Published in AJSE, Vol:23, Issue: 1
Received on 21st January 2024
Revised on 1st March 2024
Published on 25th April 2024

AIUB JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 1608 – 3679 (print) 2520 – 4890 (Online)
DOI: 10.53799/ajse.v23i1.1024

Predictions of Malaysia Age-Specific Fertility Rates using the Lee-Carter and the Functional Data Approaches

Syazreen Niza Shair, Norshahida Shaadan, Nur Amalia Badrina Meor Amirudin Fikri and Nur A'thiqah Mohd A'kashalf

***Abstract*— Global fertility has been experiencing a significant decline, reaching towards the replacement ratio. This trend, coupled with increasing life expectancies, has led to the emergence of an ageing population. In this study, we aim to analyse fertility patterns among Malaysian women, considering both their childbearing age and ethnicities. A 63-year age-specific fertility dataset, from 1958 to 2020, were collected from the Department of Statistics Malaysia. These data were fitted into the Lee-Carter model and its modified version, which is the functional data model. The models were evaluated using the out-sample forecast error measures. Results indicate that the third-order functional data model able to capture most of variation present in the actual data, consequently outperforming the Lee-Carter model in forecasting fertility rates among Chinese and Indian populations. There was a noteworthy shift in maternal ages of the highest births to older ages suggesting a trend towards delayed pregnancies among women. It is predicted that the Malay total fertility rates will likely fall to below the replacement level reaching 1.71 in 2040 whereas Chinese and Indian total fertility rates will substantially decrease to the lowest level in history below 1.0 which are 0.54 and 0.70 respectively.**

***Index Terms*— Age-specific fertility rates, Fertility rates, Functional data model, Lee-Carter model.**

I. Introduction

The global fertility rates have seen a substantial decline over the past few decades, with the average number of children per woman decreasing significantly by more than half, from 5.068 in 1960 to 2.300 in 2021 [1]. This declining trend is particularly pronounced in certain developed countries, where fertility rates have fallen below the replacement level, reaching historically low levels. For instance, Japanese total fertility rates dropped below the replacement levels as early as 1975 and hit their lowest point at 1.3 infants per woman in 2020 [2].

In Malaysia, the total fertility rates have undergone a substantial reduction from average 6.4 infants per woman in 1960 to 1.8 infants in 2020. It is noteworthy that, Malaysia's TFR exhibits a distinctive characteristic, with rates varying among ethnic groups such as Malays, Chinese, and Indians. Statistics show that Chinese and Indian total fertility rates have consistently been lower than that of Malays throughout the years. These variations in fertility patterns among different ethnic groups are likely be attributed to a multitude of factors, encompassing genetic predispositions, cultural backgrounds, and religious practices.

A continuing decline in the fertility rate across many countries could have significant implications for long-term national social and economic development. A persistently low birth rate proves inadequate to sustain the existing population, potentially leading to a substantial imbalance between elderly dependents and the working-age population. Thus, without adequate preparation, this situation may ultimately impose a substantial financial burden on the nation. One way to address these changes includes conducting a rigorous analysis and estimations of fertility rates, understanding the underlying fertility pattern and modelling the data to estimate future patterns.

While many polynomial models have been employed to model fertility rates by age groups, non-parametric models for predictive purposes have received relatively limited

Syazreen Niza Shair is Senior Lecturer of Actuarial Science at School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450, Shah Alam, Malaysia (e-mail: syazreen@uitm.edu.my).
Norshahida Shaadan is Senior Lecturer of Statistics at School of Mathematics, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450, Shah Alam, Malaysia (e-mail: shahida@uitm.edu.my).
Nur Amalia Badrina Meor Amirudin is an Executive at Credit Compliance Department, Menara Public Bank, 146, Jalan Ampang, 50450, Kuala Lumpur. (email: amaliabadrina@gmail.com)
Nur A'thiqah Mohd A'kashalf is an Executive at Wisma MCIS, Jalan Barat, 46200 Petaling Jaya (email: nurathiah77@gmail.com)

This research work is supported by College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Malaysia.

exploration. This research contributes to the existing literature by exploring bi-factor regression models including the Lee-Cater and its modified version, the functional Lee-Carter model to predict future Malaysian age-specific fertility rates in Malaysia. The predictions are extended to three different ethnic groups which are Malay, Chinese and Indian thus, providing significant knowledge about fertility differential among ethnicities. Therefore, this research aims to investigate the most accurate fertility prediction model for Malaysia, that capable of capturing fertility variations with respect to age and ethnicities. The two models will be evaluated using out-sample forecast error measures.

This research paper is organised as follows. Section II discusses the development fertility models proposed in previous research, Section III describes data that we use for this study and explains the age-specific fertility projection models and error measures for model evaluation. Section IV discusses the research outcomes in detail. Finally, Section V concludes.

II. LITERATURE REVIEW

Understanding fertility trends requires specific models tailored to age-specific fertility analysis. In literature, there are a variety of parametric fertility models used to explain the behavior of the fertility patterns by age. The class of these polynomials models includes widely recognized models such as the Hadwiger model, Hadwiger mixture model, Coale–Trussell model Beta and Gamma models, Pearson Type curve models, the scaled Weibull model, logistic model, Skew logistic model and Gompertz model [3]. These parametric models have their advantages and work as best fit under different patterns for given age groups. Nonetheless, careful attention must be given as parametric models require the specification and assumptions of age parameters, to present an accurate representation of actual fertility data [4-5]. Additionally, a limitation of parametric models lies in their inability to fully capture temporal changes in fertility, as they primarily consider age as the sole variable in the modeling process.

