<u>IOE 565</u>
<u>Birth Rates of Singapore</u>

Cleaven Yu
Chun Ming Wong
Foo Kean Lian

Content Page

Part I: Introduction

Time Series Analysis permits the user to better understand the analysis of physical systems. For this project we were given the opportunity to apply our knowledge and techniques learned during this course to proceed with the modeling and forecasting of a physical system.

We were enthralled with the possibility of creating an adequate model, thereby proceeding to forecast the Birth Rates of our country (Singapore). Being able to give a good estimation of the future Birth Rates is fundamental for a small country like Singapore. With a set of accurate future predictions, Government Officials can effectively plan for the capacity of schools or hospitals in order to cater for a future boom in Birth Rates. The ability to foresee or plan ahead cannot be underestimated for a country such as Singapore, with its small size and limited natural resources.

Our project is centered upon two sections, one is the Stationary Series aspect of the Birth Rates and, the other is the Non-Stationary Series aspect of Birth Rates. Finally we analyzed both models and gave a better insight into the capabilities and limitations of the models derived.

We proceed by giving a detailed interpretation of the data collected. In addition, we discuss our initial predictions of the Model to be derived.

Part II: Analysis of Stationary Series

2.1 Data Collection

Series Title: Expected Yearly Birthrates for Singapore (Population Statistics Section, Singapore Department of Statistics)

Data Source: SINGSTATInfo/SINGSTAT/SINGOV@SINGOV
"Report on Registration of Births and Deaths" (ISSN 0217-278X)
Registry of Births & Deaths, Singapore, Immigration and Registration.

Time Span: 1931 --- 2001 (Total of 71 years)
Modeling: 1945 -- 1990 (Total of 46 years)
Forecasting: 1991 -- 2001 (Total 11 years)

Sampling Interval: Yearly

Units: Birth Rates per 1,000 of Singapore population

Original data plot:

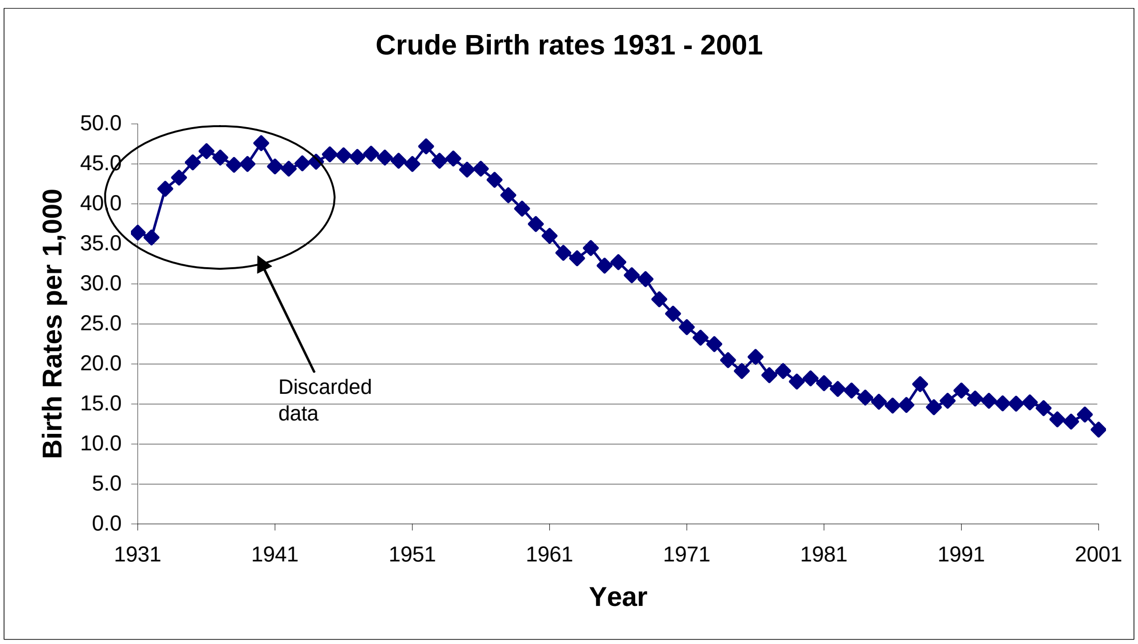


Figure 1: Crude Birth Rates (Discarded Data)

2.1.1 Discarded Data

From the analysis of our data, we decided to discard the following years: 1931 to 1944. During the period from 1931 to 1942, Singapore saw a large influx of immigrants from the Asia Pacific Region. Given the short period of time and the initial uncertainties experienced, the new settlers would not have settled in well. Therefore any prediction of Birth Rates would be inaccurate due to the constant transition of immigrants from different places.

During the period from 1942 to 1945, Singapore was occupied by Japan, during World War II. The atrocities provoked by the Japanese might have suppressed the community from starting a family. We decided that the data to be analyzed for our Model had to be one where external factors affecting Birth Rates were removed; therefore we proceeded to discard this set of data.

From the removal of the above-mentioned data sets, we believed that a good series of data was on hand to characterize the Birth Rates of a nation.

2.1.2 Analysis of Data

The ethic distribution of Singapore since 1930's to the 1980's has a mean of 76.5% Chinese, 14% Malay, and 8% Indian, the remainder composes of other minorities groups, etc Eurasians which make up 1.5% of the total population. As the population is predominantly Chinese, the total population growth is greatly influenced by the Birth Rates of the Chinese race.

From our collected data, we observed a peak in the Birth Rates occurring every 12 years, presiding during the years of 1952, 1964, 1978 and 2000. This repeated trend occurred during the year of the 'Dragon' with reference to the Chinese Horoscope.

Astrology has a long history in China and the Chinese culture, and is integrated with religious beliefs. As in the western astrological system, there are 12 zodiac signs; an auspicious animal or symbol refers to each zodiac sign. However, unlike the western system, your sign is based on the year rather than the month in which you are born. Being born or married in a particular year is believed to determine one's fortune. In this era of modern birth-control techniques, Chinese parents will often manipulate the birth dates of their children. The year of the 'Dragon', as observed from our collected data sees the biggest jump in Birth Rate as compared to any other zodiac year.

In Singapore the increased levels of education have not necessarily wiped out old beliefs in this Chinese Horoscope system. To date Singapore officials still expect a 15% to 20 % increase in births during the lunar year of the 'Dragon'.

Apart from the seasonality observed every 12 years, there is a falling trend that can also be observed from our collected data. This can be widely attributed to birth control policies and techniques introduced in Singapore over the past 30 – 50 years. An example was the introduction of the “Two-Child Families for Singapore” policy during the 1970’s, in order to further control population growth in Singapore.

