Forecasting the Patterns and Trends Age-Specific Fertility Rate in South Asia

Nasriyah Arbu[a,b], Apiradee Lim[a*], and Cherdchai Me-ead[c]

[a]Department of Mathematics and Computer Science, Faculty of Science and Technology,
Prince of Songkla University, Pattani, Thailand.
[b]Centre of Excellence in Mathematics, Commission on Higher Education,
Ratchathewi, Bangkok, Thailand.
[c]Division of Statistics, Faculty of Science, Maejo University, Chiang Mai, Thailand.
***Corresponding author, Email address: apiradee.s@psu.ac.th**

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: The total fertility rate (TFR) in South Asia has decreased remarkably over the past three decades. The decline is projected to continue in the coming years mostly due to the significant changes occurring in age specific fertility rate (ASFRs). This study aims to investigate ASFRs trends and forecast the ASFRs of India, Bangladesh, and Pakistan from 2020 to 2100. The ASFRs in South Asia data separated by 7 age groups with 5-year interval from 15 to 49 years between 1990 and 2019 were downloaded from the US Census Bureau's website. Linear regression models were used to investigate ASFRs patterns from data in that period and ASFRs data in the years 2010-2019 of each age group and were used to forecast the ASFRs in 2020 to 2100. The forecasted results show that ASFRs in India and Bangladesh in age group 15-24 years have steep declining trends whereas ASFRs in age group 25-29 in both countries and age group 45-49 years in India have gradually decreasing ASFRs trends. Age group 30-44 years in Bangladesh have slightly decreasing ASFRs trends. Pakistan with previously high fertility rates is experiencing gradually declining ASFRs in all age groups. In conclusion, the change in fertility in these three countries from low stable population to high population will occur within next 40 years from 2020.

Keywords: age-specific fertility rates, linear regression model, trends in fertility, fertility patterns

1. INTRODUCTION

Future size and composition of population are determined by fertility, mortality, and migration rates. Among these factors, fertility prediction plays a vital role due to its importance in determining and control population growth and dynamics [1]. It is a major factor for biological substitution and maintenance of mankind development [2] and also it has a big impact on the socioeconomic condition of a country [3]. Hence, in forecasting of population, fertility remains a key component. The most index usually used to predict fertility and measure population growth is the TFR [4]. The TFR measures the average number of births per woman. However, it does not provide information on the actual number of births [5]. On the other hand, the ASFRs is a measure of the annual number of births by women of a specific age (15-49 years) per 1,000 women. Fertility rates may vary among different age groups. Hence, it is pertinent to consider a mathematical model that can describe the fertility profile more accurately and precisely as it will allow the fertility forecasts to be reasonable and transparent [6].

According to the World Fertility Report, virtually all regions in Asia have experienced fertility decline in past several decades. One of them is South Asia which has decreased from 4.3 births per woman in 1990 to 2.4 in 2019 but their fertility rates remain above replacement birth rate at 2.1 births per woman. Fertility decline drives a slow rate of world's population growth. However, in 2019, Asia had 4.8 billion people, 62.8% of the world's population [7] and almost 22.43% live in India, Pakistan and Bangladesh which are among the top ten populous countries around the world (India, Pakistan, and Bangladesh in the second, fifth and seventh ranks) [3]. It has been predicted that India will overtake China as the world's most populous country by 2027 [7].

The ASFRs in age group 20-24 has been and remains the highest in India and Bangladesh [8][9] while ASFRs in age group 25-29 is highest in Pakistan [10]. There has been a reduction in ASFRs for all these countries. This reduction is attributed to different reasons. The most consistent reasons are transitions to lower fertility which are driven by two direct causes: a rise in marriage ages, thus reducing teenage childbearing, increased family size limitation by contraception and abortion [11]. There is enough evidence of continuous declining in the TFRs of South Asia countries [12].

The likelihood of having a child differs based on a woman's age. Therefore, forecasting ASFRs is useful for indicating which age groups have more effect on population change. There have been a small number of studies which have investigated and forecasted the ASFRs trend, including that conducted by [13] which forecasted population growth in West Africa using linear regression. Hence, with existing valid methods, this study aims to forecast ASFRs in the most three populous countries in South Asia. This contribution helps better understanding not only fertility transition over decades but also changing patterns in the annual number of births by specific-aged-woman in these countries.

