Parabola Volume 43, Issue 3 (2007)

Modelling Australia's Ageing Population

Bruce Henry[1]

One of the most difficult problems to be faced by developed countries throughout the world over the next fifty years is the ageing population problem. This problem is essentially a result of the 'baby boom generation' (people born in the period 1946 to 1962) having lower fertility rates but longer life expectancies compared with their predecessors. Over the next few decades the baby boom generation will move out of prime employment age into retirement, leaving fewer workers to provide for an increased aged population. As a measure of this effect, in Australia, the number of persons over the age of sixty as a percentage of working-age adults is expected to more than double from 26% to 56% over the next forty years [1]. If Australia does not take action now, it has been estimated that Government spending per annum will exceed revenues by about 5% of GDP by 2040. In today's terms this would translate into a $40 billion deficit [1].

The key considerations in Australia's ageing population can be revealed using a simple set of finite difference equations that can readily be investigated in a classroom project, both at the secondary school level and at the tertiary level. I have supervised year 11 secondary students in this project area as part of the CSIRO Student Research Support Scheme. I have also set mathematical modelling of Australia's ageing population as an assessment task for third year undergraduate students in a course called Mathematical Modelling for Real World Systems. Students enjoy the contemporary relevance of the project (the students of today will live through the major impacts of the problem) and they enjoy being able to obtain interesting and meaningful results from mathematical models.

In this article I will describe a basic model that can capture the essential features of the ageing population problem and can be investigated through very modest computer simulations using packages such as Excel, Maple or MATLAB. The model can be used to evaluate different possible strategies for alleviating the economic burden of an ageing population, such as delaying the retirement age, encouraging families to have more children or boosting immigration.

Basic Model for the Total Population

To begin with it is useful to consider a very general model for Australia's total population growth. The mathematical model for the total population, which changes in time, is an example of a dynamical system. Numbers are added to the population through birth and immigration, and numbers are removed from the population

[1] Bruce Henry is an Associate Professor in the School of Mathematics and Statistics at UNSW and the Editor of *Parabola*

through death and emigration. For example, the current birth rate in Australia is approximately 12 births per 1000 people per year [2]. For a total population of approximately 20,000,000 this corresponds to approximately 240,000 births per year or 657 births per day, or on average roughly one new-born every two minutes. This compares with the world birth rate of more than 130,000,000 per year or about four births per second [3].

In a simple model for Australia's population from one year to the next we let $x(n)$ denote the population in year n and then the population in year $n+1$ is given by

$$x(n+1) = x(n) + b(n)x(n) - d(n)x(n) + I(n) - E(n), \tag{1}$$

where $b(n)x(n)$ is the total number of newborns in year n (expressed here as a fraction multiplying the total population in year n), $d(n)$ is the fraction of the population that dies in year n, $I(n)$ is the total number of immigrants in year n and $E(n)$ is the total number of emigrants in year n. Given the population at year n we can now predict the population at year $n+1$, provided we know $b(n), d(n), I(n)$ and $E(n)$. But this is almost a mathematical tautology. How can we know $b(n), d(n), I(n)$ and $E(n)$ before the year has elapsed? One possibility is to develop independent models for $b(n), d(n), I(n)$ and $E(n)$. The simplest such model for these numbers is to assume that they do not change from one year to the next so that we can replace them by constants b, d, I and E. In this simple approximation the revised model becomes

$$x(n+1) = (1 + b - d)\, x(n) + I - E. \tag{2}$$

We should keep in mind that the predictions from this model can only be expected to be reliable if the model assumptions are valid. Making model assumptions is a standard part of the mathematical modelling paradigm.

We now have to give some thought to the values of the numbers b, d, I and E. These numbers must come from real world data. A quick Google of the Internet will reveal numerous sites with demographic data. Two sites that are especially useful in this regard are:

- The Australian Bureau of Statistics
  http://www.abs.gov.au [2]
- The CIA World FactBook
  https://www.cia.gov/library/publications/the-world-factbook/
  [3].

