Pak. J. Statist.
2015 Vol. 31(5), 547-556

FORECASTING THE TOTAL FERTILITY RATE IN MALAYSIA

Mahendran Shitan[§] and **Yung Lerd Ng**
Laboratory of Computational Statistics and Operations Research,
Institute for Mathematical Research, Universiti Putra Malaysia, Serdang, Malaysia
[§]Email: mahen@upm.edu.my

ABSTRACT

It is vital to understand the demographic development of the country as demographic changes would affect all areas of human activity. Forecasting demographic variables is important as demographic trends which are neglected could be discovered and new policies can be implemented before situation got worst. A time series model is fitted to forecast the Total Fertility Rate (TFR) in Malaysia. The Autoregressive Integrated Moving Average (ARIMA) models and ARAR models are considered and the forecasting performance of these models are evaluated by using post sample forecasting accuracy criterion. It is found that the ARAR model appeared to be the most appropriate model for forecasting the total fertility rate in Malaysia. Based on the forecasts, the TFR in Malaysia is projected to decline and will slowly level off and is expected to be approximately 1.2 (average number of children per women) for the year 2040. A 95% confidence level for year the 2040 is in the range of 0.5 to 1.9 children per women.

KEYWORDS

Time Series; Fertility decline; ARIMA model; ARAR algorithm.

1. INTRODUCTION

As demographic changes affect all areas of human activity economically, socially, culturally and even politically (McKenna, 1996), it is important for us to pay close attention to our demographic development for policy decision making and the country's future growth.

Amongst the demographic indicators, Total Fertility Rate (TFR) will be the main focus in this study. Hirschman (1994) defined TFR as the average number of children born to a woman who survives up to the age of 50 years. If TFR drops, then eventually it may lead to an aged population. World TFR has declined by 37% since 1970-1975 and for many developing countries, the decline is mainly caused by postponement of marriage and contraceptive use such as increasing age at marriage and stopping behaviour among high-parity older women (United Nations Secretariat). In addition, increased number of higher educated women, urbanisation of a country, declining mortality, growing demand in limiting family sizes also boost the decline in fertility rate.

In Malaysia too, TFR is an important concern. For instance, two states in Malaysia, namely Kelantan and Terengganu (which had very high TFR before 1990), TFR has declined sharply from 1991. The decline is partly due to the increase in education level

(Soon, 1992; Tey, 2002; Caldwell and Caldwell, 2005). According to Lim *et al.* (1987), 75% of female aged group of 15-19 in 1980 went to secondary school whereas only 15% of the previous generation were able to do so. It is found that in 2010, the proportion with tertiary education labour force is 24% (Department of Statistics Malaysia, 2010) compared to only 3% in 1980 (Tey, 2002).

The drop in TFR is also partly due to postponement of marriage by women and knowledge of fertility control and use of contraceptives (Leete and Kwok, 1986; Soon, 1992; Norville *et al.,* 2003; Brown, 2004). As mentioned in Lim *et al.* (1987), control of fertility is one of the key factor influencing fertility of married women. The prevalence rate of contraceptive methods has increased from 9% in 1966 to 58% in 1994 (Tey, 2002). Unlike the past, childbearing has lost its stand as a primary role in women's lives as it has become an option that can be scheduled and sequenced with vocational and lifestyle pursuits (Hirschman, 1994). Furthermore, since the financial crisis in 1997, many would consider appropriate family sizes since having large families can be such a burden considering the increasing education expenses and other expenditures today. Therefore, high level of household economic strains were also one of the factors which causes low fertility rate as individuals tend to postpone or not get married at all, use abortion and so on (Davis, 1963; Norville *et al.,* 2003).

Figure 1 shows a plot of 43 annual TFR (unit of measurement: average number of children per woman) observations from 1970 to 2012 (Department of Statistics, Malaysia, 2013) which clearly indicates that TFR has been declining over the last few decades and as such, it would be interesting to forecast the future TFR for Malaysia.

