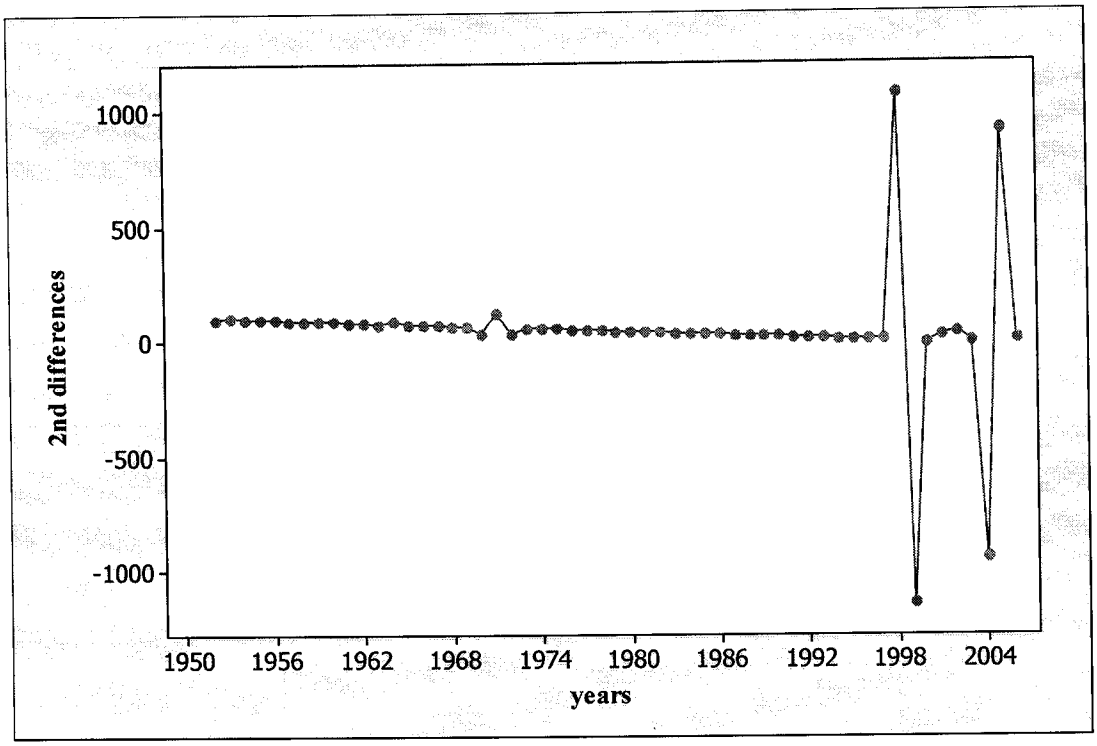


Figure 4.3 Trend of Population of Pakistan after 2nd Differencing

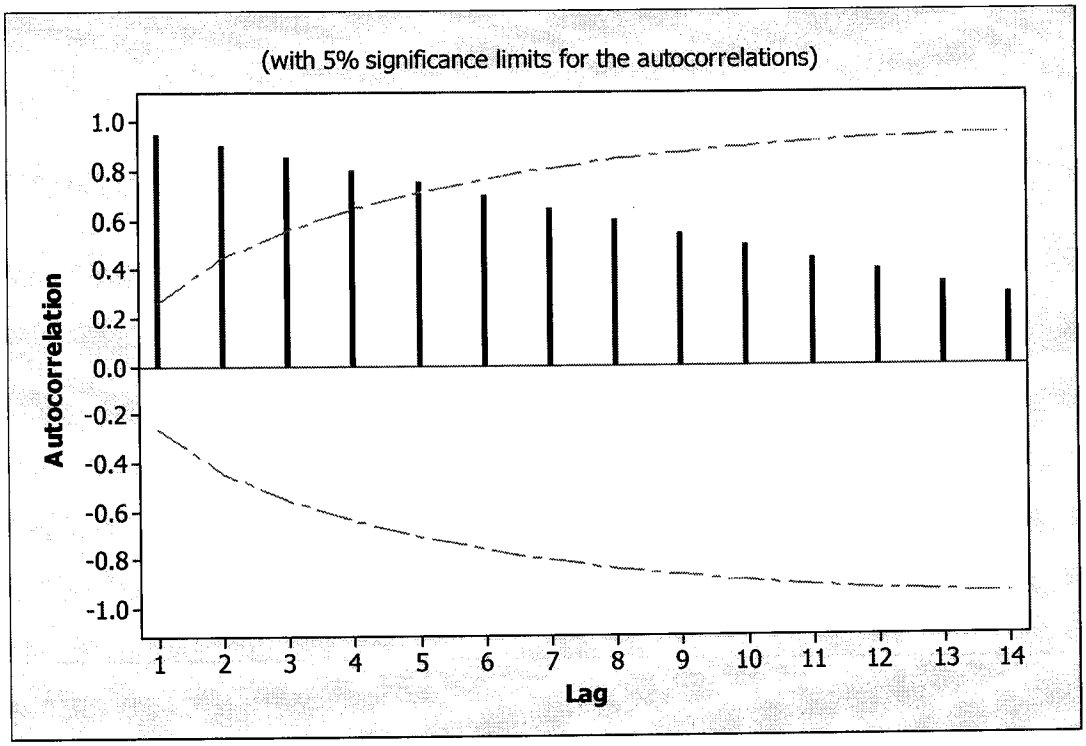


Figure 4.4 Autocorrelation Function of Population of Pakistan 1951-2007

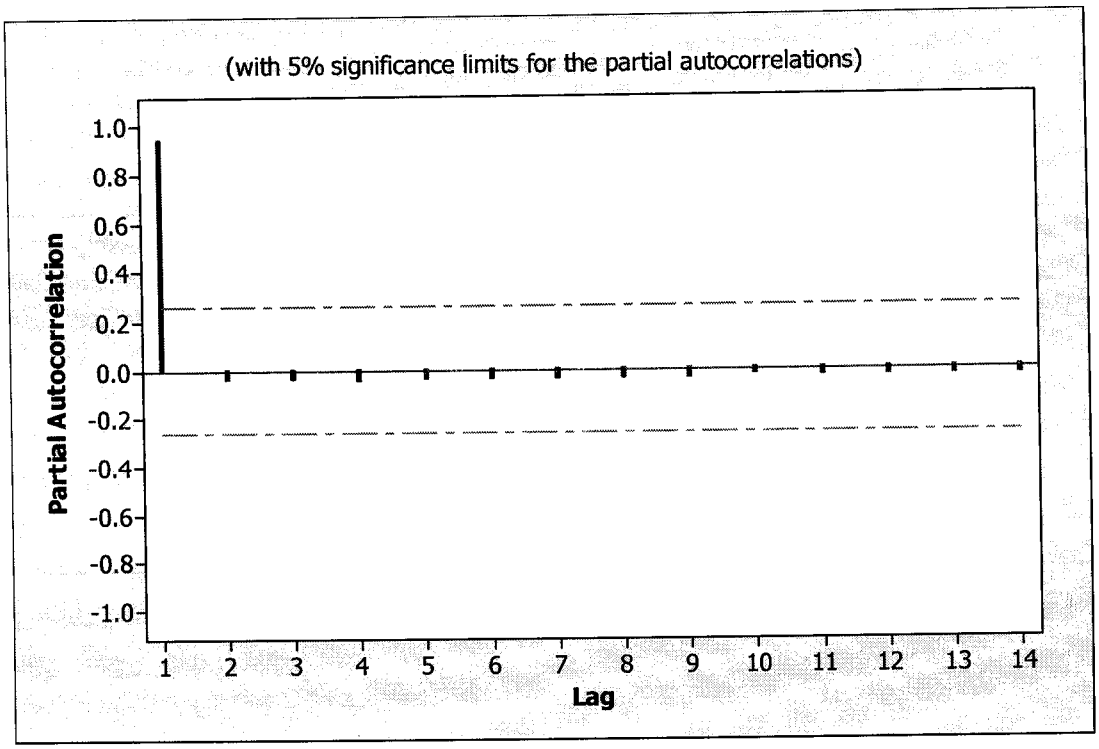


Figure 4.5 Partial Autocorrelation Function of Population of Pakistan 1951-2007

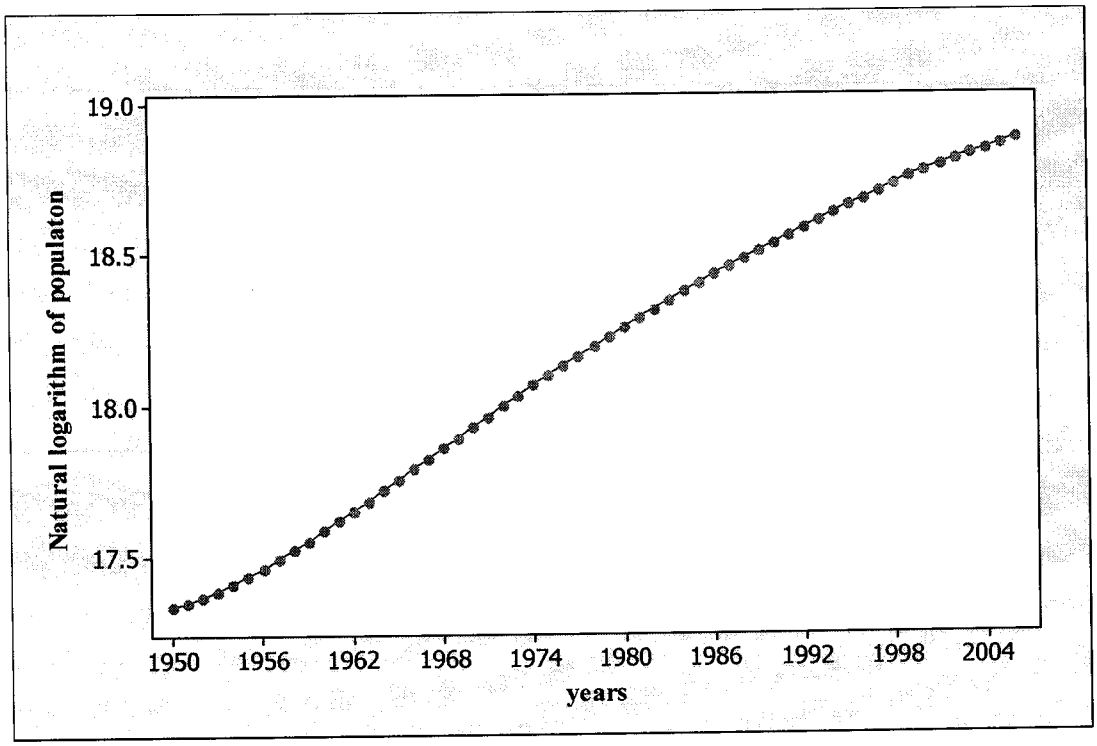


Figure 4.6 Trend of Logarithmic Population of Pakistan 1951-2007

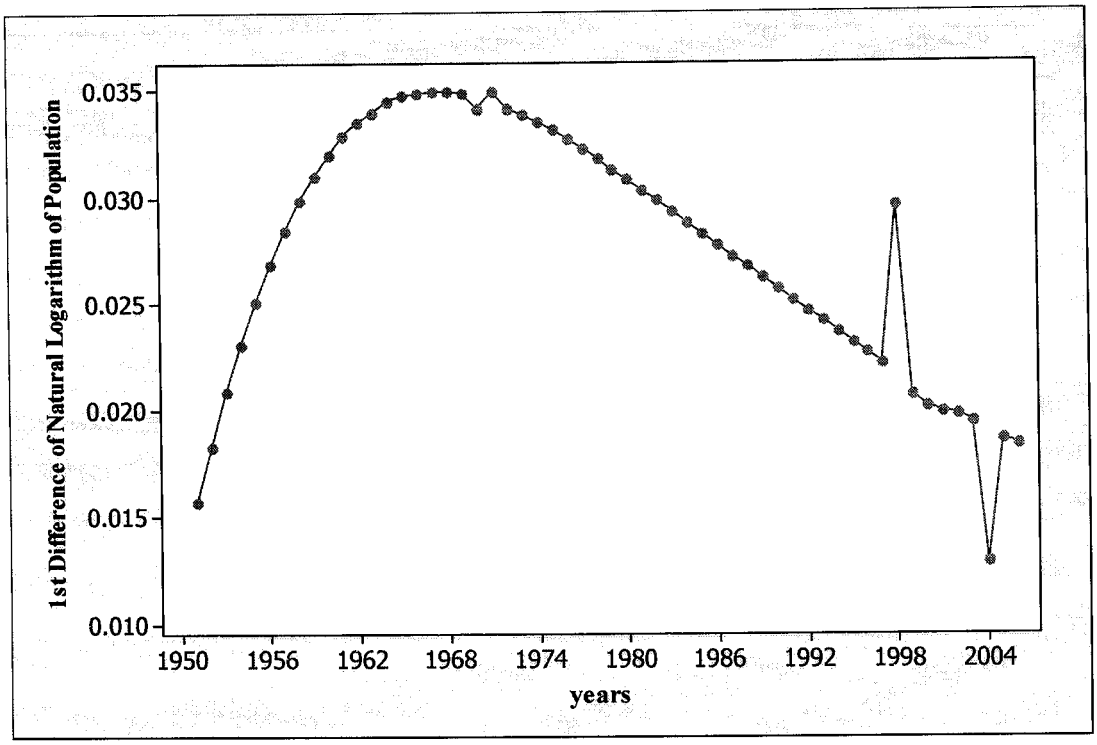


Figure 4.7 Trend of Logarithmic Population after 1st Differencing

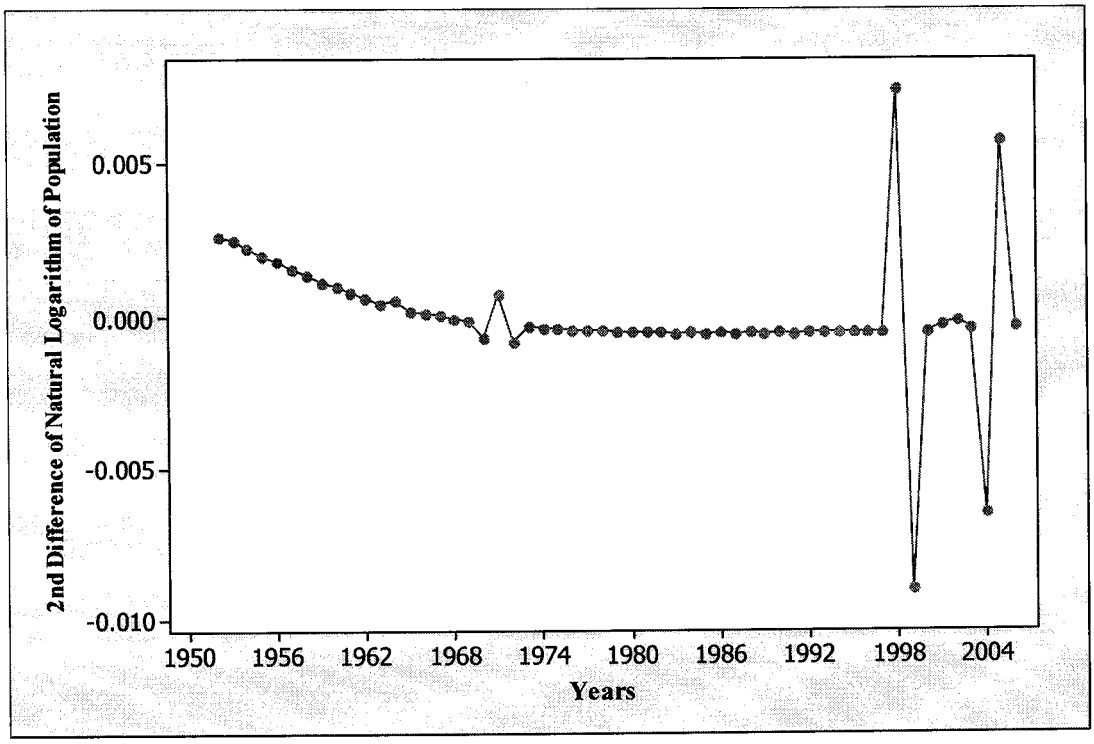


Figure 4.8 Trend of Logarithmic Population after 2nd Differencing

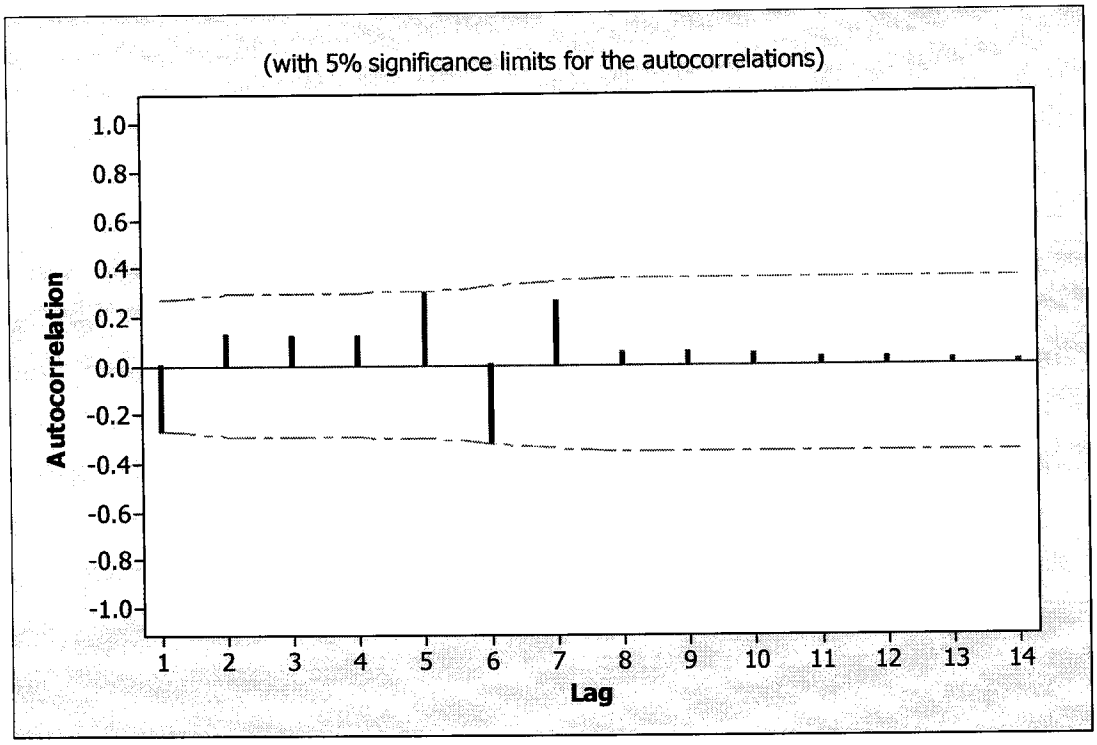


Figure 4.9 ACF of Logarithmic Population of Pakistan after 2nd Differencing

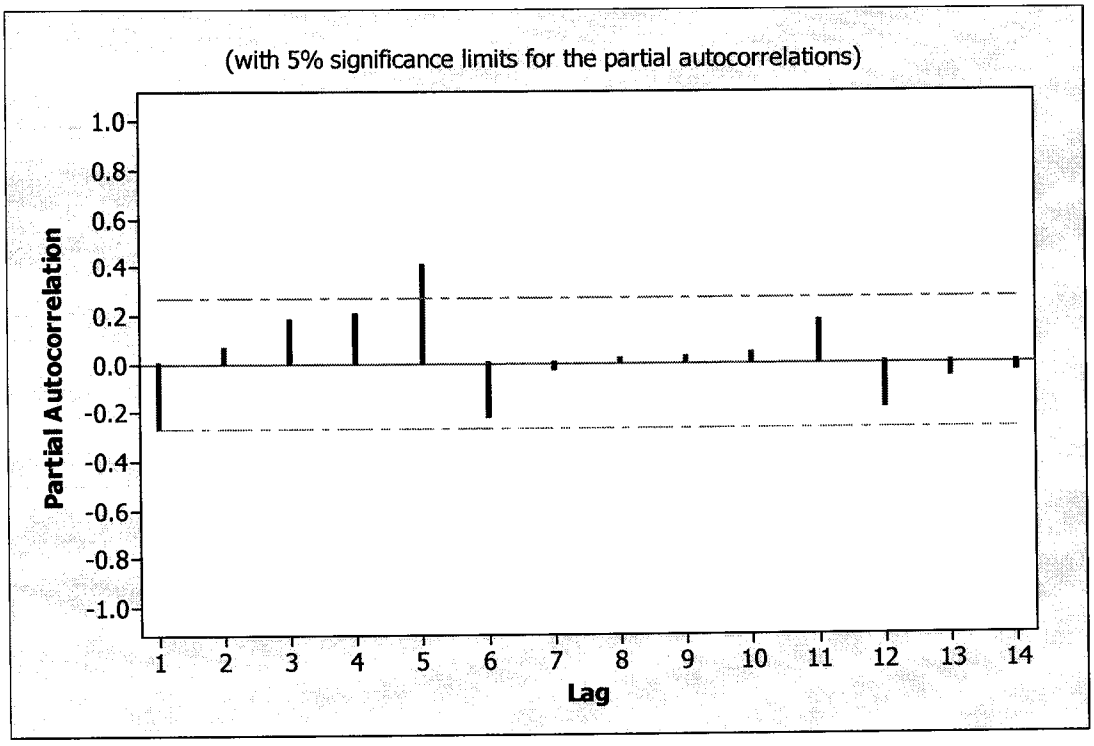


Figure 4.10 PACF of Logarithmic Population after 2nd Differencing

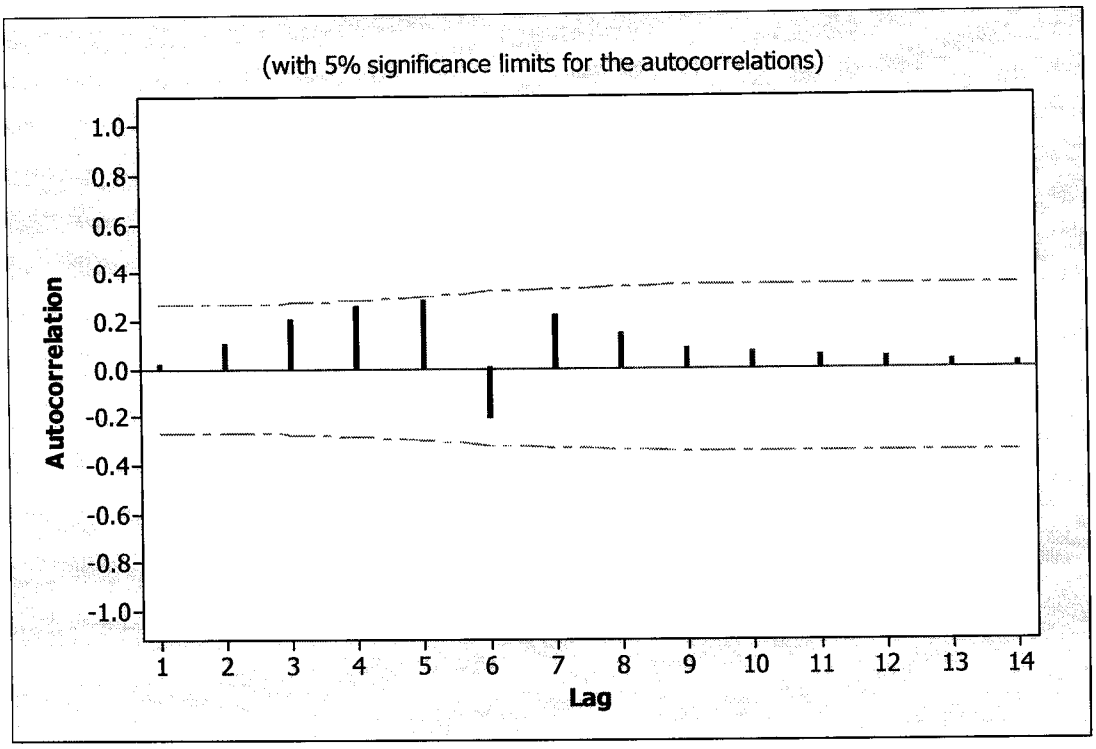


Figure 4.11 ACF of Residuals of 2nd Differencing of Logarithmic Population

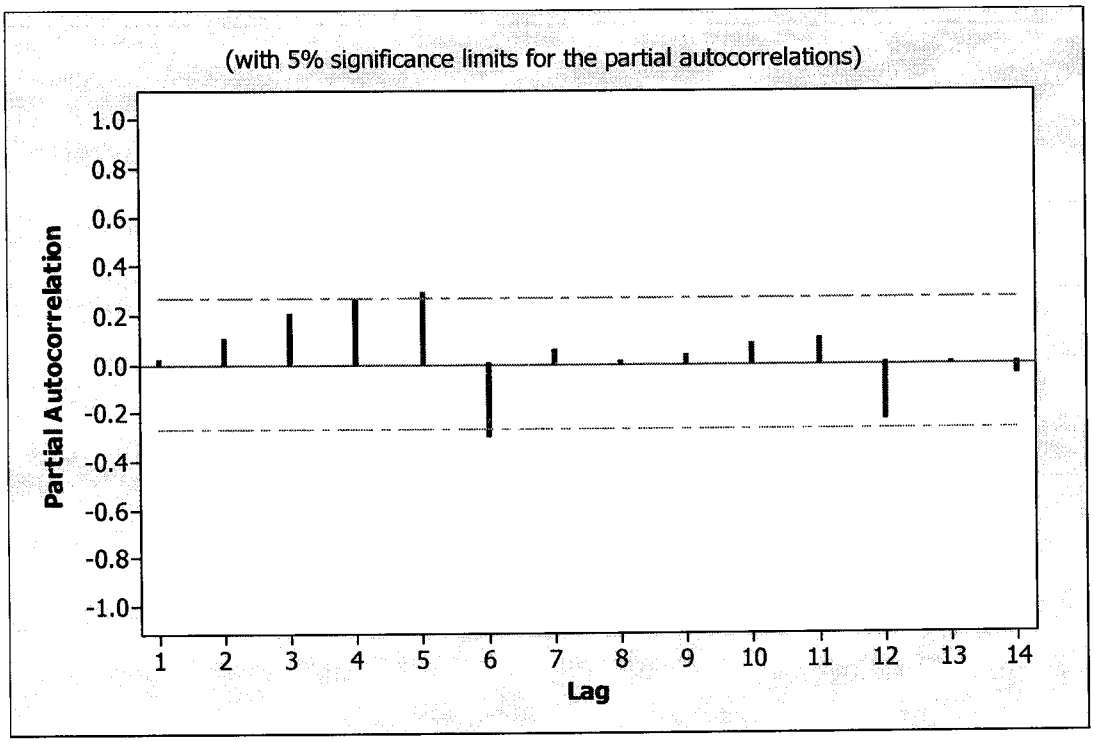


Figure 4.12 PACF of Residuals of 2nd Differencing of Logarithmic Population

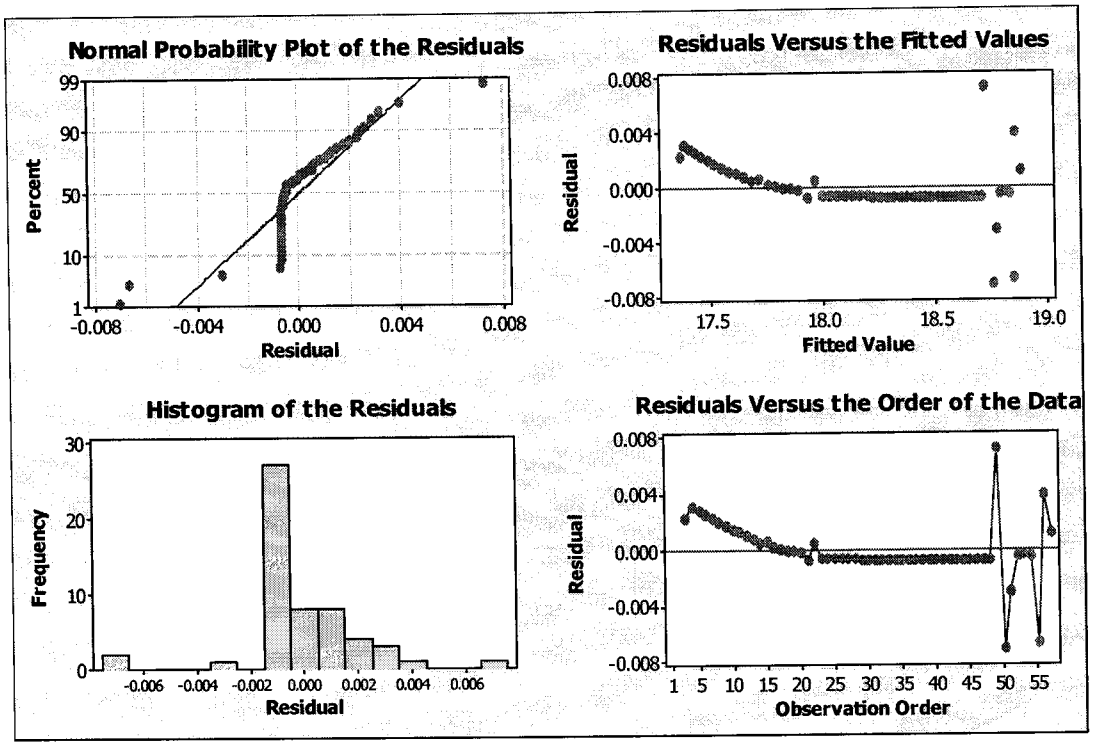


Figure 4.13 Four Residual Plots of Logarithmic Population after 2nd Differencing

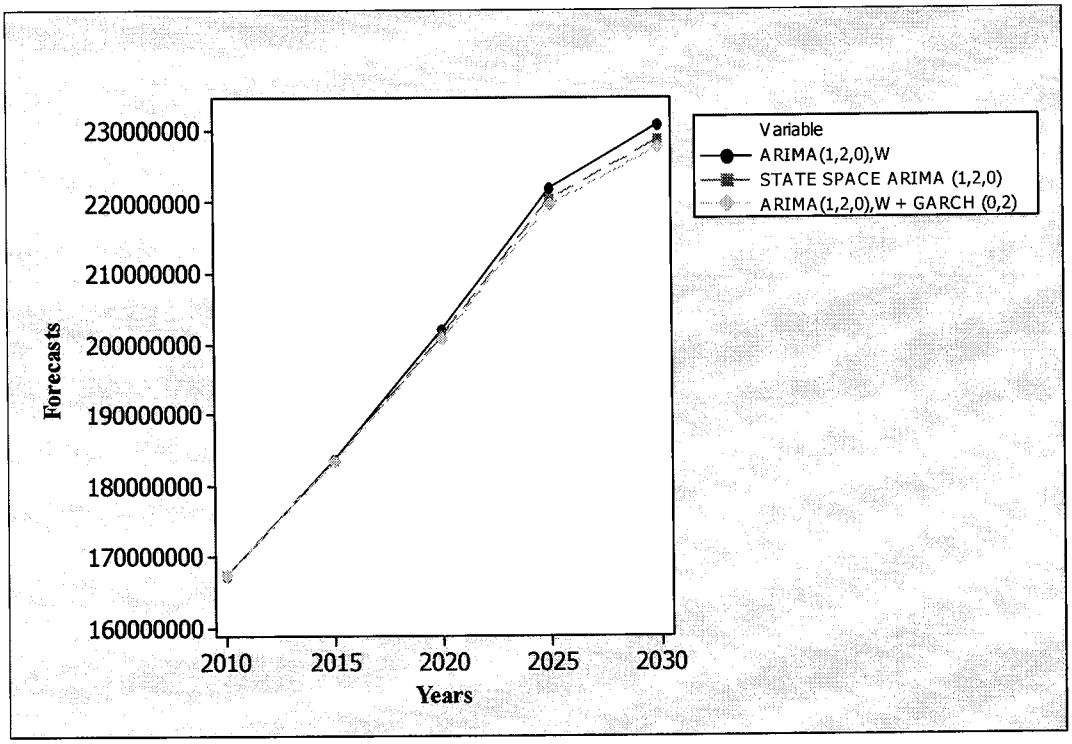


Figure 4.14 Time Series Plot of ARIMA (1, 2, 0) W, State Space ARIMA (1, 2, 0) W & GARCH (0, 2)

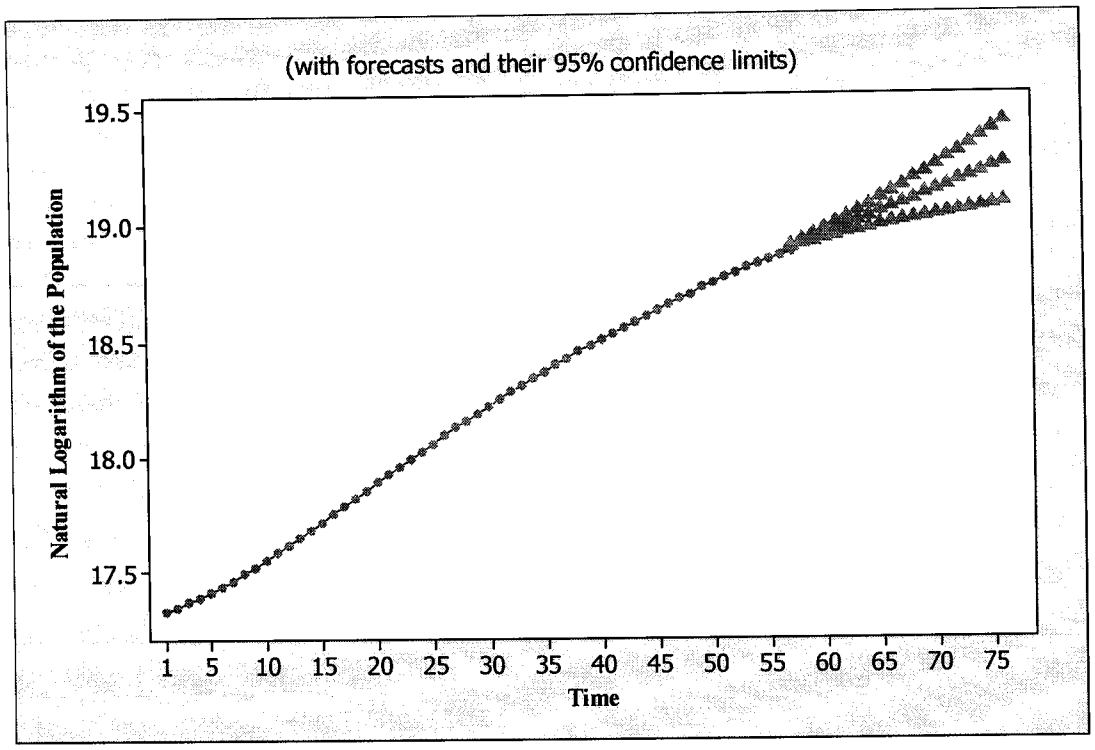


Figure 4.15 Logarithmic Trend of Predicted and Projected Population by ARIMA (1, 2, 0) W

CHAPTER 5

PROJECTIONS by TRADITIONAL MODELS

5.1 Introduction

Projection of any country plays a significant role in the planning as well as in the decision making for the socio-economic and demographic development. Today the major issue of the world is the tremendous growth of the population especially in the developing countries like Pakistan. The major innovation of this chapter is to project the population by the traditional growth models for vision 2030.