The net immigration to Australia has been reasonably constant at about 80,000 per year over many years. Here we take $I - E \approx 80,000$ as a starting approximation. The birth rates and death rates have been less constant, but as a starting point we can adopt the current numbers [3] of 12.02 births per 1000 population and 7.56 deaths per 1000 population. Thus $b = 12.02/1000 = 0.01202$ and $d = 7.56/1000 = 0.00756$. Before we can make a model prediction we need to know the current population. If we look up the entry on Australia in the CIA website we find $x(n) = 20,434,176$ (July 2007 est.). We now predict the population in July 2008 to be

$$x(n+1) = (1 + 0.01203 - 0.00756) \times 20,434,176 + 80,000 = 20,605,517.$$

We could then use this July 2008 prediction as a starting value to predict the population in July 2009 and so on.

According to the Australian Bureau of Statistics web site, Australia's population exceeded 21,000,000 in October 2007 so the above model prediction is not correct, but this is because the input data was not correct. The CIA website data underestimates Australia's current population. One of the fundamental rules of modelling is that what you get out of a model is at best only as reliable as what you put in. This being said, it would not be wise to fuss too much about the input data in the above model because the errors introduced by the other assumptions in the model (for example, a constant death rate) could be even greater than the errors introduced by inaccurate input data. It is better to use the simple model to provide a range of predictions based on a range of inputs.

When it comes to modelling the effect of Australia's ageing population it is useful to consider the population in different age-groups or cohorts. In the next section we consider a basic demographic model for Australia's population growth that takes into account some of the different characteristics of different age-groups.

Model for the Population in Different Age-groups

Here we separate Australia's population into four distinct cohorts: i) Youth 0–14, ii) Prime Working Age 15–54, iii) Semi-Retired 55–64, iv) Retired 65 –.

Suppose that we let $x_1(n)$ denote the number in the first cohort in year n, $x_2(n)$ denote the number in the second cohort in year n and so on. A simple mathematical model for the populations in each cohort in year $n+1$ is now given by

$$\begin{aligned}
x_1(n+1) &= x_1(n) - \left(\frac{1}{15}\right) x_1(n) + b_1(n) - d_1 x_1(n) + I_1 \\
x_2(n+1) &= x_2(n) - \left(\frac{1}{40}\right) x_2(n) + \left(\frac{1}{15}\right) x_1(n) - d_2 x_2(n) + I_2 \\
x_3(n+1) &= x_3(n) - \left(\frac{1}{10}\right) x_3(n) + \left(\frac{1}{40}\right) x_2(n) - d_3 x_3(n) + I_3 \\
x_4(n+1) &= x_4(n) + \left(\frac{1}{10}\right) x_3(n) - d_4 x_4(n) + I_4.
\end{aligned}$$

The fraction $\left(\frac{1}{15}\right)$ that appears in the first of these equations is the fraction of the first age-group that moves into the second age-group in a given year. The parameter $b_1(n)$ that appears in here is not a rate but is the absolute number of new-borns in year n. In a reasonable approximation we will assume that these new borns are the progeny of the females in the second age-group, i.e. only females give birth and only those females in the forty year age-range 15-54 give birth. We could then write $b_1(n) = f\left(\frac{1}{40}\right)\left(\frac{1}{2}\right) x_2(n)$

where f is the average number of surviving children per female. Of course each female will have a whole number of surviving offspring but f, being an average value, is generally not an integer, in fact current estimates put $f \approx 1.7$. We could invest a lot of time in trying to find accurate values for the remaining parameters and this is typically what students do when they start on a modelling problem. But this can be a great waste of time. It is not necessary to have precise knowledge, it is far better to have reasonable values and then put effort into examining how sensitive the model outcomes are to changes in the selected parameter values. For example, we could approximate d_4 as the fraction $\left(\frac{1}{15}\right)$ which is equivalent to assuming that all people who reach age 65 then survive until age 80, or $d_4 \approx \frac{1}{20}$, assuming people who reach age 65 go on to live to age 85. In the following we will take this more optimistic value.