2. Materials and Methods

A. Data source

Data relating to the ASFRs per 1,000 women in the form of median rates were retrieved from the US Census Bureau website. The US Census Bureau released the first complete set of estimates and projections for over 200 countries in the 1980s. The quality of the data is ensured by the demographic information on which it is based being gathered from many sources, such as censuses, surveys, vital registration, and administrative records from a variety of sources. The process of estimation and projection includes data evaluation, parameter estimation and making assumption about future change. The ASFRs data downloaded for this study covered a period of 30 years from 1990 to 2019. The ASFRs data consisted of 3 countries as follows: India, Bangladesh, and Pakistan. Basically, the age pattern of reproductive women was distributed into 7 age groups from ages 15 to 49 years with 5-years intervals at: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49 years. Thus, the data contained ASFRs of seven age groups and 3 countries for 30 years, with 630 observations and no missing values.

B. Statistical analysis

(1). Long term Limit

The minimum, maximum and median ASFRs of women in each age group and each country were calculated. These values were used to set up the limits for each age group for all countries to forecast their ASFRs using the following equation:

$$\text{Limit}=\frac{(\text{median(ASFR)}+\text{minimum(ASFR)})}{2} \tag{1}$$

Based on the data described, above the medians, and the long-term limits of ASFRs for each age group were calculated. Among almost ASFRs all countries and all age groups, the lowest and highest of limits were 1.4 and 200.7 per 1,000 women. These limits were used to constrain the long-term forecasting of the ASFRs. The ASFRs were first assessed for stability of variance and the necessity of data transformation, by fitting linear regression model to the entire ASFRs using year, age- group and country as predictors. The normal quantile-quantile (Q-Q) plots of the studentized residuals were used to evaluate the normality of the distribution. Thereafter, 3rd root transformations were applied to stabilize the variance of the ASFRs 30-year data.

Table. 1. The minimum, medium and maximum of ASFRs in each age group for individual country together with lower and upper limits

Countries	Age Group						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Median	8.7	122.4	110.6	50.0	16.1	4.5	1.4
Lower limits	4.0	63.2	98.6	72.5	37.7	11.0	1.5
Upper limits	85.9	183.2	200.7	171.6	111.6	54.7	29.5

(2). Simple linear regression

The transformed ASFRs data in the years 1990 to 2019 were plotted to illustrate the population distribution for each age group and each country. Simple linear regression was used to forecast the ASFRs in each age group and each country according to the following equation:

$$Y_t = a + b_t + e_t \tag{2}$$

where t is year, Y_t is age-specific fertility rate of each age group in year t of individual country, a is constant, b_t is slope, and e_t is error term.

There were fluctuations in fertility trends in some age groups for India, Bangladesh, and Pakistan from 1990 and 2009. Therefore, only the data from 2010 to 2019 were used to create the linear model to forecast the ASFRs for each age group and each country. Using limits as the constraints, ASFRs were forecasted for the years 2020 to 2100 based on the data for the years 2010 to 2019. In this study, the ASFR was forecasted until 2100 as the world population is forecasted to almost stop increasing at the end of this century. By applying linear regression model, the study assumes that the forecasted trends of ASFRs will remain stable in the long-term unless there is a very significant deviation in demography. Thus, the linear projections will continue until these long-term fertility limits are reached and will then remain constant. All statistical analyses and the graphical displays were carried out using the R program, version 3.6.2.

(3). Goodness of fit test

To evaluate the forecasting performance of the model, post sample forecast accuracy criterion. The goodness of fit was measured from the r-squared and standard deviations of all age groups in the three countries to determine the model accuracy.

(4). Coefficients

The coefficients of the unknown population parameters were also computed. The sign of each coefficient indicates a different meaning. A positive coefficient sign indicates an increase in the ASFRs while a negative coefficient indicates a decrease in the ASFRs.

