# MODELING AGEING POPULATION IN SRI LANKA TOWARDS BETTER CARE FOR ELDERLY

Amalie Chathurika Jayatilake

(148857H)

Degree of Master of Science

Department of Mathematics

University of Moratuwa
Sri Lanka

July 2017

# MODELING AGEING POPULATION IN SRI LANKA TOWARDS BETTER CARE FOR ELDERLY

Amalie Chathurika Jayatilake

(148857H)

Dissertation submitted in partial fulfillment of the requirements for the degree
Master of Science in Business Statistics

Department of Mathematics

University of Moratuwa
Sri Lanka

July 2017

**DECLARATION**

"I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non – exclusive right to reproduce and distribute my thesis/ dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the supervisor:

Signature of the supervisor: Date:

**ACKNOWLEDGEMENT**

First and foremost I would like to convey my utmost gratitude and thankfulness to my supervisor, Prof. T. S. G. Peiris, Professor in Applied Statistics who devoted his precious time in guiding, encouraging and assisting me so patiently by pointing out my mistakes to get the maximum out of me for a better development of this dissertation. It is not at all an exaggeration that without his great guidance this research would not become a reality.

I should specially thank Dr. A. J. Satharasinghe, Director General, Department of Census and Statistics and the Senior Statistician of the Registrar General's Department and the staff of the both departments who supported me by providing the necessary secondary data to carry out this research.

Indeed I would like to pay my sincere gratitude to all the lecturers who provided me relevant knowledge of Business Statistics through different courses during the time period which I have carried out my course work.

Also I would like to acknowledge the Sabaragamuwa University of Sri Lanka, for the financial support given to me to carry out this MSc.

Finally I would like to thank all the members of my family and all my friends for their motivations, support and encouragement rendered towards me to complete this dissertation.

**ABSTRACT**

Population ageing is an universal phenomenon and it is expected to be among the most prominent global demographic trends of the 21st century. In Sri Lanka there was a rising trend of ageing population throughout the past years and has recorded the highest number of agers within South Asia. However, no sound statistical or mathematical models were developed to project ageing population in Sri Lanka. Using the population aged 60 years and above in Sri Lanka during 1950-2016, three types of statistical models: (i) ARIMA (0, 2, 1), (ii) growth model, and (iii) double exponential smoothing model were developed. The models were compared using various statistical indicators and some statistical diagnostics tests. The comparison was done for both training set as well as validation set. Among these models the double exponential smoothing model was found as the best fitted model. According to the forecast derived from the best fitted model, it was found that the increasing trend of ageing population in the country will continue in the future and there will be approximately 2,936,000 ageing population in Sri Lanka in 2020. The information obtained in this study is beneficial for planners and decision makers in the government sector and other relevant organizations to cater to the needs of the increasing agers in the future of Sri Lanka.



**Keywords**: Ageing Population, ARIMA, Demographic, Exponential Smoothing

**TABLE OF CONTENTS**

**LIST OF FIGURES**

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| Abbreviation | Description |
|---|---|
| ACF | Autocorrelation Function |
| AIC | Akaike's Information Criteria |
| ARIMA(p,d,q) | Autoregressive Integrated Moving Averages of order p, d, and q, |
| ARMA(p,q) | Autoregressive of order p and Moving Averages of order q |
| AR | Autoregressive components |
| MA | Moving averages components |
| BIC | Schwarz Information Criterion |
| CBR | Crude Birth Rate |
| CDR | Crude Death Rate |
| DDT | *dichloro diphenyl trichloroethane* |
| DES | Double Exponential Smoothing |
| $\rho_k$ | Autocorrelation function of lag k |
| $\varphi_{kk}$ | Partial autocorrelation function for lag k |
| MAPE | Mean Absolute Percentage Error |
| SACF | Sample Autocorrelation Function |
| TFR | Total Fertility Rate |

## LIST OF APPENDIXES

# CHAPTER ONE
# INTRODUCTION

In this chapter background to the study, objectives of the study and significance of the study are described in details. Furthermore the chapter organization of this dissertation is given.

## 1.1. Definition of Ageing Population

Number of people in a country which is commonly known as “population” is a vital concept in the area of demography and related disciplines. Identifying the consistency and estimating the exact figure of future population is an important task not only in the area of demography but also in areas like social, political, educational, biological genetic geographical etc (Booth, 2006).

Population of a country is consisted of some clearly identifiable categories such as mid-year population, child population, labour force, ageing population etc. Among these categories, ageing population has become a universal phenomenon which gives more attention throughout the world. According to Gavrilov and Heuveline (2003) ageing of population (also known as, “demographic ageing”) is a summary term for shifts in the age distribution (i.e., age structure) of a population toward the older age groups.

According to United Nations (2001) age of 60 years and above is considered as the cut off for ageing. In Sri Lanka too age of 60 years and above is considered as the demarcation age in identifying the elderly population. It may be due to the fact that the most common mandatory retirement age in the government, private as well as the corporate sectors in Sri Lanka falls between ages 60 or 55 years, with an extension perhaps up to five years (Siddhisena, 2005). Nevertheless De Silva (1994) argued that although the definition of “elderly” or “aged”, varies from country to country, it is better to define as those who are of 60 years or above as ageing rather than taking 65 years and above.  Based on the above studies the elders who are in the age of 60 years and above are been considered as ageing population for this study.

## 1.2. Growth Rate of Ageing Population

Gavrilov and Heuveline (2003) mentioned that the population ageing is progressing rapidly in many industrialized countries, but in developing countries too there is a rapid increment of elderly people. Furthermore, Abeyratne, Alles, Wickremasinghe, and Hewapathirana (2014) mentioned that the speed of ageing is relatively higher in Asia and in fact in South Asia, Sri Lanka has the fastest aging population. The percentages of ageing population in South Asian countries in the year 1992 are given in Table 1.1.

Table 1.1: Percentage of population over 60 years of age in South Asian countries

| **Region/ country** | **Percentage of population age of 60 years and older** |
|---|---|
| Sri Lanka | 8.3 |
| India | 7.3 |
| South Asia | 6.7 |
| Bhutan | 6.2 |
| Nepal | 5.1 |
| Bangladesh | 4.8 |
| Pakistan | 4.6 |
| Afghanistan | 4.6 |

Source: De Silva (1994)

As per the Table 1.1 it is clear that the percentage of population of 60 years of age and older in Sri Lanka is much higher than the rest of South Asian countries. In fact, De Silva (1994) mentioned that, 8.3 percent of Sri Lanka's ageing population is relatively a large elderly population for a developing country. Prasannath (2014) mentioned that the percentage of the population of 60 years of age and above in Sri Lanka has increased to 12.2% with respect the population of 2.8 million in 2001.

Siddisena (2005) contended that the annual growth rate of Age 60+ is higher than that of the growth rate of total population in Sri Lanka. More over De Silva (1994) proved the same by mentioning that the growth rate of the aged population was more than double that of the total population during 1981 and 1991.

## 1.3. Demographic Factors which Cause Population Ageing

Manike (2014a) has mentioned that because of the change of high fertility and mortality to low fertility and mortality ("demographic transition") will cause for population ageing. She further argued that Sri Lanka is the first country in South Asia to reach replacement level fertility (an average quantum of current fertility which would be maintained by the next generation of daughters) and achieve low levels of infant mortality. More over De Silva (1994) mentioned that with rapid decrease of fertility which occurred early in time the onset of ageing process was accelerated. Aberathna et al. (2014) mentioned that the factors such as fertility control policies, vast education of reproductive practices, increases in the marital age limit of females has contributed to the decline in fertility. Furthermore they stated that effective application of DDT (*dichloro diphenyl trichloroethane*) in national efforts to eradicate malaria, improvements in the health care system and the expansion of free educational services have directly or indirectly contributed to the decline of mortality in Sri Lanka which leads to increase population over 60 years.

Gavrilov and Heuveline (2003) argued that population ageing is the inevitable result of rising life expectancy at birth. Aberathna et al. (2014) mentioned the same by providing the evidence in Sri Lankan context. Menike (2014b) also confirmed the same by mentioning that Sri Lanka is the first country in South Asia to reach high level of life expectancy. Abeyrathna et al. (2014) showed that the life expectancy at birth in Sri Lanka was 70.3 years for males and 77.9 years for females as at 2007 and this has been increased over time. The increasing trend of international migration within the working age group has become another factor which impacted heavily to increase ageing population within the country (De Silva, 1994).

## 1.4. Positive and Negative Impacts of Population Ageing

Ageing population has advantages as well as disadvantages. One negative aspect according to Aberathna et al. (2014) is that an ageing population results in an increase in the number of pensioners and retired. Eventhough the traditional custom in Sri Lanka is that the elders to be cared by their children, there is a diminishing trend of family support that elders receiving due to increased migration of young population from rural areas to urban areas and migration to overseas for employment (De Silva, 1994). The population ageing will impact to slowdown of labour force growth and its composition further will slowdown GDP growth as well (Vodopivec & Arunatilake, 2008). Governments of the countries with high ageing population have to provide necessary health care facilities, necessary programes to uplift elder’s income and to provide necessary welfare facilities (World Bank, 2008). Thus it can be concluded that ageing population is a big problem to any country with rising population over the age of 60 years and above including Sri Lanka.

One positive impact due to ageing population is their valuable experience which can be utilized fruitfully for the well being of the society (Menike, 2014a). However, it should be noted that majority of ageing population doesn't have valuable experience.

## 1.5. Common Methods Used for Population Projection in Sri Lanka

De Silva (2007) mentioned that even though number of methods are available to project population, only two common methods which are in practice: (i) mathematical model and (ii) cohort component method. Mathematical methods directly projects the total population, when initial size of the population and rates of the population growth in past, are known. Cohort component method used to project population by age and sex by considering the age and sex structure of the initial population. In this method it is assumed that population is influenced by fertility, mortality, births and migration. The age specific five year survivorship ratios and population data for midyear on five year age groups are utilized. The projection is also done for five year age brackets (Aberathna et al., 2014). These methods are discussed in details in Chapter two.

## 1.6. Projection of Ageing Population in Sri Lanka

Even though population ageing related topics has been discussed by many authors (Manike, 2014a; Manike, 2014b: Siddhisena, 2005; Prasannath, 2014; De Silva, 1994; De Silva, 2007) in Sri Lankan context all these studies are done in economics perspectives by applying descriptive statistics. These studies have various statistical drawbacks. Statistical thinking has been conspicuously absent in these papers. The validity of the models have not been carried out.

In Sri Lankan context, it was noted that only a single study has been carried out for the projection of future ageing population. It was done by De Silva (2007) and projected ageing population in Sri Lanka from 2001-2101 using cohort component method. Detail information about this study is discussed in Chapter two. Therefore it is vital to figure out an effective method to forecast the ageing population in Sri Lanka at least for short term and medium term basis.

## 1.7. Objectives of the Study

In view of the details explained above the objectives of this study are to:

a) develop a time series model for ageing population in Sri Lanka

b) validate the model

c) forecast the ageing population

## 1.8. Significance of the Study

Demographic forecasting (population, household and related other forecasting) is vital since such data are helpful for social and economic planning and are fundamental to many other forecasting exercises (Booth, 2006). Since ageing population forecasting is also a part of population projection, no doubts these projections will be helpful for government to plan necessary actions for better care of elderly population. Li, Reuser, Kraus, and Alho (2009) also claimed that it is necessary to predict the level of population ageing, to inform policy debates about the likely effect of different intervention strategies.

From Sri Lankas' point of view, forecast of ageing population in the future will lead to decide the basic requirements of elderly such as food, housing, personal care and consequently the government can take action with regard these economic matters, the health and long- term care of the elderly. Furthermore Nagarajan, Teixerria, and Silva (2013) highlighted the importance of carrying out forecasting of ageing population through the mathematical models, due to the lack of such studies carried out throughout the world.

## 1.9. Chapter Organization

This dissertation has eight chapters. The first chapter is the introduction chapter which provides the background of the study, problem statement, research objectives, and significance of the study. The second chapter is the literature review which provides the published research findings relating to the field of population projections and ageing population in Sri Lanka and other countries. Third chapter describes the materials and methods used for the study. Chapter four, five and six explain the three time series models developed to forecast the aging population in Sri Lanka and at the end of chapter six the statistical comparisons of the three methods are discussed. Conclusions and the recommendations of the study and directions of future research are given in chapter seven.

# CHAPTER TWO
# LITERATURE REVIEW

In this chapter the past research work relate to forecasting ageing population in Sri Lanka and in other countries are reviewed.

## 2.1. Overview about the Ageing Population from Various Parts of the World

Indeed population ageing is an universal phenomenon. Prasannath (2014) reported that the 21st century has been named as "The Era of Population Ageing". According to Gavrilov and Heuveline (2003) population ageing is expected to be among the most prominent global demographic trends of the 21$^{st}$ century. Therefore it is important to look after the dynamic of population ageing in the modern world. The percentages of elderly population with respect to the total populations for the two years 1950 and 2000 in the different parts of the world are shown in Table 2.1.

Table 2.1: Percentages of the elderly (65+ years) in selected regions and countries of the World: 1950 and 2000

| **Regions and countries** | **1950 (as a % of total population)** | **2000 (as a % of total population)** |
|---|---|---|
| Africa | 3.2% | 3.3% |
| Latin America and Caribbean | 3.7% | 5.4% |
| China | 4.5% | 6.9% |
| India | 3.3% | 5.0% |
| Japan | 4.9% | 17.2% |
| Europe | 8.2% | 14.7% |
| Italy | 8.3% | 18.1% |
| Germany | 9.7% | 16.4% |
| Sweden | 10.3% | 17.4% |
| U.S.A. | 8.3% | 12.3% |
| World Total | 5.2% | 6.9% |

Source: Gavrilov & Heuveline (2003)

Results in Table 2.1 indicate that the percentage of ageing population in the world with respect to the world population has increased from 5.2% (in 1950) to 6.9% (in 2000). Furthermore, it is foreseen that the percentage of ageing population has been increased in all the areas and countries with exception in Africa. Moreover, there was a significant increase of the percentage of ageing population in developed nations (Italy, Germany, Sweden and U.S.A) during the period of 1950 to 2000. Therefore, it can be mentioned that the developed nations lead the process of population ageing. Nevertheless among the given regions and countries it shows that the percentage of the ageing population of Japan has increased tremendously.

Gavrilov and Heuveline (2003) mentioned that the speed of ageing is relatively higher in Asia and in fact, Sri Lanka has the fastest ageing population among the South Asian countries. Furthermore, it has been recorded that there was 2.4 million of population in 1871 in Sri Lanka according to the records of the first housing and population census, but in the year 2012 it has been discovered that the elderly population alone in Sri Lanka was 2.5 million after 141 years of the first housing and population census (Department of Census and Statistics, Sri Lanka, 2015b). The results in Table 2.2 provide the better understanding of the changes of ageing population for the census years from 1953 to 2012.

Table 2.2: Growth of ageing population in Sri Lanka during the census years from 1953 to 2012

| **Year** | **Population(Thousands)** | | **As a percentage of total population** |
|---|---|---|---|
| | **Total** | **Age 60+** | **Age 60+** |
| 1953 | 8098 | 437 | 5.4 |
| 1963 | 10582 | 621 | 5.9 |
| 1971 | 12690 | 807 | 6.4 |
| 1981 | 14847 | 986 | 6.6 |
| 2001 | 18797 | 1738 | 9.2 |
| 2012 | 20425 | 2521 | 12.3 |

Source: Statistical Abstracts, Dept.of Census and Statistics, Sri Lanka, (1953-2012)

As per the Table 2.2 it can be seen that the ageing population in Sri Lanka has dramatically increased after the decade of 1980. To identify the changes recorded in the Sri Lankan population in three different age groups (0-14, 15-59 and 60+) during the census years of 1953 to 2012, are shown in the Table 2.3.

Table 2.3: Changes of the age structures of the population in Sri Lanka during the census year from 1953-2012

| **Year** | **As a percentage of total population in each age group** | | |
|---|---|---|---|
| | **0-14** | **15-59** | **60+** |
| 1953 | 41.7 | 57.2 | 5.4 |
| 1963 | 40.9 | 52.5 | 5.9 |
| 1971 | 38.7 | 54.7 | 6.4 |
| 1981 | 35.3 | 58.1 | 6.6 |
| 2001 | 26.3 | 64.5 | 9.2 |
| 2012 | 25.2 | 62.4 | 12.3 |

Source: Statistical Abstracts, Dept. of Census and Statistics, Sri Lanka, (1953-2012)

As depicted in Table 2.3 the proportion of the population (as a percentage of the total population) below 15 years has been decline over the years. According to Menike (2014b) these changes are occurring due to fertility changes in the reproductive age groups during the past years. In contrast percentage of ageing population over 60 years has been increased over the years.

Menike (2014b) classified that the people who are between the age of 60 and 74 years are considered "young elderly" and those who are over 75 years are called "old elderly". It has been highlighted that the oldest age category (75+) is increasing than the young age category (60-74) among the ageing population. The distribution of the population in these two categories are shown in Table 2.4.

Table 2.4: Distribution of elderly by age 1971-2012 (As a % of total elderly)

| **Year** | **60 – 74 Years** | **75+ Years** |
|---|---|---|
| 1971 | 80.5 | 19.5 |
| 1981 | 78.9 | 21.1 |
| 1991 | 78.8 | 21.2 |
| 2001 | 76.3 | 23.7 |
| 2012 | 75.9 | 24.1 |

Source: Statistical Abstracts, Dept. of Census and Statistics, Sri Lanka, (1971-2012)

Results in the Table 2.4 indicates that the percentage of oldest age category has increased from 19.5% to 24.1% during 1971 to 2012. In contrast corresponding figures in young elderly has decreased from 80.5% to 75.9% during 1971-2012.

The sex ratio (number of males per 100 females) of young elderly and older elderly in different time periods are shown in Table 2.5. Moreover another important trend of ageing population is that the proportion of the female aged population has increased rapidly. According to Manike (2014b) the main reason for this is the growth of long life expectation of women. Manike (2014b) also mentioned that as the life expectancy of the females is longer than males, a large numbers of females survive to old age compared to males.

Table 2.5: Sex ratio of the elderly population in Sri Lanka 1971-2012

| **Year** | **Sex ratios (number of males per 100 females) of different age groups** | | |
|---|---|---|---|
| | **60-64** | **75+** | **All 60+** |
| 1971 | 126.0 | 106.2 | 121.8 |
| 1981 | 114.6 | 107.2 | 112.9 |
| 2001 | 89.11 | 84.6 | 88.05 |
| 2012 | 86.4 | 67.9 | 79.43 |

Source: Statistical Abstracts, Dept. of Census and Statistics, Sri Lanka, 1971-2012

As depicted in the Table 2.5 the sex ratio among the elderly population has decreased over the years. The sex ratio among the old elderly population declined from 106 in 1971 to 68 in 2012.

Another trend in ageing population in Sri Lanka is widowhood is more prevalent among elderly women than men. The proportion of widows 60-64 years of age is about five times that of widowers in the same age group due to three reasons: (i) the life expectancy of woman is longer than that of their husbands (ii) the high life expectancy of women and (iii) the majority of the widowers get married again unlike the widows (Menike, 2014b).

## 2.2. Demographic Factors which Affected the Rise of the Ageing Population

According to Gavrilov and Heuveline (2003) there were several important demographic factors such as fertility, mortality, and migration which have affected a lot to increasing the ageing population during the past few decades. Fertility means the actual reproductive performance of an individual, a couple, a group or a population (De Silva, 2007). Mortality means the deaths as a component of population change (De Silva, 2007) and migration means the movement of people across a specified boundary for the purpose of establishing the new permanent residence (De Silva, 2007). Furthermore De Silva (1994) proved the same by mentioning that the proportion of the Sri Lankan population aged 60 and over raised from 5% in 1946 to 8% in 1991 due to the combined results of fertility, mortality and international migration trends. Apart from these factors there was a dramatic decrease of crude births (the number of births per 1,000 population in a given year) and crude deaths (the number of deaths per 1,000 population in a given year) in Sri Lanka along with the increment of life expectancy at birth (the average number of additional years a person would live if current mortality trends were to continue) due to the development of the health sector in Sri Lanka which has affected to increase the agers in the country (Department of Census and Statistics, Sri Lanka, 2015b). Furthermore, there was an increasing trend of international migration among the young population in the country. All the mentioned factors have collectively caused to increase the ageing population in the country. Table 2.6 shows how the statistical changes happened to the crude birth rate, crude death rate and total fertility rates from 1946 to 2012.