In contrast to parametric models, non-parametric methods offer greater flexibility and do not necessarily adherence to specific assumptions regarding parametric model form. The non-parametric models involve the use of smoothing techniques such as kernel smoothing, loess, B-splines and splines, known as graduation technique which dependent on historical data and solely focuses on removing random fluctuations in the data [6]. Despite being flexible, non-parametric models are not particularly popular due to their parameters are sometimes not interpretable. In addition, the spline models entail the estimation of a higher number of unknown parameters, which can potentially result in overfitting issues.

The work by Lee and Carter represents a significant milestone in the field of demographic modelling. The Lee-Carter model improves the existing polynomial model by having not only age as the factor but also time. This bi-factor model effectively extrapolates long-term mortality data as a stochastic time series and was proven accurate for US mortality data [7]. Building on this foundation, [8] extended the application of the Lee-Carter model to fertility modeling and predictive. The model's unique advantage in the context of fertility data lies in its incorporation of both age and time components, enabling it to not only account for fertility variations by maternal age but also capture fertility transitions over time. Leveraging principal component analysis (PCA) and time series analysis as its underlying framework, the Lee-Carter model has been adopted to predict future age-specific fertility rates by [9]. Its application to Malaysian fertility data has been documented in the work of [10]. However, it is noteworthy that a comprehensive comparative evaluation between the Lee-Carter model and other fertility models is yet to be undertaken, leaving the determination of the most suitable model for Malaysian fertility data an open question.

Numerous expansions and variations of the Lee-Carter model have been proposed in the existing literature. For instance, the Lee-Carter model has been extended into a functional data framework. Functional data is defined as recorded data that arise in a continuum such as time or space, represented by functions. Curves are the physical form of functional data. Functional Data Analysis (FDA) is a branch of statistics that analyses functional data [11]. FDA is one such approach to modelling time series data that has started to receive attention in the literature. In a particular context, modelling time series with functional data has a greater advantage compared to conventional point data. The benefit of the FDA is that it produces models that can be described by continuous smooth dynamics, which then enables accurate parameter estimation for use in the analysis phase, efficient data noise reduction through curve smoothing, and applicability to data with erratic time sampling schedules [12]. In addition, some useful justifications are also provided by [13] for taking into account functional data which include; it is more natural to think through modelling problems in a functional form, smoothing and interpolation techniques can produce functional representations of a finite set of observations, and analysis goals can be functional in nature, as would be the case if finite data were used to estimate an entire function, its derivatives, or the values of other functionals.

Research from [14] employing a non-parametric approach to the original Lee-Carter model to facilitate smooth age functions and robust modeling of age-specific fertility rates. This functional data methodology has found application in forecasting fertility trends in Australia [15] and Malaysia [16-17]. While [18] have explored both models using Malaysian mortality data, a comparative analysis between the functional data method and the original Lee-Carter model concerning Malaysian fertility data remains unexplored. Thus, this research aims to predict Malaysian age-specific fertility rates using both the Lee-Carter model and the functional data model and predict future ASFRs according to different ethnic groups such as Malay, Chinese and Indian, using the most accurate model. The significance of estimating ASFRs across ethnicities lies in its relevance to formulating more effective policies, as each ethnic group is experiencing a distinctive fertility transition.

III. METHODOLOGY

A. Data Preparation

The fertility data utilized in this study was obtained from the Department of Statistics Malaysia (DOSM) through a formal request submitted via the *eStatistik* online platform, accessible at https://newss.statistics.gov.my/. The dataset comprises age-specific fertility rates categorized into distinct age groups: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49. These fertility rates have been recorded for more than six decades from the year 1958 to 2020, segregating into Malaysia's primary ethnic groups--- Malay, Chinese, and Indian populations. Notably, this dataset exhibits exceptional data quality, with no instances of missing values. The data is presented in the form of [age x year] matrices, which have been subsequently transformed into .txt file format. To extract meaningful insights and conduct comprehensive demographic analyses, we employed the R software and its package namely *Demography*. The collected data were fitted into the two selected models described in the next sections.

B. The Lee-Carter Model

The Lee-carter model comprises age and time factors describe as follows:

$$ln(f_{x,t,i}) = a_{x,i} + b_{x,i}k_{t,i} + \varepsilon_{x,t,i} \quad (1)$$

where $f_{x,t,i}$ is the age-specific fertility rate for the mother's age group x in year t for i^{th} ethnic group. The $a_{x,i}$ is the average age-specific fertility over t years, $b_{x,i}$ is the age-component that describes deviation in fertility by mothers age x due to changes in $k_{t,i}$ where $k_{t,i}$ is the time-component or fertility index in t year. The $\varepsilon_{x,t,i}$ is a random error of the model.

The $b_{x,i}$ and $k_{t,i}$ parameters were estimated using the singular value decomposition of matrices $[ln(f_{x,t,i}) - a_{x,i}]$where the sum of $b_{x,i}$ is set as one whereas the sum of $k_{t,i}$ is set as zero. The estimation of $b_{x,i}$ is distinguished by the sensitivity of $ln(f_{x,t,i})$ to changes in the fertility index, $k_{t,i}$. The error term, $\varepsilon_{x,t,i}$ includes all remaining variances and the calculation of $a_{x,i}$ is defined by:

$$a_{x,i} = \frac{1}{T}\sum ln(f_{x,t,i}) \quad (2)$$

After that, the time-varying component is predicted using the univariate time series model. The forecasted fertility index is as follows:

$$\hat{k}_{t,i} = k_{t-1,i} + c + e_{t,i} \quad (3)$$

where $\hat{k}_{t,i}$ is the forecasted fertility index in year t, c a drift term, and $e_{t,i}$ is the forecasting deviation. The forecasted $\hat{k}_{t,i}$ in this formula is dependent on the drift term and is predicted to decrease approximately linearly over time. Using these forecasted time component parameters and the estimated age component parameters, the forecasted age-specific fertility rates are defined as below:

$$\hat{f}_{x,t,i} = exp(a_{x,i} + b_{x,i}\hat{k}_{t,i}) \quad (4)$$

C. The Functional Data Model

The functional data model is a mix of functional time series analysis and principal components decomposition. defined in this study as follows.

