EVALUATION OF A NEW SIMPLIFIED POPULATION PROJECTION MODEL: A CASE STUDY OF LOCAL GOVERNMENT AREA PROJECTIONS IN TASMANIA

Tom Wilson*
Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3010, Australia.
Email: wilson.t1@unimelb.edu.au

Irina Grossman
Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St, Melbourne, Vic 3010, Australia.
Email: irina.grossman@unimelb.edu.au

*Corresponding Author

ABSTRACT: Preparing local area population projections with state-of-the-art demographic models can be a challenging, time-consuming and costly task. Alternative simpler models can produce projections quickly and easily, but at the cost of less output detail, less flexibility in creating scenarios, and sometimes lower accuracy. This paper presents an evaluation of a new modelling approach which blends the conceptual sophistication of state-of-the-art cohort-component models with the low data requirements of simple models. A key feature is that no locally-specific fertility, mortality, or migration input data is necessary. The new model is tested by producing 'projections' of local government area populations by age and sex in Tasmania over recent periods, with the results then compared to actual populations. The model is shown to produce reasonably accurate projections, and out-perform a simple benchmark model. The strengths and weaknesses of the new approach are discussed.

KEYWORDS: Population projections; forecast accuracy; local government areas; Tasmania.

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

1. INTRODUCTION

The preparation of local area population projections is usually a challenging, complex, and time-consuming task. In situations where there are several hundred local areas, it can take a small team several months to complete. The input data requirements are substantial and will often necessitate the purchase of customised data tables; unavailable data may have to be indirectly estimated; and smoothing must be applied to ragged age patterns of fertility, mortality and migration to obtain the underlying demographic rates. There are no standard and easy ways of formulating assumptions about the future of local area fertility, mortality, and migration. The coding of the projection model calculations, or the use of an existing projections program, can be demanding. And the process of reviewing and fine-tuning projection outputs can take weeks.

Yet even with all the time, effort and resources which are normally dedicated to the task, projections for many areas often deviate substantially from the actual populations measured years later. Several studies have demonstrated how local area population projections can be highly erroneous even just a few years into the future, with the smallest populations experiencing the greatest errors (Rayer and Smith, 2010; Statistics New Zealand, 2008; Wilson *et al.*, 2018). In a study analysing 30 years of past local area population projections in Australia, median absolute forecast errors of the total population were found to be 2.8% after 5 years and 5.4% after 10 years (Wilson *et al.*, 2018). But the wide error distribution means that 5% of local areas experienced errors above 14.2% after 5 years and above 23.5% after 10 years.

Given that local area population projections are widely used for various planning and policy purposes, and can influence multi-million dollar decisions, it would be useful if they could achieve levels of accuracy that are within a few percent of the actual populations. Ideally, the projections would have lower errors than in the past, and with a narrower error distribution, i.e., where the highest errors are not so high. At the same time, it would be very helpful if the process of preparing the projections could be simplified and shortened. Unfortunately, it is usually the case that the more sophisticated projection models for producing age-sex projections, like the multiregional cohort-component model (Rogers, 1995), are complex, data-intensive, and time-consuming to implement. Simpler models, like the Hamilton-Perry model (Hamilton and Perry, 1962; Smith *et al.*, 2013), are data-light, easy to use, and quick to implement, but offer only basic outputs and can yield less accurate projections (Wilson, 2016). The Hamilton-Perry model projects populations by age and sex by assuming recent ratios of cohort populations at two points in time continue into the future. It is described in section 2 below.

This paper presents an evaluation of a new approach to preparing local area population projections. This employs a conceptually robust projection model which requires relatively little input data. The new approach incorporates a simplified bi-regional cohort-component model (Rogers, 1976; Wilson and Bell, 2004), but does not need any local area migration, fertility, or mortality input data to be gathered, and requires little in the way of assumption-setting (Wilson, 2022). Instead, it uses synthetic migration data derived from model migration age schedules and local area net migration age patterns. This *synthetic migration cohort-component model* has been previously tested by applying it to Statistical Area Level 3 (SA3) areas of Australia (which have populations mostly in the range of 30,000 to 130,000). Those evaluations demonstrated its ability to produce reasonably accurate projections over a 15 year projection horizon, achieving greater accuracy than the Hamilton-Perry model. Significantly, the whole projections process took only a few days.