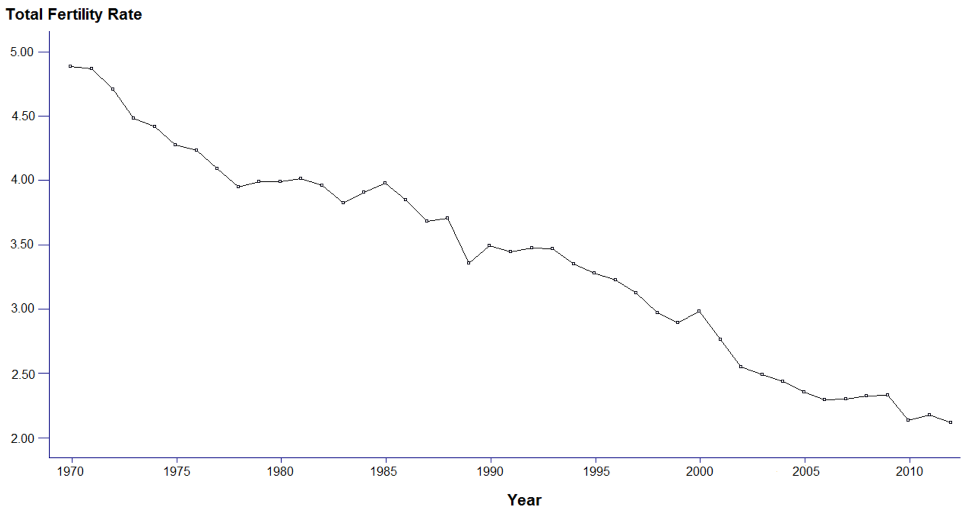

Fig. 1: Total Fertility Rate in Malaysia from 1970 – 2012

In previous literatures, demographic variables were predicted with extrapolative methods, classical cohort models or generated from structural stochastic econometric models. Other than that, forecast of demographic variables were also generated by Box-Jenkins (BJ) method which is one of the popular methods used by many researchers (Saboia, 1977; Land and Cantor, 1983; Carter and Lee, 1986; McNown and Rogers, 1989; Knudsen *et al.*, 1993; Carter, 1996). The Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model is one of the popular models used in the literature. ARIMA model has been applied by Saboia (1977), McDonald (1979), Land and Cantor (1983), Carter and Lee (1986), McNown and Rogers (1989), Knudsen *et al.* (1993) and Carter (1996) in modelling and forecasting fertility rate as well as other demographic variables. McDonald (1979) studied the relationship between classical demographic deterministic forecasting models, stochastic structural econometric models and time series models, and found that the transfer function models can provide improved forecasts as compared to univariate Autoregressive Moving Average (ARMA) models. Carter and Zellner (2003) whose study was on the ARAR error model for univariate time series and distributed lag models compared to ARAR forecasting performance with traditional ARMA model and found that ARMA routine took more iterations than ARAR routine to converge. They stated that the ARAR model is simpler than the ARMA model as ARAR model is parsimonious, allows rich time series behaviour and simplifies the inference procedure.

The main aim of this study is to fit a time series model for the TFR data and to forecast TFR in Malaysia from 2013 to 2040 by using the ARIMA (*p*,*d*,*q*) and ARAR models.

2. METHODOLOGY

The general procedure of fitting a time series model involves identification, estimation, validation and forecasting. The time series can be forecasted by several models. However, we only consider the ARIMA (*p*,*d*,*q*) and ARAR models. The ARIMA (*p*,*d*,*q*) and ARAR models (Brockwell and Davis, 2002) are defined as

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \quad (1)$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables with mean zero and variance σ^2 and is denoted by $\{Z_t\} \sim WN(0, \sigma^2)$. The time series $\{X_t\}$ is an ARMA (*p*,*q*) process with mean μ if $\{X_t - \mu\}$ is an ARMA (*p*,*q*) process. The time series $\{X_t\}$ is an ARIMA (*p*,*d*,*q*) process if $Y_t = (1 - B)^d X_t$ is a causal ARMA (*p*,*q*) process defined by

$$\phi(\mathrm{B}) Y_t = \theta(B) Z_t \quad (2)$$

where $\{Z_t\} \sim WN(0, \sigma^2)$, $\phi(\mathrm{z}) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(\mathrm{z}) = 1 + \theta_1 z + \cdots + \theta_q z^q$, B is the backward shift operator and $\phi(\mathrm{z}) \neq 0$ for $|\mathrm{z}| \leq 1$.