$$\gamma_{t,i}(x) = f_{t,i}(x) + \sigma_{t,i}(x)\varepsilon_{t,i,x} \quad (5)$$

Where$\gamma_{t,i}(x)$ is the observed fertility rates for age x in year t and for ethnic group i. The smoothed fertility rate is given by $f_{t,i}(x)$. The $\varepsilon_{t,i,x}$ is an independent and identically distributed random error and $\sigma_{t,i}(x)$ is the age-dependent error. A non-parametric technique is used to estimate the smoothed rate of $f_{t,i}(x)$, which is a median smoothing B-spline bound to be concaved. The smoothed age-specific fertility rates will then be fitted into the functional data model as indicated in (6).

$$f_{t,i}(x) = \mu_i(x) + \sum_k \beta_{t,i,k}\delta_{i,k}(x) + e_{t,i}(x) \quad (6)$$

The k variable denote the k^{th} order of the principal component analysis ($k = 1,2,3$). The $\mu_i(x)$ is the average of $f_{t,i}(x)$ for each age over the years. The age-component $\delta_{i,k}(x)$ and the time- component $\beta_{t,i,k}$ are estimated using principal component decompositions of 2 × 2 matrices of $[f_{t,i}(x) - \mu_i(x)]$. The age and time components are equivalent to the functions and coefficients of the primary components. The model errors are represented by $e_{t,i}(x)$. Using the fitted time series models, the coefficients $\beta_{t,i,1}$ $\beta_{t,i,2}$ and $\beta_{t,i,3}$ are forecasted for future. The 'Auto-arima' function from the R programming forecast package will be used to select the most accurate time series model for each coefficient. To estimate the forecast of smoothed fertility rates, the forecast of coefficients will be multipliedby the estimated principal components functions.

D. Model Evaluation

The 63-year age-specific fertility data were segmented into two distinct intervals: the fitted interval covering data from 1958 to 2009 and the evaluation interval data from the years 2010 to 2020. Both the Lee-Carter and functional data models were employed to fit the fitted interval data. Model parameters were estimated, and subsequently data predictions were made for a 10-year horizon. To assess the accuracy of both models, a comparative analysis was conducted by comparing the out-of-sample forecast values with the evaluation interval data. Evaluation was carried out using error measures including Mean Squared Error (MSE) and Root Mean Square Error (RMSE). The out-performing model was then selected for predicting Age-Specific Fertility Rates (ASFRs) and Total Fertility Rates (TFRs) from 2021 to 2040. The specific formulations of the chosen error measures are detailed as follows.:

$$MSE_i = \frac{\sum_{t=1}^{10}\sum_{x=1}^{7}(f_{x,t,i} - \hat{f}_{x,t,i})^2}{70} \quad (7)$$

$$RMSE_i = \sqrt{\frac{\sum_{t=1}^{10}\sum_{x=1}^{7}(f_{x,t,i} - \hat{f}_{x,t,i})^2}{70}}. \quad (8)$$

IV. RESULTS

Fig. 1 illustrates the observed fertility trends from 1958 to 2020. During early years from 1958 to 1962 (depicted by the red curves), Indian women aged 20 to 24 exhibited the highest peak with a fertility rate of 4.07 children per woman. Whereas, for Chinese and Malay ethnicities, the highest number of births occurred among women aged 24 to 29, with rates of 3.69 infants and 20 to 24 with 3.34 infants, respectively. Over the subsequent years, these age-specific fertility rates experienced a gradual decline in which the fertility rates of Chinese and Indian populations decreased at a faster rate compared to Malays, evident in the greater divergence between the red and purple curves for Chinese and Indians in contrast to Malays.

The analysis of recent age-specific fertility data from 2010 to 2020 (represented by the purple curves), indicates that the maternal ages of the highest births have shifted to later ages. The peak of the curves not only displayed a reduction but also delayed to later maternal ages, signifying that women today tend to give birth at older ages and have fewer children compared to women in earlier years. For instance, in 2020, Indian women had the highest number of births among those aged 30 to 34 with only 0.72 babies, as opposed to those in 1960, the highest number of births occurred among younger Indian women aged 20 to 24 with 4.07 babies, a decrease by 560% over the 60 years duration. This substantial fertility evolution can be documented with access to long-term data. According to [19], Malaysian women's marriages at later ages were the primary cause of these postponed pregnancies.

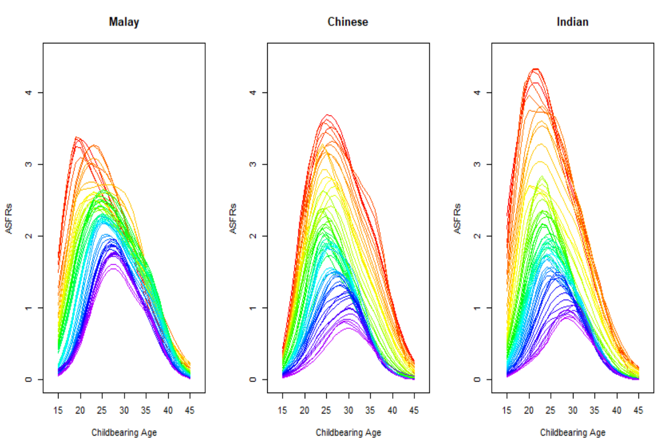

Fig. 1. Rainbow plots of age-specific fertility rates by childbearing age and ethnic groups including Malay (right), Chinese (centre) and Indian (left) from 1958 to 2020. The colours of the curves differentiate fertility rates according to years--- red (1958-1962), orange (1963-1968), yellow (1969-1972), green (1973-1988), turquoise (1989-1998), blue (1999-2009) and purple (2010-2020).