However, the new projection approach has not yet been tested on local government areas (LGAs), many of which are smaller in population than SA3 areas. Small populations are the most challenging to forecast because they are prone to sudden changes in demographic trends, and data for these areas are sparse and noisy. In this paper, we report an evaluation of the model which involved producing 'projections' of LGA populations in Tasmania over past periods. Although we could have chosen any State/Territory, Tasmania provides a useful case study because many of its 29 LGAs have very small populations which will 'stress test' the new model. The projections were then compared with actual populations and error measures were calculated. We also evaluated equivalent projections produced by the 'competitor' Hamilton-Perry model.

Following this introduction, the paper describes the data, projection models, and assessment methods employed (section 2), while the projection results are summarised in section 3. The strengths and weaknesses of the evaluated projection approaches, and the implications of the study's findings for the practice of LGA population projections, are discussed in section 4.

2. DATA AND METHODS

Synthetic Migration Cohort-Component Model

The synthetic migration cohort-component model is a simplified bi-regional cohort-component model (Wilson, 2022). It produces projections of local area populations by sex and five year age groups in five year time intervals and, to simplify its operation for users, has been programmed in an Excel/VBA program with considerable amounts of automation. Figure 1 summarises the process. For each local area, migration flows are modelled between that area and the rest of the world (the bi-regional aspect of the model). It projects two migration flows: inward migration (domestic in-migration plus immigration from overseas combined) and outward migration (domestic out-migration and emigration combined) by age and sex. Outward migration is projected using outward migration rates multiplied by local area populations-at-risk. Inward migration is projected directly as flows because the population-at-risk of inward migration is the rest of the world, which is not modelled. Births and deaths are handled in the usual way for a cohort-component model, with age-specific rates multiplied by populations-at-risk.

The model's inward and outward synthetic migration flows are estimated over a recent 5 year base period using population accounts (Rees and Willekens, 1986). While the calculations are non-trivial, the whole estimation process is automated in the Excel/VBA program and is relatively simple for users of the program. Births and deaths must be calculated before migration can be estimated. Births by sex are used if the data are available, or estimated indirectly using age-specific fertility rates multiplied by the local female population-at-risk. The age-specific fertility rates are based on the local Total Fertility Rate, which is estimated indirectly from the population age-sex structure using the xTFR method of Hauer and Schmertmann (2020), and then disaggregated to age groups using a set of model age-specific fertility rates. This indirect estimation process allows the projection model to be applied in circumstances where fertility data are unavailable for the chosen geographical areas.

Deaths by age and sex are directly input if available, or estimated from age-specific death rates multiplied by the local population-at-risk if not. Death rates are estimated indirectly from local life expectancy at birth and a national mortality surface (Wilson, 2018). The national mortality surface consists of a set of model life tables covering a wide range of mortality conditions, enabling age-specific death rates to be calculated from the point on the mortality surface which corresponds to the estimated life expectancy at birth. Once base period births and deaths are known, the remaining population change over the five year base period can be attributed to net migration.

