SJIF IMPACT FACTOR: 2.996 CRDEEPJournals
International Journal of Social Sciences Arts and Humanities Ayele & Zewdie Vol. 4 No. 3 ISSN: 2321 – 4147
International Journal of Social Sciences Arts and Humanities Vol. 4 No. 3. 2017. Pp. 71-82

Full length Research Paper

Modeling and Forecasting Ethiopian Human Population Size and it`s Pattern

Amare Wubishet Ayele[1*] and Mulugeta Aklilu Zewdie[2]

[1*]*Department of Statistics, College of Natural and Computational Sciences, Debre Markos University, Debre Markos, Ethiopia.*

[2]*Department of Statistics, College of Natural and Computational Sciences, Mekelle University, Mekelle, Ethiopia.*

Article history
Received: 16-05-2017
Revised: 05-06-2017
Accepted: 16-06-2017

Corresponding Author:
Amare W. Ayele
Assistant Department of Statistics, College of Natural and Computational Sciences, Debre Markos University, Debre Markos, Ethiopia.

Abstract:
This study is focused on time series analysis of Ethiopian population based on annual data from 1961 to 2009. Trend analysis and Box-Jenkins time series models on the population is done according to different demographic variables on the indicated year. This paper also analyze age dependency ratio, comparison of population growth rate of Ethiopia with its nine regions and two administrative regions, trend analysis of urban, rural and total population of Ethiopia and comparison of appropriate time series model which are better fit to the population of Ethiopia to forecast is done. The analysis shows the annual population growth rate of Ethiopia is estimated 2.5%, the overall dependency ratio of Ethiopia is 86 per 100 workers and the trend and the forecasted value of the population size of Ethiopia shows rapid increment within short period of time in some regions and in some state/region shows declination due to many factors so the government should looks its population policy for consistency. Moreover, the appropriate time series model, that help us to modeled the total population of Ethiopia was ARIMA (2, 1, 2), since it shows the minimum MSE. The forecast value of total population of Ethiopia shows an increasing pattern (for the coming 29 years).

***Key words**: Trend, Age Dependency Ratio, Growth Rate and Forecast*

***JEL classification**: J1, J11, C1, C22*

Introduction

The world population is the total population of humans on the planet earth, currently estimated to be 6.915128 billion by the United States census bureau, at 2009. The world population has experienced continuous growth since the end of bubonic plaque around the years 1348-1350. Population dynamics are one of the key factors to consider when thinking about development. In the past 50 years in the world has experienced unprecedented increase in population growth. A natural population increase occurs when the birth rate is greater than the death rate (www.nationsencyclopedia.com, 2011).

Moreover, a country's population growth rate depends on migration; world population growth is determined exclusively by the natural increase. The highest rate of growth increase above 1.8 % per year were seen briefly during the 1950's, for a longer period during the 1960's and 1970's the growth rate peaked up 2.2 % in 1963,and declined to 1.1% by 2009 . Annual births have reduced to 140 million since the peak at 173 million in the late 1990', and are expected to remain constant, while deaths number 57 million per year and are expected to increase 80 million per year by 2040.Current projections show a continuous increase of population, with population expected to reach between 7.5 and 10.5 million within the year 2050 (Malmberg, 2006). Ethiopian population raised by staggering 23.4 million people over the past 14 years to its current 76.9 million, according to a census approved by parliament Thursday. The new census stood at 53.4 million when the last census was conducted in 1994. The new census shows the capital Addis Ababa with a population of 2.7 million .Nearly 62 million people or 83.8 % lives in rural areas with central oromia and amhara regions growing by 3.2million and 2.4 million respectively.

Since human population is dynamics and Ethiopia is developing country, there were some problems that had been addressed in the body of this research. Some of the problems are how fast Ethiopian population grown in the past ,what it looks like Ethiopian population in the future ,how does regions of population increase as compare to the total population of Ethiopia ,what can we learn from models of future human population growth, what looks like the trend of the dependency ratio of the past 14 consecutive years, what it seems the trend of Ethiopian population size in the past consecutive years and what it looks like the distribution of a population of Ethiopia by considering residential area (urban area). The main objective of the study is to assess the trend of the population Ethiopian size based on the past historical data. Specifically the concern of this paper are to examine the trend and pattern of Ethiopian total population size over the past period of study (i.e. within the year 1961-2009); to examine the trend and pattern of urban and rural population over the past period of study of Ethiopia by considering residence (urban and rural); to compare the annual population growth rate of each region, so that the plan and program can be set for future development; to assess or examine the dependency ratio and to recommend its economic activity implication; to predict future values of population size of Ethiopia, so that the plan and program can be set for future development; to select appropriate time series model (which is best fit to explain total population size growth) of Ethiopia.

Literature Review

Basic Terminologies in Population Study

Population growth: refers to change in the size of a population which can be either positive or negative, over times depending on the balance of births and deaths. Population measured in both in absolute and relative terms. Absolute growth is difference in numbers between populations over times; on the other hand, relative growth is usually expressed as a rate or percentage (Rowland, 2003).

The population growth rate measures how populations change in size over time. The units of population growth rate are individuals per time. Population size can only be changed by four factors. Births add new individuals to a population whereas deaths remove individuals from a population. Similarly, immigration into a population adds new individuals whereas emigration out of a population removes individuals. Population growth rates are positive when more individuals are added to a population than are removed, negative when more individuals are removed than are added, and are equal to zero when an equal number of individuals are added and removed (Coale, 1974). This population size is known as the carrying capacity and is the size beyond which no significant increase can occur due to limitations of some type, e.g., food, water, space, etc. The average annual percent change in the population, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country. The rate may be positive or negative. The growth rate is a factor in determining how great a burden would be imposed on a country by the changing needs of its people for infrastructure (e.g., schools, hospitals, housing, roads), resources (e.g., food, water, electricity), and jobs. Rapid population growth can be seen as threatening (Shrock, 1976). A percent growth rate (sometimes referred to as percent change, growth rate, or rate of change) is a useful indicator to look at how much a population is growing or declining in a particular area. The percent change from one period to another is calculated from the formula:

$$PR = \frac{(Vpresent - Vpast)}{Vpast} \times 100$$

Where: PR = Percent Rate, VPresent = Present or Future Value (the population at end of period), VPast = Past or Present Value (the population at the beginning of period).