Fig. 2 demonstrates the trends of Malaysian total fertility rates according to three major ethnic groups--- Malay, Chinese, and Indian. The TFR trends show that Malaysian family size continued to decline across the generations from 1958 to 2020. In 1958, Indian women had the highest number of children with 7.3 children per woman on average, followed by Chinese 6.3 children and Malay 5.7 children. The TFRs of Indians and Chinese showed a sharp decrease afterwards, reaching every 4.8 children in 1970 while Malay's TFR declined steadily and surpassed Indian and Chinese with 5.0 children. The fertility trends continue declining with Malay laying slightly above the replacement level with 2.28 children in 2020, whereas Chinese and Indian TFRs were reaching towards below the replacement level in the same year with 1.17 and 0.98 children respectively. The replacement level of 2.1 children is commonly used as the benchmark for fertility level in one country as it refers to the point at which women give birth to babies just enough to sustain the population level.

A. *Parameter Estimations*

The interval data from 1958 to 2010 were fitted into the Lee-Carter and the third-order functional model for each ethnic group--- Malay, Chinese and Indian. The values of a_x, b_x and k_t were estimated from the Lee-Carter Model for each ethnic group. There is some constraint imposed when using the Lee-Carter model which the assumption of the summation of bx equal to 1 and summation of kt equal to 0 must be followed.

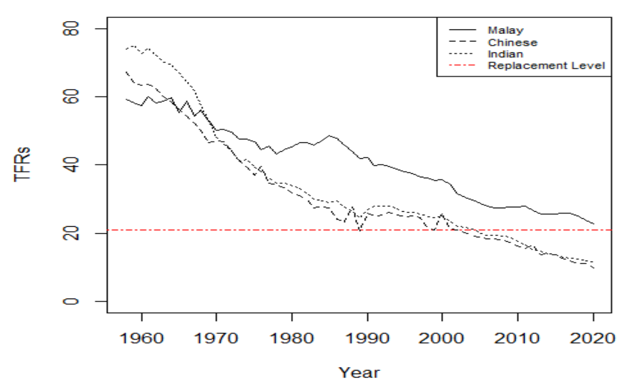

Fig. 2. Malaysia's total fertility rates (TFRs) according to ethnic groups from 1958 to 2020

Fig. 3, Fig. 4 and Fig 5 shows the trends of the estimated parameters of the Lee-Carter model which are $a_{x,i}$, $b_{x,i}$ and $k_{t,i}$ respectively. The parameter $a_{x,i}$ describes the general shape of fertility against the age of the mother of each ethnic group. The overall trend shows a bell-shaped distribution, in which the highest number of children occurred among the mid-aged women between 25 to 30 years old whereas the lowest number of children were recorded among the youngest and the oldest women. This pattern is almost similar for all ethnic groups.

The $k_{t,i}$ is a time index of the general level of fertility, showing an almost linearly decreasing pattern over the years reflecting the overall declining trends of fertility for all three sub-populations. The k_t of Chinese and Indian decrease more steeply than that of Malays. Finally, the $b_{x,i}$ parameter describes the extent to which fertility at age x changes given the overall change in the general level of fertility. The greater values of $b_{x,i}$ are associated with faster fertility change.

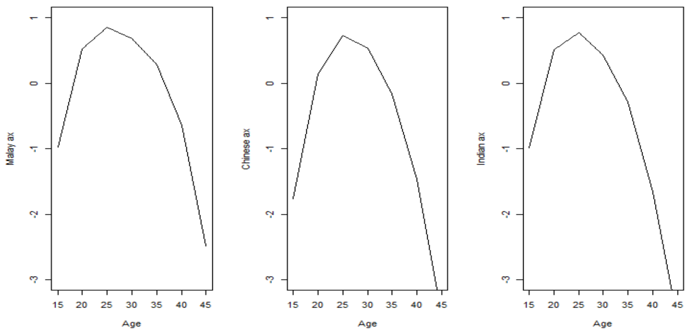

Fig. 3. The estimated parameters of the Lee-Carter model include $a_{x,i}$ for Malay (right), Chinese (center) and Indian (left).

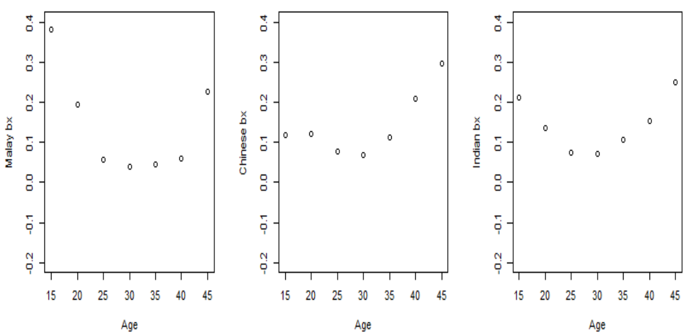

Fig. 4. The estimated parameters of the Lee-Carter model include $b_{x,i}$ for Malay (right), Chinese (center) and Indian (left).