Model Predictions

Now that we have a basic mathematical model we would like to use it to make model predictions about the impacts of different strategies on Australia's ageing population problem. In the following we will consider two strategies: (i) increasing the fertility rate from 1.7 to 3 and (ii) increasing worker participation rates in the semi-retired group. How can we assess the impact of these strategies on the ageing population problem? The total populations in each group are not as informative as the ratios of numbers in different groups. For example, if the working cohort has to economically support the retired cohort, then one measure of the impact would be the ratio of the number of people over the age of 65 to the number of people in prime working years,

$$\text{i.e.} \qquad R(n) = \frac{x_4(n)}{x_2(n)}.$$

Another dependency ratio is the ratio of youth, retired and non-working semi-retired to prime working age plus working semi-retired,

$$\text{i.e.} \qquad r(n,p) = \frac{x_1(n) + (1-p)x_3(n) + x_4(n)}{x_2(n) + px_3(n)}$$

where we have introduced a parameter p to represent the fraction of the semi-retired group that is working. In this ratio the larger the value of $r(n,p)$, the greater the burden placed on the working population to provide for the non-working population. Even with this very crude quantitative measure of the ageing population problem we can see immediately that $r(n,p)$ will decrease with increasing p, i.e. increasing participation rates in the semi-retired group will help to reduce the burden of the ageing population problem. Of course this is obvious but our mathematical model can go further by telling us in quantitative terms how much of an effect this can have. To examine this we need to carry out simulations of the model using real world data as inputs.

There are still many parameters for which we need to find values. As a starting point we might suppose that $I_1 = I_3 = I_4 = 0$ and $I_2 = 80000$. Here we consider that the net immigration numbers are constant at 80,000 per year and all immigrants are in the prime working age-group. We need to do a bit of googling to find death rates for the different cohorts and in the following we will use $d_1 = 0.0004, d_2 = 0.001, d_3 =$

$0.006, d_4 = 0.05$. For example, we suppose that six out of a thousand semi-retired people die each year. Finally we need to input starting population data. In the following simulations we set the total initial population to 21,000,000 with initially 20% in the youth age-group, 56% in the prime working age-group, 11% in the semi-retired age-group, and 13% in the over 65 age-group. These percentages can be estimated from population pyramids published by the ABS. Our mathematical model is now represented by the set of difference equations:

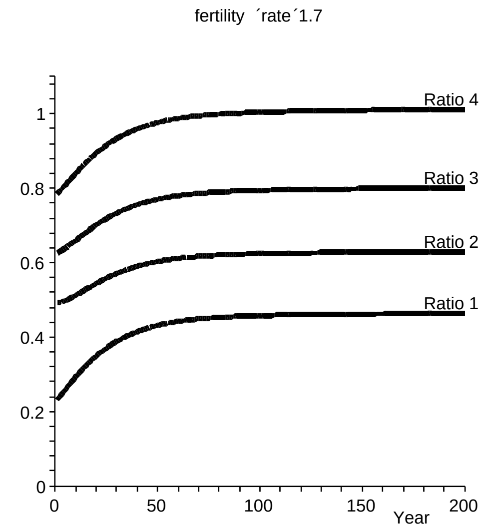
$$\begin{aligned}
x_1(n+1) &= x_1(n) - \left(\frac{1}{15}\right) x_1(n) + \left(\frac{1}{80}\right) f x_2(n) - 0.0004 x_1(n) \\
x_2(n+1) &= x_2(n) - \left(\frac{1}{40}\right) x_2(n) + \left(\frac{1}{15}\right) x_1(n) - 0.001 x_2(n) + 80000 \\
x_3(n+1) &= x_3(n) - \left(\frac{1}{10}\right) x_3(n) + \left(\frac{1}{40}\right) x_2(n) - 0.006 x_3(n) \\
x_4(n+1) &= x_4(n) - \left(\frac{1}{20}\right) x_4(n) + \left(\frac{1}{10}\right) x_3(n),
\end{aligned}$$

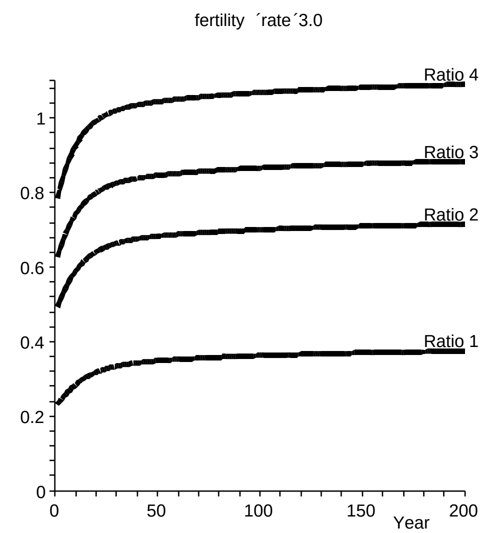
together with the initial conditions

$$\begin{aligned}
x_1(1) &= \left(\frac{20}{100}\right) 21000000 \\
x_2(1) &= \left(\frac{56}{100}\right) 21000000 \\
x_3(1) &= \left(\frac{11}{100}\right) 21000000 \\
x_4(1) &= \left(\frac{13}{100}\right) 21000000.
\end{aligned}$$