Table 2.6: CBR, CDR, and TER during the census years of 1946 to 2012

| Year | Crude Birth Rate | Crude Death Rate | Total Fertility Rate |
|---|---|---|---|
| 1946 | 37.4 | 19.8 | 5.5 |
| 1963 | 34.1 | 8.5 | 5.0 |
| 1981 | 28.2 | 5.9 | 3.8 |
| 2012 | 17.5 | 6.0 | 2.4 |

Source: Department of Census and Statistics, Sri Lanka, (2015)

As per the Table 2.6 it is clear that the crude birth rate (CBR), crude death rate (CDR) and the total fertility rate (TFR) in Sri Lanka has decreased throughout the census years of 1946, 1963, 1981 and 2012.

### 2.3. Demographic and Economic Impact of Ageing Population

It is vital to look into the economic and demographic impacts due to ageing population in any country. Prasannath (2014) argued that economic activity are reduced or stopped among aged population according to the “Theory of Activity”. This means that the aged are disengaged from their work in latter age. He also argued that economic impacts such as income support for the lengthened retirement period, changes in consumption and production, changes in the structure of labour market, changes in housing, needs of transport, and expenditure on health would arise due to problem of ageing population.

De Silva (1994) highlighted the importance of finding out ways and means to provide social and economic support for ageing population due to the social changes such as migration and urbanization, increased family female labour force participation. Due to these reasons, he further argued that generations of a family may live in different places, that they may live in a place where there is not enough space to accommodate a multi-generational family. The changes in the life-style and responsibilities of the children, the greater involvement of females in employment away from home, limit the amount of time available for caring for the elderly. Thus it can be hypothesized the above factors would contribute to the gradual decline of the traditional family-based support system.

Another impact according to Vodopivec and Arunatilake (2008) was that health system is insufficiently focused on the healthcare needs of elderly and is constrained by the lack of resources and their inequitable distribution. Thus it should be pointed out that with the temporal increase in the size of elderly population in Sri

Lanka, it is important to find out how health services can be catered to provide the needs of the elderly.

Population ageing may well translate into shrinking of the labour force, prompting questions of how to promote longer working lives and also of how to improve labour market choices for those that are forced to continue working late in their lives (Vodopivec & Arunatilake, 2008). Prasannath (2014) also mentioned that the income gained from social security benefits; formal pensions, EPF, PSPF and ETF or other means of old age protection schemes such as public assistance and safety net programmes, health or life insurance schemes, labour market participation and investment collection are not sufficient to cover expenditure for health and routine living in older years.

## 2.4. Social and Economic Implication for Population Ageing

Different authors have suggested different solutions for the implications due to the problem of ageing population. Vodopivec and Arunatilake (2008) have suggested that the importance of increasing the labour force participation of old aged workers, together with importance to make more flexible retirement age. De Silva (1994) claimed that social and economic planners must take into consideration the economic situation of all age groups in formulating employment and retirement policies for elderly people. It has also been suggested the importance of improving the skill of older workers through investments in lifelong learning and improving the choices available to old aged workers furthermore the importance of immediate formation of organized public support in the form of social security has been suggested. However, it should be pointed out that these kinds of schemes take fairly long time to establish and expand. Due to temporal increasing elderly population, the number of elderly people receiving support from pensions or provident funds and it’s amount need to be increased.

## 2.5. Population Projection Methods

Norbert (1993) mentioned that population projection has been one of the most important contributions made in demography and related disciplines. United Nations has contributed enormously as a pioneer and leading contributor to demographic estimates and projections (Norbert, 1993). Most of researchers mentioned about the importance of considering about the changes in mortality, fertility and migration in population projection (De Silva, 2007; Booth, 2006; Nobert,1993).

Different Authors used different methods of project population. The methods used in Sri Lanka are (i) mathematical method (ii) cohort component method. Mathematical method directly projects the total population, when the initial size of the population and the assumptions on future rates of population growth are given. One disadvantage of this method is that the age specific population projections cannot be ascertained.

The cohort component method, project population by age and sex, employs the age and sex structure of the initial population, together with the assumption on the future components of population change due to fertility mortality and migration. The major disadvantage of this method is that as the level of fertility, mortality and migration cannot be accurately assessed, therefore population projections are not accurate (Norbert, 1993).

Stanbery (1952) has introduced four forecasting techniques. The first method is graphical or mathematical projections of the curve of past population growth (trend based method). The second method based on relationship of population growth in an area to that in other areas (ratio method). Third method is the method of component analysis which study births, deaths, and migration separately which affects to projected population. The last technique is forecast based on specific estimates of future employment and other occasionally used methods. He mentioned that the trend based methods assume that population growth follows natural laws and, therefore, can be expressed in mathematical or graphical form. Basically, population is forecast by examining and projecting past trends into the future. Various types of expressions have been used such as linear, geometric, exponential, logarithmic, etc., to explain past historical growth and predict future growth. One advantage of this method is, easiness and flexibility to apply, but the disadvantage is the assumption of considering the same trend irrespective of time period.

According to Stanbery (1952) the second technique (ratio method) based on the past relationship between population growths in an area will be a valuable guide to project the total population. The theory behind the third method (component analysis) is that more accurate estimates can be made using the rates of change of the individual components of population that using the rates of changes of the population as a whole.

For example, it is reasonable to assume that birth, death, and migration rates for 80-year old people are different than from those for 20 year old people and that, based on historic experience, one can forecast the rates for such groups with reasonable accuracy. Within the component analysis (i) the natural increase and net migration, and (ii) cohort survival and net migration are two common methods used by many authors (Stanbery, 1952).

Zakria and Muhammad (2009) also carried out a similar review about the available population forecasting methods and according to their review they mentioned that the mathematical methods such as linear, non linear, first and higher degree regression models, simple and double exponential, logistic regression, simple decay and growth models are being used by different authors.

Booth (2006) carried out a review mentioning the demographic forecasting methods used by different researchers throughout the world during the period of 1980 to 2005. She claimed that there are main three approaches of forecasting demographic components (mortality, fertility and migration) namely (i) extrapolative methods (ii) method based on expectation and (iii) structural modeling. The extrapolative methods focus on the regularity of patterns and trends and extend these into the future without recourse to other knowledge in the form of exogenous variables. The methods based on expectation may use individual data (such as surveys of women's future birth expectations or the opinions of experts about future demographic developments). The structural modeling/ explanation methods are seeking to explain demographic processes use structural models based on theories relating demographic quantities to other variables.

Apart from these projection methods Booth (2006) mentioned that there are two other methods known as probabilistic forecasting methods (method which concern about the component of uncertainty) and direct population forecasting method for population projection (using autoregressive model, logistic function). The choice of approach depends on two factors including data availability and the purpose of the study and further mentioned the followings:

- Demographic accounting or cohort-component method is suitable for long term population forecasting
- Statistical time series method is better for short-term forecasts of demographic rates

- Structural modeling method is more appropriate for the simulation and forecasting of policy changes.

## 2.6. Population Projection Method in Sri Lanka

Moreover in Sri Lanka other than the census years, midyear ($1^{st}$ of July of the considering year to $30^{th}$ of June the next year) population estimations are carried out by the Registrar General's Department annually. De Silva (2007) pointed out the importance of carrying out this kind of population estimations due to the circumstances of holding population and housing census once in 10 years. They are using the balancing equation to project the population. The balancing equation is given in (2.1).

$$P_t = P_o + (B - D) + (I - E) \tag{2.1}$$

Where,

| | | |
|---|---|---|
| $P_t$ | = | Population at the considering period t |
| $P_o$ | = | Population at the beginning of the period |
| B | = | Birth during the period |
| D | = | Deaths during the period |
| I | = | Immigrates during the period |
| E | = | Emigrates during the period |
| (I-E) | = | Net Migration |
| (B-D) | = | Natural increase of the population |

Furthermore, the Registrar General's Department used to predict the age and sex specific population in Sri Lanka by multiplying the district wise population estimations from each age group. This method is something similar to the population projection methodology introduced by De Silva (2007).

## 2.7. Ageing Population Projections via Cohort Component Method

De Silva (2007) carried out a projection using the cohort component method (using five year age groups) to forecast the Sri Lankan population based on gender from 2001 to 2101 (for ten decade time period). For this study he used the population census (census usually carried out in Sri Lanka once in ten years) data gathered from Department of Census and Statistics, Sri Lanka. He used year 2001 censured population in Sri Lanka by age and sex as the base population for the projections. De

Silva (2007) mentioned that in the computation process separate projections are made for males and females in the five year interval of the projection period. Five year after the base year is obtained by multiplying the base population by age-sex-specific five year survivorship ratios. The formula used is given in (2.2);

$$P_{x+5}^{t+5} = P_x^t \quad S_{x\,to\,x+5}^{t+2.5} \qquad (2.2)$$

Where,

$P_x^t$ - The number of person of a given sex at mid year t in five year age group x,

$S_x^t$ -The sex specific five year survivorship ratio of a given sex.

Using the equation (2.2) he computed the ageing population as well. Table 2.7 indicates the few standard elderly population projections done by De Silva (2007) using the cohort component method.

Table 2.7: Standard elderly population projections

| **Year** | **Projected elderly population (60+ years) (in '000)** | **Actual elderly population (60+ years) (in '000)** |
|---|---|---|
| 2001 | 1,731.4 | 1738 |
| 2006 | 2,075.7 | 1834 |
| 2011 | 2,570.4 | 1930 |
| 2016 | 3,070.2 | 2623 |
| 2021 | 3,605.1 | |

As per the Table 2.7 it is clear that there will be an continuous increasing trend of ageing population in Sri Lanka during the time period of 2001 to 2021 according to the projection done by De Silva (2007). But when comparing the actual elderly population it was clear that all the ageing population forecast done by him are over estimated other than the base year (2001) aging population.

## 2.8. Population and Ageing Population Projections

Norbert (1993) carried out a study to project population growth in Sri Lanka using linear trend analysis based on population census data from 1871 to 1981.Based on his model projections has been done for the entire country for two agro climatic zones (wet and dry zones) and for the nine provinces separately.  However a statistical validity of these models or accuracy of forecast values have not been considered.

Furthermore Norbert (1993) mentioned about other population projections which has been carried out in Sri Lanka from time to time. Sarkar (1957) as cited in Norbert (1993) predicted two different projections for the period 1951-1976, based on a mathematical method and the component projection method. United Nations (1958) as cited by Norbert (1993) has prepared three projections for the period 1955-1980 under medium, low and very low growth scenarios. Furthermore, he has mentioned that the Department of Census and Statistics (1974) has carried out three projections for 1971-2000 using the component method on the basis of three different fertility assumptions combined with a single assumption of future mortality trend. Srivatsava and Abeykoon (1974) as cited in Norbert (1993) have made four population projections in which international migration has been taken into consideration and these projections dealt with the demographic situation in Sri Lanka. Norbert (1993) has concluded his review by mentioning that population projections in Sri Lanka have been mainly attempted by different authors in relation to planning and policy reviewing point of view. He further emphasized that no statistical methods have been used.

However past studies found that various approaches have been used by other countries. Abel, Bijak, and Raymer (2010) carried out a study to project population in England and Wales based on Bayesian time series approach by using autoregressive (AR) and stochastic volatility (SV) models. They have prepared these models by using historical time series data from 1841 to 2007 and predictions were done up to 2033. Furthermore they compared these projections with the projections made by the Office for National Statistics. They have applied sensitivity analysis and further they have compared the sample forecast with actual forecast and previous official forecast. They concluded the study by showing the importance of the application of Bayesian approach for population projections apart from the classical estimation methods leading to more realistic forecasts and associated uncertainty measures.

Li et al (2009) predicted the population in China up to 2060 with a focus on the process of population ageing, and quantifies the expected uncertainty using stochastic models. Furthermore they have mentioned that they used cohort-component method for forecasting and they developed three stochastic models separately for age-specific fertility, age- sex-specific mortality and net migration flows by sex and age.

According to their study they concluded that oldest – old will grow faster than any other age group in China within the years 2006-2060.

Furthermore Keilman, Pham, and Hetland (2002) have predicted the population in Norway up to 2050. They also applied three main methods to compute probabilistic forecast namely (i) time series extrapolation, (ii) analysis of historical forecast errors, and (iii) expert judgment. Moreover they demonstrated that these three methods can be combined when computing prediction intervals for population's future size and age and sex composition. Their predictions were compared with those of the official population forecast compiled by Statistics Department of Norway.

Wei, Jian, and Zhang (2015) studied two population growth models to find the factors affecting the Chinese population growth. Throughout this paper they have reviewed about the applicability of exponential growth models and logistic growth models by taking the China population as the target group. As the conclusions they have mentioned exponential growth model and logistic growth model is not suitable for predicting population growth. Furthermore they found that the two factors namely the degree of urbanization and sex ratios have significant influences on population growth in China.

Zakria and Muhammad (2009) attempted to forecast the population in Pakistan using population data from 1951 to 2007 with the application of Box Jenkins ARIMA methodology. They have considered several parsimonious models and identified ARIMA (1,2,0) model as the best fitted model and estimated that there will be approximately 230.7 million population in Pakistan in 2027. The model was validated by the criteria of MSE, AIC, p values and graphical techniques such as ACF, PACF and plots of residuals. Furthermore they mentioned that the forecasted population by them using the ARIMA (1,2,0) model was close to the projected population by different bureaus.

## 2.9. Use of Time Series Models

Paul (2011) has conducted a research to determine the exponential smoothing constant which comes under one of the time series technique of exponential smoothing. In this study he discussed in detail about the way of choosing exponential smoothing constant which is very crucial to minimize the error in forecasting. He determined the optimal value by minimizing the Mean Square Error (MSE) and Mean

Absolute Deviation (MAD) using trial and error method. But the validity of the model has not been tested even for the training dataset.

Aberathna et al. (2014) developed sex specific mortality estimation models using historical mortality data for Sri Lanka, based on the statistical time series techniques. Several alternative univariate time series models were examined for modeling males and females, as well as bivariate vector autoregressive (VAR) model. From the estimated VAR model, mortality forecast were generated for the period up to 2030 and further life tables were generated for the selected period of 2006-2008.

Cheruiyot (2015) developed ARIMA (1, 2, 0) to forecast the Kabianga university students population in Kenya. The model was validated using various methods such as Akaike Information Criterion (AIC), time series plots, Schwarz Bayesian Criterion (SBC). Results indicate that the student population will grow to 32,421 by the year 2023 when only time is considered as a factor. The results further depicted a positive steady increase of student population for University of Kabianga over the next ten years.

## 2.10. Other Studies Related to Ageing Population

Apart from the population projection studies mentioned in 2.8 it had been carried out few ageing population related studies in economics point of view. De Silva (1994) has done a study under the topic of how serious ageing in Sri Lanka and what can be done about it? In this study he clearly mentioned about the trends of ageing population in Sri Lanka and how it became a problem. He further mentioned about the economic and social implications as the solutions for the increasing trend of the population over 60 years of age. Furthermore, he produced some descriptive statistics about the ageing population, ageing population growth rates, dependency ratios and index of ageing using secondary data in Sri Lanka for the time period of 1946 to 1981. Based on these descriptive statistics, De Silva (1994) predicted some future values of ageing population.

## 2.11. Summery of the Review

The ageing population has been increasing over time in almost all countries in the world. The highest ageing population growth rate was observed in Japan while in South Asian countries it is highest in Sri Lanka. It has been found that economics impacts such as insufficient income support for the lengthened retirement period, changes in the structure of labour market, changes in housing, needs of transport and expenditure on health will arise due to the increasing trend of ageing population. Fertility, mortality, migration and birth found as the demographic factors which impact to population changes.

Some of the common population projection methods used are cohort component method, mathematical method, stochastic methods, and direct forecasting methods. Among the studies carried out to forecast the population/ageing population most of the past researchers have done projections based on the cohort component method. Furthermore, no attempt was taken to forecast ageing population in Sri Lanka using mathematical or direct method. However, various methods have been used by many authors in different countries to forecast ageing population. Nevertheless it was found most of these methods have not been statistically validated. Furthermore, less attention was given for statistical methods in particularly time series models. The results acquired from the review would be immensely useful to carry out the present study.

# CHAPTER THREE
# MATERIALS AND METHODS

In this chapter details of the secondary data and the statistical methods used for the study are described.

## 3.1. Materials

### 3.1.1. Research Methodology

Sekaran and Bougie (2010) mentioned that there are two types of research methodologies namely quantitative and qualitative. Quantitative research methodology involves application of statistical analysis which provides numerical values to make conclusions and to test specific hypotheses. In this study too it is expected to apply statistical analytical tools to make conclusions. Therefore it can be argued that the research methodology of the present study is quantitative research methodology.

### 3.1.2. Unit of Analysis

The unit of analysis refers to the level of aggregation of the data collected during the subsequent data analysis stage (Sekaran & Bougie, 2010). Furthermore, they mentioned that individual, dyads, groups, organizations, cultures, and country level as few examples for unite of analysis. Since it has been collected ageing population data in Sri Lankan from 1950-2016, the unit of analysis of the present study is the national level.

### 3.1.3. Time Horizon

According to the time horizon three types of data are available for empirical analysis namely cross sectional, pooled and both cross sectional and pooled data (Gujarati, Porter, & Gunasekar, 2012). Furthermore, they mentioned that the time series data is a set of observation on the values that a variable takes at different time (daily, weekly, monthly, quarterly, annually etc.). Since in the present study the ageing population data in Sri Lanka had been collected from 1950-2016, the time horizon of the present study is time series.

### 3.1.4. Secondary Data

Data can be obtained by either primary or secondary sources. Secondary data refer to information gathered by someone, other than the researcher conducting the current study (Sekaran & Bougie, 2010). Books, periodicals, government publications of economic indicators, census data, statistical abstracts, annual reports of companies are few examples where secondary data can be obtained.

Ageing population data from 1950 to 1991 were acquired from statistical abstracts published by the Department of Census and Statistics and the corresponding data for the period of 1992 to 2016 were acquired from Registrar General's Department. When carefully examined the gathered data during 1950-2016, it was found that it consisted of both the censured and estimated ageing population data.

According to the Department of Census and Statistics, Sri Lanka (2015a) it has been carried out only six housing and population census during 1950-2016 in Sri Lanka. They are in the years, 1953, 1963, 1971, 1981, 2001 and 2012. The estimated ageing population derived from Registrar General's Department during 1950-2016 along with the censured ageing population in Sri Lanka for said time period were used for the present study. The corresponding data set is shown in Appendix A.

## 3.2. Statistical Analysis

Since the main purpose of the present study is to model and forecast the ageing population of Sri Lanka in the future and has been applied three different time series modeling techniques as the analytical tools. These techniques are Autoregressive Integrated Moving Average, growth model and double exponential smoothing method. Data were analyzed using Minitab statistical software version 16, SPSS software version 21, version 8 of the EViews software and Microsoft Office Excel 2007.

### 3.2.1. Trend Analysis

Chandan, Singh, and Khan (2011) mentioned that it can forecast the future values based on the trend of a given data series. They mentioned that trend can determine by graphical method or through the methods of semi average, moving average method and least square method. Tulsian and Jhunjhnuwala (2010) mentioned that under least square method it can be derived three equations namely straight line trend equation, quadratic trend equation and exponential trend equation. Method of least squares is a mathematical technique employed for finding equation of a specified type of curve which best fits a given series. This is based on the principle that for the "best fitting" curve the sum of squares of difference between the observed and the corresponding estimated values obtained from the equation, should be minimum possible.

