MATEMATIKA, MJIAM, Volume 40, Number 3, 83–99

A Modified Lee-Carter Model for Age-Specific Fertility Rate Forecasting: Modelling Ethnic-Based Fertility Change in Malaysia

Syazreen N. Shair*, **Muhammad S. A. M. Nasir, Nur F. M. Fadhil, Nur F. Z. Zaidi and Muhammad A. Roslan**

Centre for Actuarial Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450, Shah Alam, Selangor

*Corresponding author: syazreen@uitm.edu.my

Article history
Received: 15 July 2024
Received in revised form: 12 November 2024
Accepted: 18 November 2024
Published on line: 1 December 2024

Abstract Accurate fertility prediction is crucial for long-term policy planning, particularly in health economics and demographics. This article proposes a modified version of the Lee-Carter model, which incorporates the skew-logistic probability density function to accurately capture the unimodal curves of age-specific fertility rates (ASFRs) in Malaysia. The model was fitted into Malaysian age-specific fertility data between 1958 and 2005, categorised by three major ethnic groups: Malay, Chinese and Indian. The forecast performances of both the modified and original Lee-Carter models were evaluated by estimating out-of-sample errors between 2006 and 2021. The model that demonstrated the highest accuracy was used to forecast ASFRs between 2022 and 2041. The analyses revealed a notable decline in overall fertility trends over the years, with particularly pronounced decreases observed among the Chinese and Indian populations. Furthermore, there has been a shift towards delayed childbirth, with the highest number of births occurring at older ages in recent years. The results indicate that the modified Lee-Carter model outperforms the original version for the Chinese and Indian populations, suggesting its ability to capture recent significant changes in fertility patterns and enhancing predictive accuracy.

Keywords Lee-Carter Model; Skew Logistic Model; Age-specific Fertility Rates; Total Fertility Rates; Ethnicities.

Mathematics Subject Classification 62P05.

1 Introduction

Human fertility patterns have undergone significant transformations on a global scale, with numerous developed nations experiencing a declining pattern in fertility rates. The total fertility rate (TFR), one of the most important demographic indicators for any population, measures the actual reproductive rate of a population based on the overall number of live births per woman. According to the Department of Statistics Malaysia's Vital Statistics 2023 report, Malaysia's TFR has notably declined over the past decade, falling from 2.1 children per woman in 2010 to 1.6 in 2022 [1]. This decline resulted from various factors, including economic challenges, lifestyle changes, delayed marriages and increased infertility issues. In addition, the age-specific fertility rate (ASFR) provides insights into the annual number of births among every 1,000 women within specific age groups, particularly those aged 15 to 49 years. The analysis of ASFR is crucial in demographic studies, given that the likelihood of childbirth varies significantly with the mother's age.

Fertility trends in Malaysia show distinct variations among the three primary ethnic groups, namely Malay, Chinese and Indian. As of 2021, Malay and Bumiputera account for 69.9 percent of the total population, while Chinese make up 22.8 percent, and Indians comprise 6.6 percent [1]. The statistics reveal significant disparities in the number of live births among these ethnic groups, with Malays recording the highest number of births, followed by Chinese and Indians. These disparities in fertility trends among ethnicities can be attributed to several factors, such as age at marriage, current family size, gender preferences and socioeconomic variables, including education, occupation, religion, contraceptive usage and the financial implications of raising children [2].

In many countries, including Malaysia, the ASFR curves generally show a unimodal and bell-shaped pattern. The reproductive age typically begins around age 15, after which the fertility level gradually increases and reaches its highest peak for women in their late twenties to early thirties. Subsequently, the ASFR curve starts to decline, reaching very low values for women over the age of 35 and typically ending at age 49, which indicates the end of the reproductive age span due to increased reproductive health risks associated with age. Various mathematical models have been proposed to analyse fertility patterns, with specific methods designed to capture the evolving patterns of fertility rates. ASFR models were developed to capture asymmetrical fertility patterns in certain countries. These include the seven-parameter Hadwiger mixture model [3], the six-parameter normal distribution [4] and the thirteen-parameter piecewise quadratic spline function incorporating a graphical intuitive technique [5]. In addition, Gaire *et al.* [6] explored the 4-degree bi-quadratic polynomial model for Napalese ASFR and proved that this model offers a better fit than traditional quadratic models.

Besides the models described above, Mazzuco and Scarpa [7] have developed an improved model that employs a skew-normal distribution to better reflect the fertility patterns of countries with bimodal and unimodal fertility rates. The model offers the advantage of having fewer parameters than previous models while ensuring the flexibility required to accommodate complex fertility patterns. Asili *et al.* [8] proposed a skew-logistic distribution model and proved that it outperformed the skew-normal model by generating a lower sum of squared residuals. The skew-logistic model has consistently been used to fit the ASFR data across nearly all Indian states due to its ability to capture asymmetric unimodal and bimodal curves of fertility rates [9]. Meanwhile, the modelling of ASFRs in Malaysia remains underdeveloped. A study by Mathivanan *et al.* [10] fitted the fertility data of Malaysian women using four mathematical models: Hadwiger, Gamma, Beta and Gompertz. However, neither the skew-normal nor the skew-logistic models have been tested.

The previously mentioned parametric models have the advantage of providing a more refined and flexible approach to modelling fertility rates because they are designed to capture age-specific fertility curves. Nevertheless, parametric models solely focus on fitting data from a specific year, whether current or historical age-specific fertility rates (ASFRs), as they are constrained by fixed functional forms of age and do not account for temporal changes. Given that fertility patterns have been seen as evolving over time, it is important to predict fertility rates while considering variations of patterns by age and fertility changes over the years.

The Lee-Carter model demonstrates a significant advancement by incorporating age factors and accommodating time-varying variables [13], which enables it to more effectively capture temporal variations, making it more suitable for projecting ASFRs. For example, Lee and Carter [14] applied the Lee-Carter model to fit ASFRs in the United States, while Hyndman and Ullah [15] employed the functional version of the Lee-Carter model for fertility rates in Australia. Moreover, the Lee-Carter model offers a natural probabilistic framework for predicting future values, and it has been used to forecast fertility rates in the United Kingdom [16]. However, only a limited number of studies have explored the application of the Lee-Carter model using Malaysian ASFR data. Hanafiah and Jemain [17] tested the Lee-Carter model to forecast Malaysian age-specific fertility curves. However, the fertility data used in their study were outdated and did not account for ethnic disparities.