The ARAR algorithm is one of the forecasting techniques that has been found to be useful for a wide range of real data sets. The idea of ARAR algorithm is to apply a memory-shortening transformation to the data followed by fitting a subset of AR model to the transformed data. The ARAR algorithm consists of three important steps. First of all, a time series is classified as long-memory process, moderately long-memory process or short-memory process according to a five-step algorithm (Brockwell and Davis, 2002). Then, for the memory-shortened series, an autoregressive process is fitted to the mean corrected series and finally, forecasting analysis is performed. The memory shortened series is expressed as

$$S_t = Y_t + \psi_1 Y_{t-1} + \cdots + \psi_k Y_{t-k} \tag{3}$$

where $\psi_0, \psi_1, \cdots, \psi_k$ are coefficients of memory-shortening filter, $t = 1, \cdots, T$, S_t is the memory-shortened series and $\bar{S}$ is denoted as the sample mean of $S_1, \cdots, S_T$. The AR model fitted to the mean corrected series $S_t - \bar{S}$ has the form

$$X_t = \phi_1 X_{t-1} + \phi_{l_1} X_{t-l_1} + \phi_{l_2} X_{t-l_2} + \phi_{l_3} X_{t-l_3} + Z_t \tag{4}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$,$l_1, l_2, l_3$ are lag values, coefficient ϕ_j and white noise variance σ^2 are found from the Yule-Walker equations (Brockwell and Davis, 2002).

3. RESULTS AND DISCUSSIONS

The TFR data from 1970 to 2012 (consisting of 43 observations) are used in the modelling process. The last 10 observations which are data from 2003 to 2012 are used in evaluation of forecasting performance of the fitted models.

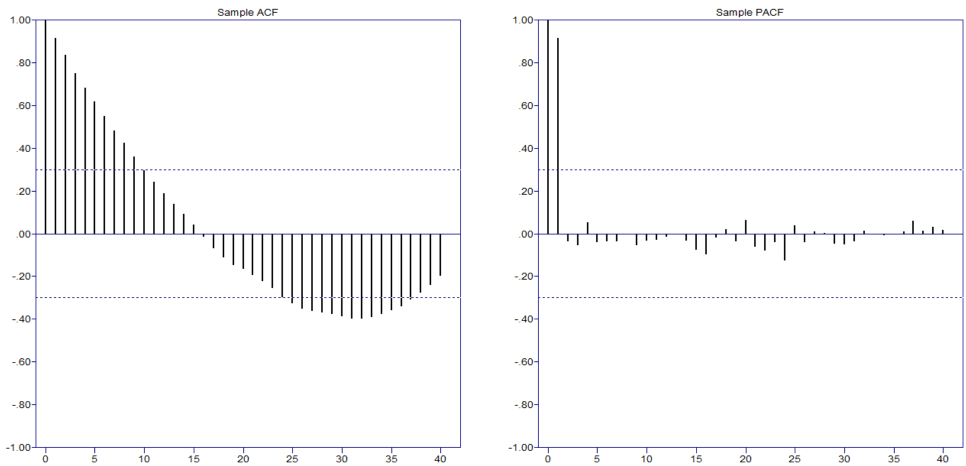

Fig. 2: ACF and PACF of the Original Time Series

In Figure 2, the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) are shown. The plot of sample ACF shows that the autocorrelation function exhibits a slow decay as the number of lag increases indicating that a trend is contained in the data and hence the series is non-stationary.

Since the time series for TFR $\{Y_t: t = 1, 2, \cdots, 43\}$ is not stationary, transformation is needed to eliminate the trend component. The time series, $\{Y_t\}$ is differenced at lag-1 twice to obtain stationarity and thus giving,

$$V_t = (1 - B)^2 Y_t \tag{5}$$

The plot of $\{V_t\}$ is shown in Figure 3.

