Regional Population Projection Program (RePPP) Model Description

Tom Wilson
Northern Institute
tom.wilson@cdu.edu.au

ISSN: 2207-3388
The author

Dr Tom Wilson is Principal Research Fellow in the Northern Institute at Charles Darwin University. He has 20 years’ experience in population projections research and practice.

Contact: tom.wilson@cdu.edu.au

Obtaining RePPP

The RePPP projection program is available for purchase from Charles Darwin University. Please contact Tom Wilson using the email address above.

Suggested citation

1. Introduction

This paper describes the equations of the projection model contained in the Regional Population Projection Program (RePPP). RePPP produces subnational population projections for between 2 and 50 regions by sex and five year age group over a projection horizon of up to 50 years. Instructions on how to use RePPP are provided in a separate User Guide (Wilson 2017).

The paper is limited to describing the mathematical details of the projection model rather than the computer program in which it is incorporated (which contains many other VBA subroutines and Excel calculations). The paper also does not discuss methods used to prepare input data and set projection assumptions. These are supplied by the user and some suggestions are provided in the User Guide (Wilson 2017).

Notation used throughout the paper is set out in Table 1 below.

Table 1: Notation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASFR</td>
<td>Age-specific fertility rate</td>
</tr>
<tr>
<td>Aus</td>
<td>Australia</td>
</tr>
<tr>
<td>B</td>
<td>Births</td>
</tr>
<tr>
<td>D</td>
<td>Deaths</td>
</tr>
<tr>
<td>E</td>
<td>Emigration</td>
</tr>
<tr>
<td>I</td>
<td>Immigration</td>
</tr>
<tr>
<td>IM</td>
<td>In-migration</td>
</tr>
<tr>
<td>N</td>
<td>Net total migration (internal + international)</td>
</tr>
<tr>
<td>NIM</td>
<td>Net internal migration</td>
</tr>
<tr>
<td>NOM</td>
<td>Net overseas (international) migration</td>
</tr>
<tr>
<td>OM</td>
<td>Out-migration</td>
</tr>
<tr>
<td>P</td>
<td>Population</td>
</tr>
<tr>
<td>Pconstraint</td>
<td>Total population constraint</td>
</tr>
<tr>
<td>SRB</td>
<td>Sex Ratio at Birth</td>
</tr>
<tr>
<td>a</td>
<td>Age group</td>
</tr>
<tr>
<td>b</td>
<td>Newly-born period-cohort</td>
</tr>
<tr>
<td>d</td>
<td>Period-cohort death rate</td>
</tr>
<tr>
<td>e</td>
<td>Period-cohort emigration rate</td>
</tr>
<tr>
<td>f</td>
<td>Females</td>
</tr>
<tr>
<td>i</td>
<td>Region</td>
</tr>
<tr>
<td>im</td>
<td>Period-cohort in-migration rate</td>
</tr>
<tr>
<td>m</td>
<td>Males</td>
</tr>
<tr>
<td>om</td>
<td>Period-cohort out-migration rate</td>
</tr>
<tr>
<td>pc</td>
<td>Period-cohort</td>
</tr>
<tr>
<td>s</td>
<td>Sex</td>
</tr>
<tr>
<td>sf</td>
<td>Scaling factor</td>
</tr>
<tr>
<td>t</td>
<td>Time t</td>
</tr>
<tr>
<td>t + 5</td>
<td>Time 5 years after t</td>
</tr>
</tbody>
</table>
2. Projection model framework

2.1. Projection model type

The projection model within RePPP is a multi-bi-regional cohort-component model. This is a simplified version of the standard multi-regional model (Rogers 1995; Rees 1997) which is created by combining multiple bi-regional models. The ‘bi-regional’ description refers to the way in which internal migration is modelled. In other words, out-migration from every region specified by the user is modelled as migration from that region (the first region) to the remainder of the country (the second “region”); in-migration is the migration flow in the opposite direction.

Empirically, the multi-bi-regional model produces results very similar to those of the standard multi-regional model but requires much less input migration data (Wilson and Bell 2004; Rogers 1976). And because the bi-regional migration flows are much larger than specific origin-destination flows, there is much less noise in the data, and therefore less need for migration age profile smoothing and adjustment.

The projection model is also deterministic. It produces a single set of numbers in each run, rather than probabilistic predictive intervals (Wilson 2013).

