Results of Appeared Candidates in Matriculation / HSLC Examination in Assam (India): A Time Series Analysis (From 1951-2014)

Geeta R. Sarmah [a], Labananda Choudhury [b] and Subrata Chakraborty* [c]

[a] Kamrup Academy Higher Secondary School, Guwahati-781001, Assam, India.
[b] Department of Statistics, Gauhati University, Guwahati-781014, Assam, India.
[c] Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India.
*Corresponding author; e-mail: subrata_arya@yahoo.co.in

Abstract

Completed Secondary education serves as a common denominator for progressing towards further education. It plays a significant role in generating the opportunities and benefits of individual, social and economic development. Therefore, it is necessary to know the nature of rate of pass percentages of female and male candidates in High School Leaving Certificate (HSLC) examination as an indicator of completion of secondary stage of education. From independence in Assam, a north-eastern state of India, the pass percentages of both male and female candidates in Matriculation / HSLC examination were not satisfactory. So, it is imperative to conduct an extensive study on completion rate of secondary education by the candidates using appropriate method. The objective of this study is to fit a time series model and forecast the pass percentages of female and male candidates in Matriculation / HSLC examination, using time series data from 1951-2014. It is observed that ARIMA(0, 1, 2) and ARIMA(0, 1, 1) models are useful for forecasting pass percentages of female and male candidates respectively. Structural approach is adopted to construct several models eliminating inappropriate ones and keeping the most suitable model. The selected models are verified in terms of model structure and forecasting accuracy. The findings of this study may be helpful for effective planning and program implementation through proper budget allocation to ensure sustained growth and development of secondary education in the state.

_____

Keywords: Male and female candidates, matriculation / high school leaving certificate (HSLC) examination, ARIMA techniques.

1. Introduction

At the end of tenth standard the Matriculation/High School Leaving Certificate (HSLC) is awarded to young adolescents of age-group 15-16 who successfully complete the curriculum and examination affiliated to Universities/Education Boards/Councils at National and State levels. In

India, High School leaving Certificate is accepted legally as proof of age and entry qualification to many jobs in the public and private sectors. Before the Passport and Permanent Account Number (PAN) or Identity Card mechanisms, the Matriculation/HSLC admit was the major Photo Identity for circulation (High School Leaving Certificate 2007). Besides getting the School Leaving Certificate, the actual learning one earns during the ten years of schooling is a lifelong resource. The performance in HSLC examination determines one's eligibility for enrolling into further education. Moreover, examination results are regarded as the key driver for systematic reform in teaching-learning process. Therefore, the prime objective of Secondary Education is to provide good quality education available, accessible and affordable to all young adolescents of relevant age group. The secondary education should adequately prepare children to join the labor market or continue to higher education. Today higher education is considered as a means of achieving an innovative society which is getting increasingly important for economic performances, productivity and competitiveness (Madore 1992).

Secondary stage of education has far-reaching benefits both at individual and social level. At individual level especially for girls, the positive externalities of secondary education on health, living conditions, age at marriage, reduced fertility, improved birth practices, lower maternal and child mortality, improved education of children and slower population growth are stronger than those of primary education (Secondary Education in India 2009). Since independence, in Assam, a north eastern state of India, the completion rate of male and female candidates in Matriculation/HSLC examination varied significantly. As in 1951, among 88 percent male and 12 percent female appeared candidates 41 percent male and 51 percent female candidates passed in the Matriculation examination. During 1961, 80 percent male and 20 percent female candidates appeared while only 48 percent male and 42 percent female candidates passed in the examination. In 1981, among 61 percent male and 39 percent female appeared candidates 45 percent male and 31 percent female candidates passed in the examination. During 2001, 54 percent male and 46 percent female candidates appeared while only 37 percent male and 20 percent female candidates qualified in the examination. In 2011, 50 percent male and 50 percent female candidates appeared in HSLC Examination. Their corresponding pass percentages were 73 and 67 percent for male and female candidates respectively. During 2013, appeared percentages of male and female candidates were 49 and 51 respectively, while their pass percentages were 73 and 68 respectively. In the year 2014 also 49 percent male and 51 percent female candidates appeared while, only 65 percent male and 58 percent female candidates passed in HSLC examination.

In Assam, though substantial progress in narrowing the gender gap in appearance rate over time in HSLC examination is achieved, yet the existence of significant gender gap in completion rate is noticed. Narrowing the gender gap in secondary education level is one of the prime goals of Millennium Development Goal (ADB 2006). However it is also evident that a large portion of both female and male candidates were unable to complete the secondary education over the period. It highlights the failure of education systems in Assam to prepare a major portion of young people with right abilities to complete the secondary education. Provision of secondary education of good quality is a crucial tool for generating the opportunities and benefits of social and economic development. So, secondary education is the focus of increasing policy debate and analysis worldwide (World Bank 2005). It is important to develop appropriate strategies to help young generation, irrespective of gender to acquire the knowledge and skills that they need to lead healthy, secured and productive lives. It will be helpful for adapting to the ever-changing society.

Therefore, in the midst of 2014, we should know the growth pattern of both female and male candidates qualified in HSLC examination in Assam for effective budget and program planning for

quality secondary and higher education using time series data. A time series is a set of observations on the values that a variable takes at different points of time. Revealing the growth pattern and making the best forecast of qualified candidates in HSLC examination in Assam using appropriate time series technique is essential to stimulate sustainable development.