Birth rates in general have decreased, resulting in the present problem of a rapidly ageing population. In 1987, the government realized their folly and implemented a new population policy. The new "Have three or more if you can afford it" policy was subsequently introduced. Unfortunately, this policy was unsuccessful in curbing the downward spiraling birth rates.

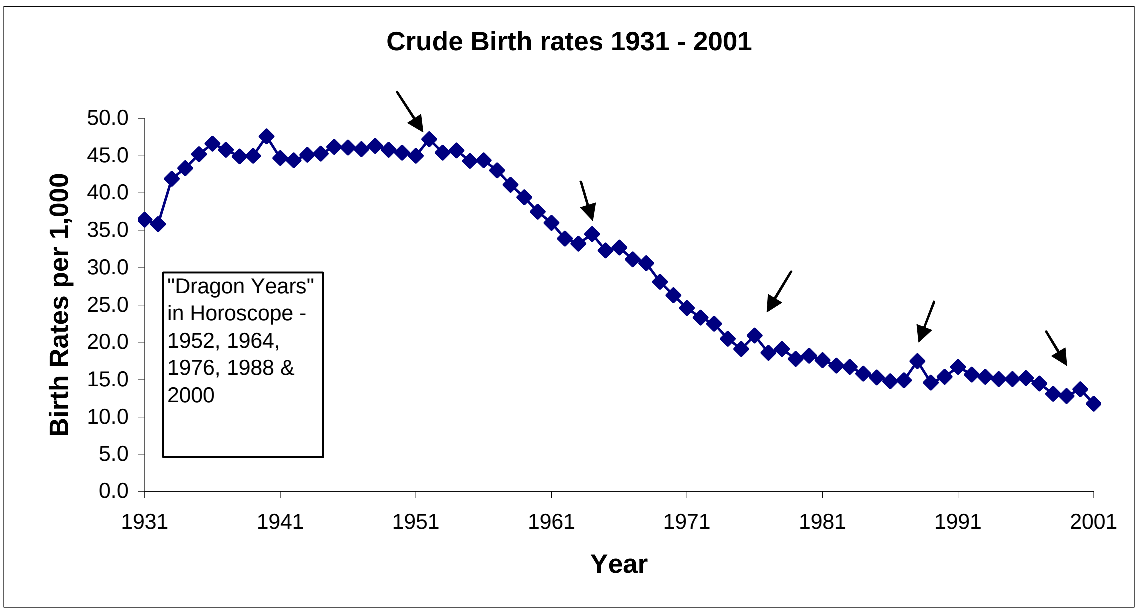


Figure 2: Crude Birth Rates (Dragon Years)

2.2 Modeling

We proceed to model our system by using the 'n' degrees of freedom vibration system as an interpretation of the ARMA models. The dependence in the series of data is approximated by a sequence of ARMA (2n, 2n-1) models. For this technique, increasing the degrees of freedom by one is equivalent to advancing the autoregressive order by two; this process is continued till an adequate model is obtained.

2.2.1 Modeling Process

An initial ARMA (2n, 2n-1) model, where n = 1, is fitted using the procedure outlined in Chapter 4 of the textbook. Every increase in the order of 'n' by steps of 1 is checked for adequacy through the F-criterion (equation 4.4.1 of text book). If the value of 'F' obtained this way is 5% of the significance level obtained from the F-distribution table, then the improvement in the Residual Sum of Squares moving from n = 1 to a n = 2 model is significant. A Matlab program, provided by Professor Dragan, was used in our calculations and derivations of an adequate ARMA model.

After a detailed analysis of the Matlab code provided, we felt that a slight alteration to the code was needed; to include the subtraction of the mean from the data values before the modeling technique was run. Although a variable was written in the code with mean subtracted, this variable was never used in the code to find the adequate model. To overcome this we inserted '**Data = iddata(ts)**' in line 25, of **PostulateARMA.m**.

The following model was obtained for our data set:

ARMA (8, 7)

Auto Regressive Part: $\mathbf{X_t + 0.5349\, X_{t-1} - 0.7783\, X_{t-2} - 0.5472\, X_{t-3} - 0.5136\, X_{t-4} - 0.4945\, X_{t-5} + 0.7131\, X_{t-6} + 0.4862\, X_{t-7} - 0.4169 X_{t-8}}$

Moving Average Part: $\mathbf{a_t + 1.702\, a_{t-1} + 1.21\, a_{t-2} + 1.002\, a_{t-3} + 0.5913\, a_{t-4} + 0.4369\, a_{t-5} + 1.046\, a_{t-6} + 0.7014\, a_{t-7}}$

Residual Sum of Squares for ARMA (8, 7) model = 18.66411

The Confidence Interval for ϕ_8 and θ_7 of this ARMA (8, 7) model is 0.4169 (± 0.4015) and 0.7014 (± 0.3859) respectively. Both Confidence Intervals do not include zero, therefore we cannot cancel the roots to simplify the model to an ARMA (7, 7) or an ARMA (8, 6) model. We conclude that the ARMA (8, 7) model is adequate. Please refer to Table 1 (Below), for the Parameters, Confidence Intervals and Residual Sum of Squares for this ARMA (8, 7) model, together with the other ARMA models used as a comparison to derive the adequate ARMA model.

Table 1: Yearly Birth Rate Data

Parameter	Order of the ARMA model			
	(4, 3)	(6, 5)	(8, 7)	(10, 9)
ϕ_1	0.588 ± 0.5791	0.5415 ± 0.6717	-0.5349 ± 0.4725	-0.1572 ± 1.894
ϕ_2	1.129 ± 0.7513	0.5472 ± 0.5312	0.7783 ± 0.4477	1.061 ± 0.9961
ϕ_3	-0.08617 ± 0.4204	0.2459 ± 0.8534	0.5472 ± 0.3637	0.02623 ± 1.84
ϕ_4	-0.6522 ± 0.4532	-0.09025 ± 0.7359	0.5136 ± 0.4768	0.1365 ±1.14
ϕ_5		0.3493 ± 0.6477	0.4945 ± 0.4813	0.2355 ± 1.078
ϕ_6		-0.6317 ± 0.4739	-0.7131 ± 0.4468	-0.2048 ± 0.987
ϕ_7			-0.4862 ± 0.3243	0.2668 ± 1.228
ϕ_8			0.4169 ± 0.4015	0.5308 ± 1.075
ϕ_9				-0.3295 ±0.834
ϕ_{10}				-0.2829 ±0.887
θ_1	-0.1278 ± 0.6154	-0.168 ± 0.7417	-1.702 ± 0.5129	-1.012 ± 1.908
θ_2	0.5726 ± 0.4402	0.08088 ± 0.6465	-1.21 ± 0.9262	-0.1645 ± 2.036
θ_3	0.5819 ± 0.3811	0.3666 ± 0.5698	-1.002 ± 0.7974	-0.4963 ± 1.445
θ_4		-0.01932 ± 0.7321	-0.5913 ± 0.7191	-0.5156 ± 1.387
θ_5		0.7337 ± 0.562	-0.4369 ± 0.6772	-0.592 ± 0.7696
θ_6			-1.046 ± 0.635	-0.5012 ± 0.960
θ_7			-0.7014 ± 0.3859	-0.1224 ± 1.029
θ_8				-0.2444 ± 0.833
θ_9				-0.0552 ± 0.901
Resident Sum of Squares (RSS)	38.03724	26.028564	18.66411	29.76038

