DEMOGRAPHIC RESEARCH

A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 46, ARTICLE 32, PAGES 919–956
PUBLISHED 17 MAY 2022
https://www.demographic-research.org/Volumes/Vol46/32/
DOI: 10.4054/DemRes.2022.46.32

Research Article

Preparing local area population forecasts using a bi-regional cohort-component model without the need for local migration data

Tom Wilson

Contents

Demographic Research: Volume 46, Article 32
Research Article

Preparing local area population forecasts using a bi-regional cohort-component model without the need for local migration data

Tom Wilson[1]

Abstract

BACKGROUND
Cohort-component models incorporating directional migration are conceptually robust demographic models which are widely employed to forecast the populations of large subnational regions. However, they are difficult to apply at the local area scale. Simpler models, such as the Hamilton–Perry model, have modest input data requirements and are much quicker, cheaper, and easier to implement, but they offer less output detail, suffer from some conceptual and practical limitations, and can be less accurate.

OBJECTIVE
The aim of this paper is to describe and evaluate the synthetic migration cohort-component model – an approach to implementing the bi-regional model for local area population forecasts without the need for any locally specific migration data.

METHODS
The new approach is evaluated by creating several sets of ‘forecasts’ for local areas of Australia over past periods. For comparison, forecasts from two types of Hamilton–Perry model are also evaluated. Error is measured primarily with an alternative Absolute Percentage Error measure for total population which takes into account how well or poorly the population age–sex structure is forecast.

RESULTS
In the evaluation for Australian local areas, the synthetic migration model generated more accurate forecasts that the two Hamilton–Perry models in terms of median, mean, and 90^{th} percentile Absolute Percentage Errors.

CONTRIBUTION
The synthetic migration model combines the conceptual and practical advantages of the bi-regional cohort-component model with the light data requirements and ease of calculation of simpler cohort models. It allows the bi-regional model to be applied in circumstances where local area migration data are unavailable or unreliable.

[1] University of Melbourne, Email: wilson.t1@unimelb.edu.au.

1. Introduction

Local area population forecasts regularly inform planning for future service provision and infrastructure development, such as kindergarten demand, school building and staffing, water supply, health services, local leisure facilities, and public transport. They are also used for commercial market assessment and site selection (Baker, Swanson, and Tayman 2021), and in some countries they are required for the revision of electoral district boundaries (e.g., AEC 2021). Forecasts help answer questions such as: Does forecast demographic change justify the construction of a new primary school, a new nursing home, or a new supermarket in the local area? Because many uses of forecasts are limited to specific age groups, age-specific population forecasts must be prepared.

The ‘gold standard’ for forecasting the populations of large subnational regions by age and sex is the multiregional cohort-component model in which internal migration is modelled between each origin and destination (Rees and Wilson 1977; Rogers 1995). However, at the local area scale, this is a very demanding model to implement. In many countries the necessary input migration data is simply unavailable. Even where it is possible to acquire the data, the size and sparsity of an origin–destination–age–sex migration matrix which includes hundreds or thousands of local areas is very challenging. Considerable time and effort have to be spent on data collation, estimating suppressed or missing cells and smoothing, assumption-setting, programming, and input and output data checking, all of which involves substantial staff time and cost. One option is to switch to the reduced-form bi-regional version of the model in which internal in- and out-migration is modelled between each local area and the rest of the country (Isserman 1993; Rogers 1976). It requires much less migration data but produces forecasts similar to that of the fully multiregional model (Rogers 1976; Wilson and Bell 2004). Both these types of directional-migration model have many strengths. They offer considerable flexibility of assumption-setting, detailed migration outputs, and, by modelling separate internal and international inward and outward migration flows, they faithfully represent actual demographic processes. Assumed future changes to migration rates are therefore likely to yield sensible population change outcomes. From a practical standpoint, directional-migration models tend to produce plausible forecasts, and applied demography textbooks recommend their use where possible (e.g., Pittenger 1976; Smith, Tayman, and Swanson 2013). However, even the reduced-form bi-regional version still requires a considerable amount of migration data and data preparation.

Alternative, simpler, cohort-component models can be used to prepare age–sex population forecasts where local area internal and/or international migration data are unavailable, unreliable, or when the application of a multiregional or bi-regional model is impractical (Wilson et al. 2021). Some cohort-component models incorporate migration through net migration, either as net migration numbers (Rowland 2003), net migration rates (Smith 1986), or a mixture of rates (for negative net migration) and numbers (positive net migration) (Wilson 2016). Net migration may be treated as a single variable, or divided into internal and international components. The advantages of net migration cohort-component models include their low data requirements and easy calculation. However, they are vulnerable to some practical limitations. Net migration rates applied to rapidly growing areas can produce implausibly high projected population growth. Adjusting age patterns of net migration to be consistent with higher or lower net migration assumptions is challenging (Pittenger 1976, 1978), and models which use net migration numbers can, in some circumstances, generate negative populations for age groups where assumed net migration is negative. In addition, from a conceptual perspective, modelling net migration is less satisfactory than directional migration because net migration is not a demographic process itself, but the net result of migration streams flowing in opposite directions.

The Hamilton–Perry model avoids dealing with net migration and simplifies the calculations further (Baker et al. 2017; Hamilton and Perry 1962; Smith, Tayman, and Swanson 2013). Population is projected by multiplying the start-of-interval population by a Cohort Change Ratio, defined as the ratio of a cohort's population at one point in time to its population size at an earlier time (often 5 years earlier) when (5 years) younger. The greatest strength of the model is the ability to create age–sex population forecasts very easily and with minimal data requirements: it only requires population estimates by age and sex at two points in time to estimate Cohort Change Ratios. Populations in the 0–4 age group are often forecast by multiplying Child/Woman Ratios by the female population aged 15–49. Limitations include the lack of demographic components (births, deaths, and migration) in the modelling process, and therefore the inability to formulate projection assumptions for fertility, mortality, and migration, and the absence of forecast outputs of these components. It is also difficult to adjust the age profile of Cohort Change Ratios in a sensible way to reflect assumed changes in overall population growth. Rapid population growth from the development of a new suburb, for example, would usually result in most growth being concentrated in the peak migratory age groups – the young adult and young childhood ages. Over the last decade or so, the Hamilton–Perry model has enjoyed a resurgence in popularity for local area forecasts in the US, with various extensions and improvements proposed (e.g., Baker et al. 2017; Baker, Swanson, and Tayman 2021; Hauer 2019; Inoue 2017; Swanson, Schlottmann, and Schmidt 2010; Tayman, Swanson, and Baker 2021). Several studies have shown that constraining to independent local area total populations substantially improves accuracy (Baker, Swanson, and Tayman 2021; Tayman, Swanson, and Baker 2021; Wilson 2016). A recent evaluation of various types of Hamilton–Perry model revealed that the lowest errors and the smallest bias were obtained from a version employing Cohort Change Ratios when a cohort population is decreasing and Cohort Change Differences when it is increasing (Wilson and Grossman 2022).

One might expect that more complex models (e.g., multiregional or bi-regional) would generate more accurate forecasts than simpler models (e.g., net migration cohort-component or Hamilton–Perry) but the evidence to date is not clear cut. In empirical tests, Smith and Tayman (2003) found little difference in the accuracy of Hamilton–Perry and cohort-component models for producing age–sex forecasts for the counties of Florida. Wilson (2016) obtained greater accuracy from the bi-regional cohort-component model than the Hamilton–Perry model in an evaluation of forecasts for local government areas in New South Wales, Australia. However, when the outputs of both models were constrained to independent total population forecasts, much of the difference disappeared and the Hamilton–Perry model came close in accuracy to the bi-regional model (except at the older ages). Several other studies have confirmed the benefits of constraining age–sex forecasts to independent forecasts of total population (e.g., Baker, Swanson, and Tayman 2021; Reinhold and Thomsen 2015; Tayman, Swanson, and Baker 2021). The recent evaluation of Hamilton–Perry models by Wilson and Grossman (2022) revealed reasonable accuracy in forecasting local age–sex populations overall, though larger errors occurred for new suburbs with rapid population growth and inner city locations with approximately static age distributions dominated by young adults.

Given the widespread use of local area forecasts for planning, budgeting, and investment decisions, it is important to produce forecasts with as much accuracy as possible. Preferably, the forecasts would be generated by a conceptually strong model to reduce the risk of implausible outputs. At the same time, it would be useful if the forecast preparation process was as simple, data-light, quick, and easy as the Hamilton–Perry models. Ideally, the projection model would incorporate directional migration in local area cohort-component calculations without the need for detailed local area migration data.

In this paper, I suggest and evaluate an approach which attempts to achieve this. The ‘synthetic migration cohort-component model’ uses directional inward and outward migration estimated for each local area from a single model migration age schedule adjusted to fit the local area’s age–sex net migration pattern for a recent period. The details of the modelling approach are set out in the next section of the paper. The model was evaluated empirically by applying it to create population ‘forecasts’ for local areas of Australia over several past periods, which were then evaluated against actual population estimates. Errors from the new model’s forecasts were also compared to those from two other ‘competitor’ models which similarly do not use local area migration input data. The third main section of the paper describes the input data and error metrics used. The results of the evaluation are presented in the next section, which includes sensitivity analyses of the forecasts to some of the simplifying assumptions. The final section summarises the key strengths and limitations of the model, explains why it works reasonably well, and sets out some avenues for further research.