Dependency ratio: is an age-population ratio of those typically not in the labor force (the dependant part) and those typically in the labor force (the productive part). It is used to measure the pressure on productive population. It measures the % of dependant people (not working age) per number of people of working age (economically active).

$$ToDR = \{NoCA(0-14) + NoP \geq 65\}/NoPWA(15-64)$$

As the ratio increases there may be an increased burden on the productive part of the population to maintain the upbringing and pensions of the economically dependent. This results in direct impacts on financial expenditures on things like social security, as well as many indirect consequences. The inverse of dependency ratio can be interpreted as how many independent workers have to provide for one dependent person (pension & expenditure on children).A high dependency ratio can cause serious problems for a country. Dependency ratio is important because it shows the ratio of economically in active compared to economically active will pay much more income tax ,corporation tax and to a lesser extent ,more sales and vat taxes. Distribution of population: refers to the arrangement of the population in space at a given time (geographically or among various types of residential areas). The Ethiopian populations always have been predominantly rural, engaging in sedentary agricultural activities such as the cultivation of crops and livestock rearing in the high lands. The distribution of Ethiopian population generally is related to climate, and soil. This physical factor explains the concentration of population in the high lands which are endowed with moderate temperature, rich soil and adequate rainfall. Population can grow at an exponential rate just as a compound interest accumulates in bank account. One way to assess the growth potential of a population is to calculate it's doubling time, the number of years it will take for a population to double in size, assuming the current rate of a population growth remain unchanged. This is done by applying the "rule of seventy ", that is seventy divided by the current population growth rate (in percent per year).The 1.4% global population growth rate in 2000 translates in to a doubling time (if growth rate remains constant) of 51 year (Ansely, 1974).

Materials and Methods

Data source

Secondary source of data recorded by central statistical authority (CSA) of Ethiopia from year 1961 to 2009 for trend analysis and for developing time series model which shows population of Ethiopia collected from 1996 to 2009 for analysis of dependency ratio and for comparison of population growth rate of each administrative regions were employed.

*Variables of the Study:*The variables in this study are the age dependency ratio, Population growth rate and the population size in urban, rural and total (sum of urban and rural) in Ethiopia.

Model Specification

In this study descriptive statistics which deals with describing (explaining) characteristics of a set (aggregate) of statistical data by methods of organizing, presenting (tables, graphs like chart, time series plot) and analyzing by using statistical measures like mean, ratio, percentage and measures of variation, and inferential statistics are employed with main focus trend analysis, time serious models (*AR, MA, ARMA, and ARIMA*).

Time Series Models

A time series is chronological sequence of observation on a particular variable (sequence of observation taken as regular intervals of time). Usually the observations are taken at regular intervals (days, months, years). Time series is a collection of observations generated sequentially through time. The special features of a time series are that the data are ordered with respect to time and that successive observations are usually expected to be dependent. We have different types of series. Stationary series who's over all behavior remains the same over the time. It only fluctuates around a constant mean, while non-stationary series is a series whose over all behavior fluctuates with a different mean .Time series may have also patterns such as seasonal pattern refers to those periodic movements that occurs regularly every year or month, cyclical pattern the series falls and rises with non-fixed it consists of cycles, which vary in amplitude and duration, trend pattern described the long term sweep of the series.

We can have a model of time series which describes the process that generate time series data using mathematical (statically) expression with the assumption that the observation (Y) is the sum of the component pattern and error .If a time series has regular pattern , then a value of the series should be a function of the previous values. If Y is the target value that we are trying to model and predict, and Y_t is the value of Y at time t, then the goal is to create a model of the form.

$$Observation = component + error$$
$$y_t = f(y_{t-1,} y_{t-2,} \dots y_{t-n}) + \varepsilon_t$$

Where: y_{t-i} is the value of the i^{th} previous observation (i^{th} lag) and y_t is current variable which is dependent on its lags.

Test of randomness

Turning point test: test the systematic oscillation of the time series data depending on the peak and trough. A peak is a value greater than its two neighboring value and a trough is a value less than its two neighboring values. In order to carry out the test we must determine the distribution of the turning points in random series. Let $P1, P2 \dots Pt$ be the total population data within each consecutive years; then the appropriate statistical hypothesis is given by:

H0: Pt, t=1, 2, 3…..T independent identically distributed random variables (the series is time independent) and *H1*: not H_0 (the series is time dependant).

Estimation of Trend

Trend refers to the long term tendency of data to move in an upward and down ward directions. Trend analysis allows describing a historical pattern (i.e. to evaluate the success of previous policy) and to describe past pattern or trends in the future (knowledge of the past can tell us a great deal about the future).In this study four types of trend model: Linear trend model, Quadratic trend model, Exponential growth trend model, and S–curve model are considered. Measure of accuracy mean absolute percentage error (MAPE), mean absolute deviation (MAD) and mean square deviation (MSD) are used to select the better trend model. Family of ARMA and ARIMA Models. Univariate time series models are a class of specifications where one attempts to model and predict time series variables using only information contained in their own past values and possibly current and past values of the error term. Time series models are usually theoretical, implying that their construction and use is not based upon any underlying theoretical model of the behaviour of a variable. Instead, time series models are an attempt to capture empirically relevant features of the observed data that may have arisen from a variety of different (but unspecified) structural models. An important class of time series models is the family of Autoregressive Integrated Moving Average (ARIMA) models, usually associated with Box and Jenkins (1976).The aim of time series analysis is to construct a model for the underlying stochastic process. This model is then used for analyzing the causal structure of the process or to obtain optimal predictions. In statistical analysis of time series, autoregressive–moving average (ARMA) model provides a parsimonious description of a weakly stationary process in terms of two polynomials, one for the auto-regression and the second for the moving average. The model is usually represented by ARMA(p, q) model, where p is the order of the autoregressive part and q is the order of the moving average part.