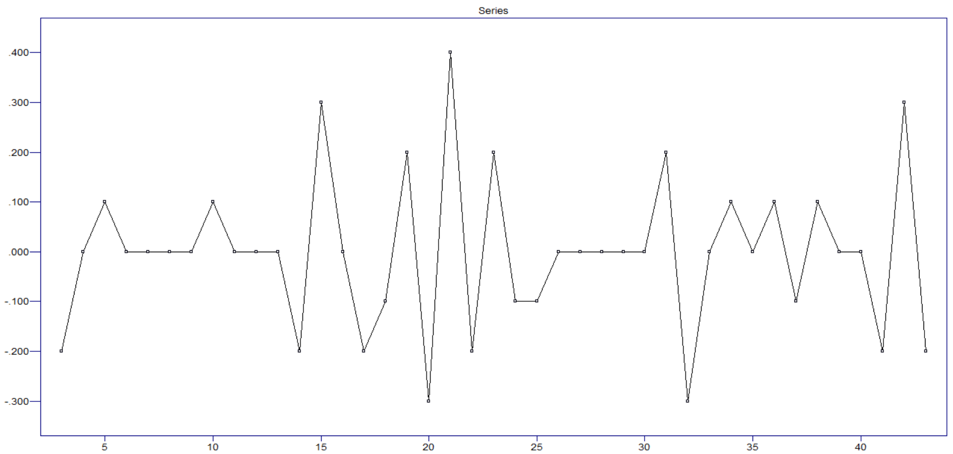

Fig. 3: Plot of time series $\{V_t\}$

To model the time series as zero mean stationary process, the mean of $V_t = -0.0024$ was subtracted from V_t to give

$$X_t = V_t - (-0.0024) \tag{6}$$

The sample ACF and sample PACF plot of X_t are shown in Figure 4.

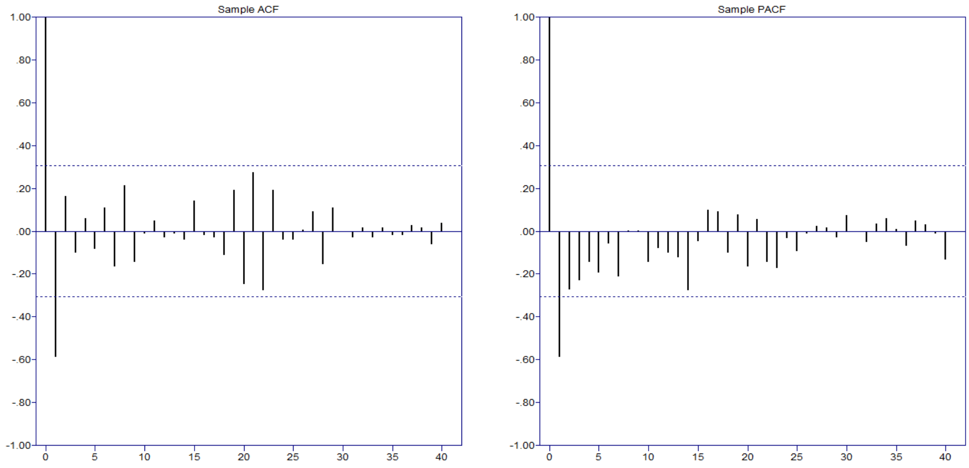

Fig. 4: Sample ACF and PACF of $\{X_t\}$

Plausible models for the transformed series can be observed from the sample ACF and sample PACF in Figure 4. Both ACF and PACF value at lag-1 which are significantly different from zero at 95% confidence level suggested that MA(1) and AR(1) models should be taken into consideration. Furthermore, mixed models of ARMA

(1,1) model is considered as well. The appropriateness of each models can be checked by using diagnostic tests namely, Ljung-Box test, McLeod-Li test, Difference-Sign test, Rank test and Jarque-Bera test such that residuals of the models have to pass these tests before they can be considered as a model for forecasting TFR in Malaysia.