2.2.2 Model Analysis

For the ARMA (8, 7) model above, the roots are:

Auto Regressive Roots are as follows:

$\lambda_1 = - 0.9904$
$\lambda_2 = -0.8884 + 0.4866i$
$\lambda_3 = -0.8884 - 0.4866i$
$\lambda_4 = -0.0035 + 0.9827i$
$\lambda_5 = -0.0035 - 0.9827i$
$\lambda_6 = 1.0058$
$\lambda_7 = 0.6167 + 0.2049i$
$\lambda_8 = 0.6167 - 0.2049i$

Moving Average Roots are as follows:

$\nu_1 = 0.6496 + 0.5847i$
$\nu_2 = 0.6496 - 0.5847i$
$\nu_3 = -0.1202 + 1.0026i$
$\nu_4 = -0.1202 - 1.0026i$
$\nu_5 = -0.8838 + 0.3542i$
$\nu_6 = -0.8838 - 0.3542i$
$\nu_7 = -0.9934$

Seasonality: From the roots λ_2 and λ_3, we calculated a seasonality of Period = 12.5 years. This seasonality of Period = 12.5 years goes against our initial data analysis that the Birth Rates would have a seasonality of Period = 12 years. The calculate period does not coincide with the Chinese Horoscope Theory we initially predicted. In order to analyze this further and test the validity of our prediction, we proceed by forcing in the Period = 12 years into our ARMA (8, 7) model, by substituting λ_2 and λ_3 with the following roots: $0.866 \pm 0.500i$, by using the operator $1 - 1.732B + B^2$ (Table 9.1 in Text Book). This operator represents the Period = 12 years.

Trend: From the roots λ_6, we see a possibility of a constant trend. In order to analyze this further, we proceed by using the operator 1 - B (Table 9.1 in Text Book).

Using the combined operator $(1 – B) (1 - 1.732B + B^2)$ for the stochastic trend and seasonality, we can reduce the autoregressive parameters by 3 and fit an ARMA (5, 7) model. The newly derived parsimonious model with the trend and seasonality operators will be used to compare with the original ARMA (8, 7) model in an adequacy comparison. The results will be used to confirm the validity of our initial predictions of whether the seasonality of Period = 12 years holds for the Birth Rate data.

2.2.3 Parsimonious Model

From, $\mathbf{Y_t = (1 - B)(1 - 1.732B + B^2)* X_t}$

We have the following: $\mathbf{Y_i = X_i - 2.732\ X_{i-1} + 2.732\ X_{i-2} - X_{i-3}}$

Using an iterative process and setting the initial values of X_{-1}, X_{-2} and X_{-3} to the value zero, we are able to calculate the new set of Y_i values for i = 1 to 46 (data points), thereby permitting us to re-calculate the Parameters and the Residual Sum of Squares for the new ARMA (5, 7) model. The new set of calculated data values, from this iterative process that was used in the derivation of the Parsimonious model is included in Appendix B.

Please refer to Table 2 (Below), for the Parameters of the Parsimonious ARMA (5, 7) model. Parsimonious Model Residual Sum of Squares = 28.51

$F = 5.2753 > F\ (3, 30)$, where $F\ (3,30) = 2.9223$

From the F-criterion, when comparing the original ARMA (8, 7) model with the Parsimonious ARMA (5, 7) model, the improvement in Residual Sum of Squares was calculated to be significant. Again we conclude that the ARMA (8, 7) model is the adequate model for Birth Rates. This gave a clear indication that our initial prediction of Period = 12 years for Birth Rates was flawed.

2.2.4 Discussion

To better understand the context of the results above, we will discuss possible factors that contribute to the deviation between our initial prediction of Period = 12 years and the actual results of Period = 12.5 years for Birth Rates. To begin, Singapore is a multi-racial country; therefore it consists of other races that do not follow the Chinese horoscope. Over past 10 to 15 years the proportion of Malays and Indians has steadily increased, approaching approximately 28% of the total population. In addition, Singapore has been active in its immigration laws to lure in foreigners to enhance our workforce and productivity. The complexity of the different cultures adds to the difficulty in creating a Birth Rate model based solely on the Chinese Horoscope theory.

Although the Chinese population still consists of approximately 70% of the total population, modern day Chinese have become less traditional and are becoming more secular. There are now many different factors and conditions to consider before starting a family, or when deciding between the family's sizes (number of children). Ambitions to rise in the corporate ladder have taken up more time for the average Chinese family, therefore leaving considerably less time for personal responsibilities.

As mentioned in our data collection, we only used a selected portion of the data. 46 samples were used to create an appropriate model to predict the following 12 years. A better model would consist of around 200 samples. As we received our data in yearly intervals, this severely restricted our ability to obtain more data. Therefore using 46 samples can be a poor indication of the characteristics and trends of the population Birth Rates of a nation.

In addition, because Singapore has a status of a developed country, she is constantly faced with an increasing cost of living. The costs of public transportation to the cost of living expenses have made it increasingly more taxing to have a large family. With this under consideration, the trend of families in the 1950's and 1960's having 3 children, have progressed to the trend of families in the 1990's having 1 to 2 children. This affects any predicted seasonality of period 12 years, as families will weigh the cost of bringing up their children when deciding to have children, over the prosperity or beliefs of having children during the auspicious years in the Chinese Horoscope.