Autoregressive (AR) Process

An autoregressive (AR) model is one where the current value of a variable Y_t depends upon on the values of the p past values of the variable plus an error term. Specifically an AR model of order p, denoted as AR (p), can be expressed as:

$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} \dots + \alpha_p Y_{t-p} + \varepsilon_t = \mu + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t$

where ε_t is a white noise disturbance term and we assume that it is independent of the past values of the response variable. The autoregressive process of order p is stationary if the roots of the characteristic equation lie outside the unit circle (or, if complex, have modulus greater than one) (Box and Jenkins, 1976, Chatfield, 1996). A useful property of an AR (p) process is that the partial ACF is zero at all lags greater than p.

Moving average (MA) process

A time series Y_t is said to be a moving average process of order q if it is a weighted linear sum of the last q random shocks /errors. In general, the moving average process of order q, denoted as MA (q), can be expressed as:

$Y_t = \mu + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \beta_3 \varepsilon_{t-3} - \cdots - \beta_q \varepsilon_{t-q} = \mu - \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t$

where q is the number of past innovations included in the moving average, $\beta_1, \beta_2, \dots, \beta_q$ are the MA parameters (coefficients) which describe the effect of the past innovations on Y_t and ε_t is the error term which is assumed as a white noise process. The order of the moving average model can be determined by analysis of the autocorrelation function (ACF) which cuts off after q lags and partial ACF that decays exponentially fast.

Autoregressive Moving average (ARMA) Process

The ARMA model states that the current value of the series Y_t depends linearly on its own previous values plus a combination of current and previous values of a white noise error term. A stationary process Y_t is called an ARMA (p, q) process, where p and q are integers, if there exist real coefficients $\alpha_o, \alpha_1, \dots . \alpha_p$; $\beta_1 \dots \dots . \beta_q$ such that,

$$Y_t = \mu + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} \qquad \forall t \in \boldsymbol{Z}$$

Using the backward shift operator:

$\boldsymbol{\alpha}(\boldsymbol{L}) Y_t = \mu + \boldsymbol{\beta}(\boldsymbol{L}) \varepsilon_t$

where ε_t is the error term which is assumed to be a white noise process; and $\alpha(L)$ and $\beta(L)$ are polynomials in L of finite order p and q, respectively, defined by:

$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_P L^P$ and $\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \dots - \beta_q L^q$

The response series is stable if the roots of the homogeneous characteristic equation $\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3 - \cdots - \alpha_P L^P = 0$ lie outside the unit circle.

Autoregressive Integrated Moving Average (ARIMA) Process

In practice many time series are non-stationary and we cannot apply stationary AR, MA or ARMA processes directly. A series which is stationary after being differenced once is said to be integrated of order 1 (denoted by I (1)). Differencing techniques are normally used to transform a non-stationary time series into stationary by subtracting each datum in a series from its predecessor. If the original the series is differenced *d* times before fitting an ARMA (p, q) process, then the model for the original indifference series is said to be an ARIMA(p, d, q) process. The general form of ARIMA (p, d, q) process is given by:

$\alpha(L)(1 - L)^d Y_t = \alpha_o + \beta(L) \varepsilon_t$

where $\alpha(L)$ and $\beta(L)$ are polynomials of order p and q as defined above.

Building ARMA Models: The Box–Jenkins Approach

Box and Jenkins (1976) introduced the first approach for estimating an ARMA models in a systematic manner. Their approach uses an iterative three-stage modeling approach.

Model identification

This is the way of making sure that the variables under the study are stationary, identifying seasonality in the dependent series (seasonal differencing if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the time series variable to decide the tentative (if any) autoregressive or moving average components in the model.

Model selection criteria

When there are multiple adequate models, the selection criterion is normally based on the likelihood function and the number of free parameters from the fitted model or on forecast errors calculated from out-of sample forecast like Akaike's Information Criterion (AIC) defined by: $AIC = -2\ln(L) + 2k$ where L is the maximized value of the likelihood function and k is the number of (free) parameters in the model (i.e. $k = p + q + 1$) and Bayesian Information Criterion (BIC) defined by $BIC = -2ln(L) + kln(T)$

Parameter estimation

Having identified the appropriate p, d and q values, and the next step is to estimate the parameters of the autoregressive and moving average terms included in the model. We can estimate the unknown parameters by ordinary least squares or by maximum likelihood (ML) methods.

Model diagnostic checking

Once we have identified and estimated the candidate ARMA models, we want to assess the adequacy of the selected models. Tests related to serial correlation and normality of residuals should be performed at this stage.

Forecasting

Forecasting the future values an observed time series is an important task in time series analysis in many areas. Once a model has been created for a time series, can use it to forecast future values beyond the end of the series.

Results and Discussion

Descriptive Statistics on population indicators

As indicated on figure 1, the urban and rural population shows increasing pattern from year to year. The size of population in rural is higher than that of urban population. The urban and rural population has linear relationship that means they have the same pattern of movement. As we can see from figure 2, the time plot shows increasing pattern of the population from time to time across residence. All urban, rural and total populations show linear relationship along the study time.