Among various time series forecasting techniques Autoregressive Integrated Moving Average (ARIMA) model pioneered by Box and Jenkins is proved to be most powerful (Bisgaard and Kulahci 2011). In the field of economics, finance, business, physical science, social science, medical science etc., Box-Jenkins techniques are extensively used to better understand the dynamics of a system and to make sensible forecasts about its future behavior (Bisgaard and Kulahci 2011). But the use of Box-Jenkins techniques in the field of education for analyzing passed candidates in any examination is very limited. Being a concurrent subject, state has responsibilities to expend on education at all levels, including secondary education (Aggarwal 1993). Therefore accurate models may be important for the state as well as for the nation for pursuing appropriate strategies for providing quality and relevant secondary and higher education to young generation for generating sustainable development.

2.   Objective of the Study

The objective of the present study is to propose models for pass percentage of both male and female candidates in Matriculation/HSLC examination in Assam from 1951 to 2014 using ARIMA techniques. Time series data on pass percentages of male and female candidates in Matriculation/HSLC examination in Assam from 1951 to 2014 are analyzed to know the stochastic properties of the data on their own under the philosophy "let the data speak for themselves" (Gujarati and Sangeetha 2011).

3.   Sources of Data

Time-Series data on male and female candidates passed in Matriculation/HSLC Examination from 1951 to 2014 were collected from different sources. Secondary education in Assam started in the year 1835. During that period administrative Headquarter of East India Company was Calcutta and Assam was a part of Bengal. There was no separate Department of Education in Assam. In 26th January 1948, Gauhati University was established. Since then the responsibilities of academic matters in secondary stage was entrusted to Gauhati University and rapid growth of secondary education started in Assam. The class X public Examination was known by different names: Entrance Examination (conducted by Calcutta University till 1947), Matriculation Examination (conducted by Gauhati University from 1948 to 1963) and High School Leaving Certificate Examination (conducted by Board of Secondary Education, Assam since 1964) (The Assam Tribune, 2009). Therefore, data were collected from Gauhati University Information Center (for 1951 to 1963), District Library, Guwahati (for 1964 to 1990) and Assam Board of Secondary Education (for 1991 to 2014). Closer observation of the collected data suggests that the pass percentages for both male and female candidates are not satisfactory from 1951 to 2014. It also reflects that though appearance rate of female candidates in HSLC Examination is increasing over time and surpassed the male candidates in 2013, the pass percentages for both male and female candidates are not satisfactory.

4. Methodology

The Box-Jenkins ARIMA(p, d, q) model is given as (Bisgaard and Kulahci 2011)

$$z_t = \sum_{i=1}^{p+d} \varphi_i z_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i}. \tag{1}$$

The three basic parameters, namely p, d and q are involved in involved in Equation (1). Here p represents the amount of autoregression, d indicates the level of systematic change over time (trend) and q represents the moving average part. In the modeling process these three parameters are estimated in an iterative way using three stages, viz. model identification, parameter estimation and diagnostic checking until the most suitable model is found (Chen 2008, Bisgaard and Kulahci 2011). Assumption of time series forecasting is that the future depends upon the present while the present depends on the past (Chen 2008).

The basis for any time series analysis is stationary time series (Chen 2008, Bisgaard and Kulahci 2011). A stationary time series has constant mean, constant variance and constant autocorrelation structure. Therefore, first step in developing an ARIMA model is to test if the series is stationary. Three ways were adopted to ascertain stationary. These were as follows:

i. Examining the plot of the raw data. To be stationary, the plot of the series should show constant location and scale.

ii. Observing the autocorrelation function (ACF) and the resulting correlograms. The ACF at lag k, denoted by ρ_k, is defined as:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{autocovariance at lag } k}{\text{variance of the time series}}, \text{ where } k \leq T/4, \ T = \text{Number of observations.} \tag{2}$$

We have 64 numbers of observations. If the ACF does not damped out within 64/4=16 lags, the process is likely not to be stationary (Bisgaard and Kulahci 2011).

iii. To provide further evidence for the nonstationary time series, we conducted Augmented Dickey-Fuller (ADF) unit root test (Gujarati and Sangeetha 2007).

If the time series is stationary, then the assumption of constant mean and homogeneity variance are met. But if the pattern presents a trend, the method of differencing advocated by Box-Jenkins can be used to remove the linear or curvilinear trend. The first order of differencing ($d=1$) is designed to remove the linear trend while the second order of differencing ($d=2$) is used to remove the curvilinear trend (Chen 2008).

The variability (if any) present in a process may be stabilized by employing logarithmic transformation before first differencing (Negron 2012 to 2015). A rough graphical check for the right transformation is the range-mean plot, which is produced by dividing the time series into smaller segments and plotting the range versus the average of each segment on a scatter plot. If the plot indicates linear relationship between average and range, then log transformation is appropriate (Bisgaard and Kulahci 2011).

Once the stationarity of the nonstationary time series for passed candidates is achieved, next step is to identify the order of ARIMA(p,d,q) model. The primary tools in identification are the Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF) and the resulting correlograms. Partial autocorrelation is the autocorrelation between z_t and z_{t-k} after removing any linear dependence on $z_1, z_2, ..., z_{t-k+1}$ and is denoted by ϕ_{kk}. After identifying the appropriate p and q values, the parameters included in the model are estimated. To choose the best model we used Akaike's Information Criterion (AIC) (Deb Roy and Das 2012). We choose the model that has minimum AIC value. For a sample size of n observations, AIC is given by:

$$AIC = -2\ln\left(\text{maximized likelihood}\right) + 2r \approx n\ln\left(\hat{\sigma}_a^2\right) + 2r, \tag{3}$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of the residual variance $\hat{\sigma}_a^2$, r is the number of parameters estimated in the model including a possible constant (Bisgaard and Kulahci 2011).