The Lee-Carter model has some limitations. For instance, the a_x parameter, which represents the average of log fertility rates for each age x over time, is usually influenced by outliers in the historical data [18]. While the k_t parameter requires a sufficient length of data to accurately evaluate the decreasing nonstationary patterns over the years. This long-term data may deviate the average into a position that does not account for current changes in fertility rates. This issue leads to the biases of a_x, parameters and could compromise the accuracy of forecast values. Moreover, the model assumes a log-linear relationship. which may not fully reflect the asymmetric patterns in ASFRs. Therefore, to address these issues, this article proposes a modified approach to the Lee-Carter model that estimates a_x based on more recent data to better reflect current fertility trends in Malaysia. In addition, it incorporates the skew-logistic distribution model applied beforehand to smooth and estimate single-age rates from the original five-year interval data. The modified approach will be compared with the original Lee-Carter model regarding out-of-sample accuracy. This research could reveal important information about reliable models for predicting Malaysia's fertility rates and enhancing the understanding of fertility patterns among Malaysian women.

2 Materials and Methods

2.1 Data

The observed ASFR datasets for Malaysian women were collected in five-year age intervals: 1519, 2024, 2529, 3034, 3539, 4044 and 4549. The minimum and maximum ages aligned with the definition of women's reproductive age as used by the World Health Organisation [19]. Data were collected from the Department of Statistics Malaysia, covering a span of 64 years from 1958 to 2021. These fertility data were disaggregated by the three major ethnic groups in Malaysia: Malay, Chinese and Indian. The datasets will be used to fit a skew-logistic distribution model, which will enable the estimation of smooth single-age fertility rates for ages 15, 16, 17, through 48 and 49. These single-age data will then be fitted into both the original and modified Lee-Carter models.

To assess the forecast performance of the Lee-Carter and modified Lee-Carter models, the 64-year ASFR data was divided into in-sample and out-of-sample datasets. The in-sample dataset comprises 75 percent of the complete data, covering the period from 1958 to 2005, while the out-of-sample dataset contains the remaining 25 percent, spanning from 2006 to 2021. For the original Lee-Carter model, parameters were estimated using the in-sample data from 1958 to 2005. In contrast, the modified Lee-Carter model was fitted differently; to be specific, the parameter $\hat{a}_x^{(B)}$ was estimated using more recent data from 1991 to 2005, while parameters $\hat{b}_x^{(A)}$ and $\hat{k}_t^{(A)}$ were estimated using data from 1958 to 2005. Both models then forecasted the ASFRs from 2006 to 2021. These forecasts were subsequently compared to the out-of-sample data to evaluate forecast accuracy.

2.2 The Skew-logistic Distribution Model

Following the study by Asili *et al.* (2014), the skew-logistic distribution model was fitted to the original data of ASFRs in five-year age intervals. This fitting allows for the estimation of smooth single-age data for each year (e.g., ages 15, 16, 17, through 48 and 49). The cumulative distribution function (CDF) and the probability density function (PDF) of the logistic distribution are defined in Equations (1) and (2), respectively, where x represents age:

$$H(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$h(x) = \frac{e^{-x}}{(1+e^{-x})^2} \tag{2}$$

Subsequently, the formula for introducing skewness into the symmetric distribution, as proposed by Azzalini and Regoli [20], is as follows:

$$f(x) = 2h(x)H(\omega) \tag{3}$$

The skew-logistic distribution function is derived by substituting $H(\omega)$ and $h(x)$ in Equation (3) with Equations (1) and (2), respectively, and $\omega = \alpha x$:

$$f_{sl}(x,\alpha) = \frac{2e^{-x}}{(1+e^{-x})^2(1+e^{-\alpha x})} \tag{4}$$

where α is the skewness parameter. The x variable is transformed to $(y-\mu)/\sigma$ in which y represents the age of child-bearing women, μ represents the location of the parameter, and σ represents the scale parameter. Thus, the skew-logistic distribution mode for unimodal curves is defined as follows:

$$f_{sl}(y:\alpha,\mu,\sigma)=\frac{2e^{-(y-\mu)/\sigma}}{(1+e^{-(y-\mu)/\sigma})^2(1+e^{-\alpha(y-\mu)/\sigma})} \tag{5}$$

The parameters of the fitted model will be estimated using the non-linear least squares method, which minimises the residual sum of squares, given as the following:

$$S_{\text{res}}=\sum_{x=15}^{49}(g(x)-f(x))^2 \tag{6}$$

The lower and upper age limits of a woman's reproductive age span are designated as $x=15$ and 49, respectively. Hence, $g(x)$ represents the fertility rate at age x as projected by the skew-logistic model, while $f(x)$ represents the observed fertility rate at the same age x.

2.3 Fitting the Lee-Carter Model

According to Lee and Carter [13], the a_x parameter represents an age-specific parameter that indicates the average level of $ln(f_{x,t})$ over time. Initial estimates of b_x and k_t are computed using the SVD method, which forms the foundation of this model. The b_x measures the degree to which the natural logarithm of the fertility rate responds to or changes according to variations in the fertility index, k_t.Finally, $\epsilon_{x,t}$ is the error term that captures all remaining variations. The in-sample datasets were fitted into the Lee-Carter model as follows:

$$\begin{aligned}ln(f_{x,t,i})&=a_{x,i}+b_{x,i}k_{t,i}+\epsilon_{x,t,i}\\ x&=15\text{to}49, t=1958,...,2005\end{aligned} \tag{7}$$

$f_{x,t}$ represents ASFR for a single age x in a given year t, a_x is the average age-specific fertility over the years, which is estimated using this formula, $a_x=\frac{1}{T}\Sigma_{t=1}^{T}f_{x,t}$. The b_x parameter indicates the deviation in fertility due to changes in k_t, which is the fertility index. These two parameters were estimated using the SVD method $[ln(f_{x,t})-a_x]$ with the constraints of $\Sigma b_x=1$ and $\Sigma k_t=0$. applied to them. The $\epsilon_{x,t}$ represents a random error.