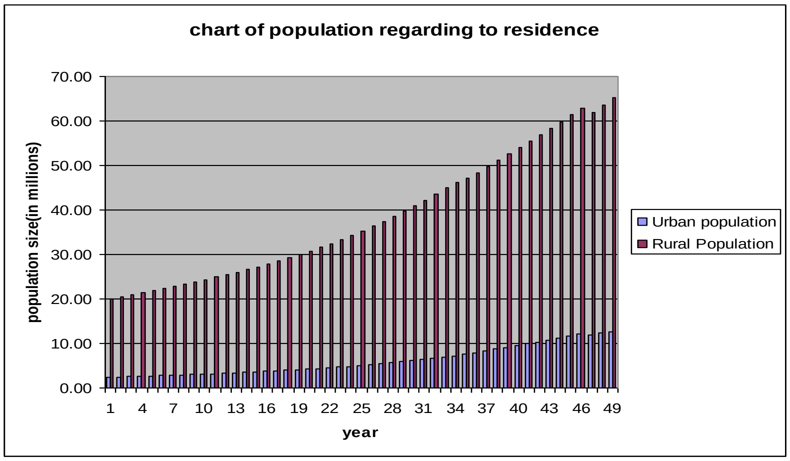


Fig 1: Distribution of population across

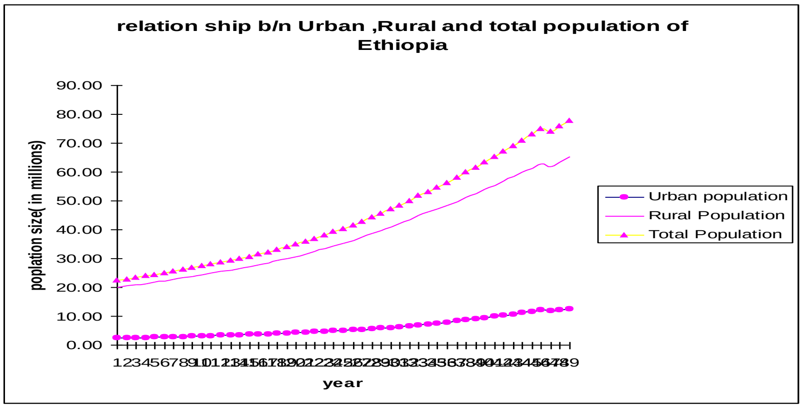


Fig 2: Distribution of population across residence from 1961 to 2009

As we can see from table 1, the annual population growth rate of Ethiopia is 2.50. It means that the population of Ethiopia grew at a rate of 2.50 percent annually. Among the city of administrative and nine regions in Ethiopia, the region that has high population growth rate are Gambela and Benshanqul Gumz region and; and the region that has low population growth rate are Afar and Amhara region along the period of 1996 to 2009. As we can see from figure 4 at the year 2007 the population growth rate is negative that means the population at that time was in the way of decreasing. Generally, the population growth rate of Ethiopia shows a decreasing pattern.

As we can see from table 2 and figure 3 the total dependency ratio of Ethiopia is in the way of decreasing. Its decrement has its own good economic implication, that is, as the ratio decreases there may be a decrease burden on the productive part of the population to maintain the upbringing and pensions of the economically dependent. Statistically between the year 1996 to 2009 population data of Ethiopia, in average the over all, the young and old age dependency ratio were 86.48; 80.84; and 5.64 respectively. Here the overall (total) age dependency ratio is indicating that for each 100 persons in the productive age group there are about 86 (young + old) dependents to be supported.

Table 1: Summary on the analysis of population growth rate of Ethiopian with its nine region and two administrative regions in the year 1996-2009 G.C

	Tigray (01)	Afar (02)	Amhara (03)	Oromia (04)	Somali (05)	Benishanqul (06)	SNNP (07)	Gambella (08)	Harare(09)	A.A	Diredawa	Ethiopia
1996	-	-	-	-	-	-	-	-	-	-	-	-
1997	2.94	2.44	2.94	3.18	2.64	2.48	3.4	2.63	4.13	2.97	4.06	**3.09**
1998	2.88	2.38	2.88	3.12	2.63	2.82	3.3	2.56	3.45	2.97	4.25	**3.04**
1999	2.83	2.41	3.07	3.08	2.62	2.75	3.22	3	2.66	2.97	4.08	**2.98**
2000	2.81	2.36	3.04	3.04	2.66	2.67	3.15	2.43	3.89	2.92	3.92	**2.95**
2001	2.79	2.22	2.99	2.99	2.68	2.6	3.1	2.37	2.37	3.75	3.77	**2.91**
2002	2.73	2.33	2.96	2.96	2.66	2.54	3.02	2.78	3.61	2.96	3.63	**2.87**
2003	2.69	2.28	2.91	2.91	2.67	2.65	2.95	2.7	3.49	2.98	4.38	**2.83**
2004	2.67	2.23	2.68	2.88	2.67	2.41	2.91	2.63	3.93	2.93	3.64	**2.80**
2005	2.67	2.18	2.66	2.86	2.65	2.69	2.87	2.56	2.7	2.92	3.78	**2.78**
2006	2.68	2.21	2.65	2.85	2.63	2.46	2.84	2.92	3.15	2.98	3.64	**2.77**
2007	-0.47	1.59	-9.96	2.28	1.44	7.33	0.94	24.25	-6.45	-7.9	-13.86	**-1.53**
2008	2.49	2.19	1.7	2.9	2.59	2.99	2.89	4.1	2.6	2.09	2.2	**2.53**
2009	2.5	2.20	1.7	2.9	2.6	3.00	2.90	4.09	2.6	2.1	2.5	**2.54**
Annual growth rate	**2.48**	**2.23**	**1.84**	**2.96**	**2.55**	**3.03**	**2.88**	**4.54**	**2.58**	**2.0**	**2.33**	**2.50**