We also applied forecast accuracy criteria, mean absolute percentage error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) to decide the better model. The RMSE, MAPE and MAE are defined as:

Root mean square error, $\qquad RMSE = \sqrt{\dfrac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \tag{4}$

Mean absolute percentage error, $\qquad MAPE = \dfrac{1}{n}\sum_{t=1}^{n}\left|\dfrac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\% \tag{5}$

Mean absolute error, $\qquad MAE = \dfrac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \tag{6}$

where y_t is the observed value; $\hat{y}_t$ is the predicted value and n is the number of predicted values. The smaller the value of RMSE, MAPE and MAE, the better the model is (Chen 2008, Shitan et al. 2014).

For further check to ascertain how well the estimated model fits the data, we conducted the usual residual diagnostic test. If the residuals estimated from the selected model are white noise, we can accept the particular fit (Gujarati and Sangeetha 2007). To confirm, we further used Ljung-Box statistic to test for non-zero autocorrelations in the residuals at lags 1-16.

The Ljung-Box (LB) statistic is defined as,

$$LB = n\left(n+2\right)\sum_{k=1}^{m}\left(\dfrac{\hat{p}_k^2}{n-k}\right) \sim \chi_m^2. \tag{7}$$

If observed χ_m^2 is less than expected χ_m^2 then there is no evidence of white noise at lags 1 to 16 (Gujarati and Sangeetha 2007). R software was used for fitting ARIMA model in this work.

5. Methodology

5.1. Female candidates

5.1.1. Tests for stationarity

Figure 1(a) represents the time series plot of pass percentages of female candidates in Matriculation/HSLC Examination in Assam for all the years from 1951 to 2014. The plot reveals that the data is nonstationary in mean and variance i.e. a chain of rapid decline in the period of 1951-1992 and then followed by a persistent increasing pattern in the period of 1993-2014. In post independence period different education committees, commissions and policies like Dr. Tara Chand Committee, 1948 (Government of India), University Education Commission, 1948 (Government of India), Secondary Education Commission, 1952 (Government of India), Indian Education Commission, 1964-1966 (Government of India), National Policy on Education, 1968 (Government of India) recommended different ways and means to bring about revolutionary changes to make entire education system including secondary education more dynamic and multidirectional (Mahanta 1997). The National Education Policy (Government of India) and Program of Action (Government of India) most correctly considered social relevance and quality of secondary

education as of prime importance. Our country has set before itself the goal of "Education for all" by 2010, providing high quality secondary education to all Indian adolescent boys and girls up to the age of 16 years by 2015 and senior secondary education up to the age of 18 years by 2020 (World Bank 2009). It is reflected in the results from 1992 to 2014 in Figure 1(a).

The non-stationary pattern was confirmed by observing the ACF and PACF plots in Figures 1(b) and 1(c). The plots depict strong positive autocorrelation. It is observed that the ACF for raw data does not die out even for large lags. Therefore, the time series is nonstationary (Bisgaard and Kulahci 2011, Gujarati and Sangeetha 2007). Autocorrelation at lag 2 and above are merely due to the propagation of the autocorrelation at lag 1. This is confirmed by Partial Autocorrelation Function (PACF) plot. The PACF plot has a significant spike at lag 1, meaning that all the higher order autocorrelation are effectively explained by the lag-1 autocorrelation (ARIMA models for time series forecasting). The nonstationarity of the time series is confirmed by using Augmented Dickey-Fuller (ADF) unit root test. With a Dickey-Fuller = -0.7838 and p-value= 0.9582, we accept that the time series for Results of appeared female candidates in Matriculation/HSLC examination is nonstationary. The nonstationarity in mean should be corrected by differencing the data.
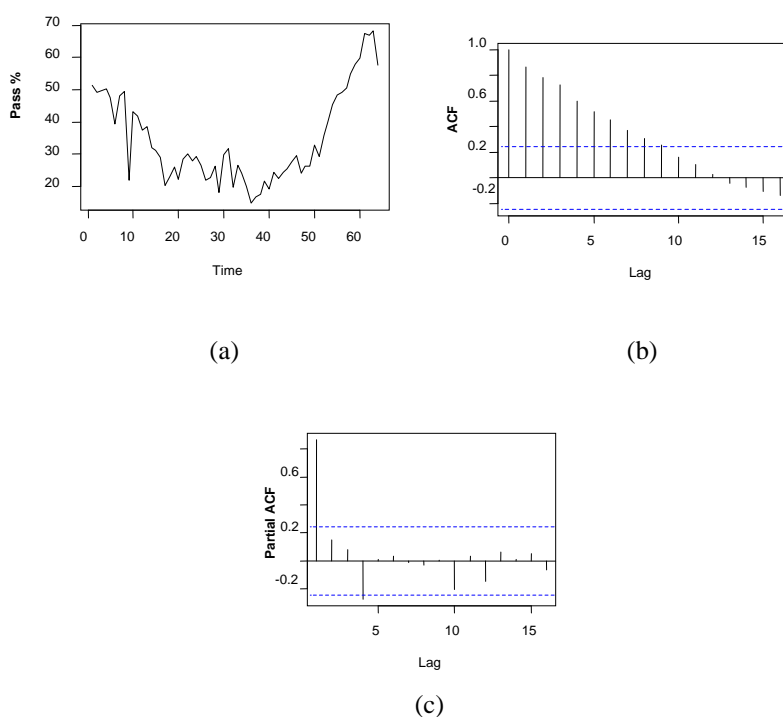

(a)   (b)

(c)

Figure 1 (a) Time series plots (b) ACF Plots (c) PACF plots of pass percentage of female candidates in Matriculation/HSLC examination in Assam (1951-2014)