(5). Data transformation

Besides, to ensure that the graphical representation in the Q-Q plots was not misinterpreted, a Box-Cox transformation was used to confirm the necessity of data transformation. The Box-Cox equation is given as follows:

$$Y_t^\lambda = \begin{cases} (Y_t^\lambda - 1)/\lambda \;; & \lambda \neq 0 \\ log(Y_t) \;; & \lambda = 0 \end{cases} \tag{3}$$

where Y_t is the data in time t and λ is a transformation parameter.

The Box-Cox transformation has been widely used in a variety of fields, especially in economics and finance. It is capable of fulfilling the basic assumptions of linearity, normality and homoscedasticity as recommended by Box and Cox who modified a family of power transformations first presented by Tukey [15] to cover discontinuity of the original function as shown below:

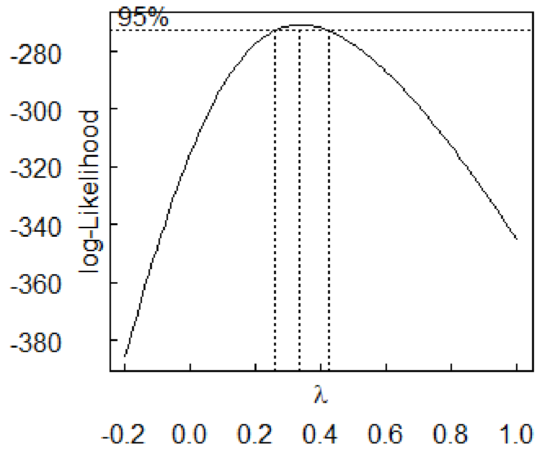

Figure 1: Box-Cox transformation plot

The X-axis of the Box-Cox plot indicates λ values representing a power transformation of data. If the confidence interval covers optimal λ value of 1, a data transformation is then unnecessary, otherwise a transformation is needed. Typically, the rounded optimal value for λ falling within the confidence limits that agrees with an understandable power transformation is used [14][15], for example the cube root (λ = 1/3), the square root (λ = 0.5) or the natural log (λ = 0). In this data, λ was equal to 0.33. Thus, the best transformation to stabilized variance was third root.

3. Result

The median ASFRs in each age group of all countries during the 30-year period are shown in Table 1. The results showed that Bangladesh had the highest ASFRs in the 15-19 age group with an ASFRs of 136.8 births per 1,000 women, whereas Pakistan had the lowest rate of 26.9 per 1,000 women. In the 20-24 age group, India had the highest ASFRs, followed by Pakistan while the lowest ASFRs was in Bangladesh. Pakistan had the highest ASFRs in the 25-29 age group, where the lowest ASFRs was found in Bangladesh. Among the 30-34 and 35-39 age group, Pakistan had the highest ASFRs with the lowest ASFRs in India. Across all the age groups, the 45-49 group had the lowest ASFRs.

Due to assumptions of linear regression, the model requires normally distributed residuals. The normal quantile-quantile (Q-Q) plot of the residuals were first examined to compare two theoretical distributions. The quantile of the second distribution is represented on the vertical axis while the quantile of the first distribution is on the horizontal axis**.** When the Q**-**Q plot follows its trend line and the two distributions are identical, it indicates that the data are normally distributed**.** With awareness of a potential outlier that may influence the regression model when analysing time series data, the Q**-**Q plot of studentized residuals was also investigated to provide more precise analysis **[**9**].**

The ASFRs were first assessed for stability of variance and the necessity of data transformation, by fitting linear regression model to the entire ASFRs using year, age-group, and country as predictors. The normal quantile-quantile (Q-Q) plots of the studentized residuals were used to evaluate the normality of the distribution. Thereafter, 3rd root transformations were applied to stabilize the variance of the ASFRs 10-year data.