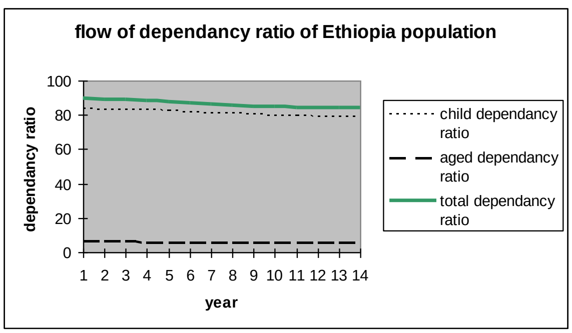


Fig 3: Dependency ratio (child, aged, and total) from 1996 to 2009

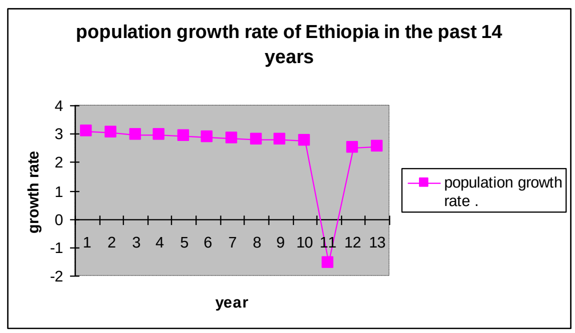


Fig 4: Population growth rate from 1996 to 2009

Table 2: *Age dependency ratio from year 1996 to 2009 GC*

Year (GC)	child dependency ratio	aged dependency ratio	Total dependency Ratio
1996	83.49	6.34	89.83
1997	83.16	6.14	89.3
1998	82.85	5.95	88.8
1999	82.54	5.76	88.3
2000	82.24	5.59	87.83
2001	81.61	5.5	87.1
2002	81.03	5.42	86.45
2003	80.8	5.34	85.82
2004	79.96	5.26	85.22
2005	79.47	5.19	84.66
2006	79.15	5.18	84.33
2007	78.6	5.81	84.41
2008	78.6	5.8	84.4
2009	78.6	5.71	84.31
Average	80.84	5.64	86.48

Inferential Statistics

Before going to analyze the data the first task is to check whether the population data is time series or not .To do this task we apply test of randomness specifically turning point test. Since the test gives $Zcal = -10$ which is much less than $Ztab = 1.96$, then we have enough information to reject the null hypothesis at 5% level of significance. It implies that the series is time dependant (the series is not random). Therefore, it is possible to apply time series analysis on the population data.

Trend Analysis of Urban, Rural and total population

Fit of different trend models such as linear trend model, Quadratic trend model, exponential trend model and s-curve trend model was done. Depending up on measures of accuracy for urban population data the Quadratic trend model given by $Yt = 2.71214 - 2.27E - 02 * t + 4.73E - 03 * t ** 2$; for rural population data the Quadratic trend model given by $Yt = 19.7519 + 0.307117 * t + 1.33E - 02 * t ** 2$, and for total population data exponential trend model given by $Yt = 20.9072 * (1.02761 ** t)$ are more potential (appropriate) trend models to describe their pattern, since their measure of accuracy is very small as we compare with other trend models (Table3).

Table 3: Summary table for different trend fit analysis equation of population of Ethiopian under different category and there measure of accuracy.

Population category	Model Type	Measure of Accuracy			Fitted trend Equation
		MAPE	MAD	MSD	
Urban	Linear trend model	15.87	0.765	0.752	Yt = 0.700153 + 0.213969*t
	Quadratic trend model	3.036	0.157	0.0363	Yt =2.71214 - 2.27E-02*t + 4.73E-03*t**2
	Exponential growth model	3.629	0.217	0.077	Yt = 2.14042*(1.03692**t)
	S-curve model	1.623	0.164	0.188	Yt = (10**2)/(-12.2472 + 54.1170*(0.978465**t-1))
Rural	L linear trend model	6.688	2.161	6.020	Yt = 14.0822 + 0.974141*t
	Quadratic trend model	1.032	0.435	0.332	Yt = 19.7519 + 0.307117*t + 1.33E-02*t**2
	Exponential growth model	1.718	0.66	0.588	Yt = 18.8312*(1.02621**t)
	S - curve model	1.413	0.642	1.282	Yt=(10**3)/(-8.33817 + 59.0343*(0.980256**t-1))
Total	Linear trend model	7.726	2.921	10.8941	Yt = 14.7836 + 1.18808*t
	Quadratic trend model	2.001	0.938	1.0200	Yt = 22.4638 + 0.284525*t + 1.81E-02*t**2
	Exponential growth model	1.904	0.831	0.93092	Yt = 20.9072*(1.02761**t)
	S-curve model	1.403	0.772	2.18304	Yt = (10**3)/(-9.81370 + 54.9913*(0.980811**t-1)

One of the use of trend analysis permits us to describe past pattern or trends in the future (knowledge of the past can tell us a great deal about the future).The future urban, Rural and total population of Ethiopia is forecasted by the appropriate fitted trend model given above. As we can see from table 4, the forecasted values using the appropriated fitted trend model shows strictly increasing pattern from the origin of forecasting 2010.