Before differencing we have checked for right log transformation to obtain homogeneous variability, by dividing the time series into small segments of five years, computed the five years' range and average and plotted as scatter plot in Figure 2(a) for pass percentage of female candidates. It is observed from the Figure 2(a) that, there exists no relationship between averages and ranges of segments in the time series. Therefore, log transformation is not necessary. For

confirmation we conducted log-transformation to the time series of pass percentage of female candidates and presented the respective graph in Figure 2(b). Comparing Figure 2(b) with the Figure 1(a) it is observed that although the scale has changed, it does not appear to have less variability than the untransformed plot.
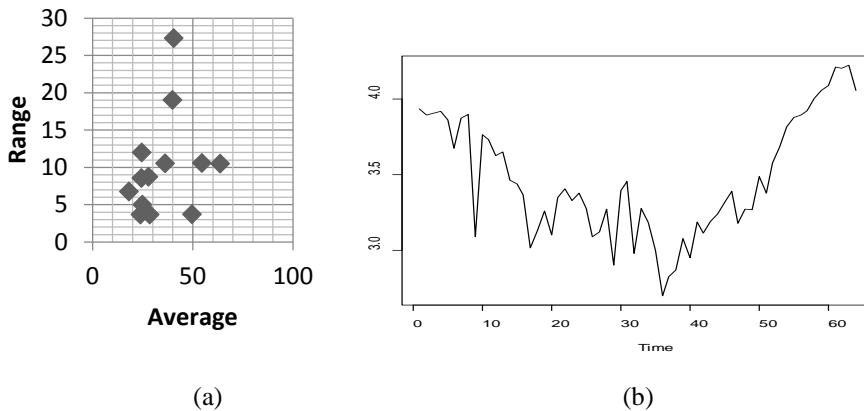

(a)   (b)

Figure 2 (a) Range-mean plot (b) Log transformed series of pass percentage of female candidates in Matriculation/HSLC examination in Assam (1951-2014)

Figure 3(a) represents first differenced time series plot. Figures 3(b) and 3(c) represent the ACF and PACF plots of first differenced series. The figures suggest that, as a result of first differenced, nonstationarity is considerably reduced. For confirmation we apply ADF test to the first differenced series. It is found that Dickey-Fuller = -4.3529 with p-value 0.01. From the ACF and PACF plots and ADF test we confirmed that the series is now stationary.
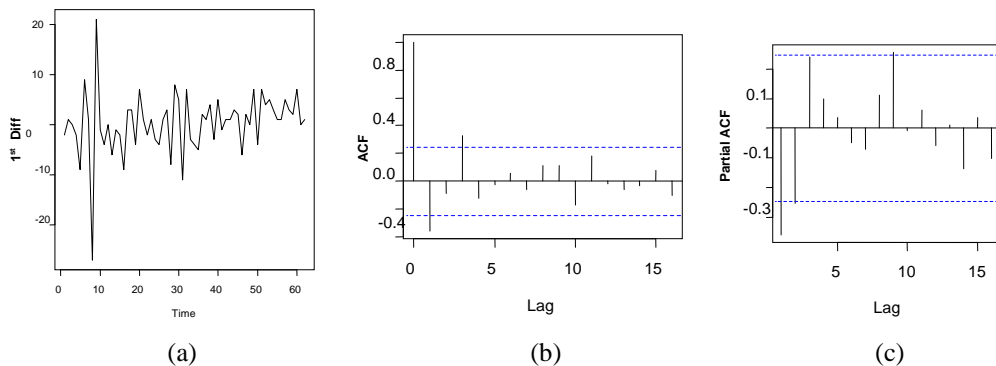

(a)   (b)   (c)

Figure 3 Plot of (a) First differenced time series (b) ACF of first differenced (c) PACF of first differenced of percentage of female candidates in Matriculation/HSLC examination in Assam (1951-2014)

5.1.2.   Model identification
For identifying the ARIMA(p, d, q) model, we examine the ACF and PACF plots of stationary time series of female candidates passed in Matriculation/HSLC examination in Figure 3(b) and 3(c).

Here we differenced the series only once, so $d = 1$. For identifying the order of autoregressive component p we observed the PACF plot in Figure 3(c). It is observed that, there are two large autocorrelation at lag 1 and lag 2. At lag 2 it touches the significant bounds. All other autocorrelations cut off after lag 2 .Therefore PACF plot suggests that an AR(2) could be accurate in the present time series.

To get idea about the order of moving average component, we examined the ACF plot in Figure 3(b). Here we found, there is a significant correlation at lag 1. After lag 1 all other correlations drop near to zero except at lag 3, as lag 3 exceeds significant bound. Therefore ACF plot suggests an MA(1) or MA(2) could be accurate. We have identified p, d, q so our potential models may be:

(1) ARIMA(1, 1, 0) as PACF is zero after lag 1.
(2) ARIMA(0, 1, 1) as ACF is zero after lag 1.
(3) ARIMA(0, 1, 2) as ACF is zero after lag 1 and at lag 2 ACF touches significant bound.
(4) ARIMA(1, 1, 1) a combination model of (1) and (3).
(5) ARIMA(1,1,2) a combination model of (1) and (4).

We tried all the models and from the point of view of parsimony and forecast accuracy we selected the model with the lowest value of AIC, MAPE, RMSE and MAE. The results are presented in Table 1 below.

Table 1 The AIC, MAPE and MAE values for the time series of participated female candidates (1951-2014)

Model	AIC	RMSE	MAE	MAPE (%)
ARIMA(1,1,0)	406.30	5.84	4.20	14.25
ARIMA(0,1,1)	405.14	5.79	4.22	14.24
ARIMA(0,1,2)	404.44	5.65	4.10	13.87
ARIMA(1,1,1)	406.58	5.76	4.22	14.25
ARIMA(1,1,2)	405.06	5.59	4.00	13.56

In the present study, both ARIMA(0, 1, 1) and ARIMA(0, 1, 2) models are found to be adequate for pass percentage of female candidates appeared in the examination. But in residual check we found existence of non-zero autocorrelation in the forecast errors at lags 1-16 in ARIMA(0, 1, 1) model and hence preferred the ARIMA(0,1,2) model over ARIMA(0, 1, 1).