Table 2: Yearly Birth Rate Data

Parameter	Order of the ARMA model	
		$(1 - B)(1 - 1.732B + B^2)$
	(8, 7)	(5, 7)
ϕ_1	-0.5349 ± 0.4725	1.573 ± 2.086e-008
ϕ_2	0.7783 ± 0.4477	- 0.1445 ± 8.553e-009
ϕ_3	0.5472 ± 0.3637	- 0.5852 ± 6.071e-009
ϕ_4	0.5136 ± 0.4768	- 0.4221 ± 8.552e-009
ϕ_5	0.4945 ± 0.4813	0.5787 ± 2.086e-008
ϕ_6	-0.7131 ± 0.4468	
ϕ_7	-0.4862 ± 0.3243	
ϕ_8	0.4169 ± 0.4015	
θ_1	-1.702 ± 0.5129	1.027 ± 2.924e-010
θ_2	-1.21 ± 0.9262	0.5368 ± 2.796e-010
θ_3	-1.002 ± 0.7974	- 0.1406 ± 2.74e-010
θ_4	-0.5913 ± 0.7191	- 0.5922 ± 2.658e-010
θ_5	-0.4369 ± 0.6772	- 0.269 ± 2.588e-010
θ_6	-1.046 ± 0.635	0.1087 ± 2.511e-010
θ_7	-0.7014 ± 0.3859	0.3393 ± 2.374e-010
(RSS)	18.66411	28.51

2.3 Forecasting

2.3.1 Forecasting Data

The data collected is represented by an ARMA (8, 7) model. The parameters are:

$\phi_1 = -0.5349$

$\phi_5 = 0.4945$

$\phi_2 = 0.7783$

$\phi_6 = -0.7131$

$\phi_3 = 0.5472$

$\phi_7 = -0.4862$

$\phi_4 = 0.5136$

$\phi_8 = 0.4169$

$\theta_1 = -1.7020$

$\theta_5 = -0.4369$

$\theta_2 = -1.2100$

$\theta_6 = -1.046$

$\theta_3 = -1.0020$

$\theta_7 = -0.7014$

$\theta_4 = -0.5913$

$\sigma_a^2 = 0.6221$

With these parameters, the Matlab program (Predict Function) was used to obtain a forecast for points up to 15 steps ahead of the original data with origin at t = 1990. The forecasted data is shown in Table 3 (Below) and a plot of this data including the 95% Confidence Intervals (CI) for each of the forecasted is represented in Figure 3 (Below).

The forecast data was calculated using the origin at t = 1990 so that the forecast data may be compared and verified against the original data from 1991–2001. Forecast data for 2002-2005 was also obtained as an estimate of the birth rates the Singapore Government may expect in the coming years.

A 95% Confidence Interval for the forecast data was included to obtain an interval with a 95% probability that the actual birth rates will fall within the interval.

Table 3: Forecast at Different Lags

l	G(l)	t=1990	Year	Xt	Xt(l)	$1.96\sqrt{Var[e_t(l)]}$	Upper 95% CI	Lower 95% CI
		t-3	1987	14.9				
		t-2	1988	17.5				
		t-1	1989	14.6				
0	1	t	1990	15.4				
1	1.1671	t+1	1991	16.7	14.6	1.545962968	16.14596	13.05404
2	1.364	t+2	1992	15.7	13.6	2.376021065	15.97602	11.22398
3	1.728	t+3	1993	15.4	13.3	3.176800959	16.4768	10.1232
4	1.881	t+4	1994	15.1	13.15	4.150731329	17.30073	8.999269
5	2.616	t+5	1995	15.1	13.15	5.068015456	18.21802	8.081985
6	2.621	t+6	1996	15.2	13.28	6.483876214	19.76388	6.796124
7	2.608	t+7	1997	14.5	13.08	7.645855286	20.72586	5.434145
8	2.774	t+8	1998	13.1	13.12	8.643789112	21.76379	4.476211
9	2.845087	t+9	1999	12.8	13.14	9.649162315	22.78916	3.490838
10	3.091179	t+10	2000	13.7	13.2	10.60435076	23.80435	2.595649
11	2.654737	t+11	2001	11.8	13.23	11.63140773	24.86141	1.598592
12	2.900314	t+12	2002		13.09	12.33424072	25.42424	0.755759
13	2.995792	t+13	2003		13.3	13.12393865	26.42394	0
14	2.948609	t+14	2004		13.22	13.91716513	27.13717	0
15	2.943252	t+15	2005		13.18	14.64468714	27.82469	0

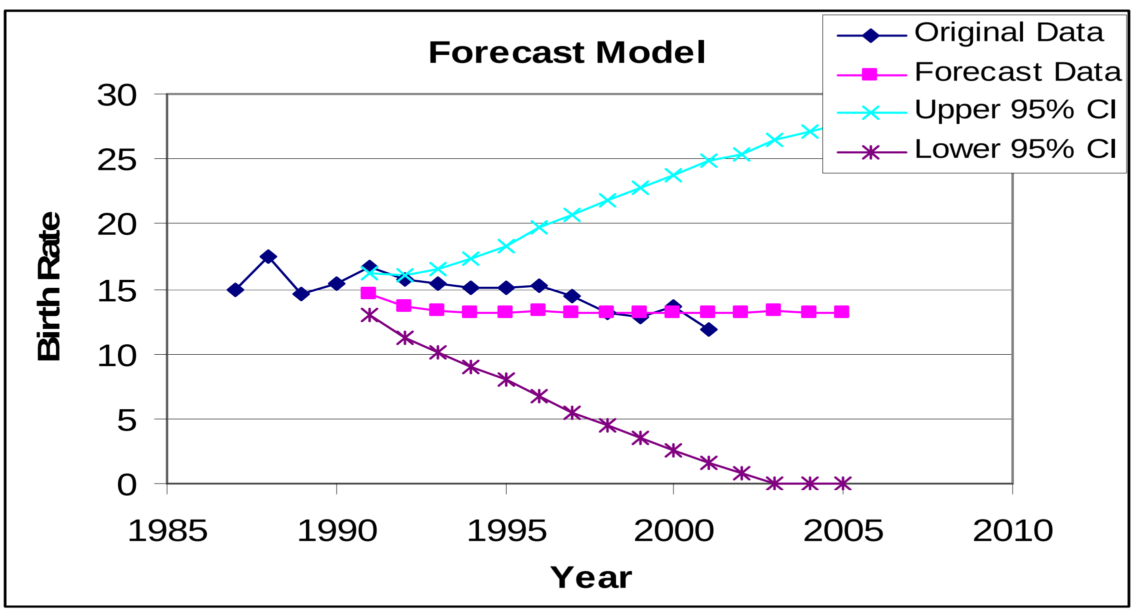


Figure 3: Plot of Forecast Data (Birth Rate per 1000)

2.3.2 Analysis of Forecast Data

It was observed that the 95% Confidence Interval widened as we forecasted data further into the future. This is acceptable, as the estimates will always become more uncertain when it is forecasted further into the future. Furthermore, the probability limits were based on the Green's Function, G, which characterizes the memory of the dynamics of the systems. As the data is based on an ARMA (8, 7) model, the system has a strong memory and hence, the limits will be wider.