2. The synthetic migration cohort-component model

The synthetic migration cohort-component model combines the strengths of directional migration cohort-component modelling and local area net migration patterns, whilst minimising input data requirements, preparation time, and costs. The model is presented in two versions, one which deals with a single area of interest, and the other which includes all local areas within a country or state/province. These versions match the most common likely uses of the model. The term ‘model’ here is used to cover the full set of projection calculations, including the cohort-component model, the associated ways of preparing simplified fertility, mortality, and migration inputs, and constraining to separate forecasts of local area total populations. The model calculations of the single area synthetic migration model are explained first, followed by a description of how the input data are prepared. Then the multiple area version of the model is presented.

2.1 Single area version

The synthetic migration cohort-component model incorporates directional migration, but in the simplest way possible. Migration flows consist of inward migration (immigration from other countries plus internal in-migration) and outward migration (emigration to other countries plus internal out-migration). It is therefore a bi-regional model comprising the region of interest and a rest-of-the-world ‘region’. The accounting basis of the model is movement population accounts in which the migration data consists of migration events rather than transitions (Rees 1984, 1985). To keep the modelling simple, it uses 5-year age groups and 5-year time intervals. The population accounting equation is written as:

$$P^i_{s,a+5}(t+5) = P^i_{s,a}(t) - D^i_{s,a\to a+5}(t,t+5) - O^i_{s,a\to a+5}(t,t+5) + I^i_{s,a\to a+5}(t,t+5) \qquad (1)$$

where:

P = population
D = deaths
O = outward migration
I = inward migration

i = local area
s = sex
t = point in time
$t, t+5$ = the projection interval between t and 5 years after t
a = age group
$a \rightarrow a+5$ = the change in a cohort's age from a to $a+5$ during the t to $t+5$ projection interval. It refers to the parallelogram-shaped period-cohort space in a Lexis diagram (Bell and Rees 2006).

For the newly born infant cohort, the initial population in Equation 1 is replaced by births. Births are projected in the conventional way by multiplying age-specific fertility rates by age-specific female populations-at-risk, with disaggregation by sex via the sex ratio at birth. Deaths and outward migration are projected using occurrence/exposure rates multiplied by the area's population-at-risk. However, inward migration is handled as flows rather than rates because the population-at-risk is less clear cut than for outward migration, and it keeps the modelling simpler than if rates and a rest-of-the-world population-at-risk were used. The projection equation may be summarised as:

$$P^i_{s,a+5}(t+5) = \frac{\left(1 - \frac{5}{2} d^i_{s,a\rightarrow a+5} - \frac{5}{2} o^i_{s,a\rightarrow a+5}\right)}{\left(1 + \frac{5}{2} d^i_{s,a\rightarrow a+5} + \frac{5}{2} o^i_{s,a\rightarrow a+5}\right)} P_{s,a}(t) + \frac{1}{\left(1+\frac{5}{2} d^i_{s,a\rightarrow a+5}+\frac{5}{2} o^i_{s,a\rightarrow a+5}\right)} I^i_{s,a\rightarrow a+5} \qquad (2)$$

where:

d = death rate
o = outward migration rate.

However, in coding the model I took a 'deconstructed' approach in which each component is calculated separately in an iterative calculation scheme. This keeps the code relatively simple and transparent, and permits any constraining to be applied easily. Thus, deaths and outward migration are forecast as:

$$\begin{aligned} D^i_{s,a\rightarrow a+5}(t,t+5) &= d^i_{s,a\rightarrow a+5}\ \tfrac{5}{2}\left(P^i_{s,a}(t) + P^i_{s,a+5}(t+5)\right) \\ O^i_{s,a\rightarrow a+5}(t,t+5) &= o^i_{s,a\rightarrow a+5}\ \tfrac{5}{2}\left(P^i_{s,a}(t) + P^i_{s,a+5}(t+5)\right) \end{aligned} \qquad (3)$$

where end-of-interval populations are set to zero in the first iteration and then updated in successive iterations. The iterative approach is taken because initial migration flows are adjusted to be consistent with an independent total population forecast. In the application shown later in this paper, an average of four extrapolative models was selected to generate this total population forecast, though other models could be used; for example, a housing-unit model (Foss 2002), a simple extrapolative model such as a linear model (Smith, Tayman, and Swanson 2013: chapter 8), or forecasts from a machine learning model (Grossman et al. 2022).

The adjustment to the total population constraint consists of three steps. First, the required volume of net migration to obtain the total population constraint is calculated using the population accounting relationship:

$$N^i = P^i(t+5) - P^i(t) - B^i + D^i \tag{4}$$

where:

N = total net migration
P = total population
B = total births
D = total deaths.

Second, the total amount of inward migration which is consistent with the required total net migration is calculated simply as total outward migration plus the required total net migration:

$$I^i = O^i + N^i.$$

Third, inward migration by sex and period-cohort is scaled by the ratio of the required inward migration total to the sum of preliminary inward migration over sex and period-cohort:

$$I^i_{s,a\to a+5}[2] = I^i_{s,a\to a+5}\ \frac{I^i}{\sum_s \sum_a I^i_{s,a\to a+5}} \tag{5}$$

where $[2]$ refers to an updated value.

The iterative calculations then continue, updating the end-of-interval populations using Equation 1 and forecast births, deaths, and migrations until convergence is achieved.

In theory, both outward and inward migration could be adjusted proportionally to match the required net migration value. However, if net migration assumptions embedded in the total population forecasts (Equation 4) were substantially lower than in the base period, it would result in outward migration being increased substantially and inward migration being greatly reduced. At the local area scale, outward migration rates in some ages are already very high. Thus, this approach would risk increasing outward migration to the extent that negative populations would result. For practical reasons therefore, only inward migration is adjusted.

2.2 Base period population accounts and data preparation

The outward migration rates and preliminary inward migration flows for the forecasts are prepared from a set of base period population accounts which cover the 5-year period up to the jump-off point of the forecasts. The method described in this section draws upon the principles of population accounting data preparation (Rees and Wilson 1977; Rees and Willekens 1986) and elements of the data adjustment approach of Simpson and Snowling (2011). The data inputs required to calculate the population accounts are listed in Table 1. At a minimum, the data preparation does not require any local fertility, mortality, or migration data, just local area population estimates by sex and 5-year age group for the jump-off year and 5 years earlier. Some national fertility and mortality rates and a set of model migration rates are required.

Table 1: Data required to calculate the base period population accounts

Variable	Summary
Populations	Local area population estimates by sex and 5-year age group for the start and end of the base period.
Births	Either: recorded local area base period births by sex (if available). Or: model age-specific fertility rates (e.g., national rates) to estimate births indirectly.
Deaths	Either: recorded base period deaths by sex and period-cohort (if available). Or: a national life table for the period, or a mortality surface of ${}_nL_x$ values, to estimate deaths indirectly.
Net migration	No input additional data are required (net migration by sex and period-cohort for the base period is calculated as an accounting residual).
Outward and inward migration	Synthetic base period outward and inward migration flows by sex and period-cohort are estimated from three data inputs: (1) model migration rates by sex and period-cohort, (2) a crude migration turnover rate (inward + outward migration divided by the total population-at-risk) which can be area-specific or universal, and (3) local area net migration by sex and period-cohort (just calculated).

Population estimates by sex and 5-year age group are required for the jump-off year of the forecasts and 5 years earlier. These form the starting and ending populations of the base period population accounts.

In many cases, recorded births by sex for the base period will be available. If not, births can be estimated indirectly. First, the Total Fertility Rate is estimated from the jump-off year age–sex population estimates using the xTFR estimation method of Hauer and Schmertmann (2020, Equation 5) which derives the TFR from the population age–sex distribution. The estimated TFR is:

$$xTFR = \left(10.65 - 12.55\,\frac{W_{25-34}}{W_{15-49}}\right)\frac{C_{0-4}}{W_{15-49}} \tag{6}$$

where:

C = children
W = women.

Age-specific fertility rates are estimated by scaling a set of model fertility rates (e.g., national rates for the base period) to the TFR estimate. These rates are multiplied by base period female populations-at-risk to estimate the number of births. Total births are then divided into males and females by an assumed (e.g., national) sex ratio at birth.

If recorded base period deaths by sex and period-cohort are available, those data can be used. If not, they can be estimated indirectly by multiplying modelled period-cohort death rates by base period populations-at-risk. The simplest model death rates are national death rates obtained from a life table. Assuming national mortality levels is clearly an approximation, but it will generally produce reasonable results unless local mortality is known to be very different from the national average. If local life expectancy at birth statistics are available, but age-specific deaths data are not (or are too sparse or noisy), then it is possible to derive approximate local death rates via model life tables or a national mortality surface (Wilson 2018). The mortality surface consists of a series of past and forecast national life tables. By matching local life expectancy at birth to the point on the mortality surface with equivalent T_0 values it is possible to calculate death rates by sex and period-cohort from the ${}_nL_x$ values. A more detailed description is given in Wilson (2018) and a simple illustration is available from https://doi.org/10.6084/m9.figshare.17088539.v1.

At the local area scale, migration is generally the most important demographic component of change. The way this is estimated is the defining feature of the synthetic migration cohort-component model. The estimation process starts with model migration-age schedule rates and finishes with synthetic inward and outward migration flows by sex and period-cohort which are consistent with the base period net migration age–sex pattern. An illustration of how migration-age profiles develop over the migration estimation process is shown in Figure 1 using the example of female migration in and out of one inner city local area. The calculation consists of seven steps.