2.2. Accounting framework

The organising framework of the model is that of movement population accounts as described by Rees (1984, 1986). It therefore uses movement rather than transition migration data (Rees and Willekens 1986). Movement migration data comprise counts of moves over boundaries over a period of time; transition migration data are changes of location between two points in time. Because the two types of migration data are measured from different perspectives they tend to give very different migration flow numbers even though they measure the same phenomenon of population mobility.

The model employs a typical approach to the projection of demographic components of change in a movement accounting framework. Most demographic events are projected as:

\[
\text{event} = \text{demographic rate} \times \text{person-years at risk.}
\]

The only exception is immigration, which is input directly as immigration numbers.

2.3. Calculation scheme

The projections are computed in an iterative calculation scheme programmed in Excel VBA. This offers a practical way of incorporating migration constraints and outputting detailed demographic components. The fact that end-of-interval populations are included on the right-hand side of many of the model’s equations is therefore unimportant because these populations are updated in successive iterations. Iteration stops when the projected population in every age-sex-region category varies by less than a very small amount from the previous iteration.

2.4. Age and time

The projection model proceeds in five year intervals and uses a five year age group breakdown of the population up to a final open-ended age group of 85+.

Deaths and migration are modelled in period-cohorts, defined as the age-time space occupied by a cohort over a projection interval (Figure 1 below; see also Rees and Woods 1986, pp. 305-312). For example, the period-cohort labelled ‘10-14 to 15-19’ refers to the cohort aged 10-14 at the start of a projection interval which ages to 15-19 by the end of it.

Births are projected using age groups (not period-cohorts).

2.5. Geography

The geographical areas in the model are subnational regions chosen by the user. These can vary in number between 2 and 50
and must cover the whole country. Users who wish to produce projections for regions in just one part of the country (e.g. SA4 regions in one State of interest) can do so but must add a final ‘rest of Australia’ region to cover the rest of the country.

There is only one geographical scale in the model. There are no higher level geographies to which the regional projections are constrained, or lower level geographies of small areas.

Figure 1: Period-cohorts used in the projection model
3. Accounting equations

Population accounting equations lie at the heart of the projection model. For all period-cohorts except newly-born infants this is:

\[ P_{s,a}^i(t+5) = P_{s,a}^i(t) - D_{s,pc}^i - O M_{s,pc}^i - E_{s,pc}^i + I M_{s,pc}^i + I_{s,pc}^i. \]

For newly-born infants the “initial” population comprises the number of births that occur during the interval:

\[ P_{s,0-4}^i(t+5) = B_{s}^i - D_{s,b}^i - O M_{s,b}^i - E_{s,b}^i + I M_{s,b}^i + I_{s,b}^i. \]

These accounting equations apply irrespective of the user-selected options, input data, and assumptions. Sections 4 and 5 describe how the various accounting equation terms are calculated. Births and deaths are always calculated in the same way. However, the way in which the migration terms are computed varies according to the user-selected migration option.
4. Births and deaths equations

4.1. Births

Births are calculated by multiplying age-specific fertility rates (for ages 15-19, 20-24, …, 45-49) by the number of female age-specific person-years at risk:

\[ B^i_a = ASFR^i_a \times \frac{s}{2} \left( \frac{P^i_{f,a}(t) + P^i_{f,a}(t + 5)}{2} \right). \]

The projected births are summed over age groups and then split into male and female births using the sex ratio at birth:

\[ B^i_f = B^i \times \frac{100}{SRB + 100}, \]
\[ B^i_m = B^i \times \frac{SRB}{SRB + 100}. \]

4.2. Deaths

Deaths are projected by multiplying death rates by sex and period-cohort by the number of person-years at risk. For all periodcohorts except newly-born infants the equation is:

\[ D^i_{s,pc} = d^i_{s,pc} \times \frac{s}{2} \left( \frac{P^i_{s,pc}(t) + P^i_{s,pc}(t + 5)}{2} \right) \]

whilst for newly-born infants the person-years approximation of Willekens and Drewe (1984) is applied:

\[ D^i_{s,b} = d^i_{s,b} \times \frac{s}{2} \left( \frac{P^i_{s,0-4}(t) + P^i_{s,0-4}(t + 5)}{2} \right) \]
5. Migration equations

The projection of migration varies according to the migration option selected by the user in the RePPP Excel workbook.