5.1.3. Model estimation
After identifying the model, the parameters are estimated. The summary information for the fitted model is incorporated in Table 2.

Table 2 Estimated Coefficients for an ARIMA(0, 1, 2) Model (Pass% Female)

Variables	Coefficients	Standard Errors	σ^2	Log likelihood	p-value
MA(1) θ_1	-0.5882	0.1502	32.44	-199.22	0.000
MA(2) θ_2	0.3281	0.1966			0.003

5.1.4.   Diagnostic checking

For further check how well the ARIMA(0, 1, 2) model fits the data, we conducted the usual residual diagnostic checks. In Figure 4 (a), (b) and (c) the ACF, PACF and Ljung-Box Chi square value of residuals after fitting ARIMA(0, 1, 2) model to the time series of pass percentage of female candidates in the Examination are presented. It is observed that none of these plots show any significant autocorrelation. The nature of residuals in Figure 3 indicates that the ARIMA(0, 1, 2) model fits the data well.
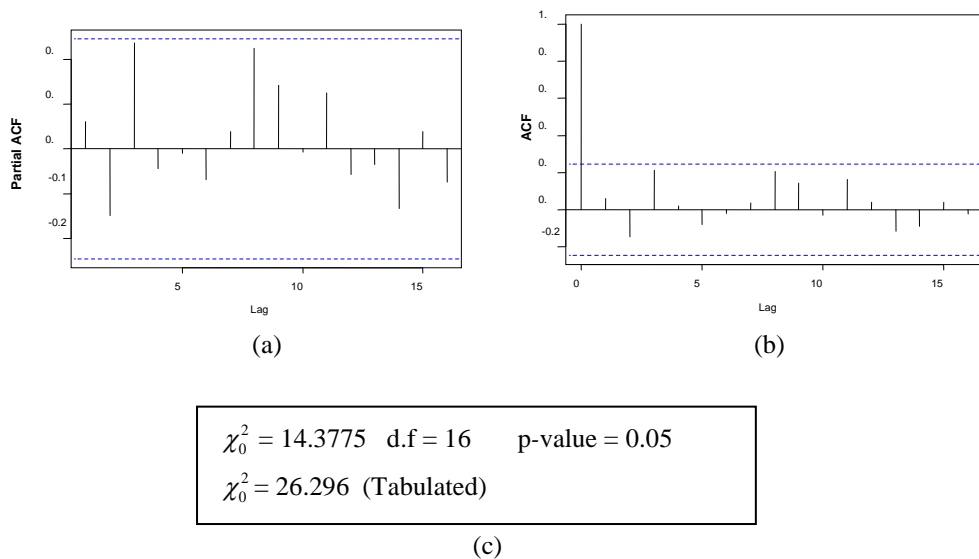

(a)   (b)

$$\chi_0^2 = 14.3775 \quad \text{d.f} = 16 \qquad \text{p-value} = 0.05$$
$$\chi_0^2 = 26.296 \ \ (\text{Tabulated})$$

(c)

Figure 4 (a) ACF Plot (b) PACF plot (c) Chi Square value of Box-Jung test of Residuals after fitting ARIMA(0, 1, 2) model to pass percentages of female candidates in Matriculation/HSLC examination in Assam (1951-2014) (Diagnostic Check)

5.1.5.   Forecasting

The estimated ARIMA(0, 1, 2) model was used to forecast the future pass percentages of female candidates in the Examination. The observed data can be treated as a realization from an infinite population of such time series that could have been generated by the stochastic process (Box et al. [21]). Therefore exact prediction of future values is not possible. Therefore, a prediction interval and the probability with which the future observation will lie within the interval can be provided (Bisgaard and Kulahci 2011). If $\sigma_e^2(\ell)$ be the variance of prediction errors, then 95% prediction interval is given by

$$\hat{z}_t \pm 1.96 \sigma_e\left(\ell\right) \tag{8}$$

We attempted to forecast the pass percentage of female candidates five periods ahead, i.e., for 2015, 2016, 2017, 2018 and 2019. The results are presented in Table 3. The comparison graph of forecasts, the 95% prediction intervals and actual observations are plotted in Figure 5. From visual inspection, it is obvious that the chosen model is reasonably good as the predicted series is very close to the observed series and within the 95% prediction limits.

Table 3 Forecast with 95% prediction intervals for pass % of female candidates
after fitting ARIMA(0, 1, 2) model to the data from 1951 to 2014