According to the World Population Prospects [WPP] (2006), the population of Pakistan would be 173.351 million, 190.659 million, 208.315 million and 224.956 million for the years 2010, 2015, 2020 and 2025 respectively. NIPS (2006) projected the population of Pakistan for the years 2010, 2015, 2020, 2025 respectively and given in the chapter of review of literature. Projections made by NIPS are slightly less than the projections reported by all other national as well as international agencies. Jan et al. (2007) projected the population of North West Frontier Province (NWFP) of Pakistan by Modified Exponential model and reported that it would be 61.12 million in 2053 whereas, it was 21 million in 2008. It indicates about 2.75 times increase than that of the population of 2008. Such a tremendous increase in NWFP in only 45 years was an alarm bell for NWFP,

the social scientists; population associated departments as well as the Government of Pakistan.

The population of Pakistan drastically increased from 34 million to 158 million during 1951 to 2007. The reality is that Pakistan witnessed a very high growth rate in its early decades after gaining independence. Although the population growth rate of Pakistan is decreased from the past. NIPS (2006) reported 1.86% growth rate of Pakistan during 2006 which is about 83% lower than the growth rate of 1998 i.e. 2.69%. Even then, there are still 156 countries out of 229 in the world that have less growth rate as compared to Pakistan (Population Growth Rate, 2008). The more reduction in growth rate is indispensable to maintain a balance between the population and the available resources of the country.

Relationship between Literacy, Education & Demographic (2009) advocated the relationship between women's education level and population growth. The educated women have less number of children than the uneducated women. An extra year of schooling reduces female fertility by as much as 5 to 10 percent. Education, particularly of girls and women, helps to control excessive population growth by promoting the concepts of family planning, collective health and well-being. An educated family makes informed choices with respect to having a child as well as for maintaining the health of the whole family. Moreover, the decrease in growth rate might be due to the increased female literacy rate. If the current female literacy rate continues, there may be more decrease in the growth rate of Pakistan.

Nobody can deny this fact that there is a negative relationship between education and the growth rate especially the female education. Most of the educated community prefers a

small family size and consequently the growth rate of population is decreased as compared to the past. Looking at the current trend of the growth rate, it is expected that in future the growth rate will decrease or at least remain same. If it continues, the population will grow with same age and sex distribution.

Due to such drastic increase in the population, limited available resources of the country, poor knowledge of future population and infeasible planning and management policies, the previous as well as the current Government of Pakistan remained in trouble since its independence. That is why; Pakistan could never have managed its future planning properly. Consequently, the citizens of Pakistan have been deprived from the basic necessities of life e.g. the quality food, water, health, education, employment, electricity, gas, transportation and other manufacturing and utility goods etc. The irregular conduct of population census and inadequate forecasting of the population might be the main cause of this distortion of the country. It seems that without regular census and adequate forecasting of the population, the solution of population problems as well as the stability of the elected democratic Governments is impossible. Moreover, Pakistan can neither stand in the row of the developed countries during the 21st century nor take the right decisions about its current and future planning. Although a large number of national and international scientists and agencies projected the population of Pakistan and its territories for different years. But the optimum forecasting of the population is the only key of success to take the right decisions regarding the future planning and to honour the commitments at the national as well as international levels.

The choice of a parsimonious model depends on the nature and population trend, the polynomial model may be of the type linear and nonlinear including the first and higher

degree regression models, simple exponential and Modified Exponential, Gompertz and Logistic growth models. Using such models, population of different countries is projected by different scientists (Shryock et al. 1973; Agrawal 2000; Jan et al. 2007). Srinivasan (1998) discussed the component method of population projection to project the population. According to one school of thought, the forecasting for vision 2030 in human population is preferable as the population growth rate does not remain constant for too long a period ahead. Klosterman (1990) advocated that 2 to 4 years forecasting is assumed to be short term and for more years a long term forecasting.

5.2 Objectives

The objectives of this chapter are:

- Projection of the population of Pakistan using some traditional growth models
- Comparison of projections with that of ARIMA model for vision 2030

5.3 About the Data

The data in this chapter is spread over the 57 years from 1951 to 2007 with regular interval of one year. Most of the data is taken from (Kemal et al., 2003; Iqbal, 2007) and population census reports of Pakistan (Anonymous, 1967, 1972, 1984, 2001).

5.4 Methodology

The Population is projected year wise from 2008 to 2032 using the traditional growth models i.e. Logistic growth model, Gompertz model, Modified exponential curve and Exponential growth model. The Goodness of fit of the models is assessed using the Mean Absolute Percentage Error (MAPE). Minitab-14 and SPSS-16 statistical applications are used for analysis purposes.

The mathematical forms of used models are:

Logistic Growth curve: The curve is of the form

$$Y = \frac{1}{\frac{1}{U} + A\,B^{\,t}}$$

Where Y is the response variable, A and B are the parameters of logistic model (SPSS-16, 2007). This curve is not recommended for too long a period ahead forecasting and for the population that is decreasing (Shryock et al., 1973).

Gompertz Curve: The logistic curve closely resembles the half normal curve whereas the Gompertz curve is not normal but a skewed one. The curve is of the form

$$Y = K\,A^{B^{\,t}}$$

The Gompertz curve is exactly the same as that of the Modified exponential curve except that it is the increase in the logarithms of the y values which are decreased by a constant proportion (Shryock et al., 1973).

Modified Exponential Curve: The form of the modified exponential curve is

$$Y = K + A\,B^{\,t}$$

Which yields an ascending asymptotic curve, the value of B lies between 0 and 1 whereas A assumes the negative values (Shryock et al., 1973).

Exponential Growth Model: Exponential growth model can be characterized by a constant percentage increase in the of population over time

$$Y = P_o\,e^{B\,t}$$

Where P_o equals the initial population at time t = 0, B represents percentage rate of growth, t is the time measured in the appropriate unit of one or five years and e is the base of the natural system of logarithms (Shryock et al., 1973).

Autoregressive Integrated Moving Average (ARIMA): Verbeek (2005) gave the following general form of ARMA (*p, q*) model

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

ARMA (*p, q*) is the combination of autoregressive and moving average specification which consists of the autoregressive part of order p and moving average part of order q. Where Y_t is the population at time t and is treated as response variable, Y_{t-1} is the population at lagged one and so on. On the other hand, ε_t is a white noise process at time t and ε_{t-1} is the residual at lag one.

Goodness of Fit Criteria

Mean Absolute Percentage Error (MAPE)

It is an evaluation statistic which is used to assess the goodness of fit of different models in national and sub national population projections. This statistic is expressed in percentage. The concept of mean absolute percentage error (MAPE) seems to be very simple but is of great importance in selecting a parsimonious model than the other statistics e.g. Coefficient of relative variation (CRV) and mean error (ME). A model with smaller MAPE is preferred to the other models.

The mathematical form of the MAPE is as under

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Yt - \widehat{Y_t}}{Yt}\right| * 100$$

Where Yt , $\widehat{Y_t}$, *and* n are the actual, fitted and number of observation of the (dependent variable) population respectively.

5.5 Results and Discussion

The reported and projected population by traditional and time series models are given in Table 5.1 for the years 2012, 2017, 2022, 2027, 2032. Box Cox transformation was applied to the data for the stationary purposes which gave the value of $\lambda = 0.2207$ along with its interval (-0.4051, 0.8466). Since the interval contains the value zero, it suggested that the log transformation is appropriate choice to make our series stationary before differencing of the series for the application of ARIMA models (Box & Jenkins, 1976). The fitted population given in table 5.1 for the year 2007 is 162.23 million and 158.08 million using Modified exponential and ARIMA (1, 2, 0) models respectively. These estimates are not only close to each other but also close to (NIPS 2006; Iqbal 2007) whereas these estimates are less than the other three traditional models given in the Table 5.1.

Similarly, the projected population for the year 2027 is 250.68 million and 230.68 million by Modified exponential model and ARIMA (1, 2, 0) respectively. The projected population by Modified exponential growth model is higher than the projected population by ARIMA (1, 2, 0) model. If the current growth rate continues, the projected population would also be approximately 250 million using compound growth model as that of the Modified exponential model. Consequently, the population would be doubled during the next 37 years.

The projected population by logistic model is 364.16 million which is more than double as compared to the population of 2007 during the next 25 years. It might be the overestimation of the population. It indicates that the growth rate in future will be greater than 3, which seems to be impossible and contradictory to the real situation. The logic

behind this fact can be seen. In 1998 census, the Government of Pakistan's expectations about growth rate was around 3% but after computation, it was announced 2.69% whereas according to the NIPS, the growth rate was 1.86% in 2006. If the current population growth rate continues, the projected population by ARIMA (1, 2, 0) W would be 254.09 million in 2032. It seems too much increase in population that is unaffordable for a third world country like Pakistan. The projection by ARIMA (1, 2, 0) W is satisfactory up to 2027. The real challenge to Pakistan is to decrease the growth rate or to limit the population size as well as to increase the country's resources to fulfill the ever increasing needs of the population in future. The decrease in growth rate is an easiest route to limit the population and this target might be achieved by only increasing the literacy rate in the female sector of the population.

The Mean Absolute Percentage errors are given in Table 1. Mean absolute percentage error of ARIMA (1, 2, 0) W is 0.49% which is minimum whereas MAPE (4.28%) is maximum of Logistic model. The mean absolute percentage errors of other models are between these two limits. On the basis of this selection criterion, ARIMA (1, 2, 0) W model can be preferred to the other growth models for population projection of Pakistan. Jan et al. (2007) used seven traditional growth models to project the population of NWFP province of Pakistan from 2003 to 2053. Jan et al. reported 61.12 million population of NWFP province in 2053 and recommended the Modified exponential growth model as a parsimonious model using the same evaluation statistics for goodness of fit of the model. The reason might be the trend differences between the population of province NWFP.

Figure 5.1 presents the trend of the reported population of Pakistan during the years 1972-2007. The line graph does not show any clear cut clue about the linear or quadratic

trend of the population but it seems to be a nonlinear trend of the population of Pakistan. Moreover, univariate time series model may also be tried to project the population of Pakistan.

Figure 5.2 compares the projected population of Pakistan computed by different models. The projected population trend by the Modified exponential growth and ARIMA (1, 2, 0) W models are approximately close to each other whereas the trend of other three traditional models are internally close to each other but different from the Modified exponential growth and ARIMA (1, 2, 0) W models.

Figure 5.3 compares the trend of reported and projected population by time series model ARIMA (1, 2, 0) W model. The fitted population of first 57 years is exactly the same as that of the original population. Figure 5.4, 5.5, 5.6, 5.7, and 5.8 present the residual plots of logistic, Modified Exponential, Gompertz and Exponential growth models as well as ARIMA (1, 2, 0) W respectively. The residual plots 5.4, 5.5, 5.6, and 5.7 almost have the same pattern but all are different from Figure 5.8. The analysis the residual plot of ARIMA (1, 2, 0) W indicates that most of the residuals move around one. The residual plot of ARIMA model is approximately random. The model fitting in social sciences is considered as an art than that of mathematical science. That is why, sometime the robust estimates in time series data are acceptable according to their national resources and circumstances.

5.6 Conclusion and Recommendations

The projected population for the year 2032 is 364.16 million, 356.46 million 341.93 million, 277.98 million and 254.09 million using logistic, Gompertz, Exponential, Modified exponential and ARIMA model respectively. The projection by exponential method is slightly less than the Gompertz and the logistic but higher than the Modified exponential model. The Modified exponential growth model projection is 277.97 million and 250.68 million population for the years 2032 and 2027 respectively which is minimum as compared to the other three traditional models. On the other hand, the ARIMA (1, 2, 0) W projected 230.68 million population for the year 2027 which is more close to the other national and international scientist's forecast (NIPS, 2006; WPP, 2006). Logistic model has 4.28% MAPE, which is highest among all the five models whereas the ARIMA (1, 2, 0) W has 0.49% MAPE which is minimum. MAPE of other models are between these two limits.

The projected population by traditional growth model (Modified exponential model) is close to the projection by time series ARIMA (1, 2, 0) W model. The MAPE of these models are 1.0578% and 0.485797% respectively. In the light of this statistic, again the ARIMA (1, 2, 0) W model among the other traditional models might be declared as parsimonious model. Since the same model is declared parsimonious model as that of given in chapter 4. The projection by ARIMA (1, 2, 0) W may be helpful for the future planning and projects of the Non Government Organizations as well as the government of the country. It is clear that increase in population projection is not as much as in the past. This decrease in growth rate might be due to the increased literacy rate especially of female education.

Briefly, It is the need of modern era. Since the private educational institutes are out of reach of common man, subsidized education must be provided to everyone at government educational institutes.

Table 5.1 Population Projection of Pakistan Using Different Growth Models

Year	population (in millions)	Projected Population (in millions)				
		Logistic	Gompertz	Exponential Growth	Modified Expo. Growth	ARIMA (1,2,0)
1972	65.31	62.02	63.75	62.80	64.87	63.10
1977	74.64	71.88	72.90	72.33	75.45	74.68
1982	87.29	83.30	83.50	83.30	86.99	87.35
1987	100.82	96.54	95.80	95.94	99.56	100.89
1992	114.94	111.89	110.09	110.49	113.26	115.03
1997	129.39	129.67	126.73	127.24	128.20	129.48
2002	144.80	150.28	146.13	146.54	144.48	144.86
2007	158.28	174.17	168.80	168.77	162.23	158.08
2012		201.86	195.32	194.37	181.57	173.65
2017		233.94	226.40	223.85	202.66	190.70
2022		271.13	262.91	257.80	225.63	209.64
2027		314.22	305.85	296.90	250.68	230.68
2032		364.17	356.46	341.93	277.98	254.09
Statistical Evaluation Techniques (klosterman,1990)						
Mean Absolute % Error (MAPE)		4.28 %	3.48 %	3.71 %	1.06 %	0.49 %
Coefficient of Relative variation (CRV)		12.2891	13.3784	17.5473	14.6356	