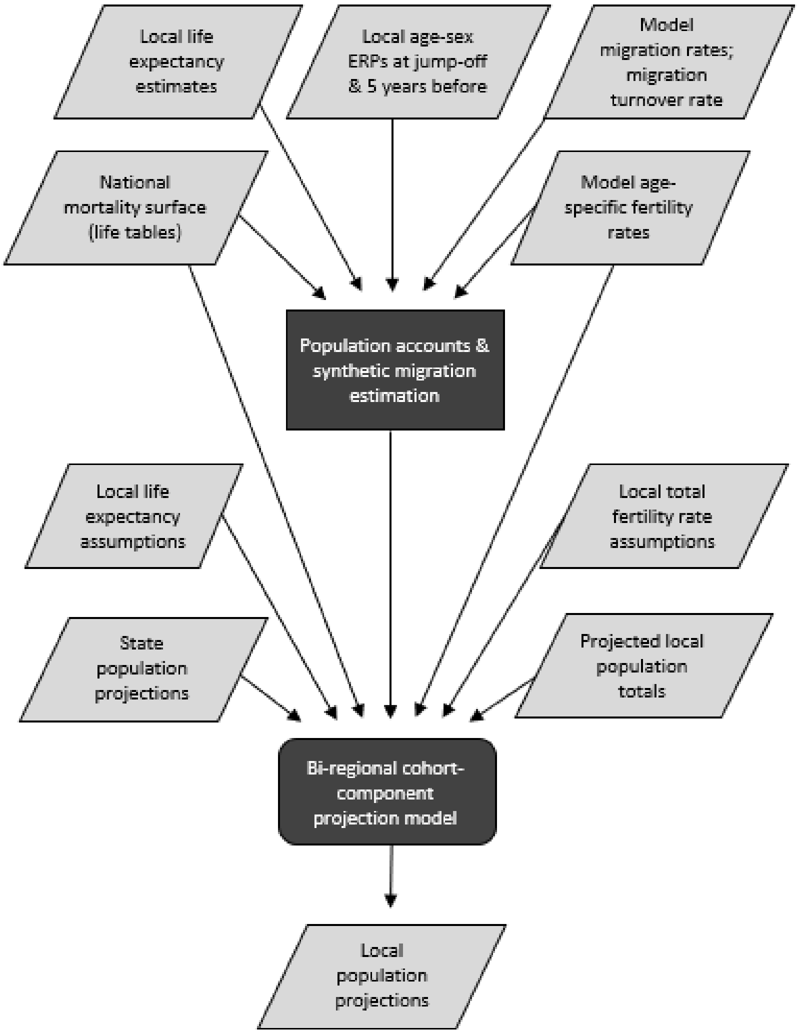

Figure 1. A Summary of the Projections Process Using the Synthetic Migration Cohort-Component Model. Source: the Authors.

Synthetic migration flows are then estimated from three main inputs:

a) a set of model migration rates by age and sex used for all areas;
b) base period net migration for each local area by age and sex (just calculated); and

c) an estimated crude migration turnover rate for local areas, defined as total inward plus outward migration divided by the total population. This can be universal or area-specific.

Preliminary migration for the base period is obtained by multiplying the model migration rates by local area base period populations-at-risk for each age-sex group. Base period total migration turnover is estimated by multiplying the crude migration turnover rate by local area total populations-at-risk. The age-sex migration estimates are then scaled to match total migration turnover, and split evenly into separate inward and outward migration flows. The two flows are then differentiated by adjusting them proportionally to be consistent with base period total net migration. Finally, the two flows are adjusted by age and sex so that inward minus outward migration by age and sex matches the base period net migration age-sex values. It is important to note that these are not estimates of real migration flows; they are synthetic migration values with plausible age patterns which are consistent with base period net migration.

The cohort-component population projections are subject to two sets of constraints. The first is a set of independent projections of local area population totals. This is included because previous research has demonstrated that constraining small area age-sex projections to independent totals generally improves accuracy (Baker *et al.,* 2021; Reinhold and Thomsen, 2015; Tayman *et al.*, 2021; Wilson, 2016). The projection model adjusts inward and outward migration so that the cohort-component calculations produce projections which match these independent total populations. The second constraint is a set of State population projections. It is generally good practice to ensure that local area projections are fully consistent with projections at higher geographies. Local area births by sex, deaths by age and sex, and net migration by age and sex are constrained to those of the State projections. With all the demographic components of change constrained, local area projected populations by age and sex are automatically consistent with those at the State level.

Test Projection Assumptions

Two sets of historical test projections for Tasmania's LGAs were produced using the synthetic migration model and the benchmark Hamilton-Perry model. They were:

(1) 2006-based projections out 10 years to 2016, and

(2) 2012-based projections out 5 years to 2017.

These jump-off years were selected because they are the same as those of ABS population projections (Australian Bureau of Statistics [ABS], 2008, 2013a) and thereby enable the use of the Australian Bureau of Statistics (ABS) projections for Tasmania as constraints.

Fertility assumptions were specified in the form of Total Fertility Rates (TFRs). In the synthetic migration model Excel/VBA program, TFRs for the base period are estimated indirectly (Hauer and Schmertmann, 2020). For the test projections, it was assumed that these TFRs would remain unchanged into the future.

Mortality assumptions were formulated in terms of male and female life expectancy at birth. Although it involves some approximation, the simplest option is to assume all areas experience the same life expectancies as the State as a whole, which was the chosen approach in this case. The life expectancy assumptions for Tasmania from the ABS population projections were used. Age-specific death rates were calculated from the assumed life expectancies using the mortality surface of past and projected national life tables.