Figure 1: Base period inward and outward migration-age profiles at various points in the synthetic migration estimation process, Adelaide City SA3 area females

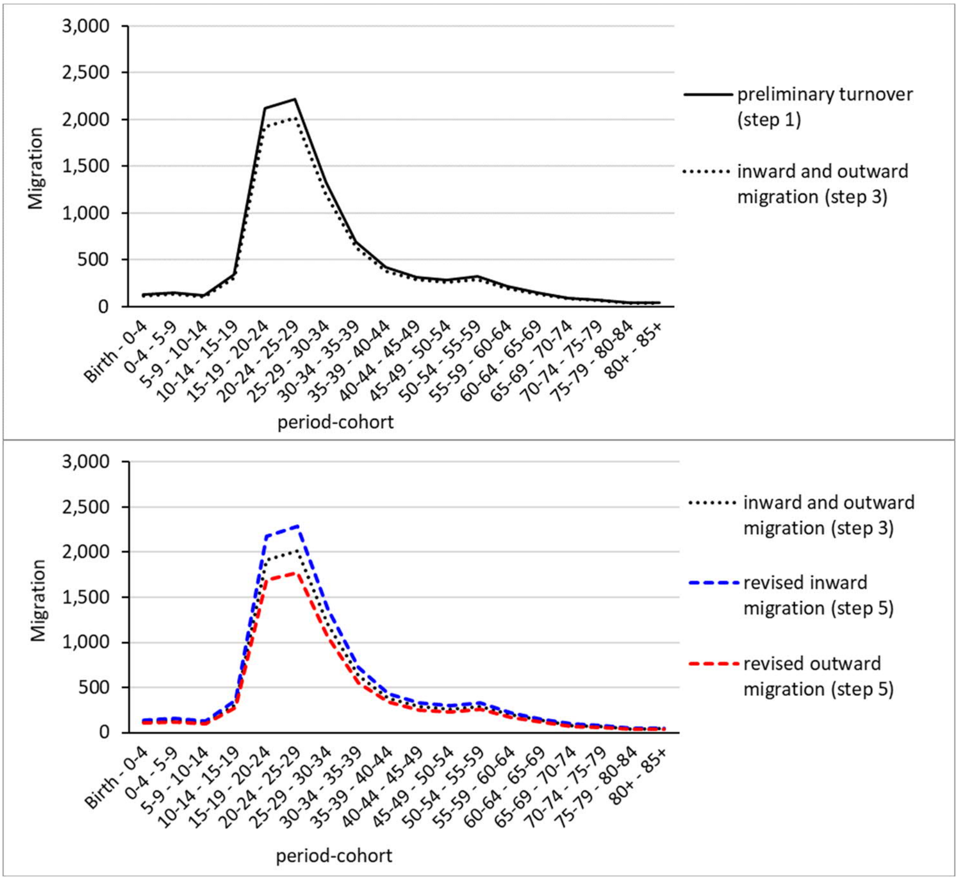
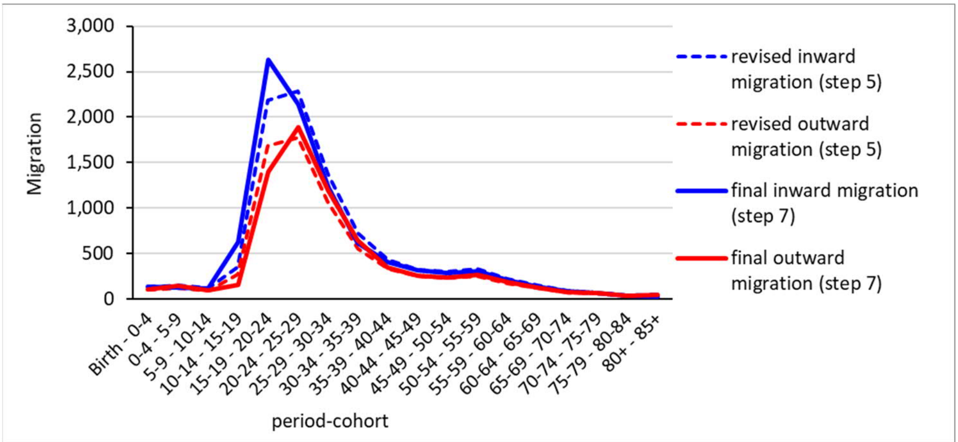


Figure 1: (Continued)



First, preliminary base period migration turnover (inward plus outward migration combined) by sex and period-cohort is prepared by multiplying the model migration rates by the local area populations-at-risk:

$$M^i_{s,a\rightarrow a+5}(t-5,t) = m_{s,a\rightarrow a+5}\ \tfrac{5}{2}\left(P^i_{s,a}(t-5) + P^i_{s,a+5}(t)\right) \tag{7}$$

where:

M = preliminary migration turnover flow
i = local area
m = model migration rate.

The model migration rates were created by fitting a model migration schedule to male and female internal migration between all SA3 areas over the period 2006–2011 using data from the 2011 census. The rates are provided in the Appendix.

The second step estimates the aggregate migration turnover for the local area. Use is made of the crude migration turnover rate, defined as total inward plus outward migration divided by total population (re-named from the original ‘population turnover’ measure introduced by Dennett and Stillwell 2008). The migration turnover rate may be a universal value estimated for all areas or it may be area-specific. Total migration turnover for the local area is calculated as the product of this rate and the total population -at-risk:

$$TO^i(t-5,t) = to^i \ \tfrac{5}{2}\left(P^i(t-5) + P^i(t)\right) \tag{8}$$

where:

to = crude migration turnover rate
TO = migration turnover flow.

Actual total migration turnover could be used instead where such data are available or can easily be estimated. However, in many cases these data will not be available.

Third, preliminary base period migration turnover by sex and period-cohort (from step 1) is scaled to aggregate migration turnover (step 2) to give migration turnover by sex and period-cohort specific to the local area:

$$TO^i_{s,a\to a+5}(t-5,t) = M^i_{s,a\to a+5}(t-5,t) \ \frac{TO^i(t-5,t)}{\sum_s \sum_a M^i_{s,a\to a+5}(t-5,t)} \tag{9}$$

These values are then split in half to give preliminary estimates of inward and outward migration:

$$\begin{aligned} I^i_{s,a\to a+5}(t-5,t) &= \tfrac{1}{2}\,TO^i_{s,a\to a+5}(t-5,t) \\ O^i_{s,a\to a+5}(t-5,t) &= \tfrac{1}{2}\,TO^i_{s,a\to a+5}(t-5,t) \end{aligned} \tag{10}$$

At this stage, inward and outward migration flows are identical (Figure 1, top graph). In the fourth step, total net migration is estimated as an accounting residual:

$$N^i(t-5,t) = P^i(t) - P^i(t-5) + D^i(t-5,t) - B^i(t-5,t) \tag{11}$$

where:

N = base period total net migration.

Fifth, preliminary directional migration is adjusted up and down to create different inward and outward migration flows which are consistent with base period total net migration. The scaling factor is calculated using the quadratic equation:

$$f^i = \frac{N^i + \sqrt{(N^i)^2 + 4\,\frac{1}{2}TO^i\ \frac{1}{2}TO^i}}{2\,\frac{1}{2}TO^i} \tag{12}$$

with time labels omitted for clarity. This is used in preference to the plus–minus method (Smith, Tayman, and Swanson 2013) because it can cope with very large adjustments. Revised inward and outward migration flows by sex and period-cohort are then created from the preliminary flows (step 3) adjusted by the scaling factor:

$$I^{i}_{s,a\rightarrow a+5}(t-5,t)[2] = I^{i}_{s,a\rightarrow a+5}(t-5,t)\ f^{i} \tag{13}$$

$$O^{i}_{s,a\rightarrow a+5}(t-5,t)[2] = \frac{O^{i}_{s,a\rightarrow a+5}(t-5,t)}{f^{i}}$$

The inward and outward migration flows at this stage possess different values but identical age–sex patterns (Figure 1, middle graph).

Sixth, base period net migration by sex and period-cohort is calculated as an accounting residual using population estimates at the start and end of the base period plus the births and deaths just estimated:

$$N^{i}_{s,a\rightarrow a+5}(t-5,t) = P^{i}_{s,a+5}(t) - P^{i}_{s,a}(t-5) + D^{i}_{s,a\rightarrow a+5}(t-5,t) \tag{14}$$

For the infant period-cohort, births replace the start-of-period population. In the projections created for the evaluation presented later, residual net migration was smoothed by averaging over sex up to age 65 in order to reduce noise but retain area-specific migration–age patterns.

In the final seventh step, the revised inward and outward migration flows (from step 5) are constrained to be consistent with net migration by sex and period-cohort (step 6). In this step scaling factors are applied to migration for each sex and period-cohort to proportionally adjust inward and outward migration. The scaling factors are calculated using the quadratic equation:

$$f^{i}_{s,a\rightarrow a+5} = \frac{N^{i}_{s,a\rightarrow a+5} + \sqrt{\left(N^{i}_{s,a\rightarrow a+5}\right)^{2} + 4\ I^{i}_{s,a\rightarrow a+5}[2]\ O^{i}_{s,a\rightarrow a+5}[2]}}{2\ I^{i}_{s,a\rightarrow a+5}[2]} \tag{15}$$

with time labels omitted here to aid clarity. The scaling factors are smoothed over age above age 45 using three-term moving averages to reduce noisy patterns. They are then applied to adjust inward and outward migration:

$$I^{i}_{s,a\rightarrow a+5}(t-5,t)[3] = I^{i}_{s,a\rightarrow a+5}(t-5,t)[2]\ f_{s,a\rightarrow a+5} \tag{16}$$

$$O^{i}_{s,a\rightarrow a+5}(t-5,t)[3] = \frac{O^{i}_{s,a\rightarrow a+5}(t-5,t)\ [2]}{f_{s,a\rightarrow a+5}}.$$

These final synthetic inward and outward migration flows are now consistent with base period net migration patterns (Figure 1, bottom graph).