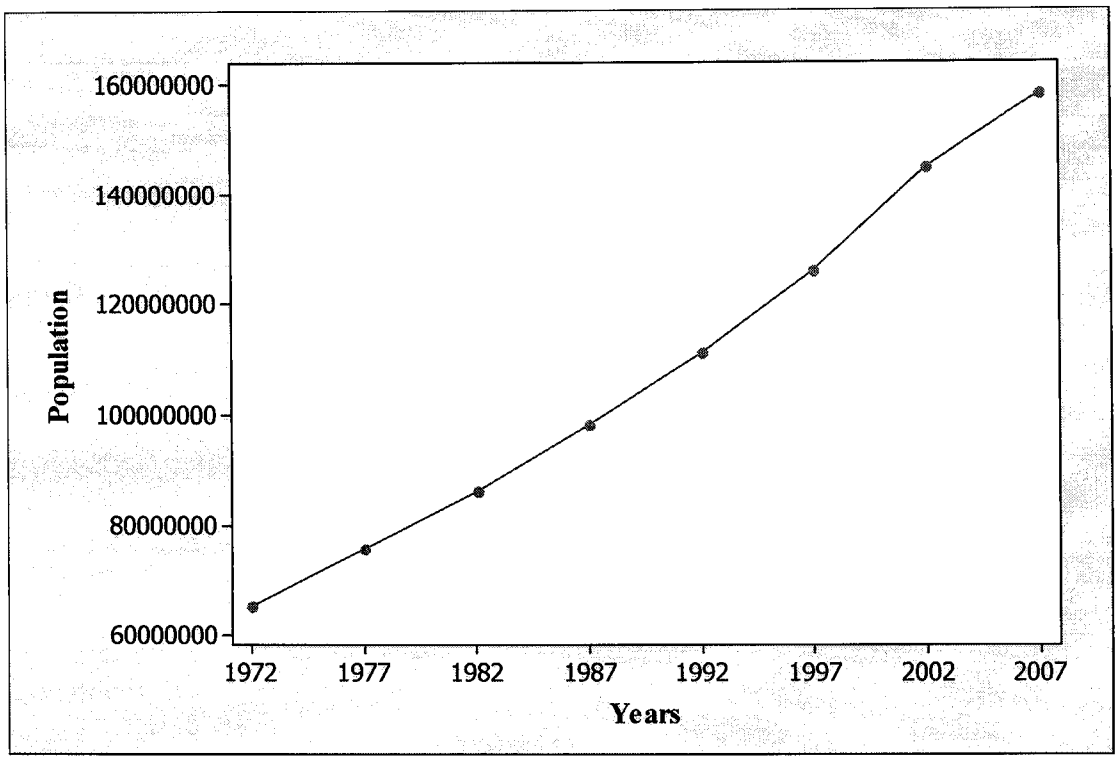


Figure 5.1 Population Trend of Pakistan during the Years 1972- 2007

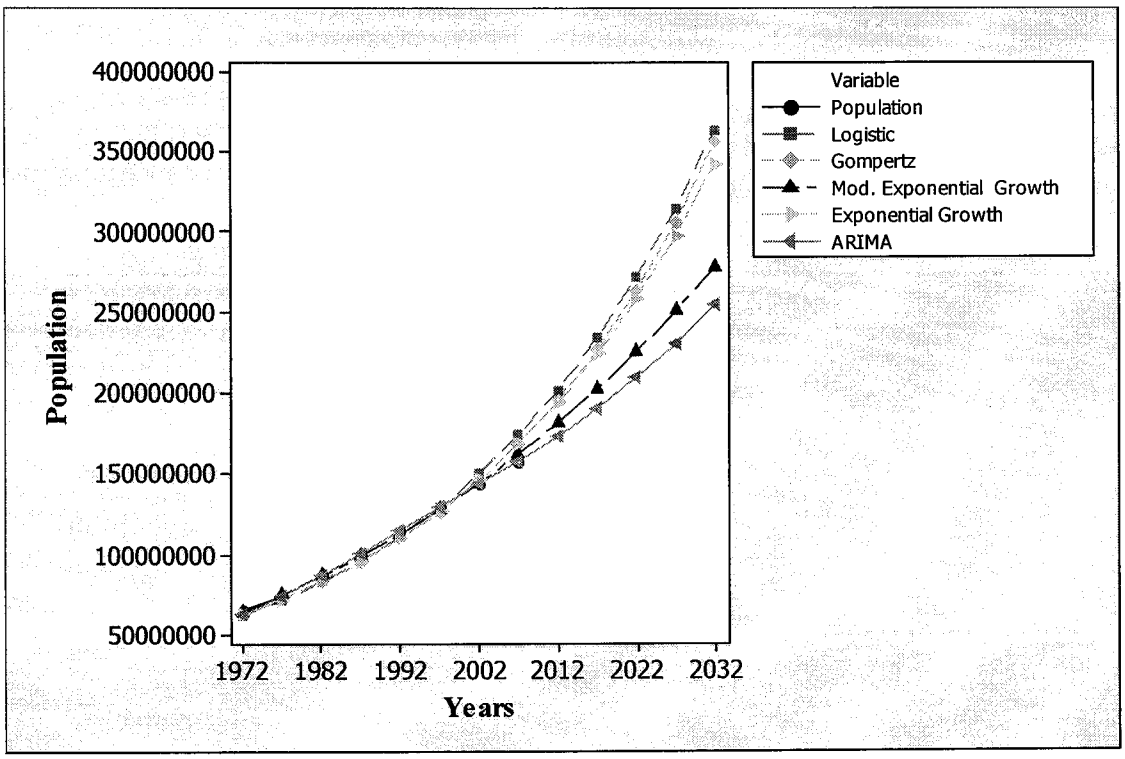


Figure 5.2 Projected Populations by Different Models 1972-2032

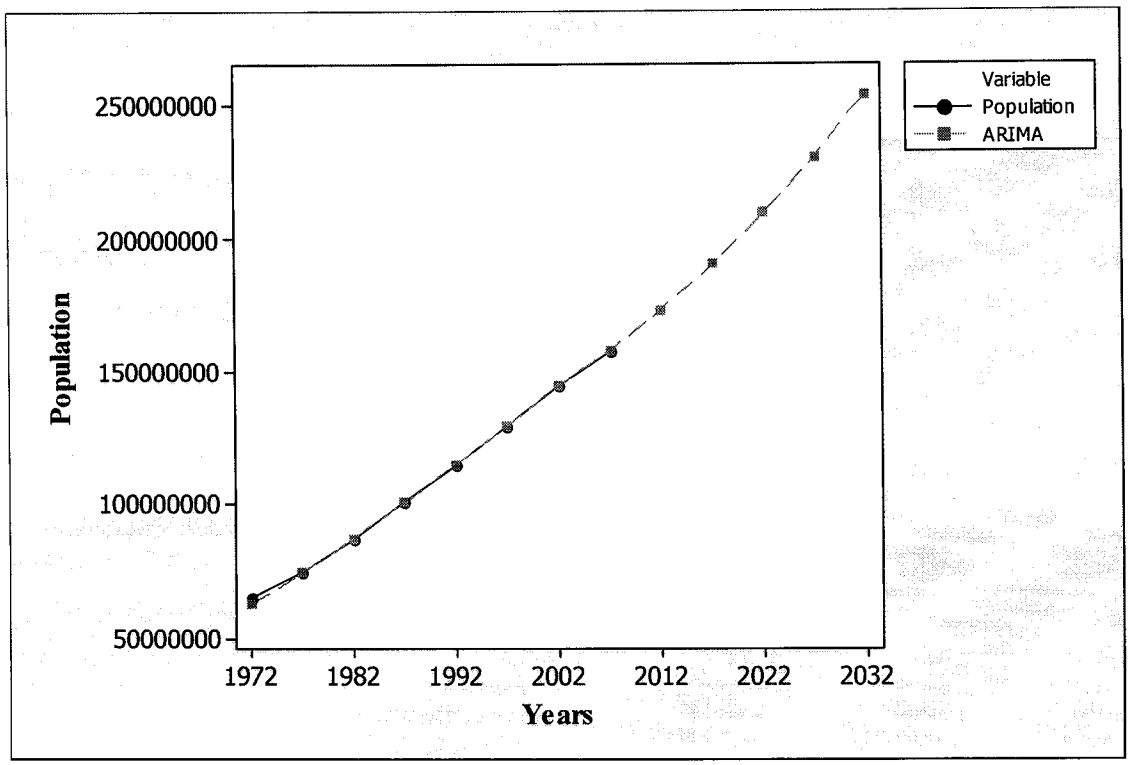


Figure 5.3 **Projected Populations by ARIMA (1, 2, 0) W Model 1972-2032**

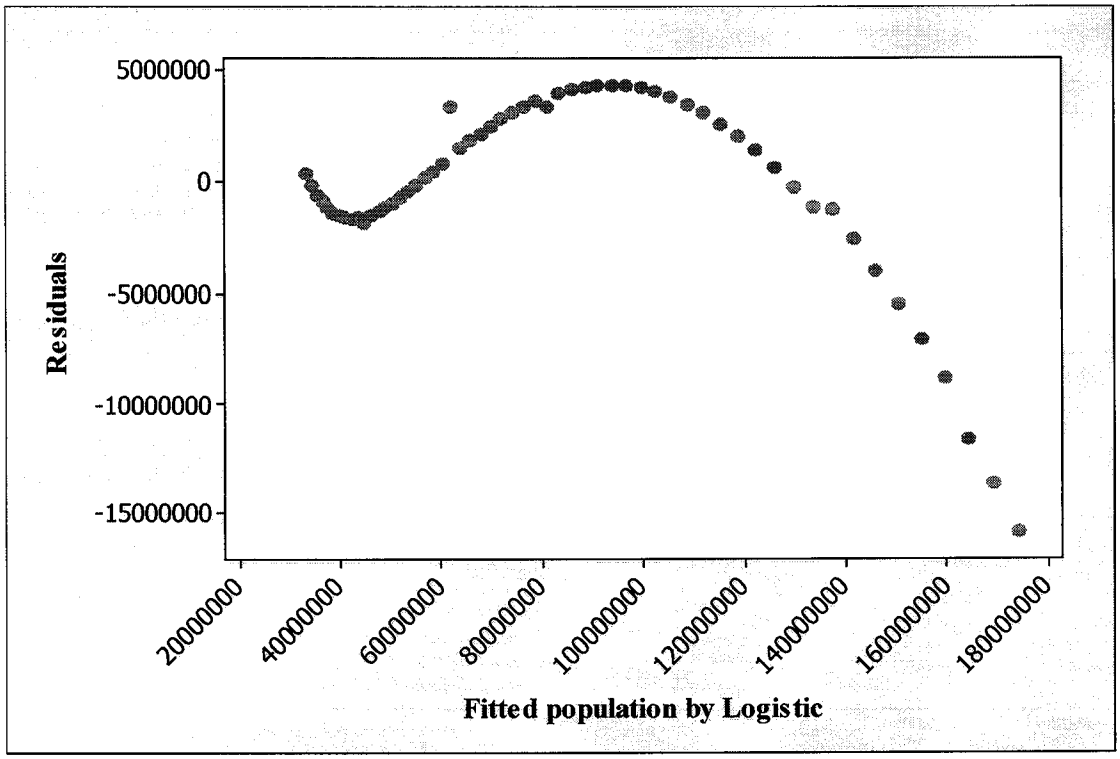


Figure 5.4 **Residual Plot of Logistic Model 1972-2032**

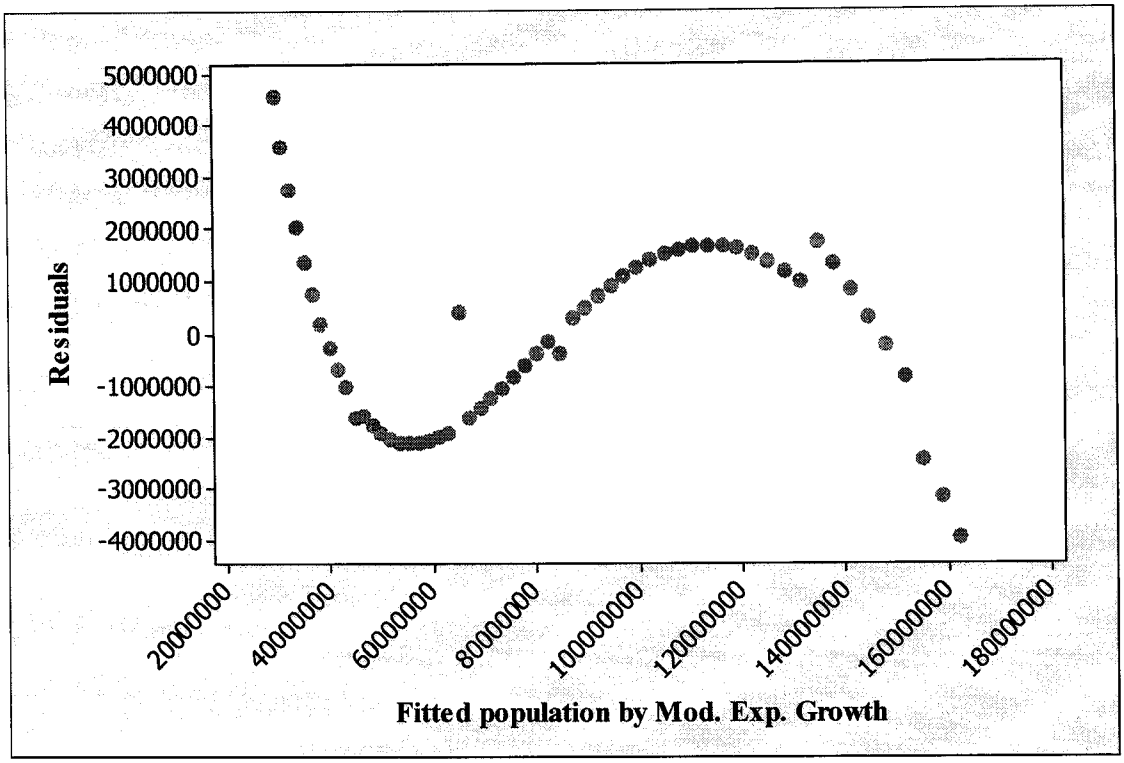


Figure 5.5 Residual Plot of Modified Exponential Growth Model

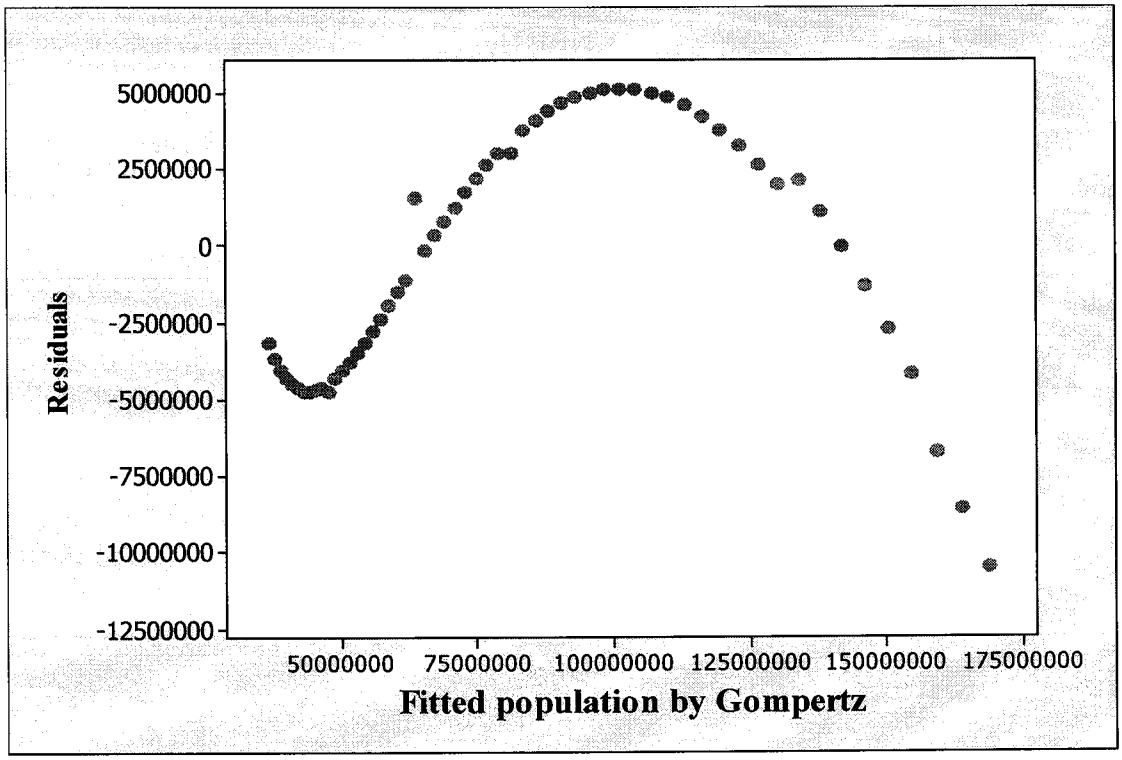


Figure 5.6 Residual Plot of Gompertz Growth Model

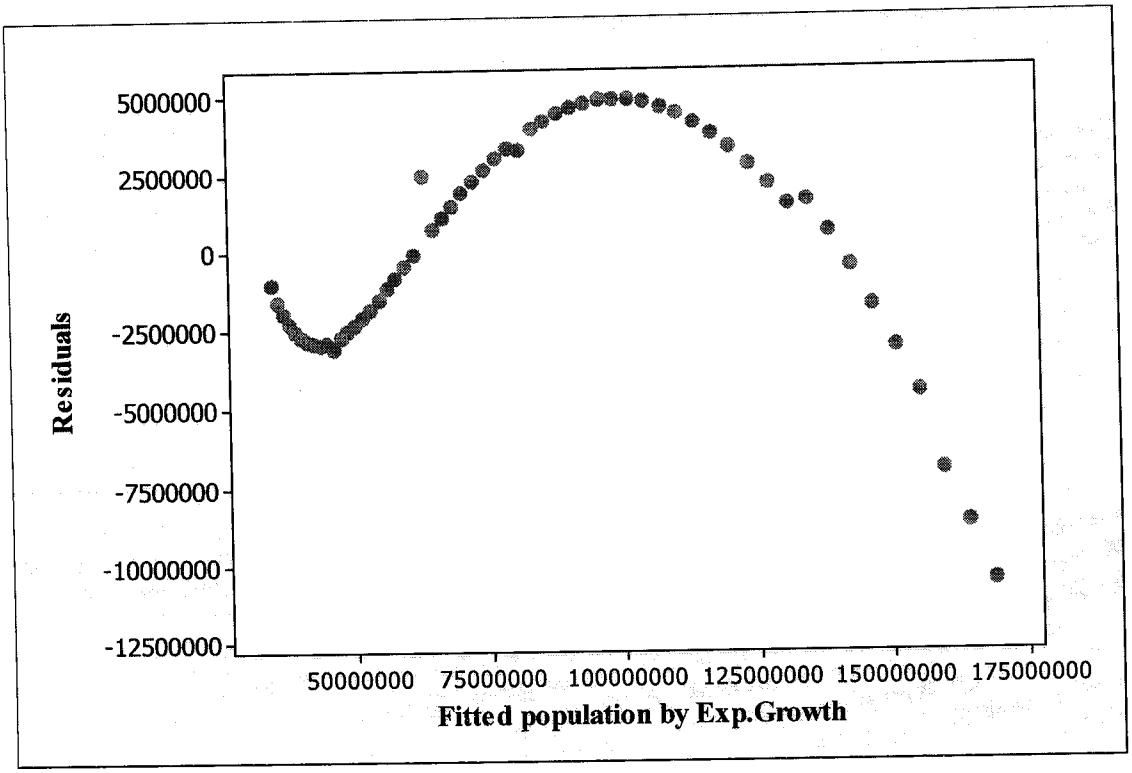


Figure 5.7 Residual Plot of Exponential Growth Model

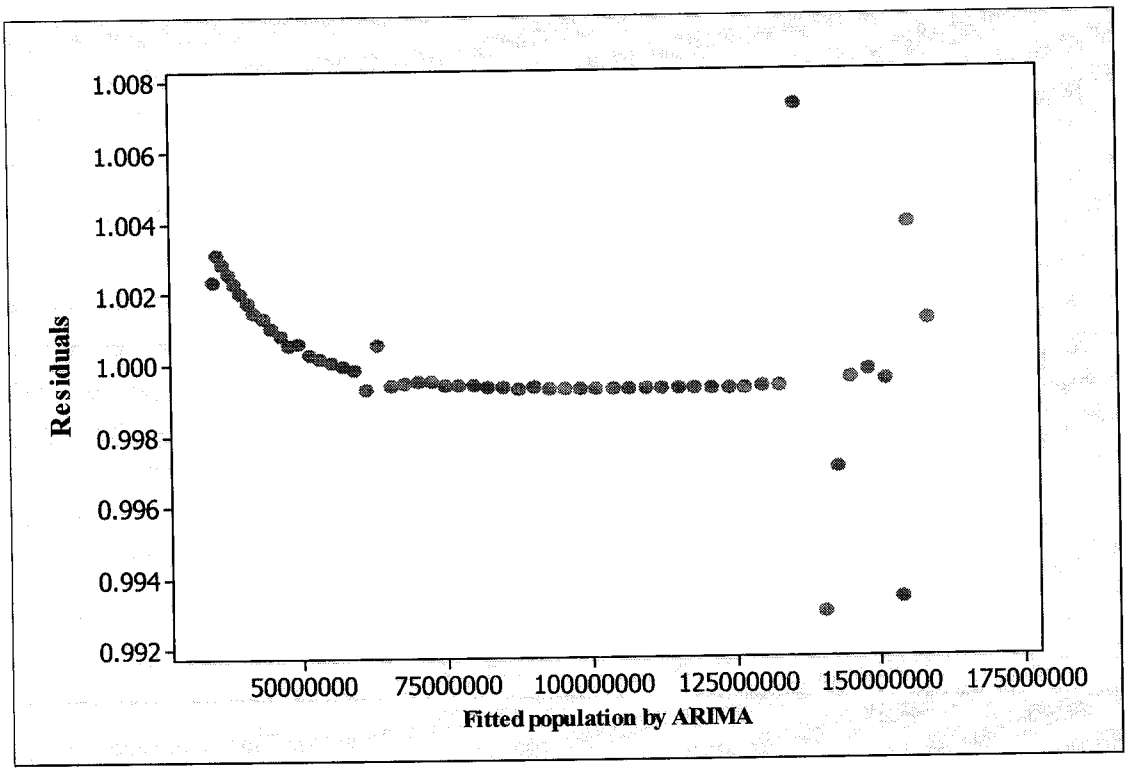


Figure 5.8 Residual Plot of ARIMA (1, 2, 0) W Model

of the dependent population (<15 and 65+) is also projected and found to be 44.3%, 61.2% and 68.7% for the years 2005, 2031, 2056 respectively in the presence of net international migration. Similarly, these percentages would be 68.0% and 80.9% in 2031 and 2056 respectively in the absence of net international migration.

Keeping in view the importance of population projection, in this chapter, it is tried to project the age sex distribution of population, the population of important age segments e.g. childhood (0-4) and dependent population (<15 and 65+), working group (15-64) and ageing population (65+) as well as the total population for vision 2030. The inequality overtime has also been examined within age distribution of actual population of different censuses as well as of projected population. Population is usually projected by using the component and exponential growth models. In brief, Modified Markov Chain model is used to project the age sex distribution of population whereas to examine the age inequality, the Gini coefficients, confidence intervals are computed and Lorenz curves are drawn.

6.2 Objectives

The objectives of this chapter are:

- Projection of age sex distribution
- Projection of total population
- Projection of significant segments of the population
- Measuring the inequality of the age sex distribution of actual censuses
- Measuring the inequality of the age distribution of projected population

6.3 About the data:

The basic source of the data is population censuses held in Pakistan e.g. 1951, 1961, 1972, 1981 and 1998 (Anonymous, 1967, 1972, 1984, 2001). The total population in all censuses was distributed in five years of age groups with last age group 75+ in the last three censuses 1972, 1981, 1998 whereas in the first two censuses 1951 and 1961, the last age group of population was 60+. Kemal et al. (2003) reported the poor quality of population censuses data of Pakistan and pointed out that without smoothing the population census data of Pakistan should not be used for further population analysis.

6.4 Modified Markov chain model

The system of equation approach (Keyfitz 1964; Markov Chains 2008) is modified to project the population of different age groups as well as the total population of Pakistan. The system of equation is

$$x' = x'A \qquad (6.1)$$

Where A is the matrix of one step transition probabilities a_{ij} i.e. the probability of moving from state i to state j in step one. Since the probabilities are nonnegative and the process must make a transition into some state, such that

$$a_{ij} \geq 0, \qquad i,j \geq 0\,; \quad \sum_{j=0}^{n} a_{ij} = 1, \;\; i = 0, 1, \; . \; . \; . \; , n$$

Let A denote the matrix of one step transition probabilities a_{ij} , so that

$$A = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & \cdots & \cdots & \cdots & \cdots & a_{0n} \\ a_{10} & a_{11} & a_{12} & a_{13} & \cdots & \cdots & \cdots & \cdots & a_{1n} \\ a_{20} & a_{21} & a_{22} & a_{23} & \cdots & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i0} & a_{i1} & a_{i2} & a_{i3} & \cdots & \cdots & \cdots & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n0} & a_{n1} & a_{n2} & a_{n3} & \vdots & \vdots & \vdots & \vdots & a_{nn} \end{bmatrix} \tag{6.2}$$

and x' is an initial row vector

$$x' = [\ x_o \quad x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad \cdots \quad x_{n-1} \quad x_n] \tag{6.3}$$

In real life population phenomena, the situation is slightly different with that of Markov Chain (2008) i.e. the new born babies are more in numbers than persons died in all age groups (0-70+) that compensate the population. No doubt, the existing population in different age groups belongs to different birth cohorts in cross sectional population censuses but the new babies are always positioned in first age group.

To overcome this complexity and to project the age distribution of population, the following A_l empirical transition probability matrix of order 9 x 9 is developed taking into account the situation of real life population phenomena. Each row and each column of this matrix is considered as separate state. The last row and column of matrix A_l transition probability matrix is an additional state (9th state) which is called a source /sinks state. Except the last state, the remaining 8x8 transition probability matrix contains the probabilities of survival of different age group people. In real life population phenomena, it is clear that an individual can never returned to previous state but may go to next state or may remain in the same state (passed away). The last column (9th column) contains the probabilities of dying in each state except its last element. The first and last elements of 9th state (row) are "r" and "1-r" respectively. Where "r" is the proportion of

the total population by which new born babies are positioned in 1st state (age group 0-9) and compensate the passed away population. The computational procedure of r is given on page 110 line 6. The resulting matrix A_1 consists of nine non overlapping states with respect to the age groups e.g. 1st state indicates the group of people having age 0-9, 2nd state indicates the group of people having age 10-19, 3rd state having people of age 20-29, 4th state having people of age 30-39, 5th state people having age 40-49, 6th state indicates the group of people having age 50-59, 7th state and 8th state indicate the group of people having age 60-69 and 70+ respectively. However, a_1, a_2, a_3, a_4, a_5, a_6, a_7 are the transition probabilities in the following matrix A_1 where a_1 is the probability of survival of the individuals from 1st state (age group 0-9) to 2nd state, a_2 is the probability of survival of the individuals from 2nd state (age group 10-19) to 3rd state, similarly 3rd state to 4th state, 4th state to 5th state, 5th state to 6th state and 7th state to 8th state respectively. a_1 may also be interpreted as the proportion by which the population aged 0-9 go into the 2nd state (10-19). Similarly other proportion can be interpreted in the same fashion. The population of 8th state are assumed to have zero probability of survival or one (1) probability of dying (definitely/ eventually death).

$$A_1 = \begin{bmatrix} 0 & a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 1-a_1 \\ 0 & 0 & a_2 & 0 & 0 & 0 & 0 & 0 & 1-a_2 \\ 0 & 0 & 0 & a_3 & 0 & 0 & 0 & 0 & 1-a_3 \\ 0 & 0 & 0 & 0 & a_4 & 0 & 0 & 0 & 1-a_4 \\ . & . & . & . & . & \vdots & . & . & . \\ . & . & . & . & . & \cdots & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{n-2} & 1-a_{n-2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-r \end{bmatrix} \tag{6.4}$$

Where a_1 is the transition probability of moving from State 1 (Babies) to State 2 (Adolescents) during the time period of ten years?

a_2 is the transition probability of moving from State 2 (Adolescents) to State 3 (Adults) during the time period of ten years.

In general, a_k is the transition probability of moving from State k to State k+1 in one time period of ten years.

State n-1 is the state of peoples of an age 70+.

Here "r" is the average growth rate of population of all provinces of Pakistan including Islamabad during 1972-1998 multiplied by the length of age group.

The resulting Modified Markov chain became irreducible/regular, aperiodic and hence a stationary distribution (limiting distribution).

The resulting system of equation is as follows

$$x_1 = rx_n$$

$$x_2 = a_1 x_1$$

$$x_3 = a_2 x_2$$

$$\vdots$$

$$x_{n-1} = a_{n-2} x_{n-2}$$

One extra equation that all probabilities must add up to one, namely

$$x_1 + x_2 + x_3 + x_4 + \cdots\cdots\cdots + x_n = 1$$

Solving the system of equations, we get

$$x_k = \frac{r \prod_{j=0}^{k-1} a_j}{r \sum_{i=0}^{n-2} \prod_{j=o}^{i} a_j + 1} \qquad k = 1, 2, \ldots, n\text{-}1$$

$$x_n = \frac{1}{r \sum_{i=0}^{n-2} \prod_{j=o}^{i} a_j + 1} \qquad (6.5)$$

Where $a_o = 1$ (for convenience)

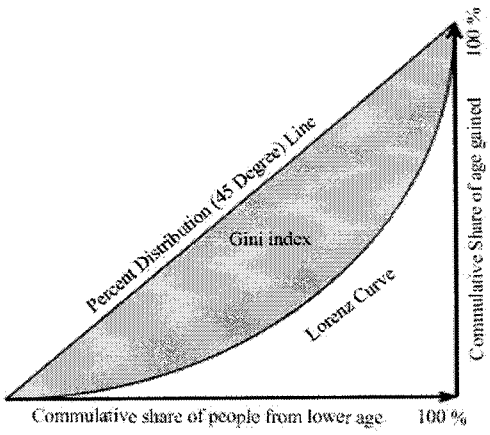
For population projections of Pakistan, the initial row vector x' consists of the population of different age groups of 1972 population census each group of interval of 10 years whereas the last element is the total of all groups' population of 1972 without federally administered tribal area population. To project the population of 1981, the initial row vector x' is multiplied with that of matrix A_2 (6.6). The first eight elements of the resulting vector (1x9) are the projected population of different age groups of 1981 and their sum is the projected population total of 1981. Later on, this sum will become the 9th element of the next initial vector to project the population for 1991. The resulting vector x' of order 1x9 is multiplied with that of the same matrix A_2 (6.6) to project the population of different age groups of 1991 and so on. In this way, each multiplication of this row vector and matrix will give the forecasted population for the next ten years. Additionally, Spectrum demographic software is used to project the population of different age segments e.g. the projection of babies, teenager, working, dependent and ageing population etc.

6.5 Inequality Measures of Age Sex Distribution

It is unanimously acceptable that population of different age groups have different requirements like food, medicine, industry, accommodation, marriages etc. Similarly, the planning and management of a country from micro to macro level depends on the age sex distribution of the population of the country. If the population of different age groups is projected precisely, then it would be highly supportive for a first-rate knowledge to determine the population needs that is inevitable for planning and administration of the state. It means that without the optimum knowledge of age sex distribution of population of a country; the optimum and outstanding planning of the state is impossible.

To measure the disparity of the age sex distribution of projected population by Modified Markov chain during (1981-2031) and of projected population of significant age segments (2010-2030) by component growth method, the Gini coefficients, and confidence intervals are computed.

The Mathematical formula of the Gini coefficient is $\sum_{k=1}^{n}(X_k - X_{k-1})(Y_k + Y_{k-1})$ and graph of the inequality i.e. Lorenz curve is given on the right hand side:



The less concave in the Lorenz curve away from the line of equality indicates less degree of age disparity. The Gini coefficient presents the area of concentration between the Lorenz curve and the line of perfect equality as it expresses a proportion of the area enclosed by the triangle defined by the line of perfect equality and the line of perfect inequality (Gini Coefficient 2009). The

Gini coefficient (G) lies between 0 and 1. The minimum value of G is zero when all measurements are equal and the theoretical maximum of G is one for an infinitely large set of observations which is the ultimate inequality. Cowell (1995) wrote that different scientists applied these measures into other areas than income and wealth, but mostly within economics. According to Stuart, & Ord (1994), the closer the value of the coefficient is to 1, the more unequal the distribution. Pan American Health Organization (2001) measured the inequality by computing the biased and unbiased Gini coefficients along with their confidence intervals using different data sets e.g. GNP per capita, infant mortality rate, live births and infant deaths etc. Brown (1994) used the Gini-style indices to evaluate the spatial patterns of health practitioners and used Alberta data for theoretical considerations as an application. Slack & Rodrigue (2008) computed the traffic inequality at different terminals using index of dissimilarity (ID), Gini coefficient (G) and Gini's means difference (GMD).

6.6 Results and discussion

To project the population , 1972 population census data is substituted into the equation (6.3) and the following x' vector became an initial row vector (Anonymous 1972) i.e.

$$x' = [19544973 \quad 13168302 \quad 9208929 \quad 7289623 \quad 5390373 \quad 3497042 \quad 4362641 \quad 1876447 \quad 62309340]$$

The first element of x' is the number of individuals of Pakistan having age 0-9, second element is the number of individuals of Pakistan having age 10-19, the 2nd last element is the number of individuals of Pakistan having age 70+. Similarly the other elements of x' can be interpreted except the last element. The last element of x' is the total population of 1972 which plays a significant role in multiplication with that of matrix A_2. After

substituting the transition probabilities and the value of r in matrix A_2 (6.4), the resulting transition probability matrix becomes A_2 is of the following form:

$$A_2 = \begin{bmatrix} 0 & 0.9499 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0501 \\ 0 & 0 & 0.889 & 0 & 0 & 0 & 0 & 0 & 0.1110 \\ 0 & 0 & 0 & 0.9572 & 0 & 0 & 0 & 0 & 0.0428 \\ 0 & 0 & 0 & 0 & 0.9522 & 0 & 0 & 0 & 0.0478 \\ 0 & 0 & 0 & 0 & 0 & 0.9160 & 0 & 0 & 0.0840 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8490 & 0 & 0.1510 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5800 & 0.4200 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6000 \end{bmatrix} \quad (6.6)$$

The entries above the main diagonal 0.9499, 0.889, 0.9572, 0.9522, 0.9160, 0.8490 and 0.58 are the probabilities of survival of the preceding states (age groups) e.g. first value is a_1= 0.9499 which is the probability of survival of 1st state/age group (0-9) or probability to move into the next age group 10-19. Similarly the other probabilities can also be interpreted. These probabilities are computed from two consecutive population censuses 1972 and 1981 (Anonymous 1972; 1981). The elements of last column of matrix A_2 (6.6) are the probabilities of dying in all states except the last element i.e. one minus r. Matrix A_2 may call a Modified Markov chain transition probability matrix.

The projected population of 1981 is obtained by multiplying the matrices x' and A_2 . The total projected population for the year 1981 is obtained by adding the projected population from 0-9 to 70+ age groups. The projected population of each age group from 0-9 to 70 + is exactly the same as that of the actual population census 1981 of Pakistan. Population projection for the year 1991 is obtained by multiplying the matrices x' and A_2 after updating the initial vector x'. Similarly the projection for each next ten years is obtained after updating the initial vector x' each time and by multiplying with that of same matrix A_2. The main advantage of this method is that, only the vector x' has

to be updated before multiply with that of matrix A_2. The population projection depends only on the accuracy of probabilities of survival between the two censuses, the value of r and the population of different age groups taken in an initial row vector. The more accurate the data set, the more accurate the projections would be.

Finally, it is concluded and recommended that in each next projection, the initial row vector of population x' has to be updated before multiplying with that of the transition probability matrix A_2. Usually, the population census is held after every 10 year in almost all over the world. That is why; the population is projected with ten years of intervals. If population census becomes possible after every five years then projection can be made for five years of interval.

The disparity of age distribution of actual population censuses of Pakistan i.e. 1951, 1961, 1972, 1981 and 1998 has been measured by using the Gini Coefficients along with their confidence intervals. The Lorenz Curves are drawn for male, female and for both sexes separately. These measures are also computed for comparison purposes using projected population age sex distributions during 1981-2031. StatsDirect (2008) software is used to compute the Gini coefficients, confidence intervals and to draw the Lorenz curves.

Table 6.1 reveals the projected population for the years 1981-2031 with ten years of intervals in each age group. The projected population for the years 1981, 1991, 2001, 2011, 2021 and 2031 is 82.55 million, 106.96 million, 138.77 million, 179.97 million 233.99 million and 306.75 million population respectively. The projected populations are close to the projections (WPP 2008; People Facts & Figures 2009; Total Population by Country 2009) and greater than (NIPS 2006; IDB 2008). The Modified Markov Chain

Model gives satisfactory results for only 30-40 years ahead from which the probability of survival is derived. Projection of human population more than 30-40 years is not encouraging due to the instability of growth rate of population of modern era. According to Klosterman (1990), twenty years forecasting is assumed to be a long term forecasting.

Table 6.2 presents the Gini coefficients along with their confidence intervals of projected population for the years 1981, 1991, 2001, 2111, 2021 and 2031. The Gini coefficient of the projected population for the years 1981 is 0.4725 with 95% confidence interval (0.4016 - 0.6383) which is approximately close to that of the actual population census of 1981 i.e. 0.4552 (0.4137- 0.5408). Gini coefficients of projected population for the years 1991 and 2001 are 0.4727, 0.4717 respectively which are although close to the Gini coefficients of actual population censuses but the confidence intervals of projected population are slightly wider. Similar pattern is observed among the Gini coefficients 0.4712, 0.4690 and 0.4716 of the year 2111, 2021 and 2031 respectively but overall the population size is changed. These measures indicated the medium level of concentration among different age groups of the population of different projections. Moreover, these coefficients do not specify the lower and higher side of more concentration among different age groups of population.

Table 6.3 consists of the Gini coefficients along with confidence intervals of population censuses 1951-1998 for male, female and for both sexes separately. The Gini coefficient for 1951 and 1961 are less and almost the same which indicate less level of concentration among different age groups of the population. Gini coefficients of population censuses 1972, 1981, 1998 lie within 0.4337--0.4777 and of projected population 1981, 1991, 2001, 2011, 2021 and 2031 are 0.4725, 0.4727, 0.4717, 0.4712, 0.4690, and 0.4716

respectively. This is an indication that if the growth rate of 1998 as well as associated social and cultural values continues; then the shape of population distribution will remain stable but the population size would surely be increased. On the other hand, if the Total Fertility Rate, life expectancy, sex ratio and international migration are changed in near future, the age and sex distribution of population would definitely be changed.

Table 6.4 presents the percentage of momentous segments projected population for selected years. In 1998, 14.82%, 28.60%, 53.10%, 3.50% population of ages 0-4, 5-14, 15-64 and 64+ are reported respectively. These percentages are being projected 12.20%, 21.96%, 61.93% and 3.91% of ages 0-4, 5-14, 15-64 and 64+ respectively up to the year 2020 whereas these would be 10.13%, 20.51%, 64.89% and 4.46% corresponding to above mentioned age groups up to the year 2030. It shows approximately 1.6% and 7% decrease in age group 0-4 and 5-14 up to 2020 and 5% and 8% in age groups 0-4 and 5-14 up to 2030 respectively.

Table 6.5 Gini Coefficients of Component Growth Model Projections 1998-2030

Population Census	Sex	Gini Coefficient	95% Confidence Interval
1998	Both sex	0.5418	(0.4065 , 0.7799)
2010	Both sex	0.6010	(0.5275 , 0.8006)
2015	Both sex	0.6042	(0.5227 , 0.7933)
2020	Both sex	0.6127	(0.5286 , 0.7877)
2025	Both sex	0.6238	(0.5366 , 0.7802)
2030	Both sex	0.6390	(0.5510 , 0.7721)

Table 6.5 reveals the Gini coefficients of projected population of significant age segments for different years 1998-2030. These coefficients indicate the more disparity into

different age groups of population. According to Total Population by Country (2009), the projected population of age 0-14 is about 37.20% and ranked the Pakistan at 46th position in the world which was approximately 43% in 1998. It seems that it is only due to the decreasing trend of growth rate. The Projected population of working age group (15-64) in Pakistan is about 58.60% and ranked at 98th position. It means that there are 97 countries in the world that are healthier than Pakistan as they have more working population. On the other hand, with 4.20% population of age 65+, Pakistan is ranked at 94th positions whereas this percentage was 3.50% in 1998. It clearly indicates the increase in aged population.

According to the World Population Prospects (2008), the projected population of age 0-4 would be 13.2%, 12.7%, 11.6%, 10.6% and 9.8% in the years 2010, 2015, 2020, 2025 and 2030 respectively whereas the population of age 5-14 would be 23.4%, 22.5%, 21.9%, 20.9% and 19.5% in the years 2010, 2015, 2020, 2025 and 2030 respectively. It indicates that baby and teenager population will decrease during the next 10-20 years.. On the other hand, working population (15-64) will increase approximately 9% and 12% up to 2020 and 2030 respectively as compared to 1998. This is a good indication that the population of Pakistan will be healthier and vigorous in future. Similarly, the ageing population (65+) would be 4.1%, 4.3%, 4.6%, 5.1% and 5.8% in the years 2010, 2015, 2020, 2025 and 2030 respectively. Ryan & Willits (2007) also reported that the aging population of United States is being increased. In 1900, the percentage of elderly population (75+) was 29% which is expected 56% in 2040. At the same time, the percentage of 64+ will be 3.91% and 4.46% up to 2020 and 2030 respectively which was 3.50% in 1998? It indicates the increase in ageing population during the next 10-20 years.

It seems that there would be a problem of ageing population in future and it would be a big hindrance in the country development. Although Pakistan is a religious country and people feel pleasure to look after their elders even then the ageing population would be affected. It will happen only due to increased literacy rate and busy scheduling of life in the modern era. In the light of these estimates, it is emphasized that Government should thresh out the expected social problems of ageing population, its management and the needs of ageing population well in time. It is indispensable to overcome unavoidable circumstances of ageing population in future.

Figures 6.1, 6.2, 6.3 and 6.4, 6.5, 6.6 are the Lorenz curves of 1972 and 1981 actual population census for both sexes as well as for male and female separately which indicate medium level of concentration among 1972 and 1981 population censuses. Figures 6.7, 6.8, 6.9 present the Lorenz differences for 1998 population census which have slightly high level of concentration as compare to 1972 as well as 1981 population census.

Figures 6.10, 6.11 and 6.12 are the Lorenz curves of projected population for the years 2111, 2021, 2031. These plots are slightly different as that of the plots of census 1998.

Figures 6.13, 6.14, 6.15, 6.16, 6.17 and 6.18 are the Lorenz curves of the projected population of important age segments for the years 1998, 2010, 2015, 2020, 2025 and 2030. These plots are quite different and have a greater Lorenz differences during the next 20 years as compare to the distribution of population segments of 1998.

Figure 6.19 presents a comparison of age and sex distribution from the years 2010 to 2030. Over all the population of Pakistan will increase to approximately 73 million from 2010 to 2030 with almost similar increase in each age group except the age groups 4-15,

15-64. It might be due to improved medication and domestic hospitality; the ageing population would also be increased.

Figures 6.20 and 6.21 are the population pyramids for the years 2020 and 2030 with that of the 1998 population pyramids. These pyramids indicate the increase and decrease in the distribution of the population in 2020 as well as in 2030. According to the population pyramid of 2020, there would be 12.20% population under age 5 and 21.96% population of age 5-14 which is about 2.62% and 6.64% less as compared to 1998 respectively. Similarly, this percentage would be 10.13% and 20.51% respectively in 2030 which indicates 4.69% and 8.30% decrease as compared to the 1998. The more decrease is expected during the first decade 2010-2020 than the next decade 2020-2030. Pyramids also present the increase in working and ageing population, although the increase in working group is more than ageing group.

6.7 Conclusion and Recommendations

The Modified Markov chain method has been found appropriate to project the age sex distribution of population 40 years ahead since 1981. The total projected population for the years 2011, 2021 are close to the population projection (WPP 2008; People Facts & Figures 2009; Total Population by Country 2009) and greater than (NIPS 2006; IDB 2008). (Populous Pakistan 2009; PRB 2008) reported the projected population for the year 2050 only, not for the years 2011, 2021 and 2031 but it seems that projection would be the same for these years. The measures of age disparity i.e. the Gini coefficients and the confidence intervals of the population censuses 1951-1961 are approximately same and less as compared to the coefficients of 1972, 1981, 1998. These coefficients are also computed for the projected populations that are slightly different with respect to the population census 1998. Moreover, the magnitudes of the coefficients indicate the medium level of concentration during the next 20 years.

On the other hand, the population of significant segments is also projected by population component growth model which indicates, that babies (0-4) and teenager population (5-14) will decrease whereas working and ageing population will increase during the next 20 years. These projections are approximately close to the projection (WPP, 2008). These projections predict the decrease in growth rate, increase in life expectancy and stable sex ratio of the population. Gini coefficients and Lorenz curves are also computed for projected age distribution that indicate high level of concentration and increase in ageing population. Being a religious society, Pakistanis feel proud to look after their elders and it is supposed that ageing population would not become a significant problem. However the ageing population might be affected due to busy schedule of modern life and it would

become an inevitable problem in the international competition era. The Government should start working on the socio-economic problems of ageing population from all aspects; their needs and managements well in time. Otherwise it would become difficult to handle the problems and facilities of the elders. The Government should also publicize and subsidize the science education, especially the female education and private technical education to cater the job requirements of qualified and technical graduates.