Migration turnover assumptions were estimated from census migration data. One year interval out-migration from each LGA to anywhere else in Australia, in-migration to each LGA, and immigration from overseas, were obtained. Emigration was assumed to be half the value of immigration. These migration flows were then used to estimate approximate crude migration turnover rates for each area. Although there are several conceptual and empirical approximations in this approach, the model has been shown to be insensitive to the values of the crude migration turnover rate (Wilson, 2022), so approximate estimates are sufficient.

Population Projections Used as Constraints

The State population projection constraints consisted of the ABS 2006-based and 2012-based Series B population projections (ABS, 2008, 2013a). Fortunately, the "re-casting" of Estimated Resident Populations (ERPs) made by the ABS following the 2011 Census, in which ERPs between 1991 and 2011 were revised, involved only small revisions being made to Tasmania's population in 2006 (ABS, 2013b). The recast 2006 ERP was judged sufficiently close to the old ERP (which formed the jump-off populations in the projections) to avoid the need for the ABS projections to be adjusted to be consistent with the recast 2006 ERPs.

Local area total population constraints were prepared by a simple Linear/Exponential model (LIN/EXP) based on population change over the 10 year period up to the jump-off year. In this model, populations are projected using linear extrapolation (LIN) if population change over the last decade has been positive, and exponential extrapolation (EXP) if it has been negative. It, therefore, avoids projecting extreme or implausible values (e.g. exponential growth applied to an area growing rapidly would give extremely high growth, while linear extrapolation applied to a rapidly declining population could result in zero population in the long run). The LIN/EXP extrapolations were constrained to sum across LGAs to the projected State total population from the ABS projections.

The Hamilton-Perry Model

The Hamilton-Perry model (Hamilton and Perry, 1962) is a parsimonious projection model which produces population projections by age group and sex. In its most basic form, it requires only populations by age and sex at two points in time, usually 5 or 10 years apart (Smith *et al.*, 2013). Projections are calculated using Cohort Change Ratios, which are defined as the ratio of a cohort at one point in time to its size 5 (or 10) years earlier. For example, the female population of a local area aged 50-54 in 2021 divided by the population aged 45-49 in 2016 is the Cohort Change Ratio for the cohort which ages from 45-49 to 50-54 over the 2016-2021 interval. Projections are created by multiplying the current population by a Cohort Change Ratio, e.g. the female population aged 50-54 in 2026 is the population aged 45-49 in 2021 multiplied by the Cohort Change Ratio. The 0-4 year old population is calculated via a Child/Woman Ratio, usually defined as the 0-4 year old population divided by the female population aged 15-49. More details about the Hamilton-Perry model are available in Baker *et al.* (2017) and Smith *et al.* (2013).

The Hamilton-Perry model, with various refinements and extensions, has enjoyed a resurgence in recent years, particularly in the United States (Baker *et al.*, 2017; Baker *et al.*, 2021; Hauer, 2019; Swanson *et al.*, 2010; Tayman *et al.*, 2021; Wilson and Grossman, 2022). In the implementation of the Hamilton-Perry used for this study, projections were subject to the same two sets of constraints as the synthetic migration model: projected LGA population totals from the LIN/EXP model, and ABS age-sex projections for Tasmania.

Forecast Error Measures

The test projections were compared to ERPs prepared by the ABS (ABS, 2021a, 2021b) using two main error measures. Projected total populations produced by the extrapolative model were evaluated using Absolute Percentage Error (APE). It is defined as:

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

where F = forecast and A = actual population. Mean and median values across all LGAs are reported.

The error in a local area's projected age-sex populations was measured by a modified APE measure (Wilson, 2022). The $APE_{age\text{-}sex}$ summarises in a single metric the error across all age-sex projected populations for a local area, and is calculated as:

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

where s= sex and a = age group. When viewing projected and actual populations in a population pyramid, the numerator is the difference in area between the projected population age-sex structure and the actual age-sex structure. The $APE_{age\text{-}sex}$, therefore, highlights errors in a population's projected age-sex structure even when the total population is projected accurately. Effectively, $APE_{age\text{-}sex}$ is a population-weighted mean of all individual age-sex APEs for a local area. Mean and median values of $APE_{age\text{-}sex}$ across all LGAs are reported.