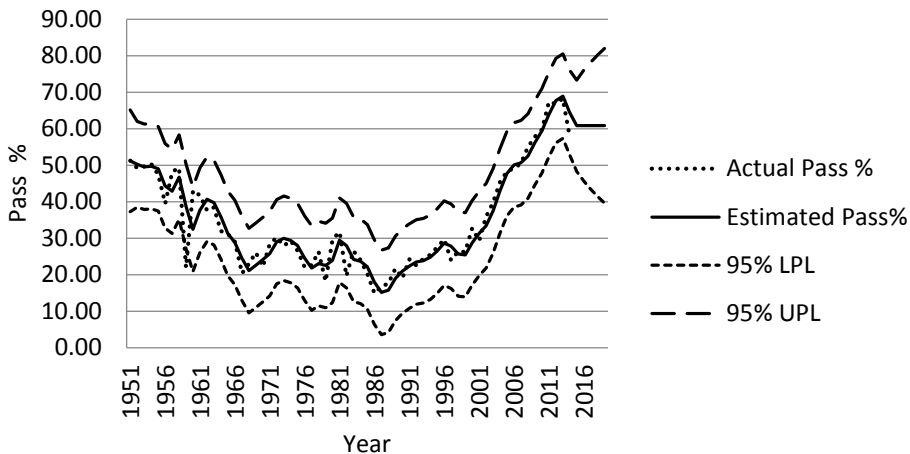| Year	Forecast (Pass %)	95% LPL	95% UPL
2015	64.49	53.33	75.65
2016	60.82	48.74	72.89
2017	60.82	46.19	75.44
2018	60.82	44.02	77.61
2019	60.82	42.10	79.54

Figure 5 Comparison graphs of actual Vs forecasted pass percentage values of female
candidates in Matriculation / HSLC Examination in Assam (1951-2019)
(With 95% Upper and Lower Prediction Limits)

5.2. Male candidates

In this section we consider pass percentage of male candidates in Matriculation/HSLC examination from 1951 to 2014 in Assam and carry out test of stationarity, model identification, model estimation, diagnostic checking and finally forecasting, exactly in the similar sequence as was done for the females in Section 5.1. After meticulously conducting all the steps it is found that the ARIMA(0, 1, 1) is best model for forecasting the pass percentage of male candidates based on the data from 1951 to 2014. The results are summarized in Figures 6-10 and Tables 4-6. From the study it is observed that for forecasting pass percentages of male candidates in the examination ARIMA(0, 1, 1) is useful model.

5.2.1.   Tests for stationarity

(a)   (b)

Dickey-Fuller= -1.5977,
p-value= 0.738.
Thus the null hypothesis
of non stationary is
accepted.

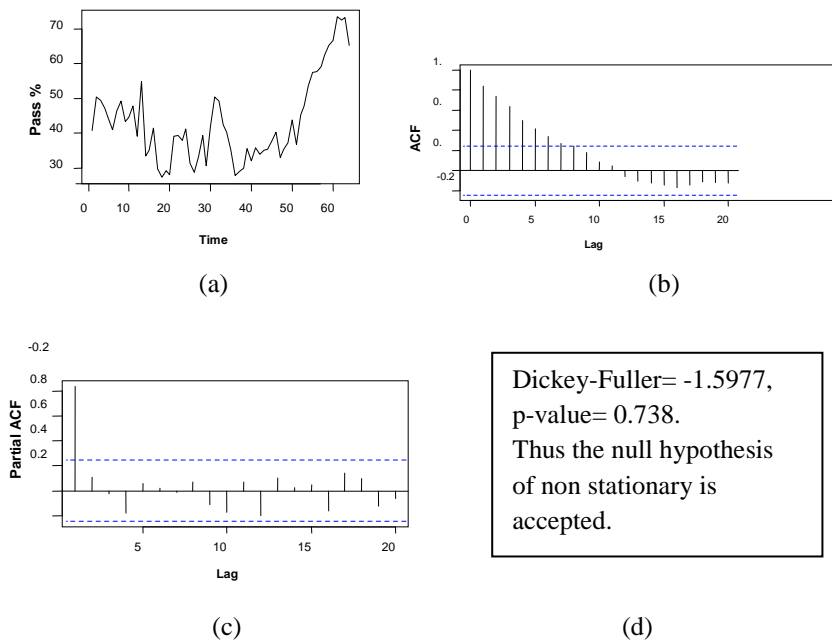(c)   (d)

Figure 6 (a) Time series plot (b) ACF plot (c) PACF plot and (d) Dickey-Fuller test result of pass percentages of male candidates in Matriculation/HSLC examination in Assam (1951-2014)
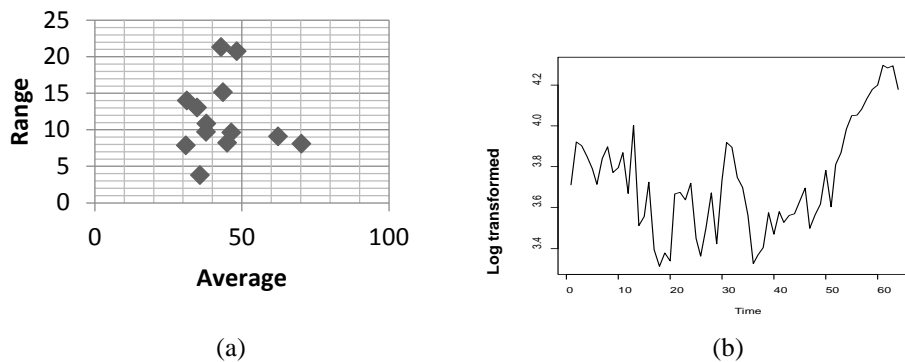

(a)   (b)