Table 6.1 Age Distribution of Projected Population by Modified Markov Chain

Age group	1981 (in millions)	1991 (in millions)	2001 (in millions)	2011 (in millions)	2021 (in millions)	2031 (in millions)
0-9	24.98	32.14	41.69	54.12	70.26	92.35
10-19	18.57	23.73	30.53	39.60	51.40	68.91
20-29	11.83	16.50	21.09	27.14	35.20	45.70
30-39	8.99	11.32	15.80	20.19	25.98	33.70
40-49	6.71	8.56	10.78	15.04	19.72	24.74
50-59	4.77	6.15	7.84	9.87	13.78	18.07
60-69	3.05	4.05	5.22	6.65	8.38	11.70
70 +	1.45	1.77	2.35	3.03	3.86	4.86
Fata	2.20	2.74	3.48	4.33	5.40	6.72
Total	82.55	106.96	138.77	179.97	233.99	306.75

Table 6.2 Gini Coefficients of Projected Population by Modified Markov Chain 1981-2031

Projection year	Sex	Gini Coefficient	95% Confidence Interval
1981	Both sex	0.4725	(0.4016, 0.6383)
1991	Both sex	0.4727	(0.3991, 0.6291)
2001	Both sex	0.4717	(0.4023, 0.5920)
2011	Both sex	0.4712	(0.3938, 0.6002)
2021	Both sex	0.4690	(0.3965, 0.6132)
2031	Both sex	0.4716	(0.3991, 0.6463)

Table 6.3 Gini Coefficients of Population Censuses 1951-1998

Population Census	Sex	Gini Coefficient	95% Confidence Interval
	Both	0.3776	(0.3311, 0.4583)
1951	Male	0.3633	(0.3122, 0.4709)
	Female	0.3942	(0.3495, 0.4823)
	Both sex	0.3963	(0.3383, 0.5043)
1961	Male	0.3818	(0.3398, 0.4606)
	Female	0.4132	(0.3515, 0.5149)
	Both sex	0.4504	(0.3895, 0.5512)
1972	Male	0.4337	(0.3711, 0.5553)
	Female	0.4695	(0.4063, 0.5728)
	Both sex	0.4552	(0.4137, 0.5408)
1981	Male	0.4400	(0.3891, 0.5319)
	Female	0.4720	(0.4071, 0.5873)
	Both sex	0.4720	(0.3957, 0.5927)
1998	Male	0.4667	(0.3971, 0.5744)
	Female	0.4777	(0.3954, 0.5834)

Table 6.4 Age Distribution of Population for Selected Years 1998-2030

Year	Percentage of Population in different age groups			
	0-4	5-14	15-64	65+
1998	14.82	28.60	53.10	3.50
2010	13.03	23.20	60.23	3.53
2015	12.89	22.41	60.98	3.73
2020	12.20	21.96	61.93	3.91
2025	11.20	21.58	63.07	4.15
2030	10.13	20.51	64.89	4.46

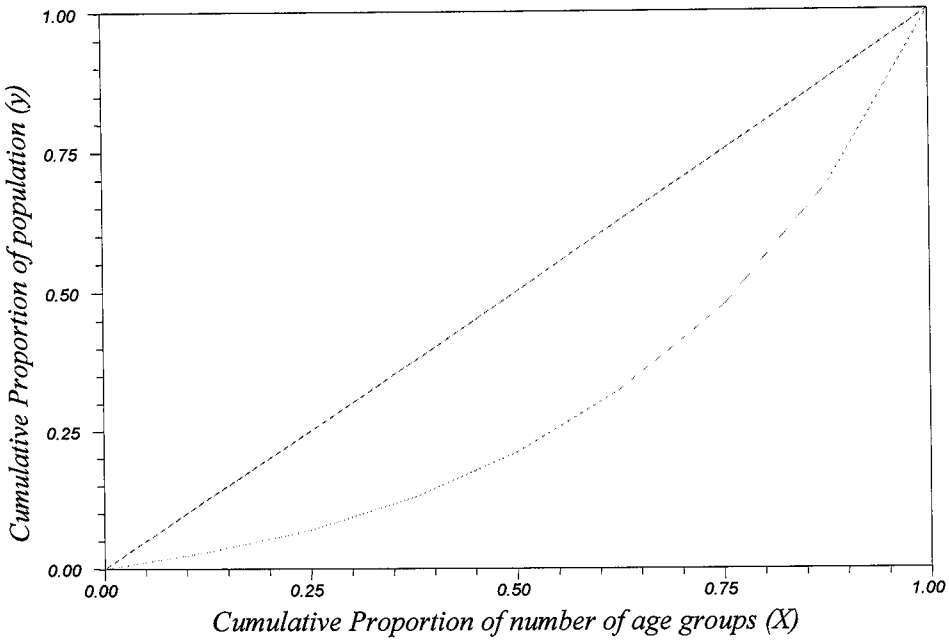


Figure 6.1 Lorenz Curve of Population Census 1972 (Both Sexes)

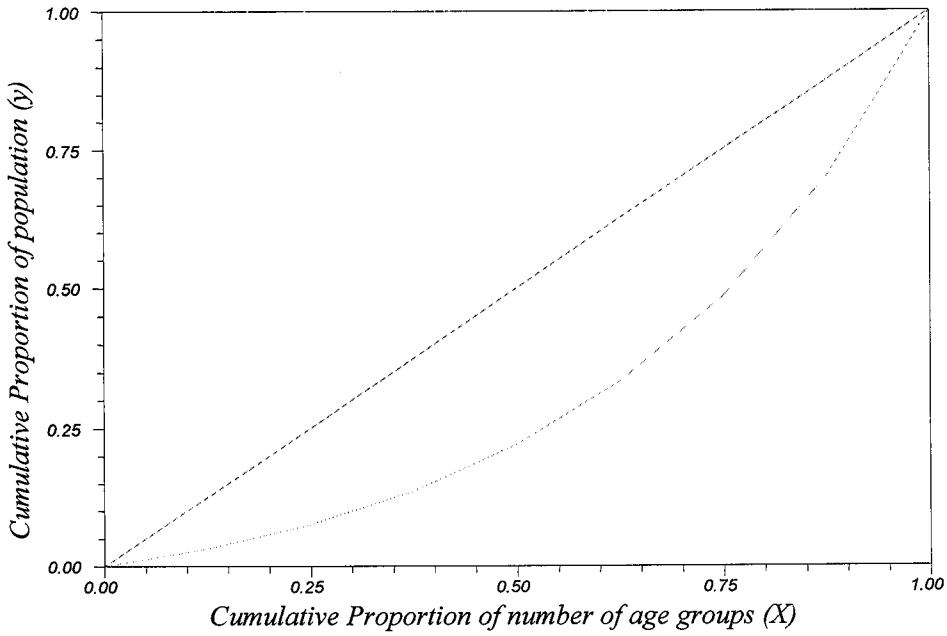


Figure 6.2 Lorenz Curve of Population Census 1972 (Male)

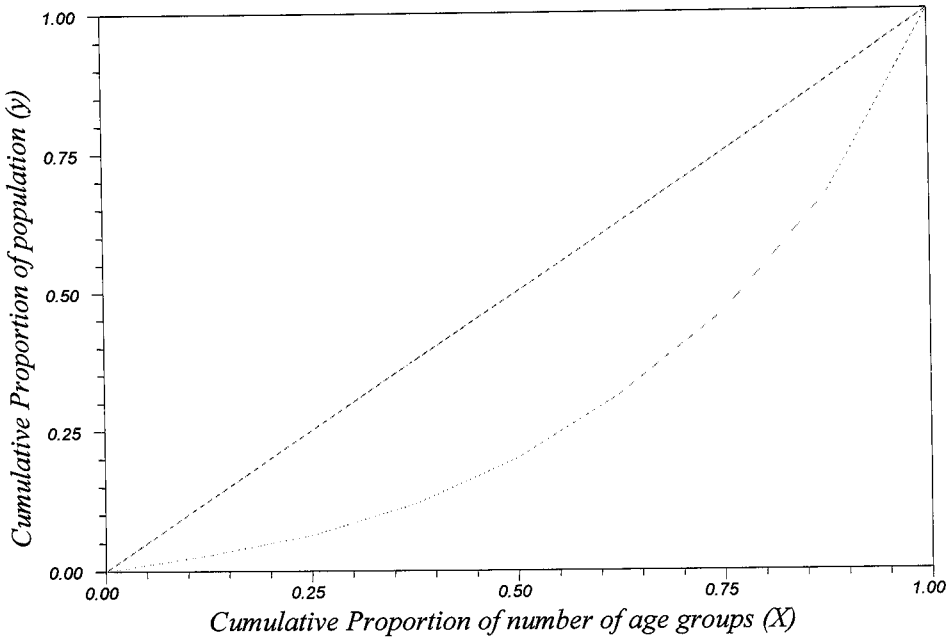


Figure 6.3 Lorenz Curve of Population Census 1972 (Female)

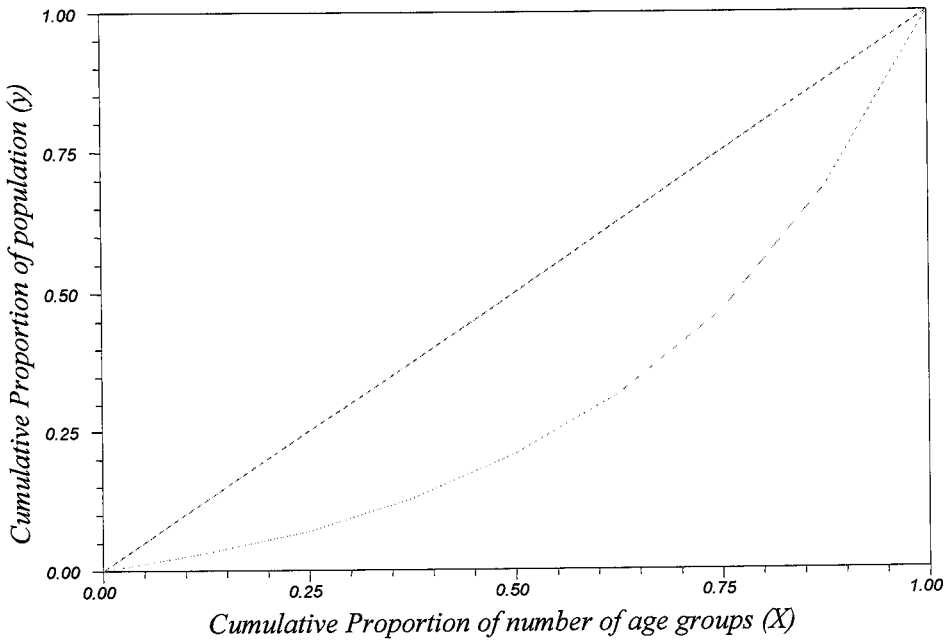


Figure 6.4 Lorenz Curve of Population Census 1981 (Both Sexes)

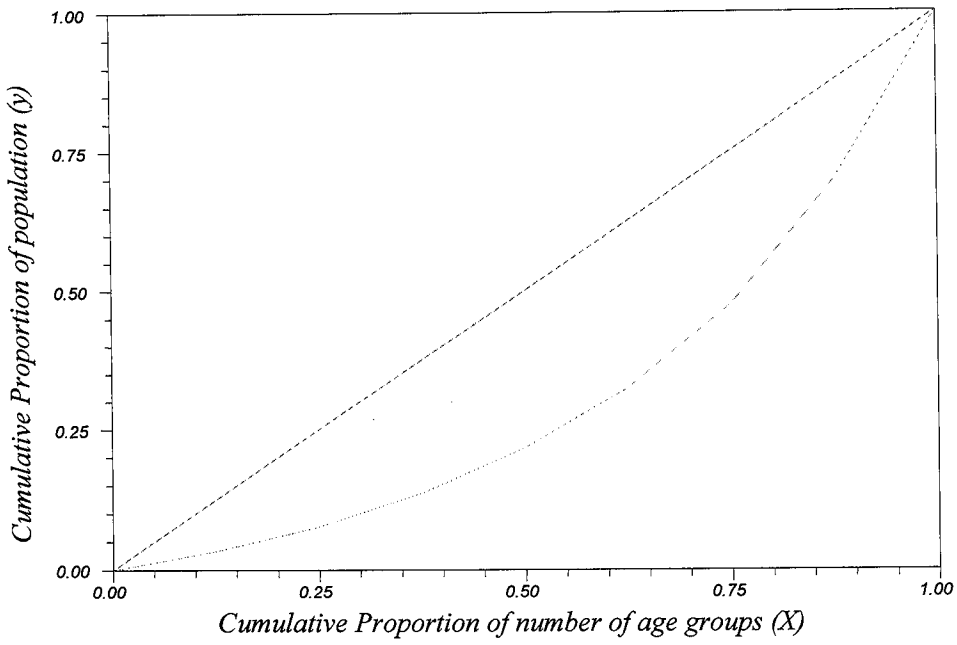


Figure 6.5 Lorenz Curve of Population Census 1981 (Male)

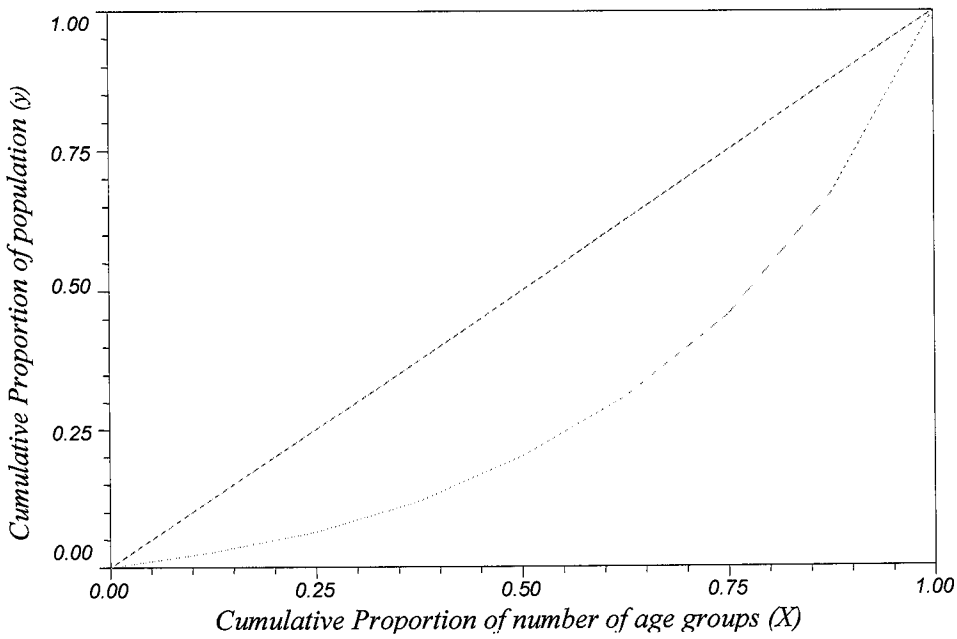


Figure 6.6 Lorenz Curve of Population Census 1981 (Female)

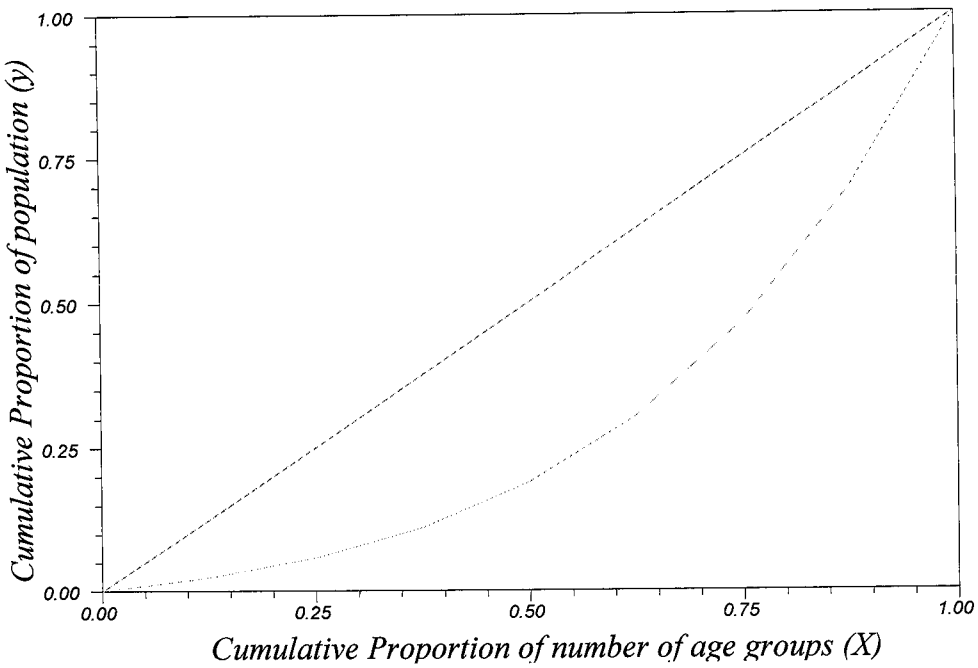


Figure 6.7 Lorenz Curve of Population Census 1998 (Both Sexes)

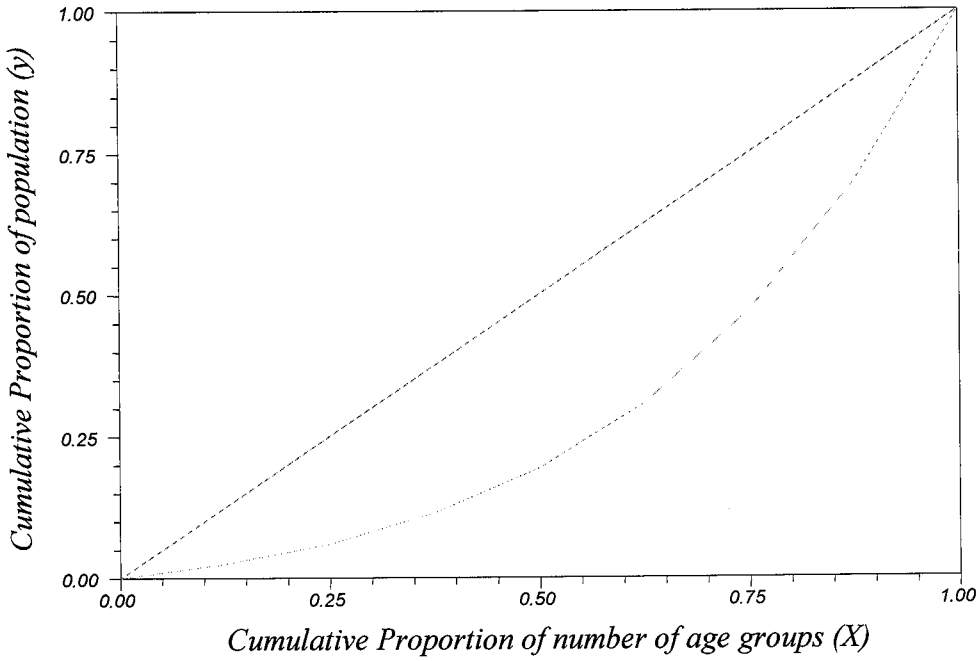


Figure 6.8 Lorenz Curve of Population Census-1998 (Male)

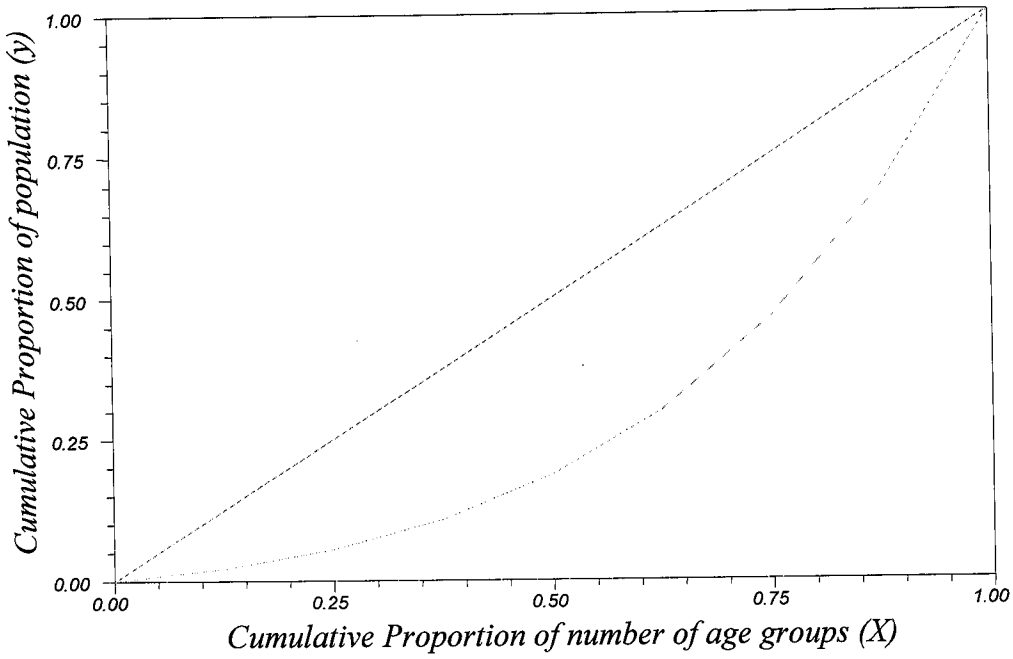


Figure 6.9 Lorenz Curve of Population Census 1998 (Female)

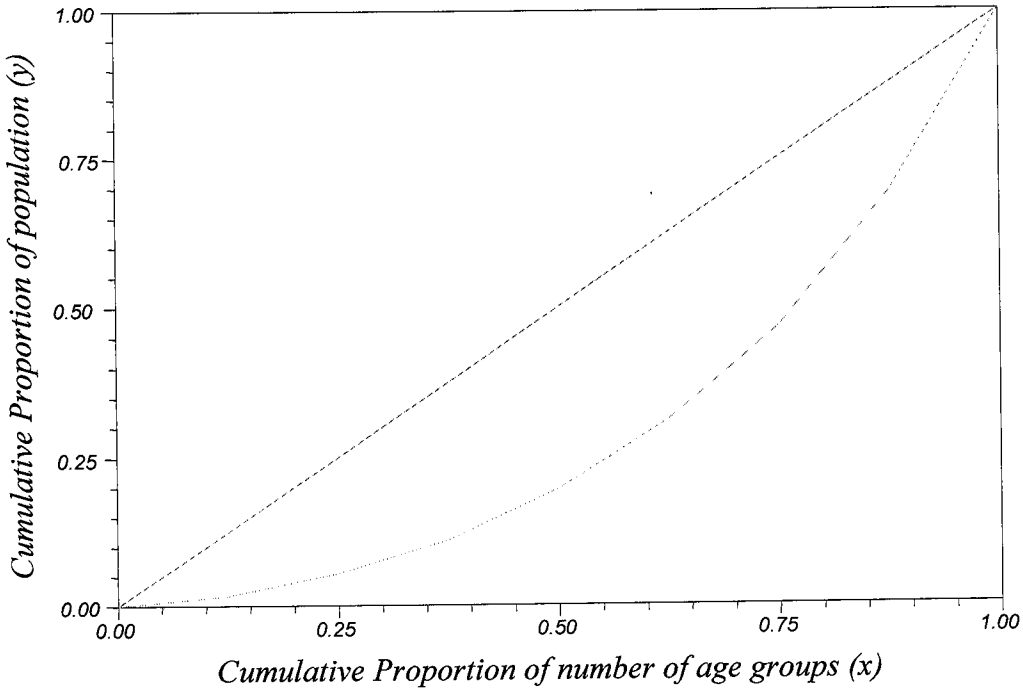


Figure 6.10 Lorenz Curve of Projected Population 2011 (Both Sex)

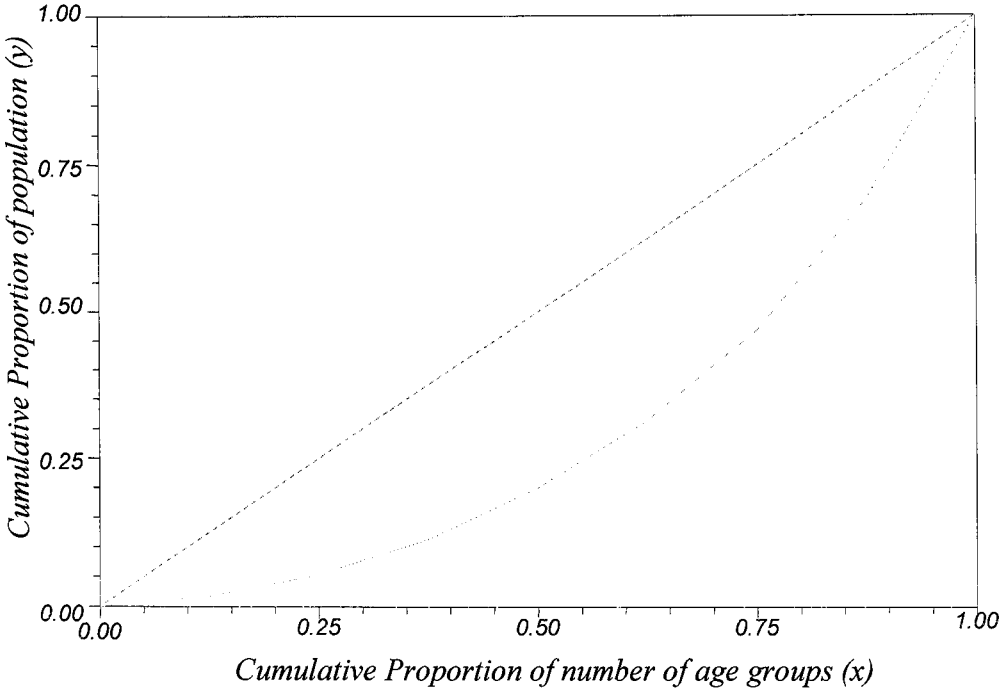


Figure 6.11 Lorenz Curve of Projected Population 2021 (Both Sex)

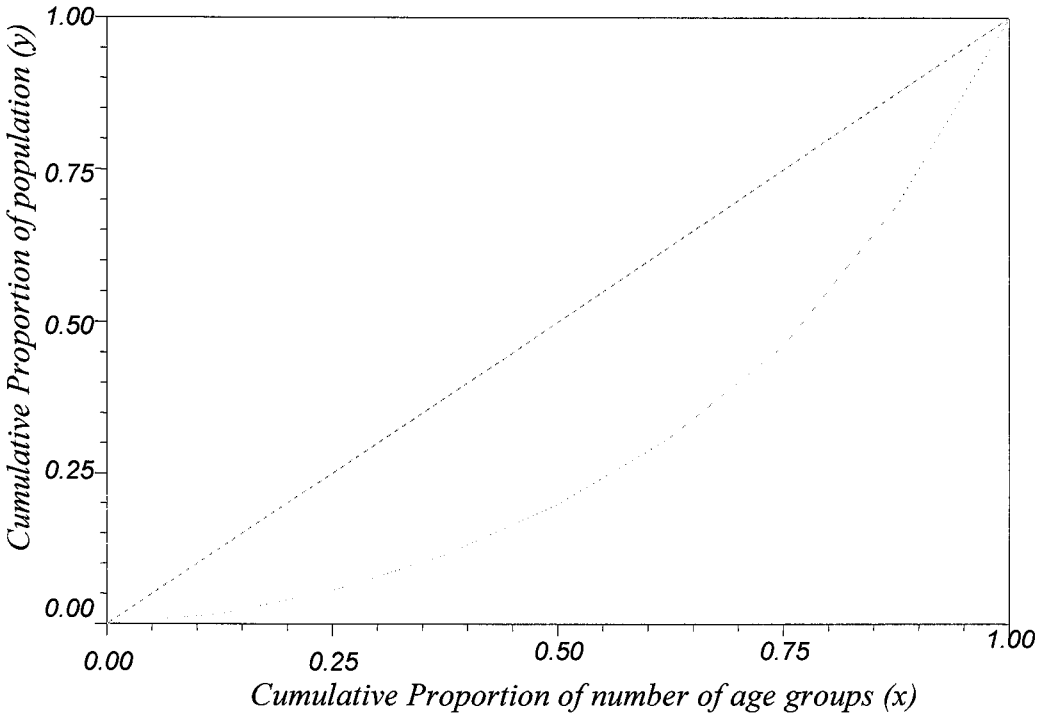


Figure 6.12 Lorenz Curve of Projected Population 2031 (Both Sex)

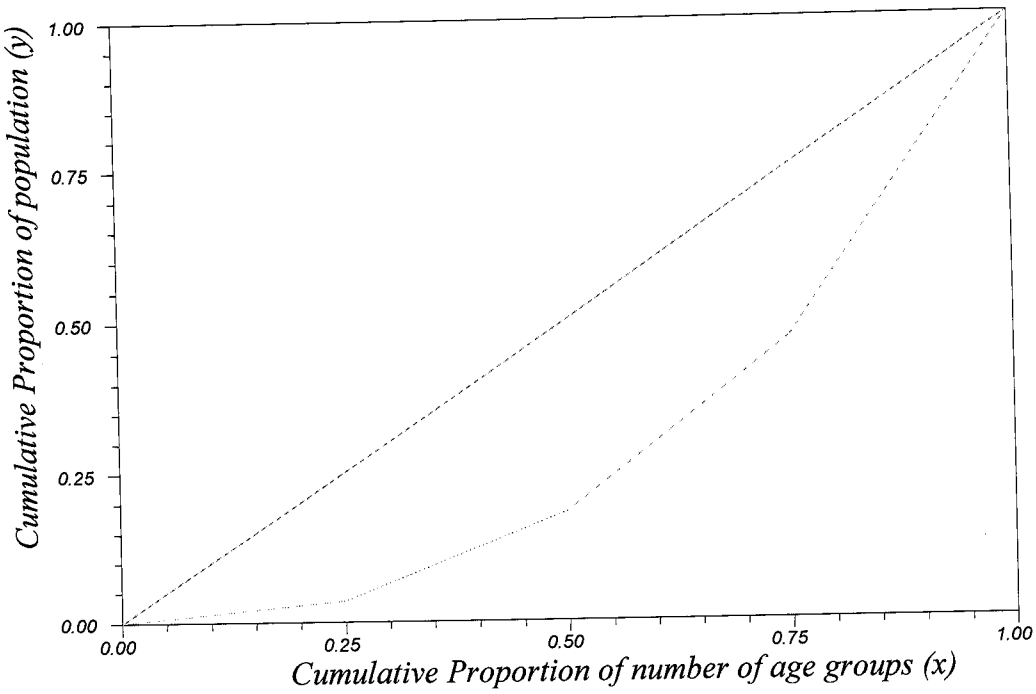


Figure 6.13 Lorenz Curve of Component Projected Population 1998 (Both Sex)

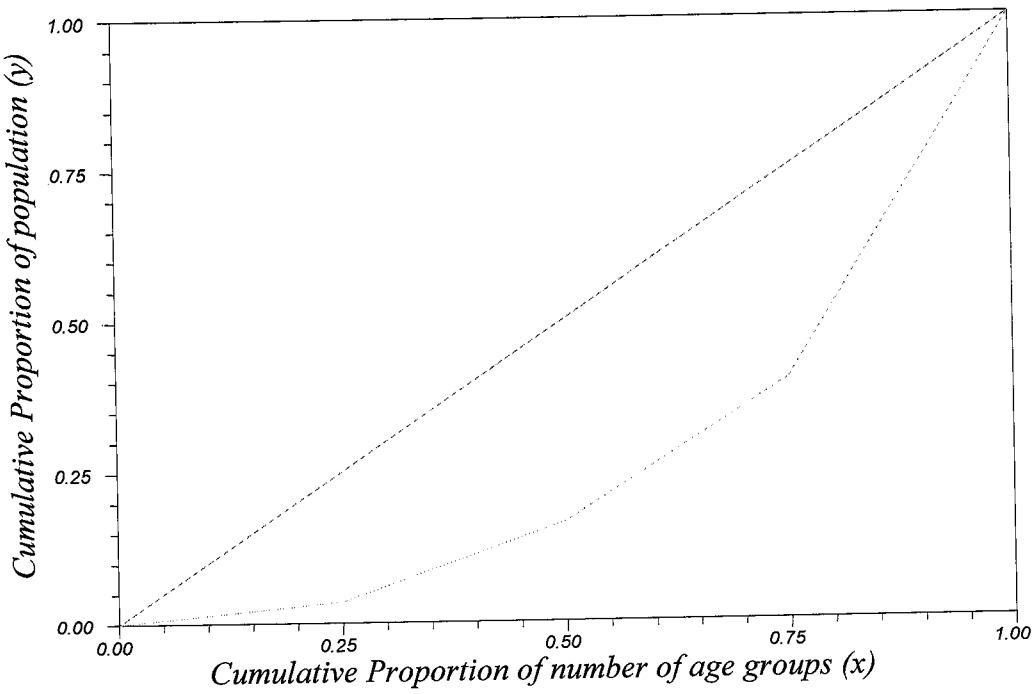


Figure 6.14 Lorenz Curve of Component Projected Population 2010 (Both Sex)