The linear trend equation is as follows,

$$Y_t = a + bt + e_t$$

Where, $Y_t$ = Value of the series at time t

$e_t$ = error and it is assumed $e_t \sim ii\ D\ N(0, \sigma^2)$

a and b are the parameters and can be estimated as:

$$\hat{a} = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$\hat{b} = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Other types of the trend models generally used are;

Quadratic trend: $Y_t = a + bt + ct^2$

Growth model: $Y_t = ab^t$ or $\text{Log } Y_t = \text{Log } a + t^*(\text{Log } b)$

### 3.2.2. Exponential Smoothing

Exponential smoothing is one of the time series technique which is widely used in forecasting. Single exponential smoothing and double exponential smoothing are two famous methods come under exponential smoothing technique. Exponential smoothing is a very popular methodology to produce a smoothed time series (Paul, 2011).

#### 3.2.2.1. Single Exponential Smoothing

Rani and Raza (2012) mentioned that single in exponential smoothing is suitable for forecasting time series with no trend or seasonal pattern. The smoothing equation can be written as,

Next period forecast = weight (present period observation) + (1-weight)*present period forecast

That is, $F_{t+1} = \hat{y}_{t+1} = \alpha y_t + (1-\alpha) F_t$

Where $F_{t+1}$ = Forecast value at time t+1

$\hat{y}_{t+1}$ = Estimator for $y_{t+1}$ and $\alpha$ is the smoothing constant.

The smoothing equation is based on averaging (smoothing) past values of a series in decreasing (exponential) manner. The observations are weighted, with more weights given to the more recent observations. The weights used are $\alpha$ for the most recent observation, $\alpha(1-\alpha)$ for the next most recent, $\alpha(1-\alpha)^2$ for the next, and so on. At each time , the weighted observation along with the weighted estimate for the present period are combine to produce a new period forecast. Therefore the equation can extend as follows,

$$F_{t+1} = \hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 t_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \ldots\ldots\ldots. + \alpha(1-\alpha)^{t-1} y_t$$

According to Paul (2011) depending on the correct choice of $\alpha$ it is possible to tell the accuracy of this method. Furthermore he has pointed out some advantages and disadvantages of exponential smoothing. The major advantages are relatively good short term accuracy, simplicity, low cost, not require large amount of historical data and forecast are easy to obtain and updating only depends on last data point. The start up time requiring to find the "best" $\alpha$ along with process of continuously monitoring and updating the value of $\alpha$ and only one period ahead can be forecasted are two major disadvantages of this method.

#### 3.2.2.2. Determination of the Exponential Smoothing Constant

Different authors have introduced different methods to determine this exponential smoothing constant α. Paul (2011) came up with a finding of determining the smoothing constant. He suggested, choosing an appropriate value of exponential smoothing constant is very crucial to minimize the errors in forecasting. He demonstrated how to select the exponential smoothing constant by minimizing Mean Square Error (MSE) and Mean Absolute Deviation (MAD) through "Trial and Error method" from his research work. He further mentioned that through the trial and error method MSE and MAD should be calculated separately by putting different values within 0 to 1 for exponential smoothing constant.

#### 3.2.2.3. Double Exponential Smoothing (DES)

Double exponential smoothing can be able to apply to a time series with a trend (Rani & Raza, 2012). According to Siregar, Butar, Rahmat, Andayani, and Fahmi (2016) double exponential smoothing method smoothed trend component separately using the two parameters namely $\alpha$ and $\beta$ . Therefore he further mentioned that these two parameters need to be optimized so the search for the best combinations of parameters is more complicated than using only one parameter. In this method the original series is smoothed two times. The basic idea behind DES is to introduce a term to take in to account the possibility of a series exhibiting some form of trend. Equation of DES is as follows:

$$\hat{Y}_{t+1} = a_t + b_t$$

Where , $a_t = 2S_t^{'} - S_t^{''}$ (Updated intercept/the estimated level at period t)

$$b_t = \frac{\alpha}{1-\alpha}(S_t^{'} - S_t^{''})$$ (Updated slope/the estimated trend at period t )

t= number of the time period ahead.

$S_t^{'}$ *and* $S_t^{''}$ are the single and double smoothing statistics found by applying the exponential smoothing equation. The value of $S_t^{'}$ at period t is given by

$$S_t^{'} = \alpha\, y_t + (1-\alpha) S_{t-1}^{(1)}$$

$$S_t^{''} = \alpha S_t + (1-\alpha) S_{t-1}^{(2)}$$

In this case both smoothing constants were taken as the same. However some authors have taken two different values for smoothing constants (Handanhal, 2013; Paul, 2011).

### 3.2.3. Box- Jenkins Approach for ARIMA Methodology

This is a forecasting procedure which has been developed by Box and Jenkins (Cheruiyot, 2015).The main reason behind using Box and Jenkins technique for modeling is that it has been shown to give relative accurate forecast (Jha, Sinha, Arkatkar, & Sarkar, 2013). Furthermore Cheruiyot (2015) mentioned that in the analysis of future data ARIMA model has proven its viability when it comes to giving adequate results due to its predictive power. According to Gujarati et al. (2012) one of the reasons for the popularity of the ARIMA modeling is its success of forecasting particularly in short term forecasting

Box and Jenkins introduced four types of time series models such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA)**.** The ARIMA model has three parameters p, d, and q and is often written as ARIMA (p,d,q). The order of AR part is p where that the current value of the series $\{y_t\}$ can be explained as a function of p number of steps in to the past. The parameter d represents the level of differencing the original time series needs to undergo to become it stationary. The order of moving average model is q (MA(q)) assume the white noise of the right hand side off the defining equation are combined linearly to form the observed data (Smith & Arawal, 2015).There are several stages of setting up a Box –Jenkins forecasting model and it can be illustrated as per in the Figure 3.1

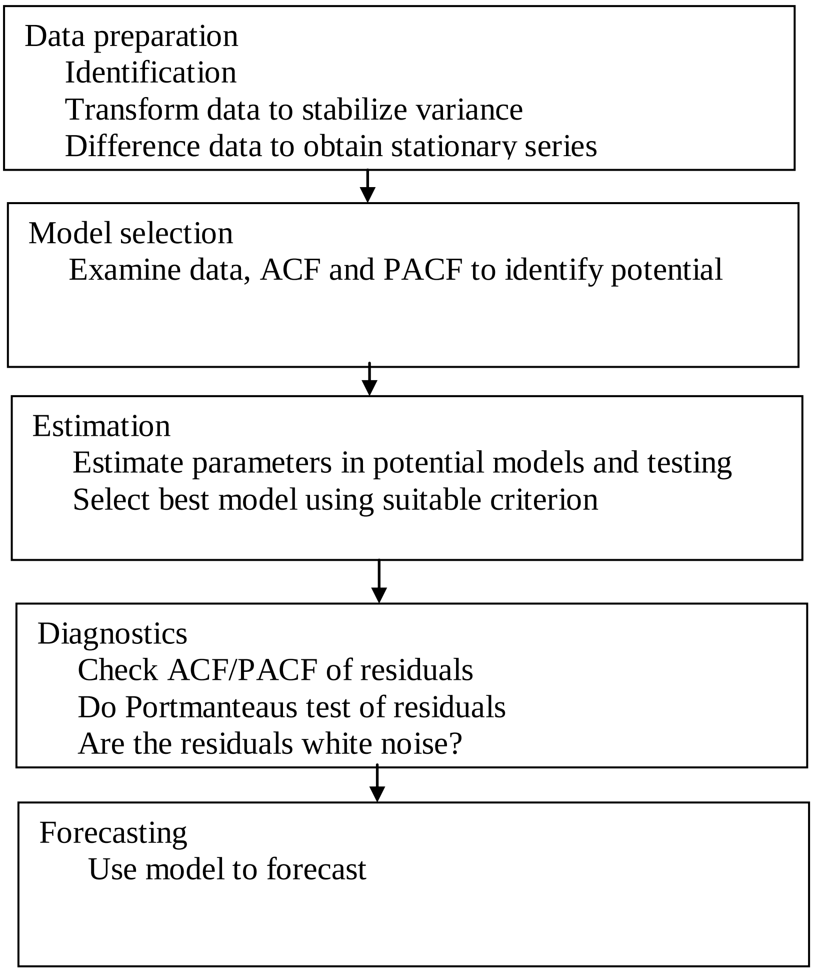



Figure 3.1: The Box – Jenkins methodology

As per the Figure 3.1 the first step is to prepare data. In Box –Jenkins methodology it is compulsory to make the series stationary by applying necessary data transformation methods or differencing the series. As the second step it should select the model after carefully examination of the ACF and PACF. Then have to estimate the parameters and testing, select the best model using suitable criteria. Then have to conduct diagnostic checks for the residuals. Last step in Jenkins methodology is to use the fitted model for forecasting.

Autocorrelation and partial autocorrelation plots are heavily used in time series analysis and forecasting.

#### 3.2.3.1. Autocorrelation Function (ACF) and Sample Autocorrelation Function (SACF)

According to Gujarati et al. (2012) autocorrelation function at lag k is defined by, $\rho_k = \frac{\gamma_k}{\gamma_o}$ and the sample autocorrelation function at lag k is as follows,

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_o}$$

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=b}^{n}(y_i - \bar{y})^2}$$

The order of significant autocorrelation of a stationary series is considered as the order of MA part.

#### 3.2.3.2. Partial Auto Correlation Function (PACF)

The partial autocorrelation function between $Y_t$ and $Y_{t+k}$ is the conditional correlation between $Y_t$ and $Y_{t+k}$ and the partial autocorrelation function is as follows,

$$\phi_{kk} = Corr(Y_t, Y_{t+k} / Y_{t+1,} Y_{t+2,} .........., Y_{t+k-1})$$

In time series analysis PACF gives the partial correlations of time series with its own lagged values controlling for the values of the time series at all shorter lags. The partial autocorrelations at lag k (k= 1, 2 , 3 ……….) for AR model is generally computed using Yule Walker equation (Gujarati et al. , 2012). The order of significant partial autocorrelation of a stationary series is considered as the order of AR part.

## 3.3. Stationary and Non Stationary Time Series

It is said a given time series is a stationary, (weakly stationary) series only if its statistical properties (mean and variance) are constant through time, otherwise (mean and variances are not constant) such a series is stated as non stationary series. Classical Box – Jenkins model describes stationary time series. Therefore it is compulsory to transfer a non stationary time series into stationary series prior to applying Box – Jenkins model (Adikari & Agrawal, 2013)

The visual plot of time series is often enough to say that the data are stationary or not. The autocorrelation of stationary data drop to zero relatively, while for a non stationary series they are significantly different from zero for several time lags. It can be able to identify non stationary series graphically if autocorrelation data decreases slowly as the number of time lags increase. Otherwise statistical test of Augmented Dickey – Fuller test (ADF) can be used to identify the stationarity of a given series (Adikari & Agrawal, 2013).

### 3.3.1. Augmented Dickey – Fuller (ADF) Test

Augment Dickey Fuller (ADF) is used to test whether the given series is a stationary or not (Kwiatkowski, Philip, Schmidt & Shin, 1992). This was developed by Dickey and Fuller in 1979.The null hypothesis ($H_0$) of ADF test is the series is not stationary or series has a unit root $(\phi_1 = 1)$ and the alternative hypothesis ($H_1$) is the series is stationary or series don't has a unit root $(\phi_1 < 1)$.

### 3.3.2. Removing Non – Stationarity in a Time Series

One way of removing non – stationary is through the method of differencing (Smith & Agrawal, 2015). It is able to define the differenced series as the change between each observation in the original series: If the time series values are as $y_1, y_2, y_3, \ldots\ldots\ldots\ldots\ldots, y_n$ and the series is not stationary, then it is necessary to take the first differenced series $\nabla y_t = y_t - y_{t-1}$.

If the first difference of the original series is also non stationary, then it is necessary to take the second differenced series;

That is, $Z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = (1 - 2B + B^2) y_t$ or

Where the operator B such that $B^q y_t = y_{t-q}$.

### 3.4. Basic Description of ARIMA (p,d,q) models

#### 3.4.1. An Autoregressive (AR) Process of Order p

A time series $\{Y_t\}$ is said to be AR process of order p, if $Y_t$ can be expressed in the form of,

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots\ldots\ldots.. + \phi_p Y_{t-p} + e_t$$

Where $\phi$'s are real constants and $\{e_t\}$ is purely random process with mean zero and constant variance (Gujarati et al., 2012).

#### 3.4.2. ACF of AR (1) Process

ACF of AR(1) process: $Y_t = \phi_1 Y_{t-1} + e_t$ is given by $\rho_k = \phi_1^k$, k=0,1,………… Thus ACF of AR (1) is exponentially decay, The theoretical ACF and PACF of AR(1) is shown in Figure 3.2.

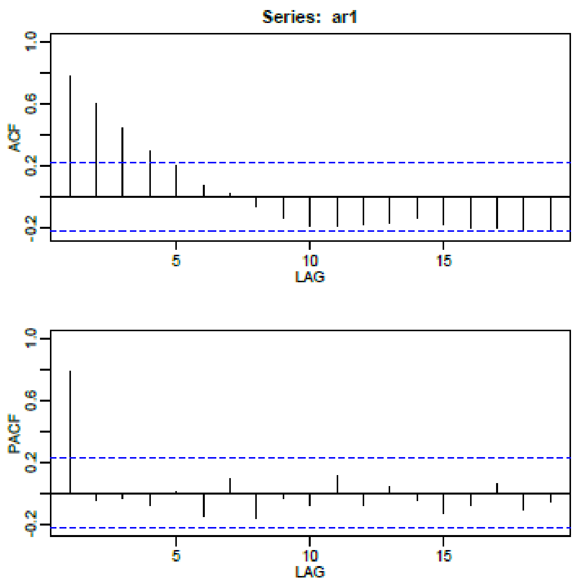



Figure 3.2: Theoretical ACF and PACF for AR (1)

According to the Figure 3.2, AR (1) process can be identified if the PACF has one large spike at lag 1 (positive or negative side) with exponential decay of ACF (Gujarati et al., 2012).

### 3.4.3. A Moving Average (MA) Process of Order q

This model can be represented by,

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ..... - \theta_q e_{t-q}$$

Where $\theta's$ s are real constants and $\{e_t\}$ is purely random process with mean zero and constant variance (Gujarati et al. , 2012).

### 3.4.4. ACF of MA (1) process

The autocorrelation function of MA(1): $Y_t = e_t - \theta_1 e_{t-1}$ is given by,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{-\theta_1}{(1+\theta_1^2)}, & k=1 \\ 0, & k>1 \end{cases}$$

Figure 3.3 depicts theoretical ACF and PACF of MA (1)

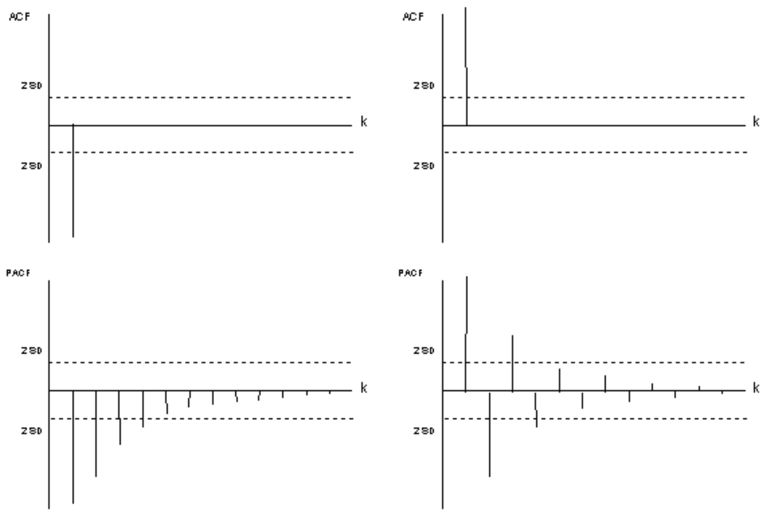



Figure 3.3: Theoretical ACF and PACF for MA (1)

According to the Figure 3.3 if the ACF has one large spike at lag 1 (positive or negative side) with exponential decay of PACF it can be easily identified as MA(1) process (Gujarati et al., 2012).

### 3.4.5. ARMA (p,q) Process

In time series analysis Autoregressive Moving Average (ARMA) contains both AR and MA parameters. ARMA (p,q) model can be represented by,

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ....... + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ..... - \theta_q e_{t-q}$$

### 3.4.6. Theoretical Patterns of ACF and PACF

Gujarati et al. (2012) has given the guidelines to identify the relevant stochastic process (Table 3.1).

Table 3.1: Theoretical Patterns of ACF and PACF

| Types of Model | Typical pattern of ACF | Typical Pattern of PACF |
|---|---|---|
| AR(p) | Decay exponentially or with damped sine wave pattern or both | Significant spike through lag p |
| MA(q) | Significant spikes through lag q | Declines exponentially |
| ARMA(p,q) | Exponential decay | Exponential decay |

Source: Gujarati et al. (2012)

According to Gujarati et al. (2012) ACFs and PACFs of AR(p) and MA(q) processes have opposite patterns; in the AR(P) case the ACF declines geometrically or exponentially but the PACF cut off after a certain number of lags, whereas the opposite happens to an MA(q). Furthermore Gujarati et al. (2012) mentioned that in practice the estimated ACF and PACF will not exactly match to theoretical ACF and PACF. Therefore ARIMA modeling requires a great deal of skill.

### 3.4.7. Criteria for Model Comparison

Zakria and Muhammad (2009) mentioned that Akaike Information Criteria (AIC) and the Schwartz's Bayesian Criterion (SBC) are used to select the best model which gives the lowest AIC and SBC values after thorough comparison of the fitted models. These two indicates are as follows:

1. Akaike Information Criteria: $AIC(k) = n\ \ln(\hat{\sigma}^2) + 2k$
2. Schwartz's Bayesian Criterion: $SBC(k) = n\ \ln(\hat{\sigma}^2) + k\ \ln(n)$

### 3.6. Diagnostics of Errors (Assumptions on Error Terms)

After estimation of parameters of the fitted model, and prior to use this model for forecasting some diagnostic checks should be carried out for the random error terms and there are some assumptions to be fulfilled. These assumptions are the error terms should be normal and independent with constant variance. Therefore it is necessary to do the diagnostic test to check whether those assumptions are satisfied by the fitted models or not (Gujarati et al., 2012).

### 3.6.1. Normality of Error Terms

The normality of the error term is tested by using the statistical tests such as Anderson Darling Test, Shapiro-Wilk test, Lilliefors test and Jarque – Bera(JB) test (Cheruiyot, 2015). The most common one is J-B test statistic which is given as:

$$JB = n\left(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2-3)^2}{24}\right)$$

Where, skewness $= \sqrt{b_1} = \frac{m_3}{(m_2)^{3/2}}$

kurtosis $= b_2 = \frac{m_4}{(m_2)^3}$

$m_2$, $m_3$, $m_4$ are the second, third and fourth central moments respectively.

The null hypothesis, $H_o$: normality of the distribution and alternative hypothesis, $H_1$: Distribution not follows a normal distribution. Under $H_o$ JB ~ $\chi_2^2$.

### 3.6.2. Serial Correlation

Breusch Godfrey Serial Correlation Lagrange's Multiplier (LM) test is used to test the serial correlation existed in error series. The null hypothesis ($H_o$) of LM test is that there is no serial correlation of any order (Kwiatkowski et al, 1992). Test statistics of LM test is as follows;

$$LM = \sum_{i=1}^{T} s_t^2 / \hat{\sigma}_i^2$$

Where $S_t = \sum_{i=1}^{t} e_i \qquad t = 1,2,\ldots\ldots\ldots,T$

### 3.6.3. Constant Variance of Errors

White's General Test is used to check constant variance of error terms in time series. The null hypothesis, $H_o$: Homoscedasticity and the alternative hypothesis, $H_1$: Heteroscedasticity.