Outward migration rates for use in the projections are calculated by dividing outward migration by the base period populations-at-risk:

$$o^{i}_{s,a\rightarrow a+5}(t-5,t) = \frac{O^{i}_{s,a\rightarrow a+5}(t-5,t)[2]}{\frac{5}{2}\left(P^{i}_{s,a}(t-5)+P^{i}_{s,a+5}(t)\right)} \qquad (17)$$

while inward migration is used directly in the projections as a migration flow.

2.3 Multiple area version

The multiple area version of the synthetic migration cohort-component model involves the same calculations as the single area version, except for the additional feature of constraining to independent national forecasts (or those of a state or similar larger region). After births by sex have been projected for all local areas, they are constrained by multiplying by the ratio of national projected births to the sum of local projected births:

$$B^{i}_{s}[2] = B^{i}_{s}\ \frac{B^{nat}_{s}}{\sum_{i} B^{i}_{s}} \qquad (18)$$

where:

nat = national.

Similarly, deaths are constrained to national projected deaths by sex and period-cohort:

$$D^{i}_{s,a\rightarrow a+5}[2] = D^{i}_{s,a\rightarrow a+5}\ \frac{D^{nat}_{s,a\rightarrow a+5}}{\sum_{i} D^{i}_{s,a\rightarrow a+5}} \qquad (19)$$

Note that the constraining of local births and deaths to the national projections means that actual local fertility rates and life expectancies will differ slightly from the original assumptions.

Local inward and outward migration flows by sex and period-cohort are then constrained. Iterative proportional fitting is used to constrain local directional migration by sex and period-cohort to (1) local total net migration, and (2) national (or state) net migration by sex and period-cohort. The constraining does not use a conventional iterative proportional fitting algorithm. Instead, for both sets of net migration constraints, initial inward and outward migration flows are adjusted by the same proportional amount so that they are consistent with each net migration constraint. In rare circumstances (usually where outward migration is very high), it is possible for this adjustment to generate negative populations. Where this occurs, the constraining algorithm intervenes and changes inward and outward migration flows to avoid generating negative populations.

3. Forecast evaluation

3.1 Data inputs and forecasts

Population estimates by sex and 5-year age group for SA3 areas of Australia are available on a consistent set of boundaries from 1991 to 2016 (ABS 2021). In 2016 the median SA3 area population was 61,539, with the 95% range spanning 11,600 to 192,073. Making full use of this dataset, four sets of SA3 population 'forecasts' were produced, 1996-based, 2001-based, 2006-based, and 2011-based, with all forecasts extending out to 2016. The accompanying base periods for which population accounts were prepared were, respectively, 1991–1996, 1996–2001, 2001–2006, and 2006–2011. For the purposes of this study, three SA3 areas of New South Wales with very small populations were combined into one, and the SA3 areas which form Other Territories were excluded from the analysis because not all were officially included as part of Australia for the whole study period. In total, forecasts were produced for 328 SA3 areas.

Two types of forecasts were prepared: (1) forecast-constrained, in which local area forecasts were constrained to separate forecasts of local area population totals and national population forecasts by age and sex, and (2) estimate-constrained, in which local area forecasts were constrained to actual local area population totals and national population estimates. The purpose of the first type was to create a set of forecasts as close to a real application as possible; the purpose of the second type was to remove the effect of incorrect constraining forecasts and thus reveal the accuracy of the age–sex forecasting ability of the synthetic migration cohort-component model.

For the forecast-constrained set, local area forecasts of total population were prepared by taking the average of four extrapolative models. Several studies have demonstrated that averaging the outputs of multiple forecast models often yields gains in accuracy over any of the individual models (e.g., Armstrong 2001; Goodwin 2009; Rayer and Smith 2010; Smith and Shahidullah 1995; Wilson 2017). The four models were:

1) a constant share of population model in which each small area's share of the national population at the jump-off year is held constant;

2) a linear/exponential model in which population is forecast using linear extrapolation if population change over the previous decade was positive, and an exponential model if it was negative;
3) a variable share of growth model in which preliminary projected population growth from the linear/exponential model is adjusted to match national projected growth;
4) a modified exponential model (Baker et al. 2008) in which ceiling and floor limits to projected populations are imposed.

The forecasts from models (2) and (4) were constrained to sum to national projected growth; models (1) and (3) produce forecasts consistent with the national projections as part of their calculations. Population growth rates used in models (2) to (4) were based on the decade prior to the jump-off year.

For all SA3 areas, the simplest assumptions were selected for the cohort-component forecasts. Base period TFRs, calculated using the xTFR estimation method, were assumed to remain unchanged into the future. Age-specific fertility rates were calculated by scaling base period national fertility rates to match each area's TFR. National forecast mortality rates were used for all areas. For all areas, a crude migration turnover rate of 0.1744 was assumed, which was the approximate migration turnover rate for all SA3 areas estimated from 2011 census migration data. Base period inward migration and outward migration rates for each area were applied in all forecast intervals. All forecasts were constrained to national population forecasts or estimates.

National population forecasts were prepared which mimicked the official Australian Bureau of Statistics (ABS) medium series projections of the time. These projections used the same TFR, life expectancy at birth, and overseas migration assumptions as assumed by the ABS at the jump-off year of each of the forecasts. It was not possible to use actual ABS projections because of the lack of projected demographic components by sex and period-cohort in some of the projection publications, and differences in jump-off populations due to a major revision of the 1991–2011 population estimates time series following the 2011 census (ABS 2013).

To compare the synthetic migration cohort-component model with other models which do not require local area migration data, SA3 area population forecasts by age and sex were also prepared using two types of Hamilton–Perry model (Hamilton and Perry 1962; Baker et al. 2017). The first comparative model is the standard Hamilton–Perry model using Cohort Change Ratios (denoted HP CCR). The second is an alternative which uses CCRs when base period cohort population change is negative and CCDs when the population change is positive (HP CCR/CCD). Smoothing of CCR and CCD age profiles occurred in both Hamilton–Perry models. Populations aged 0–4 were created via

Child/Woman Ratios. The outputs of both models were constrained to SA3 total populations and national population forecasts or estimates by age and sex.

3.2 Error measures

Errors for total populations and individual age–sex group forecasts were measured using Absolute Percentage Error (APE):

$$APE = \frac{|F - A|}{A} 100$$

where F denotes forecast and A the actual population. Median, 90th percentile, and mean values of APE are reported.

To summarise forecast errors across age–sex groups of a population in a single metric, an alternative Absolute Percentage Error measure for total population was used. The conventional APE for total population does not account for the quality of a forecast by age and sex. It could be close to zero – suggesting an accurate forecast – in situations where there are large offsetting age–sex errors, such as substantial over-forecasts of younger age groups and severe under-forecasts of older age groups. To overcome this, an alternative measure, $APE_{age-sex}$, was calculated by summing absolute errors for each age–sex group and then dividing by the total actual population; i.e.,

$$APE_{age-sex} = \frac{\sum_s \sum_a |F_{s,a} - A_{s,a}|}{A} 100.$$

Graphically, the numerator is the absolute difference in area between the forecast population pyramid and the actual population pyramid. $APE_{age-sex}$ is equivalent to Weighted Mean Absolute Percentage Error and the Mean Absolute Deviation/mean ratio (Kolassa and Schutz 2007) but is calculated differently so that it can accommodate actual age–sex populations of zero, which can occur in small populations. If all age–sex forecast errors take the same sign (all positive or all negative), then the alternative $APE_{age-sex}$ will be the same as the conventional total population APE. If, however, there is a mix of over- and under-forecasts across age–sex populations, then $APE_{age-sex}$ will be greater, indicating a less successful forecast in terms of age–sex structure. Note that the value of $APE_{age-sex}$ is dependent on the number of age–sex groups and their width in age. Its value calculated using single-year age–sex groups will differ from that using 5-year age–sex groups. All comparisons between populations should use $APE_{age-sex}$ calculated with the same age–sex population breakdown. Median, 90th percentile, and mean values of this measure across local areas are reported in this paper.

4. Results

4.1 Total population forecast errors

The national and SA3 area total population forecast errors are briefly reported first, because they influence the age–sex forecast errors. The SA3 total populations in the estimate-constrained forecasts have errors of zero, by definition. Errors in predicting total population from the average of four extrapolative models used in the forecast-constrained set of forecasts are summarised in Table 2. Also included in the table are Absolute Percentage Errors of the national population forecasts. Errors in forecasting the national population are quite low except for the 2001-based forecast, which was due to under-projections of both the fertility rate and net overseas migration. For the SA3 area forecasts, median APEs are 2%–3% after 5 years, 4%–6% after 10 years, and 7%–9% after 15 years, though, as the 90th percentile APEs show, the upper end of the distribution contains errors which are quite high.

Table 2: Total population errors in the forecast-constrained set of forecasts

	Forecast series			
Forecast horizon	**1996-based**	**2001-based**	**2006-based**	**2011-based**
	SA3 area projections *Median APE (%)*			
5 years	2.4	2.7	2.1	2.7
10 years	4.4	5.8	4.0	
15 years	7.4	8.8		
	90th percentile APE (%)			
5 years	8.3	9.0	6.6	8.1
10 years	15.2	15.3	11.8	
15 years	21.3	22.2		
	MAPE (%)			
5 years	3.7	4.0	3.3	4.0
10 years	6.6	7.7	6.0	
15 years	10.0	11.2		
	National forecast *APE (%)*			
5 years	0.1	1.2	1.3	1.3
10 years	1.2	5.9	2.2	
15 years	5.5	9.9		

4.2 Age–sex population forecast errors

Table 3 summarises the age–sex population forecast errors of the synthetic migration cohort-component model and the two implementations of the Hamilton–Perry model as measured by $APE_{age\text{-}sex}$. The lowest errors for each set of forecasts by forecast horizon are indicated by shading. The main finding is that average errors from the synthetic model are the lowest for all forecasts evaluated, whether measured as median or mean errors. The percentage point reductions in error between HP CCR and the synthetic model increase with forecast horizon, giving modest reductions in median error of up to 1 percentage point after 5 years but larger reductions of 1–3 percentage points after 15 years. Reductions in mean error are generally larger than median error, with the greatest being a 4.7 percentage point improvement after 15 years (for the 2001-based estimate-constrained forecast).