Table 1
Diagnostic Tests for ARIMA models

No	**Model**	**Ljung-Box**	**McLeod-Li**	**Difference Sign**	**Rank**	**Jarque-Bera**
1	ARIMA (1,2,0)	18.377 (0.563)	38.042 (0.013)	18.000 (0.285)	406.000 (0.928)	1.601 (0.449)
2	ARIMA (0,2,1)	16.851 (0.663)	23.589 (0.313)	19.000 (0.593)	425.000 (0.736)	0.623 (0.732)
3	ARIMA (1,2,1)	17.633 (0.612)	21.417 (0.495)	20.000 (1.000)	419.000 (0.840)	1.229 (0.541)

Table 1 shows the test statistics and p-values (in parentheses) of diagnostic tests for the plausible ARIMA models considered in this study. From Table 1, it can be seen that the residuals from the ARIMA(1,2,0) does not pass the McLeod-Li test, since the residuals of ARIMA(1,2,0) model is statistically significant at 5% level. This indicates that the residuals may not be identically and independently distributed. Therefore the ARIMA(1,2,0) was discarded from our study.

To evaluate the forecasting performance of the models, post sample forecast accuracy criterion such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are used. The MAE, RMSE and MAPE are defined as

$$\text{Mean Absolute Error, } MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{7}$$

$$\text{Root Mean Square Error, } RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

$$\text{Mean Absolute Percentage Error, } MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|}{n} \times 100\% \tag{9}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value and n is the number of predicted values. The model with the smallest values of MAE, RMSE and MAPE will be the most appropriate model for forecasting. Ten points were used in order to evaluate the forecast accuracy and the results are shown in Table 2.

Table 2
MAE, RMSE and MAPE values for the time series considered in this study

No	**Model**	**MAE**	**RMSE**	**MAPE(%)**
1	ARIMA(0,2,1)	0.308	0.383	13.909
2	ARIMA(1,2,1)	0.303	0.381	13.728
3	ARAR Forecast	0.075	0.083	3.292

Results show that ARAR model appeared to be the best model and out-performed the ARIMA models, as it has the smallest MAE, RMSE and MAPE values which are 0.075, 0.083 and 3.292% respectively. Based on the forecast accuracy criterion, MAPE for the ARAR model indicates that on average, our forecasts were off by less than 4% error.

Hence, based on the ARAR model, we predicted 28 periods ahead together with the prediction bounds. The forecasted values from year 2013 to 2040 together with the 95% forecast intervals are shown in Table 3 and a plot of these forecasted values are shown in Figure 5.

Table 3
Forecasted values for TFR in Malaysia from 2013 to 2040

Year	Prediction	95% Prediction Bound	
		Lower	Upper
2013	2.087	1.892	2.282
2014	2.036	1.786	2.286
2015	2.004	1.707	2.300
2016	1.973	1.638	2.308
2017	1.931	1.562	2.298
2018	1.861	1.464	2.259
2019	1.859	1.436	2.283
2020	1.816	1.368	2.263
2021	1.774	1.295	2.252
2022	1.743	1.244	2.242
2023	1.707	1.194	2.221
2024	1.676	1.147	2.206
2025	1.641	1.098	2.184
2026	1.604	1.048	2.161
2027	1.582	1.014	2.151
2028	1.550	0.969	2.130
2029	1.515	0.923	2.108
2030	1.489	0.887	2.090
2031	1.459	0.848	2.070
2032	1.431	0.811	2.052
2033	1.403	0.774	2.032
2034	1.375	0.737	2.013
2035	1.350	0.705	1.996
2036	1.323	0.670	1.977
2037	1.296	0.635	1.957
2038	1.272	0.605	1.939
2039	1.247	0.573	1.921
2040	1.223	0.542	1.903