### 3.6.4. Autocorrelation of Errors

The Ljung and Box Q statistic (Portmanteaus test) is often used to test whether the series is white noise (Adikari & Agrawal, 2013). The null hypothesis, $H_o$: there is no autocorrelation up to order $k$ and test statistics is, $Q = n(n+2)\sum_{j=1}^{k}(n-j)^{-1}\hat{\rho}_j^{\,2}$ .

Under $H_o$ Q~ $\chi^2$ .

### 3.7. Forecasting Types to Developed ARIMA Models

Jha et al. (2013) mentioned that after developed ARIMA models there are two types of techniques available to forecast the future values. One is statistic forecasting and the second one is dynamic forecasting. The statistic forecasting or the simple one step ahead forecast will only forecast a single time period ahead at time. Dynamic forecast in the other hand is used for forecasting for a longer horizon.

### 3.8. Accuracy of the Fitted Model

Once the model is identified the accuracy of the model is tested for both training set as well as the validation set. Several criteria suggested are:

(i) Mean Absolute Percentage Error (MAPE)

$$MAPE = \sum_{t=1}^{n}\left|\frac{e_t}{n}\right| * 100$$

(ii) Mean Absolute Deviation (MAD)

$$MAD = \frac{\sum_{i=1}^{n}|e_t|}{n}$$

(iii) Mean Square Deviation (MSD)

$$MSD = \frac{\sum_{t=1}^{n}e_t^2}{n}$$

Where $e_t$ is the error at time t (Adikari & Agrawal, 2013). Siregar et al. (2016) mentioned that the smaller these measures the better the forecast is made. Many researchers (Siregar, 2016; Smith & Agrawal, 2015; Rani & Raza 2012) used these accuracy measurements comparatively to select the best fitted model among different time series models.

# CHAPTER FOUR

# DEVELOPMENT OF ARIMA MODEL

The annual aging population can be considered as a time series at equal intervals. Therefore in this chapter ARIMA model developed to forecast ageing population is discussed.

## 4.1. Trends of Ageing Population

The temporal variability of ageing population during the study period of 1950 to 2012 is depicted in Figure 4.1.

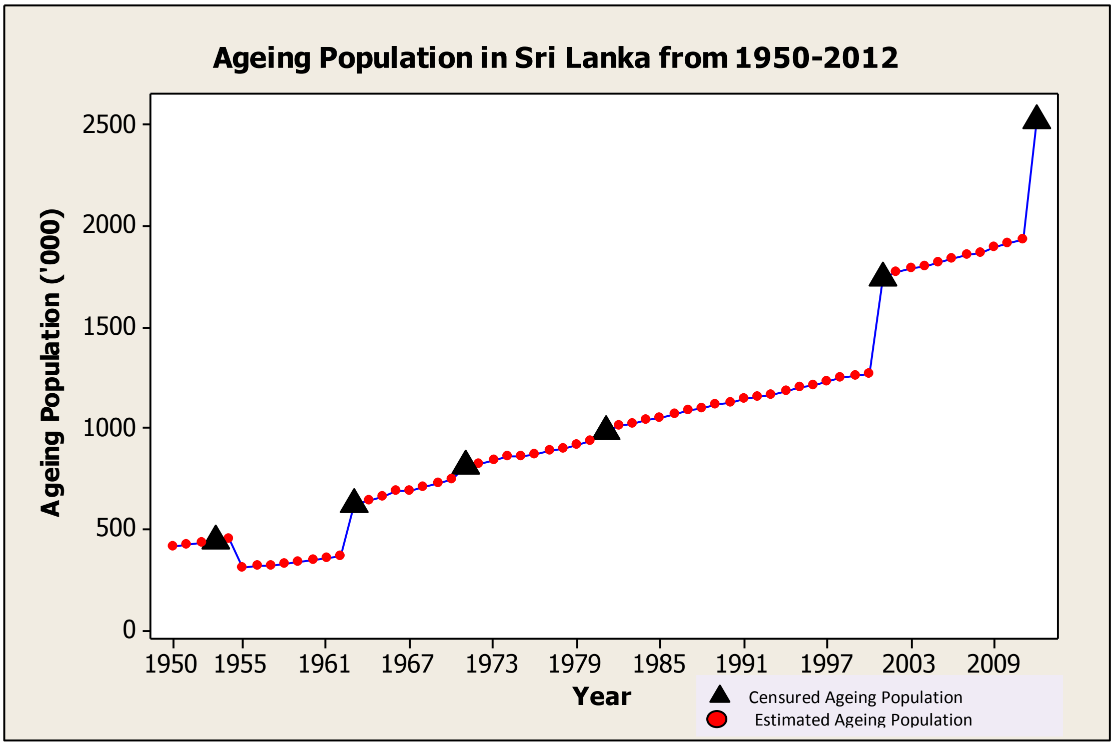



Figure 4.1: Ageing population in Sri Lanka from 1950-2012

The values during 1953, 1963,1971,1981,2001 and 2012 were actual data obtained during census years (1953, 1963,1971,1981,2001 and 2012) and remaining values has been estimated by the Registrar General's Department using cohort component method as described in Chapter two. Norbert (1993) pointed out that if the factors which determine the level of fertility, mortality and migration have not been accurately assessed, the projections have not been correct and that is the disadvantage of cohort component method.

However, as depicted in Figure 4.1 it is clear that there is an increasing trend of ageing population in Sri Lanka from the year 1950 to 2012 except during the period from 1955 to 1961. Furthermore, it can be seen that there is a higher annual increasing rate during 2001 to 2012. As there is increasing temporal trend it is clear that the data series is not stationary. Thus series has to be made stationary using smoothing method before decide ARIMA models.

## 4.2. Stationary of the Series

In developing ARIMA model data series during 1950-2012 was used as a training set and that 2013-2016 was used as a validation set. The plot of ACF of training set is shown in Figure 4.2.

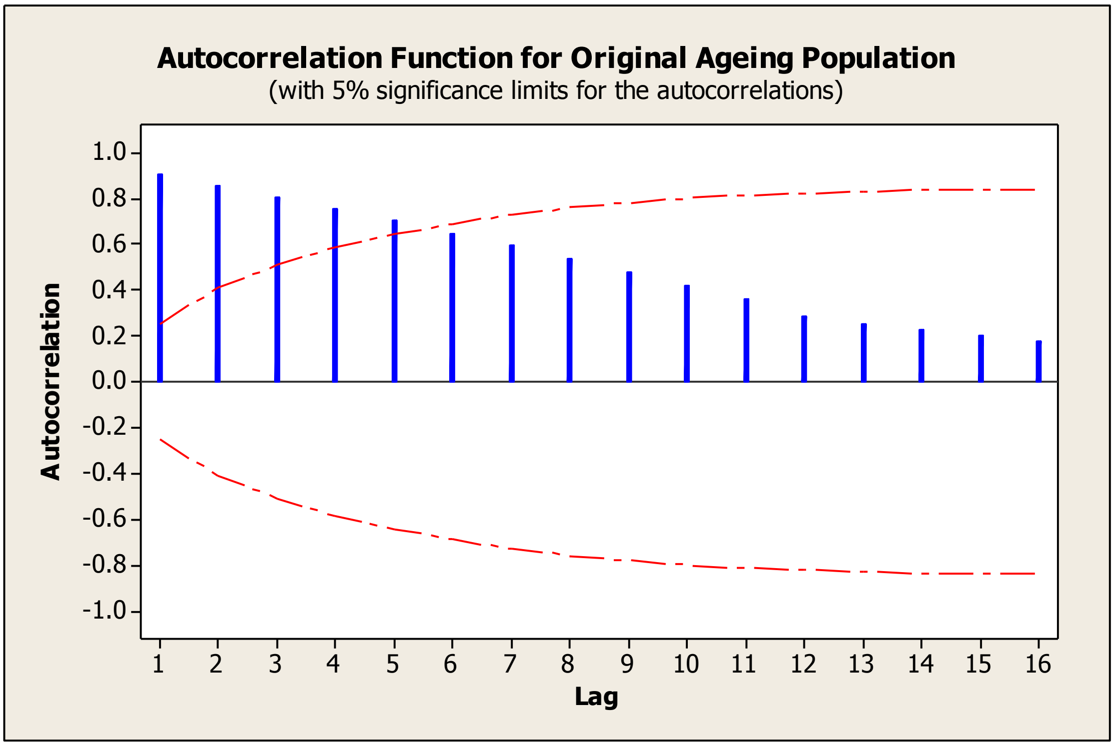



Figure 4.2: Sample ACF of ageing population in Sri Lanka from 1950-2012

According to the Figure 4.2 it shows that all auto correlations are positive also it can be seen that $1^{st}$ $2^{nd}$ $3^{rd}$ $4^{th}$ and $5^{th}$ autocorrelations are significantly different from zero while all other autocorrelations are not significantly different from zero. Therefore it can be concluded that the observed series is not a stationary series.

To confirm the graphical evidence, the Augmented Dickey - Fuller (ADF) test was applied to test the stationary of the series. The output of the ADF test is shown in Table 4.1.

Table 4.1: Results of ADF test of the original ageing population data (1950-2012)

Null Hypothesis: ORIGINAL_AGING has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -0.797412 | 0.9601 |
| Test critical values: | 1% level | -4.113017 | |
| | 5% level | -3.483970 | |
| | 10% level | -3.170071 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ORIGINAL_AGING)
Method: Least Squares
Date: 04/16/17 Time: 11:09
Sample (adjusted): 1951 2012
Included observations: 62 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| ORIGINAL_AGING(-1) | -0.077360 | 0.097014 | -0.797412 | 0.4284 |
| C | 3.577228 | 30.57734 | 0.116990 | 0.9073 |
| @TREND("1950") | 3.416549 | 2.612878 | 1.307581 | 0.1961 |

| | | | |
|---|---|---|---|
| R-squared | 0.074473 | Mean dependent var | 34.06452 |
| Adjusted R-squared | 0.043099 | S.D. dependent var | 100.0183 |
| S.E. of regression | 97.83924 | Akaike info criterion | 12.05171 |
| Sum squared resid | 564778.4 | Schwarz criterion | 12.15463 |
| Log likelihood | -370.6029 | Hannan-Quinn criter. | 12.09212 |
| F-statistic | 2.373738 | Durbin-Watson stat | 1.539664 |
| Prob(F-statistic) | 0.101971 | | |

As per the Table 4.1, it is clear that the original series is not a stationary series, since the p value (0.9601) of the ADF test statistics is greater than 0.05.Thus it can be concluded with 95% confidence that the original series is not stationary.

Smith and Agrawal (2015) mentioned that by taking the difference of the original series it is possible to remove the non – stationarity of a given series. Therefore first difference was taken to convert the series stationary and the results of the ADF test for the first difference series is shown in Table 4.2.

Table 4.2: Results of ADF test of the first differenced ageing population series

Null Hypothesis: D(ORIGINAL_AGING) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -5.763830 | 0.6701 |
| Test critical values: | 1% level | -4.115684 | |
| | 5% level | -3.485218 | |
| | 10% level | -3.170793 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ORIGINAL_AGING,2)
Method: Least Squares
Date: 04/16/17 Time: 11:11
Sample (adjusted): 1952 2012
Included observations: 61 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| D(ORIGINAL_AGING(-1)) | -1.062435 | 0.184328 | -5.763830 | 0.3542 |
| C | -11.15743 | 26.34570 | -0.423501 | 0.6735 |
| @TREND("1950") | 1.474163 | 0.727995 | 2.024964 | 0.0475 |

| | | | |
|---|---|---|---|
| R-squared | 0.374270 | Mean dependent var | 9.524590 |
| Adjusted R-squared | 0.352693 | S.D. dependent var | 123.1470 |
| S.E. of regression | 99.07840 | Akaike info criterion | 12.07763 |
| Sum squared resid | 569358.7 | Schwarz criterion | 12.18144 |
| Log likelihood | -365.3677 | Hannan-Quinn criter. | 12.11832 |
| F-statistic | 17.34587 | Durbin-Watson stat | 1.545166 |
| Prob(F-statistic) | 0.000001 | | |

As per the Table 4.2, it can be concluded that the first differenced series is not a stationary series, since the p value (0.6701) of the ADF test is greater than 0.05. Thus ADF test was further applied for the second differenced series to check the stationary of the series. Results of the ADF test for the corresponding series is shown in Table 4.3.

Table 4.3: Results of ADF test of the second differenced series

Null Hypothesis: D(ORIGINAL_AGING,2) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -9.991966 | 0.0000 |
| Test critical values: | 1% level | -4.118444 | |
| | 5% level | -3.486509 | |
| | 10% level | -3.171541 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ORIGINAL_AGING,3)
Method: Least Squares
Date: 04/16/17 Time: 15:52
Sample (adjusted): 1953 2012
Included observations: 60 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| D(ORIGINAL_AGING(-1),2) | -1.493074 | 0.149427 | -9.991966 | 0.0000 |
| C | -19.09513 | 31.52371 | -0.605739 | 0.5471 |
| @TREND("1950") | 0.887760 | 0.856013 | 1.037087 | 0.3041 |

| | | | |
|---|---|---|---|
| R-squared | 0.639318 | Mean dependent var | 9.483333 |
| Adjusted R-squared | 0.626662 | S.D. dependent var | 187.9310 |
| S.E. of regression | 114.8284 | Akaike info criterion | 12.37346 |
| Sum squared resid | 751577.3 | Schwarz criterion | 12.47818 |
| Log likelihood | -368.2039 | Hannan-Quinn criter. | 12.41442 |
| F-statistic | 50.51689 | Durbin-Watson stat | 1.806059 |
| Prob(F-statistic) | 0.000000 | | |

By comparing observed test statistics (-9.991) with critical values at all three significance levels as shown in Table 4.3, it can be confirmed with 95% confidence that the second differenced series is a stationary series. That is if the observed series is $\{y_i\}$ then $\{y_t - y_{t-1}\} - \{y_{t-1} - y_{t-2}\} = \{1 - 2B + B^2\} y_t$ is stationary, where the operator B is such that $B^p y_t = y_{t-p}$.

## 4.3. ACF of the Stationary Series

The sample ACF of second differenced series is shown in Figure 4.3.

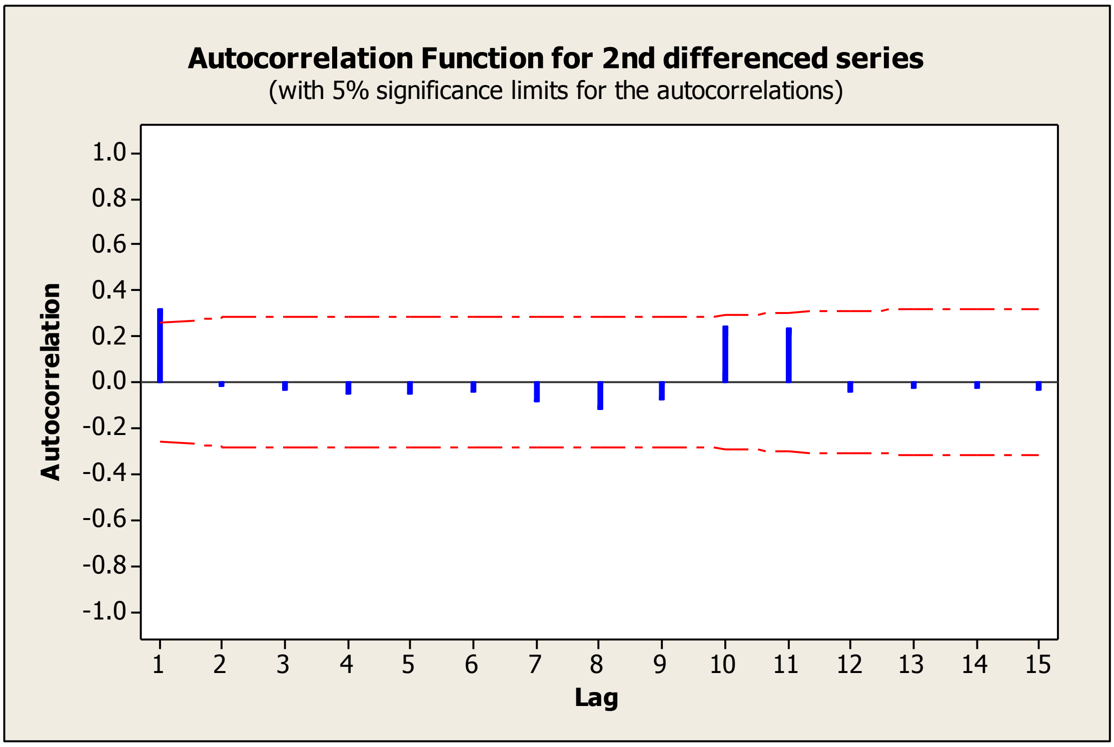


Figure 4.3: Sample ACF of the second differenced series of ageing population

As depicted in the Figure 4.3 the ACF of the second difference series gives evidence of stationary as only the first autocorrelation is significantly different from zero, while all others are not significantly different from zero.

## 4.4 PACF of the Stationary Series

Partial autocorrelation function of the stationary (second differenced) series is shown in Figure 4.4.

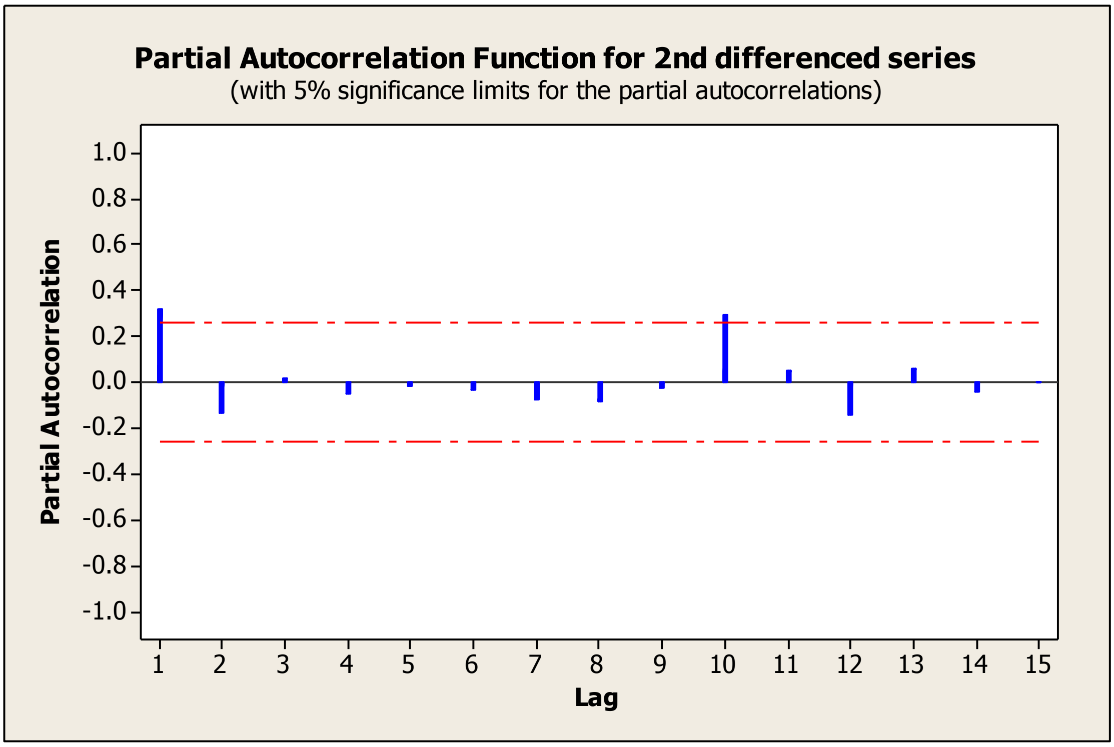



Figure 4.4: Sample PACF of the second differenced series of ageing population

Figure 4.4 clearly shows that the first partial autocorrelation is significant ($p<0.05$). Furthermore, partials at lag 2 to 9 are not significantly difference from zero. However, partial autocorrelation at lag 10 is significant. Such a thing could happen due to the random behavior of data, but physical reason for this phenomenon cannot be explained.