The forecast data from 1991-2001 was found to be relatively close to the actual birth rates recorded during these years. The biggest discrepancy was found to be only 2.1 and it occurred in the first 3 years (1991-1993). This slight inaccuracy could be due to the small sample size that we used for our forecast. Overall, as the exact forecast data calculated is close to the actual birth rates, it suggests that the forecast model is a good estimate for the birth rates in Singapore.

Apart from 1991, the forecast data seems to follow a constant trend since it falls within the range of 13-14. This is consistent with the ARMA (8, 7) model obtained earlier, as it contained a constant trend (with one of the roots very close to 1 in value). Since the forecast data was projected up to the year 2005, it will be fair to say that the birth rates in Singapore during the coming decade will fall within this range unless something drastic happens.

Part III: Analysis of Non-Stationary Series

3.1 Alternate Model

The alternative method of modeling is obtained by decomposing a model into two parts, deterministic and stochastic. Non-stationarity can be removed by removing the deterministic trend in the data, thereby permitting the transformed stationary series to be modeled using the above mentioned ARMA modeling techniques.

3.2 Analysis of Deterministic and Stochastic Parts

3.2.1 Deterministic Part

The collected data from 1945 to 1990 is observed to have a strong linear trend. Along that linear trend, the data points seem to be randomly scattered (Scattered along a straight line). As such it is a good estimate to use linear regression to model the deterministic part of the data, i.e. $\mathbf{Y_t = \beta_0 + \beta_1 t + \varepsilon_t}$ (Section 2.1.1 of Textbook).

We used the statistical software 'R', a feely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, from the Carnegie Mellon University site. By entering the appropriate coding, given in Appendix A, we obtained a linear regression, Birth Rates = 1745.74255 - 0.87191 * Year, where Year has a range of between 1945 and 1990 (inclusive). Therefore, the deterministic portion of the model can be expressed in the following form:

$\mathbf{\beta_0 + \beta_1 t = 49.8776 - 0.87191t}$**, where** $\mathbf{0 \leq t \leq 46}$

From the statistical software, R^2 has a range of: $0 \leq R^2 \leq 1$ (R^2 represents Correlation) When the calculated R^2 has the value of 1, the model can be described as being able to accurately explain all data points. In our obtained model, the value of $R^2 = 0.9578$. This is a suggestion that the linear model derived will be able to explain the data points accordingly.

3.2.2 Residuals

Successive data points are shown in Figure 4 (Below), Plot of Residuals versus Fitted data. As can be observed, the residuals, ε_t, have a tendency to decrease once they start decreasing and similarly, increase once they start increasing. In addition, most of the residuals are considerably small when compared to the corresponding observations (Collected Data). The majority of the residuals are approximately between 5 to 8% of the corresponding observations as observed in Figure 5 (Below). Included in Appendix A is a statistical QQ plot that describes the model following the 'Normal' assumption.