As would be expected, the estimate-constrained forecasts, which use actual SA3 total populations and actual national populations, produce more accurate age-sex forecasts. In addition, the reduction in error from the Hamilton–Perry models is greater than for the forecast-constrained forecasts. It demonstrates that if national forecasts and local area total populations are forecast well, then the synthetic model delivers quite decent improvements in accuracy over the Hamilton–Perry models. Errors from the synthetic model's forecasts also have a slightly narrower error distribution than the Hamilton–Perry models. The 90^{th} percentile of the $APE_{age\text{-}sex}$ distribution is lower for all but one of the forecasts, and there is a greater reduction in 90^{th} percentile error values in the estimate-constrained forecasts.

For each age–sex group, Absolute Percentage Errors were also calculated. Mostly – but certainly not for every age group – the average synthetic model errors are lower. Figure 2 illustrates the Median Absolute Percentage Errors across the 1996-based and 2001-based forecasts combined for the female population at a forecast horizon of 15 years. Errors are shown for both the projection-constrained forecasts (Figure 2, part a) and the estimate-constrained forecasts (Figure 2, part b). The high errors in the childhood ages in the projection-constrained forecasts are due to large errors in the assumed fertility rates of the 2001-based national forecasts. The only difference between the two types of forecast is the constraining populations. The difference in errors evident between the two graphs is therefore the result of errors in the local total population and national population constraints.

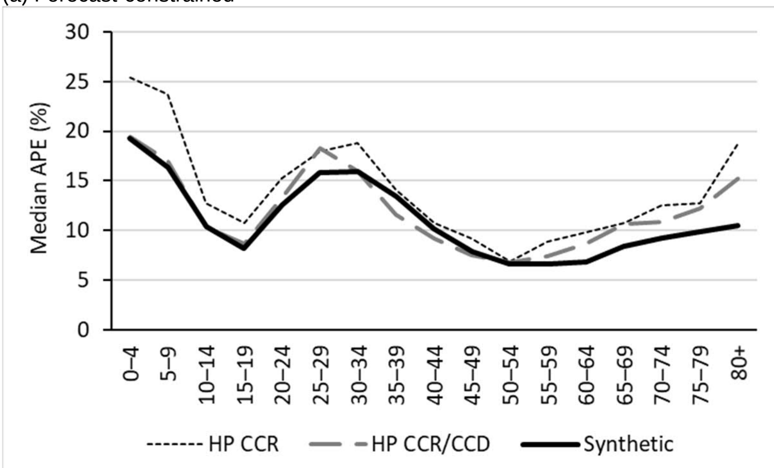
Table 3: Age–sex population forecast errors by projection model measured by $APE_{age\text{-}sex}$

		Forecast-constrained				Estimate-constrained			
Forecast horizon	**Model**	**1996-based**	**2001-based**	**2006-based**	**2011-based**	**1996-based**	**2001-based**	**2006-based**	**2011-based**
		Median $APE_{age\text{-}sex}$ (%)							
5 years	HP CCR	6.1	6.1	4.8	5.0	5.0	4.8	4.0	3.9
	HP CCR/CCD	6.0	5.7	4.4	4.8	4.8	4.5	3.7	3.7
	Synthetic	5.7	5.3	4.4	4.7	4.2	4.0	3.4	3.3
10 years	HP CCR	9.3	11.9	7.7		7.6	8.4	6.0	
	HP CCR/CCD	9.0	10.5	7.2		7.3	7.5	5.4	
	Synthetic	8.5	9.5	7.1		6.1	6.0	5.1	
15 years	HP CCR	13.6	15.3			10.3	10.6		
	HP CCR/CCD	12.5	13.8			9.6	9.3		
	Synthetic	11.9	12.6			7.7	7.7		
		90th percentile $APE_{age\text{-}sex}$ (%)							
5 years	HP CCR	15.1	14.4	9.6	9.8	11.0	10.2	7.4	7.1
	HP CCR/CCD	14.5	12.7	9.0	9.4	10.4	9.1	6.8	6.6
	Synthetic	14.3	11.1	8.7	9.0	10.0	6.8	5.8	5.5
10 years	HP CCR	20.7	24.5	14.8		16.1	17.9	11.2	
	HP CCR/CCD	18.5	20.7	14.0		14.2	15.0	10.4	
	Synthetic	18.8	19.1	13.6		11.9	11.3	8.5	
15 years	HP CCR	26.3	32.8			19.0	23.1		
	HP CCR/CCD	24.8	30.1			17.3	19.8		
	Synthetic	23.6	26.2			14.3	15.0		
		Mean $APE_{age\text{-}sex}$ (%)							
5 years	HP CCR	8.0	8.0	5.9	6.3	6.7	6.3	4.9	4.9
	HP CCR/CCD	7.6	7.4	5.6	6.0	6.2	5.6	4.5	4.4
	Synthetic	7.4	6.9	5.5	5.9	5.6	4.7	4.1	4.0
10 years	HP CCR	12.0	14.0	9.6		9.7	10.5	7.5	
	HP CCR/CCD	11.2	12.7	9.1		8.7	9.1	6.8	
	Synthetic	10.8	11.7	8.8		7.5	7.2	5.8	
15 years	HP CCR	16.1	18.6			12.5	13.4		
	HP CCR/CCD	14.8	16.9			10.9	11.4		
	Synthetic	14.2	15.5			9.1	8.8		

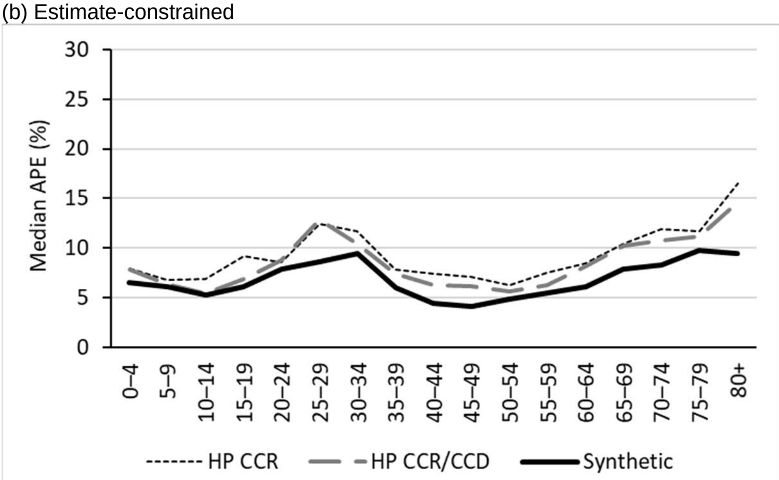
Note. The lowest errors are indicated by the shaded cells.

Figure 2: Median Absolute Percentage Errors for female age-specific SA3 area population forecasts at a forecast horizon of 15 years (1996-based and 2001-based forecasts combined)

(a) Forecast-constrained



(b) Estimate-constrained



To provide an indication of the relative accuracy of forecasts by area classifications typically used in forecast evaluation studies, median $APE_{age\text{-}sex}$ errors are presented for areas classified by (1) broad urban/rural region type – either within a state capital city metropolitan region or in the rest of each state, (2) base period growth rate, and (3) jump-off population size. Errors are shown for the 1996-based estimate-constrained forecasts only to keep the table size manageable. (Error patterns for the other forecasts which extend out 15 years are similar, with the synthetic model's advantage being greater for the estimate-constrained forecasts.)

Table 4 presents the errors. For each area classification and forecast horizon, the synthetic model gives the lowest errors. Local areas located within the rest of state regions tend to experience lower forecast errors than those within capital cities, with the forecast error differences between the two broad region types smaller for the synthetic model. Forecast errors by base period growth rate exhibit a u-shape pattern for the Hamilton–Perry models, in common with many other studies (Tayman 2011), but not for the synthetic model. Errors by jump-off population size category show the expected decline in error with increasing population size (Wilson et al. 2018).