Judging what level of error is acceptable to users is difficult because it is likely to vary depending on each specific use of the projections. Nonetheless, for this paper we consider errors of APE and $APE_{age\text{-}sex}$ under 2.5% to be excellent, 2.5-5% to be good, 5-10% acceptable, and over 10% as poor. Given that forecast error generally increases as population size decreases (Tayman, 2011), we expected the age-sex projections to be more erroneous than population totals.

3. RESULTS

Projected Population Totals

Forecast errors of projected LGA population totals from the extrapolative LIN/EXP model are summarised in Table 1. The table presents Absolute Percentage Errors (APEs) of LGA populations out 5 and 10 years from 2006 and out 5 years from 2012. Average error values are given at the bottom of the table. The median APEs for the 2006-based projections were 1.8% after 5 years and 3.7% after 10 years, while the median APE for the 2012-based projections was 2.6%. Compared to local area population projections in Australia generally, these projections are relatively accurate. Previously calculated median APEs for local area total populations in

Australia for projections prepared from the 1980s to the 2010s are 2.8% after 5 years and 5.4% after 10 years (Wilson *et al.*, 2018).

Table 1. Absolute Percentage Errors of Projected Population Totals from the LIN/EXP Extrapolative Model, 2006- and 2012-Based Projections. Source: the Authors.

	2006-based		2012-based
LGA	**5 years ahead**	**10 years ahead**	**5 years ahead**
Break O'Day	2.4	11.9	6.9
Brighton	2.2	1.7	1.5
Burnie	0.4	5.6	5.3
Central Coast	1.6	2.9	3.4
Central Highlands	3.5	1.8	6.8
Circular Head	0.4	3.2	3.0
Clarence	1.6	3.9	2.7
Derwent Valley	1.9	2.8	0.2
Devonport	1.1	2.9	2.8
Dorset	2.3	9.2	2.7
Flinders	8.3	9.4	17.5
George Town	0.5	0.4	0.6
Glamorgan-Spring Bay	3.8	7.4	1.6
Glenorchy	1.7	3.5	2.6
Hobart	1.5	2.4	2.3
Huon Valley	3.1	2.0	0.3
Kentish	0.1	7.4	6.1
King Island	2.5	2.2	2.1
Kingborough	2.4	1.0	0.8
Latrobe	4.8	3.4	0.9
Launceston	0.6	2.7	2.0
Meander Valley	4.0	10.9	2.2
Northern Midlands	3.9	6.5	1.2
Sorell	2.0	3.0	1.1
Southern Midlands	1.8	5.3	5.7
Tasma	0.1	6.4	4.2
Waratah-Wynyard	1.0	3.8	3.5
West Coast	0.7	6.3	5.4
West Tamar	1.5	6.2	0.5
Median APE	**1.8**	**3.5**	**2.6**
Mean APE	**2.1**	**4.7**	**3.3**

As is usually the case for population projections, the test LGA population projections became less accurate the further they extended into the future. For the 2006-based projections, the total population projections 5 years out were either excellent or good (under 5% APE) for 28 out of 29 LGAs (97%), while at 10 years out projections for only 17 out of 29 LGAs (59%) achieved the same level of accuracy. Two LGAs, Break O' Day and Meander Valley, experienced projections 10 years out which were more than 10% too high (classified as poor quality). In the 2012-based projections, 22 out of 29 LGAs (76%) had total population projections in the excellent or good categories.

Projected Age-Sex Populations

Errors in the projections of LGA age-sex populations, as measured by $APE_{age\text{-}sex}$, are shown in Table 2. Given the commonly observed negative relationship between error and population size, it is not surprising to find that errors for age-sex-specific populations are higher than those for total populations. Average errors of the synthetic model's projections of age-sex-specific populations are moderately lower than those of the Hamilton-Perry model. The median $APE_{age\text{-}sex}$ for the synthetic model's 2006-based projections is 5.2% after 5 years (compared to 6.4% for the Hamilton-Perry model) and 7.7% after 10 years (compared to 9.8%). For the 2012-based projections, the median $APE_{age\text{-}sex}$ out 5 years is 5.7% for the synthetic model compared to 6.3% for the Hamilton-Perry.

Figure 2 provides an illustration of an average level of accuracy in the age-sex projections generated by the synthetic model. It shows the projected population of West Tamar LGA in 2016 (shaded bars), along with the 2016 ERPs (solid black lines) and the jump-off ERPs of 2006 (dashed lines). The projection has been reasonably successful in predicting the extent of population ageing, though it has over-projected the childhood age populations (due to errors in both births and net migration) and adult populations in the 30s and 40s age groups. Thus, the *broad* age-sex structure of this population has been projected moderately well 10 years ahead, but errors for some individual age groups are quite high.