These equations are solved in a serial fashion for $x_1(2), x_2(2), x_3(2), x_4(2)$ and then $x_1(3), x_2(3), x_3(3), x_4(3)$ and so on. It is also possible to find an explicit solution to the mathematical model. The set of difference equations is a linear system that can be solved using standard techniques in linear algebra to give $x_1(n), x_2(n), x_3(n)$ and $x_4(n)$ as explicit functions of n. This is an interesting exercise for undergraduate students.

We have carried out simulations using the above input data to forecast populations 200 years into the future with fertility rates at $f = 1.7$ and $f = 3.0$. The results of these simulations are summarized in the figure below. In these graphs, Ratio 1 $= R(n)$, Ratio 2 $= r(n, p = 1)$, Ratio 3 $= r(n, p = 1/2)$, and Ratio 4 $= r(n, p = 0)$ are plotted as a function of year. The results for the current fertility rates are shown in the left plot and those with increased fertility rates are shown in the right plots. The starting values, which are the same in both plots, represent economically affordable values for these ratios (since the budget is in surplus) but future increases in these ratios could cause economic hardship. A detailed economic model would be required to estimate the level of the hardship.





The ageing population problem is most evident in the plot of Ratio 1 with current fertility rates. Recall that Ratio 1 shows the ratio of people over the age of 65 to the ratio of the prime working age-group. According to the model prediction in the left-hand plot, this ratio will more than double over the next fifty years. The strategy of increasing fertility rates to $f = 3$ has a clear positive impact on this ratio (Ratio 1 is lower in the right-hand plot than the left-hand plot) but it would be unwise to conclude that this would be good policy. For a start more babies mean more dependents, as shown in Ratios 2, 3, 4 where the youth age-group is combined with the over 65 age-group as a proportion of the working population. These ratios, taking into account youth dependencies as well as aged dependencies, are higher with higher fertility rates. But the overall numbers in the population will also be considerably greater with greater fertility rates and the possibilities of greater economies of scale may partially compensate the dependency ratios. Again this would require an economic model. As an aside, it is interesting to reflect on the magnitude of the growth in population with the increased fertility rates. We can readily find the values from our model simulations. With a fertility rate of $f = 1.7$ our model predicts a population of slightly less than thirty million in two hundred years time. However the fertility rate of $f = 3.0$ results in a model prediction of more than one hundred and fifty million in two hundred years time.

There is an answer to the ageing population problem that can be spotted in the above plots. The answer lies in increasing worker participation rates in the semi-retired cohort. For example, if participation rates in this cohort are currently at 50% ($p = 0.5$) then Ratio 3 is the most relevant dependency ratio and there is no gain in increasing fertility rates, as discussed above. However if participation rates were suddenly changed from 50% to 100% then the most relevant dependency ratio curve would jump from Ratio 3 to Ratio 2. But the curve for Ratio 2, with current fertility rates, at all times remains below the starting value on the curve for Ratio 3 representing a currently affordable option (assuming the current 50% participation in the semi-retired group).

It is interesting to explore the model for other parameters too. Here are a few suggestions. Use the above model to assess the impact of the following: (i) a ten-year increase in life expectancy; (ii) twenty percent worker participation among the retired group; (iii) forty percent of the aged group self sufficient; (iv) a threefold increase in skilled immigration numbers.

References

[1] *Australia's Demographic Challenges* Social Policy Division, The Treasury, Commonwealth of Australia, 2004,

http://demographics.treasury.gov.au/content/discussion.asp

[2] *Census Profile Activities – Australia's Ageing Population* Australian Bureau of Statistics, Commonwealth of Australia,

http://www.abs.gov.au

[3] *CIA - World Fact Book*

https://www.cia.gov/library/publications/the-world-factbook/