Fitting Family of ARIMA Models

ARIMA models (autoregressive integrated moving average model is a powerful class of models which can be applied to many real time series. Procedures that we follow in developing ARIMA models are:-

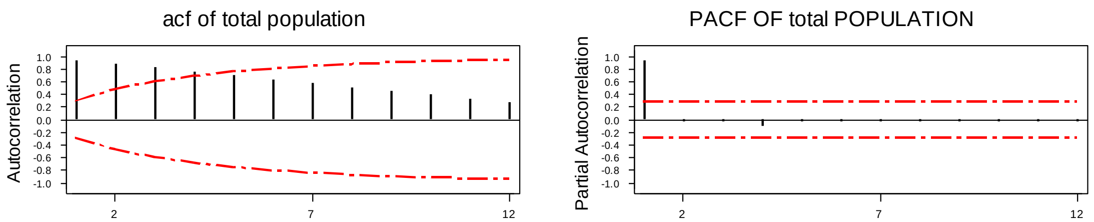
Checking Stationary of the series

Hypothesis $H_{0:}$ The series is stationary (r_k=0) versus $H_{1:}$ The series is not stationary ($r_k \neq 0$)

Table 4: Forecasted values for the urban, rural and total population of Ethiopia by the appropriate trend model

Year	Forecasted value of Total population (In millions)	Forecasted value of Rural population (In millions)	Forecasted value of Urban population (In millions)
2010	81.868	68.4589	**13.4106**
2011	83.977	70.1134	**13.8660**
2012	86.123	71.7946	**14.3309**
2013	88.305	73.5025	**14.8052**
2014	90.523	75.2370	**15.2891**
2015	92.778	76.9983	**15.7823**
2016	95.068	78.7862	**16.2851**
2017	97.394	80.6008	**16.7973**
2018	99.757	82.4420	**17.3190**
2019	102.156	84.3100	**17.8501**
2020	104.591		
2021	107.062		
2022	109.569		
2023	112.113		
2024	**114.692**		

Decision rule: Stationary (test of Stationary)
If the time series is stationary the sample acf and pacf will lies with a limit {-2s.e (r_k) ; 2 s.e(r_k)},and{-2S.E(**Φ_{kk}**);2SE (Φ_{kk}),respectively. if not the series is not stationary(i.e. non stationary). As we can see from figuire 5, the ACF & PACF plot the sample acf and pacf value will lies without a limit {-2s.e (r_k) ; 2 s.e(r_k)},and{-2S.E(**Φ_{kk}**);2SE (Φ_{kk}),respectively, it implies that Ho is rejected ,then the series is not stationary. So to change nonstationary series into stationary series we apply first order differencing. After we apply first order differencing the series becomes stationary (d=1).



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.94	6.59	46.16	8	0.51	1.18	243.79
2	0.88	3.71	87.68	9	0.45	1.01	256.59
3	0.83	2.78	124.77	10	0.39	0.86	266.54
4	0.76	2.23	157.01	11	0.34	0.73	273.98
5	0.70	1.87	184.71	12	0.28	0.60	279.30
6	0.64	1.59	208.19				
7	0.57	1.37	227.77				

Lag	PAC	T	Lag	PAC	T
1	0.94	6.59	8	-0.03	-0.22
2	-0.03	-0.18	9	-0.03	-0.22
3	-0.03	-0.18	10	-0.03	-0.22
4	-0.09	-0.65	11	-0.03	-0.22
5	-0.03	-0.22	12	-0.03	-0.22
6	-0.03	-0.21			
7	-0.03	-0.20			

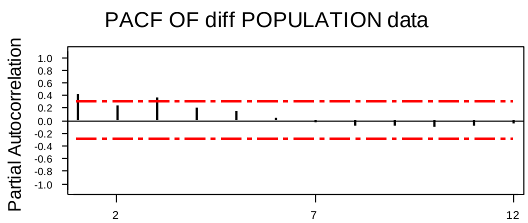
Fig 5: The acf and pacf of the total population series from 1960-2009

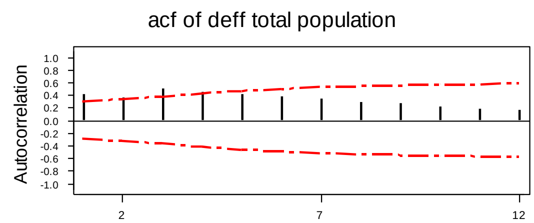
Model Selection (Identification)
As we can see from figure 6, the PACF of the differenced series displays a sharp cutoff after lag 2, then the number of AR term becomes 2, ACF of the differenced series displays a sharp cutoff after lag 2, the number of MA term becomes 2. Therefore the tentatively identified model becomes ARIMA (2, 1, 2). The estimation Parameter estimation (calculate parameter values and goodness of fit statistic).

ARIMA Model: Total population (in millions) variable of interest
ARIMA model for total population of Ethiopia (in millions): Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	1.7814	0.1360	13.10	0.000
AR 2	-0.7952	0.1323	-6.01	0.000
MA 1	1.8097	0.0837	21.61	0.000
MA 2	-0.9062	0.1089	-8.32	0.000
Constant	0.01365	0.01029	1.33	0.192

Differencing: 1 regular difference, Number of observations: Original series 49, after differencing 48
Residuals: SS = 9.79348 (back forecasts excluded), MS = 0.22776 DF = 43
The above ARIMA (2, 1, 2) model converged after twenty one iterations. The AR (1), AR (2), MA (1), and MA (2) coefficients had a t-value of 13.10, -6.01, 21.61 and -8.32 respectively. As a rule of thumb, we can consider values over two as indicating that the associated coefficients can be judged as significantly different from zero. To select the appropriate or most potential model we should modeled other ARIMA model which has the closest order with the tentatively selected model ARIMA (2, 1, 2), for the purpose of developing better fit for the series (Table 5).





Lag	PAC	T	Lag	PAC	T
1	0.42	2.90	8	-0.07	-0.51
2	0.24	1.64	9	-0.08	-0.58
3	0.37	2.59	10	-0.09	-0.64
4	0.21	1.46	11	-0.08	-0.56
5	0.14	1.00	12	-0.04	-0.29
6	0.04	0.28			
7	-0.03	-0.22			

Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.42	2.90	8.95	8	0.30	1.13	72.14
2	0.37	2.21	16.13	9	0.27	0.98	76.57
3	0.51	2.76	29.94	10	0.22	0.80	79.72
4	0.46	2.17	41.43	11	0.18	0.65	81.90
5	0.42	1.81	51.21	12	0.16	0.57	83.65
6	0.39	1.57	59.76				
7	0.34	1.32	66.63				

Fig 6: ACF and PACF plots of the total population series from 1960-2009 (First differenced).