Table 4: Age–sex population forecast errors measured by median $APE_{age\text{-}sex}$ for the 1996-based estimate-constrained forecasts for selected categories of area

		Broad region		Base period growth				Jump-off population		
Forecast horizon	Model	Capital city	Rest of State	<0% p.a.	0–1% p.a.	1–2% p.a.	2+% p.a.	<30,000	30,000–59,999	60,000+
				Median $APE_{age\text{-}sex}$ (%)						
5 years	HP CCR	5.2	4.7	4.7	4.5	5.1	6.5	6.2	4.9	4.4
	HP CCR/CCD	5.1	4.5	4.6	4.3	4.8	5.7	5.6	4.6	4.1
	Synthetic	4.2	4.0	3.9	3.9	4.8	4.5	4.8	3.9	3.4
10 years	HP CCR	8.7	6.6	7.7	6.4	7.3	9.8	10.2	7.0	7.1
	HP CCR/CCD	8.0	6.6	7.2	6.4	7.1	8.0	8.4	7.0	6.6
	Synthetic	6.3	5.8	6.2	5.6	6.2	6.2	6.9	5.9	5.2
15 years	HP CCR	11.3	8.7	10.4	8.6	9.8	13.5	14.0	9.6	9.1
	HP CCR/CCD	10.4	8.6	10.3	8.3	9.7	11.2	11.2	9.1	8.5
	Synthetic	7.8	7.4	8.1	7.3	8.2	7.4	9.5	7.8	6.6
				90th percentile $APE_{age\text{-}sex}$ (%)						
5 years	HP CCR	13.6	10.1	10.5	9.1	11.7	12.0	19.2	9.3	9.2
	HP CCR/CCD	12.1	8.9	10.4	8.4	10.5	10.5	13.6	9.0	9.0
	Synthetic	12.0	8.0	9.7	8.5	10.7	10.2	12.7	8.2	8.7
10 years	HP CCR	17.1	13.9	18.1	12.2	16.8	16.5	23.5	13.9	13.3
	HP CCR/CCD	15.7	12.1	16.4	11.4	13.2	15.3	16.9	12.3	12.7
	Synthetic	13.6	10.3	14.2	10.6	11.3	13.0	14.7	10.4	11.8
15 years	HP CCR	19.1	18.6	22.5	16.0	18.3	19.4	27.8	18.2	9.1
	HP CCR/CCD	18.4	15.9	22.8	15.2	16.1	17.0	21.7	15.4	8.5
	Synthetic	15.4	12.6	16.0	13.7	12.9	13.3	16.2	13.4	6.6
				Mean $APE_{age\text{-}sex}$ (%)						
5 years	HP CCR	7.3	5.9	6.9	5.7	6.7	7.7	9.1	6.2	5.6
	HP CCR/CCD	6.7	5.5	6.4	5.6	6.3	6.7	7.8	5.9	5.4
	Synthetic	6.1	4.9	5.8	5.3	5.7	5.8	6.9	5.3	5.1
10 years	HP CCR	10.6	8.4	10.9	7.8	9.6	11.0	13.5	8.8	7.9
	HP CCR/CCD	9.5	7.6	9.8	7.5	8.6	9.2	11.0	8.2	7.5
	Synthetic	8.2	6.7	8.6	6.9	7.3	7.5	9.3	7.2	6.6
15 years	HP CCR	13.5	11.2	14.1	9.9	13.0	13.9	17.4	11.4	10.3
	HP CCR/CCD	11.7	9.9	12.7	9.4	10.7	11.5	13.4	10.5	9.7
	Synthetic	9.5	8.5	10.7	8.5	8.9	8.5	11.0	9.0	7.9

Note. The lowest errors are indicated by the shaded cells.

4.3 Sensitivity of the forecasts to alternative migration turnover assumptions

To test the sensitivity of the synthetic migration model to alternative migration turnover assumptions, two additional sets of forecasts were produced, one with a 50% reduction in the migration turnover rate used earlier, and the other with a 50% increase in the rate. Table 5 presents the forecast errors alongside the 'standard' assumptions used for the main set of forecasts reported in Table 3. For such large changes to the migration turnover rate, the differences in forecast accuracy are small. According to the median age–sex APE, the lower migration turnover assumption generated slightly more accurate forecasts overall, but if measured by the mean age–sex APE, the lower and standard assumptions were more evenly matched. The higher migration turnover assumption proved marginally less accurate in terms of mean or median error. The main point is that the population forecasts are quite insensitive to migration turnover assumptions. In practical applications, therefore, when very little migration data are available, even a rough ‘guesstimate’ of the migration turnover rate should be sufficient to generate reasonable forecasts.

4.4 Incorporating local assumptions

The effect of using local (rather than national) life expectancy at birth by sex was evaluated by creating alternative 2006-based estimate-constrained forecasts. The 2006 jump-off year was selected because of the availability of SA3 area deaths data by age and sex for the 2001–2006 base period. Abridged life tables for SA3 areas over the 2001–2006 period were prepared. Local differences with national life expectancy were calculated, and then projected life expectancy for each SA3 area was assumed to be national projected life expectancy plus or minus the 2001–2006 difference. In this example, the impact on forecast error after 10 years proved to be negligible. The median SA3 $APE_{age\text{-}sex}$ in 2016 using local life expectancy values only changed from 5.12% to 5.11% (the mean value remained unchanged), with the median absolute difference in SA3 $APE_{age\text{-}sex}$ between the two sets of forecasts being just 0.03 percentage points.

Table 5: Sensitivity of age–sex population forecast errors to alternative migration turnover assumptions

		Forecast-constrained				**Estimate-constrained**			
Forecast horizon	**Migration turnover assumption**	**1996-based**	**2001-based**	**2006-based**	**2011-based**	**1996-based**	**2001-based**	**2006-based**	**2011-based**
		Median $APE_{age\text{-}sex}$ (%)							
5 years	50% lower	5.6	5.4	4.4	4.6	4.2	4.1	3.4	3.3
	Standard	5.7	5.2	4.4	4.7	4.1	4.0	3.4	3.3
	50% higher	5.8	5.3	4.6	4.8	4.3	4.0	3.6	3.5
10 years	50% lower	8.4	9.7	6.9		6.0	6.0	5.0	
	Standard	8.5	9.5	7.1		6.1	6.0	5.1	
	50% higher	8.9	9.5	7.4		6.4	6.3	5.6	
15 years	50% lower	11.6	12.5			7.4	7.5		
	Standard	11.8	12.6			7.7	7.7		
	50% higher	12.3	12.6			8.3	8.0		
		90th percentile $APE_{age\text{-}sex}$ (%)							
5 years	50% lower	14.4	11.9	8.8	9.0	10.1	7.6	6.1	5.7
	Standard	14.3	11.1	8.7	9.0	10.0	6.8	5.8	5.5
	50% higher	14.2	10.8	8.7	9.1	9.8	6.9	5.9	5.8
10 years	50% lower	18.5	19.8	13.7		12.5	12.3	8.6	
	Standard	18.8	19.1	13.6		11.9	11.3	8.5	
	50% higher	18.3	19.1	13.8		12.2	10.9	8.7	
15 years	50% lower	23.7	27.5			14.4	15.1		
	Standard	23.6	26.2			14.3	15.0		
	50% higher	23.9	26.1			15.1	14.4		
		Mean APEage-sex (%)							
5 years	50% lower	7.4	7.0	5.5	5.8	5.6	4.9	4.1	4.0
	Standard	7.4	6.9	5.5	5.9	5.6	4.7	4.1	4.0
	50% higher	7.5	6.9	5.7	6.0	5.7	4.8	4.3	4.2
10 years	50% lower	10.7	11.9	8.7		7.5	7.4	5.8	
	Standard	10.8	11.7	8.8		7.5	7.2	5.8	
	50% higher	11.0	11.7	9.0		7.9	7.3	6.1	
15 years	50% lower	14.0	15.7			8.8	9.0		
	Standard	14.2	15.5			9.1	8.8		
	50% higher	14.6	15.5			9.6	8.9		

Note: The lowest errors are indicated by the shaded cells.

A possible explanation for this minimal difference in error is the fact that there are minor differences in the disaggregation of cohort change into mortality and net migration between the local and national mortality assumptions in the base period population accounts. In the forecasts using national mortality rates there is a mix of slight over- and under-estimation of mortality across local areas. However, remaining cohort change is assumed to be net migration, so that even though mortality and net migration are imperfectly estimated, cohort change is fully accounted for. Using local mortality assumptions gives better mortality estimates in the base period accounts, which also results in better estimates of net migration. But the amount of cohort change in the base period remains the same. In most cases this leads to similar forecasts.

To evaluate the impact of using locally varying crude migration turnover rates, SA3-specific rates were estimated using ABS 2011 census data. Census counts of 1-year-interval internal in-migration, out-migration, and immigration were obtained. The values of in-migration, out-migration, immigration and then half of immigration again (as a proxy for emigration) were combined to give an estimate of migration turnover, and then divided by SA3 census populations to create a rate. Although there are several conceptual and empirical approximations inherent in this approach, it is simple to implement and yields much more realistic rates than simply using a universal rate for all local areas. The estimated migration turnover rates were then used to generate alternative 2006-based estimate-constrained forecasts. Ideally, migration turnover rates would have been estimated from the 2006 census, but the statistical geography prior to 2011 was radically different and did not include SA3 areas.

The use of local migration turnover assumptions had little effect on overall error. At a forecast horizon of 10 years, the median SA3 $APE_{age\text{-}sex}$ declined only marginally, from 5.12% to 5.06% (the mean $APE_{age\text{-}sex}$ declined from 5.82% to 5.71%). The median absolute difference in individual SA3 $APE_{age\text{-}sex}$ values between the two sets of forecasts was 0.19 percentage points. Yet many SA3-specific migration turnover rates varied considerably from the national average, with the median absolute relative change in migration turnover rates from the national average assumption being 27%. Despite sizeable changes to migration turnover rates, the overall effect on the forecasts, and forecast error, was small.

However, for a select few areas with very high migration turnover rates there were moderate error reductions from using local migration assumptions. Notable improvements occurred for the inner city SA3 areas of Sydney Inner City (with $APE_{age\text{-}sex}$ declining from 10.0% to 6.0%), Adelaide City (from 16.4% to 11.5%), and two rapidly growing SA3 areas in Canberra with very small populations. Population pyramids for the Sydney and Adelaide SA3 areas at a forecast horizon of 10 years are shown in Figure 3. The light blue bars indicate forecasts generated using local migration turnover rates, the dotted outline bars represent forecasts using national average turnover rates, and actual

2016 population estimates are shown by the black outline bars. As the graph shows, the Sydney forecast is reasonably accurate; the Adelaide forecast is not quite as good (especially at ages 20–24), but the use of local migration turnover rates improves the forecast.