To forecast fertility trends using the Lee-Carter model, k_t must be predicted over time, typically as a stochastic process. Since k_t represents a univariate time series, $ARIMA(p,d,q)$ models may accurately forecast the rates. The projection of k_t using the $ARIMA(p,d,q)$ model is expressed as follows:

$$\begin{aligned}\Delta k_t=\mu+\phi_1\Delta k_{t-1}+\phi_2\Delta k_{t-2}+\phi_3\Delta k_{t-3}+...+\phi_p\Delta k_{t-p}-\\ \theta_1\epsilon_{t-1}-\theta_2\epsilon_{t-2}-\theta_3\epsilon_{t-3}-...-\theta_q\epsilon_{t-q}\end{aligned} \tag{8}$$

k_t denotes fertility index, μ denotes constant term of fertility index, and ϕ_i denotes the coefficient of AR model that describes the effect of a past observation k_{t-1} on the current observation k_t, where $i=1,2,3p$, while p determines the number of past values (lags) used. Additionally, the θ_j is the coefficient of MA model that represents the effect of past errors ϵ_{t-j} on the current observation, where $j=1,2,3...q$, while q specifies the number of past error terms included.

The forecasted ASFRs are estimated using the projected fertility index $\hat{k}_t$, along with the estimated a_x and estimated b_x. The forecast of ASFRs over a 16-year horizon, from 2006 to 2021 (s year from the forecast origin) can be obtained from the equation provided below:

$$\hat{f}_{x,t+s} = \exp(a_x + b_x \hat{k}_{t+s}) \tag{9}$$

2.4 Fitting the Modified Lee-Carter Model

For the modified Lee- Carter model, two models, A and B, were proposed, each based on different fitting periods were proposed. The estimation of model A is similar to the original Lee-Carter model, using complete in-sample data from 1958 to 2005. In contrast, the estimation of model B includes more recent development, using data from 2005 to 2021. The combination of the two models is expected to produce the predicted $f_{x,t}$ that forms a spiked fertility distribution, avoiding a deviation to the left side, as stated by impach and Arltova [18]. The development of $f_{x,t}$ using the modified Lee-Carter is described as follows:

$$f_{x,t} = \exp(\hat{a}_x^{(B)} + \hat{b}_x^{(A)} \hat{k}_t^{(A)}) \tag{10}$$

$\hat{a}_x^{(B)}$ represents the average age-specific fertility from model B, while $\hat{b}_x^{(A)}$ represents the deviation in fertility due to changes in k_t from model A, and $\hat{k}_t^{(A)}$ is the fertility index in years t from model A. Similar to the original Lee-Carter model, to predict the future $f_{x,t}$ it is essential to forecast the values of parameter k_t, usually calculated using $ARIMA(p, d, q)$ models. As the values of the parameters a_x and b_x are not impacted by time, causing the predictions using the Lee-Carter model purely extrapolative.

2.5 The Out-sample Forecast Error Measures

The out-sample forecast errors were measured as follows.
Mean Absolute Error:

$$MAE = \frac{|\Sigma_t \Sigma_x (\hat{f}_{x,t} - f_{x,t})|}{nm} \tag{11}$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{\Sigma_t \Sigma_x (\hat{f}_{x,t} - f_{x,t})^2}{nm}} \tag{12}$$

In the above equations, $\hat{f}_{x,t}$ denotes the forecasted ASFRs for age x and out-sample forecast years from 2006 to 2021 whereas $f_{x,t}$ denotes the observed rates for the same period. The variable n indicates the total number of years, which is 16, and m represents the total number of ages from 15 to 49, which is 35.

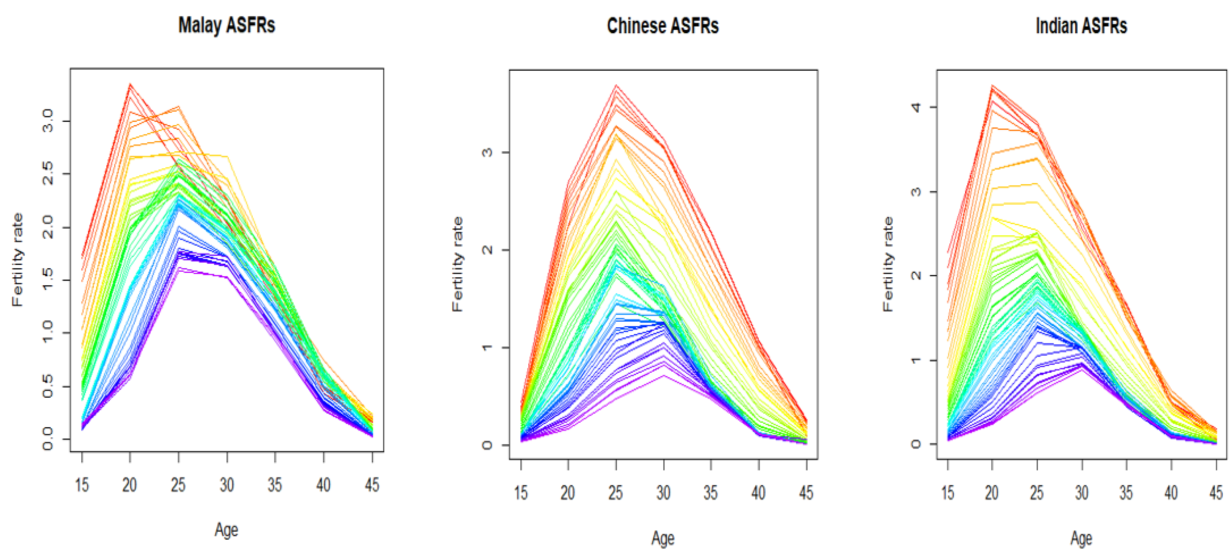
The model that results in the lowest out-of-sample forecast errors will be considered the most accurate and will be selected for projecting fertility rates 20 years into the future, covering the period from 2022 to 2041. R software was used to perform the analyses and modelling in this study, and the "demography" package, which is suitable for demographic analysis and forecasting, was specifically utilised for the Lee-Carter modelling.

3 Results and Discussion

The observed ASFRs for each ethnic group are presented in a 'rainbow plot' format, as shown in Figure 1. As the rates were based on five-year age intervals, the plots appear less smooth.

3.1 Trends of Observed ASFRs by Ethnicity

Figure 1 displays the rainbow plots of fertility rates for the following age groups: 1519, 2024, 2529, 30-34, 3539, 4044 and 4549. These plots show which age groups record the highest and lowest rates, while the colours indicate how the fertility rate declines over the years. As the years progress from the earlier periods, shown by red curves, to more recent years, shown by violet curves, the ASFR trends are seen to be declining steadily. Furthermore, over the years, the most common ages for childbirth among all three ethnic groups shifted to older ages. For instance, in 1958, Indian women recorded their highest birth rates between the ages of 20 and 25, while in 2021, this shifted to the age range of 30 to 35 years, indicating that women today often delay having children.