## 4.5. Identification of Parsimonious Models

The order of MA and AR are decided by the significant autocorrelations of the ACF and PACF respectively. For the stationary series it was found that only the 1st ACF (Figure 4.3) and the 1st PACF (Figure 4.4) are significant in the stationary series. In order to decide possible ARIMA models to the stationary series the observed patterns of ACF and PACF of stationary series (Figure 4.3 & Figure 4.4) were compared with the theoretical ACF and PACF of AR(1), MA(1) and ARMA(1,1) are shown in Appendix B.

It can be hypothesized that the stationary series would have come from ARMA (1, 0, 0) or ARMA (0, 0, 1). Furthermore, it was assumed that the observed patterns of ACF and the PACF in the stationary series could have come from ARMA

(1,1) as well. Therefore the following three models: ARIMA (1, 2, 0), ARIMA(0,2,1) and ARIMA(1,2,1) were considered as three possible parsimonious models for the original series $\{y_t\}$. The Parameter estimates of those three parsimonious models are given in the Table 4.4, 4.5 and 4.6 respectively.

Table 4.4: Parameter estimation of ARIMA (1, 2, 0)

Dependent Variable: D(ORIGINAL_AGING,2)
Method: Least Squares
Date: 04/16/17 Time: 16:03
Sample (adjusted): 1953 2012
Included observations: 60 after adjustments
Convergence achieved after 3 iterations

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 6.531204 | 9.949717 | 0.656421 | 0.5141 |
| AR(1) | -0.493939 | 0.149522 | -3.303446 | 0.0016 |

| | | | |
|---|---|---|---|
| R-squared | 0.158356 | Mean dependent var | 9.666667 |
| Adjusted R-squared | 0.143845 | S.D. dependent var | 124.1812 |
| S.E. of regression | 114.9032 | Akaike info criterion | 12.35882 |
| Sum squared resid | 765759.1 | Schwarz criterion | 12.42863 |
| Log likelihood | -368.7647 | Hannan-Quinn criter. | 12.38613 |
| F-statistic | 10.91275 | Durbin-Watson stat | 1.773200 |
| Prob(F-statistic) | 0.001640 | | |

| | |
|---|---|
| Inverted AR Roots | -.49 |

Table 4.5: Parameter estimation of ARIMA (0,2,1)

Dependent Variable: D(ORIGINAL_AGING,2)
Method: Least Squares
Date: 04/16/17 Time: 17:09
Sample (adjusted): 1952 2012
Included observations: 61 after adjustments
Convergence achieved after 11 iterations
MA Backcast: 1951

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.575016 | 1.133259 | 1.389812 | 0.1698 |
| MA(1) | -0.945720 | 0.030814 | -30.69095 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.343003 | Mean dependent var | 9.524590 |
| Adjusted R-squared | 0.331867 | S.D. dependent var | 123.1470 |
| S.E. of regression | 100.6596 | Akaike info criterion | 12.09360 |
| Sum squared resid | 597809.1 | Schwarz criterion | 12.16281 |
| Log likelihood | -366.8549 | Hannan-Quinn criter. | 12.12073 |
| F-statistic | 30.80251 | Durbin-Watson stat | 1.587498 |
| Prob(F-statistic) | 0.000001 | | |

| | |
|---|---|
| Inverted MA Roots | .95 |

Table 4.6: Parameter estimation of ARIMA (1,2,1)

Dependent Variable: D(ORIGINAL_AGING,2)
Method: Least Squares
Date: 04/16/17 Time: 17:03
Sample (adjusted): 1953 2012
Included observations: 60 after adjustments
Convergence achieved after 18 iterations
MA Backcast: 1952

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.550648 | 1.110176 | 1.396759 | 0.1679 |
| AR(1) | -0.068181 | 0.186779 | -0.365037 | 0.7164 |
| MA(1) | -0.944866 | 0.031988 | -29.53824 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.344608 | Mean dependent var | 9.666667 |
| Adjusted R-squared | 0.321612 | S.D. dependent var | 124.1812 |
| S.E. of regression | 102.2810 | Akaike info criterion | 12.14203 |
| Sum squared resid | 596300.2 | Schwarz criterion | 12.24675 |
| Log likelihood | -361.2610 | Hannan-Quinn criter. | 12.18299 |
| F-statistic | 14.98542 | Durbin-Watson stat | 1.526766 |
| Prob(F-statistic) | 0.000006 | | |

| | |
|---|---|
| Inverted AR Roots | -.07 |
| Inverted MA Roots | .94 |

According to the Table 4.4 it is clear that the parameter of AR of the ARIMA (1, 2, 0) model is significantly different from zero and the constant term is not significantly different from zero. As per the Table 4.5 it is clear that the parameters of MA of the ARIMA (0, 2, 1) model is significantly different from zero but the constant term is not significantly different from zero. According Table 4.6 it is clear that the parameter of AR term and the constant term of the ARIMA (1, 2, 1) model are not significantly different from zero, but MA term is significantly different from zero.

Therefore based on the significance of the parameters ( Table 4.4, 4.5, and 4.6) ARIMA(1,2,1) was rejected and ARIMA(1,2,0) and ARIMA(0,2,1) were considered for further comparison.

### 4.6. Comparison of Significant Parsimonious Models

Among ARIMA (1,2,0) model and ARIMA(0,1,2) model, to identify the best model several diagnostic tests were carried out. Table 4.7 shows the comparison of residuals of the two models at lags 12, 24, 36 and 48.

Table 4.7: Comparison of randomness of the errors of the parsimonious models

| **Model** | **Box-Pierce (Q) Statistics at lag** | | | |
|---|---|---|---|---|
| | **12** | **24** | **36** | **48** |
| ARIMA (1,2,0) | 15.8<br>P = 0.104 | 16<br>P = 0.818 | 17.2<br>P = 0.993 | 34.7<br>P = 0.993 |
| ARIMA (0,2,1) | 14.5<br>P = 0.150 | 15.1<br>P = 0.857 | 16.5<br>P = 0.995 | 25.8<br>P = 0.993 |

According to the Box-Pierce (Q) statistics as in the Table 4.7 the errors at lag 12, 24, 36 and 48 are not significant in both models which leads to the conclusion that the autocorrelations up to lag 48 are not significantly different from zero. Thus it can be concluded the errors are randomly distributed in both models. Furthermore, when carefully observing the p values of Box-Pierce (Q) statistics it's clear that the p values of ARIMA (0, 2,1) model is higher than that of the ARIMA (1,2,0) model at all lags. It indicates that the change at rejecting $H_o$: $\rho_1 = \rho_2 = ....... = \rho_{48}$ is more in the model ARIMA(0,2,1) than that in the model ARIMA(1,2,0).Therefore it can be considered the ARIMA (0, 2,1) model as the better model than the ARIMA(1,2,0) model.

More diagnostic test related to residuals were carried out to further confirm the selected model. The corresponding results are shown in table 4.8.

Table 4.8: Further diagnostic checks

| **Model** | **AIC** | **SBC** | **p-values of LM test** | **MSE** | **AD Test** |
|---|---|---|---|---|---|
| ARIMA (1,2,0) | 12.36 | 12.43 | 0.07 | 12979 | 0.005 |
| ARIMA (0,2,1) | 12.09 | 12.16 | 0.87 | 9950 | 0.005 |

The results in Table 4.8 also shows that there is no serial correlation in the ARIMA (1,2,0) model and ARIMA (0,2,1) since the Breusch-Godfrey Serial Correlation Lagrange Multiplier (LM) test statistics are not significant at 5% level, but it is significant at 10% for the model ARIMA (1,2,0).Therefore it leads not to reject $H_o$: there is no serial correlation in the error series and ARIMA (0,2,1) can be considered as the best fitted model out of the two models selected. Furthermore, among the two significant models there is a large MSE in ARIMA(1,2,0) model compared with the ARIMA(0,2,1) model. Based on the lower MSE too it can be confirmed that the ARIMA (0, 2, 1) model is the best model.

As per the Table 4.8 lowest Akaike Information Criteria (AIC) and lowest Schwartz's Bayesian Criterion (SBC) were observed for the ARIMA (0,2,1) model. This criteria too suggested that ARIMA (0,2,1) model as the best model out of the two model selected.

According to Appendix C, The p values of Q statistics of ARIMA (0, 2,1) model are greater than 0.05. Thus it can be concluded that there is no autocorrelation of residuals at any lag which further confirmed the randomness. In other words error terms of the ARIMA (0, 2, 1) model is distributed independently.

However it was noted that AD test as shown in Table 4.8 confirmed that the, assumption of the normality of error term is violating in both models. Finally, it can be concluded that ARIMA (0,2,1) is the best fitted model for the original series of annual ageing population.

## 4.7. ARIMA Model

The 2nd difference series of $\{ Y_t \}$ is given by

$Z_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - 2B + B^2)\, y_t$ where B is an operator such that $B^p y_t = y_{t-p}$

The fitted model $(1 - 2B + B^2) y_t = (1 - \theta_1) e_t$ can be written as;

$$Y_t - 2Y_{t-1} + Y_{t-2} = e_t - \hat{\theta}_1 e_{t-1} + \delta$$

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t - \hat{\theta}_1 e_{t-1} + \delta$$

Thus using the results in Table 4.5

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t + 0.9457 e_{t-1} + 1.575$$

## 4.8. Testing Model for an Independent Set

The percentage errors of the ARIMA (0, 2,1) model for the validation data set were computed (Table 4.9).

Table 4.9: Percentage errors for the validation data set

| **Year** | **2013** | **2014** | **2015** | **2016** |
|---|---|---|---|---|
| Estimated ageing population (in '000) | 2599 | 2678 | 2758 | 2840 |
| Actual ageing population (in '000) | 2548 | 2571 | 2593 | 2623 |
| % Error | -2.00% | -4.16% | -6.36% | -8.27% |

As per the Table 4.9 it can be seen that percentage errors are below $\pm 10\%$ for all points, but there is an increasing trend of percentage errors of the validation dataset with time. To check the accuracy of the fitted ARIMA (0, 2, 1) model, MAPE for the training set and validation set separately were compared and given in the Table 4.10.

Table 4.10: Accuracy of the fitted model

| **Accuracy Measurement** | **MAPE** | **MAD** | **MSD** |
|---|---|---|---|
| Training set | 4.5% | 43.49 | 9624 |
| Validation set | 5.2% | 135 | 22091 |

According to the Table 4.10 the MAPE for the training dataset and the validation dataset is 4.5% and 5.2% respectively. Based on the above results ARIMA (0,2,1) can be considered as a more appropriate model.

## 4.9. Forecasted Ageing Population in Sri Lanka from 2017-2020

Using the fitted ARIMA (0, 2, 1) model ageing population in Sri Lanka for the years 2017, 2018, 2019 and 2020 were predicted (Table 4.11).

Table 4.11: Forecasted ageing population in Sri Lanka from 2017-2020

| **Year** | **Forecasted ageing population (in '000)** |
|---|---|
| 2017 | 2,923 |
| 2018 | 3,008 |
| 2019 | 3,094 |
| 2020 | 3,181 |

As per the Table 4.11 the increasing trend of ageing population will be continue in the future. There will be approximately 3,181,000 ageing population in the year 2020 according to the estimations derived through ARIMA (0, 2,1).

## 4.10. Comparison Forecasted and Actual Ageing Population with Relevant Confidence Intervals

Cohen (1986) recommended the estimation of confidence interval using the fitted model due to the uncertainty not only to the point forecasts of future population, but also to the estimates of those forecasts' uncertainty. Therefore 95% confidence intervals for the forecast values during 2013 to 2016 were computed (Figure 4.5) based on the developed ARIMA model.

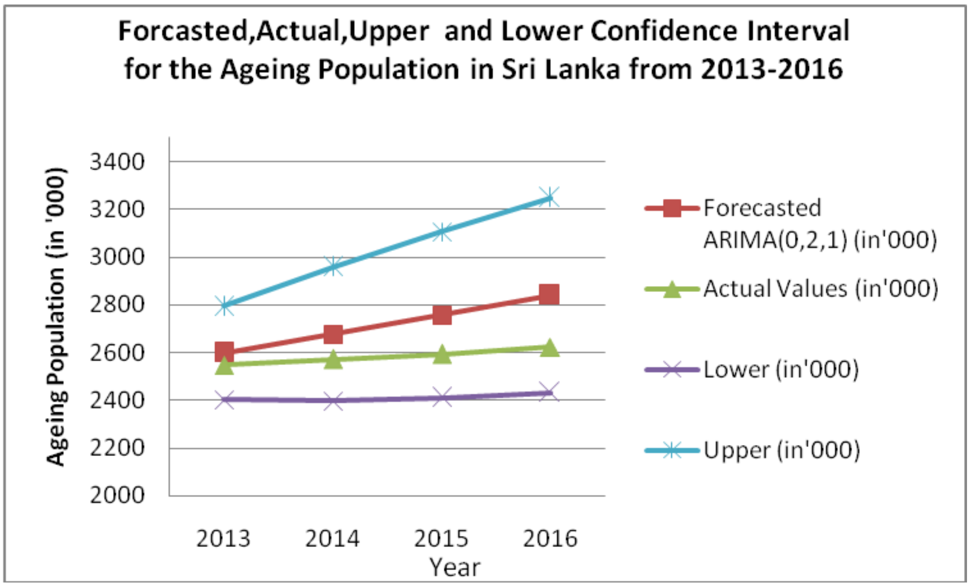



Figure 4.5: Comparison of forecasted and actual ageing population in Sri Lanka with related confidence intervals from 2013-2016

According to the Figure 4.5 it is clear that the actual aging population from year 2013 to 2016 (validation set) is close to the predictions made through ARIMA(0,2,1) model and has the similar trend.

## 4.11. Summary of Chapter Four

ARIMA(0,2,1) was identified as the best fitted ARIMA model to forecast annual ageing population in Sri Lanka. The model was developed using data from 1950-2012 and validated using data from 2013-2016.The residuals of the fitted model was tested using various diagnostic tests and confirmed all assumptions (constant variance, and zero mean (in Appendix c)) are valid except normality (Table 4.8). Slight deviations from normality does not effect seriously on the model (Ghasemi & Zahediasl, 2012). The percentage errors for the validation dataset is very low and varied from 2% to 8%. The best fitted model is,

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t + 0.945e_{t-1} + 1.575$$

The short term predictions were calculated for the period of 2017 to 2020, and the expected ageing population in Sri Lanka in 2020 will be approximately 3,181,000.

# CHAPTER FIVE

# GROWTH MODEL FOR AGEING POPULATION

In this chapter growth model developed for the ageing population in Sri Lanka is discussed. The model was trained using data from 1950 to 2012 (63 years) and it was validated for the period of 2013-2016.

## 5.1. Growth Model

Chandan et al. (2011) mentioned that it is possible to forecast the future values based on the trend of a given data series. Therefore growth model was used to capture the trend of the series and to forecast the future ageing population. Based on the pattern in Figure 4.1 it is very clear that a linear trend is not suitable. As exponential models are fitted in chapter six, a growth model given in (5.1) was considered as a suitable trend model.

$$Y_t = ab^t \qquad (5.1)$$

To make the series linear the log transformation was applied and then the trend model is given in equation 5.2.

$$\text{Log } Y_t = \text{Log } a + t \text{ Log } b \qquad (5.2)$$

It is now clear that there is linear relationship between log $Y_t$ and t and thus the theory of linear regression model was applied to estimate two parameters log a and log b. Output is given in the Table 5.1 and Table 5.2.

Table 5.1: Analysis of Variance of the fitted model

| **Source** | **DF** | **Sum of Square** | **Mean Square** | **F** | **P** |
|---|---|---|---|---|---|
| Regression | 1 | 17.556 | 17.56 | 671.716 | 0.000 |
| Residuals | 61 | 1.594 | 0.026 | | |
| Total | 62 | 19.150 | | | |

$R^2$ = 91.7%, $R^2$ (adj.) = 91.5%

The results in Table 5.1 (p = 0.00, F = 671.716) indicate that the fitted model is significant and able to explain 91% of the variability of log series of the observed values.

Table 5.2: Parameter estimation of the fitted trend model

| | **Unstandardized Coefficients** | | **T Value** | **Sig. of the t-statistics** |
|---|---|---|---|---|
| | Coefficient | Std. Error | | |
| Log a | 5.8603 | 0.0412 | 142.15 | 0.000 |
| Log b | 0.0290 | 0.0011 | 25.92 | 0.000 |

As per the Table 5.2 the two coefficients of the fitted growth model are significantly different from zero since the p values of both the coefficients are less than 0.05. Thus it is further confirmed that the fitted model as well as the two parameters are significantly different from zero. Using the estimated values of the parameters in the model (Table 5.2) a and b were estimated as; $\hat{a} = 350.829$ and $\hat{b} = 1.029$ respectively.

Then the model for the original series can be written as in the equation 5.3:

$$Y = 350.833*(1.029)^t \qquad (5.3)$$

t = 1, 2,…………….., 62 (where, when t=1 year 1950 and when t=62 year 2012)

Same model can be displayed with the log transformation as shown in the equation 5.4:

$$\text{Log } Y = \text{Log } 350.833 + t \log 1.029 \qquad (5.4)$$

Graphical representation of the fitted growth model of the ageing population is shown in Figure 5.1.

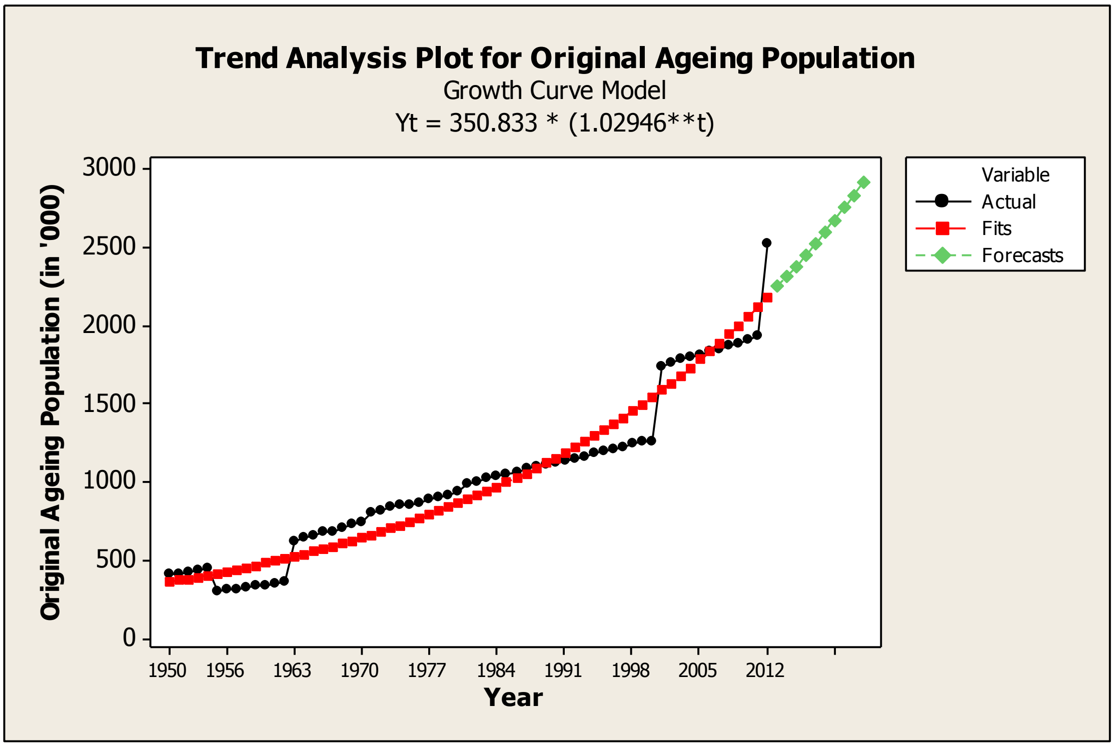



Figure 5.1: Graphical representation of the fitted growth model for the ageing population in Sri Lanka

According to the Figure 5.1 it is clear that the growth curve is nearly close to the actual ageing population.