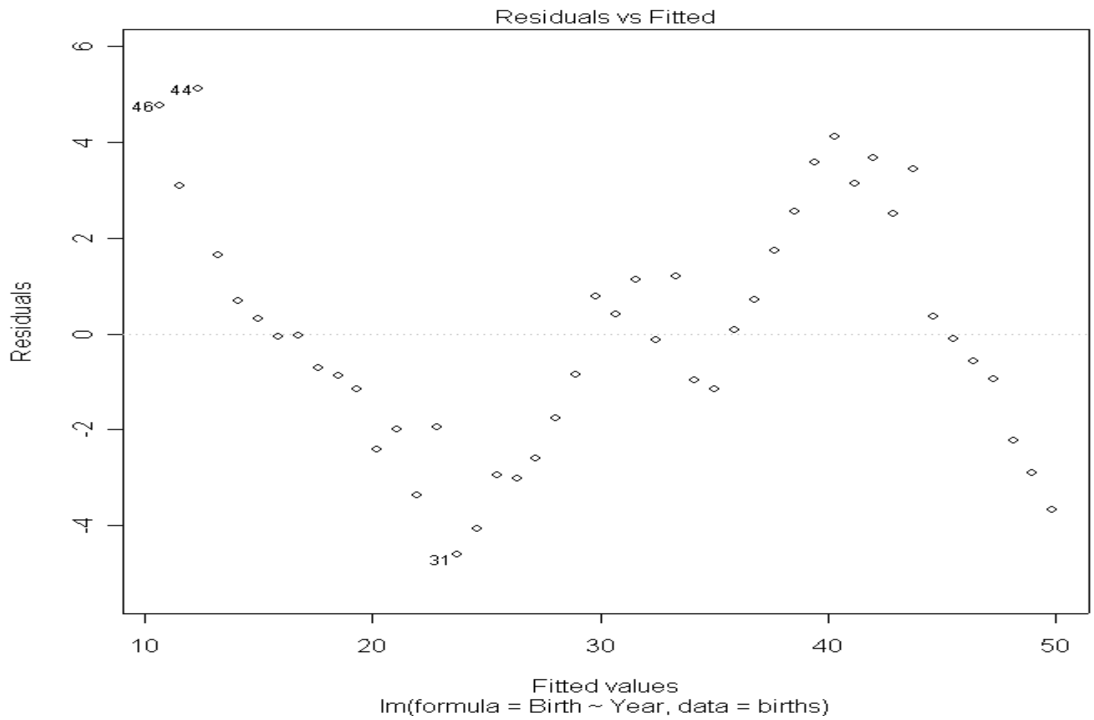


Figure 4: Residuals versus Fitted

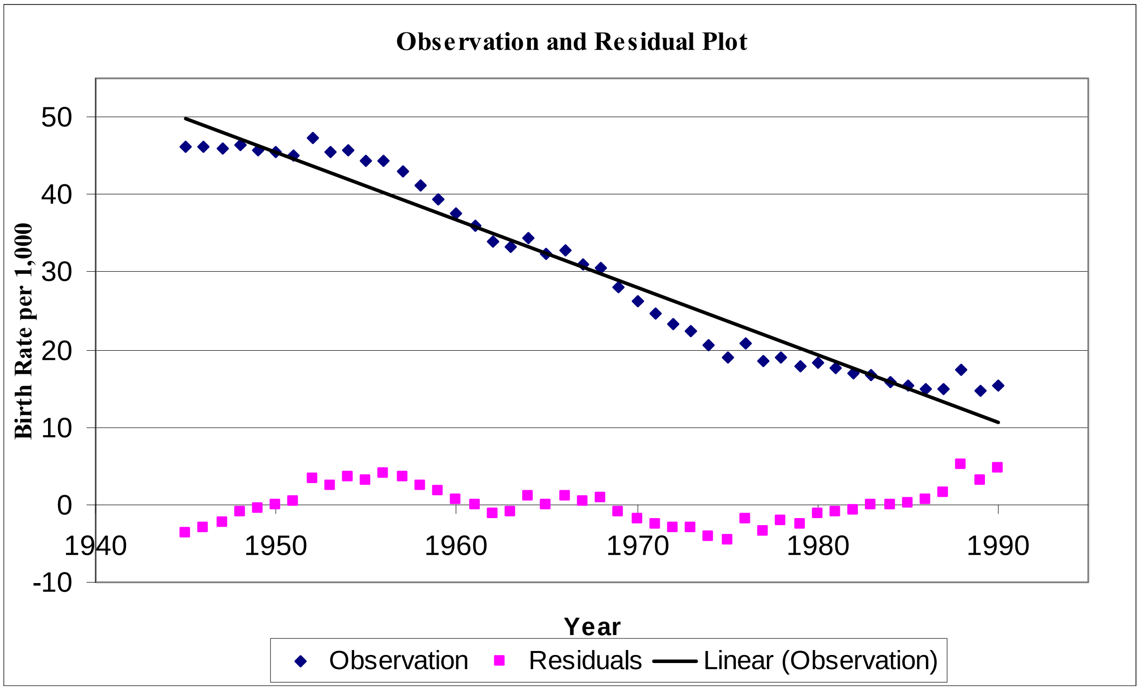


Figure 5: Observation and Residuals versus Time

3.2.3 Stochastic Part

The residuals can be modeled by the above-mentioned method addressed in Section 2.2.1 of the report, to derive the stationary part of our model. The complete model follows the form $\mathbf{Y_t = \beta_0 + \beta_1 t + X_t}$, where X_t represents the stationary part of the model.

By subtracting the deterministic part from the collected set of data, i.e. $X_t = Y_t - (\beta_0 + \beta_1 t)$, the stationary series can be used to find an adequate ARMA model. Appendix C contains the residuals used to obtain the stationary part of the model.

Through Matlab, the Stochastic function takes the form: $\mathbf{X_t = 0.8687X_{t-1} + a_t}$

Please refer to Table 4 (Below), for the Parameters, Confidence Intervals and Residual Sum of Squares for this ARMA (1) model, together with the other ARMA models used as a comparison to derive the adequate ARMA model.

3.2.4 Combined Model

The combined model of deterministic part and stochastic part takes the form:

$$\mathbf{Y_t = 49.8776 - 0.87191t + 0.8687X_{t-1} + a_t}$$

Table 4: Stochastic Part (ARMA Model)

Parameter	Order of the ARMA model		
	ARMA (4, 3)	AR (1)	ARMA (2, 1)
ϕ_1	0.4155 ± 0.4046	0.8687 ± 0.08636	-0.1354 ± 0.09452
ϕ_2	1.185 ± 0.4703		0.9137 ± 0.09105
ϕ_3	0.009151 ± 0.3615		
ϕ_4	-0.6979 ± 0.3473		
θ_1	-0.2868 ± 0.4137		0.9958 ± 0.08396
θ_2	0.616 ± 0.2882		
θ_3	0.7004 ± 0.2942		
RSS	50.16456	77.49764	63.75432

3.3 Forecasting

3.3.1 Forecasting Data

A forecast for the non-stationary model was done based on the stochastic part of it which was found to follow an AR (1) model. Using the same methods to forecast the data as mentioned earlier in the report for the ARMA (8,7) model, the Matlab program (Predict Function) was also used to obtain a forecast 15 years ahead starting form an origin at t = 1990. The forecasted data is shown in Table 5 (Below). Figure 6 (Below) represents a plot of this data including the 95% Confidence Intervals (CI) for each forecasted point.

Forecasted data was done for the years 1991 – 2005 using the origin at t = 1990 to allow comparisons against the original data for the years 1991-2001, as well as to compare with the forecasted data obtained using the ARMA (8,7) model as shown in the earlier part of the report. Similarly, forecast data for 2002-2005 obtained from this non-stationary model may be used as an estimate for the birth rates the Singapore Government may expect in the coming years.

A 95% Confidence Interval for the forecast data was included to obtain an interval with a 95% probability that the actual birth rates will fall within the interval.

Table 5: Forecast at Different Lags

l	G(l)	t=1990	Year	Xt	Xt(l)	$1.96\sqrt{Var[e_t(l)]}$	Upper 95% CI	Lower 95% CI
		t-3	1987	14.9				
		t-2	1988	17.5				
		t-1	1989	14.6				
0	1	t=1990	1990	15.4				
1	0.8687	t+1	1991	16.7	15.4	2.601201356	18.0012	12.7988
2	0.75464	t+2	1992	15.7	15.21	3.445624495	18.65562	11.76438
3	0.429753	t+3	1993	15.4	15.33	3.965548958	19.29555	11.36445
4	0.03411	t+4	1994	15.1	15.24	4.120099645	19.3601	11.1199
5	4.62E-08	t+5	1995	15.1	15.35	4.121054882	19.47105	11.22895
6	9.69E-45	t+6	1996	15.2	15.22	4.121054882	19.34105	11.09895
7	0	t+7	1997	14.5	15.43	4.121054882	19.55105	11.30895
8	0	t+8	1998	13.1	15.37	4.121054882	19.49105	11.24895
9	0	t+9	1999	12.8	15.28	4.121054882	19.40105	11.15895
10	0	t+10	2000	13.7	15.23	4.121054882	19.35105	11.10895
11	0	t+11	2001	11.8	15.27	4.121054882	19.39105	11.14895
12	0	t+12	2002		15.32	4.121054882	19.44105	11.19895
13	0	t+13	2003		15.31	4.121054882	19.43105	11.18895
14	0	t+14	2004		15.32	4.121054882	19.44105	11.19895
15	0	t+15	2005		15.24	4.121054882	19.36105	11.11895