5.1. Migration option 1

If migration option 1 is chosen all migration variables are simply set to zero:

\[
OM_{s,pc}^i = 0 \\
E_{s,pc}^i = 0 \\
IM_{s,pc}^i = 0 \\
I_{s,pc}^i = 0
\]

5.2. Migration option 2

If migration option 2 is selected migration is projected using the multi-bi-regional model and the migration rates and immigration flows supplied by the user in the RePPP Excel workbook.

First, out-migration from each region to the remainder of the country is projected. Out-migration for all period-cohorts except newly-born infants is calculated as:

\[
OM_{s,pc}^i = om_{s,pc}^i \frac{s}{2} \left( p_{s,pc}^i(t) + p_{s,pc}^i(t + 5) \right)
\]

and for newly-born infants as:

\[
OM_{s,b}^i = om_{s,b}^i \frac{s}{2} p_{s,0-4}^i(t + 5)
\]

Second, preliminary in-migration to each region from the remainder of the country is calculated. In-migration for all period-cohorts except newly-born infants is computed as:

\[
IM_{s,pc}^i(1) = im_{s,pc}^i \frac{s}{2} \left( p_{s,pc}^{Aus-i}(t) + p_{s,pc}^{Aus-i}(t + 5) \right)
\]

and for newly-born infants as:

\[
IM_{s,b}^i(1) = im_{s,b}^i \frac{s}{2} p_{s,0-4}^{Aus-i}(t + 5)
\]

In-migration for each sex and period-cohort summed across all regions must equal out-migration summed across all regions. In-migration is adjusted by the model to ensure the total numbers of internal in- and out-migrations are the same. Adjusted in-migration is thus:

\[
IM_{s,pc}^i(2) = IM_{s,pc}^i(1) \frac{\sum_i OM_{s,pc}^i}{\sum_i IM_{s,pc}^i(1)}
\]

The amount of adjustment required is usually very small.

Immigration numbers supplied by the user are input directly into the projections. Emigration is calculated in the same way as out-migration:

\[
E_{s,pc}^i = e_{s,pc}^i \frac{s}{2} \left( p_{s,pc}^i(t) + p_{s,pc}^i(t + 5) \right)
\]

for all period-cohorts except newly-born infants, and

\[
E_{s,b}^i = e_{s,b}^i \frac{s}{2} p_{s,0-4}^i(t + 5)
\]
for newly-born infants.

### 5.3. Migration option 3

If migration option 3 is selected preliminary migration flows are first calculated from user-supplied migration rates and immigration flows (as for migration option 2). But they are then adjusted to match user-specified net internal migration and net overseas (international) migration totals. Adjusted immigration, emigration, in-migration and out-migration totals (summed over sex and period-cohort) are calculated before adding sex and period-cohort detail.

Overseas migration is computed first. A scaling factor $sf$ for each region is estimated such that immigration ($I$) multiplied by $sf$ minus emigration ($E$) divided by $sf$ equals the user-specified value of net overseas migration (NOM):

$$ I sf - \frac{E}{sf} = NOM. $$

A quadratic equation is used to find $sf$:

$$ sf = \frac{NOM + \sqrt{(NOM^2 + 4 I E)}}{2 I}. $$

Then new immigration and emigration totals for each region are calculated as:

$$ I'(2) = I^i sf^i $$

$$ E'(2) = E^i / sf^i. $$

If total NOM is higher than implied by the immigration numbers and emigration rate assumptions, the method will increase immigration and decrease emigration. Conversely, if the set NOM is lower, it will reduce immigration and increase emigration.

Immigration and emigration by sex and period-cohort are then adjusted to the new $I'(2)$ and $E'(2)$ totals:

$$ I_{s,pc}^i(2) = I_{s,pc}^i(1) \frac{I(2)}{\sum_s \sum_{pc} I_{s,pc}^i(1)} $$

$$ E_{s,pc}^i(2) = E_{s,pc}^i(1) \frac{E(2)}{\sum_s \sum_{pc} E_{s,pc}^i(1)}. $$

Internal migration is calculated next. Preliminary out-migration and in-migration projections by sex and period-cohort are calculated using user-supplied rates and populations-at-risk (as for migration option 2).