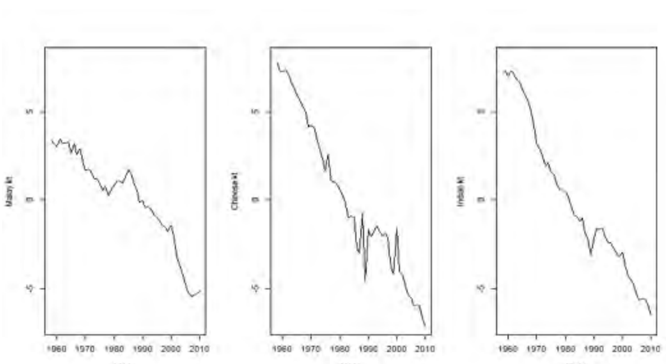

Fig. 5. The estimated parameters of the Lee-Carter model include $k_{t,i}$(for Malay (right), Chinese (center) and Indian (left).

The estimated parameters of the third-order functional data model for Malay, Chinese and Indian include $\mu_i(x)$, the coefficients which are time components ($\beta_{t,1,i}$, $\beta_{t,2,i}, \beta_{t,3,i}$) and the associated basis functions which are age components ($\delta_{i,1}(x), \delta_{i,2}(x), \delta_{i,3}(x)$), were displayed in Figure 6, Fig 7 and Fig 8 respectively. The main effects represent the $\mu_i(x)$ which is the average bell curve of age-specific fertility rates for each sub-population over the years The first-order time components ($\beta_{t,1,i}$) depicted a decreasing pattern in all ethnic groups indicating that the fertility rates are declining over the historical period while the first order of age components ($\delta_{i,1}(x)$) reflects how the fertility rates by age change as fertility rates decrease over time. The second and third orders of coefficient basis functions explained the remaining data variation.

The total percentages of variation explained by each model are displayed in Table 1. The first-order functional data models explain less variation for Malay and Chinese compared to the Lee-Carter model indicating the parameters of the functional data model are not sufficient to represent the actual data variation. However, the second and third-order functional model shows that the percentage of variation improved and was higher than the Lee-Carter for all ethnic groups. This indicates that the third-order functional data model is able to capture most of variation representing in the actual data. The fourth and higher order of functional data models were not considered in this study due to the additional percentages of variation being minuscule and too many parameters may result in complex and overfitting modelling.

Once the data were fitted into the functional data model, only the coefficient values will be forecasted using the time series models. We used the R program to run the analysis and adopted auto. arima function to predict the coefficient values. For example, the first-order coefficients for Malay data were forecasted using the ARIMA (1,1,0) with drift, whereas the second and third orders coefficients were forecasted using the ARIMA (0,2,1) and ARIMA (1,0,1), respectively. The forecasted coefficients with 95% confidence intervals (yellow) are shown in Figure 4. The fitted bases and the forecasted coefficients will be used to predict the 10-year out-sample age-specific fertility rates from the year 2011 to 2020 using the equation (6). Subsequently, the TFRs out-sample forecast values were calculated based on the predicted ASFRs.

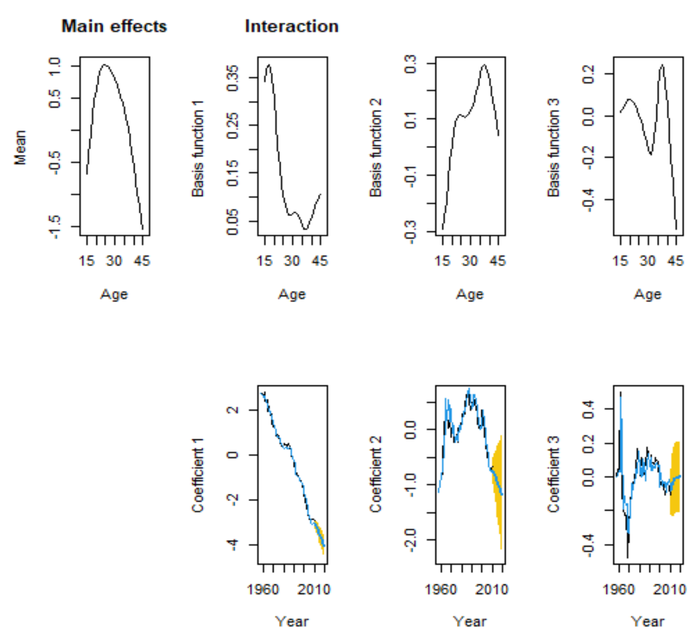

Fig. 6. Fitted coefficients and basis functions of the third-order functional data model for Malay data.

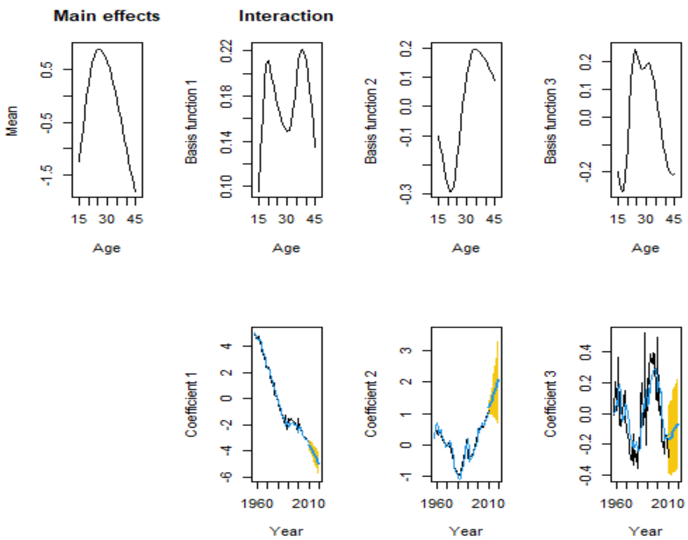

Fig. 7. Fitted coefficients and basis functions of the third-order functional data model for Chinese data.