## 5.2. Comparison of the Percentage Errors

It has been taken the percentage errors through the developed growth model for the validation dataset and the output is given in the Table 5.3.

Table 5.3: Percentage errors for the validation dataset

| **Year** | **2013** | **2014** | **2015** | **2016** |
|---|---|---|---|---|
| Estimated ageing population (in '000) | 2249 | 2315 | 2383 | 2454 |
| Actual ageing population (in '000) | 2548 | 2571 | 2593 | 2623 |
| % Error | 11.7% | 9.9% | 8.1% | 6.5% |

As per the Table 5.3 there is a decreasing trend of percentage errors of the validation set with respect to time.

### 5.3. Accuracy of the Fitted Growth Model

Accuracy of the fitted growth model was tested by taking the accuracy measures (MAPE, MAD, MSD) for the training set and the validation set separately and is shown in Table 5.4.

Table 5.4: Accuracy of the fitted growth model

| Accuracy Measurement | MAPE | MAD | MSD |
|---|---|---|---|
| Training set | 13.6% | 104.66 | 14677 |
| Validation set | 9.1% | 234 | 57049 |

As depicted in the Table 5.4 the MSD of the training set is smaller than the MSD in the validation set. But both the MAPE and MAD are smaller in the validation set comparing with the same two measurements generated from the training set. Higher MSD in validation set indicate that the greater effect of outliers. Since the MAPE of the validation set (9.1%) is less than 10% it can be recommended that the fitted model as the accurate model.

### 5.4. Forecasted Ageing Population in Sri Lanka from 2017-2020 from the Growth Model

Using the fitted growth model it has been forecasted that the ageing population in Sri Lanka from 2017-20120 as shown in Table 5.5.

Table 5.5: Forecasted ageing population in Sri Lanka from 2017-2020

| Year | Forecasted Ageing Population (in '000) | Predicted Confidence Intervals for the Ageing Population (in '000) | |
|---|---|---|---|
| | | Lower | Upper |
| 2017 | 2526 | 1806 | 3533 |
| 2018 | 2600 | 1858 | 3640 |
| 2019 | 2677 | 1912 | 3749 |
| 2020 | 2756 | 1967 | 3861 |

As per the Table 5.5 the forecasted ageing population through the growth model in Sri Lanka from 2017 – 2020 is continuously increasing. According to the estimations there will be approximately 2,756,000 ageing population in Sri Lanka in 2020.

### 5.5. Summary of the Chapter Five

The fitted growth model $Y_t = 350.833*(1.029)^t$ is also considered as suitable model to predict ageing population in Sri Lanka and was able to explain 91% of the temporal variability of the annual ageing population. The model was trained using data from 1950-2012 and validated using data from 2013-2016. The percentage errors varied from 12% to 7% for the validation dataset and the Mean Absolute Percentage Error was 9% for the validation dataset. The expected ageing population in 2020 was estimated as 2,756,000.

# CHAPTER SIX

# DEVELOPMENT OF DOUBLE EXPONENTIAL SMOOTHING MODEL

The two types of time series models namely ARIMA (0,2,1) and growth model: $Y_t = ab^t$ were developed to predict annual ageing population series in the Chapter four and Chapter five respectively. Both models have advantages as well as disadvantages. In this chapter a double exponential smoothing method was applied to develop another forecasting model and then comparison is made among these three models.

## 6.1. Determination of Smoothing Constant

Since it was found that there was an increasing trend of ageing population in Sri Lanka as per the Figure 4.1, double exponential smoothing was applied to smooth the series. According to Siregar (2016), double exponential smoothing method smoothed trend component separately using the two parameters namely $\alpha$ and $\beta$. He further mentioned that double exponential smoothing uses a dynamic trend component that works well for the series having shift in the trend. Smoothing constants are the key to success of exponential smoothing. Therefore prior to applying double exponential smoothing it is necessary to decide those two smoothing constants. Thus in order to select the right constants a simulation study through trail and error method was carried out by changing $\alpha$ starting from 0.063613 to 0.963613 with an increment of 0.1 and that of $\beta$ starting from 0.944218 to 0.044218 with an increment of 0.1. Best combination of $\alpha$ and $\beta$ selected based on the minimum mean absolute percentage error (MAPE). Various authors (Handanhal, 2013; Paul, 2011) have discussed the consequences of two smoothing constants $\alpha$ and $\beta$. In this study larger $\alpha$ and smaller $\beta$ values were taken as initial values for the simulation study. The results are shown in Table 6.1.

Table 6.1: MAPE for different values of smoothing constants (α and β) in DES model

| **Double Exponential Smoothing constant α** | **Double Exponential Smoothing constant β** | **Mean Absolute Percentage Error (MAPE)** | **Double Exponential Smoothing constant α** | **Double Exponential Smoothing constant β** | **Mean Absolute Percentage Error (MAPE)** |
|---|---|---|---|---|---|
| 0.063613 | 0.944218 | 19.5 | 0.563613 | 0.944218 | 07.4 |
| 0.063613 | 0.844218 | 19.0 | 0.563613 | 0.844218 | 07.3 |
| 0.063613 | 0.744218 | 18.3 | 0.563613 | 0.744218 | 07.3 |
| 0.063613 | 0.644218 | 17.5 | 0.563613 | 0.644218 | 07.3 |
| 0.063613 | 0.544218 | 16.5 | 0.563613 | 0.544218 | 07.4 |
| 0.063613 | 0.444218 | 15.5 | 0.563613 | 0.444218 | 07.4 |
| 0.063613 | 0.344218 | 14.6 | 0.563613 | 0.344218 | 07.3 |
| 0.063613 | 0.244218 | 14.0 | 0.563613 | 0.244218 | 07.0 |
| 0.063613 | 0.144218 | 14.1 | 0.563613 | 0.144218 | 06.7 |
| 0.063613 | 0.044218 | 14.3 | 0.563613 | 0.044218 | 06.9 |
| 0.163613 | 0.944218 | 16.4 | 0.663613 | 0.944218 | 06.9 |
| 0.163613 | 0.844218 | 16.3 | 0.663613 | 0.844218 | 06.9 |
| 0.163613 | 0.744218 | 16.0 | 0.663613 | 0.744218 | 6.8 |
| 0.163613 | 0.644218 | 15.6 | 0.663613 | 0.644218 | 6.6 |
| 0.163613 | 0.544218 | 14.9 | 0.663613 | 0.544218 | 6.5 |
| 0.163613 | 0.444218 | 14.3 | 0.663613 | 0.444218 | 6.6 |
| 0.163613 | 0.344218 | 13.6 | 0.663613 | 0.344218 | 6.5 |
| 0.163613 | 0.244218 | 13.1 | 0.663613 | 0.244218 | 6.3 |
| 0.163613 | 0.144218 | 12.6 | 0.663613 | 0.144218 | 06.0 |
| 0.163613 | 0.044218 | 12.3 | 0.663613 | 0.044218 | 06.2 |
| 0.263613 | 0.944218 | 12.7 | 0.763613 | 0.944218 | 06.7 |
| 0.263613 | 0.844218 | 12.5 | 0.763613 | 0.844218 | 06.5 |
| 0.263613 | 0.744218 | 12.2 | 0.763613 | 0.744218 | 06.4 |
| 0.263613 | 0.644218 | 12.2 | 0.763613 | 0.644218 | 06.3 |
| 0.263613 | 0.544218 | 12.1 | 0.763613 | 0.544218 | 06.1 |
| 0.263613 | 0.444218 | 11.9 | 0.763613 | 0.444218 | 06.0 |
| 0.263613 | 0.344218 | 11.4 | 0.763613 | 0.344218 | 05.7 |
| 0.263613 | 0.244218 | 10.8 | 0.763613 | 0.244218 | 05.4 |
| 0.263613 | 0.144218 | 10.6 | 0.763613 | 0.144218 | 05.6 |
| 0.263613 | 0.044218 | 10.4 | 0.763613 | 0.044218 | 05.6 |
| 0.363613 | 0.944218 | 10.1 | 0.863613 | 0.944218 | 06.6 |
| 0.363613 | 0.844218 | 10.2 | 0.863613 | 0.844218 | 06.5 |
| 0.363613 | 0.744218 | 10.2 | 0.863613 | 0.744218 | 06.4 |
| 0.363613 | 0.644218 | 10.2 | 0.863613 | 0.644218 | 06.3 |
| 0.363613 | 0.544218 | 10.0 | 0.863613 | 0.544218 | 06.2 |
| 0.363613 | 0.444218 | 09.7 | 0.863613 | 0.444218 | 06.1 |
| 0.363613 | 0.344218 | 09.5 | 0.863613 | 0.344218 | 05.8 |
| 0.363613 | 0.244218 | 09.2 | 0.863613 | 0.244218 | 05.4 |
| 0.363613 | 0.144218 | 09.0 | 0.863613 | 0.144218 | 05.0 |
| 0.363613 | 0.044218 | 08.9 | 0.863613 | 0.044218 | 05.2 |

Table 6.1 (Continued)

| | | | | | |
|---|---|---|---|---|---|
| 0.463613 | 0.944218 | 08.3 | 0.963613 | 0.944218 | 06.6 |
| 0.463613 | 0.844218 | 08.4 | 0.963613 | 0.844218 | 06.5 |
| 0.463613 | 0.744218 | 08.5 | 0.963613 | 0.744218 | 06.4 |
| 0.463613 | 0.644218 | 08.6 | 0.963613 | 0.644218 | 06.3 |
| 0.463613 | 0.544218 | 08.6 | 0.963613 | 0.544218 | 06.2 |
| 0.463613 | 0.444218 | 08.5 | 0.963613 | 0.444218 | 06.1 |
| 0.463613 | 0.344218 | 08.2 | 0.963613 | 0.344218 | 05.8 |
| 0.463613 | 0.244218 | 07.9 | 0.963613 | 0.244218 | 05.4 |
| 0.463613 | 0.144218 | 07.7 | 0.963613 | 0.144218 | 04.9 |
| 0.463613 | 0.044218 | 07.8 | 0.963613 | 0.044218 | 03.1 |

Based on the results in Table 6.1 it is clear that the best combination is $\alpha = 0.963613$ and $\beta = 0.044218$ where it gets minimum MAPE.

## 6.2. Fitting Double Exponential Smoothing Model

Having decided the two smoothing constants, the initial starting point values and the initial forecast are very important. However, in this case the default option in Minitab software was used. Relevant formulas to calculate the initial value of trend and initial value of level were included in Appendix D. According to calculations it was found that the initial value of level is 409.03 and initial value of trend is 19.24. Therefore the initial forecast for the first observation of the series was taken as 409.03 + 19.24 = 417.27. The double exponential smoothing plot obtained from Minitab is shown in Figure 6.1.

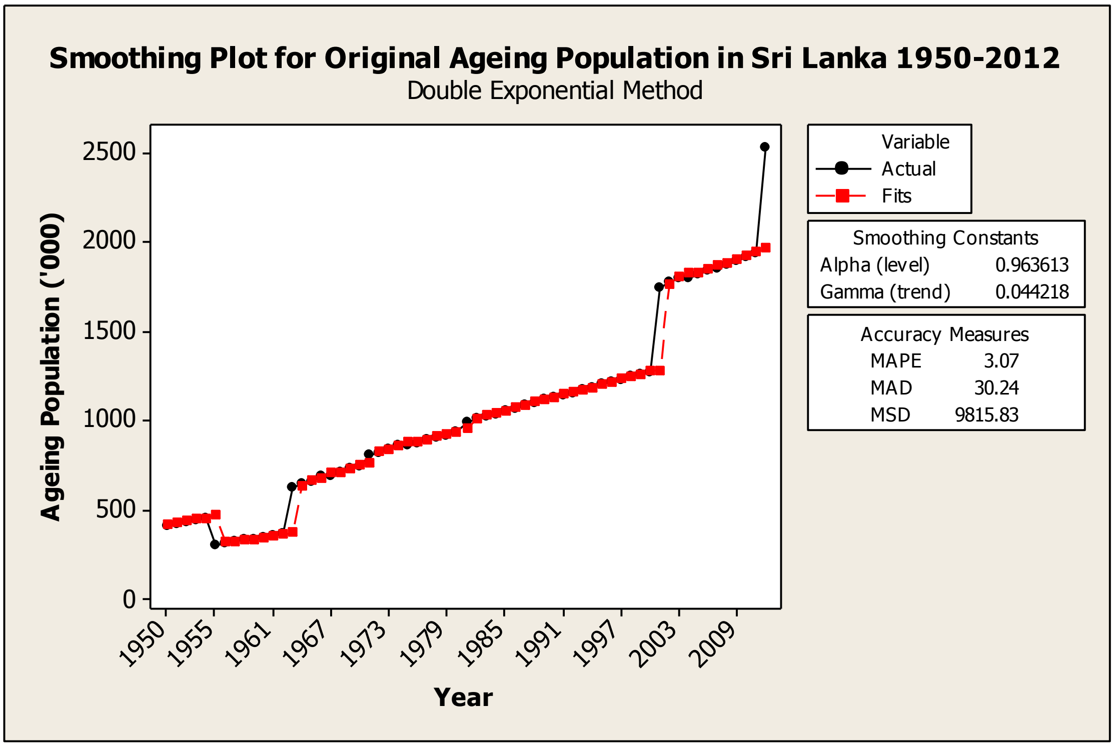



Figure 6.1: Double exponential smoothing curve for the ageing population in Sri Lanka

As per the Figure 6.1 it is clear that fitted double exponential smoothing curve is very much close to the original ageing population data. It was found that correlation coefficient between the actual and the predicted values for the training data set was 0.982 (p = 0.000). As per the Appendix E the percentage errors for the training data set was varied from -3% to 39% with exceptional in the year of 1955 (-53%).

### 6.3. Comparison of the Percentage Errors of the Validation Set

It has been taken the percentage errors through the developed double exponential smoothing model for the validation dataset and the output is given in the Table 6.2.

Table 6.2: Percentage errors for the validation dataset

| **Year** | **2013** | **2014** | **2015** | **2016** |
|---|---|---|---|---|
| Estimated ageing population (in ‘000) | 2555 | 2609 | 2664 | 2718 |
| Actual ageing population (in ‘000) | 2548 | 2571 | 2593 | 2623 |
| % Error | -0.3% | -1.5% | -2.7% | -3.6% |

The percentage error for the independent data set is also very low (less than $\pm 10$), but it can be seen that there is an increasing trend of percentage errors of the validation dataset with respect to time (Table 6.2).

### 6.4. Accuracy of the Fitted Double Exponential Smoothing Model

To confirm the suitability of the double exponential smoothing technique to forecast the future ageing population in Sri Lanka, validation measurements (MAPE, MSD and MAD) has been taken for the training set and the validation set separately and the results are given in the Table 6.3.

Table 6.3: Accuracy of the fitted double exponential smoothing model

| Accuracy Measurement | MAPE | MAD | MSD |
|---|---|---|---|
| Training set | 3.1% | 30.94 | 10135 |
| validation set | 2% | 53 | 3855 |

As per the Table 6.3 the two accuracy measures MAPE and MSD are smaller in validation dataset when comparing with the same measurements in the training set. But the measurement MAD of the training set is smaller than the same measurement of the validation dataset. Nevertheless lower MSD value in validation dataset indicated that there is lower effect of outliers. Since the MAPE value of the validation dataset (2%) is less than 10% and came in to a conclusion that the model developed through the double exponential smoothing method is accurate to forecasting the future ageing population in Sri Lanka.

### 6.5. Short-Term Forecasting of Ageing Population (2017-2020)

Using the developed double exponential smoothing model forecasting was carried out for the ageing population in Sri Lanka for the years 2017, 2018, 2019 and 2020 and the results are shows in Table 6.4.

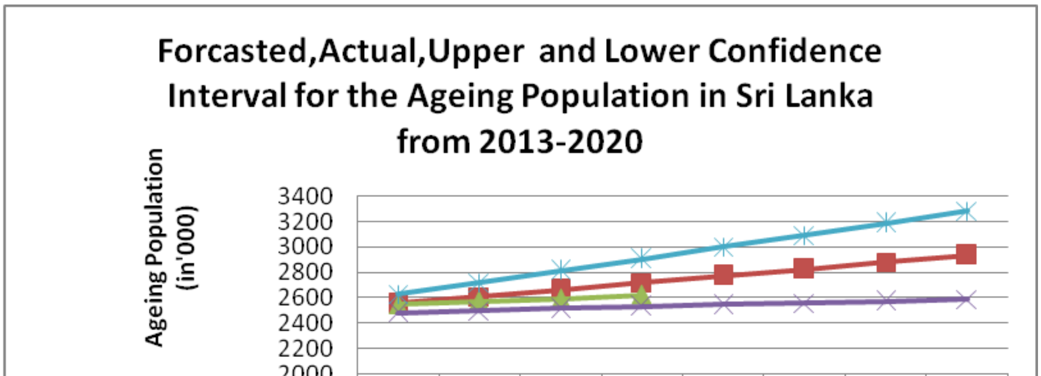
Table 6.4: Forecasted Aging Population in Sri Lanka from 2017-2020

| Year | Forecasted Ageing Population (in ‘000) |
|---|---|
| 2017 | 2,772 |
| 2018 | 2,827 |
| 2019 | 2,881 |
| 2020 | 2,936 |

As per the Table 6.4 the forecasted ageing population in Sri Lanka from 2017 to 2020 is further increasing according to the forecast done through the double exponential smoothing method. There will be approximately 2,936,000 ageing population in Sri Lanka in 2020 according to the estimations.

## 6.6. Confidence Intervals for the Predicted Values (2013-2020)

The confidence intervals for the predicted values 2013-2020 were computed and the corresponding figure is shown in Figure 6.2.




| | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|
| Forecasted throgh Double Exponential Smoothing model (in'000) | 2555 | 2609 | 2664 | 2718 | 2773 | 2827 | 2881 | 2936 |
| Actual Values (in'000) | 2548 | 2571 | 2593 | 2623 | | | | |
| Lower (in'000) | 2481 | 2499 | 2515 | 2530 | 2545 | 2560 | 2574 | 2589 |
| Upper (in'000) | 2629 | 2720 | 2813 | 2906 | 3000 | 3094 | 3188 | 3283 |

Figure 6.2: Forecasted, actual, upper and lower confidence intervals for the ageing population (2013-2020) through DES model

It can be seen that intervals has been widening as time increase. This is due to the fact that as the time moves away from the mean confidence interval used to increase. However, based on the above confidence intervals it can be concluded with 95% confidence that the ageing population in Sri Lanka by 2020 can vary within 2,589,000 and 3,283,000. Similar conclusion can be made for other years as well.

## 6.7. Comparison of Three Models

### 6.7.1. Percentage Errors

The percentage errors for the validation dataset obtained using the three models were compared simultaneously (Table 6.5).

Table 6.5: Comparison of the percentage errors of three models for the validation set

| Model | Ageing Population ('000) | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|
| ARIMA(0,2,1) | Estimated | 2599 | 2678 | 2758 | 2840 |
| | Actual | 2548 | 2571 | 2593 | 2623 |
| | % Error | -2.0% | -4.2% | -6.7% | -8.8% |
| Growth Model | Estimated | 2249 | 2315 | 2383 | 2454 |
| | Actual | 2548 | 2571 | 2593 | 2623 |
| | % Error | 11.7% | 9.9% | 8.1% | 6.5% |
| Double Exponential Smoothing | Estimated | 2555 | 2609 | 2664 | 2718 |
| | Actual | 2548 | 2571 | 2593 | 2623 |
| | % Error | -0.3% | -1.5% | -2.7% | -3.6% |

As per the Table 6.5 it is clear that the percentage errors of the ARIMA (0,2,1) and double exponential smoothing models are increasing with respect to time except the growth model. It can be seen that the percentage errors for all years are the highest for growth model and that are lowest for double exponential smoothing method. Thus in respect to the percentage errors it can be concluded that double exponential smoothing model is better than the other two. Similar results were obtained for the training sets of the three models as in Appendix E, F and G.