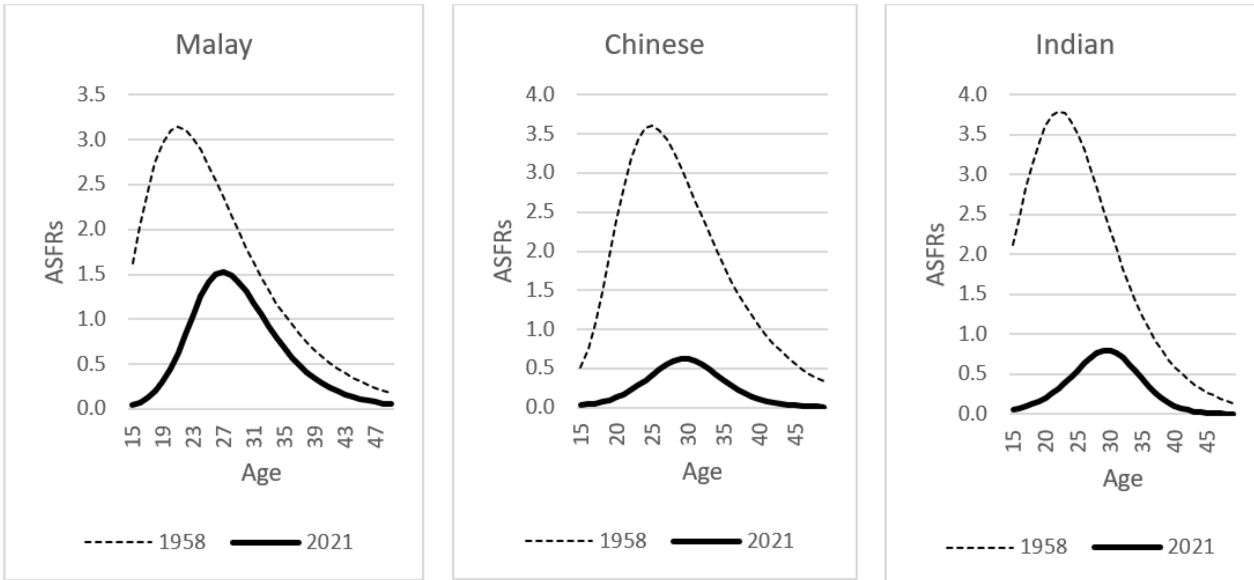


(a) (b) (c)

Figure 1: Rainbow plot of the observed Malay (a), Chinese (b) and Indian (c) ASFRs for age groups 15-45 from 1958-2021.

3.2 Estimating Single-age ASFRs for Different Ethnic Groups

Estimated single-age of ASFRs were obtained by fitting the observed 5-year age group data to a skew-logistic function for each year from 1958 to 2021, across three ethnic groups. This procedure was applied to each observed year, resulting in 64 curves. Figure 2 displays the estimated single-age fertility rates only for two years including 1958 (the earliest year) and 2021 (the latest year). The curves for single ages in Figure 2 display a smoother pattern compared to the observed curves. Moreover, fertility rates have dropped substantially, with the mode of the curves shifting to the right from 1958 to 2021. A comparison by ethnic groups showed that Chinese and Indian fertility rates have declined at a faster rate than those of the Malay population, as shown by a larger gap between the 1958 and 2021 curves.



(a) (b) (c)

Figure 2: Estimated Single-aged ASFRs plots for Malay (a), Chinese (b) and Indian (c) for ages 15 to 49 years old from 1958-2021, using the Skew-logistic distribution model.

Table 1 provides a comparison of estimated ASFRs for Malay, Chinese and Indian women in 1958 and 2021, along with the calculated percentage of declines in fertility rates over the period between 1958 and 2021. This analysis highlights substantial changes in fertility patterns across different age groups and ethnicities, showing overall declines in fertility rates. The declines in fertility rates are significant among younger women aged 15 and 20 across all three ethnic groups, with declines exceeding 90 percent in most cases. At age 25, the declines vary as Malay women experience a decline of 48 percent, while Chinese and Indian women experience declines of 88 percent and 84 percent, respectively. This suggests that Chinese and Indian women may have experienced more rapid fertility declines than Malay women during this peak fertility age. For older age groups, particularly those aged 40 and 45, the percentages of declines in fertility rates remain high, indicating a significant drop in fertility among these women, who historically may have had later-life childbirth.

The estimation of parameters was performed for each year across 64 curves using a non-linear least squares method. The range of parameters (minimum and maximum values) for each ethnic group is displayed in Table 2. The skewness values of the α parameters for Malaysian fertility range from -0.79 to 5.57 for all ethnic groups. This range is consistent with the findings of Mishra *et al.* [9], who reported that the α values ranged from 0 to 5 for four states in India. The skewness parameters for the Malay ethnic group were found to be positive for all years, indicating a positively skewed fertility curve. In contrast, the skewness parameter values for the Chinese and Indian ethnic groups showed both positive and negative values, with more recent years curves showing slight negative skewness trends. The goodness-of-fit was evaluated

Table 1: Reduction Percentages of Age-specific Fertility Rates from 1958 to 2021 by Ethnic Groups

Age	Malay			Chinese			Indian		
	1958	2021	Fertility Reduction (%)	1958	2021	Fertility Reduction (%)	1958	2021	Fertility Reduction (%)
15	1.627	0.046	97%	0.512	0.031	94%	2.123	0.061	97%
20	3.102	0.442	86%	2.398	0.131	95%	3.617	0.207	94%
25	2.718	1.405	48%	3.602	0.416	88%	3.513	0.550	84%
30	1.795	1.306	27%	2.860	0.623	78%	2.307	0.801	65%
35	1.051	0.679	35%	1.810	0.351	81%	1.228	0.437	64%
40	0.577	0.287	50%	1.032	0.111	89%	0.591	0.104	82%
45	0.307	0.113	63%	0.557	0.029	95%	0.273	0.017	94%

Table 2: Range of Estimated Parameters from the Skew-logistic Model from 1958 to 2021 by Ethnic Groups

Ethnic	Skewness (α) (min, max)	Location (μ) (min, max)	Scale (σ) (min, max)	R^2 (min, max)	TFR (1958, 2021)
Malay	$(0.82, 5.57)$	$(15.27, 25.22)$	$(4.27, 8.33)$	$(0.9536, 0.9962)$	$(5.91, 2.28)$
Chinese	$(-0.79, 3.14)$	$(19.66, 32.47)$	$(3.37, 7.48)$	$(0.9652, 0.9980)$	$(6.71, 0.98)$
Indian	$(-0.466, 4.52)$	$(16.43, 31.57)$	$(3.38, 7.38)$	$(0.9755, 0.9992)$	$(7.39, 1.18)$

using the coefficient of determination, R^2, which yielded values greater than 95% for all curves, indicating the variation in the dependent variable can be well explained by the parameters of the model.