Figure 7 (a) Range-mean plot (b) log transformed series of pass percentage of male candidates in Matriculation/HSLC examination in Assam (1951-2014)
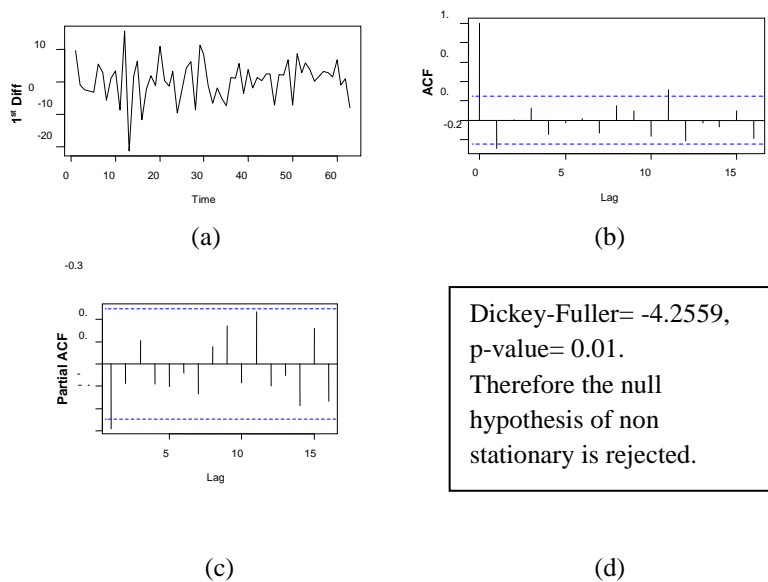
(a)

(b)

(c)

(d)

Dickey-Fuller= -4.2559, p-value= 0.01. Therefore the null hypothesis of non stationary is rejected.

Figure 8 (a) Time series plot (b) ACF plot (c) PACF plot and (d) Dickey-Fuller test reusly of first differenced time series (pass percentages of male Candidates in the Examination 1951-2014)

5.2.2.   Model identification

Table 4 Results of model identification for pass percentages of male candidates in the examination
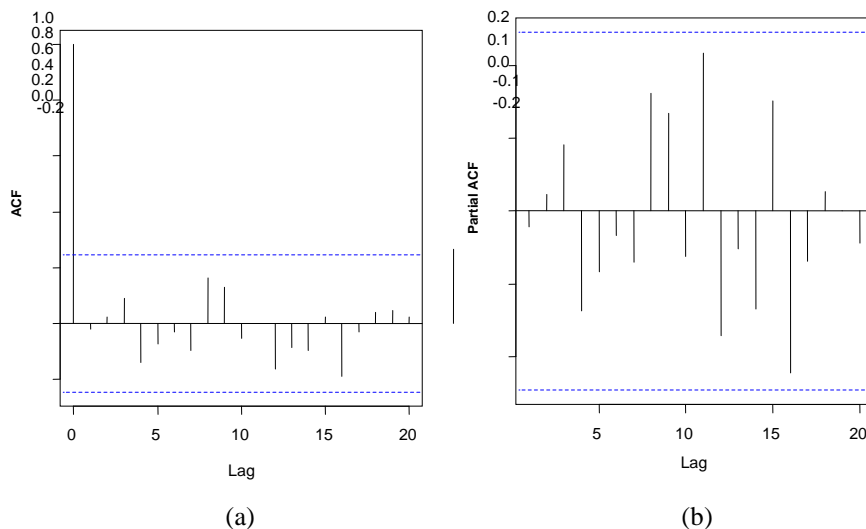
(1) ARIMA (1,1,0)		as PACF is zero after lag 1		
(2) ARIMA (0,1,1)		as ACF is zero after lag 1		
(3) ARIMA (1,1,1)		combination of models (1) and (2)		
Models with AIC, RMSE, MAE and MAPE (%)				
Model	AIC	RMSE	MAE	MAPE (%)
ARIMA(1,1,0)	405.24	5.79	4.68	11.61
ARIMA(0,1,1)	405.11	5.79	4.64	11.44
ARIMA(1,1,1)	406.96	5.79	4.66	11.53

5.2.3.   Model estimation

Table 5 Estimated coefficients for ARIMA(0, 1, 1) model (Pass percentages of male candidates)

Variables	Coefficients	Standard Errors	σ^2	Log likelihood	p-value
MA(1) θ	-0.3036	0.1168	34.03	-200.55	0.016

5.2.4. Diagnostic check

(a)   (b)

$$\chi_0^2 = 19.03 \quad \text{d.f} = 16 \qquad \text{p-value} = 0.05$$
$$\chi_0^2 = 26.296 \ \text{(Tabulated)}$$

Figure 9 (a) ACF Plot (b) PACF plot (c) Chi Square value of Box-Jung test of residual after fitting ARIMA(0,1,1) model to pass percentages of male candidates in the examination (1951-2014)

5.2.5. Forecasting

Table 6 Forecast with 95% prediction intervals for pass percentages of male candidates after fitting ARIMA(0, 1, 1) model to the data from 1951 to 2014

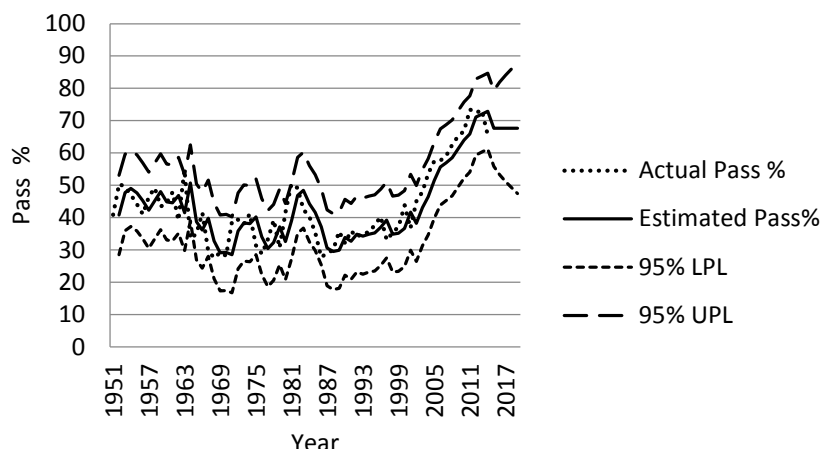| Year	Forecast (Pass %)	95% LPL	95% UPL
2015	67.64	56.20	79.07
2016	67.64	53.70	81.57
2017	67.64	51.59	83.69
2018	67.64	49.72	85.55
2019	67.64	48.03	87.24