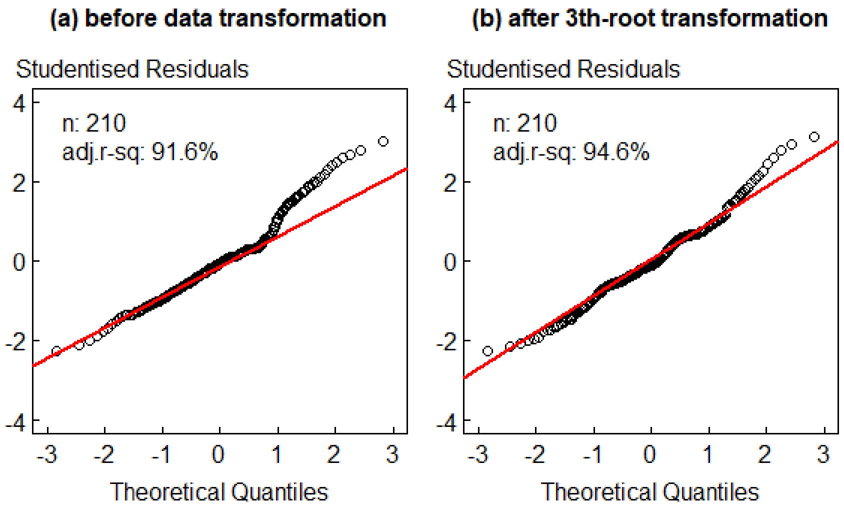

Figure 2: Comparison normal Q-Q plots of the studentized residuals between before (a) and after ASFRs transformation (b)

The Figure 3 shows the results of ASFRs forecasting in each reproductive age group of the three countries with the high levels of fertility. The solid-coloured lines illustrate the trend of ASFRs in each age group between 1990 and 2019. The superimposed solid black lines represent the values produced by fitting separate linear models to the data for the ten most recent years (2010-2019) in each age group. The dotted coloured lines are forecasts given by projecting these model values from the year 2020 to 2100 in each age group and country. R-squared values as percentages from each model is presented in the legend of each plot. The solid-coloured circles at the end of each forecasting line are the limit in that age group.

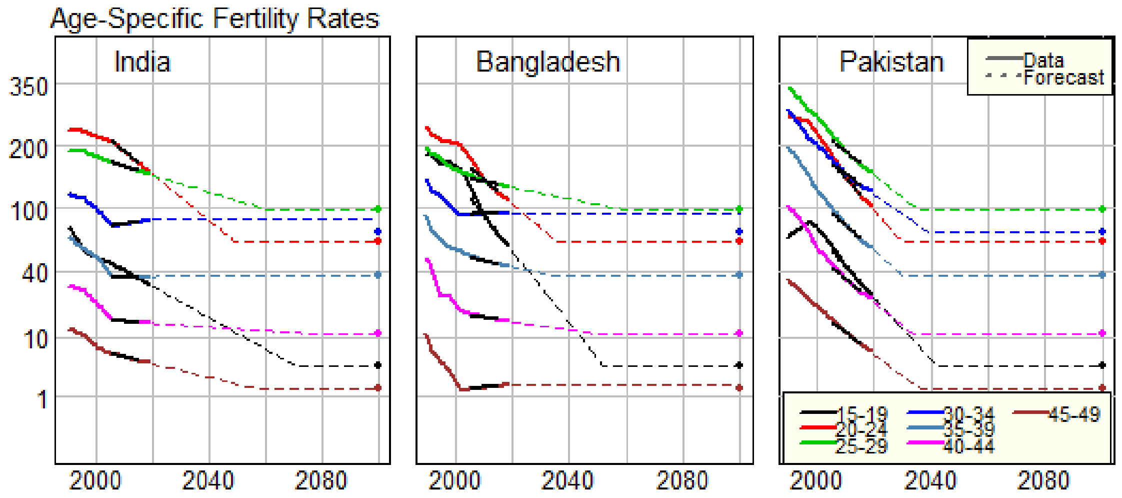

Figure 3: The forecasting trend of ASFRs in each age group of the countries with high level of fertility

The forecasted trends of ASFRs in this group had hit their own individual limits for all age groups. The forecasted ASFRs for all the countries in this group had decreasing trends with a sharply decreasing trend being observed in the 15-19 age group in Bangladesh and Pakistan. The only exception to this trend was in the 45-49 age group in Bangladesh which had an increasing trend. In India, the 30-39 age group had an ASFRs exceeding the limit forecast as did the 30-34 age group in Bangladesh. The median durations to reach the limits in the groups with decreasing trends in their ASFRs in Pakistan, Bangladesh, and India were 18, 32 and 32 years, respectively while the median duration to reach the limit in the only group with an increasing trend (the 45-49 age group in Bangladesh) was15 years.