### 6.7.2. Accuracy Measures

Siregar et al. (2016) mentioned that smaller the accuracy of measures the better the forecast. The three accuracy measures namely MAPE, MAD and MSD obtained for were compared simultaneously for the training set as well as for the validation set. The corresponding figures are shown in Table 6.6.

Table 6.6: Comparison of the accuracy of the models through training set and validation set

| Type of the dataset | Accuracy measurement | ARIMA(0,2,1) model | Exponential trend model | Double exponential smoothing model |
|---|---|---|---|---|
| Training dataset | MAPE | 4.5% | 13.6% | 3.1% |
| | MAD | 43.49 | 104.66 | 30.94 |
| | MSD | 9624 | 14677 | 10135 |
| Validation dataset | MAPE | 5.2% | 9.1% | 2% |
| | MAD | 135 | 234 | 53 |
| | MSD | 22091 | 57049 | 3855 |

According to the Table 6.6 the two measurements MAPE and the MAD derived from the training set of the double exponential smoothing model are smaller, comparing with the same measurements derived from the two other models. Simultaneously the smaller MSD value of the training set belongs to the ARIMA (0, 2, 1) model. When considered about the validation set, all the three accuracy measurements (MAPE, MAD, and MSD) derived from double exponential smoothing model are comparatively small. Based on these reasons it can be mentioned that the model fitted through the double exponential smoothing method is the best fitted model compared with the other two models. It should be highlighted that MAPE are less than 4% for the double exponential model. In fact, Siregar (2016) mentioned that if MAPE is less than 10% the fitted model is said to be excellent. Therefore with respect to three accuracy measures it can be concluded that double exponential method is the best among the selected three models.

### 6.8. Summary of the Chapter Six

The exponential smoothing is one of the most popular forecasting techniques. It is easy to understand and easy to use. In this Chapter it was shown that double exponential smoothing model with the smoothing constants $\alpha = 0.963613$ and $\beta = 0.044218$ and initial forecast 417.27 was found to be the best model to forecast ageing population in Sri Lanka for short-term forecasting. Those values depend on the training data set and in this study model was trained using data from 1950 to 2012. If one wants to change the period of data then smoothing constant and the use of initial forecast should be determined. It was shown that this model is more statistically sound compared with other two models used in this study. The best feature of this model is that, MAPE for both validation and training dataset is less than 4%. The estimated ageing population in Sri Lanka in 2020 is 2,936,000.

# CHAPTER SEVEN

# CONCLUSIONS AND RECOMMENDATIONS

Based on the results obtained by the statistical analysis of this study the followings conclusions and recommendations can be made.

## 7.1. Conclusions

- The double exponential smoothing method gave better forecast values for ageing population in Sri Lanka during 1950 to 2016. This model is more superior to other two models with respect to the statistical indicators.
- The forecasted ageing populations for 2017 to 2020 are 2.772, 2.827, 2.881 and 2.936 million respectively.
- So far no statistical model is available in Sri Lanka to forecast ageing population. Until a further model is developed this methodology can be used to forecast ageing population in Sri Lanka, as short term basis.
- The forecasted ageing populations up to 2020 would be very useful for policy makers to implement various projects to care elderly persons in Sri Lanka.

## 7.2. Recommendations

- Double exponentially method is generally good where there is a underlying linear trend. However, the choice of two smoothing constants and the choice of initial forecast have a significant impact on the performance of the forecast. Thus further development for the model is required by changing the initial forecast values.
- Due to the continuously rising trend of future ageing population in Sri Lanka, the government or the relevant authorities should cater the needs of the elderly population by rethinking and developing necessary welfare facilities, designing pension schemes, developing long term care systems and health care facilities, along with necessary programmes to uplift the elders’ income.
- As explained in Chapter four there are some drawbacks in estimating annual ageing population for non-census years. A better method need to be developed to estimate ageing population in those years.

## 7.3. Suggestions

- The possibility of neural network models and multivariate time series techniques such as Vector Autoregression or Bayesian Autoregression models are suggested to be investigated. The factors such as mortality, fertility, births, deaths and migrations can be incorporated in those techniques.

## REFERENCE LIST


Abel, J. A., Bijak, J., & Raymer, J. (2010). A comparison of official population projections with Bayesian time series forecast for England and Wales. *Centre for Population Change Working Paper* 7/2010 (ISSN2042-4116). Retrieved from http://eprints.soton.ac.uk/160957/1/Working_paper_7.pdf

Aberathna, W., Alles, L., Wickremasinghe, W. N., & Hewapathirana, I. (2014). Modeling and forecasting mortality in Sri Lanka. *Sri Lanka Journal of Applied Statistics*, 15(3), 141-170. Retrieved from http:// sljastats.sljol.info/ articles/?app=&order=&f=428&f=28f=332&page=2

Adhikar, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. Retrieved from https://gejza.nipax.cz/_media/stochasticke_procesy:1302.6613.pdf

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3): 547–581. doi: 10.1016/j.ijforecast.2006.04.001

Chandan, J. S., Singh, J., & Khanna, K. K. (2011). *Business Statistics*, (2nd revised ed.). Delhi: Vikas Publishing House Pvt. Ltd.

Cheruiyot, K. J. (2015). Analysis and forecasting of university students population using Autoregressive Integrated Moving Average model. *Journal of Mathematical Theory and Modeling*, 5(13), ISSN 2224-5804

Cohen, J. E. (1986). Population forecast and confidence intervals for Sweden: a comparison of model - based and empirical approaches. *Journal of Demography*, 23 (1), 105-126. doi:10.2307/2061412

Department of Census and Statistics, Sri Lanka (1997, 1992, 1986, 1979). *Statistical Abstracts of the Demographic Socialist Republic of Sri Lanka*, Department of Government Press, Colombo

Department of Census and Statistics, Sri Lanka (1955, 1960, 1962, 1963, 1969). *Statistical Abstracts of Ceylon*, Government Press Ceylon

Department of Census and Statistics, Sri Lanka (2015a). *Census of Population and Housing 2012*, ISBN:978-955-577-940-1. Retrieved from http://www.statistics.gov.lk/PopHouSat/CPH2011/Pages/Activities/Reports/FinalReport/FinalReportE.pdf

Department of Census and Statistics, Sri Lanka (2015b). *Increasing elderly population in Sri Lanka*, ISBN:978-955-577-941-8. Retrieved from http://www.statistics.gov.lk/samplesurvey/Thematic01.pdf

De Silva, W. I. (1994). How serious is ageing in Sri Lanka and what can be done about it?. *Asia-Pacific Population Journal*, *9*(1), 19-36.

De Silva, W. I. (2007). A population projection of Sri Lanka for the new millennium 2001-2101 trends and implications. Institute for health policy, Colombo.

Gavrilov, L. A., & Heuveline, P. (2003). The aging of population. *Manuscript of article in the encyclopedia of population*, Retrieved from http://www.galegroup.com/servlet/ItemDetailServlet?region=9&imprint=000&titleCode=M333&type=4&id=174029

Ghasemi, A., & Zahediasl, S. (2012). Normality test for statistical analysis: a guide for non statisticians. *International Journal of Endocrinology & Metabolism*, 10(2): 486-489. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/

Gujarati, D. N., Porter, D. C., Gunasekar, S. (2012). *Basic Econometrics* (4th ed.). New Delhi: Tata McGraw Hill Education Private Ltd.

Handanhal, V. R., (2013). Determining the optimal values of exponential smoothing constants – Does solver really work?. *American Journal of Business Education*, 6(3). Retrieved from http://files.eric.ed.gov/fulltext/EJ1054363.pdf

Jha, K., Sinha, N., Arkatkar, S. S. & Sarkar, A. K. (2013). Modeling growth trend and forecasting techniques for vehicular population in India. *International Journal for Traffic and Transport Engineering,* 3(2), 139-158. Retrieved from http://dx.doi.org/10.7708/ijtte

Keilman, N., Pham, D. Q., & Hetland, A. (2002). Why population forecast should be probabilistic - illustrated by the case of Norway. *Online Journal of Demographic Research*, 6(15), Retrieved from https://ideas.repec.org/a/dem/demres/v6y2002i15.html

Kwiatkowski, D., Philip, P. C.B., Schmidt, P., & Shin, Y. (1992).Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics*, 54(1992) 159-178. Retrieved from https://www.journals.elsevier.com/journal-of-econometrics/

Li, O., Reuser, M., Kraus, C., & Alho, J. (2009). Aging of a giant: a stochastic population forecast for China 2006-2060. *Journal of Population Research*, 26, 21-50. doi:10.1007/s12546-008-9004-z

Menike, H. R. A. (2014a). Age structure transition and demographic bonus in Sri Lanka. *International Journal of Social Sciences and Entrepreneurship, 1* (12), 421-431.

Menike, H. R. A. (2014b). Important features of the elderly population in Sri Lanka. *Research Process,* 2(2), 29-38.

Minitab Inc. (2010). Minitab technical support document, Retrieved from https://it.minitab.com/enus/support/documentation/Answers/DoubleExpSmoothTrendLevel.pdf

Nagarajan, R., Teixeria, A. A. C., & Silva, S. (2013). The impact of population ageing on economic growth: an in- depth bibliometric analysis. *FEP Working Papers*, ISSN: 0870-8541. Retrieved from http://wps.fep.up.pt/wps/wp505.pdf

Norbert, S. A. (1993). Population growth in Sri Lanka: A trend projection analysis. *Economic Review*. Retrieved from http://dl.nsf.ac.lk/bitstream/handle/1/13989/ER-18 (11)-11.pdf? sequence=2

Paul, S. K., (2011). Determination of exponential smoothing constant to minimize mean square error and mean absolute deviation. *International Research Journal*,11(3). ISSN: 0975-5861. Retrieved from http://globaljournals.org/GJRE_Volume11/7-Determination-of-Exponential-Smoothing-Constant-to.pdf

Prasannath, V. (2014). Trends in aging population and its consequences in Sri Lanka. *Proceedings of second Annual Research Conference, Faculty of Commerce and Management, South Eastern University of Sri Lanka*, 100-104. Retrieved from http://ir.lib.seu.ac.lk/bitstream/handle/123456789/720/12%20Pages%20100-105%20Proceeding%202014.03.14%20(Final%20Version)12.pdf?sequence=1

Rani, S., & Raza, I. (2012). Comparison of trend analysis and double exponential smoothing method for price estimation of major pulses in Pakistan. *Pakistan Journal of Agricultural Research*, 25(3): 2012. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=PK2013000223

Registrar's General Department. Age and sex specific midyear population sizes of Sri Lanka - 1992-2016.

Sekaran, U., Bougie, R. (2010). *Research methods for business: A skill building approach* (5th ed.). West Sussex, UK: John Wiley & Sons Ltd Publications.

Siddhisena, K. A. P. (2005). Social- economic implications of aging in Sri Lanka: an overview, *Oxford Institute of Aging Working Papers*, 105. Retrieved from *http://www.ageing.ox.ac.uk/download/98*

Siregar, B., Butar, I. A. B., Rahmat, R. F., Andayani, U., & Fahmi, F. (2016). Comparison of exponential smoothing methods in forecasting palm oil real production. *Journal of Physics*, 801 (2017). doi:10.1088/1742-6596/801/1/012004

Smith, M., & Agrawal, R. (2015). A Comparison of time series model forecasting methods on patent groups, Retrieved from http://ceur-ws.org/Vol-1353/paper_13.pdf

Stanbery, V. B. (1952). *Population forecasting methods a report on forecasting and estimation methods*. United States Department of Transportation. Retrieved from https://www.fhwa.dot.gov/planning/tmip/publications/other_reports/population_forecasting_methods/pfm.pdf

Tulsian, P. C., Jhunjhnuwala, B. (2010). Business Statistics, (2nd ed.). New Delhi: S. Chand & Company Ltd.

Vodopivec, M., & Arunatilake, N. (2008). The impact of population aging on the labour market: the case of Sri Lanka, *IZA Discussion Paper* No.3456. Retrieved from http://ftp.iza.org/dp3456.pdf

Wei, H., Jian, Y., & Zhang, Y. (2015). A review of the two population growth models and an analysis of factors affecting the Chinese population growth. *Asian Journal of Economics Modeling*, 3(1):8-20. ISSN(e):2312-3656/ISSN(p):2313-2884. Retrieved from http://www.aessweb.com/journals/5009

World Bank. (2008). *Sri Lanka Addressing the needs of an Aging Population*, Human Development Unit South Asian Region, Report No.43396 LK

Zakria, M., & Muhammad, F. (2009). Forecasting the population of Pakistan using ARIMA models. Pakistan Journal of Agricultural Sciences, 46 (3), 214-223, ISSN 2076-0906. Retrieved from http://pakjas.com.pk/papers/97.pdf

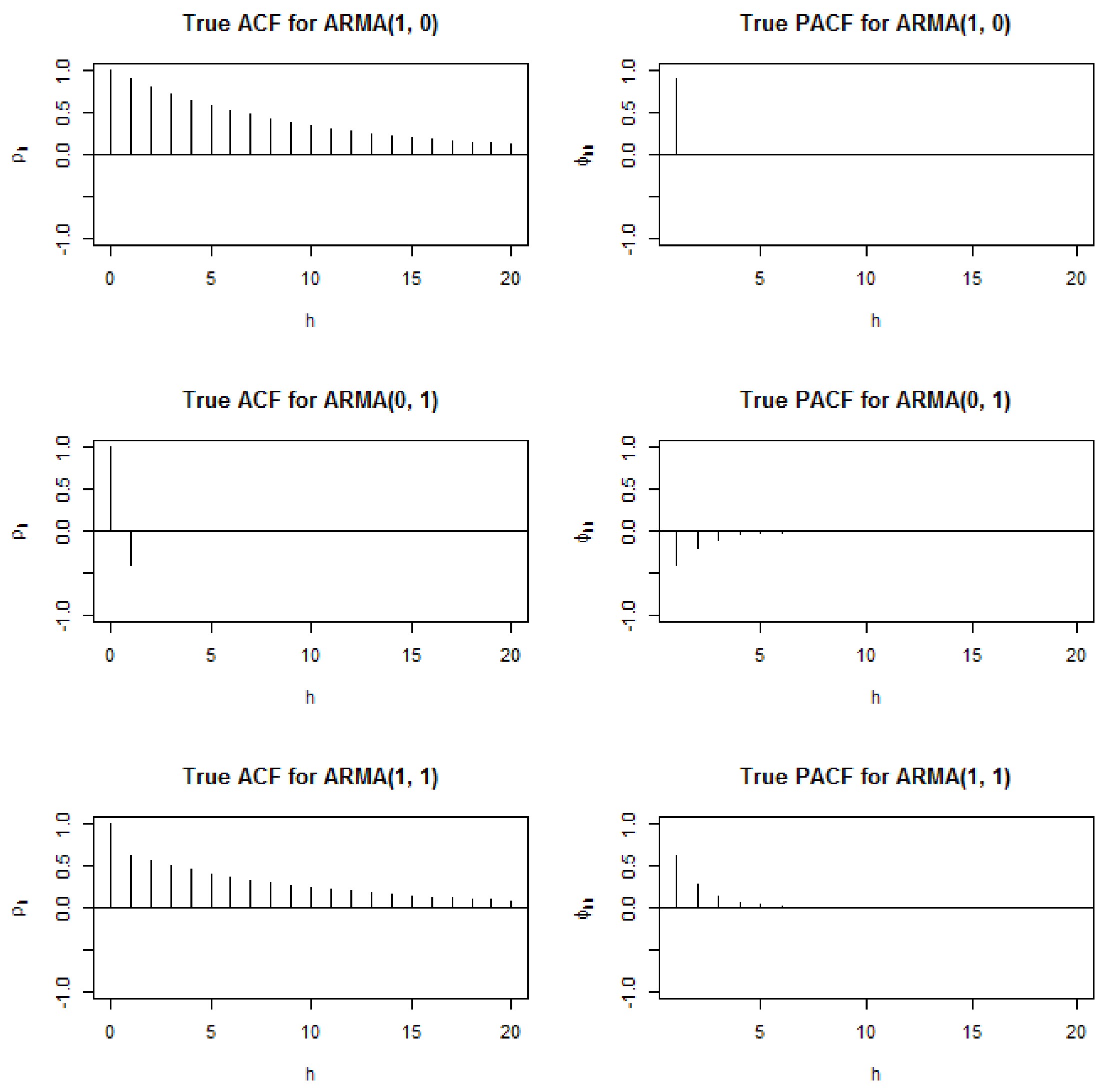
**BIBLIOGRAPHY**

United Nations. (2001).World population prospects: the 2000 revision. New York Nations

Appendix – A: Ageing population dataset

| Year | Aging Population (in '000) | Year | Aging Population (in '000) | Year | Aging Population (in '000) |
|---|---|---|---|---|---|
| 1950 | 409 | 1972 | 817 | 1994 | 1184 |
| 1951 | 419 | 1973 | 841 | 1995 | 1199 |
| 1952 | 430 | 1974 | 856 | 1996 | 1212 |
| 1953 | 437 | 1975 | 856 | 1997 | 1226 |
| 1954 | 453 | 1976 | 872 | 1998 | 1244 |
| 1955 | 305 | 1977 | 888 | 1999 | 1258 |
| 1956 | 313 | 1978 | 900 | 2000 | 1264 |
| 1957 | 319 | 1979 | 917 | 2001 | 1738 |
| 1958 | 327 | 1980 | 936 | 2002 | 1770 |
| 1959 | 337 | 1981 | 986 | 2003 | 1793 |
| 1960 | 346 | 1982 | 1008 | 2004 | 1795 |
| 1961 | 354 | 1983 | 1022 | 2005 | 1814 |
| 1962 | 365 | 1984 | 1036 | 2006 | 1834 |
| 1963 | 621 | 1985 | 1052 | 2007 | 1852 |
| 1964 | 643 | 1986 | 1067 | 2008 | 1870 |
| 1965 | 658 | 1987 | 1083 | 2009 | 1893 |
| 1966 | 686 | 1988 | 1099 | 2010 | 1909 |
| 1967 | 690 | 1989 | 1113 | 2011 | 1930 |
| 1968 | 706 | 1990 | 1129 | 2012 | 2521 |
| 1969 | 728 | 1991 | 1141 | 2013 | 2548 |
| 1970 | 743 | 1992 | 1152 | 2014 | 2571 |
| 1971 | 807 | 1993 | 1167 | 2015 | 2593 |
| | | | | 2016 | 2623 |

Appendix – B: Theoretical autocorrelation and partial autocorrelation plots for AR(1), MA(1) and ARMA(1,1)



