

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

Nasriyah Arbu^a, Apiradee Lim^b, Cherdchai Me-ead^c

^{a,b}Department of Mathematics and Computer Science, Faculty of Science and Technology,
Prince of Songkla University, Pattani, Thailand.

^aCentre of Excellence in Mathematics, Commission on Higher Education,
Ratchathewi, Bangkok, Thailand.

^cDivision 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 (ASFR). This study aims to investigate ASFR trends and forecast the ASFR 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 ASFR patterns from data in that period and ASFR data in the years 2010-2019 of each age group 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 ASFR trends. Age group 30-44 years in Bangladesh have slightly decreasing ASFR trends. Pakistan with previously high fertility rates are 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 ASFR 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] [7]. It has been predicted that India will overtake China as the world's most populous country by 2027 [7].

The ASFR in age group 20-24 has been and remains the highest in India and Bangladesh while ASFR in age group 25-29 is highest in Pakistan [8]. There has been a reduction in ASFR for all of 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; and increased family size limitation by contraception and abortion [9]. There is enough evidence of continuous declining in the TFRs of South Asia countries.

The likelihood of having a child differs based on a woman's age. Therefore, forecasting ASFR 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 ASFR trend, including that conducted by [10] which forecasted population growth in West Africa using linear regression. Hence, with existing valid methods, this study aims to forecast ASFR 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 ASFR 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 ASFR data downloaded for this study covered a period of 30 years from 1990 to 2019. The ASFR data consisted of 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:

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

Based on the data described, above the medians, and the long-term limits of ASFR for each age group were calculated. Among almost ASFR 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 ASFR. 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 ASFR 30-year data.

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

ountries	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
L. limits	4.0	63.2	98.6	72.5	37.7	11.0	1.5
U. limits	85.9	183.2	200.7	171.6	111.6	54.7	29.5

*L.limits is Lower imit and U.limit is Upper limit

(2). Simple linear regression

Therefore, the ASRF of the recent decade, 2010-2019, which were relatively low volatility in each individual nation were transformed with 3rd-root and used to fit the simple linear regression model in order for projecting long-term ASFR trends in year 2020 to 2100 of individual country as following equation:

$$Y_t = a + b_t + e_t \quad (1)$$

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 ASFR 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). Data transformation

Besides, to ensure that the graphical representation in the normal quantile-quantile Q-Q plots of the residuals was not misinterpreted, Box-Cox transformation was used to confirm the necessity of data transformation.

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 [11] who modified a family of power transformations first presented by [12] to cover discontinuity of the original function as shown below:

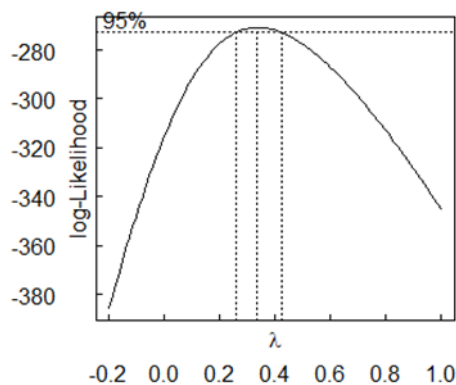


Figure 1: Box-Cox transformation plot

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 (Box and [11][13] for example the cube root ($\lambda = 1/3$), the square root ($\lambda = 0.5$) or the natural log ($\lambda = 0$)). 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 ASFR in the 15-19 age group with an ASFR 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 ASFR, followed by Pakistan while the lowest ASFR was in Bangladesh. Pakistan had the highest ASFR in the 25-29 age group, where the lowest ASFR was found in Bangladesh. Among the 30-34 and 35-39 age group, Pakistan had the highest ASFR with the lowest ASFRs in India. Across all the age groups, the 45-49 group had the lowest ASFR.

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 analyzing 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 ASFR 10-year data.

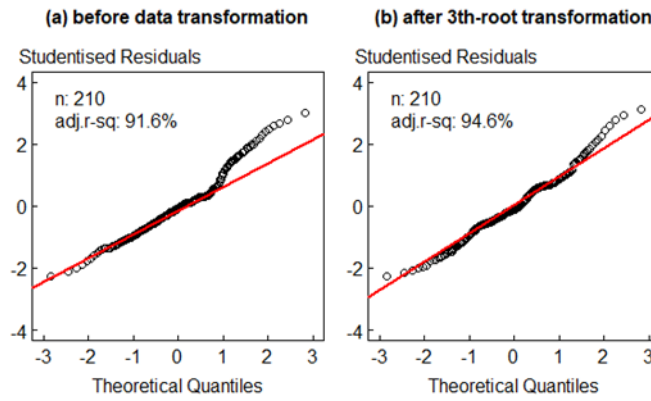


Figure 2: Comparison normal Q-Q plots of the studentized residuals between before (a) and after ASFRs transformation (b)

The Figure 2 shows the results of ASFR forecasting in each reproductive age group of the three countries with the high levels of fertility. The solid-colored lines illustrate the trend of ASFR 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 colored 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 are presented in the legend of each plot. The solid-colored circles at the end of each forecasting line are the limit in that age group.