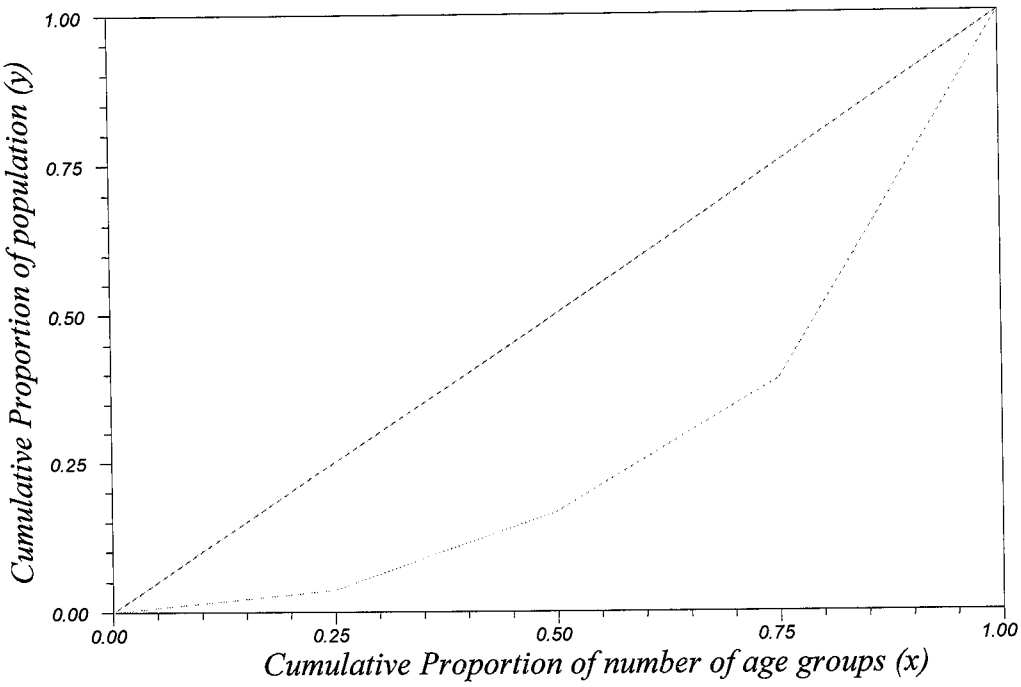


Figure 6.15 Lorenz Curve of Component Projected Population 2015 (Both Sex)

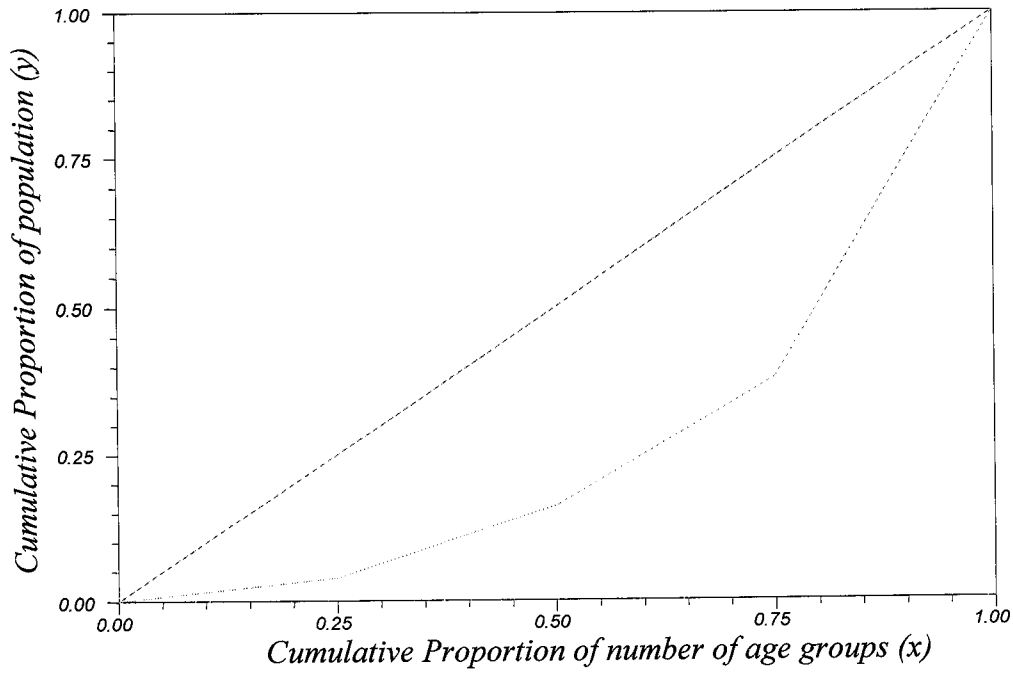


Figure 6.16 Lorenz Curve of Component Projected Population 2020 (Both Sex)

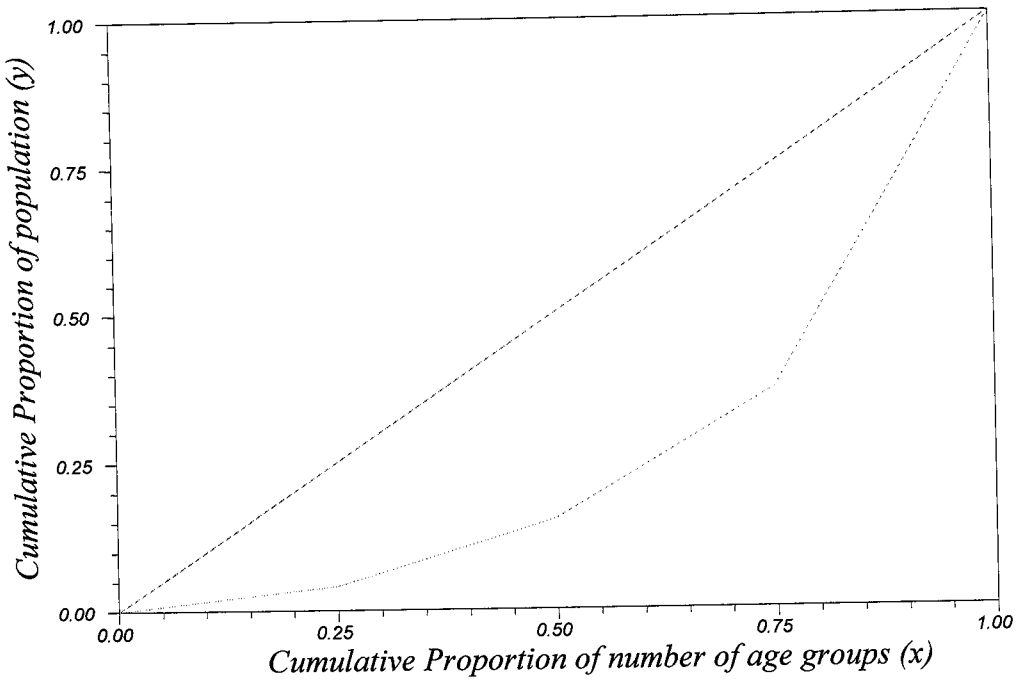


Figure 6.17 Lorenz Curve of Component Projected Population 2025 (Both Sex)

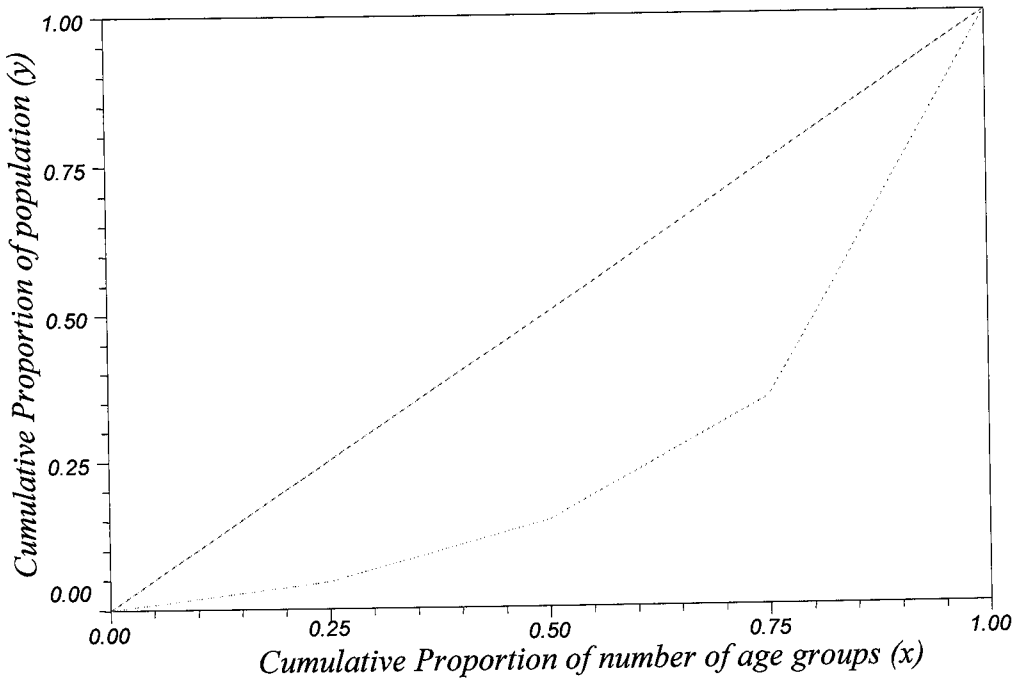


Figure 6.18 Lorenz Curve of Component Projected Population 2030 (Both Sex)

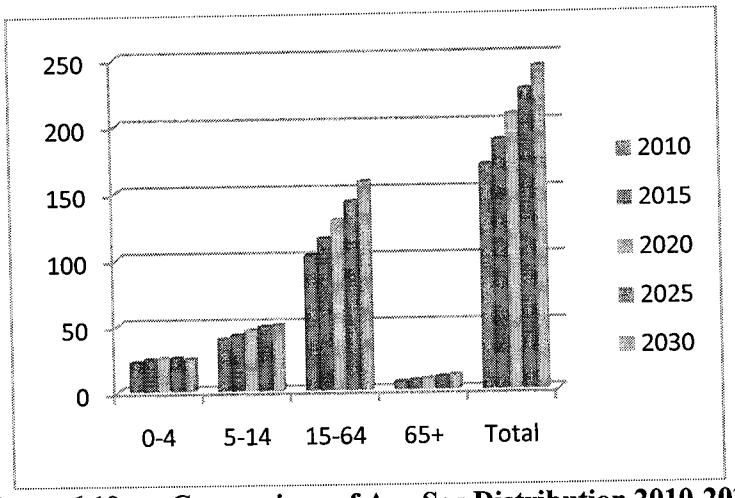


Figure 6.19 Comparison of Age Sex Distribution 2010-2030

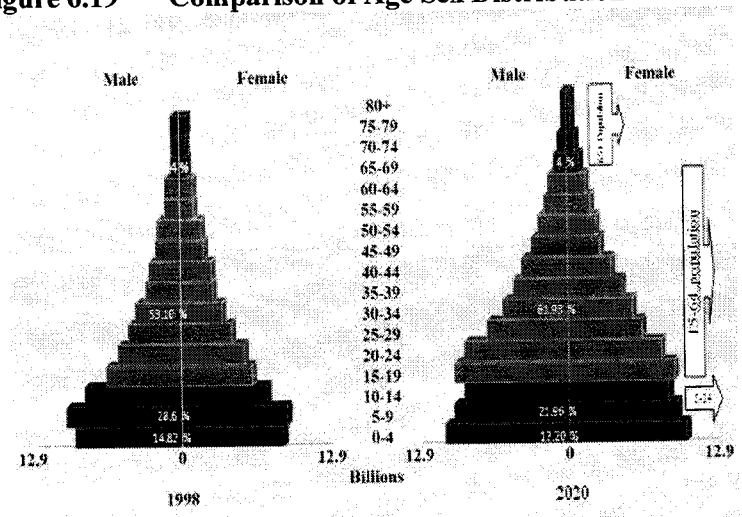


Figure 6.20 Population Pyramids of 1998 & 2020

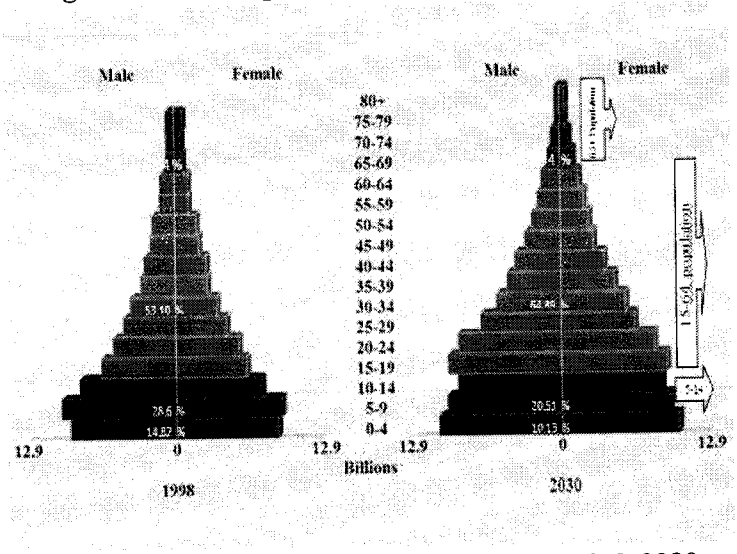


Figure 6.21 Population Pyramids of 1998 & 2030

CHAPTER7

FERTILITY ANALYSIS

7.1 Introduction

In demographic contexts, fertility refers the actual production of offspring and is measureable; it is the opposite of fecundity or infertility (the inability to produce offspring i.e. in a woman it is an inability to conceive; in a man it is an inability to impregnate) and it is not measureable. Basically, the fertility can be measured in two different ways i.e. period measures and cohort measures. Period measures refer to a cross section of the population in one year i.e. Crude birth rate (CBR), General fertility rate (GFR), Child women ratio (CWR) whereas, cohort measures refer to the same people usually over a period of decades i.e. Age specific fertility rate (ASFR), Total fertility rate (TFR), Gross reproduction rate (GRR), Net reproduction rate (NRR) etc. All the demographic measures have their own practical significance according to the situation or requirement. But in this chapter, our main focus is over the cohort measures.

Total fertility rate (TFR) of a country means the average number of children a women have during their childbearing age 15 to 49. Gross reproduction rate (GRR) is the number of baby girls a synthetic cohort will have during the childbearing age. It assumes that all of the baby girls will grow up and live up to at least age 49 whereas Net reproduction rate

(NRR) starts with the GRR and adds the realistic assumption that some of the women will die before age 49. These measures vary from country to country with respect to own country's health status (Shryock et al., 1973).

According to Pakistan Demographic Health Survey, the total fertility rate of Pakistan was 4.1 (PDHS, 2006) whereas Nation Master (2008) reported the total fertility rates of Pakistan, Bangladesh, and India, i.e. 3.73, 3.08 and 2.76 children per woman on the average respectively. Although fertility has been decreasing in Pakistan day by day but it is still very high as compared to the other countries of the world. It seems, it is due to the increased literacy rate and contraceptive measures. Kabir & Mosleh Uddin (1987) used the age specific fertility rates to examine the fertility trend of Bangladesh during 1953-1986. The greater increase has been seen in age specific fertility rate during the period 1953-1974 whereas the age specific fertility rates (ASFRs) started to decrease from 1974 to 1986. The pattern of total fertility rates of Bangladesh was also examined and found to be exactly the same pattern during the same period. Bairagi & Datta (2001) reported that the TFR of Bangladesh decreased up to 1966, then increased up to 1974, then further started to decrease up to 1998. Islam & Ali (2004) computed the demographic cohort measures using the age specific fertility rates of rural Bangladesh during 1980-1998 and fitted the polynomial models taking the age as an explanatory variable. Nasir et al. (2009) used the age specific fertility rates of Pakistan and estimated the demographic cohort measures. Nasir et al. also fitted the second and third degree polynomial models on the ASFRs and tested the goodness of fit of polynomial models using Cross Validity Prediction Power (ρ^2_{cv}). This technique was also used by Islam & Ali (2004) for model

validation. Cohort measures of fertility are computed for Pakistan as well as its rural urban regions for the years 1984-2005. These measures are presented in Table 7.5.

7.2 Objectives

The objectives of this chapter are:

- Computation and comparison of the cohort measures i.e. GRR, NRR and TFR of rural and urban population of Pakistan
- Modeling the age specific fertility rates of urban rural population of Pakistan during 1984-2005
- Testing the goodness of fit of the models

7.3 About the data

Data on live birth population as well as age specific fertility rates (ASFR) of Pakistan and its rural, urban regions are taken from Pakistan Demographic Surveys (PDS, 1984-1986, 1988-1992, 1995-1997, 1999-2001, 2003 and 2005). The available data on live birth population is presented in Table 7.1 and for the years i.e. 1987, 1993, 1994, 1998, 2002-04, data is not available. There are total 16 years for which the ASFR are available (Federal Bureau of Statistics [FBS], 1984-2005).

7.4 Methodology

The age specific fertility rates of Pakistan and its rural, urban regions are used to compute the different demographic cohort measures i.e. Total fertility rate, Gross reproduction rate, Mean age of childbearing. Different polynomial models are also fitted on the ASFRs (Nasir 2009; Shryock et al. 1973; Montgomery et al. 1982; Keyfitz 1968). Minitab-14 statistical software is used to compute the demographic measures as well as and for

model fitting. StatDirect statistical package is used to compute the inequality among age specific fertility rates during different years. The mathematical form of the total fertility rate, using seven 5-year age groups is given by

$$TFR = 5 \sum_{i=1}^{7} \frac{B_i}{P_i} \, x \, 1{,}000$$

Where B_i is the number of live births registered during the year to mothers of age group i, i is an interval of 5 years, P_i is the mid-year population of women of the same age (Shryock, 1973).

The mathematical form of the gross reproduction rate is given by

$$GRR = \sum_{w_1}^{w_2} \frac{B_x}{P_x} . K$$

Where B_x is the number of female infants born to mothers of age x, P_x is the number of women of age x in the midyear population, w_1 and w_2 are, respectively, the lower and upper limits of the childbearing period, and K is a constant-unity (1), or 100, or 1000 (Shryock, 1973).

If the computation uses 5-year age groups, the formula of GRR becomes

$$GRR = 5 \sum_{i=1}^{7} \frac{B_i}{P_i} . K$$

The mathematical form of the mean age of childbearing is given by

$$\bar{a} = \frac{\sum_a a f_a}{\sum_a f_a}$$

Where a represents the midpoint of each age interval and f_a represents an age-specific birth rate for a 5-year age group (Shryock, 1973). The mathematical form of the nth degree polynomial model for the age specific fertility rates, forward cumulative age specific fertility rates and backward cumulative age specific fertility rates respectively are

$$y = b_o + \sum_{j=1}^{p} b_j z^j + \varepsilon$$

$$y^f = b_o + \sum_{j=1}^{p} b_j z^j + \varepsilon$$

$$y^b = b_o + \sum_{j=1}^{p} b_j z^j + \varepsilon$$

Where z is the mid value of age group in years, y is the age specific fertility rates, y^f is the forward cumulative age specific fertility rates, y^b is the backward cumulative age specific fertility rates, b_o is the constant, b_j is the coefficient of z^j (j =1, 2, 3, 4,.....,p) and ε is the stochastic error term of the model (Nasir, 2009).

7.5 Results and discussion

Table 7.2, Table 7.3 and Table 7.4 reveal the available age specific fertility rates of the rural, urban regions as well as of Pakistani women respectively during their childbearing age from 15 to 49 years during the years 1984-2005. ASFR of rural women of Pakistan was 72.96 in 1984 for the age 15-49 and it remained almost same up to 1988 with a little bit variation. A slight increase has been seen during the years 1989-1992 and then started to decrease from 1995 to to-date. The reason might be the increase of female literacy rate as well as the trend of late marriages. In 2005, the age specific fertility rate was 26.20 which is about 65% less as compared to the age specific fertility rate of 1984. The age specific fertility rate remains high during the age 20-39 years whereas the most fertile period of women is 25-34 years. The greater reduction has been seen in the ASFRs during the ages 15-19 and 40-49.

In 1984, the age specific fertility rates of urban women and of Pakistan of the ages 15-19 were 52.18 and 65.8 respectively whereas, in 2005, these rates were reported 10.70 and

20.3 respectively for the same age and regions. Similarly, in 1984, these rates were noted 35.06 and 37.9 respectively for the same regions but of the ages 45-49 whereas in 2005, these rates were 11.60 and 18.1 respectively for the same regions. It is evidently clear that greater reduction in fertility has been observed in urban areas as compared to the rural areas as well as of Pakistan (FBS, 1984-2005).

Table 7.5 presents the total fertility rates, gross reproduction rates and mean age of childbearing of rural, urban regions as well as of Pakistan. In 1984, Total fertility rate of rural population of Pakistan was 7.27 whereas in 2005, it was 4.09 which is about 43.74% decrease in total fertility rate during the 21 years. Similarly total fertility rate of urban area was 6.24 and 3.29 for the same years which shows approximately 47.27% decrease during the same time period. On the other hand, the total fertility rate of Pakistan during the same years was 6.95 and 3.79 respectively which shows 45.46% decrease as compared to the year 1984. Since a long time, the government of Pakistan has started different contraceptive measures to decrease the fertility of the country, although fertility decreased in the country but less reduction has been examined in the rural areas as compared to urban as well as of Pakistan whereas, the reduction in gross reproduction rates of rural, urban and of Pakistan has almost similar. Mean age of childbearing is also computed for the same year which was around 30 years in 1984 and around 29.75 years in 2005, it indicates the decreasing trend in fertility as well as in mean age of childbearing. The reason of less decrease in fertility in rural areas might be the lack of facilities of education, health, exposure to mass media, early marriages, poor information of family planning methods and greater influence of religious leaders over innocent residents as well as male dominant societies etc.

Table 7.6 consists of the Gini coefficients along with the confidence intervals of the age specific fertility rates of 1990, 1995, 2000 and 2005 of Pakistan as well as its rural, urban areas. Table 7.6 also consists of the Gini coefficients of total fertility rates of Pakistan and its rural urban areas for the years 1984-2005. The Gini coefficients of the year 1990, 1995 and 2000 are less than the Gini coefficient of 2005 in the three regions. The reason might be the greater variation in the age specific fertility rates of 2005, especially in the ages 15-19 and 45-49.

Table 7.7 reveals the estimated R^2 and Cross Validity Prediction Power (ρ^2_{cv}) of different models fitted on the age specific fertility rates, forward cumulative age specific fertility rates, backward cumulative age specific fertility rates of rural area, urban area and of Pakistan. The third degree polynomial is fitted on the age specific fertility rates of rural area, urban area and of Pakistan which gives almost the same value of R^2 (0.99) for three regions rural, urban and Pakistan. Similarly, the magnitude of Cross Validity Prediction Power (ρ^2_{cv}) is 0.958, 0.87, and 0.953 respectively for the same regions. The value of Cross Validity Prediction Power of Pakistan and its rural area is close to each other but differ from the urban area.

In the middle of Table 7.7, a second degree polynomial is fitted on the forward cumulative age specific fertility rates of Pakistan and its regions (rural and urban). The magnitude of R^2 and Cross Validity Prediction Power (ρ^2_{cv}), of rural area (0.994, 0.984), urban area (0.99, 0.97) and of Pakistan are (0.993, 0.979) respectively. The magnitude of R^2 is same as that of the Cross Validity Prediction Power (ρ^2_{cv}) is same for the three regions.

In the last column of Table 7.7, the same third degree polynomial is fitted on the backward cumulative age specific fertility rates of rural area, urban area and of Pakistan. The computed value of R^2 for all three regions is exactly same i.e. 0.998 whereas the Cross Validity Prediction Power (ρ^2_{cv}) value is 0.987 which is also same for three regions. Consequently, the third degree polynomial is recommended for Pakistan and its rural urban regions using the age specific fertility rates.

Figure 7.1 shows the trend of age specific fertility rates of rural area of women of ages 15-49 during the years 1984-2005. It indicates that at both ends of childbearing age i.e. 15-19, 45-49; the fertility is least as compared to the other childbearing ages. More precisely, the most fertile period of rural women is 20-39 years. Graph also indicates the most fertile period of women age which is 25-29 years during (1984-2005). Although fertility decreased with the passage of time but the most fertile period remained same. This might be due to increased literacy rate, improved health sciences and contraceptive measures etc. The fertility pattern of Pakistan and its rural urban regions is similar to the reciprocal of V shape. Figures 7.2, 7.3 and 7.4 present the predicted polynomial fit on the age specific fertility rates, forward cumulative age specific fertility rates, backward cumulative age specific fertility rates of rural areas women of Pakistan of the year 2005 respectively.

Figure 7.5 displays the trend of ASFRs of urban women which is little bit flatter than that of the trend of rural women given in Figure 7.1. Figures 7.6, 7.7, and 7.8 reveal the polynomial fit of order third, second and third on the age specific fertility rates, forward cumulative age specific fertility rates, backward cumulative age specific fertility rates of urban women of Pakistan of the year 2005 respectively. Figure 7.9 shows the ASFRs

pattern of the Pakistani women of the year 2005 which is approximately similar pattern as that of rural and urban women of the same year. whereas Figures 7.10, 7.11 and 7.12 present the trend of polynomial fit on the age specific fertility rates, forward cumulative age specific fertility rates and backward cumulative age specific fertility rates of Pakistani women of the year 2005 respectively.

Figure 7.13, 7.14 and 7.15 reveal the trend of total fertility rate and gross reproduction rates of Pakistan and its rural, urban regions for the years (1984-2005) respectively. A greater reduction has been seen in the urban area as compared to Pakistan and its rural area. Figure 7.16 presents the comparison of total fertility rates of Pakistan and its two geographical regions i.e. rural and urban. There is smooth decline in total fertility rates of Pakistan during 1984-2005 whereas greater variation has been examined in the total fertility rates of urban population especially during the period from 1990 to 2005. In the last decade, a gradual reduction is visualized in Pakistan, its rural and urban areas.

Figures 7.17, 7.18, 7.19 and 7.20 reveal the Lorenz curve of the age specific fertility rates of 1990, 1995, 2000 2005 of rural areas of Pakistan. Almost similar increase in variation is seen in each five years gap but in 2005, there is a greater variation as compared to 1984. Similarly, the Figures 7.21, 7.22, 7.23 and 7.24 present the Lorenz curve of the age specific fertility rates of 1990, 1995, 2000 2005 of urban areas of Pakistan whereas Figures 7.25, 7.26, 7.27 and 7.28 also present the Lorenz curve of the age specific fertility rates of 1990, 1995, 2000 and 2005 of Pakistan. Approximately similar change in variation has been examined in Pakistan and its urban areas during the same years.

7.6 Conclusion and recommendations

In this chapter, the fertility of Pakistan and its two main geographical regions urban and rural has been studied during 1984 to 2005. The age specific fertility rates have a traditional reciprocal of V shape pattern. Fertility remained very high in the rural areas than the urban areas throughout the period 1984-2005. The Total fertility rates (TFR) of urban, rural and of Pakistan were 7.27, 6.24 and 6.95 respectively in 1984 whereas these rates were decreased up to 4.09, 3.29 and 3.79 in 2005 for the same regions respectively. Approximately 43.7% decrease in TFR has been examined in rural areas, 47.3% decrease in urban areas and 45.5% decrease in Pakistan during 1984-2005. Age inequality has also been examined and presented by computing the Gini coefficient and Lorenz curves. The variation among the Gini coefficients of 1990, 1995 and 2000 is less as compared to the Gini coefficient of 2005 for rural, urban and of Pakistan. It indicates that in earlier time the fertility started from early ages and continued till their childbearing age. The Gini coefficient for the year 2005 indicates that age inequality is increased especially in urban areas. It might be due to increased female literacy rate as well as the increase mean age of child bearing. Different polynomial models are also fitted on the age specific fertility rates, backward cumulative age specific fertility rates, forward cumulative age specific fertility rates of Pakistan and its rural urban regions. Third degree polynomial model is found to be a good and may be recommended to fit on the age specific fertility rates of Pakistan and its rural urban regions.