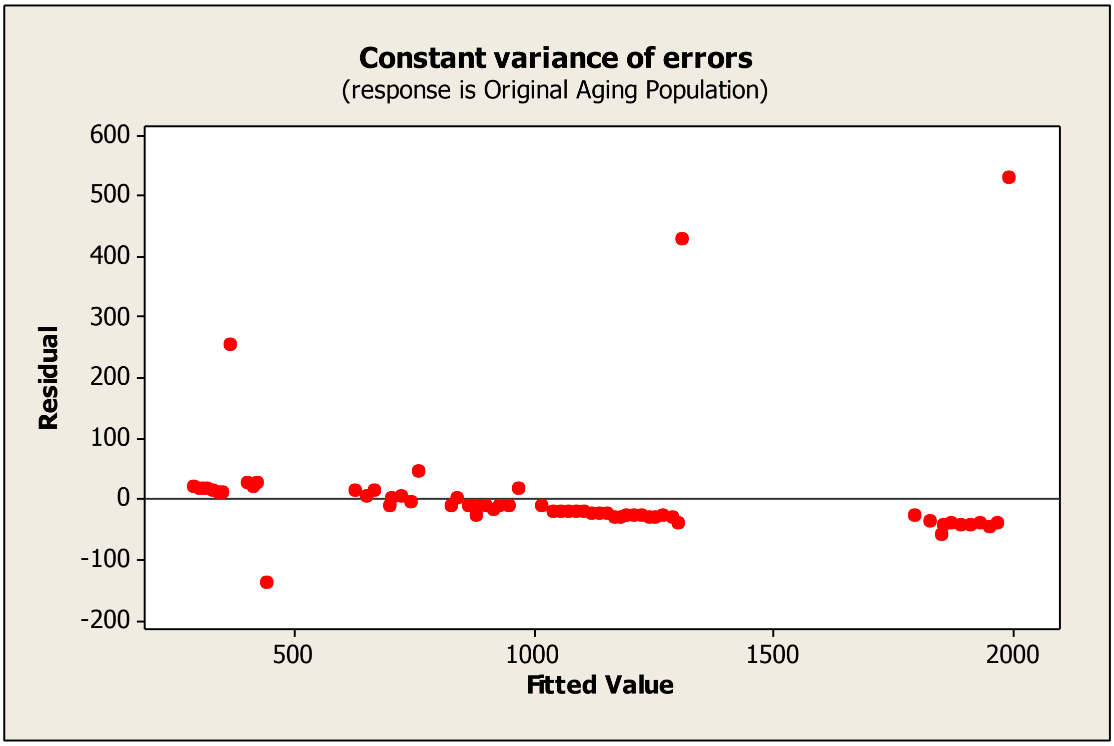
Appendix – C: Diagnostic checks of residuals of ARIMA(0,2,1) model

Q statistics

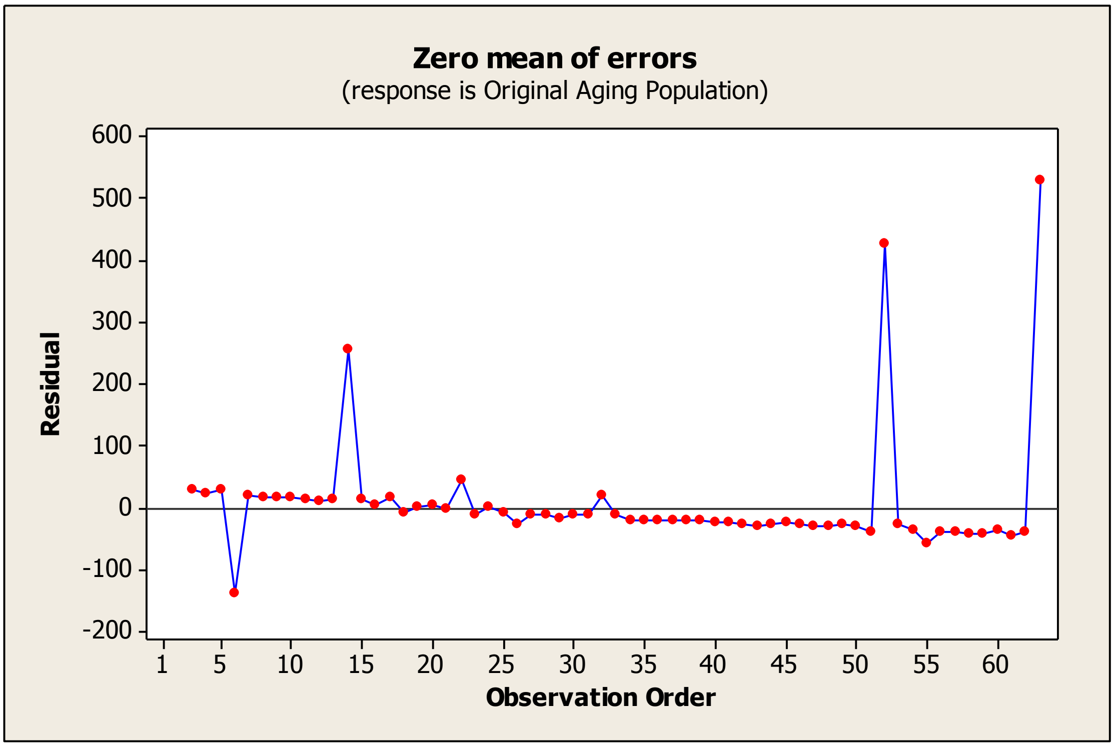
Date: 05/04/17 Time: 22:59
Sample: 1950 2012
Included observations: 61
Q-statistic probabilities adjusted for 1 ARMA term

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| . \| . \| | . \| . \| | 1 | -0.037 | -0.037 | 0.0859 | |
| . \| . \| | . \| . \| | 2 | -0.044 | -0.045 | 0.2115 | 0.646 |
| . \| . \| | . \| . \| | 3 | -0.048 | -0.052 | 0.3668 | 0.832 |
| . \| . \| | . \| . \| | 4 | -0.054 | -0.060 | 0.5619 | 0.905 |
| . \| . \| | . \| . \| | 5 | -0.047 | -0.057 | 0.7167 | 0.949 |
| . \| . \| | . \| . \| | 6 | -0.047 | -0.061 | 0.8728 | 0.972 |
| . \| . \| | .*\| . \| | 7 | -0.049 | -0.067 | 1.0442 | 0.984 |
| .*\| . \| | .*\| . \| | 8 | -0.112 | -0.135 | 1.9461 | 0.963 |
| . \| . \| | .*\| . \| | 9 | -0.059 | -0.096 | 2.2054 | 0.974 |
| . \| . \| | .*\| . \| | 10 | -0.040 | -0.087 | 2.3256 | 0.985 |
| . \|*** \| | . \|*** \| | 11 | 0.401 | 0.371 | 14.701 | 0.143 |
| . \| . \| | . \| . \| | 12 | -0.025 | -0.025 | 14.748 | 0.194 |
| . \| . \| | . \| . \| | 13 | -0.014 | -0.009 | 14.764 | 0.255 |
| . \| . \| | . \| . \| | 14 | -0.013 | -0.013 | 14.777 | 0.321 |
| . \| . \| | . \| . \| | 15 | -0.017 | 0.001 | 14.800 | 0.392 |
| . \| . \| | . \| . \| | 16 | -0.029 | -0.024 | 14.873 | 0.461 |
| . \| . \| | . \| . \| | 17 | -0.013 | -0.000 | 14.887 | 0.533 |
| . \| . \| | . \| . \| | 18 | -0.004 | 0.015 | 14.889 | 0.603 |
| . \| . \| | . \|*. \| | 19 | -0.010 | 0.079 | 14.897 | 0.669 |
| . \| . \| | . \| . \| | 20 | 0.009 | 0.058 | 14.904 | 0.729 |
| . \| . \| | . \| . \| | 21 | -0.020 | 0.014 | 14.943 | 0.780 |
| . \| . \| | **\| . \| | 22 | -0.018 | -0.215 | 14.976 | 0.824 |
| . \| . \| | . \| . \| | 23 | -0.026 | -0.025 | 15.042 | 0.860 |
| . \| . \| | . \| . \| | 24 | -0.021 | -0.034 | 15.088 | 0.891 |
| . \| . \| | . \| . \| | 25 | -0.022 | -0.025 | 15.139 | 0.917 |
| . \| . \| | .*\| . \| | 26 | -0.046 | -0.071 | 15.372 | 0.932 |
| . \| . \| | . \| . \| | 27 | -0.025 | -0.019 | 15.442 | 0.949 |
| . \| . \| | . \| . \| | 28 | -0.020 | -0.037 | 15.487 | 0.962 |

Results of constant variance of ARIMA (0,2,1) model




Results of zero mean of ARIMA (0,2,1) model



Appendix – D: Calculating initial values for level and trend in double exponential smoothing using Minitab

According to Minitab technical support document (2010) Minitab calculates back to the initial observation, using data from later observations:

$$p_2 = 2x_3 - x_4 - (MA_1 * e_3) - (MA_2 * e_4)$$

$$e_2 = x_2 - p_2$$

$$p_1 = 2x_2 - x_3 - (MA_1 * e_2) - MA_2 * e_3$$

$$e_1 = x_1 - p_1$$

Where:

$pi$ = the predicted value of the i$^{th}$ smoothed observation

$xi$ = the value of the i$^{th}$ observation in your time series

$ei$ = the value of the i$^{th}$ residual, stored from the ARIMA

Minitab calculates the initial value for level ($L_1$) according to this formula:

$$L_1 = p_1 + w_1 * (e_1)$$

Minitab calculates the initial value for trend ($T_1$) according to this formula:

$$T_1 = p_2 - L_1$$

Appendix – E: Percentage Errors for training dataset (DES model)

| Year | Ageing Population | Fitted through DES | % Errors | Year | Ageing Population | Fitted through DES | % Errors |
|---|---|---|---|---|---|---|---|
| 1950 | 409 | 417.27 | -2.02 | 1982 | 1008 | 1005.02 | 0.30 |
| 1951 | 419 | 428.55 | -2.28 | 1983 | 1022 | 1028.17 | -0.60 |
| 1952 | 430 | 435.94 | -1.38 | 1984 | 1036 | 1042.24 | -0.60 |
| 1953 | 437 | 446.65 | -2.21 | 1985 | 1052 | 1055.98 | -0.38 |
| 1954 | 453 | 453.37 | -0.08 | 1986 | 1067 | 1071.73 | -0.44 |
| 1955 | 305 | 469.02 | -53.78 | 1987 | 1083 | 1086.56 | -0.33 |
| 1956 | 313 | 319.98 | -2.23 | 1988 | 1099 | 1102.36 | -0.31 |
| 1957 | 319 | 321.97 | -0.93 | 1989 | 1113 | 1118.21 | -0.47 |
| 1958 | 327 | 327.70 | -0.21 | 1990 | 1129 | 1132.06 | -0.27 |
| 1959 | 337 | 335.59 | 0.42 | 1991 | 1141 | 1147.85 | -0.60 |
| 1960 | 346 | 345.57 | 0.12 | 1992 | 1152 | 1159.69 | -0.67 |
| 1961 | 354 | 354.62 | -0.18 | 1993 | 1167 | 1170.40 | -0.29 |
| 1962 | 365 | 362.64 | 0.65 | 1994 | 1184 | 1185.10 | -0.09 |
| 1963 | 621 | 373.63 | 39.83 | 1995 | 1199 | 1201.96 | -0.25 |
| 1964 | 643 | 631.25 | 1.83 | 1996 | 1212 | 1216.91 | -0.40 |
| 1965 | 658 | 662.33 | -0.66 | 1997 | 1226 | 1229.77 | -0.31 |
| 1966 | 686 | 677.73 | 1.21 | 1998 | 1244 | 1243.57 | 0.03 |
| 1967 | 690 | 705.62 | -2.26 | 1999 | 1258 | 1261.43 | -0.27 |
| 1968 | 706 | 709.83 | -0.54 | 2000 | 1264 | 1275.43 | -0.90 |
| 1969 | 728 | 725.23 | 0.38 | 2001 | 1738 | 1281.23 | 26.28 |
| 1970 | 743 | 747.11 | -0.55 | 2002 | 1770 | 1757.66 | 0.70 |
| 1971 | 807 | 762.19 | 5.55 | 2003 | 1793 | 1806.35 | -0.74 |
| 1972 | 817 | 826.32 | -1.14 | 2004 | 1795 | 1829.72 | -1.93 |
| 1973 | 841 | 837.89 | 0.37 | 2005 | 1814 | 1831.02 | -0.94 |
| 1974 | 856 | 861.57 | -0.65 | 2006 | 1834 | 1848.65 | -0.80 |
| 1975 | 856 | 876.65 | -2.41 | 2007 | 1852 | 1867.94 | -0.86 |
| 1976 | 872 | 876.32 | -0.50 | 2008 | 1870 | 1885.31 | -0.82 |
| 1977 | 888 | 891.54 | -0.40 | 2009 | 1893 | 1902.63 | -0.51 |
| 1978 | 900 | 907.36 | -0.82 | 2010 | 1909 | 1925.01 | -0.84 |
| 1979 | 917 | 919.18 | -0.24 | 2011 | 1930 | 1940.56 | -0.55 |
| 1980 | 936 | 935.90 | 0.01 | 2012 | 2521 | 1960.92 | 22.22 |
| 1981 | 986 | 954.82 | 3.16 | | | | |

Appendix – F: Percentage Errors for training dataset (Growth model)

| Year | Ageing Population | Forecasted | % Errors | Year | Ageing Population | Forecasted | % Errors |
|---|---|---|---|---|---|---|---|
| 1950 | 409 | 361.2 | 11.70 | 1982 | 1008 | 914.4 | 9.28 |
| 1951 | 419 | 371.8 | 11.26 | 1983 | 1022 | 941.4 | 7.89 |
| 1952 | 430 | 382.8 | 10.99 | 1984 | 1036 | 969.1 | 6.46 |
| 1953 | 437 | 394.0 | 9.83 | 1985 | 1052 | 997.6 | 5.17 |
| 1954 | 453 | 405.6 | 10.46 | 1986 | 1067 | 1027.0 | 3.75 |
| 1955 | 305 | 417.6 | -36.91 | 1987 | 1083 | 1057.3 | 2.38 |
| 1956 | 313 | 429.9 | -37.34 | 1988 | 1099 | 1088.4 | 0.96 |
| 1957 | 319 | 442.5 | -38.73 | 1989 | 1113 | 1120.5 | -0.67 |
| 1958 | 327 | 455.6 | -39.32 | 1990 | 1129 | 1153.5 | -2.17 |
| 1959 | 337 | 469.0 | -39.17 | 1991 | 1141 | 1187.5 | -4.07 |
| 1960 | 346 | 482.8 | -39.54 | 1992 | 1152 | 1222.4 | -6.11 |
| 1961 | 354 | 497.0 | -40.41 | 1993 | 1167 | 1258.4 | -7.84 |
| 1962 | 365 | 511.7 | -40.19 | 1994 | 1184 | 1295.5 | -9.42 |
| 1963 | 621 | 526.8 | 15.18 | 1995 | 1199 | 1333.7 | -11.23 |
| 1964 | 643 | 542.3 | 15.67 | 1996 | 1212 | 1372.9 | -13.28 |
| 1965 | 658 | 558.2 | 15.16 | 1997 | 1226 | 1413.4 | -15.28 |
| 1966 | 686 | 574.7 | 16.23 | 1998 | 1244 | 1455.0 | -16.96 |
| 1967 | 690 | 591.6 | 14.26 | 1999 | 1258 | 1497.9 | -19.07 |
| 1968 | 706 | 609.0 | 13.73 | 2000 | 1264 | 1542.0 | -21.99 |
| 1969 | 728 | 627.0 | 13.88 | 2001 | 1738 | 1587.4 | 8.66 |
| 1970 | 743 | 645.4 | 13.13 | 2002 | 1770 | 1634.2 | 7.67 |
| 1971 | 807 | 664.5 | 17.66 | 2003 | 1793 | 1682.3 | 6.17 |
| 1972 | 817 | 684.0 | 16.28 | 2004 | 1795 | 1731.9 | 3.52 |
| 1973 | 841 | 704.2 | 16.27 | 2005 | 1814 | 1782.9 | 1.72 |
| 1974 | 856 | 724.9 | 15.31 | 2006 | 1834 | 1835.4 | -0.08 |
| 1975 | 856 | 746.3 | 12.82 | 2007 | 1852 | 1889.5 | -2.02 |
| 1976 | 872 | 768.3 | 11.90 | 2008 | 1870 | 1945.1 | -4.02 |
| 1977 | 888 | 790.9 | 10.94 | 2009 | 1893 | 2002.4 | -5.78 |
| 1978 | 900 | 814.2 | 9.54 | 2010 | 1909 | 2061.4 | -7.98 |
| 1979 | 917 | 838.2 | 8.60 | 2011 | 1930 | 2122.1 | -9.95 |
| 1980 | 936 | 862.8 | 7.82 | 2012 | 2521 | 2184.6 | 13.34 |
| 1981 | 986 | 888.3 | 9.91 | | | | |

Appendix – G: Percentage Errors for training dataset (ARIMA 0,2,1)

| Year | Ageing Population | Forecast | % Error | Year | Ageing Population | Forecast | % Error |
|---|---|---|---|---|---|---|---|
| 1950 | 409 | | | 1982 | 1008 | 1019.07 | -1.10 |
| 1951 | 419 | | | 1983 | 1022 | 1042.13 | -1.97 |
| 1952 | 430 | 402.46 | 6.40 | 1984 | 1036 | 1056.94 | -2.02 |
| 1953 | 437 | 415.62 | 4.89 | 1985 | 1052 | 1071.72 | -1.87 |
| 1954 | 453 | 424.61 | 6.27 | 1986 | 1067 | 1088.54 | -2.02 |
| 1955 | 305 | 442.79 | -45.18 | 1987 | 1083 | 1104.30 | -1.97 |
| 1956 | 313 | 292.25 | 6.63 | 1988 | 1099 | 1121.08 | -2.01 |
| 1957 | 319 | 302.22 | 5.26 | 1989 | 1113 | 1137.83 | -2.23 |
| 1958 | 327 | 310.08 | 5.18 | 1990 | 1129 | 1152.50 | -2.08 |
| 1959 | 337 | 319.93 | 5.06 | 1991 | 1141 | 1169.21 | -2.47 |
| 1960 | 346 | 331.80 | 4.11 | 1992 | 1152 | 1181.78 | -2.59 |
| 1961 | 354 | 342.58 | 3.23 | 1993 | 1167 | 1193.32 | -2.25 |
| 1962 | 365 | 352.28 | 3.49 | 1994 | 1184 | 1208.94 | -2.11 |
| 1963 | 621 | 365.02 | 41.22 | 1995 | 1199 | 1226.61 | -2.30 |
| 1964 | 643 | 629.67 | 2.07 | 1996 | 1212 | 1242.21 | -2.49 |
| 1965 | 658 | 653.42 | 0.70 | 1997 | 1226 | 1255.73 | -2.42 |
| 1966 | 686 | 669.93 | 2.34 | 1998 | 1244 | 1270.26 | -2.11 |
| 1967 | 690 | 699.76 | -1.41 | 1999 | 1258 | 1288.89 | -2.46 |
| 1968 | 706 | 704.86 | 0.16 | 2000 | 1264 | 1303.39 | -3.12 |
| 1969 | 728 | 722.27 | 0.79 | 2001 | 1738 | 1309.65 | 24.65 |
| 1970 | 743 | 745.81 | -0.38 | 2002 | 1770 | 1797.19 | -1.54 |
| 1971 | 807 | 762.11 | 5.56 | 2003 | 1793 | 1829.80 | -2.05 |
| 1972 | 817 | 828.76 | -1.44 | 2004 | 1795 | 1853.13 | -3.24 |
| 1973 | 841 | 839.80 | 0.14 | 2005 | 1814 | 1854.86 | -2.25 |
| 1974 | 856 | 865.22 | -1.08 | 2006 | 1834 | 1874.07 | -2.18 |
| 1975 | 856 | 881.33 | -2.96 | 2007 | 1852 | 1894.31 | -2.28 |
| 1976 | 872 | 881.99 | -1.15 | 2008 | 1870 | 1912.49 | -2.27 |
| 1977 | 888 | 899.08 | -1.25 | 2009 | 1893 | 1930.66 | -1.99 |
| 1978 | 900 | 916.14 | -1.79 | 2010 | 1909 | 1953.96 | -2.36 |
| 1979 | 917 | 929.06 | -1.32 | 2011 | 1930 | 1970.06 | -2.08 |
| 1980 | 936 | 947.10 | -1.19 | 2012 | 2521 | 1991.30 | 21.01 |
| 1981 | 986 | 967.16 | 1.91 | | | | |