Figure 3: Population forecasts for Sydney Inner City and Adelaide City SA3 areas in 2016 (2006-based estimate-constrained forecasts)

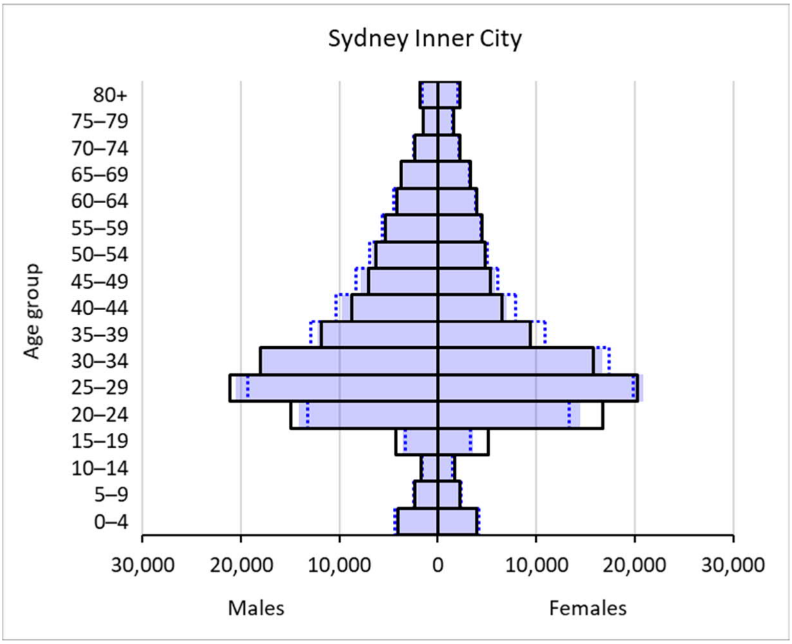
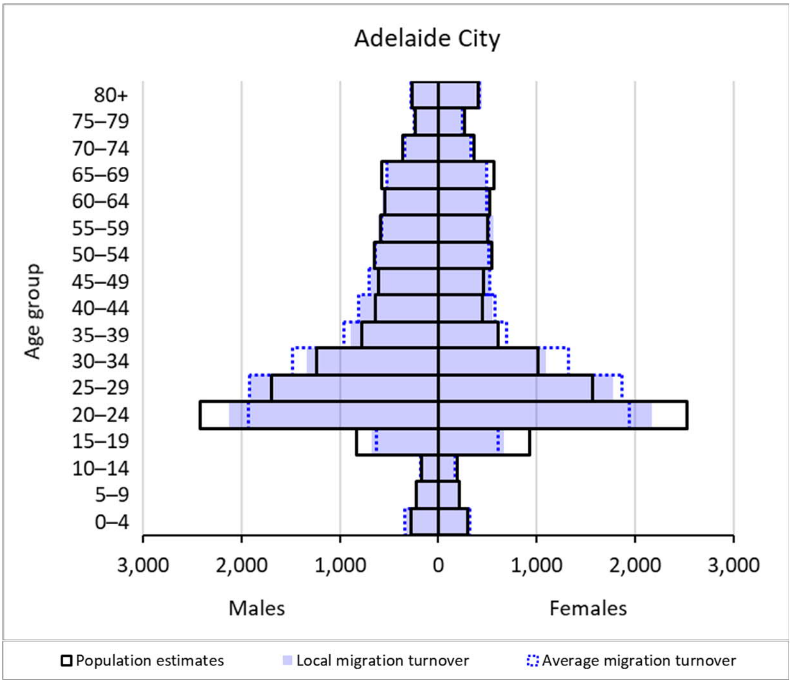


Figure 3: (Continued)



5. Discussion and conclusions

5.1 Main contributions

This paper has described the synthetic migration cohort-component model – an approach to creating local area population forecasts by age and sex using a bi-regional cohort-component model but without the need for locally specific migration, fertility, or mortality input data. It combines the conceptual and empirical strengths of the bi-regional cohort-component model and the important data on net migration age patterns contained in each local area's base period population accounts. It offers the flexibility of setting area-specific mortality and migration assumptions, but still produces good quality forecasts without them. It was shown that modest improvements in accuracy for some inner city and rapidly growing areas can be obtained by estimating local crude migration turnover rates.

Importantly, the model can be applied to countries without local migration data. It can also be used in situations where such data does exist but is unreliable, very noisy, or prohibitively expensive, or where local migration data are not available for the selected geographical units under consideration. Due to the low data requirements, local area age–sex population forecasts can be prepared quickly, easily, and cheaply. An Excel workbook which incorporates the multiple-area version of the model and contains illustrative local area population forecasts for Australia is available for download at https://doi.org/10.6084/m9.figshare.19372784.v1. This version incorporates minor modifications based on user feedback (including an extension to age 85+ and a few cosmetic changes) and therefore differs very slightly from the workbooks used for this paper.

When applied to create 'forecasts' for SA3 local areas of Australia over past periods, the synthetic migration model was shown to produce more accurate forecasts than the comparison Hamilton–Perry models. This accuracy advantage increases with forecast horizon length. For short forecast horizons of 5 years there are only modest accuracy differences between the synthetic migration and Hamilton–Perry models, but for horizons of 15 years the differences are sizeable, especially if total populations are forecast accurately (see Table 2, estimate-constrained forecasts).

5.2 Why does the synthetic migration model work as well as it does?

The forecast tests reported above demonstrate that, overall, the synthetic migration model generated more accurate forecasts than the two comparison Hamilton–Perry models. But why is this the case, given that the synthetic migration assumptions at first glance appear to be quite approximate? In essence, it is because the synthetic migration model is a little better at forecasting varying cohort population growth over time than models using CCRs or CCDs. The key to this outcome is the creation of directional migration estimates which are consistent with cohort-specific base period population accounts for the local area. The synthetic inward and outward migration estimates are not 'true' migration flows, but they comprise realistic migration age profiles which are consistent with the base period net migration age–sex pattern. In the forecasts, the model generates age patterns of inward, outward, and net migration flows which vary over time according to the area's forecast assumptions and changing population age structure.

The total population constraints in the synthetic migration model are applied by adjusting migration only (with no adjustments made to fertility or mortality rates). The preliminary migration projections are adjusted so that the cohort-component population forecasts sum over age and sex to the independent total population forecast. The inward migration age profile is shifted up or down as much as necessary. Outward migration forecasts, projected as a function of outward migration rates, change according to the size of populations-at-risk. Adjustments to preliminary migration forecasts will generally be greatest in the younger adult age groups and smallest at the oldest ages, which reflects the real pattern of population change in most cases. The effective CCR age pattern will therefore undergo some modification. For the Hamilton–Perry forecasts prepared for this study, constraining to the total population forecast was achieved by adjusting projected populations across all age groups by the same factor, effectively scaling all CCRs by the same amount. (An alternative approach in which cohort population change was constrained instead did not improve the Hamilton–Perry forecasts).

5.3 Limitations of the synthetic migration model

Although the synthetic migration model possesses several useful features, it is important to note its limitations. The projected synthetic inward and outward migration flows form an important part of the projection calculation process but cannot be regarded as reliable directional migration forecasts in their own right. Net migration values should be reported instead. There are also several simplifications in the directional migration modelling. International and internal migration are not modelled or output separately; and while outward migration is projected via rates and an origin population-at-risk, inward migration is not. Furthermore, migration adjustments made to achieve consistency with independent total population forecasts occur primarily through the adjustment of inward migration flows. All these features are designed to simplify the modelling and minimise data inputs, but the trade-off is some inevitable approximation of reality. Finally, like all demographic models, if trends and patterns change radically from those of the base period, forecasts using base period rates are likely to be inaccurate.

5.4 Limitations of the study

Limitations of the study include the restriction of the forecast evaluation to one type of local area (SA3s) in one country over a period of just two decades. The model would benefit from being tested in other countries. The synthetic migration model forecasts prepared for this study were also not compared to those of a fully multi-regional model or bi-regional model using recorded local area directional migration data. This was due to data limitations. The necessary internal migration data for the selected statistical geography are only available from one census (2011 census), and sub-state emigration estimates have only been available since 2016 (ABS 2018). In addition, the smoothing techniques used in creating the synthetic migration input data are quite basic, and would probably benefit from some enhancement.

5.5 Next steps

The obvious next step is to evaluate the model in other countries under a variety of demographic environments to ascertain whether it offers similar accuracy gains as in Australia. Finessing of the synthetic migration estimation process would also be useful. The current methods used to smooth the migration-age profile adjustment factors and the base period net migration age patterns are rudimentary. The smoothing outcomes could probably be improved by employing the latest statistical methods, which vary the amount of smoothing according to need – more for smaller populations where there is considerable noise, and less for larger populations. It would also be useful to determine whether the synthetic migration estimates could be improved by extending the temporal and spatial extent of the estimation process; e.g., using the average results of two 5-year base periods, and incorporating data from neighbouring areas. Furthermore, it would be worth investigating whether the model could provide any more detail in the outputs. Are there ways of creating a single-year-of-age version without compromising too much on the simplicity and low data requirements of the 5-year age group model? A single-year-of-age breakdown would also provide projections in single-year time intervals. But alternative ways of obtaining annual projection numbers by interpolating the five yearly projections could also be tested.

6. Acknowledgements

This research was supported by the Australian Research Council’s Discovery Projects scheme (project DP200101480).

References

Armstrong, J.S. (2001). Combining forecasts. In: Armstrong, J.S. (ed.) *Principles of forecasting: A handbook for researchers and practitioners*. Boston, MA: Springer: 417–439. doi:10.1007/978-0-306-47630-3_19.