Table 7.1 Live Births Population Sex and Region Wise of Pakistan during 1984-2005

Years	Rural			Urban			Pakistan		
	Both Sexes	Male	Female	Both Sexes	Male	Female	Both Sexes	Male	Female
1984	3515	1809	1706	2588	1372	1216	6103	3181	2922
1985*	3553	1834	1719	2774	1414	1360	6327	3248	3079
1986*	3618	1867	1751	2820	1469	1351	6438	3336	3102
1988**	2286538	1169144	1117394	908388	497437	410951	3194926	1666581	1528345
1989	2530364	1307976	1222388	1045595	551313	494283	3575959	1859289	1716671
1990	2631840	1382878	1248961	976837	509169	467669	3608678	1892048	1716631
1991	2597137	1354620	1242516	996167	512350	483817	3593303	1866970	1726333
1992	2659778	1356939	1302839	969020	493873	475148	3628799	1850812	1777987
1995	3002793	1554875	1447918	1263670	646608	617062	4266463	2201483	2064980
1996	3618	1867	1751	2820	1469	1351	6438	3336	3102
1997	2945221	1521859	1423362	1198286	627177	571109	4143507	2149035	1994472
1999	2280311	1191850	1088461	1596347	828696	767651	3876658	2020546	1856112
2000	2291364	1197281	1094083	1515624	778924	736700	3806988	1976204	1830784
2001	2527504	1332090	1195414	1192190	610755	581435	3719694	1942485	1776849
2005	2546343	1340012	1206331	1226151	653480	572671	3772494	1993492	1779002

Table 7.2 Age Specific Fertility Rates (Per 1000 women) of Rural Areas

Years	Age Specific Fertility Rates						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1984	72.96	274.60	373.82	325.38	250.23	118.09	39.06
1985	64.76	268.25	360.43	344.10	246.40	124.47	55.03
1986	59.38	280.81	359.57	303.17	239.57	141.01	57.34
1988	77.60	281.40	341.50	285.60	219.70	126.60	49.90
1989	90.10	273.70	328.00	273.50	218.60	114.00	49.20
1990	90.50	294.90	320.40	294.10	195.90	111.30	35.20
1991	81.90	275.60	322.70	272.30	211.60	95.70	34.00
1992	90.80	285.90	324.90	272.60	184.90	82.80	35.30
1995	68.60	252.70	312.00	258.80	160.70	101.40	34.30
1996	58.00	293.70	301.00	267.30	157.60	79.00	27.50
1997	59.60	251.90	281.10	222.00	156.30	75.90	36.80
1999	45.00	222.50	265.00	213.00	133.20	76.40	38.60
2000	41.20	219.00	258.70	220.50	141.10	66.00	35.10
2001	27.80	178.40	251.30	208.70	133.00	67.10	27.30
2003	29.70	178.90	236.10	201.60	126.50	57.80	23.70
2005	26.20	173.60	233.90	185.90	115.40	62.00	21.90

Source: Federal Bureau of Statistics (1984-2005) * Sample; ** Estimated

Table 7.3 Age Specific Fertility Rates (Per 1000 women) of Urban Areas

Years	Age Specific Fertility Rates						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1984	52.18	255.13	353.39	289.34	173.81	89.99	35.06
1985	48.21	283.15	329.41	289.68	211.06	70.00	31.61
1986	44.67	234.11	361.81	303.00	196.68	93.51	40.37
1988	45.40	229.40	315.00	262.50	167.30	80.50	22.70
1989	49.00	249.80	314.10	276.00	151.30	75.30	24.70
1990	46.80	233.50	297.70	239.20	133.50	60.70	20.20
1991	44.60	223.60	299.80	230.80	134.50	54.00	13.60
1992	40.50	214.40	287.20	217.10	115.70	57.50	11.50
1995	42.30	225.00	290.90	208.30	123.50	67.60	19.80
1996	46.50	202.80	285.10	231.90	113.50	37.00	16.00
1997	38.10	189.70	258.00	189.80	116.60	52.60	17.60
1999	27.10	186.20	247.20	191.80	101.20	45.70	11.00
2000	24.00	168.80	226.90	183.30	85.90	40.50	9.30
2001	18.60	135.80	227.90	176.70	96.20	41.90	12.60
2003	14.20	137.10	219.30	169.20	93.40	34.20	11.20
2005	10.70	132.10	210.90	169.60	92.50	31.00	11.60

Source: Federal Bureau of Statistics (1984-2005)

Table 7.4 Age Specific Fertility Rates (Per 1000 women) of Pakistan

Years	Age Specific Fertility Rates						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1984	65.8	268.3	367.6	314.4	226.1	109.6	37.9
1985	59.2	273.0	350.8	327.0	235.3	108.6	47.9
1986	54.3	265.8	360.3	303.1	226.2	126.0	52.2
1988	66.0	263.6	333.0	278.3	203.3	111.2	41.8
1989	75.7	265.8	323.4	274.3	197.1	102.0	41.6
1990	75.5	274.8	313.2	276.0	175.9	97.0	30.5
1991	69.0	258.2	315.4	259.0	186.5	82.3	27.4
1992	73.3	261.4	312.9	254.5	162.6	74.5	27.8
1995	59.1	243.4	305.1	241.9	148.1	90.1	29.6
1996	54.7	258.2	295.9	255.4	143.0	65.5	23.2
1997	52.3	231.0	273.2	211.2	142.9	68.4	30.7
1999	36.2	205.6	256.9	203.6	118.3	61.7	25.8
2000	32.9	195.1	244.2	203.8	114.5	54.4	22.9
2001	24.2	162.0	242.9	197.2	118.5	57.9	21.9
2003	23.7	163.1	229.6	190.0	112.7	49.0	18.8
2005	20.3	157.6	225.5	179.9	106.6	50.1	18.1

Source: Federal Bureau of Statistics (1984-05)

Table 7.5 TFR, GRR and Mean Age of Childbearing (MAC) of Pakistan and its Rural Urban Regions

Years	Rural			Urban			Pakistan		
	TFR	GRR	MAC	TFR	GRR	MAC	TFR	GRR	MAC
1984	7.27	3.53	30.15	6.24	2.93	29.75	6.95	3.33	30.05
1985	7.32	3.54	30.53	6.32	3.10	29.65	7.01	3.41	30.29
1986	7.20	3.49	30.27	6.37	3.05	30.20	6.94	3.34	30.49
1988	6.91	3.38	30.14	5.61	2.54	29.71	6.49	3.10	30.05
1989	6.74	3.25	29.95	5.70	2.70	29.44	6.40	3.07	29.83
1990	6.71	3.18	29.55	5.16	2.47	29.14	6.21	2.96	29.47
1991	6.47	3.09	29.62	5.00	2.43	29.02	5.99	2.88	29.47
1992	6.39	3.13	29.21	4.72	2.31	28.97	5.84	2.86	29.17
1995	5.94	2.87	29.66	4.89	2.39	29.19	5.59	2.70	29.53
1996	5.92	2.87	29.19	4.66	2.39	28.81	5.30	2.55	29.18
1997	5.42	2.62	29.48	4.31	2.06	29.23	5.05	2.43	29.42
1999	4.97	2.37	29.77	4.05	1.95	29.07	4.54	2.17	29.48
2000	4.91	2.34	29.75	3.69	1.80	29.01	4.34	2.09	29.46
2001	4.47	2.11	30.08	3.55	1.73	29.62	4.12	1.97	29.94
2003	4.26	*	29.82	3.39	*	29.49	3.93	2.03	29.71
2005	4.09	1.94	29.83	3.29	1.54	29.59	3.79	1.79	29.75

* indicates the missing data (unavailable) (Nasir, 2009)

Table 7.6 Inequality Measures of ASFRs of Pakistan and its Regions 1990-2005

Region	Years	Gini Coefficient	Confidence Interval
Rural	1990	0.3594	(0.2540-0.5286)
	1995	0.3828	(0.2917- 0.5189)
	2000	0.4015	(0.2553- 0.4974)
	2005	0.4343	(0.2828- 0.5695)
Urban	1990	0.4492	(0.3239- 0.5982)
	1995	0.4499	(0.3431- 0.6102)
	2000	0.4962	(0.3369- 0.6664)
	2005	0.5152	(0.2927- 0.6841)
Pakistan	1990	0.3827	(0.2809- 0.5490)
	1995	0.4018	(0.3044- 0.5569)
	2000	0.4403	(0.2959- 0.5685)
	2005	0.4612	(0.3009- 0.6109)
Rural (TFR)	1984-2005	0.1099	(0.0909- 0.1341)
Urban (TFR)	1984-2005	0.1285	(0.1111- 0.1566)
Pakistan (TFR)	1984-2005	0.1198	(0.1026- 0.1451)

Table 7.7 Model Fitting and Goodness of Fit on the ASFR, Forward Cumulative ASFR, Backward Cumulative ASFR of 2005 of Pakistan & its Rural Urban Regions

Model	*Model of Age Specific Fertility Rates* $Y = \beta_o + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$				*Model of Forward Cumulative ASFR* $Y = \beta_o + \beta_1 X + \beta_2 X^2$			*Model of Backward Cumulative ASFR* $Y = \beta_o + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$			
Rural	β_o	β_1	β_2	β_3	β_1	β_2	β_3	β_o	β_1	β_2	β_3
Coeff. (P-value)	-1848 (0.0013)	190.10 (0.0013)	5.56 (0.0019)	0.050 (0.0028)	-1141 (0.0007)	81.28 (0.0005)	-0.84 (0.0025)	-526.70 (0.234)	176.90 (0.0173)	-6.98 (0.010)	0.074 (0.020)
	$R^2 = 0.99$, $\rho^2_{cv} = 0.958$				$R^2 = 0.994$, $\rho^2_{cv} = 0.984$			$R^2 = 0.998$, $\rho^2_{cv} = 0.986$			
Urban	β_o	β_1	β_2	β_3	β_o	β_1	β_2	β_o	β_1	β_2	β_3
Coeff. (P-value)	-1814 (0.006)	184.60 (0.0064)	-5.41 (0.0088)	0.049 (0.0129)	-1042 (0.0019)	73.43 (0.0016)	-0.79 (0.006)	-828.70 (0.060)	187.40 (0.0077)	-7.14 (0.005)	0.08 (0.0048)
	$R^2 = 0.98$, $\rho^2_{cv} = 0.87$				$R^2 = 0.99$, $\rho^2_{cv} = 0.97$			$R^2 = 0.998$, $\rho^2_{cv} = 0.987$			
Pakistan	β_o	β_1	β_2	β_3	β_o	β_1	β_2	β_o	β_1	β_2	β_3
Coeff. (P-value)	-1838 (0.0019)	188.30 (0.002)	-5.51 (0.0028)	0.05 (0.0042)	-1104 (0.001)	78.34 (0.0008)	-0.82 (0.0036)	-645.00 (0.1401)	181.30 (0.0126)	-7.05 (0.0077)	0.07 (0.0074)
	$R^2 = 0.99$, $\rho^2_{cv} = 0.953$				$R^2 = 0.993$, $\rho^2_{cv} = 0.979$			$R^2 = 0.998$, $\rho^2_{cv} = 0.987$			