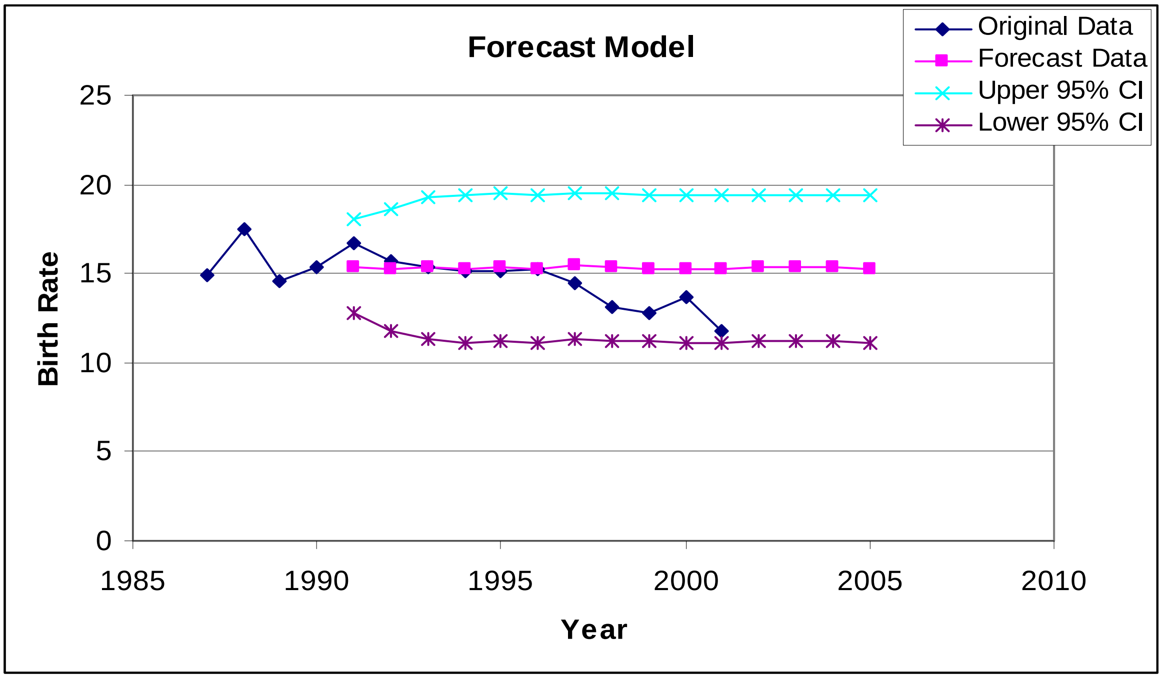


Figure 6: Plot of Forecast Data (Birth Rate per 1000)

3.3.2 Analysis of Forecast Data

Just as it was observed in the ARMA (8, 7) model, the 95% Confidence Interval for this forecast also widened as we forecasted data further into the future (Refer to Part IV for the comparison between both models). Hence, the reasons for an increasing interval are similar to the reasons given earlier. Estimates will always become more uncertain when it is forecasted further into the future. Also, the probability limits were based on the Green's Function, G, which characterizes the memory of the dynamics of the systems.

The forecasted data show that the birth rates remain within the range of 15.21 - 15.4. It is intuitively impossible for the birth rates to remain within such a small range. Furthermore, a comparison against the original data for the years 1991-2001 shows a huge discrepancy especially in the year 2001 where the data differed by 3.8. Hence, we conclude that the forecast data based on this model is considered inaccurate. This inaccuracy could be due to a number of factors.

Firstly, the sample sized used for this forecast is too small. While a sample size of 200 is generally required to obtain a good forecast, a sample size of only 46 was used for this analysis. This could have led to the inaccuracy of the forecast.

Furthermore, the decision to base the deterministic part of the non-stationary model on a linear trend was primarily based on observation. It is highly possible a linear trend was not suitable. Instead, the deterministic part may be based on an exponential or polynomial trend.

Part IV: Conclusion and Analysis

The first notable difference between the forecast model obtained by the ARMA (8, 7) model and the non-stationary model is that the latter model shows a smaller increase in the confidence intervals. This is because the deterministic part of the non-stationary model can be predicted without error so that its mean squared error for long-term forecasts is limited to the error of its stochastic part. Since the stochastic part of the model is stable (with mean equal to zero) and the forecasted data for the non-stationary model is based primarily on the stochastic part, its variance and confidence interval is finite. Hence, the confidence interval for the forecast based on the non-stationary model levels off, where as the ARMA (8, 7) model shows a constantly increasing variance and confidence interval.

In addition, the forecast for the years 1991 - 2001given by the ARMA (8, 7) model gives figures closer to the actual birth rates witnessed in Singapore during these years. This suggests that the ARMA (8, 7) model will also give a better forecast of the birth rates in the coming years. It is believed that the non-stationary model did not give an accurate forecast as the deterministic part of the model was wrongly chosen to be a linear trend. Had it been chosen correctly, the non-stationary model may have given a more accurate forecast. In fact, an F-criterion test comparing the RSS values for the ARMA (8, 7) model against the non-stationary model showed that there was a significant difference.

$$F = [\ (RSS_1 - RSS_2)/13\ ] \ /\ [\ RSS_2/(46-16)\ ]$$
$$= 7.274 > F_{0.95,13,30} \qquad \text{,where } F_{0.95,13,30} = 2.063$$

In conclusion, non-stationary models based on polynomial trends should be considered and tested using the same tests and procedure as was done with the analysis based on a linear trend. It may have resulted in a more accurate model and forecast for the birth rates for Singapore.

Appendix A

"R" Software Coding and Output

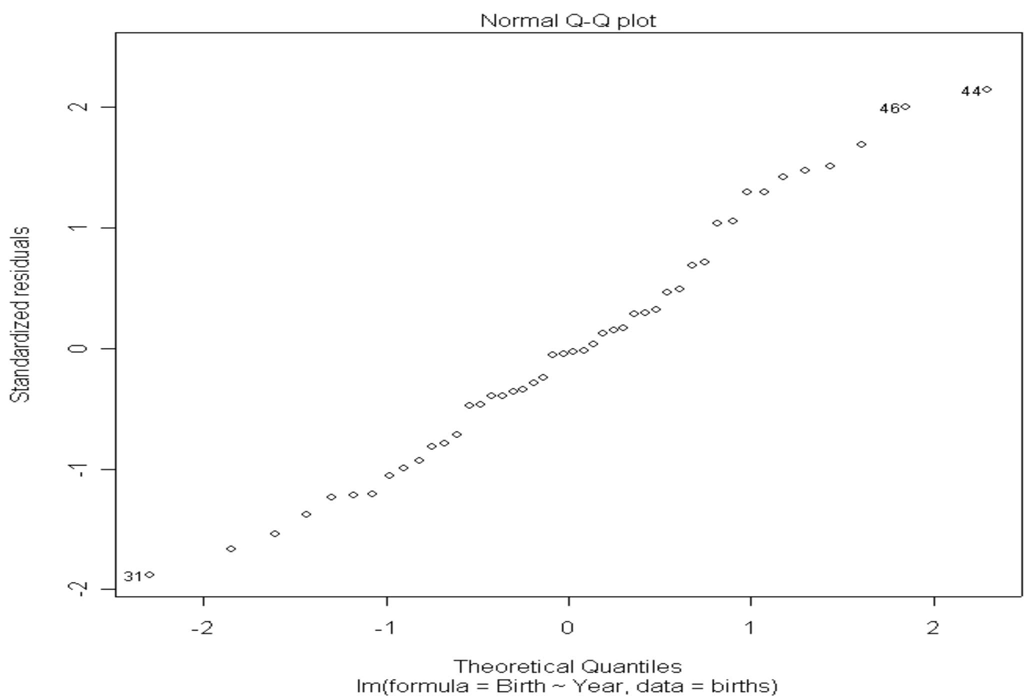
```
> births <- read.table("Birth.txt", header=T)
> Model <- lm(Birth ~ Year, births)
> summary(Model)

Call:
lm(formula = Birth ~ Year, data = births)

Residuals:
     Min       1Q   Median       3Q      Max
-4.61064 -1.89660 -0.08596  1.53883  5.12426

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1745.74255   54.27049   32.17   <2e-16 ***
Year          -0.87191    0.02758  -31.61   <2e-16 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Residual standard error: 2.484 on 44 degrees of freedom
Multiple R-Squared: 0.9578,     Adjusted R-squared: 0.9569
F-statistic: 999.2 on 1 and 44 DF,  p-value: < 2.2e-16
```