3.3 Fitting of ASFRs using the Lee-Carter and Modified Lee-Carter

This part presents a detailed overview of the applications of both the original Lee-Carter model and its modified version. The estimated smoothed single-age fertility rates derived from the skew-logistic distribution model were applied to both models. The key distinction between the two models lies in the fitting period for the a_x parameter. The original model uses data from 1958 to 2005, while the modified version uses a more recent period from 1991 to 2005, encompassing the latest 15 years of data. According to [18], the modification of the Lee Carter model was designed to capture recent trends in fertility rates, which have shown a decrease and a slight shift to the right.

Figure 3 depicts a concave downward shape of a_x parameters for both the Lee-Carter and modified Lee-Carter models, representing the overall age fertility profile. The values of a_x are the lowest at the youngest and oldest reproductive ages, while they are higher between the ages of 20 and 35. Moreover, the value of parameter a_x for the modified Lee Carter model is shown to be lower than that of the original Lee-Carter model and is slightly shifted towards older ages. These changes can be attributed to the recent trend of declining fertility rates and the shift in the age at which women experience their highest fertility.

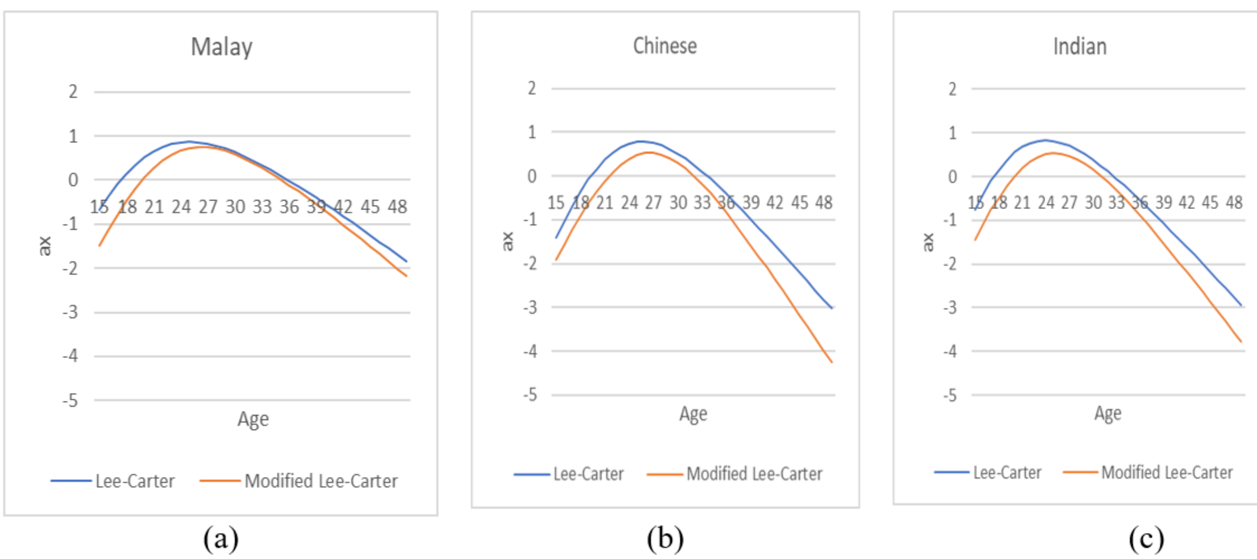
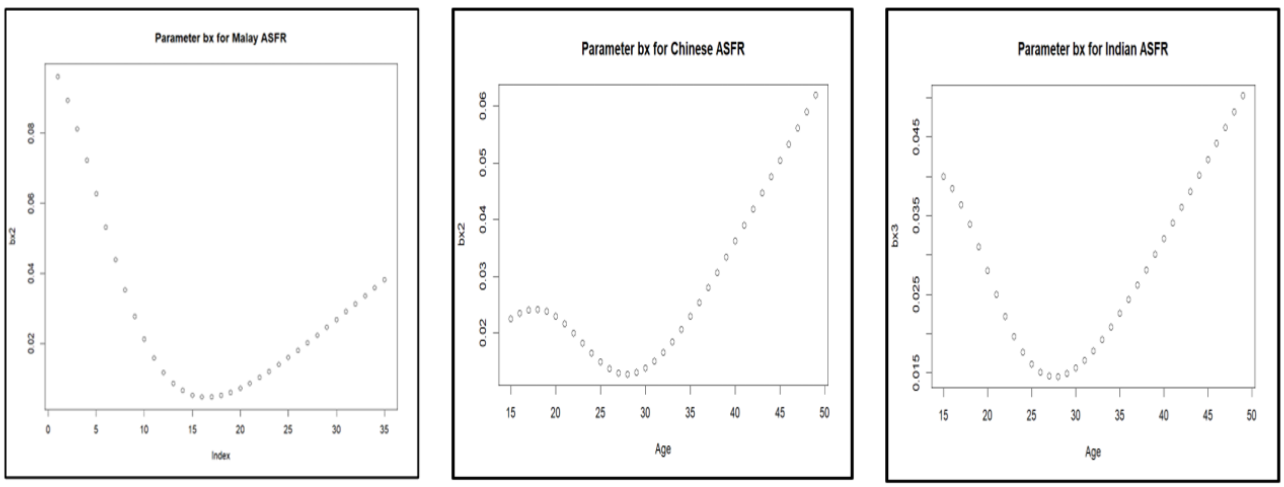


Figure 3: Estimated a_x parameters from the Lee-Carter (blue curves) and modified Lee-Carter models (orange curves) for Malay (a), Chinese (b) and Indian (c).