Once the data were fitted into the functional data model, only the coefficient values will be predicted using the time series models. We used the R program to run the analysis and adopted auto. arima function to predict the coefficient values. For example, the first-order coefficients for Malay data were forecasted using the ARIMA (1,1,0) with drift, whereas the second and third orders coefficients were forecasted using the ARIMA (0,2,1) and ARIMA (1,0,1), respectively. The forecasted coefficients with 95% confidence intervals (yellow). The fitted bases and the forecasted coefficients will be used to predict the 10-year out-sample age-specific fertility rates from the year 2011 to 2020 using (6). Subsequently, the TFRs out-sample forecast values were calculated based on the predicted ASFRs.

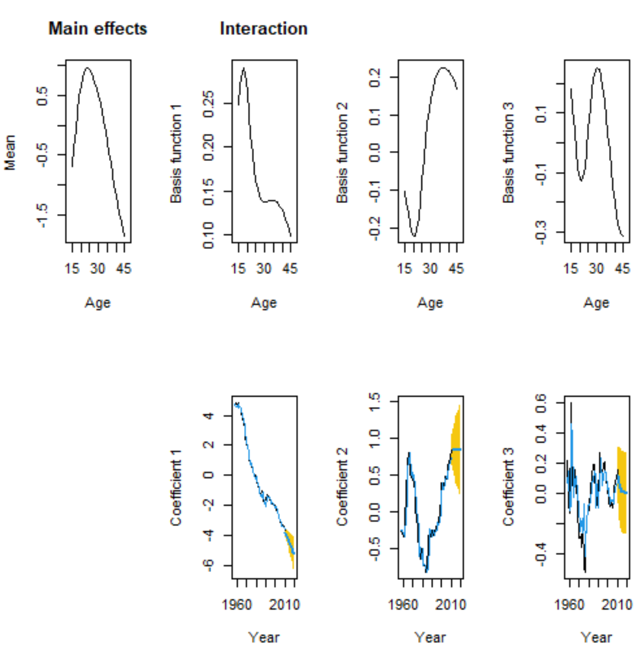

Fig. 8. Fitted coefficients and basis functions of the third-order functional data model for Indian data.

TABLE 1

THE TOTAL PERCENTAGE OF FERTILITY VARIATION EXPLAINED BY LC (LEE-CARTER) AND FD (FUNCTIONAL DATA) MODELS

Ethnicity	LC	FD ($K = 1$)	FD ($K = 2$)	FD ($K = 3$)
Malay	92.40%	91.60%	97.65%	98.55%
Chinese	95.70%	94.50%	98.47%	99.26%
Indian	92.70%	95.40%	98.56%	99.17%

Subsequently, the out-sample ASFR forecast values from the Lee-Carter model and the third-order functional data model will be evaluated and compared with the observed values of the same year. The error results are shown in Table 2 and Table 3. The third-order functional data model consistently out-performed the Lee-Carter model in predicting both Chinese and Indian ASFRs due to the out-sample errors of the functional data model were lower than the Lee-Carter model for these ethnic groups. In contrast, the Lee-Carter model work better to predict Malay ASFRs than that of the functional data model. One may use different model to predict the fertility rates of different ethnic groups to get more accurate results. However, by taking average errors from the three sub-populations, results indicate that functional data is the most accurate model to forecast the Malaysian fertility rates by ethnic group.

TABLE 2

THE COMPARISON OF OUT-SAMPLE MEAN SQUARED ERROR (MSE) OF ASFR FROM LC AND FD MODELS

Ethnicity	Malay	Chinese	Indian	Overall
LC	**0.0086**	0.0342	0.0357	0.0262
FD ($K = 3$)	0.0129	**0.0213**	**0.0276**	**0.0206**

TABLE 3

THE COMPARISON OF OUT-SAMPLE ROOT MEAN SQUARED ERROR (MSE) OF ASFR FROM LC AND FD MODELS

Ethnicity	Malay	Chinese	Indian	Overall
LC	**0.0929**	0.1849	0.1891	0.1556
FD ($K = 3$)	0.1134	**0.1458**	**0.1662**	**0.1418**

In this research, we produced a 20-year age-specific fertility rate projection from 2021 to 2040 using the third-order functional data model, which is the most accurate model for overall Malaysian fertility rates. As shown in Fig. 9., Age-specific fertility rates continue to decline from 2021 to 2040, which is substantial for Chinese and Indian women between the ages of 15 and 30. Results show that, Malay women aged 28 will be having the highest number of children in 2040 which is 1.410 children per woman, whereas Chinese and Indian women would give the highest number of birth to 0.634 and 0.691 children each at the age of 32. These results indicate that fertility rates of Malaysian women continue declining for each women childbearing age in the future which is in line with research from [10] that found a decline in Malaysia fertility curves over the years occurred to mothers ages 20s and 40s.