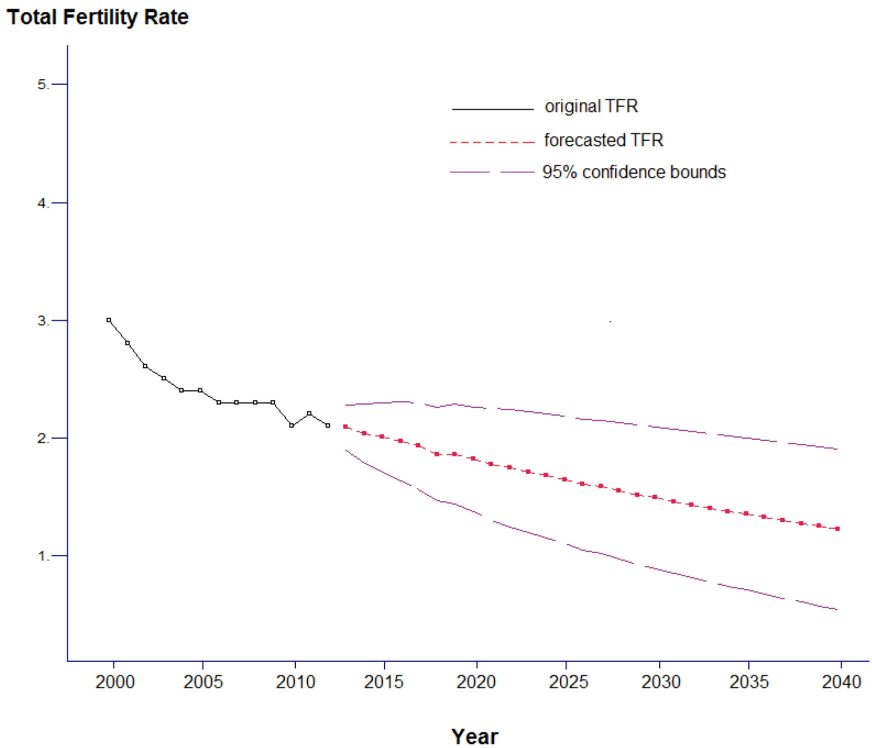

Fig. 5: Forecast for TFR in Malaysia from 2013 to 2040

From Figure 5, it can be seen that the TFR in Malaysia from 2013 to 2040 slowly drops and gradually levels off and is expected to be in the range of 0.5 to 1.9 births per woman at 95% confidence level for year 2040.

4. CONCLUSION

The objective of this study is to find an appropriate time series model to forecast total fertility rate in Malaysia from 2013 to 2040 and various time series models such as ARIMA and ARAR models have been considered.

We found that the ARAR model is an appropriate model for forecasting TFR in Malaysia with MAE of 0.075, RMSE of 0.083 and MAPE of 3.292%. Based on this model, the TFR in Malaysia is projected to decline and will slowly level off and is expected to be approximately 1.2 (average number of children per women) for the year 2040. A 95% confidence level for year the 2040 is in the range of 0.5 to 1.9 children per women.

Due to public health awareness, therapeutic medicine, etc., the mortality rate has decreased and the life expectancy has increased. Therefore, if the TFR keeps decreasing as projected in our ARAR model, eventually Malaysia will have population ageing problem.