Table 5: Summary of models with closest order with the tentatively selected model ARIMA (2, 1, 2) model (comparison of better fit).

Number	Model type	Type	Coef	SE Coef	T	P	SSE	MSE
1	ARIMA(1,1,4)	AR 1	0.9604	0.5179	1.85	0.071	12.47	**0.30**
		MA 1	0.9043	0.6292	1.44	0.158	Df=42	
		MA 2	0.1427	0.2279	0.63	0.535		
		MA 3	-0.6387	0.4887	-1.31	0.198		
		MA 4	0.5524	0.3653	1.51	0.138		
		Constant	0.04496	0.03107	1.45	0.155		
2	ARIMA(1,2,4)	AR 1	-0.3494	1.9629	-0.18	0.860	11.89	**0.29**
		MA 1	0.6656	1.9812	0.34	0.739	Df=41	
		MA 2	0.3547	2.0097	0.18	0.861		
		MA 3	-0.8379	0.4417	-1.90	0.065		
		MA 4	0.0424	1.0364	0.04	0.968		
		Constant	-0.00517	0.05273	-0.10	0.922		
3	ARIMA(2,1,2)	AR 1	1.7814	0.1360	13.10	0.000	9.793	0.227
		AR 2	-0.7952	0.1323	-6.01	0.000	Df=43	
		MA 1	1.8097	0.0837	21.61	0.000		
		MA 2	-0.9062	0.1089	-8.32	0.000		
		Constant	0.01365	0.01029	1.33	0.192		

In order to select the most potential model, I considered the MSE for selection of the better fit. As we can see from table 5, ARIMA (2, 1, 2) has the smallest MSE, as we compare with the remaining models. Therefore the better potential modeled which is use to modeled the total population data of Ethiopia is ARIMA (2; 1; 2) is given below.

$$Yt \;= 0.01365 + (1 + 1.7814)\,Yt-1 - 0.7952\,Y\,t-2 + \varepsilon t - 1.8097\,\varepsilon t-1 + 0.9062\,\varepsilon t-2$$
$$= 0.01365 + 2.17814\,Yt-1 - 0.7952\,Y\,t-2 + \varepsilon t - 1.8097\,\varepsilon t-1 + 0.9062\,\varepsilon t-2$$

Test of significant for the ARIMA (2, 1, 2) parameters

As we can see from table 5, for the selected ARIMA (2, 1, 2) model the coefficients of AR (1), AR (2), MA (1) and MA (2) have no significant p-value (the p-values are small). This shows that the associated parameters are significantly different from zero. Therefore we can conclude that the parameters are in the model. On the contrary the constant (μ) have a significant (large) p-value; this shows that the associated parameters are not significantly differ from zero. Then we can remove from the model. Therefore the model becomes:-

$$Yt = 2.17814\, Yt - 1 - 0.7952\, Y\, t - 2 + \varepsilon t - 1.8097\, \varepsilon t - 1 + 0.9062\, \varepsilon t - 2$$

Diagnostic Checking

Examining the acf and pacf plots of the residual: normality test of residuals

As we seen from figure 7, the ACF and PACF of the residual have no any structure (that is there is no significant spike); this shows that the residual of the model are white noise.

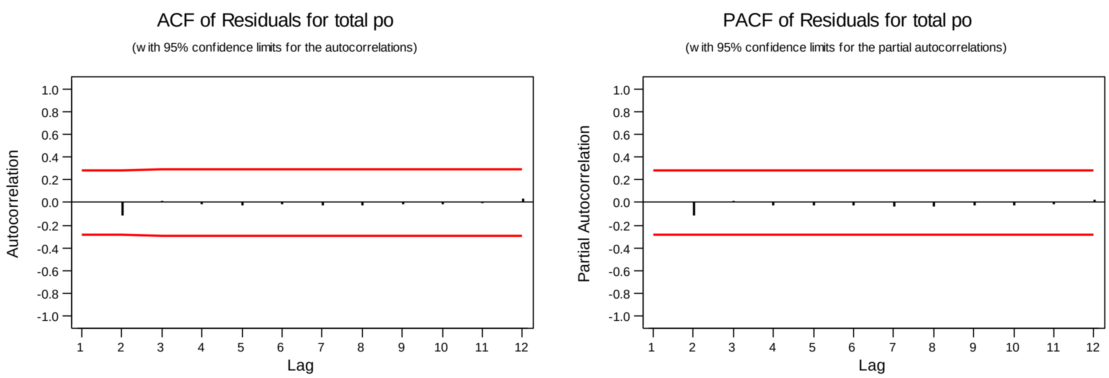


Fig 7: Residual plot for ACF and PACF of the total population data.

Apartmanteau lack of fit test (Ljung-Box) Chi-Square statistic

This examines the autocorrelation coefficients of residual by taking a number of starting from 1 lags say 10, 15, or 25.Table 6: summary table for modified box-pierce (Ljung-box) chi- square statistic

ARIMA(2,1,2)

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	**48**
Chi-square	0.9	2.4	3.4	*
Df	7	19	31	*
p-value	**0.996**	**1.000**	**1.000**	*

From the test given above the Ljung-Box statistics give no significant p-values, indicating that the residuals appeared too uncorrelated. The ARIMA (2, 1, 2) model appears to be adequate (fit well), we can use it to forecast the population size in the future.