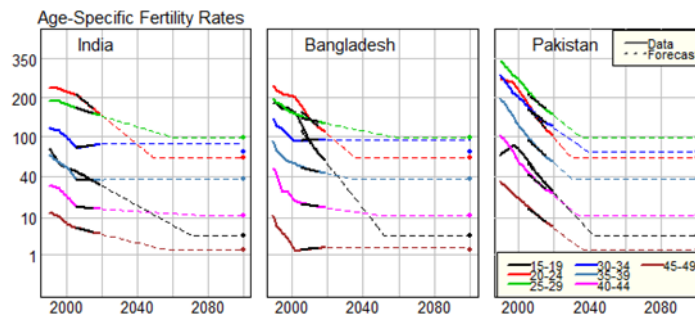


Figure 3: The forecasting trend of ASFR in each age group of the countries with low level of fertility

The forecasted trends of ASFR 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 ASFR 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 ASFR 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) was 15 years.

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

Age group	R-square (%)		
	India	Bangladesh	Pakistan
Age 15-19	99.86	99.61	99.58
Age 20-24	99.82	99.32	99.36
Age 25-29	99.65	98.81	99.13
Age 30-34	99.35	97.06	99.11
Age 35-39	97.95	98.27	99.25
Age 40-44	97.32	98.32	99.46

Age 45-49	99.48	72.39	99.69
-----------	-------	-------	-------

From Table 2, overall, India had the highest and lowest r-squared value of 99.86% and 72.39.32%, respectively occurring in age groups 15-19 and 40-44. Regarding the RMSE value, Pakistan had the highest value (0.0366) occurring in age group 15-19 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 paper an approach to forecasting the ASFR is proposed for the 3 most populous countries in South Asia using simple linear regression models, which had R-squared values ranging from 97.32 - 99.86% with the exception of one age groups in Bangladesh with low R-squared values of 72.39 %. The models had reasonably good fit to the data as most of the ASFRs in the three countries and all the age groups followed linear trends, except for the 45-49 age groups in Bangladesh, for which the ASFR slightly departed from a linear trend. This good fit is consistent with the results of the study conducted by [10]. 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. However, in the past two decades young women have enjoyed a more open society with a greater chance of leaving their homes to work in other cities where they may live with their partners [14] and might be due to young couples delaying the birth of their first child or increasing the spacing between births, as well as the cessation of childbearing among older women especially in India [15].

Among countries with a high level of fertility, there were decreasing trends in fertility in all age groups with sharp decreases among teenagers and young women in all countries. Except Bangladesh, with a slightly increasing trend in the 45-49 age group. This can be explained by the transition to lower fertility being driven by a rise in marriage age [16]. This in turn leads to a reduction in teenage childbearing²⁰ and a tendency for family size to be limited (ref) 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. The significant contribution of this study is the forecasting of fertility in all age groups separately 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 ASFR which do not follow a linear trend.

5. Conclusion

In conclusion, the structure of the population in the three South Asia countries studied has changed from a younger to an older population particularly in countries, with high fertility rates where fertility has decreased among younger age groups while increasing in older age groups as Bangladesh. Meanwhile, in Pakistan fertility rates show a dramatically decreasing trend across all age groups in countries with previously high rates of fertility. Therefore, the structure of the population will change over coming decades and all those countries should focus their policies on health care and the well-being of the elderly whereas policies related to the labor force, family planning and education should be emphasized in countries with high rates of fertility.

Acknowledgement

We are grateful to Emeritus Professor Dr. Don McNeil from Macquarie University, Sydney, Australia, and Assoc. Prof. Dr. Apiradee Sae-lim my advisor for initiative and guidance on forecasting techniques used in this statistical modeling. Moreover, we wish to thankful to supported by the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand..

References

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. Lutz, W., Testa, M. R., & Penn, D. J. (2006). Population density is a key factor in declining human fertility. *Population and Environment*, 28(2), 69-81.
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. Ali, S. M., Hussain, J., & Chaudhry, M. A. (2001). Fertility Transition in Pakistan: Evidence from Census [with Comments]. *The Pakistan Development Review*, 537-550.

5. Keyfitz, N. (1981). The limits of population forecasting. *Population and development review*, 579-593.
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. *World population ageing 2019*. New York: United Nations; 2019a.
8. 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.
9. Sathar, Z. A., & Phillips, J. F. (2001). *Fertility Transition in South Asia*. Oxford University Press.
10. Lee, B., Tongkumchum, P., & McNeil, D. (2018). Forecasting monthly world tuna prices with a plausible approach.
11. 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.
12. Box, G.E.P. and Cox, D.R. 1964. An analysis of transformations. *J. R. Stat. Soc.*
13. *Series B*. 26, 211-252.
14. Box, G.E.P. and Jenkins, G.M. 1970. *Time series analysis: forecasting and control*.
15. San Francisco: Holden-Day.
16. Sakia, R. M. 1992. The Box-Cox transformation technique: a review. *The Statistician*.
17. 41, 169-178.
18. Visalakshi, S., & Geetha, R. (2018). Modelling age specific fertility rate in India through fertility curves. *Bulletin of Pure & Applied Sciences-Mathematics and Statistics*, 37(2), 391-405.
19. Nasir, J. A., Tahir, M. H., & Riaz, M. (2010). Measuring and modeling the fertility profile of indigenous people in Pakistan: A study of the Arians. *Pakistan Journal of Commerce and Social Sciences (PJCSS)*, 4(2), 132-146.
20. 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.