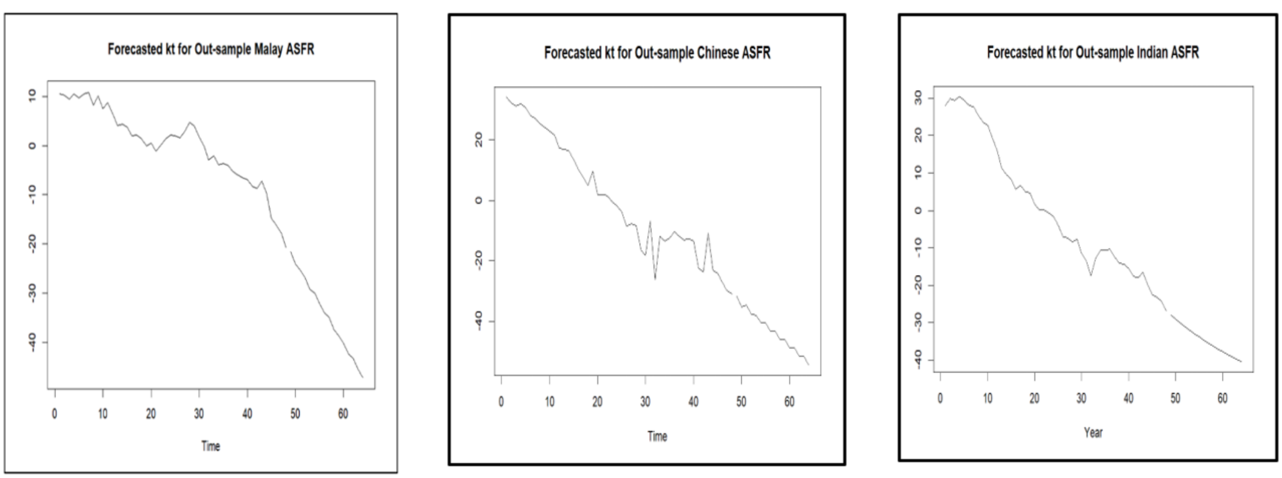
Figure 4 and Figure 5 present the estimates of b_x and k_t parameters of both the Lee-Carter model and the modified Lee-Carter model. The k_t parameter provides insights into how the fertility rate evolves over time, while b_x parameter describes the extent of changes in fertility by age when k_t changes. The trends for the estimated k_t parameters show a decline over the years, with the Chinese and Indian data portraying nearly linear decreasing patterns. The estimated

k_t parameters were forecasted for the out-sample or evaluation period from 2006 to 2021 using the best ARIMA(p, d, q) models. These models derived the estimates for Malay, Chinese, and Indian populations: ARIMA$(2, 2, 3)$ for Malay, ARIMA$(5, 2, 3)$ for Chinese and ARIMA$(1, 1, 1)$ for Indian, and were selected based on the lowest AIC values, as displayed in Table 3.



(a) (b) (c)

Figure 4: Estimated b_x parameters from the Lee-Carter model for Malay (a), Chinese (b) and Indian (c).



(a) (b) (c)

Figure 5: Estimated k_t parameters from the Lee-Carter model for Malay (a), Chinese (b) and Indian (c).

Table 3: The AIC values for $ARIMA(p, d, q)$ models for each ethnic group

Malay		Chinese		Indian	
Model	AIC	MODEL	AIC	Model	AIC
ARIMA (0,1,0)	181.00	ARIMA (0,1,2)	280.90	ARIMA (0,1,0)	205.75
ARIMA (1,2,3)	179.15	ARIMA (6,2,5)	281.63	ARIMA (0,1,1)	202.62
ARIMA (3,2,3)	182.11	ARIMA (5,2,4)	282.45	ARIMA (1,1,1)	**197.03**
ARIMA (2,2,2)	178.84	ARIMA (5,2,3)	**280.55**	ARIMA (1,1,0)	199.80
ARIMA (2,2,3)	**176.29**	ARIMA (6,2,4)	281.83	ARIMA (0,1,2)	200.94

The Lee-Carter model's equation (9) and the modified Lee-Carter model's equation (10) were used to estimate the ASFRs forecast values, using the estimated values of a_x and b_x parameters and the projected values of k_t. The out-of-sample errors were then calculated by comparing these ASFR forecasts with the observed values to assess the forecasting accuracy. The performance of the Lee Carter and modified Lee Carter models for all ethnic groups is indicated by the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values presented in Table 4. The modified Lee Carter model outperformed the original Lee-Carter model for Chinese and Indian populations, as indicated by lower error levels. The model is superior to the original version for predicting substantial declines in ASFRs within Chinese and Indian populations in recent years. Conversely, for the Malay population, the modified model underperformed the Lee-Carter model due to larger error values. This difference arises because the fertility rates among the Malay population have decreased slower compared to the other two ethnic groups. Consequently, recent trends have not given sufficient weight to the overall fertility decline in this group.

Table 4: The Out-sample ASFR Forecast Errors from the Lee-Carter (LC) model and the modified Lee-Carter (MLC) model, using the error metrics of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), for three ethnic groups in Malaysia.

Ethnic	MAE		RMSE	
	LC	MLC	LC	MLC
Malay	**0.09663**	0.1506	**0.1461**	0.2136
Chinese	0.1267	**0.1260**	0.1810	**0.1746**
Indian	0.1389	**0.1275**	0.2119	**0.1743**

The extended forecasts spanning a 20-year period for ASFRs from 2022 to 2041 were performed across the entire population and within three major ethnic groups in Malaysia: Malay, Chinese and Indian. The results, presented in Figure 6, distinctly indicate a downward trajectory in ASFRs over the years, from earlier years (represented by the red curves) to later years (represented by the violet curves). Among the three groups, the Malay ethnic group is predicted to have the highest future ASFRs. Conversely, the Chinese and Indian ethnic groups are predicted to experience relatively lower ASFRs over the next 20 years. The Malay ethnic group's fertility is declining at a relatively slower rate compared to the others. Therefore, their TFR is projected to fall below the replacement threshold of 2.1 for the first time in 2031, with an expected TFR value of 2.07 that year. The higher fertility rate among Malay women can be attributed to factors such as an earlier average age at first marriage, a higher proportion of women entering marriage and relatively lower levels of birth control practices. Notably, the trend of delaying the first birth among Malaysian women is indicated by a gradual shift in the peak ages for childbirth to older ages. The ASFRs for Chinese and Indian women are predicted to be lower than those of the Malay group, as seen by the shorter rainbow curves for the Chinese and Indian groups in Figure 6 compared to those for the Malays. The gap between the curves for 2022 and 2041 indicates that Chinese fertility rates are declining more rapidly than those of the Malay and Indian populations. For the Chinese population, the TFR fell below 1.0 in 2021, reaching 0.98, and this is expected to fall further from 0.82 in 2022 to 0.23 in 2041. In the case of the Indian, the TFR was 1.18 in 2021 and is projected to drop from 0.97 in 2022 to 0.62 in 2041.