From QQ-plot we see our model follows the "Normal" assumption.

Appendix B

Data used to derive ARMA (5, 7) Model

Data set calculated from the iterative process described in **Section 2.2.3** of report.

1945	46.2	1987	298.3065
1946	126.2184	1988	218.1898
1947	218.6103	1989	125.7981
1948	298.6146	1990	45.8926
1949	344.7902		
1950	344.762		
1951	298.5376		
1952	218.5052		
1953	126.1133		
1954	46.12308		
1955	-0.02813		
1956	0.028196		
1957	46.27697		
1958	126.3235		
1959	218.7154		
1960	298.6915		
1961	344.8183		
1962	344.7338		
1963	298.4607		
1964	218.4		
1965	126.0082		
1966	46.04621		
1967	-0.0562		
1968	0.056456		
1969	46.35398		
1970	126.4286		
1971	218.8204		
1972	298.7683		
1973	344.8463		
1974	344.7055		
1975	298.3836		
1976	218.2949		
1977	125.9032		
1978	45.96938		
1979	-0.0842		
1980	0.08478		
1981	46.43104		
1982	126.5338		
1983	218.9255		
1984	298.8451		
1985	344.8743		
1986	344.6772		

Appendix C **Calculation of Residual Values**

$y_t = \beta_0 + \beta_1 t$	Birth Rates = 1745.74255 - 0.87191 * Year					
Year	Birth	Deterministic Birth rates	Residual	Residual^2	% diff	Absolute
1945	46.2	49.8776	-3.6776	13.52474	8%	8%
1946	46.1	49.00569	-2.90569	8.443034	6%	6%
1947	45.9	48.13378	-2.23378	4.989773	5%	5%
1948	46.3	47.26187	-0.96187	0.925194	2%	2%
1949	45.8	46.38996	-0.58996	0.348053	1%	1%
1950	45.4	45.51805	-0.11805	0.013936	0%	0%
1951	45	44.64614	0.35386	0.125217	-1%	1%
1952	47.2	43.77423	3.42577	11.7359	-7%	7%
1953	45.4	42.90232	2.49768	6.238405	-6%	6%
1954	45.7	42.03041	3.66959	13.46589	-8%	8%
1955	44.3	41.1585	3.1415	9.869022	-7%	7%
1956	44.4	40.28659	4.11341	16.92014	-9%	9%
1957	43	39.41468	3.58532	12.85452	-8%	8%
1958	41.1	38.54277	2.55723	6.539425	-6%	6%
1959	39.4	37.67086	1.72914	2.989925	-4%	4%
1960	37.5	36.79895	0.70105	0.491471	-2%	2%
1961	36	35.92704	0.07296	0.005323	0%	0%
1962	33.9	35.05513	-1.15513	1.334325	3%	3%
1963	33.2	34.18322	-0.98322	0.966722	3%	3%
1964	34.5	33.31131	1.18869	1.412984	-3%	3%
1965	32.3	32.4394	-0.1394	0.019432	0%	0%
1966	32.7	31.56749	1.13251	1.282579	-3%	3%
1967	31.1	30.69558	0.40442	0.163556	-1%	1%
1968	30.6	29.82367	0.77633	0.602688	-3%	3%
1969	28.1	28.95176	-0.85176	0.725495	3%	3%
1970	26.3	28.07985	-1.77985	3.167866	7%	7%
1971	24.6	27.20794	-2.60794	6.801351	11%	11%
1972	23.3	26.33603	-3.03603	9.217478	13%	13%
1973	22.5	25.46412	-2.96412	8.786007	13%	13%
1974	20.5	24.59221	-4.09221	16.74618	20%	20%
1975	19.1	23.7203	-4.6203	21.34717	24%	24%
1976	20.9	22.84839	-1.94839	3.796224	9%	9%
1977	18.6	21.97648	-3.37648	11.40062	18%	18%
1978	19.1	21.10457	-2.00457	4.018301	10%	10%
1979	17.8	20.23266	-2.43266	5.917835	14%	14%
1980	18.2	19.36075	-1.16075	1.347341	6%	6%
1981	17.6	18.48884	-0.88884	0.790037	5%	5%
1982	16.9	17.61693	-0.71693	0.513989	4%	4%
1983	16.7	16.74502	-0.04502	0.002027	0%	0%
1984	15.8	15.87311	-0.07311	0.005345	0%	0%
1985	15.3	15.0012	0.2988	0.089281	-2%	2%
1986	14.8	14.12929	0.67071	0.449852	-5%	5%
1987	14.9	13.25738	1.64262	2.6982	-11%	11%
1988	17.5	12.38547	5.11453	26.15842	-29%	29%
1989	14.6	11.51356	3.08644	9.526112	-21%	21%
1990	15.4	10.64165	4.75835	22.64189	-31%	31%
			Rss =	271.4093		8%

References

Data Analysis:
http://www.singstat.gov.sg/index.html
http://www.hongkonghotelsearch.com/directories/hongkong/chinazodiac.html
http://china.tyfo.com/int/cdnews/national/20000201n-2.htm

Falling Birth Rates:
http://www.thecore.nus.edu.sg/sea/students/bbonus/p2.html

Statistical Software, Carnegie Mellon University:
http://lib.stat.cmu.edu/R/CRAN/