5. REFERENCES

1. Brockwell, P.J. and Davis, R.A. (2002) *Introduction to Time Series Forecasting* (2nd Edition), New York: Springer-Verlag.
2. Brown. R.L. (2004). Baby Boom and Baby Bust-Fertility Rates and Why They Vary, *Contingencies January/February*, 17-20.
3. Caldwell, J.C. and Caldwell, B.K. (2005). The Causes of the Asian Fertility Decline, *Asian Population Studies*, 1(1), 31-46.
4. Carter, L.R. (1996). Forecasting U.S. Mortality: A Comparison of Box-Jenkins ARIMA and Structural Time Series Models. *The Sociological Quarterly*, 37(1), 127-140.
5. Carter, L.R. and Lee, R.D. (1986). Joint Forecasts of U.S. Marital Fertility, Nuptiality, Births, and Marriages Using Time Series Models, *Journal of the American Statistical Association*, 81(396), 902-911.
6. Carter, R.A.L. and Zellner, A. (2003). The ARAR Error Model for Univariate Time Series and Distributed Lag Models, *Studies in Nonlinear Dynamics and Econometrics*, 8(1), 1-42.
7. Davis, K. (1963). The Theory of Change and Response in Modern Demographic History. *Population Index*, 29(4), 345-356.
8. Department of Statistics Malaysia (2010). *http://www.statistics.gov.my/portal/download_Labour/files/BPTMS/ringkasan_perangkaan_2010.pdf.* Accessed August 20, 2013.
9. Department of Statistics Malaysia (2013). *http://www.statistics.gov.my/portal/download_Economics/files/DATA_SERIES/2011/pdf/21Perangkaan_Penduduk.pdf.* Accessed July 22, 2013.
10. Department of Statistics Malaysia, *Vital Statistics, Malaysia* (2010). *http://www.statistics.gov.my/portal/images/stories/files/LatestReleases/vital/Vital_Statistics_Malaysia_2010.pdf.* Accessed March 28, 2014.
11. Department of Statistics Malaysia, *Vital Statistics*, *Malaysia* (2011). *http://www.statistics.gov.my/portal/images/stories/files/LatestReleases/vital/Vital_Statistics_Malaysia_2011.pdf.* Accessed March 28, 2014.
12. Department of Statistics Malaysia, *Vital Statistics, Malaysia* (2012). *http://www.statistics.gov.my/portal/images/stories/files/LatestReleases/vital/Vital_Statistics_Malaysia_2012.pdf.* Accessed March 28, 2014.
13. Hirschman, C. (1994). Why Fertility Changes, *Annual Review of Sociology*, 20, 203-233.
14. Knudsen, C., McNown, R. and Rogeres, A. (1993). Forecasting Fertility: An Application of Time Series Methods to Parameterized Model Schedules, *Social Science Research*, 22, 1-23.
15. Land, K.C. and Cantor, D. (1983). ARIMA Models of Seasonal Variation in U.S. Birth and Death Rates, *Demography*, 20(4), 541-568.
16. Leete, R. and Kwok, K.K. (1986). Demographic Changes in East Malaysia and Their Relationship with Those in Peninsula 1960-1980, *Population Studies*, 4(1), 63-100.
17. Lim, L.L., Jones, G.W. and Hirschman, C. (1987). Continuing Fertility Transitions in Plural Sociecy: Ethnic Trends and Differentials in Peninsular Malaysia, *Journal of Biosocial Science*, 190, 405-425.

18. McDonald, J. (1979). A Time Series Approach to Forecasting Australian Total Live-Births, *Demography*, 16(4), 575-601.
19. McKenna, F. (1996). *New Brunswick at the Dawn of A New Century*. Canada: Province of New Brunswick.
20. McNown, R. and Rogers, A. (1989). Forecasting Mortality: A Parameterized Time Series Approach, *Demography*, 26(4), 645-660.
21. Norville, C., Gomez, R. and Brown, R.L. (2003). *Some Causes of Fertility Movements*, IIPR Insurance and Pension Reports No. 2003-02, Institute of Insurance and Pension Research, University of Waterloo.
22. Saboia, J.L.M. (1977). Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting, *Journal of the American Statistical Association*, 72(358), 264-270.
23. Soon, L.Y. (1992). Determinants of Fertility in Malaysia: How Much Do We Know? *Journal of Southeast Asian Studies*, 23(1), 112-132.
24. Tey, N.P. (2002). Social, Economic and Ethnic Fertility Differentials in Peninsular Malaysia, Paper presented at the *IUSSP Conference on Southeast Asia's Population in a Changing Asian Context*, held at the Siam City Hotel, Bangkok, Thailand. 10-13 June 2002.
25. United Nations Secretariat. Fertility Levels and Trends in Countries with Intermediate Levels of Fertility. http://www.un.org/esa/population/publications/ completing/ RevisedFFPSPOPDIVpaper.PDF. Accessed September 29, 2013.