Cross Validity Prediction Power = ρ^2_{cv}

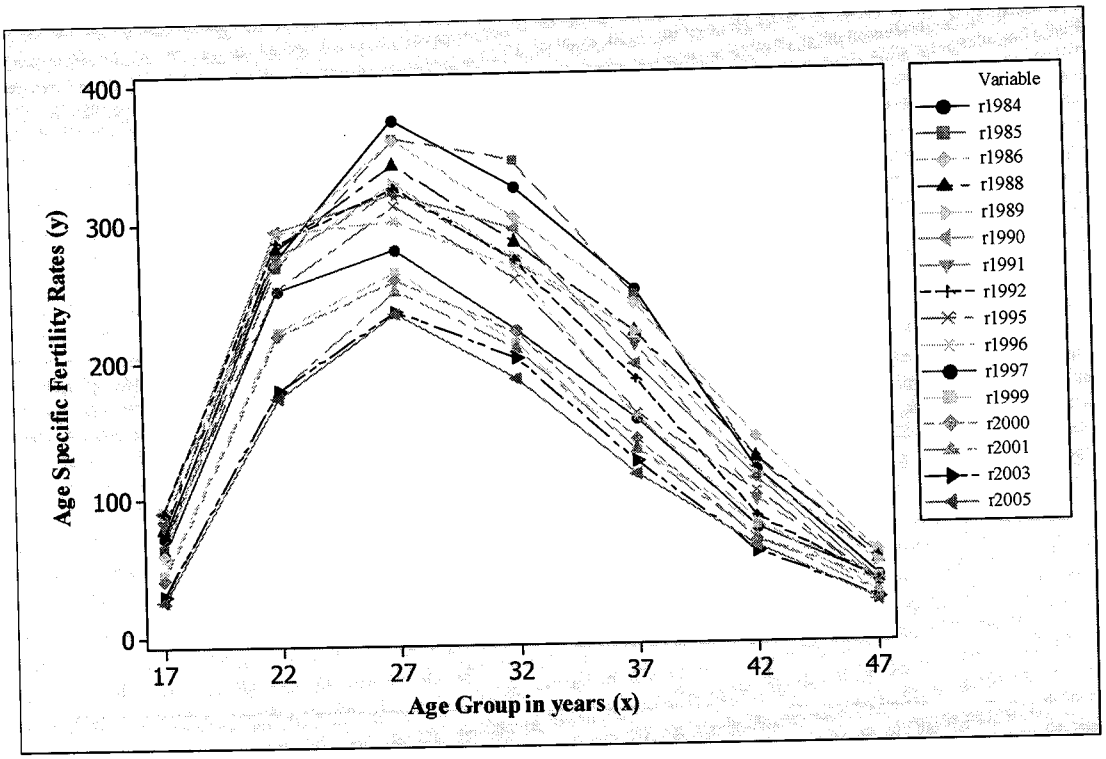


Figure 7.1 Age Specific Fertility Rates Trend of Rural Pakistan 1984-2005

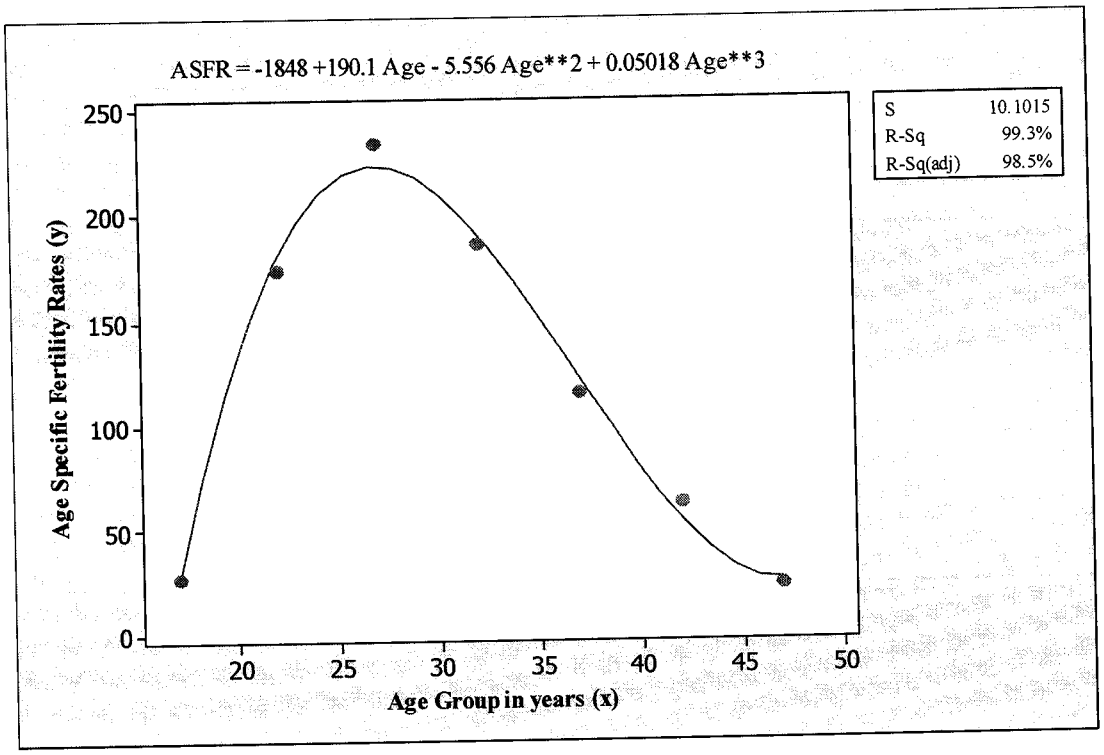


Figure 7.2 Modeling the Age Specific Fertility Rates of 2005 of Rural Areas

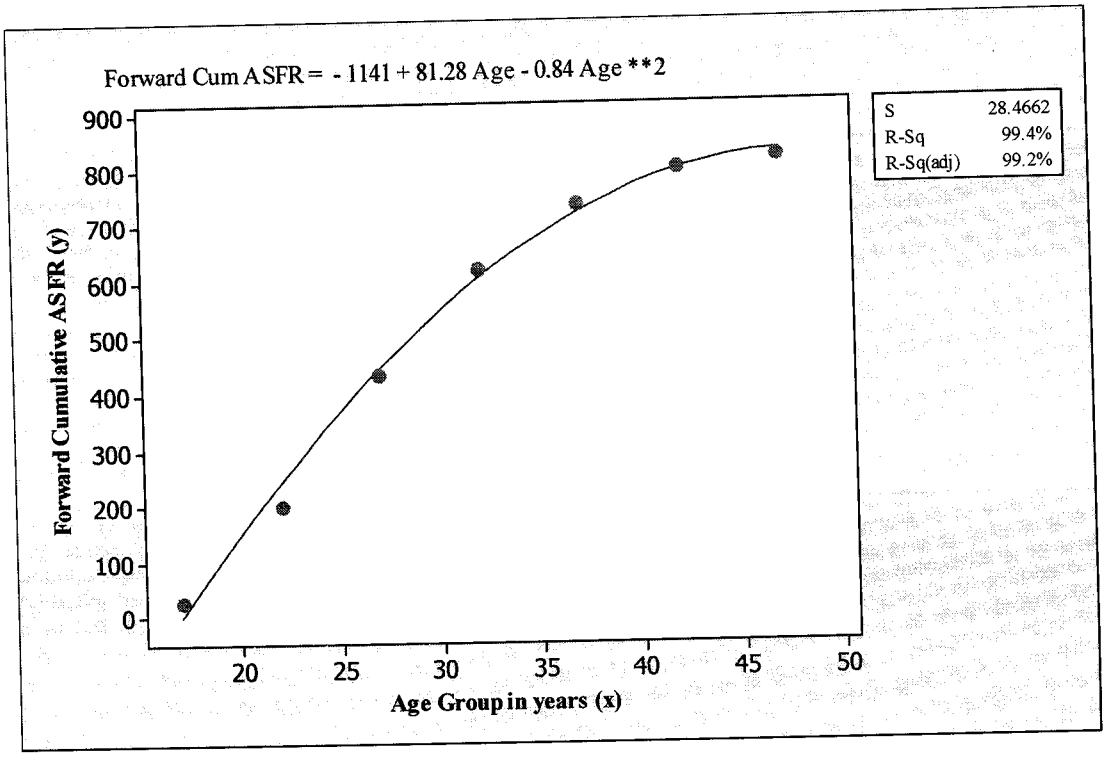


Figure 7.3 Modeling the Forward Cumulative ASFR of 2005 of Rural Areas

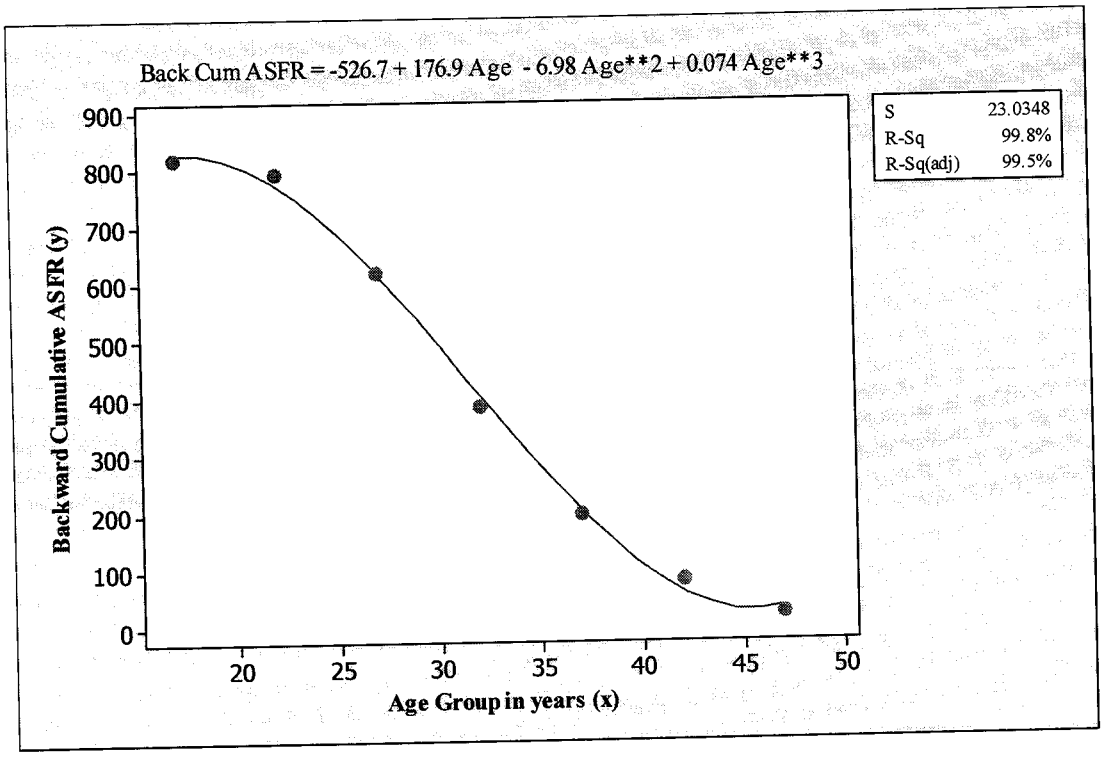


Figure 7.4 Modeling the Backward Cumulative ASFR of 2005 of Rural Areas

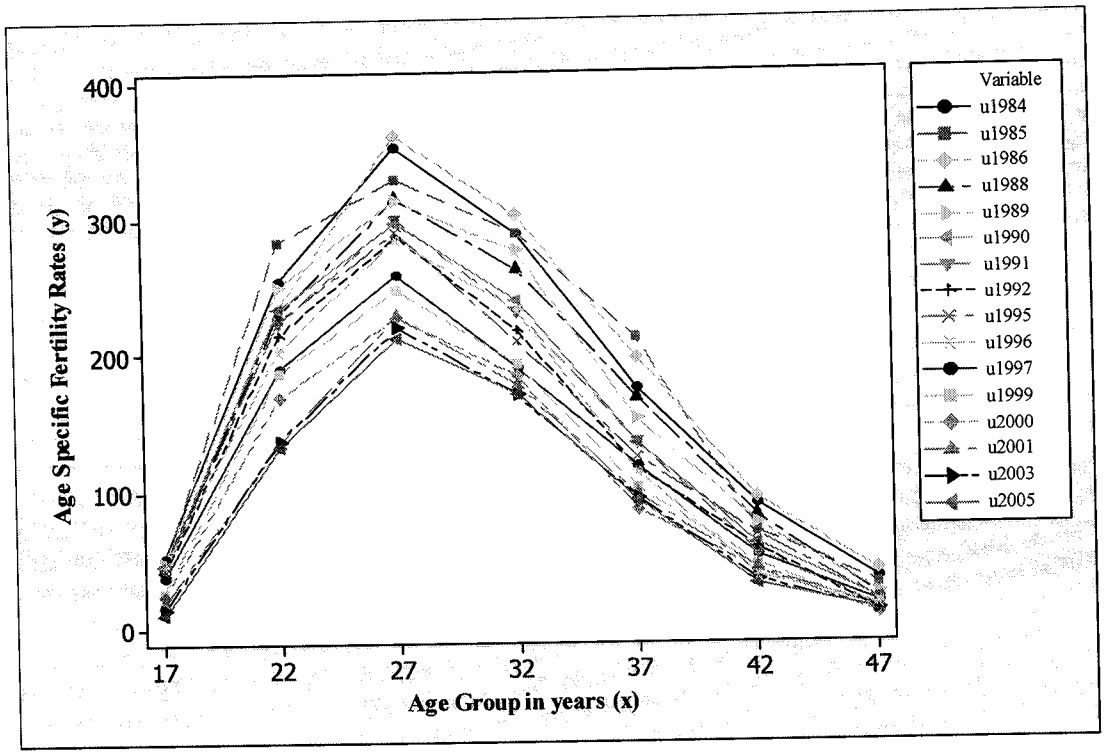


Figure 7.5 **Age Specific Fertility Rates Trend of Urban Areas 1984-2005**

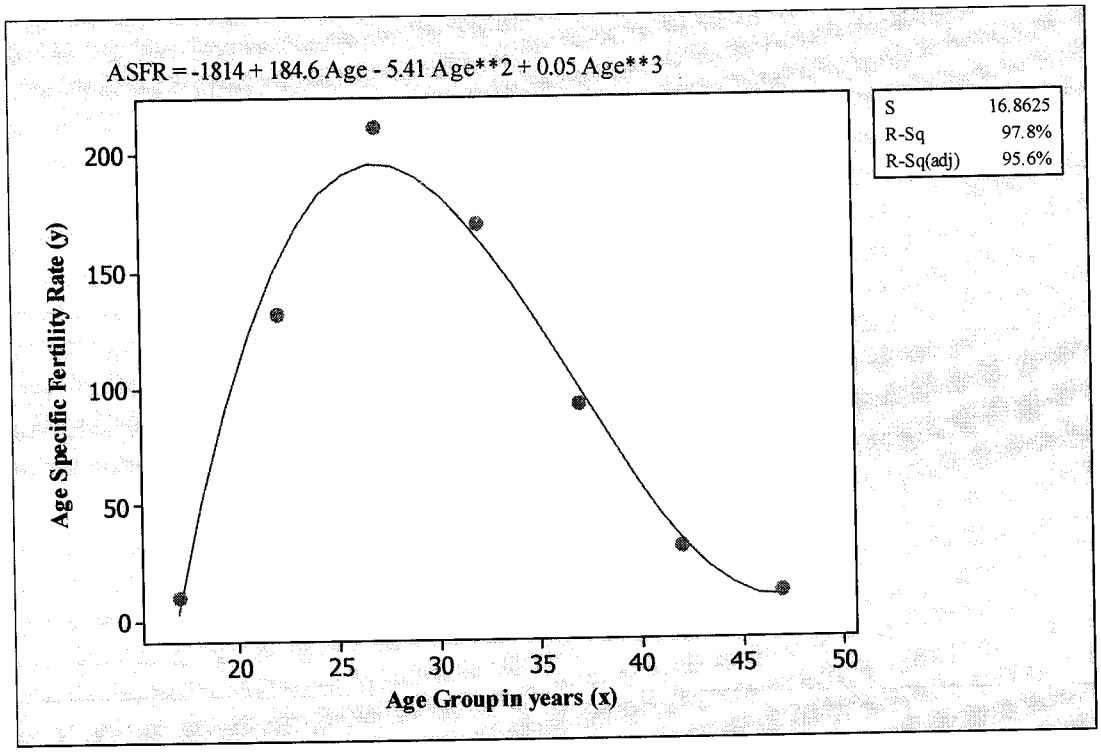


Figure 7.6 **Modeling the Age Specific Fertility Rates of 2005 of Urban Areas**

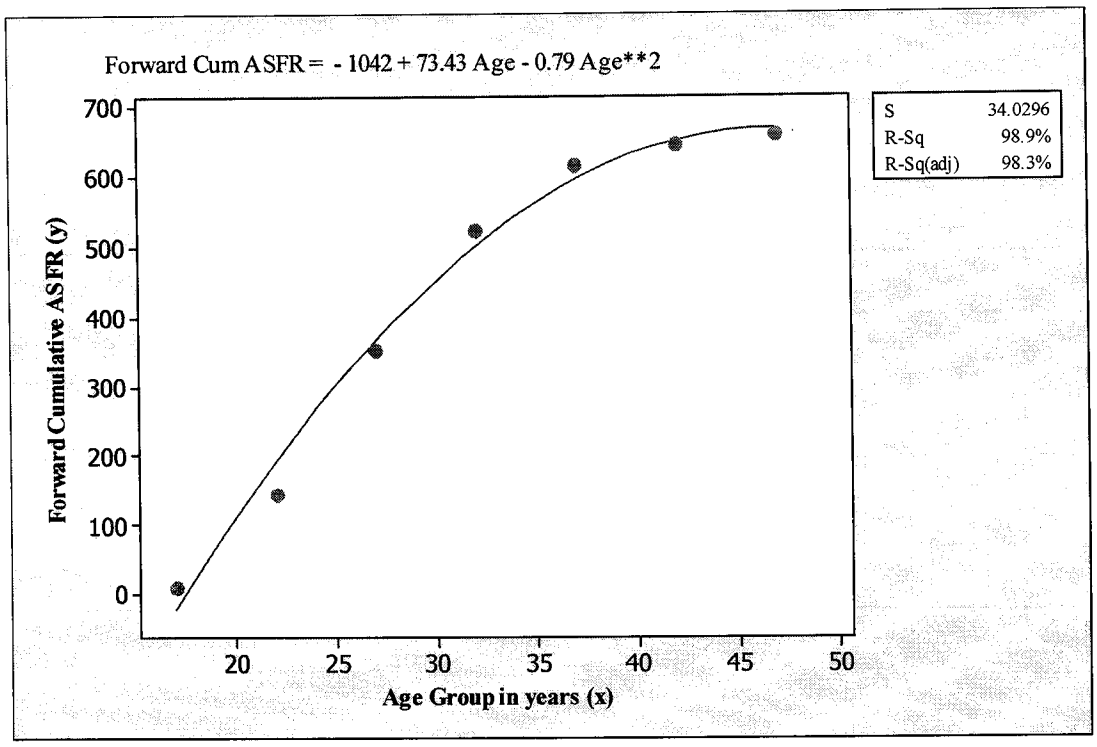


Figure 7.7 Modeling the Forward Cumulative ASFR of 2005 of Urban Areas

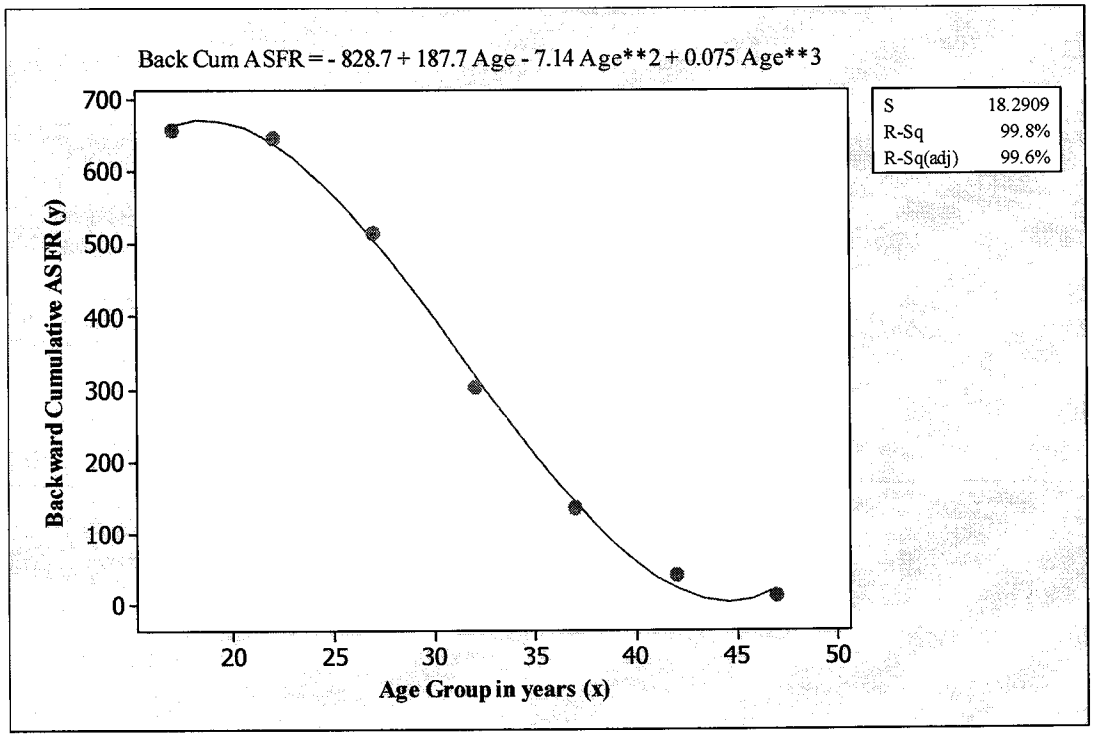


Figure 7.8 Modeling the Backward Cumulative ASFR of 2005 of Urban Areas

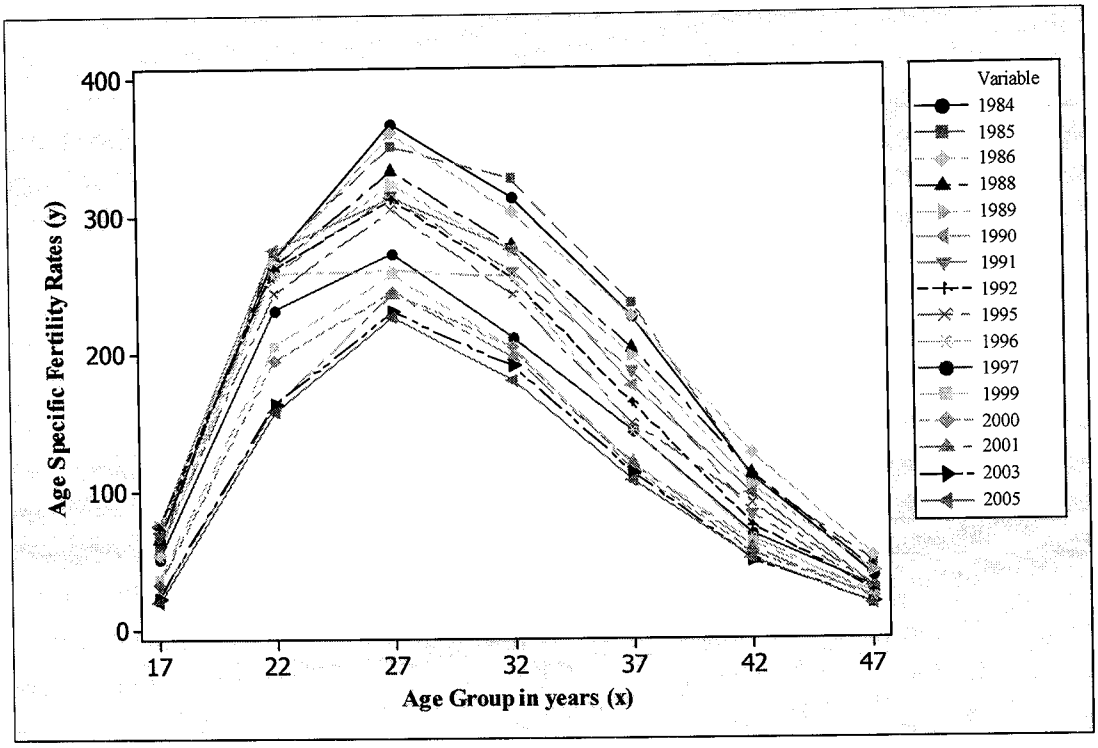


Figure 7.9 Age Specific Fertility Rates Trend of Pakistan 1984-2005

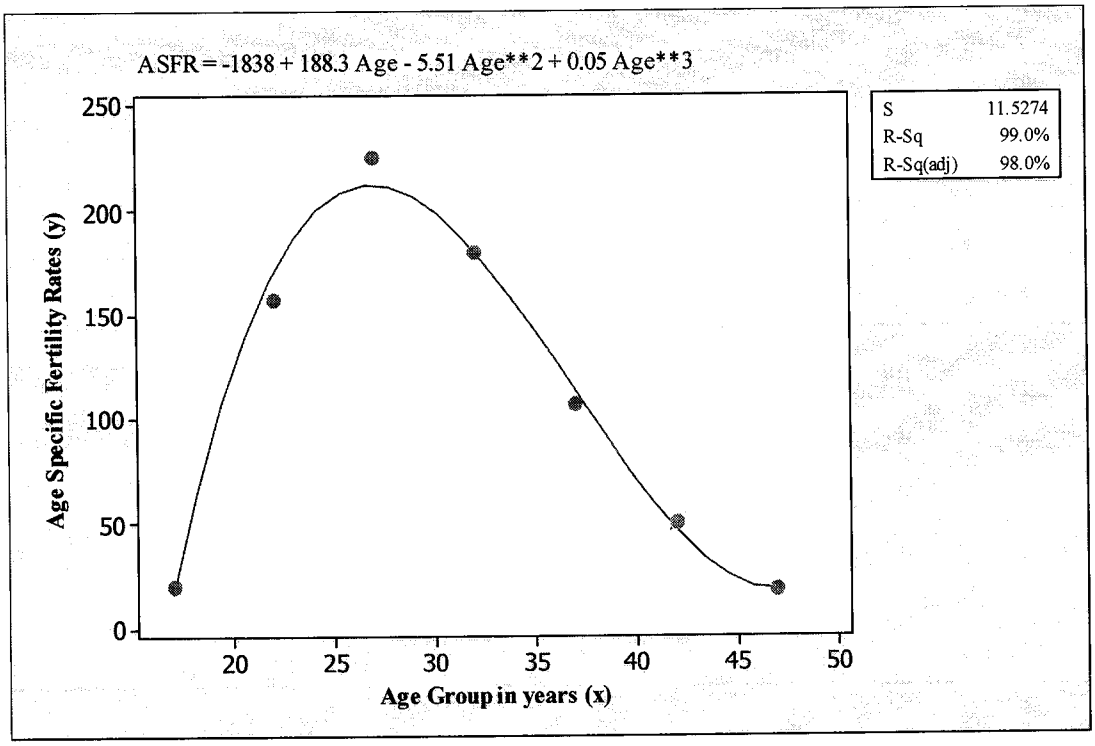


Figure 7.10 Modeling the Age Specific Fertility Rates of 2005 of Pakistan

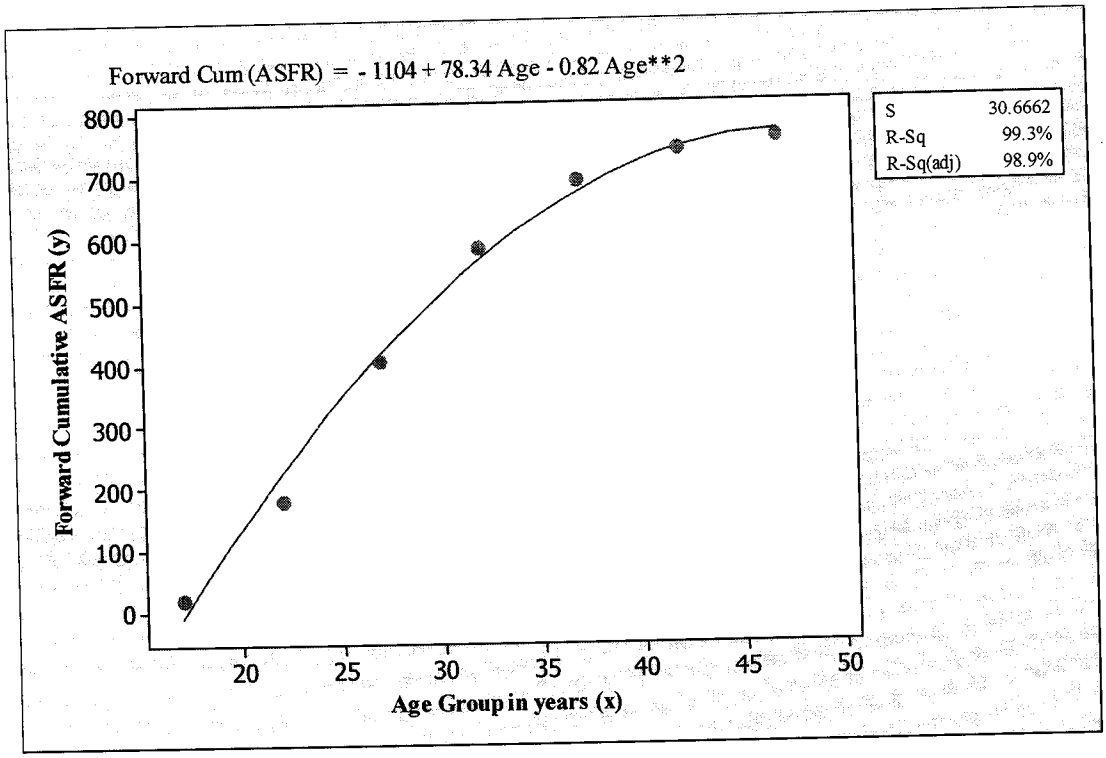


Figure 7.11 **Modeling the Forward Cumulative ASFR of 2005 of Pakistan**

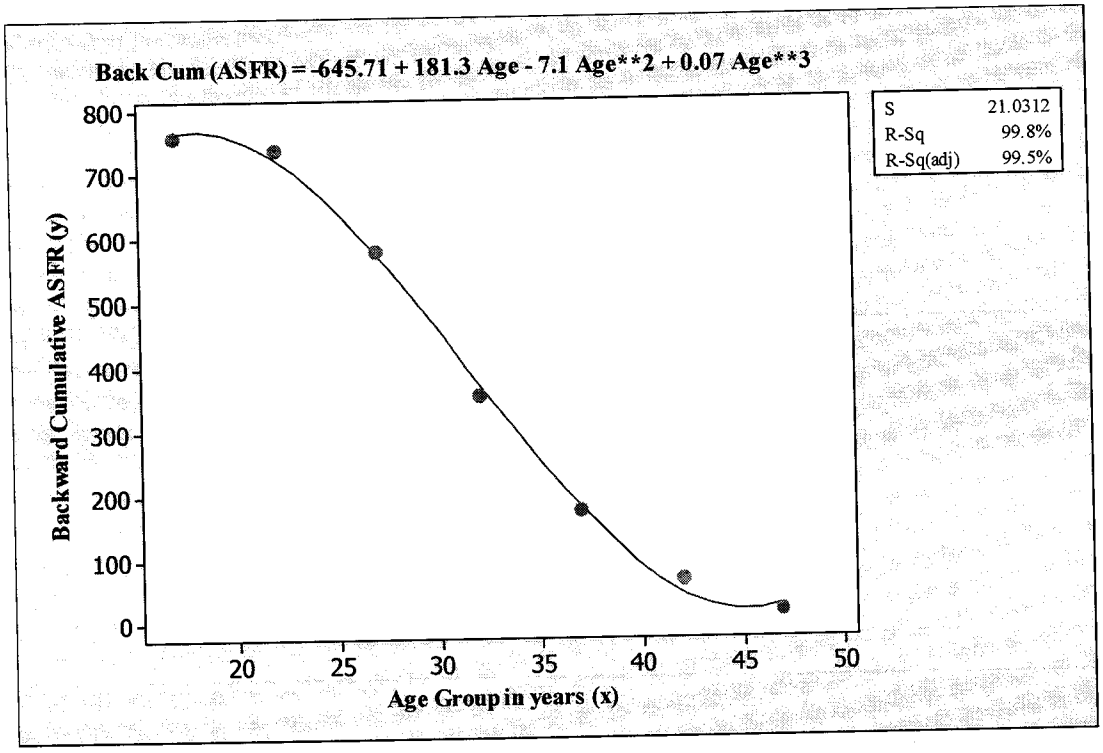


Figure 7.12 **Modeling the Backward Cumulative ASFR of 2005 of Pakistan**

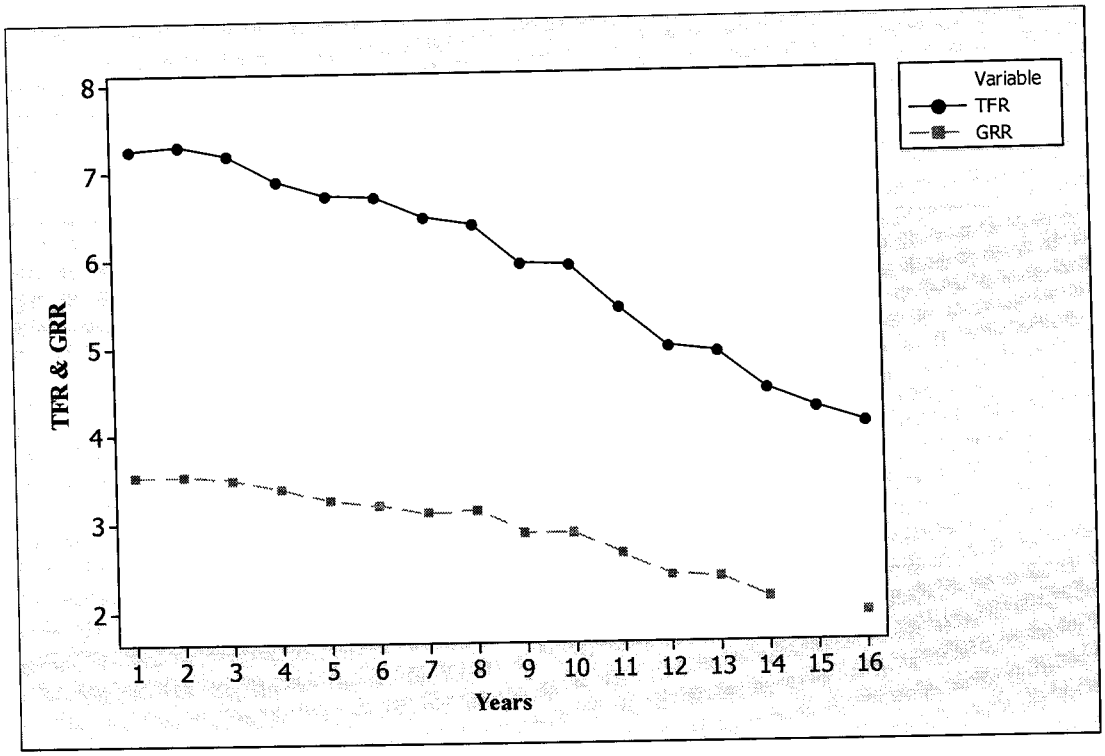


Figure 7.13 Trend of TFR and GRR Rate of Rural Areas 1984-2005

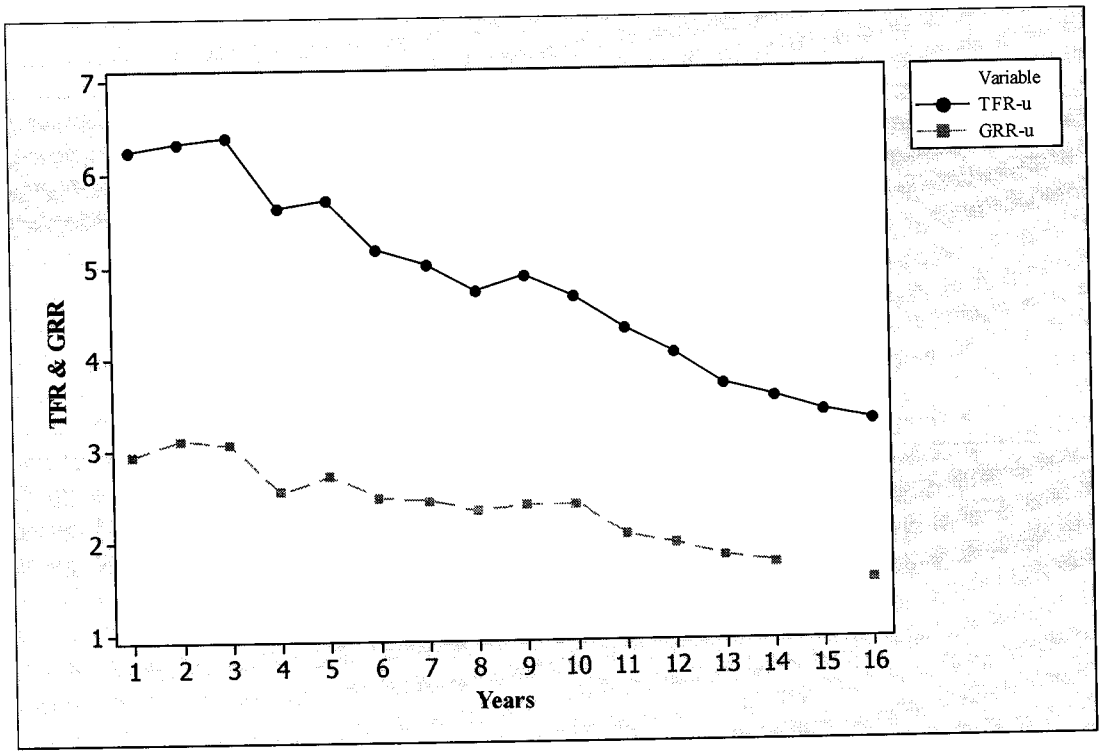


Figure 7.14 Trend of TFR and GRR of Urban Areas 1984-2005

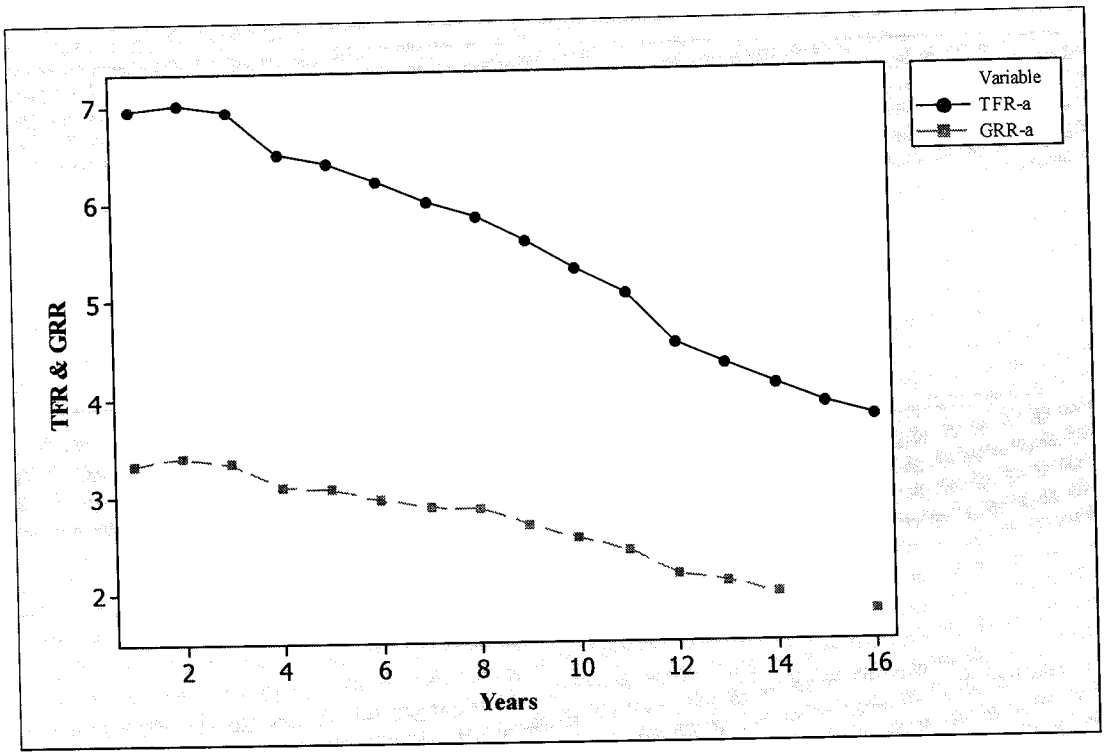


Figure 7.15 Trend of TFR and GRR of Pakistan 1984-2005

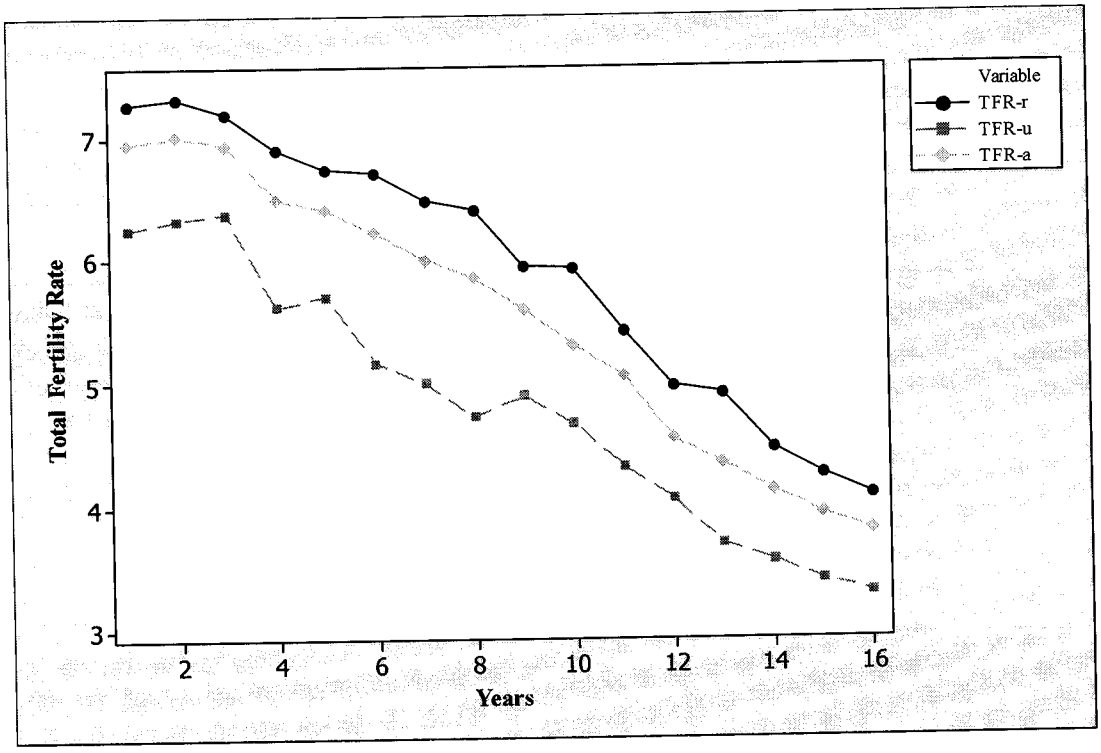


Figure 7.16 Trends of TFR of Rural, Urban and Pakistan 1984-2005

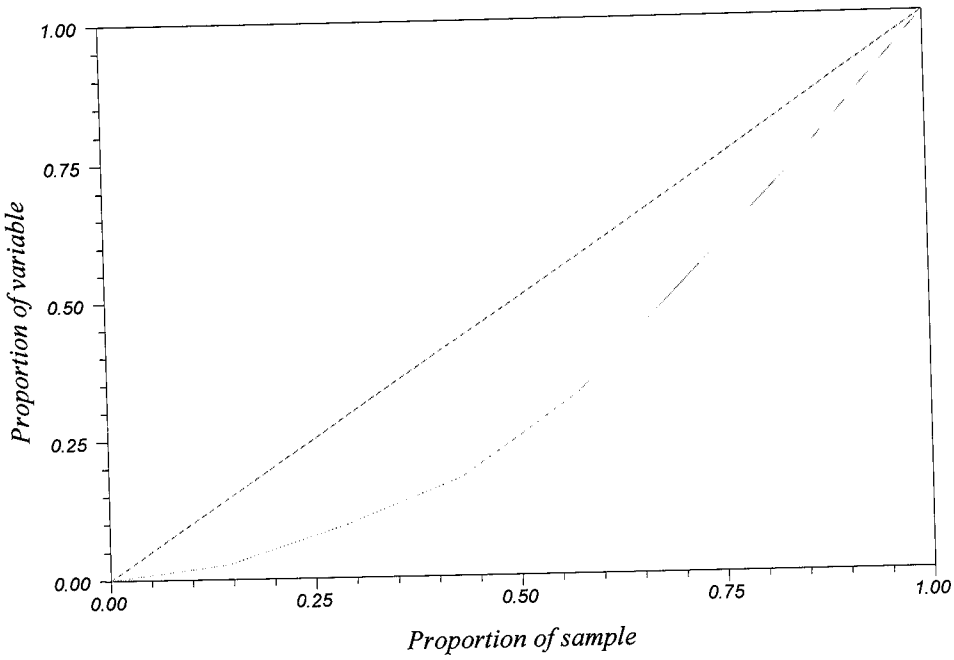


Figure 7.17 Lorenz Curve for Age Specific Fertility Rate 1990 of Rural Areas

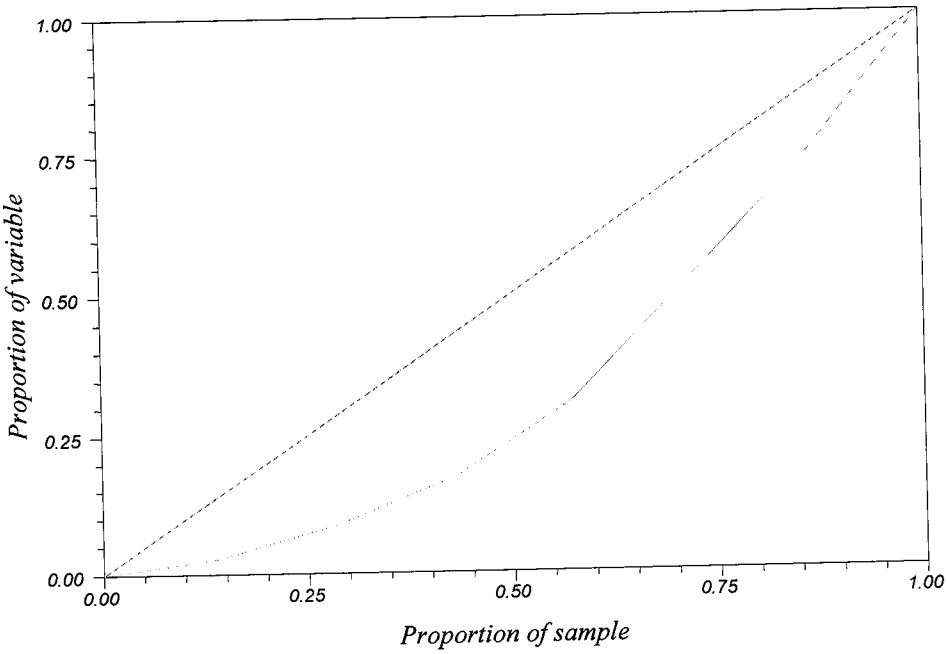


Figure 7.18 Lorenz curve for Age Specific Fertility Rate 1995 of Rural Areas

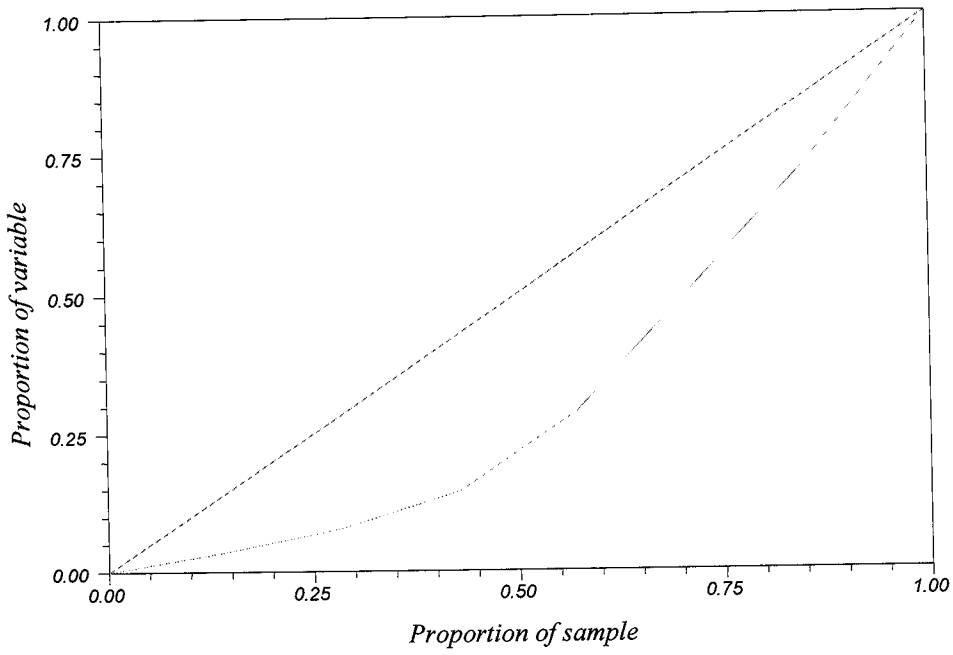


Figure 7.19 Lorenz Curve for Age Specific Fertility Rate 2000 of Rural Areas

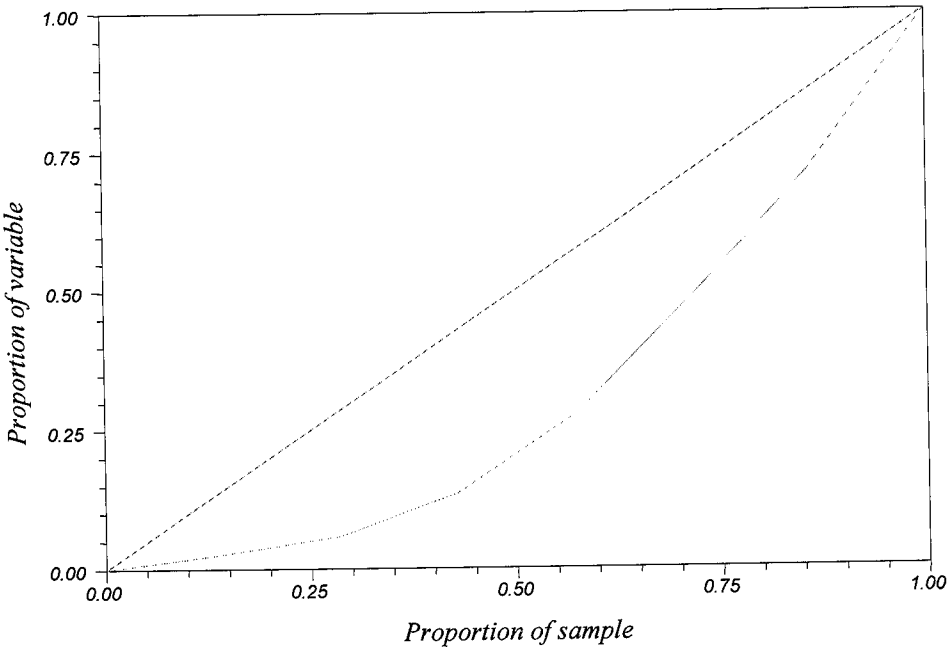


Figure 7.20 Lorenz Curve for Age Specific Fertility Rate 2005 of Rural Areas

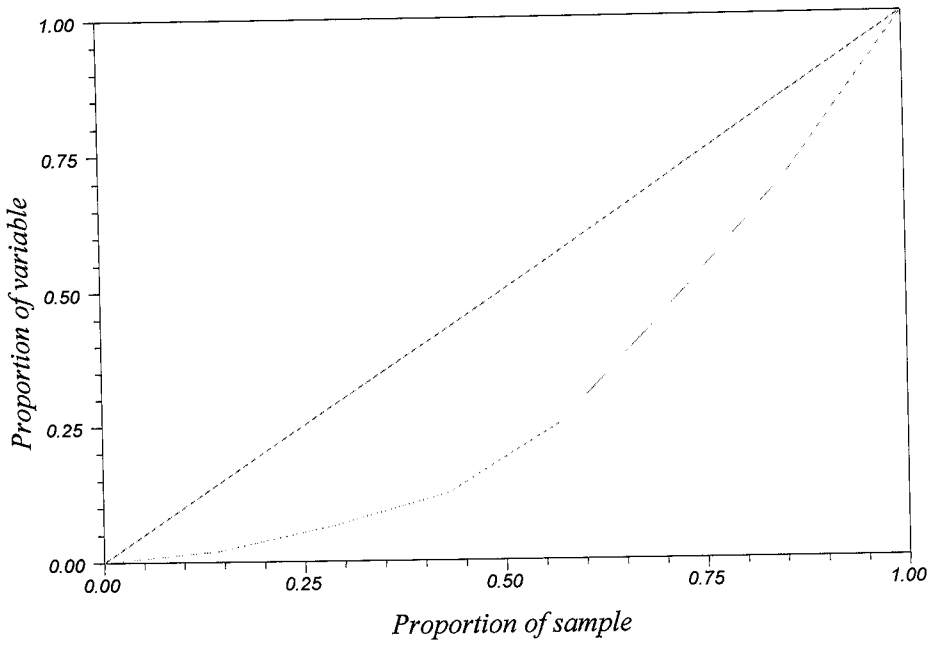


Figure 7.21 Lorenz Curve for Age Specific Fertility Rate 1990 of Urban Areas

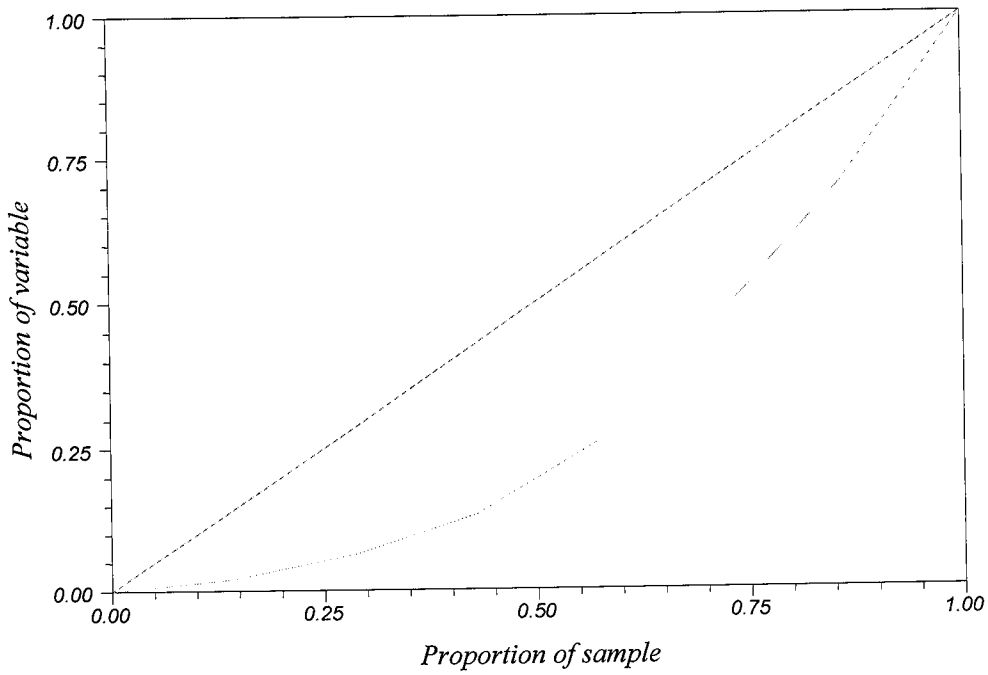


Figure 7.22 Lorenz Curve for Age Specific Fertility Rate 1995 of Urban Areas

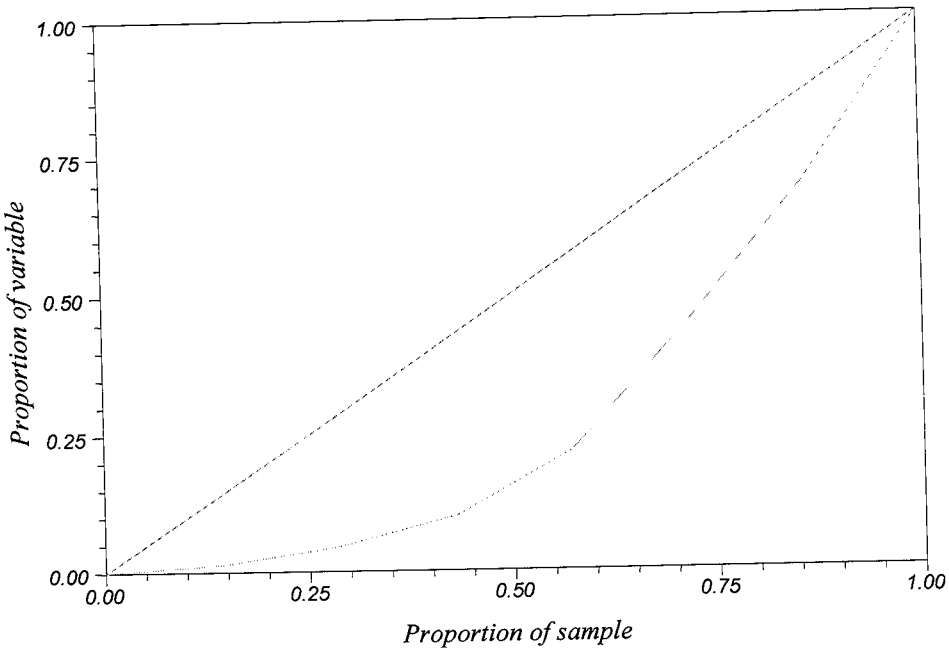


Figure 7.23 Lorenz Curve for Age Specific Fertility Rate 2000 of Urban Areas

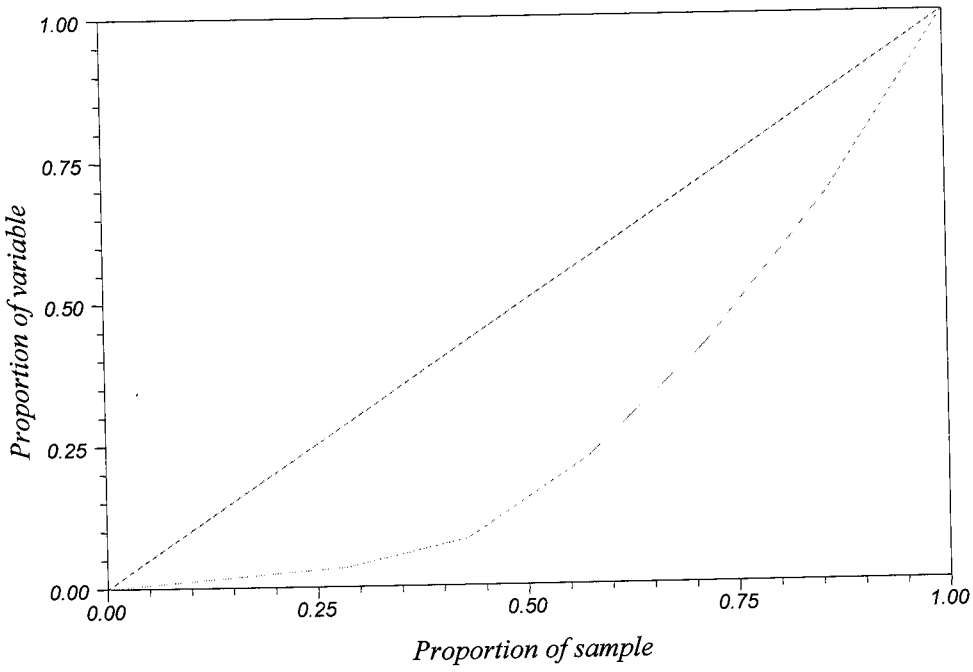


Figure 7.24 Lorenz Curve for Age Specific Fertility Rate 2005 of Urban Area

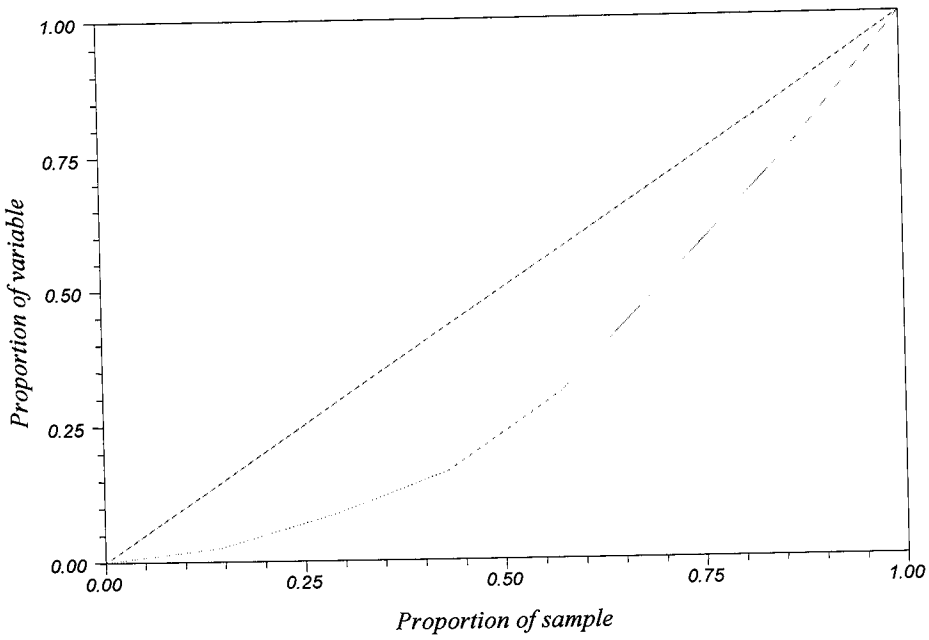


Figure 7.25 Lorenz curve for Age Specific Fertility Rate 1990 of Pakistan

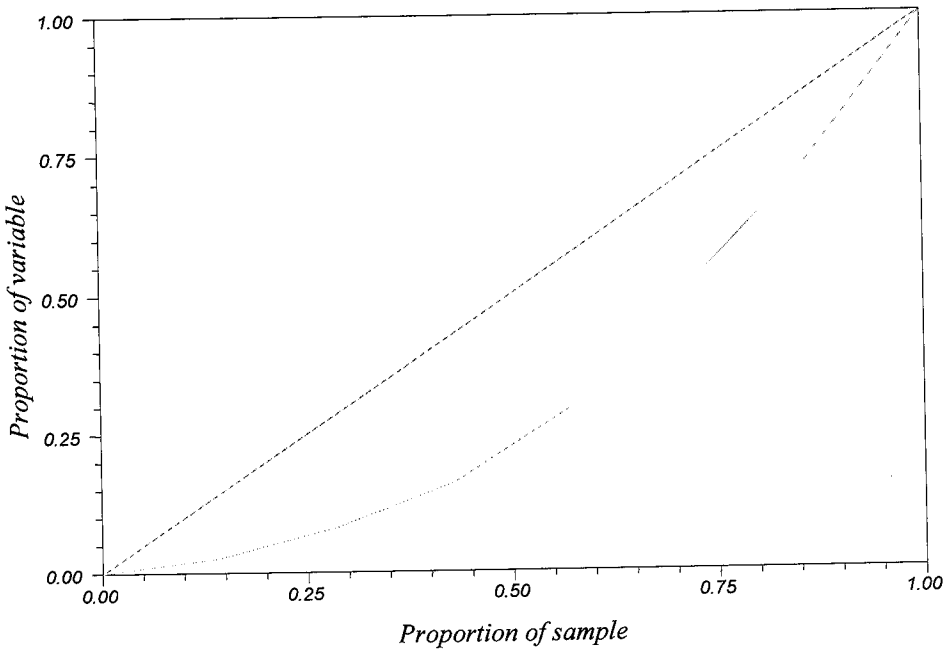


Figure 7.26 Lorenz curve for Age Specific Fertility Rate 1995 of Pakistan

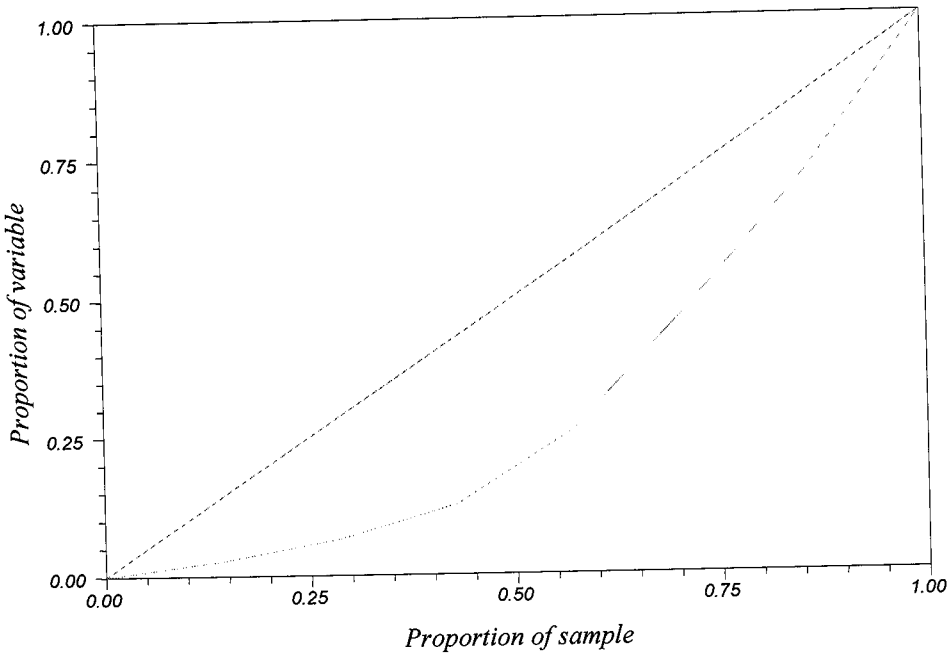


Figure 7.27 Lorenz curve for Age Specific Fertility Rate 2000 of Pakistan

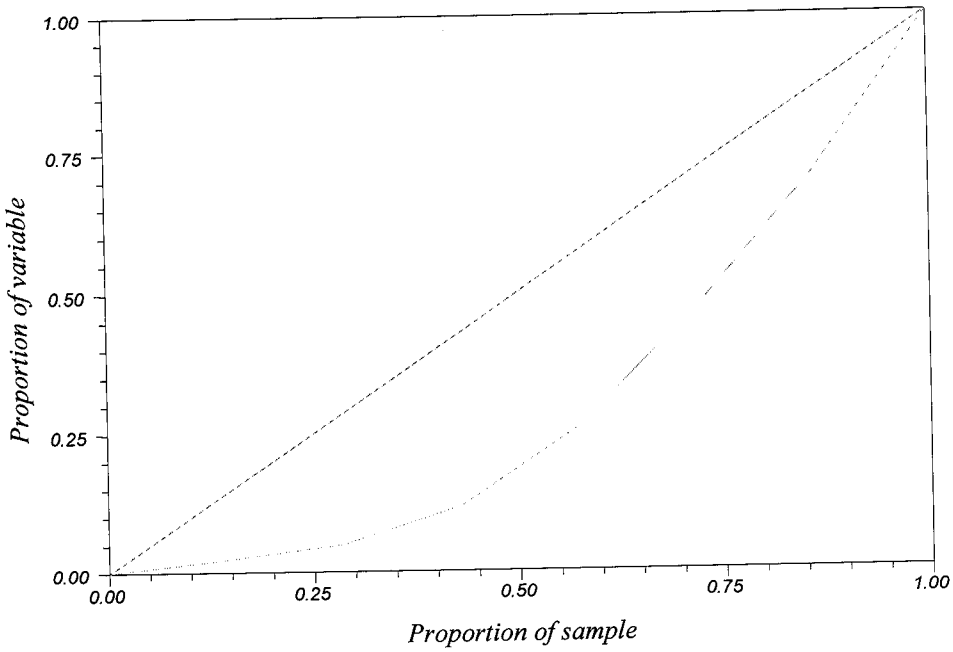


Figure 7.28 Lorenz curve for Age Specific Fertility Rate 2005 of Pakistan

SUMMARY

The Projection of age-sex distribution and population of Pakistan vision 2030, estimation of the inequality of the actual and projected age-sex distribution, reproductive cohort measures and fertility trends of the population during the last 20 years are the major objectives of the present study. The population censuses data 1951, 1961, 1972, 1981 and 1998 of Pakistan as well as some other data sets are used to achieve the objectives of the study.

The quality of the population census data has been checked in chapter 3 using different digit preference indexes. The Whipple indexes of 1972, 1981, 1998 censuses for both sexes are 347, 330 and 186 respectively which is too high than the acceptable range 105-125. The indexes are also computed for males and females separately. The results indicated that there is greater age misreporting among males during 1972 and 1981 than females which is against the usual expectations. On the other hand, the age misreporting among females is greater than male during 1998 census which is not against the usual prospects. The significant digit preference was seen among the male as well as female at ages ending 0 and 5. The results indicated that 1972 census data is highly inaccurate as compared to 1981 and 1998. The 1998 census data is also inaccurate but lesser extent than the preceding two censuses. Myers blended and other indexes are also computed for males and females separately and exactly the same conclusions are drawn. It is factual that digit preference can reduce, not eliminate. Since the quality of Pakistan census data

is not only poor but extremely poor. The strong smoothing technique is used to smooth the data as this technique is recommended for highly inaccurate data.

In the light of these results, it is suggested that the age based questions should be improved and increased for cross checking. Enumerators should be prepared with new and tactful trainings which will helpful in asking the questions from the respondents especially age based questions.

Chapter 4 consists of the population projection of Pakistan using autoregressive integrated moving average (ARIMA) for the next 20 years. After testing the stationary of the data, identification and estimation as well as using the criterion Mean Squared Error (MSE), Akaike Information Criteria (AIC), Schwarz Criterion (SC), P-values for goodness of fit of the model, the model ARIMA (1, 2, 0) W was found to be parsimonious. If the current growth rate continues, the population of Pakistan would be approximately 230.7 million in 2027 along with 95% confidence limits 193.33 million and 275.25. P-value of this model indicates the independency and randomness of the residuals.

According to the parsimonious model, there will be 74.29% increase in the Population till 2027 as compared to the 1998 population census whereas 45.74% increase as compared to the estimates (Iqbal, 2007). Population is also projected for different years i.e. 2010, 2015, 2020, 2025, and 2027 and given in Table 4.4. The projections by ARIMA (1, 2, 0) W is close to the projections by different bureaus i.e. Population Reference Bureau (2007), United States Census Bureau (2008), Pakistan reality (2008) and Population of Pakistan (2008). These bureaus reported 229 million population of Pakistan for the year

2025. The projections by ARIMA (1, 2, 0) W are equally important for the government of Pakistan as well as Non Government Organizations for future planning and projects.

Chapter 5 includes the projection of population for the years 2032 using the traditional growth models i.e. logistic, Gompertz, Exponential, and Modified exponential. The projections by said models are 364.16 million, 356.46 million 341.93 million, 277.98 million respectively. The Modified exponential model projected 277.98 million which is minimum among the other three projections but greater than the projection of ARIMA (1, 2, 0) W model. The projection by Modified exponential growth model for the years 2027 is 250.68 million whereas by ARIMA (1, 2, 0) W is 230.68 million respectively. Mean absolute percentage error (MAPE) is used as an evaluation statistics to identify the appropriate model. MAPE (4.28%) is maximum for Logistic model and minimum (0.49%) for ARIMA (1, 2, 0) W model respectively. The MAPE of Modified exponential model is 1.0578% which is greater than the time series models. The model ARIMA (1, 2, 0) W is again identified as a parsimonious model after comparing with that of traditional models. The projection by ARIMA (1, 2, 0) W is more close to the other national and international scientist's forecast (WPP, 2006).

The total population and age sex distribution is also projected in chapter 6 by using Modified Markov Chain model for 40 years since 1981. The projected population for the years 2011, 2021 are closer to the population projection (WPP 2008; People facts and figures & Total population by country 2009) and greater than (NIPS 2006; IDB 2008). The age disparity of the actual population censuses is measured by the Gini coefficients which is almost same for 1951 and 1961 but less than the coefficients of 1972, 1981,

1998. Gini coefficients of the projected population indicate the medium level of concentration during the next 20 years.

The projected population of different age groups indicated i.e. babies (0-4) and teenager population (5-14) will decrease whereas working and ageing population will increase. These projections are approximately close to the projection (WPP, 2008). Gini coefficients of ageing population indicate the high level of concentration. In the light of these findings, the decrease in the growth rate, increase in the life expectancy and the stable sex ratio of the population might be expected. Since ours are a religious society, the ageing population may not pose a significant problem. However the ageing population might be affected due to busy schedule of their spouse's life in the modern era. So, the government should start working on the socio-economic problems of the ageing population from all aspects; their needs and management well in time. Otherwise it would become difficult to handle their problems and to facilitate the elders.

In the fertility analysis, given in the chapter 7, the fertility pattern of two major geographical regions of Pakistan i.e. urban and rural during 1984-2005 showed greater fertility in the rural areas than the urban ones during the period 1984-2005. In 1984, the total fertility rates (TFR) of urban, rural and of Pakistan were 7.27, 6.24 and 6.95 respectively, whereas in 2005, those were 4.09, 3.29 and 3.79 respectively. Approximately 43.7%, 47.3% and 45.5% decrease in TFR has been seen in rural areas, urban areas and in Pakistan during 1984-2005. Gini coefficients of the age specific fertility rates for the years 1990, 1995, 2000 and 2005 for three regions rural, urban and Pakistan showed that greater variation exists in 2005 as compared to the preceding years. It indicated that the fertility started from premature ages and continued till their

childbearing age from mid 1980s to 2000. The Gini coefficient for the year 2005 indicates that age inequality is increased especially in urban areas. The reason might be the increased female literacy rate and mean age of child bearing. The third degree polynomial model was found to be appropriate when different polynomial models were fitted on the age specific fertility rates, backward cumulative age specific fertility rates, forward cumulative age specific fertility rates of Pakistan and its rural urban regions.

Finally, the findings may be helpful for the future planning and projects of the government as well as Non Government Organizations of the country. The decreasing trend in population might be due to the increased literacy rate especially the female education. The education of IT, science and technology may be provided on priority to everyone at government educational institutes at subsidized rates in order to utilize the maximum potentials of our youth. The private and the government educational institutions of the less developed cities, villages as well as the remote and far flung areas of our country must fall within the most targeted zones. It is the dire need of the modern era.

Suggestions for further research

New researchers may extend this work in the following aspects:

- Projections of the age sex distribution of population using Modified Markov chain method by making a transition probability matrix (tpm) of higher order from 5 years age groups data.
- Projections may be made by using age specific fertility rates instead of average growth rate in transition probability matrix (tpm).

- Projection of population at micro levels using Time series models at province as well as sub levels which will be more helpful for better doorstep management and planning.
- Fertility analysis pattern of the respondents may be studied along with its education and employment.

REFERENCES

Agrawal, U. D. (2000). *Population projections and their accuracy*. Delhi: B.R. Publication Corporation.

Angeles, G., Guilkey, D. K., & Mroz, T. A. (1998). Purposive program placement and the estimation of family planning program effects in Tanzania. *Journal of the American Statistical Association*, *93*(443), 884-899.