Forecasting

Forecasting the future values of an observed time series is the main objective of this paper, which is given in table 7 and figure 8. As you can see from the table the forecast values using ARIMA (2, 1, 2) model shows strictly increasing pattern from the origin of forecasting 2009 up to 2029.

Table 7: 20 year forecasting values for the total population of Ethiopia using ARIMA (2,1, 2) model

Period/year in G.C	**Forecast value (in millions)**	**95 Percent Limits** **Lower Upper limit**
2010	79.299	78.363 , 80.234
2011	80.645	79.340 , 81.949
2012	81.872	80.249 , 83.495
2013	83.002	81.048 , 84.956
2014	84.052	81.727 , 86.377
2015	85.039	82.288 , 87.789
2016	85.974	82.737 , 89.211
2017	86.87	83.084 , 90.655
2018	87.735	83.343 , 92.128
2019	88.578	83.524 , 93.633
2020	89.46	83.642 , 95.170
2021	90.223	83.707 , 96.739
2022	91.035	83.731 , 98.339
2023	91.844	83.722 , 99.966

2024	92.654	83.689 , 101.619
2025	93.468	83.640 , 103.295
2026	94.286	83.581 , 104.991
2027	95.11	83.518 , 106.703
2028	95.942	83.455 , 108.429
2029	96.782	83.397 , 110.167

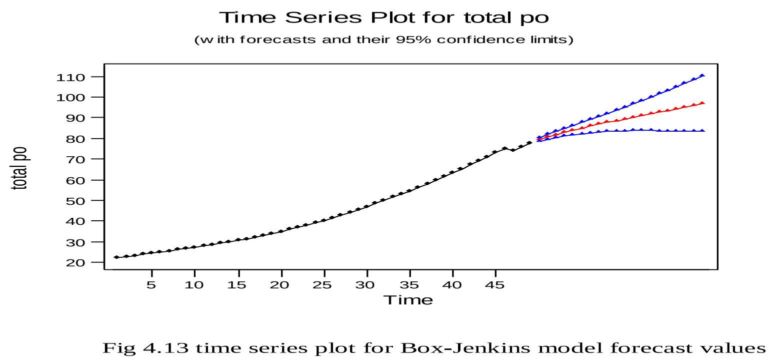


Fig 8: Forecast values of Ethiopian population size from 2009 to for the coming 20 years

Conclusion

Based on the result of this study, the annual population growth rate of Ethiopia is estimated around 2.5%. The amhara region has the smallest annual population growth rate which is 1.84 %; and Gambella region has the highest annual population growth rate which is 4.54 %; between the year 1996 -2009. Based on the past 14 years (1996-2009) population data of Ethiopia the overall dependency ratio shows a slowly decrement from year to year. Due to the dependency ratio shows a decrement trend, the economically active society will increase a much more income tax, corporation tax, and to a lesser extent, more sales and VAT taxes. In Average the overall age dependency ratio of Ethiopia is 86.48, that means that for each 100 persons in the productive age group, there are about 86 dependants to be supported. The trend of urban, rural and total population of Ethiopia shows an increasing pattern (shows increment pattern from year to year).The appropriate time series model which modeled the total population of Ethiopia is ARIMA (2, 1, 2).The forecast value of total population of Ethiopia shows that the total population of Ethiopia are increasing for the future (for the coming 29 years).

Recommendation

Based on the finding of this study, the following possible recommendations are forwarded which are believed to better solution to some of the problems in this study. Since the population growth of amhara region shows a very high decrement pattern, as we compare with other regions, so the administrative should be asses the causes of low population growth rate like disasters, migration and Gambela region shows the highest annual population growth rate than others then the government should be satisfies the needs like health service, education service, and others. Since the pattern of Ethiopian population shows an increasing pattern so the government of the country should be formulate the future policy of the country by considering as the population is in increasing way. The forecast value of urban, rural, and total population of Ethiopia shows an increment in size, then the administrative body of the country should be increase the infrastructures like education service ,health service, road construction and also other necessities.

References

Box, G. and Jenkins, G. (1976). Time Series Analysis: Forecasting and Control, *San Francisco: Holden-Day.dd.*

Chatfield, C. (1996, 2003). ***The Analysis of Time Series***, Fifth Edition, Chapman & Hall, London.Coale, A.J (1974). History of Human Population. *Scientific American*. 231 40.51

Douglas c. Montgomery, Cheryl l. Jennings, and Murat Kulahci (2008) Introduction to Time Series Analysis and Forecasting(a john wiley &. sons, inc., publication)

Fuller,W.A (1976). Introduction to time series.

Hood, R.P (2003). Statistics for Business and Economics (third edition)

http://www.nationsencyclopedia.com/africa/ethiopia-location-size-and-extent.html#ixzz1owkpbsjj

Jonathan D. Cryer and Kung-Sik Chan(2008) Time Series Analysis with Applications in R Second Edition(Springer Science+Business Media, LLC)

Malmberg. Bo. (2006). Global Population Ageing, Migration and European External Policies, Institute for Futures Studies, Stockholm Sweden.

Peter J. Brockwell and Richard A. Davis (2002) Introduction to time series and forecasting 2nd ed (Springer texts in statistics)

Richard Harris and Robert Sollis(2003) Applied Time Series Modelling and Forecasting(John Wiley & Sons Ltd)

Robert H. Shumway and David S. Stoffer(2011) Time Series Analysis and Its Applications With R Examples Third edition(springer text in statistics)
Rowland, D.T(2003). Demographic Methods and concepts, Oxford University press, USA.
Shrock,H.S,and Siegel,J.S (1976):The methods and materials of demography studies in population, academic press, California, USA.
Vandele.w(1983). Applied time series and Box –Jenkins models.
Wei(2006) Time series analysis univariate and multivariate methods (congress catalog publication)