The quadratic equation is used again to find a scaling factor $sf$ for each region to adjust in- and out-migration totals:

$$ sf = \frac{NIM + \sqrt{(NIM^2 + 4 IM OM)}}{2 IM}. $$

Then new in-migration and out-migration totals are calculated:

$$ IM^{i}(2) = IM^{i} sf^{i} $$

$$ OM^{i}(2) = OM^{i} / sf^{i}. $$

A nice feature of this adjustment method is that adjusted in- and out-migration grand totals (summed over region, sex and period-cohort) are identical, even if the preliminary grand totals are not.

Out-migration by sex and period-cohort is adjusted:

$$ OM_{s,pc}^{i}(2) = OM_{s,pc}^{i}(1) \frac{OM(2)}{\sum_s \sum_{pc} OM_{s,pc}^{i}(1)} $$

Adjusted in-migration by sex and period-cohort is found by iterative proportional fitting because it must:
(i) sum over sex and period-cohort to total in-migration $IM_i(2)$, and

(ii) sum across regions to equal out-migration by sex and period-cohort summed across regions.

### 5.4. Migration option 4

If migration option 4 is chosen, migration flows are determined indirectly via an independent total population constraint. This might be obtained, for example, from a separate dwelling-led model, an employment-led model, or an extrapolative model of total populations.

Preliminary migration flows are calculated from user-supplied migration rates and immigration flows as before. Then total net migration (net internal and net international combined) is calculated as a residual:

$$N^i = P_{\text{constraint}}(t + 5) - P(t) - B + D.$$  

A quadratic equation is used to find a scaling factor $sf$ for each region to adjust internal and international migration totals:

$$sf = \frac{N + \sqrt{(N^2 + 4(I + IM)(E + OM))}}{2(I + IM)}.$$  

Then new directional migration totals are calculated:

$$IM^i(2) = I^i sf^i$$

$$EM^i(2) = E^i sf^i$$

$$OM^i(2) = OM^i sf^i.$$  

However, the grand in-migration and out-migration totals will differ to some extent. They need to be adjusted so that they are equal. A simple average is taken:

$$IM(3) = \frac{[IM(2) + OM(2)]}{2}$$

$$OM(3) = \frac{[IM(2) + OM(2)]}{2}.$$  

The immigration and emigration grand totals also need adjustment to ensure that the grand totals of in-migration – out-migration + immigration – emigration equal total net migration summed across all regions. Thus:

$$I(3) = I(2) - (IM(3) - IM(2))$$

$$E(3) = E(2) - (OM(3) - OM(2)).$$  

The matrix of regional migration totals now looks like this.

<table>
<thead>
<tr>
<th>Region</th>
<th>In-migration</th>
<th>Out-migration</th>
<th>Immigration</th>
<th>Emigration</th>
<th>Total net mig</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$IM^1(2)$</td>
<td>$OM^1(2)$</td>
<td>$I^1(2)$</td>
<td>$E^1(2)$</td>
<td>$N^1$</td>
</tr>
<tr>
<td>2</td>
<td>$IM^2(2)$</td>
<td>$OM^2(2)$</td>
<td>$I^2(2)$</td>
<td>$E^2(2)$</td>
<td>$N^2$</td>
</tr>
<tr>
<td>3</td>
<td>$IM^3(2)$</td>
<td>$OM^3(2)$</td>
<td>$I^3(2)$</td>
<td>$E^3(2)$</td>
<td>$N^3$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>$IM(3)$</td>
<td>$OM(3)$</td>
<td>$I(3)$</td>
<td>$E(3)$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

Special iterative proportional fitting is applied to constrain the matrix values to the row and column marginal (in the grey cells). Then out-migration, immigration and emigration by sex and period-cohort are constrained to the newly fitted totals:
\[
OM_{s,pc}(3) = OM_{s,pc}(1) \frac{OM^i(3)}{\sum_s \sum_{pc} OM_{s,pc}(1)}
\]

\[
I_{s,pc}(3) = I_{s,pc}(1) \frac{I^i(3)}{\sum_s \sum_{pc} I_{s,pc}(1)}
\]

\[
E_{s,pc}(3) = E_{s,pc}(1) \frac{E^i(3)}{\sum_s \sum_{pc} E_{s,pc}(1)}
\]

Adjusted in-migration by sex and period-cohort is found by iterative proportional fitting because it must:

(i) sum over sex and period-cohort to total in-migration \(IM^i(3)\), and
(ii) sum across regions to equal out-migration by sex and period-cohort summed across regions.
References