In terms of the classification of error values for the age-sex projections, none were excellent and only a minority were good. In the 2006-based projections, the number of LGAs whose age-sex projections can be classified as good was 10 after 5 years, and 3 after 10 years. The equivalent number of LGAs from the Hamilton-Perry projections were 6 and 0. Fortunately, the majority of LGA projections achieved $APE_{age\text{-}sex}$ values under 10%, placing them in the good or acceptable categories. In the 2012-based projections, the synthetic model generated good age-sex projections for 12 LGAs, while the Hamilton-Perry managed good projections for 10 LGAs. A further 13 LGAs from both the synthetic and Hamilton-Perry projections fell in the acceptable category, ensuring that the majority of LGA age-sex projections were at least acceptable.

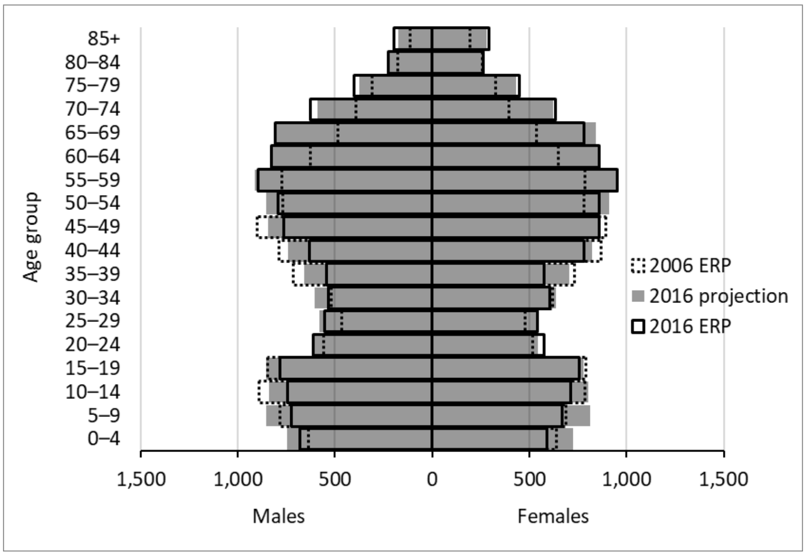

Figure 2. The Age-Sex Structure of the Population of West Tamar in 2016 as Projected by the Synthetic Migration Model. Source: ABS; the Authors.

The age-sex population forecast error generated by both models has a fairly strong association with the error of the extrapolative projected population totals. For the synthetic model, R = 0.65 for the 2006-based projections out 5 years and 0.58 out 10 years, while R = 0.81 for the 2012-based projections out 5 years. As would be expected, it is generally the case that the better the independent total population projection, the better the age-sex projection.

Table 2 shows that the worst projections of age-sex populations from the synthetic model (with values of $APE_{age\text{-}sex}$ consistently above 10%) were obtained for Flinders, Tasman, and King Island. These are the LGAs with the smallest populations in Tasmania, and their populations are subject to the largest random fluctuations in demographic rates. In fact, when considering all LGAs in the State there is a clear association between $APE_{age\text{-}sex}$ and the logarithm of total population (not shown), a finding common to many other local area forecast evaluation studies (Smith, 1987; Tayman, 1996; Wilson *et al.*, 2018). Equivalent errors from Hamilton-Perry model for all three small LGAs were higher.

Table 2. $APE_{age\text{-}sex}$ of Population Age-Sex Projections From the Synthetic Migration and Hamilton-Perry Models, 2006- and 2012-Based Projections.
Source: the Authors.