Figure 10 Comparison graphs of actual vs forecasted pass percentage values of male
candidates in Matriculation/HSLC examination in Assam (1951-2019)
(with 95% upper and lower prediction limits)

6. Discussion

The Box-Jenkins (ARIMA) model technique has been used to study the pass percentages of
male and female candidates in Matriculation/HSLC examination in Assam for the period 1951 to
2014. It is found that ARIMA(0, 1, 2) and ARIMA(0, 1, 1) models are adequate for female and
male candidates respectively in our study. After estimating the parameters, we conducted diagnostic
test followed by residual analysis. We found that the assumptions of model adequacy are not
violated except in one case in Figure 8(b) but there too the outliers are not too far away (Bisgaard
and Kulahci 2011). Prediction of pass percentages for five years namely 2015, 2016, 2017, 2018
and 2019 in the examination along with their 95% prediction intervals have been forecasted.
According to the result of HSLC examination 2015 the pass percentages of female and male
candidates were found to be 59.9 and 66.5 percent respectively. The corresponding forecasted pass
percentage for females is 64.49 with 95 percent lower and upper prediction limits 53.33 and 75.65
respectively. The forecasted pass percentages of male candidates is 67.64 percent with lower and
upper prediction limits 56.20 and 79.07 respectively. In a time series, generated by stochastic
process, it is not always possible to exactly predict the output of the process in the future.
Therefore, a prediction interval and the probability with which the future observation will lie in that
interval are provided (Bisgaard and Kulahci 2011). It is observed that the actual pass percentages lie
in the estimated 95% prediction intervals for both female and male candidates.

The present study on HSLC examination results of Assam depicts that though the pass
percentages for male candidates were better than those of female candidates, yet the results were
not satisfactory for both male and female candidates over the period. In Figure 5 and Figure 10, we
have noticed a decreasing trend in near future. Since completed secondary education serves as a
common denominator for progressing towards further education, the outcomes of our investigation
may be helpful for the policy makers for adopting appropriate strategic planning for investing in
quality secondary education which will lead to more rapid and sustainable growth and development
of our young generation.

Acknowledgments

References

Aggarwal JC. Development and Planning of Modern Education, Ahmedabad, Gujarat, India Vikas Publishing House Pvt. Ltd.; 1993.

ARIMA models for time series forecasting. Available from: people.duke.edu~rnau/411arim3.htm. [accessed on March 12, 2014].

Available in the old news paper cell of the District Library in Guwahati, Assam,India.

Bisgaard S and Kulahci M. Time Series Analysis and Forecasting by Example, Hoboken, New Jersey, Wiley, A John Wiley & Sons, INC.; 2011.

Board of Secondary Education, Assam Results of High School Leaving Certificate Examination & Assam High Madrassa Examination. Assam. (1991-2014).

Box GEP, Jenkins GM and Reinsel GC. Time Series Analysis, Forecasting and Control, Third Edition, Delhi, India Pearson Education; 2003.

Chen C. An Integrated Enrollment Forecast Model. Association for Institutional Research. IR Applications, 2008; 15:1-17.

Deb Roy T and Das KK. Time Series Analysis of Dibrugarh Air Temperature. Journal of Atmospheric and Earth Environment. 2012; 1(1), 30-34.

Document of the World Bank. Secondary Education in India Universalizing Opportunity. Human Development Unit South Asia Region. 2009. Available from: http://datatopics. worldbank.org/hnp/files/edstats/INDstu09b.pdf [accessed March 12, 2014]

Expanding Opportunities and Building Competencies for Young People, A New Agenda for Secondary Education. The World Bank. 2005. Available from: http://siteresources. worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-10990. [accessed January 15, 2013].

Gauhati University Examination Results, Compiled in the Statistical Unit of the Gauhati University under supervision of the Statisticians, Gauhati University, Assam. (1948-1968).

Government of India, Ministry of Human Resource Development, Department of Education, Available from: http://www.teindia.nic.in/mhrd/50yrsedu/home.htm [accessed March 13, 2014].

Gujarati ND and Sangeetha G. Basic Econometrics, New Delhi, Tata McGraw Hill Education private Limited; 2007.

High School Leaving Certificate, India, 2007. Wikipedia, the free encyclopedia. 2007. Available from: http://en.wikipedia.org [accessed March 12, 2014].

Madore O. Post-Secondary Education: An Imperative for Canada's Future. November, 1992. Available from: http://publications.gc.ca/Collection-R/LoPBdP/BP/bp319-e.htm BP-319 E, Publication.gc.ca [accessed March 10, 2014]

Mahanta NN. Secondary Education Issues and Problems, Guwahati-781022. Kashyap Publishing House; 1997.

Negron, A. ARIMA Models in a Nutshell, (2012-2015) Available from: http://altons.github.io /r/2013/05/12/arima-models-in-a-nutshell/ [accessed March 12, 2014)].

Pursuing Gender Equality through the Millennium Development Goals in Asia and the Pacific. Asia pacific MDG Study Series. (ADB, 2006).

Human Development Unit South Asia Region. Secondary education in India: Universalizing opportunity. 2009 [cited 2013 Jan 15]; Available from http://datatopics.worldbank.org/hnp/files/edstats/INDstu09b.pdf.

Shitan M, Karmoker PK, Lerd NY. Time series modeling and forecasting of Ampang line passenger ridership in Malaysia, Pak J Stat., 2014; 30(3), 385-396.

Class-X public examination-a burden or boon? The Assam Tribune. 2009 Nov 11.