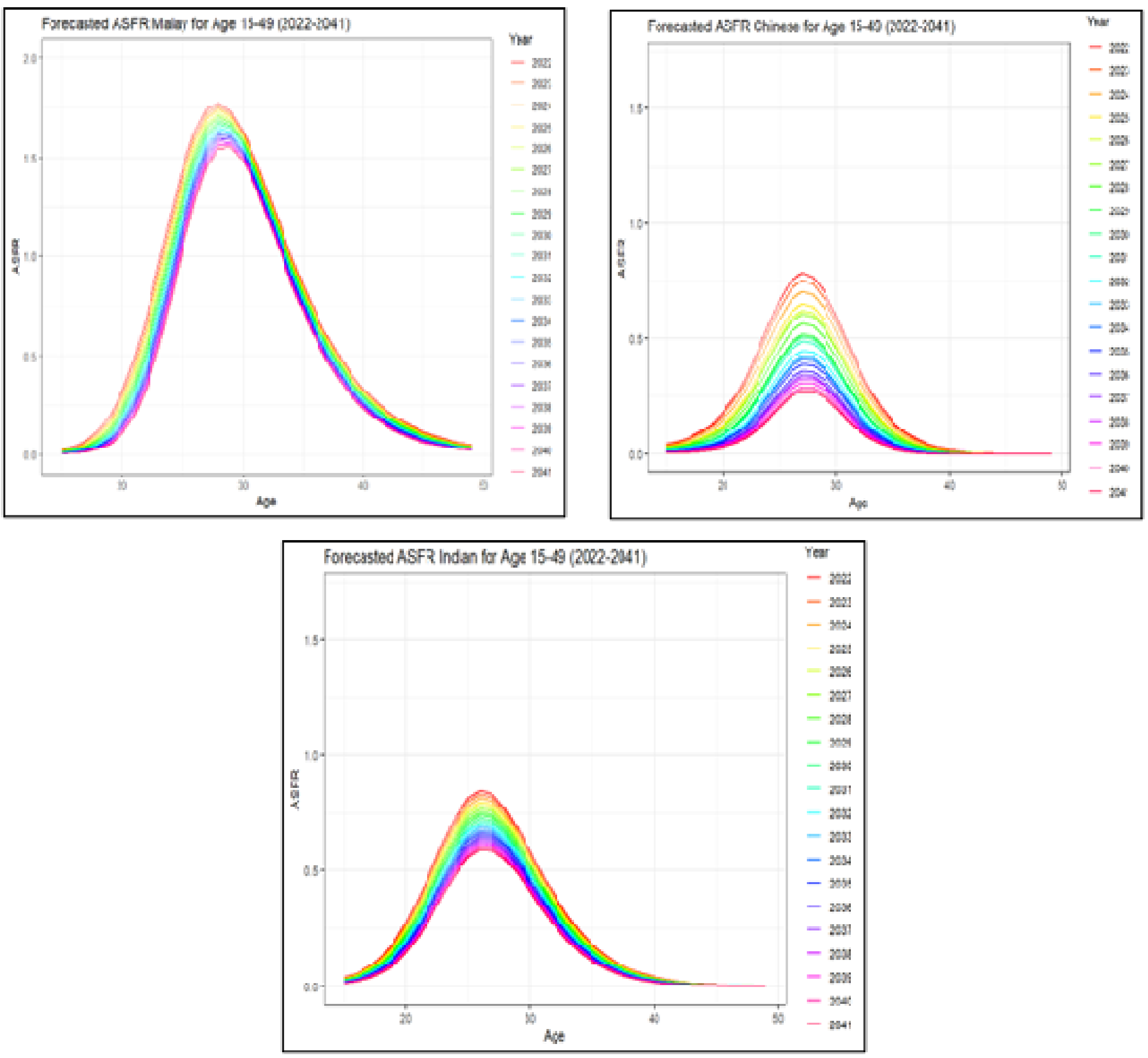


Figure 6: Forecasted ASFRs Malay, Chinese and Indian from 2022 to 2041.

TFR values below replacement ratios have recently been observed in several nations, particularly in more developed regions, such as Northern America and Europe. However, TFR values falling below 1.0 raise significant concerns since this suggests that each couple has only one child to care for or replace them. This trend is exemplified in South Korea, where the TFR was recorded at 0.78 [21] in 2023. The question is raised as to whether it is appropriate for the Chinese TFR projections to go below 0.5. Compared to other nations, a smaller population groups in Xiangyang district of Jiamusi city, China, recorded the lowest TFR anywhere in the world, which was only 0.41 in the year 2000 [22] .

The convergence of decreasing fertility rates and changing mortality trends is expected to contribute to an aging demographic in Malaysia [23]. Consequently, the government may face challenges in providing adequate support for a growing elderly population and may be obliged to impose higher taxes on a smaller working-age population. Given the projections from this research, it is imperative for the government to implement comprehensive initiatives that promote reproductive awareness and education among the public. Responding to the declining birth rates, the government announced in Budget 2020 that funds would be allocated for children's facilities, especially in hospitals and schools. Government agencies are also encouraged to open childcare services and schools for young children [24]. Additionally, parents receive tax relief when they utilise registered childcare services, and couples who seek fertility treatments, such as IVF, also receive tax relief. Maternity leave has also been extended from 60 days to 90 days to ensure that mothers are completely recovered before returning to work. These initiatives aim to lessen the financial burdens on couples seeking medical assistance to conceive and subsequently to reverse the declining fertility pattern in Malaysia.