	2006-based				2012-based	
	5 years ahead		10 years ahead		5 years ahead	
LGA	**Synth**	**H-P**	**Synth**	**H-P**	**Synth**	**H-P**
Break O'Day	7.9	9.0	16.2	17.5	8.9	10.4
Brighton	5.1	6.7	7.3	9.8	4.5	4.5
Burnie	3.5	4.1	7.4	7.6	6.0	6.6
Central Coast	4.4	5.3	6.6	7.8	5.6	5.7
Central Highlands	11.9	14.4	14.0	18.0	15.2	14.8
Circular Head	4.5	5.8	7.9	10.4	6.4	6.3
Clarence	2.9	3.1	4.3	5.0	3.5	3.7
Derwent Valley	5.2	6.5	7.6	10.2	5.8	6.3
Devonport	3.7	4.1	6.6	8.0	4.6	4.2
Dorset	6.8	6.4	11.2	11.2	5.6	6.8
Flinders	34.8	41.6	26.1	34.4	22.5	29.0
George Town	5.0	5.7	7.5	9.7	7.7	7.7
Glamorgan-Spring Bay	9.2	11.2	11.5	16.7	7.1	9.4
Glenorchy	4.9	5.3	7.3	7.6	3.7	4.4
Hobart	4.8	5.4	6.6	7.6	3.4	4.0
Huon Valley	5.6	6.4	7.2	6.9	5.0	5.3
Kentish	7.6	7.0	11.7	12.9	9.0	9.2
King Island	13.7	16.6	16.0	20.1	13.8	14.2
Kingborough	4.4	4.9	4.3	6.0	3.3	3.5
Latrobe	5.9	6.5	9.2	9.9	6.3	6.4
Launceston	2.7	3.3	4.6	5.4	3.5	3.2
Meander Valley	5.7	6.1	11.4	11.1	4.2	5.0
Northern Midlands	6.2	7.2	9.6	9.4	4.7	4.7
Sorell	5.1	7.4	7.2	9.6	4.9	5.4
Southern Midlands	6.4	8.7	10.9	13.3	8.2	9.3
Tasma	15.6	19.0	18.8	26.4	9.9	13.5
Waratah-Wynyard	5.1	5.6	6.2	8.4	5.7	5.9
West Coast	6.5	8.4	12.0	14.9	12.6	13.9
West Tamar	3.9	4.7	7.7	7.6	4.7	4.8
Median $APE_{age\text{-}sex}$	**5.2**	**6.4**	**7.7**	**9.8**	**5.7**	**6.3**
Mean $APE_{age\text{-}sex}$	**7.2**	**8.5**	**9.8**	**11.8**	**7.1**	**7.9**

Note: Synth = synthetic migration cohort-component model; H-P = Hamilton-Perry model

The two influences of total population size and error in the separate projections of the total population together account for much of the variation in age-sex population forecast error. Multiple linear regression shows that the logarithm of total population size at the jump-off year and the APE in the LIN/EXP projections of the total population are able to predict $APE_{age\text{-}sex}$ fairly well. Summary regression outputs are shown in Table 3. For the synthetic model, the adjusted R^2 values are 0.71 and 0.84 for the 2006-based projections out 5 years and 10 years, respectively, and

0.91 for the 2012-based projections out 5 years. The Shapley decomposition (Shapley, 1953) indicates that the natural log of the jump-off population contributes most to the R-squared.

Table 3. Modelling the Value of $APE_{age\text{-}sex}$ from the Synthetic Migration Model. Source: the Authors.

Variable	Coefficient	Standard error	p value	% contribution†
After 5 years				
2006-based				
Intercept	36.37	6.35	5.02E-06	
Ln(jump-off population)	-3.49	0.64	1.08E-05	61
Total APE	1.49	0.39	0.00076	39
Adjusted R-squared	0.71			
No. observations	29			
2012-based				
Intercept	28.41	2.80	1.56E-10	
Ln(jump-off population)	-2.49	0.28	2.32E-09	57
Total APE	0.58	0.09	8.08E-07	43
Adjusted R-squared	0.91			
No. observations	29			
After 10 years				
2006-based				
Intercept	39.18	3.64	4.57E-11	
Ln(jump-off population)	-3.45	0.37	7.47E-10	74
Total APE	0.55	0.13	0.00021	26
Adjusted R-squared	0.84			
No. observations	29			

Note: Ln(jump-off population) = natural logarithm of total population size at the jump-off year; Total APE = the APE in the LIN/EXP projections of the total population. † Shapley decomposition of the adjusted R-squared, with the standardised Shapley value reported.