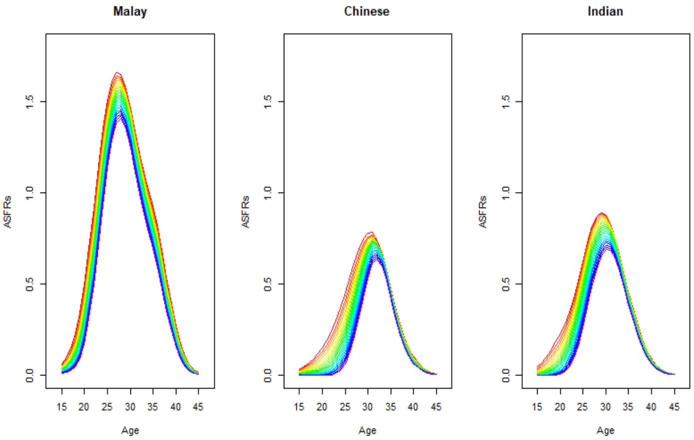

Fig. 9. The 20-year forecast values of age-specific fertility rates from the third-order functional data model for Malay (right), Chinese (centre) and Indian (left) for the year 2021 to 2040.

Fig. 10. illustrates the projected values for Malaysia's total fertility rate by ethnicity would continue to decline by 2040. Chinese TFRs are declining on average by 3.14% annually, which is more rapidly than Malay and Indian TFRs. The TFRs of Malay will likely fall from 2.28 in 2020 to below the replacement level to reach 1.71 in 2040. In addition, it is projected that by 2040, the TFRs of Chinese and Indian women will have substantial decrease to a level below 1.0 which are 0.54 and 0.70 respectively. This finding is consistent with research from [20] that concluded Chinese and Indian Malaysians will be having an "ultra-low fertility level".

Fertility trends that below 1.0 already occurred in few countries recently such as South Korea, Hong Kong and Japan. It is a high time for the Malaysian government to benchmark these nations and take lessons from them to improve the current birth related policy. The future population structure of Malaysia will be affected by the alarmingly low birth rate. It is important that the government establish policies for preventing or reversing the declining trends. For instance, in order to address this issue, the Korean government has put into place a number of action programmes from the early 2000s. All infants will receive a allowance per month in the first year after birth as part of the plans, while pregnant parents will receive a cash bonus to aid with prenatal expenses [21].

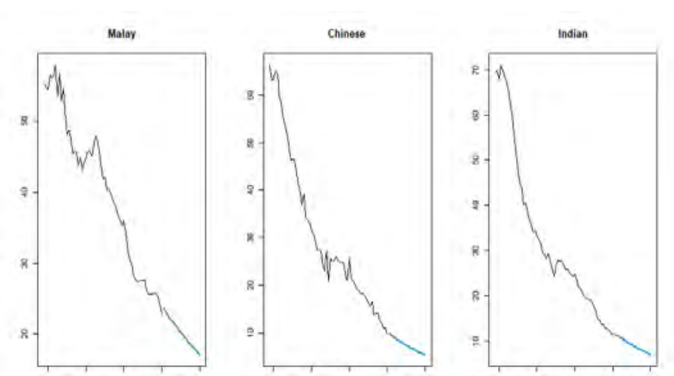

Fig. 10. The Malaysian total fertility rate (TFR) forecasts for every ten women of Malay (a), Chinese (b) and Indian (c) from 2021 to 2030

V. CONCLUSIONS

Our study extended the application of two demographic forecasting models, the Lee-Carter model and the functional data model to 63-year Malaysian age-specific fertility rates (ASFRs) from 1958 to 2020 segregated by ethnic groups. The access to this long-term data provides evidence of significant fertility rates change among Malaysian women. The trends observed by ASFRs provides evidence that Malaysian women tend to postpone their pregnancy in which the maternal age of the highest births has shifted to later ages over the years. In addition, this study proves that the Malaysian total fertility rates (TFRs) have decreased substantially over the observation period to fall below the replacement level by 2020, particularly among Chinese and Indians. The Malay TFR laying slightly above the replacement level by 2020, whereas Chinese and Indian TFRs were reaching towards below the replacement level in the same year.

This study provides empirical evidence that Malaysian sub-population fertility data by ethnicities can best be fitted into the third-order functional data model as the model could explain up to 99% of data variation, higher than total variation of the Lee-Carter model. The out-sample forecast errors of ASFRs and TFRs showed that the third-order functional data model outperformed the Lee-Carter model overall. The 20-year forecasts indicate that Malay total fertility rates will likely fall from 2.28 in 2020 to below the replacement level reaching 1.71 in 2040. Chinese and Indian total fertility rates will have substantial decrease to a level below 1.0 which are 0.54 and 0.70 respectively by 2040.

The evolution in Malaysian fertility rates is an alarming fact as, together with low mortality rates, it may impact the Malaysian population structure such as smaller number of the young generation with larger number of elderlies, will increase the old-age dependency ratios. The Malaysian government need to change views and policies on fertility to increase and encourage more births. Thus, it is recommended that future works to analyse Malaysian fertility using other Lee-Carter extensions and consider government policy that impacted fertility rates by looking at the effectiveness of policies implemented from other countries.

REFERENCES

[1] United Nations Department of Economic and Social Affairs, Population Division, "World Population Prospects 2022: Summary of Results," UN DESA/POP/2022/TR/NO. 3, 2022.

[2] United Nations, "Government response to Low fertility in Japan," United Nations Expert Group Meeting on Policy Responses to Low Fertility, Policy Brief No. 11, 2015.

[3] U. Srivastava, K. K. Singh, A. Pandey, and N. Narayan, "Experiments in modeling recent Indian fertility pattern," Scientific Reports, [Online]. Available: https://doi.org/10.1038/s41598-021-85959-z, 2021.

[4] M. R. Islam, "Modeling of Age-Specific Fertility Rates of Jakarta in Indonesia: A Polynomial Model Approach," International Journal of Scientific & Engineering Research, vol. 2, no. 11, pp. 1-5, 2011.