Table. 2. R-square from regression model for each age group and each country

	India			Bangladesh			Pakistan		
Age group	Intercept	R-squared	Std. Error	Intercept	R-squared	Std. Error	Intercept	R-squared	Std. Error
15-19	-0.03021	0.99800	0.00045	-0.06912	0.99510	0.00172	-0.06023	0.99460	0.00157
20-24	-0.04428	0.99780	0.00073	-0.04668	0.99290	0.00140	-0.06200	0.99290	0.00186
25-29	-0.01636	0.99660	0.00034	-0.01107	0.98910	0.00041	-0.04405	0.99130	0.00146
30-34	0.00843	0.99440	0.00022	0.00202	0.97400	0.00012	-0.03825	0.99110	0.00128
35-39	0.00199	0.98180	0.00010	-0.01117	0.98410	0.00050	-0.04767	0.99200	0.00152
40-44	-0.00354	0.97600	0.00020	-0.00776	0.98460	0.00034	-0.04437	0.99370	0.00125
45-49	-0.01287	0.99480	0.00033	0.00422	0.75540	0.00085	-0.04161	0.99590	0.00094

From Table 2, overall, India had the highest and lowest r-squared value of 99.80% and 75.54%, respectively occurring in age groups 15-19 and 40-44. Regarding the Std. Error, Pakistan had the highest value (0.00186) occurring in age group 20-24 while India had the lowest (0.0001) from age group 35-39. In India, an increase in coefficient values occurred in age groups 30-34 and 35-39. However, the decrease occurred in all other age groups. In Bangladesh, the increase in coefficient values were recorded for age groups 30-34 and 45-49 whereas there were decreasing trends in all other age groups. In contrast to India and Bangladesh, the coefficients values in all age groups in Pakistan had decreasing trends.

4. Discussion

In this study, an approach to forecasting the ASFRs is proposed for the 3 most populous countries in South Asia using simple linear regression models, which had R-squared values ranging from 75.54 - 99.80 %. The model had reasonably good fit to the data as most of the ASFRs in the 3 countries and all the age groups followed linear trends. This good fit is consistent with the results of the study conducted by Owusu et al [13]. However, this forecasting method assumes constant conditions, such as there being no significant change in family planning policies, no extreme disasters, or major economic changes.

The results from this study revealed that the 3 countries with high fertility levels had decreasing forecasted trends of ASFRs. Sharply decreasing trends were observed for women in age groups 15-19 and 20-24 in India and Bangladesh but in all age groups in Pakistan. Among the three countries, India was the first country to show a decline in fertility beginning in the 1970s [11]. Population growth has long been a concern of the Indian government and they believed that similar to developed nations, a rise in the standard of living and industrialization can only be achieved when there is a reduction in the population growth rate. Thus, the government began a massive program to lower the birth rate from 41 per 1,000 to a target of 20-25 per 1,000 by the mid-1970s [17]. Most of age groups in India had a decreasing fertility trends with a slightly increasing trends in age groups 30-34 and 35-39 years. Bangladesh has experienced a dramatic fertility decline since 1985 and the reasons for this rapid decline have been attributed primarily to family planning programs, expanding female education and other socio-economic factors [18]. The forecast result showed a decreasing trend in all age groups with a sharply decreasing trend in teenagers (15-19), working age (20-24).

The more-direct causes of declining fertility in Pakistan includes existing desires for smaller families, increasing use of modern contraception for the purpose of limiting family size and a reduction in the social, cultural, and costs of contraception [19]. Thus, all age groups had a sharply decreasing trend. Among the 3 countries, there were decreasing trends in fertility in all age groups with sharp decreases among teenagers and young women in all countries.

However, age group 44-49 in Bangladesh had a slightly increasing trend. This can be explained by the transition to lower fertility being driven by a rise in marriage age. This in turn leads to a reduction in teenage childbearing [9] and a tendency for family size to be limited due to factors such as the use of contraception and abortion, and a longer time being spent in education, particularly in the context of rapid economic transformation in developing countries [7].