4 Conclusion

The global landscape of fertility has undergone significant shifts, particularly in developing countries including Malaysia. Declining fertility rates have led to lower overall population growth, indicating that these countries are undergoing critical challenges in maintaining the ability to replace the current population. This research proposes a modified version of the Lee-Carter model that integrates the Skew-logistic distribution method to estimate single-age ASFRs before fitting the data into the Lee-Carter model. In addition, the a_x parameters were modified to capture the significant decline in fertility observed in recent year. Overall, the skew-logistic model showed a good fit for the ASFRs across each ethnic group. The original Lee-Carter model was the best fit for the Malay population for forecasting over the next 20 years, while the modified Lee-Carter model proved more well suited for the Chinese and Indian populations, both of which have experienced major fertility changes in recent years. Projections provided insights that the TFR for the Malay population is expected to fall below the replacement ratio for the first time in 2031, whereas the TFR for the Chinese and Indian populations are projected to reach their lowest historical values by 2041. Hence, it is vital for the government to regularly monitor the population's fertility rates and periodically assess the effectiveness of current fertility and childbirth policies to ensure that they align with national agendas.

References

[1] Department of Statistics Malaysia. (2022, February 1). Current Population Estimates, Malaysia, 2022.
https://v1.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=dTZXanV6UUdyUEQ0SHNWOVhpSXNMUT09

[2] Visalakshi, S., and Geetha, R. Modelling age specific fertility rate in India through fertility curves. Bulletin of Pure & Applied Sciences - Mathematics and Statistics.
https://doi.org/10.5958/2320-3226.2018.00042.5

[3] Chandola T., Coleman D.A., and Hiorns R.W. Recent European fertility patterns: fitting curves to 'distorted' distributions. Popul Stud (Camb). 1999. Nov; 53(3):317-29.
https://doi.org/10.1080/00324720308089

[4] Peristera, P and Kostaki, A. Modeling fertility in modern populations. Demographic Research. 2007. 16(6):169–196.
https://doi.org/10.4054/DemRes.2007.16.6

[5] Schmertmann, C. P. A system of model fertility schedules with graphically intuitive parameters. Demographic Research. 2003. 9:81-110.

[6] Gaire, A. K., Thapa, G. B., and KC, S. Mathematical Modeling of Age-Specific Fertility Rates of Nepali Mothers: A Polynomial Approach. Pakistan Journal of Statistics and Operation Research. 2022. 18(2):417–426. https://doi.org/10.18187/pjsor.v18i2.3319

[7] Mazzuco, S., and Scarpa, B. Fitting Age-Specific Fertility Rates by a Flexible Generalized Skew Normal Probability Density Function. Journal of the Royal Statistical Society Series A (Statistics in Society). 2014. 178(1). https://doi.org/10.1111/rssa.12053

[8] Asili, S., Rezaei, S. and Najjar, L. Using Skew-Logistic Probability Density Function as a Model for Age-Specific Fertility Rate Pattern. BioMed Research International. 2014. Article ID 790294 https://doi.org/10.1155/2014/790294

[9] Mishra, R., Singh, K. K., & Singh, A. (2017). A Model for Age-Specific Fertility Rate Pattern of India Using Skew-Logistic Distribution Function. American Journal of Theoretical and Applied Statistics. 2017. 6(1):32–37. https://doi.org/10.11648/j.ajtas.20170601.14

[10] Mathivanan, N. M. N., Ghani, P. A. and Ghani, N. A. M. Tracing Mathematical Function of Age Specific Fertility Rate in Peninsular Malaysia. The Indonesian Journal of Electrical Engineering and Computer Science. 2018. 9(3):637–642. https://doi.org/10.11591/ijeecs.v9.i3.pp637-642

[11] Shitan, M. and Ng, Y. L. Forecasting the total fertility rate in Malaysia. Pakistan Journal of Statistics. 2015. 31(5):547–556.

[12] Wang, Y. Analysis of China's Birth Rate Prediction based on Time Series. Advances in Economics, Management, and Political Sciences. 2023. 55(1):205–217. `https://doi.org/10.54254/2754-1169/55/20231009`

[13] Lee, R. D. and Carter, L. R. Modeling and forecasting US mortality. Journal of the American Statistical Association. 1992. 87(419):659–671. `https://doi.org/10.1080/01621459.1992.10475266`

[14] Lee, R. D. Modeling and forecasting the time series of US fertility: age distribution, range, and ultimate level. International Journal of Forecasting. 1993. 9(2):187–202. `https://doi.org/10.1016/0169-2070(93)90004-7`

[15] Hyndman, R. J. and Ullah, M. S. Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis. 2007. 51(10):4942–4956. `https://doi.org/10.1016/j.csda.2006.07.028`

[16] Winiowski, A., Smith, P. W., Bijak, J., Raymer, J. and Forster, J. J. Bayesian population forecasting: Extending the Lee-Carter method. Demography. 2015. 52(3):1035–1059. `https://doi.org/10.1007/s13524-015-0389-y`

[17] Hanafiah, H. and Jemain, A. A. Structural modeling of age specific fertility curves in Peninsular Malaysia: An approach of Lee Carter method. AIP Conference Proceedings. 2013. 1571(1). 1062-1068. `https://doi.org/10.1063/1.4858794`

[18] impach, O. and Arltova, M. An increasing of prediction power of the Lee-Carter model: The case of Czech and Spanish age-specific fertility rates' forecasting. In Proceedings ITISE 2016. Granada 27-29, June, 2016.

[19] World Health Organization. Women of reproductive age (15-49 years) population (thousands). Global Health Observatory. Retrieved [28 Oct 2024] from `https://www.who.int/data/gho/indicator-metadata-registry/imr-details/women-of-reproductive-age-(15-49-years)-population-(thousands)`

[20] Azzalini, A and Regoli, G. Some properties of skew-symmetric distributions. Annals of the Institute of Statistical Mathematics. 2010. 64(4). `https://doi.org/10.1007/s10463-011-0338-5`

[21] McCurry, J. South Korea's birthrate sinks to fresh record low as population crisis deepens. The Guardian. 2023. Retrieved from `https://www.theguardian.com/world/2023/feb/22/south-koreas-birthrate-sinks-to-fresh-record-low-as-population-crisis-deepens`

[22] Terrell, H. K. M. Fertility in China in 2000: A County Level Analysis (Master's thesis). Texas A & M University.2005. Retrieved from `https://web.archive.org/web/20221214155317/https://oaktrust.library.tamu.edu/handle/1969.1/3892`

[23] Ibrahim, N. S. M., Lazam, N. M. and Shair, S. N. Forecasting Malaysian mortality rates using the Lee-Carter model with fitting period variants, J. Phys.: Conf. Ser., vol. 1988, no. 1, p. 012103, 2021.

[24] Inland Revenue Board of Malaysia. (2020). *Budget 2020* https://phl.hasil.gov.my/pdf/pdfam/Budget_2020.pdf