Australian Bureau of Statistics (ABS) (2013). Australian Demographic Statistics December 2012; Feature article 2: Recasting 20 years of ERP. Canberra: Australian Bureau of Statistics. https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/3101.0Feature%20Article2Dec%202012?opendocument&tabname=Summary&prodno=3101.0&issue=Dec%202012&num=&view.

Australian Bureau of Statistics (ABS) (2018). Regional population growth, Australia, 2016–17. Canberra: Australian Bureau of Statistics. https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3218.0Main+Features12016-17?OpenDocument.

Australian Bureau of Statistics (ABS) (2021). Estimated resident population. Canberra: Australian Bureau of Statistics. https://explore.data.abs.gov.au/.

Australian Electoral Commission (AEC) (2021). Victoria small area population and enrolment projections. Canberra: Australian Electoral Commission. https://www.aec.gov.au/Electorates/Redistributions/2021/vic/files/victoria-small-area-population-and-enrolment-populations.pdf.

Baker, J., Ruan, X., Alcantara, A., Jones, T., Watkins, K., McDaniel, M., Frey, M., Crouse, N., Rajbhandari, R., Morehouse, J., Sanchez, J., Inglis, M., Baros, S., Penman, S., Morrison, S., Budge, T., and Stallcup, W. (2008). Density-dependence in urban housing unit growth: An evaluation of the Pearl-Reed model for predicting housing unit stock at the census tract level. *Journal of Economic and Social Measurement* 33(2–3): 155–163. doi:10.3233/JEM-2008-0301.

Baker, J., Swanson, D.A., Tayman, J., and Tedrow, L.M. (2017). *Cohort change ratios and their applications.* Cham, Switzerland: Springer. doi:10.1007/978-3-319-53745-0.

Baker, J., Swanson, D., and Tayman, J. (2021). The accuracy of Hamilton–Perry population projections for census tracts in the United States. *Population Research and Policy Review* 40(6): 1341–1354. doi:10.1007/s11113-020-09601-y.

Bell, M. and Rees, P. (2006). Comparing migration in Britain and Australia: Harmonisation through use of age-time plans. *Environment and Planning A: Economy and Space* 38(5): 959–988. doi:10.1068/a35245.

Dennett, A. and Stillwell, J. (2008). Population turnover and churn: Enhancing understanding of internal migration in Britain through measures of stability. *Population Trends* 134: 24–41.

Foss, W. (2002). Small area population forecasting. *The Appraisal Journal* 70(2): 163–172.

Goodwin, P. (2009). New evidence on the value of combining forecasts. *Foresight: The International Journal of Applied Forecasting* 12: 33–35.

Grossman I, Bandara K, Wilson T, and Kirley M (2022) Can machine learning improve small area population forecasts? A forecast combination approach. *Computers, Environment and Urban Systems* 95: 101806. doi:10.1016/j.compenvurbsys.2022.101806.

Hamilton, C.H. and Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces* 41(2): 163–170. doi:10.2307/2573607.

Hauer, M.E. (2019). Population projections for US counties by age, sex, and race controlled to shared socioeconomic pathway. *Scientific Data* 6(1): 1–15. doi:10.1038/sdata.2019.5.

Hauer, M.E., and Schmertmann, C.P. (2020). Population pyramids yield accurate estimates of total fertility rates. *Demography* 57(1): 221–241. doi:10.1007/s13524-019-00842-x.

Inoue, T. (2017). A new method for estimating small area demographics and its application to long-term population projection. In: Swanson D.A. (ed.): *The frontiers of applied demography*. Cham: Springer: 473–489. doi:10.1007/978-3-319-43329-5_22.

Isserman, A.M. (1993). The right people, the right rates: Making population estimates and forecasts with an interregional cohort-component model. *Journal of the American Planning Association* 59(1): 45–64. doi:10.1080/01944369308975844.

Kolassa, S. and Schutz, W. (2007). Advantages of the MAD/mean ratio over the MAPE. *Foresight* 6: 40–43.

Pittenger, D.B. (1976). *Projecting state and local populations*. Cambridge, Mass: Ballinger.

Pittenger, D.B. (1978). On making flexible projections of age-specific net migration. *Environment and Planning A: Economy and Space* 10(11): 1253–1272. doi:10.1068/a101253.

Rayer, S. and Smith, S.K. (2010). Factors affecting the accuracy of subcounty population forecasts. *Journal of Planning Education and Research* 30(2): 147–161. doi:10.1177/0739456X10380056.

Rees, P. (1984). Spatial population analysis using movement data and accounting methods: theory, models, the 'MOVE' program and examples. Leeds: School of Geography, University of Leeds. (Working paper 404).

Rees, P. (1985). Does it really matter which migration data you use in a population model? In: van der Knapp, G.A. and White, P.E. (eds.) *Contemporary studies of migration*. Norwich: GeoBooks: 55–77.

Rees, P. and Willekens, F. (1986). Data and accounts. In: Rogers, A. and Willekens, F.J. (eds.) *Migration and settlement: A multiregional comparative study*. Dordrecht: D. Reidel: 19–58.

Rees, P. and Wilson, A.G. (1977). *Spatial population analysis*. London: Edward Arnold.

Reinhold, M. and Thomsen, S.L. (2015). Subnational population projections by age: An evaluation of combined forecast techniques. *Population Research and Policy Review* 34(4): 593–613. doi:10.1007/s11113-015-9362-0.

Rogers, A. (1976). Shrinking large-scale population-projection models by aggregation and decomposition. *Environment and Planning A: Economy and Space* 8(5): 515–541. doi:10.1068/a080515.

Rogers, A. (1995). *Multiregional demography: Principles, methods and extensions*. Chichester: Wiley.

Rowland, D.T. (2003). *Demographic methods and concepts*. Oxford: Oxford University Press.

Simpson, L. and Snowling, H. (2011). Estimation of local demographic variation in a flexible framework for population projections. *Journal of Population Research* 28(2): 109–127. doi:10.1007/s12546-011-9060-7.

Smith, S.K. (1986). Accounting for migration in cohort-component projections of state and local populations. *Demography* 23(1): 127–135. doi:10.2307/2061413.

Smith, S.K. and Tayman, J. (2003). An evaluation of population projections by age. *Demography* 40(4): 741–757. doi:10.1353/dem.2003.0041.

Smith, S.K. and Shahidullah, M. (1995). An evaluation of population projection errors for census tracts. *Journal of the American Statistical Association* 90(429): 64–71. doi:10.1080/01621459.1995.10476489.

Smith, S.K., Tayman, J., and Swanson, D.A. (2013). *A practitioner's guide to state and local population projections*. Dordrecht: Springer. doi:10.1007/978-94-007-7551-0.

Swanson, D.A., Schlottmann, A., and Schmidt, B. (2010). Forecasting the population of census tracts by age and sex: An example of the Hamilton–Perry method in action. *Population Research and Policy Review* 29(1): 47–63. doi:10.1007/s11113-009-9144-7.

Tayman, J. (2011). Assessing uncertainty in small area forecasts: State of the practice and implementation strategy. *Population Research and Policy Review* 30(5): 781–800. doi:10.1007/s11113-011-9210-9.

Tayman, J., Swanson, D.A., and Baker, J. (2021). Using synthetic adjustments and controlling to improve county population forecasts from the Hamilton–Perry method. *Population Research and Policy Review* 40(3): 1355–1383. doi:10.1007/s11113-021-09646-7.

Wilson, T. (2016). Evaluation of alternative cohort-component models for local area population forecasts. *Population Research Policy Review* 35(2): 241–261. doi:10.1007/s11113-015-9380-y.

Wilson, T. (2017). Does averaging yield more accurate local and regional population forecasts? *Applied Spatial Analysis and Policy* 10(4): 497–513. doi:10.1007/s12061-016-9194-2.

Wilson, T. (2018). Evaluation of simple methods for regional mortality forecasts. *Genus* 74:14. doi:10.1186/s41118-018-0040-z.

Wilson, T. and Grossman, I. (2022). Evaluating alternative implementations of the Hamilton–Perry model for small area population forecasts: The case of Australia. *Spatial Demography*. doi:10.1007/s40980-021-00103-9.

Wilson, T. and Bell, M. (2004). Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research* 21(2): 127–160. doi:10.1007/BF03031895.

Wilson, T., Brokensha, H., Rowe, F., and Simpson, L. (2018). Insights from the evaluation of past local area population forecasts. *Population Research Policy Review* 37(1): 137–155. doi:10.1007/s11113-017-9450-4.

Wilson, T., Grossman, I., Alexander, M., Rees, P., and Temple, J. (2021). Methods for small area population forecasts: State-of-the-art and research needs. *Population Research and Policy Review*. doi:10.1007/s11113-021-09671-6.

Appendix

Table A1: Model migration schedule rates

Period-cohort	Females	Males
Birth to 0–4	0.27641	0.27927
0–4 to 5–9	0.22832	0.22753
5–9 to 10–14	0.17435	0.17020
10–14 to 15–19	0.18610	0.16218
15–19 to 20–24	0.35814	0.29384
20–24 to 25–29	0.45111	0.40990
25–29 to 30–34	0.39706	0.39627
30–34 to 35–39	0.30038	0.31805
35–39 to 40–44	0.21962	0.23831
40–44 to 45–49	0.16786	0.18051
45–49 to 50–54	0.14147	0.14773
50–54 to 55–59	0.13116	0.13401
55–59 to 60–64	0.12571	0.12842
60–64 to 65–69	0.11573	0.11945
65–69 to 70–74	0.09726	0.10085
70–74 to 75–79	0.07271	0.07464
75–79 to 80–84	0.04793	0.04793
80+ to 85+	0.02800	0.02699

Source: Model migration schedule fitted to 2006–2011 migration between all SA3 areas from the ABS 2011 census