The significant contribution of this study is the forecasting of fertility in all age groups separated for each country, which provides a clearer picture of fertility transition in each country. More detailed information relating to fertility is useful in setting proper policies and in economic and social planning. However, the findings of this study are limited by the linear model employed not being suitable for some age groups with high fluctuations in their ASFRs which do not follow a linear trend.

5. Conclusion

In conclusion, the structure of the population in the three South Asian countries studied has dramatically decreasing trend across almost all age groups. Therefore, the structure of the population will change over coming decades and polices related to the labour force and education should be emphasized in countries with these countries.

Acknowledgement

We are grateful to Emeritus Professor Don McNeil from Macquarie University, Sydney, Australia, for his initiative and guidance on forecasting techniques used in this statistical modelling. This study was funded by the Graduate School, Prince of Songkla University, Thailand was partially funded by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand.

Reference

1. Verma, A., Singh, G. P., & Singh, A. (2018, April). Probabilistic Projections of the Age- Specific Fertility Rates in India. In PAA 2018 Annual Meeting. PAA.
2. General Statistics Office. (2010). The 2009 Vietnam population and housing census: completed results.
3. Rayhan, I., Akter, K., & Islam, M. S. (2018). Determinants of fertility rate decline in the south Asian countries: A panel data approach. International Journal of Development Research, 8(07), 21583-21589.
4. Shitan, M., & Ng, Y. L. (2015). Forecasting the Total Fertility Rate in Malaysia. Pakistan Journal of Statistics, 31(5).
5. Vanella, P., & Deschermeier, P. (2019). A Principal Component Simulation of Age-Specific Fertility–Impacts of Family and Social Policy on Reproductive Behavior in Germany. Population Review, 58(1).
6. Hanafiah, H., & Jemain, A. A. (2013, November). Structural modeling of age specific fertility curves in Peninsular Malaysia: An approach of Lee Carter method. In AIP Conference Proceedings (Vol. 1571, No. 1, pp. 1062-1068). American Institute of Physics.
7. United Nations. (2019). World population prospects 2019: highlights. Department of Economic and Social Affairs, Population Division.
8. Ponnapalli, K. M., & Soren, R. K. (2020). Estimation of Age-Specific Fertility Rates for Districts in India from Median Age of the total Population: A New Approach. Demography India, 49(1), 4-25.
9. Streatfield, P. K., Kamal, N., Ahsan, K. Z., & Nahar, Q. (2015). Early marriage in Bangladesh: Not as early as it appears. Asian Population Studies, 11(1), 94-110.
10. Akhtar, J. A. N. M., & Tahir, M. H. (2009). Pak. J. Statist. 2009 Vol. 25 (3), 251-263 Reproductivity and Age-Specific Fertility Rates in Pakistan after 1981. Pak. J. Statist, 25(3), 251-263.
11. Sathar, Z. A., & Phillips, J. F. (2001). Fertility Transition in South Asia. Oxford University Press.
12. Rele, J. R. (1992). Fertility levels and trends in South Asia: An assessment and prospects. Genus, 133-154.
13. Owusu, B. A., Lim, A., Makaje, N., Sama, A., Owusu, B. E., & Arbu, N. (2018). Age-Specific Fertility Rate Projections in West Africa. Journal of Population and Social Studies [JPSS], 26(2), 119-127.
14. Box, G. E., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26(2), 211-243.
15. Sakia, R. M. (1992). The Box-Cox transformation technique: a review. Journal of the Royal Statistical Society: Series D (The Statistician), 41(2), 169-178.
16. Lee, B., Tongkumchum, P., & McNeil, D. (2018). Forecasting monthly world tuna prices with a plausible approach.
17. Ledbetter, R. (1984). Thirty years of family planning in India. Asian Survey, 24(7), 736-758.
18. Bora, J., Saikia, N., & Lutz, W. (2019). Revisiting the causes of fertility decline in Bangladesh: Family planning program or female education?
19. Sathar, Z. A., & Casterline, J. B. (2001). The Onset of Fertility Transition in Pakistan. Pakistan Development Review, 40(3), 237-238.