[5] N. Arbu, A. Lim, and C. Me-ead, "Forecasting the Patterns and Trends in Age-Specific Fertility Rate in South Asia," Turkish Journal of

Computer and Mathematics Education, vol. 8, no. 12, pp. 2595-2600, 2021.
[6] K. K. Lam and B. Wang, "Robust Non-Parametric Mortality and Fertility Modeling and Forecasting: Gaussian Process Regression Approaches," Forecasting, vol. 3, no. 1, pp. 201-227, 2021, DOI: 10.3390/forecast3010013.
[7] G. D. Lee and L. R. Carter, "Modeling and forecasting U.S. mortality," Journal of the American Statistical Association, vol. 87, no. 419, pp. 659-671, 1992.
[8] R. D. Lee, "Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level," International Journal of Forecasting, vol. 9, pp. 187-202, 1993.
[9] P. Vanella and P. Deschermeier, "A Principal Component Simulation of Age-Specific Fertility – Impacts of Family and Social Policy on Reproductive Behavior in Germany," Population Review, vol. 58, no. 1, 2019, DOI: 10.1353/prv.2019.0002.
[10] H. Hanafiah and A. A. Jemain, "Change of Age-Specific Fertility Curves in Peninsular Malaysia: An Approach of Functional Data Analysis," in Proceedings of the International Conference on Science, Technology and Social Sciences (ICSTSS) 2012, Springer, Singapore, 2014, DOI: 10.1007/978-981-287-077-3_19.
[11] J. O. Ramsay and B. W. Silverman, Functional Data Analysis, Springer, New York, 2005.
[12] S. Ullah and C. F. Finch, "Application of Functional Data Analysis: A systematic review," BMC Medical Research Methodology, vol. 13, no. 43, pp. 1-12, 2013.
[13] J. O. Ramsay and C. J. Dalzell, "Some tools for functional data analysis," Journal of the Royal Statistical Society, Series B (Statistical Methodology), vol. 53, pp. 539–572, 1991.
[14] R. J. Hyndman and M. S. Ullah, "Robust forecasting of mortality and fertility rates: A functional data approach," Computational Statistics & Data Analysis, vol. 51, no. 10, pp. 4942–4956, 2007.
[15] R. J. Hyndman and H. Booth, "Stochastic population forecasts using functional data models for mortality, fertility, and migration," International Journal of Forecasting, vol. 24, pp. 323-342, 2008.
[16] H. Hanafiah and A. A. Jemain, "Change of Age-Specific Fertility Curves in Peninsular Malaysia: An Approach of Functional Data Analysis," Proceedings of the International Conference on Science, Technology, and Social Sciences (ICSTSS) 2012, Springer, Singapore, 2014, DOI: 10.1007/978-981-287-077-3_19.
[17] S. N. Shair, A. Y. Yusof, S. N. M. Khair, and N. I. Sharan, "Malaysian fertility transitions: analyses and projections of ASFR and TFR by ethnicity," International Journal of Engineering & Technology, vol. 8, no. 1.7, pp. 68-74, 2019.
[18] S. N. Shair, N. A. Zolkifli, N. F. Zulkefli, and A. Murad, "A functional data approach to the estimation of mortality and life expectancy at birth in developing countries," Pertanika Journal Science & Technology, vol. 27, no. 2, pp. 797-814, 2019.
[19] A. Bakar and N. Abdullah, "Labor force participation of women in Malaysia," in International Economic Conference on Trade & Industry, Penang, Malaysia, 3-5 December 2007.
[20] T. N. Peng. Bracing for Low Fertility in Malaysia, ISEAS, pp 1-11, 2020.
[21] I. Sajid, "South Korea announces incentives for more births: Government to dole out cash to expecting parents," Anadolu Agency, 2020.

Syazreen Niza Shair was born in Johor Bahru, Johor, Malaysia. She received the BSc degree in Actuarial Science from the Universiti Teknologi MARA in 2002 and MSc degree in Statistics in University of Malaya in 2007, in 2017, received her PhD in Applied Finance and Actuarial Studies from Macquarie University, Sydney Australia.

Currently, she serves as a Senior Lecturer of Actuarial Science in Universiti Teknologi MARA, Malaysia. She is the author of articles and research papers included the exploration of various models, such as the Lee-Carter model and functional data models, to gain insights into mortality and fertility transitions among different ethnic groups in Malaysia. With a passion for understanding population dynamics, Dr Syazreen has undertaken extensive research to shed light on the factors influencing fertility rates in Malaysia and their implications for policy planning. Dr Syazreen was a recipient of the Malaysia Fundamental Research Grant Scheme Award 2019 and 2023.

Norshahida Shaadan, was born in Mentakab, Pahang Malaysia in 1973. She received the BSc. degree in Mathematics and Statistics with 1st class honor from the University of Bradford United Kingdom. She was awarded as the best student in the graduation year 1996 from the University, and her achievement was also acknowledged by the Institute of Mathematics United Kingdom. She obtained her Master's and PhD degrees in Statistics from the Universiti Kebangsaan Malaysia in 2007 and 2015 respectively.

Currently, she is a senior lecturer at the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Selangor, Malaysia. Over 23 years of service as a lecturer, she teaches various courses such as Research Methodology, Applied Statistics, Probability and Statistics, and Quality Control. Her research interests include environmental (i.e., air pollution and water quality) modelling, climate change studies as well as social science issues. Her main expertise is applied statistics, control charts, and Functional Data Analysis. Dr Norshahida is an Editor of the Malaysian Journal of Computing and has been appointed as a reviewer in several reputable journals.