Bairagi, R., & Datta, A. K. (2001). Demographic transition in Bangladesh: What happened in the twentieth century and what will happen next? *Asia Pacific Population Journal, 16*(4), 3-16.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Brown, M. (1994). Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on the Alberta data. *Social Science and Medicine, 38*(9): 1243-1256.

Burnham, K. P., & Anderson D. R. (2002). *Model selection and multi model inference: A practical information theoretic approach.* New York: Springer Science and Business Media.

Carter, L. R. (1996). Forecasting U.S. mortality: A comparison of Box-Jenkins ARIMA and structural time series models. *The Sociological Quarterly*, *37*(1), 127-144.

Census. (2009). *Definition of census.* Retrieved March 6, 2009, from http://encyclopedia2.thefreedictionary.com/Population+census

Chatfield, C. (1996). Simple descriptive techniques: In C. Chatfield & J. V. Zidek (Eds.). *The analysis of time series: An introduction.* 9-26. New York: Chapman and Hall.

Chen, M. H., Dey, D. K., & Sinha, D. (2000). Bayesian analysis of multivariate mortality data with large families. *Journal of Applied Statistics, 49*(1), 129-144.

Cohen, J. E. (1986). Population forecasts and confidence intervals for Sweden: A comparison of Model-based and empirical approaches. *Demography, 23*(1), 105-126.

Cowell, F. A. (1995). *Measuring Inequality.* (2nd ed., draft 3rd ed. May 2000). http://darp.lse.ac.uk/Frankweb/Frank/pdf/measuringinequality2.pdf) Hemel Hempstead: Harvester Wheatsheaf.

Elliott, M. R., & Little, R. J. A. (2005). A Bayesian approach to 2000 Census evaluation using ACE survey data and demographic analysis. *Journal of the American Statistical Association, 100* (470), 380-387.

Eltigani, E. E. (2009). Towards replacement fertilize in Egypt and Tunisia. *Studies in Family Planning, 40*(3), 215-226.

Federal Bureau of Statistics Division. (1985). *Population Demographic Survey-1985.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1986). *Population Demographic Survey-1986.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1988). *Population Demographic Survey-1988.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1989). *Population Demographic Survey-1989.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1990). *Population Demographic Survey-1990.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1991). *Population Demographic Survey-1991.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1992). *Population Demographic Survey-1992.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1995). *Population Demographic Survey-1995.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1996). *Population Demographic Survey-1996.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1997). *Population Demographic Survey-1997.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (1999). *Population Demographic Survey-1999.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (2000). *Population Demographic Survey-2000.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (2001). *Population Demographic Survey-2001.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (2003). *Population Demographic Survey-2003.* Islamabad, Government of Pakistan.

Federal Bureau of Statistics Division. (2005). *Population Demographic Survey-2005.* Islamabad, Government of Pakistan.

Feeney, G., & Alam, I. (2003). New estimates and projections of population growth in Pakistan. *Population and Development Review, 29*(3), 483-492.

Feldman, B. S., Zaslavsky, A. M., Ezzati, M., Peterson, K. E., & Mitchell, M. (2009). Contraceptive use, Birth spacing, and Autonomy: An analysis of the Oportunidades program in rural Mexico. *Studies in Family Planning, 40*(1), 51-62.

Fujiwara, M., & Caswell, H. (2002). Estimating population projection matrices from multi stage mark- recapture data. *Ecology, 83*(12): 3257-3265.

Gini coefficient. (2009). *Gini coefficient.* Retrieved January 6, 2009, from http://en.wikipedia.org/wiki/Gini_coefficient

Goesling, B., & Firebaugh, G. (2004). The trend in international health inequality. *Population Development Review, 30*(1), 131-146.

Heilig, G. K. (2006). Many Chinas? The economic diversity of China's provinces *Population and Development Review, 32*(1), 147-161.

Hinde, A., & Mturi, A. J. (2000). Recent trends in Tanzanian fertility. *Population Studies,* 54(2), 177-191.

Hussain, R., & Bittles, A. H. (1999). Consanguineous marriage and differentials in age at marriage, contraceptive use and fertility in Pakistan. *Journal of biosocial science, 31*(2), 121-138.

International Data Base. (2008). International Data Base (IDB): Information gateway, U.S. Census Bureau. Retrieved May 5, 2009, from http://www.census.gov/idb/ranks.html

Iqbal, Z. (2007, December 29). Pakistan: The housing market thrives as more Pakistan capital returns homes. *Dawn*. P.1. Retrieved February 22, 2008, from AIOU Newspapers database.

Ishida, K., Stupp, P., & Melian, M. (2009). Fertility Decline in Paraguay. *Studies in family planning, 40*(3), 227-234.

Islam, M. R., & Ali, M. K. (2004). Mathematical modeling of age specific fertility rates and study the reproductivity in the rural area of Bangladesh during 1980-1998. *Pakistan Journal of Statistics, 20*(3), 379-392.

Jan, B., Ishfaq A., & Shuhrat S. (2007). Selecting of mathematical model for projections of NWFP population. *The Journal of Humanities and Social Sciences, XV* (2), 69-78.

Kabir, M., & Mosleh Uddin M. (1987). Fertility transition of Bangladesh: Trends and determinants. *Asia Pacific Population Journal, 2*(4), 54-74.

Keilman, N. (1998). How accurate are the United Nations world population projections. *Population and Development Review, 24*(1), 15-41.

Keilman, N. (2001). Data quality and accuracy of United Nations population projections, 1950-95. *Population studies, 55*(2), 149-164.

Kemal, A. R., Irfan, M., & Mahmood, N. (Eds.). (2003). *Population of Pakistan: An analysis of 1998 population and housing census*. Islamabad: Pakistan Institute of Development Economics.

Keyfitz, N. (1964). The population projection as a matrix operator. *Demography, 1*(1), 56-73.

Keyfitz, N (1968). *Introduction to the mathematics of population*. Addison Wesley: Reading, Massachusetts.

Klosterman, R. E. (1990). *Community analysis and planning techniques.* Lanham, MD: Rowman & Lettlefield.

Kostaki, A., & Paraskevi, P. (2007). Modeling fertility in modern populations. *Demography, 16*(6), 141-194.

Krueckeberg, D. A., & Silvers, A. L. (1974). *Urban planning analysis: Methods and* models. John Wiley & Sons.

Lam, D., & Marteleto, L. (2008). Stages of the demographic transition from a child's perspective: Family size, cohort size, and children's resources. *Population Demographic Review, 34*(2), 225-252.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association, 87*(419), 659-671.

Lee, Y., & Nelder, J. A. (2000). The relationship between double-exponential families and extended quasi-liklihood families, with application to modelling Geissler's human sex ratio data. *Journal of Applied Statistics, 49*(3), 413-419.

Loh, S., & George, M. V. (2007). Projected population size and age structure for Canada and provinces: With and without international migration. *Canadian Studies in Population, 34*(2), 103-127.

Markov Chains. 2008. *Markov Chain.* Retrieved January 5, 2008, from http://www.sosmath.com/matrix/markov/markov.html

McDonald, J. (1979). A time series approach to forecasting Australian total live-births. *Demography, 16*(4), 575-601.

McNay, K., Arokiasamy, P., & Cassen, R. (2003). Why are educated women in India using contraception? A multilevel analysis. *Population Studies, 57*(1), 21-40.

People Facts & Figures. (2009). *People facts and figures*. Retrieved April 10, 2009, from http://www.os-connect.com/pop/p2a.asp?whichpage=1&pagesize=20&sort=2050

Population. (2009). *Definition of population.* Retrieved September 10, 2009, from http://dictionary.reference.com/browse/population/

Population Census Organization. (1967). *Projections of population of Pakistan 1961 to 1981. (Census Bulletin No.7),* Ministry of home & Kashmir affairs, Home Affairs Division, Islamabad. Government of Pakistan.

Population Census Organization. (1972). *Statistical report of Pakistan: Population census of Pakistan 1972*. Statistics Division, Islamabad. Government of Pakistan.

Population Census Organization. (1984). *1981 Census report of Pakistan*, Statistics Division. Islamabad. Government of Pakistan.

Population Census Organization. (2001). *1998 Census report of Pakistan*, Statistics Division. Islamabad. Government of Pakistan.

Population Growth Rate. (2008). *Population growth rate*. Retrieved December 12, 2008, from http://www.indexmundi.com/pakistan/population_growth_rate.html/

Population of Pakistan. (2008). *Population of Pakistan*. Retrieved January 4, 2009, from http://www.indexmundi.com/g/r.aspx?t=0&v=21&l=en

Populous Pakistan. (2009). *Populous Pakistan-The Boston Globe*. Retrieved May 3, 2009, from http://www.boston.com/news/world/asia/articles/2007/04/01/populous_pakistan/

Population Policy of Pakistan. (2002). *Population Policy of Pakistan.* Retrieved September 5, 2008, from *http://www.mopw.gov.pk/event3.html*

Population Reference Bureau. (2007). *Population reference bureau.* Retrieved September 10, 2008, from http://www.prb.org/Datafinder/Geography/Summary.aspx?region=145&regiontype=2

Population Reference Bureau. (2008). *Population reference bureau.* Retrieved January 4, 2009, from http://www.prb.org/Datafinder/Geography/Summary.aspx?region=148&region_type=2

Population Resources Centre providing the demographic dimensions of public policy. (2007). *Instability in Pakistan: Demographic Factors.* Retrieved September 8, 2008, from http://www.prcdc.org/events/15

Pullum, T. W. (2006). Statistical methods to adjust for date and age misreporting to improve estimates of vital rates in Pakistan. *Statistics in Medicine, 10*(2), 191-200.

Raab, G. M., & Donnelly, C. A. (1999). Information on sexual behaviour when some data are missing. *Journal of the Royal Statistical Society, 48*(1), 117-133.

Rauf Textile & Printing Mills (Pvt) Ltd. (2008). *Pakistan facts and figures.* Retrieved September 13, 2008, from http://rauf.com/Facts_figures.html

Relationship between Literacy, Education & Demographic. (2009). *Basic Education.* Retrieved April 3, 2009, from http://www2.unescobkk.org/elib/publications/TrainingManual/

Ruggles, S., & Heggeness, M. (2008). Intergenerational coresidence in developing countries. *Population Development Review, 34*(2), 253-281.

Ryan, A. K., & Willits, F. K. (2007). Family ties, physical health, and psychological well being. *Journal of aging and health, 19*(6), 907-920.

Sathar, Z. A. (1993). The much-awaited fertility decline in Pakistan: wishful thinking or reality. *International Family Planning Perspectives, 19*(4), 142-146.

Sathar, Z. A., & Casterline, J. B. (1998). The Onset of fertility transition in Pakistan. *Population Development Review, 24*(4), 773-796.

Sathar, Z. A., Crook, N., Callum, C., & Kazi, S. (1988). Women's status and fertility change in Pakistan. *Population Development Review, 14*(3), 415-432.

Sathar, Z. A., & Kazi, S. (1990). Women work and reproduction in Karachi. *International Family Planning Perspectives, 16*(2), 66-69.

Shryock, H. S., Seigel, J. S., & Associates (1973).*The methods and materials of demography*. U. S. Bureau of the Census. Washington, DC: Government Printing Office.

Slack, B., & Rodrigue, J. P. (2008). *Gini coefficient.* Retrieved September 13, 2008, from http://people.hofstra.edu/geotrans/eng/ch4en/meth4en/ch4m1en.html

Smith, S. K., & Rayer, S. (2008). *An evaluation of sub county population forecasts in Florida*. Bureau of Economic and Business Research, University of Florida. USA.

Smith, S. K., & Shahidullah, M. (1995). An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association, 90*(429), 64-71.

Smith, S. K., & Sincich, T. (1988). Stability over time in the distribution of population forecast errors. *Demography, 25*(3), 461-473.

Smith, S. K., & Tayman, J. (2003). An evaluation of population projections by age. *Demography, 40*(4), 741-757.

SPSS 16 (2007). Chicago, IL: SPSS. Inc.

Srinivasan, K. (1998). *Basic demographic techniques and applications*. Delhi: Sage Publications.

World Population Prospects. (2006). *World Population Prospects: The 2006 Revision. .DESA. Department of economic and social affairs of the United Nations.* Retrieved September 20, 2008, from http://esa.un.org/unpp

World Population Prospects. (2008). *World Population Prospects: The 2008 Revision. population database.* Retrieved March 14, 2009, from *http://esa.un.org/unpp/index.asp?panel=3*