4. DISCUSSION AND CONCLUSIONS

The results of the test projections reveal the synthetic migration cohort-component model to be capable of producing reasonable projections of Tasmania's LGA populations by age and sex up to a decade ahead. And this was achieved using the simplest approach to formulating assumptions about the future of fertility, mortality and migration turnover. The synthetic migration model proved moderately more accurate than the Hamilton-Perry model. These findings supplement previous work which has shown that the synthetic model produces reasonable projections for SA3 areas and with lower errors than alternative simple cohort models (Wilson, 2022). Of course, reasonable model performance in the past is not an absolute guarantee of reasonable performance in the future. But it at least suggests that it is possible.

We argue that the synthetic migration model is a useful addition to the population forecasting toolbox for preparing local area projections. This is for several reasons:

1) it requires relatively little input data, much of which is freely available from the ABS website (e.g. local area ERPs by sex and five year age group)
2) projections can be prepared easily and quickly using the freely available Excel/VBA projection program
3) it produces local area population projections by sex and five year age group which are consistent with independent State-level projections
4) it outputs projected local area births, deaths and net migration
5) assumption-setting is relatively simple; assumptions are prepared in terms of the TFR, life expectancy at birth, the crude migration turnover rate, and projected population totals
6) the model is conceptually strong, being based on the directional migration bi-regional cohort-component model
7) it has been shown to produce more accurate projections than the simple Hamilton-Perry model in this evaluation study and others (Wilson, 2022).

In summary, the model offers a lot for relatively little effort and cost: the projections output quality and detail are high in relation to the sum of all the required inputs (demographic data inputs, time needed, staffing, level of expertise, project costs, etc.).

The model could be useful for preparing local area projections where projected populations in five year age groups in five year time intervals are sufficient for stakeholders. It can handle various types of local area geography, including LGAs, SA3 areas, and SA2 areas, although the smaller the population, the more challenging the projections become due to noisy data. The model may also prove useful in circumstances where the resources available for producing local area projections are very limited, or where they have to be prepared very quickly. Furthermore, the model could also play a role as a validation tool to compare results with those of a more complex and detailed projection system.

However, several limitations of this model should be noted. Internal and overseas migration are not handled separately, and so separate internal and overseas projection assumptions cannot be prepared, and separate outputs are not available. Inward migration is projected directly as a migration flow, and not as a rate multiplied by a population-at-risk. The synthetic inward and outward migration flows are not accurate estimates of actual inward and outward migration in themselves. They comprise migration flows with plausible age patterns which are consistent with the base period net migration values. They cannot be interpreted as robust projections of directional migration. In addition, the calculation of net migration as a residual in the base period to act as a constraint on the inward and outward migration flows means that the synthetic migration values depend on the accuracy of the base period ERPs by age and sex.

Furthermore, like all population projection models, demographic trends are likely to vary from the assumed trajectories to some extent. In particular, the assumption of base period outward migration rates and inward migration age patterns (though not levels) remaining unchanged will prove unrealistic in areas undergoing major changes in terms of housing type and socio-economic composition. Errors at the local area scale may be higher than many users expect, or at least higher than they find easy to handle in their decision-making. Examining how past projections have turned out can therefore be a useful guide to the approximate magnitude of error that can be expected in the future. As this study has demonstrated, projections for areas with the smallest populations are likely to be the least accurate. For populations under about 5,000 people, we would recommend that projections are accompanied by a warning that they are highly error-prone.

The synthetic migration model cannot offer the same sophistication and detail of a multiregional or bi-regional cohort-component model with single year of age detail and a distinction between overseas, interstate, and intra-state migration flows. This more complex type of model offers the greatest flexibility in assumption-setting and projection output detail. But it requires a lot more input data, much more effort in data preparation and assumption-setting, greater expertise in projection methodology, more data validation, and more cost and time in total. Where this is not possible, the synthetic migration cohort-component model offers a good alternative.

Projection Program

The Excel/VBA implementation of the synthetic migration cohort-component model is available for download at https://doi.org/10.6084/m9.figshare.19372784.v1.

REFERENCES

Australian Bureau of Statistics (2008). TABLE B9. Population projections, By age and sex, Australia - Series B. Online version accessed March 2022, https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3222.02006%20to%202101?OpenDocument.

Australian Bureau of Statistics (2013a). TABLE B9. Population projections, By age and sex, Australia - Series B. Online version accessed March 2022, https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3222.02012%20(base)%20to%202101?OpenDocument.

Australian Bureau of Statistics (2013b). Feature Article 2: Recasting 20 Years of ERP. Online version accessed March 2022, https://www.abs.gov.au/ausstats/abs@.nsf/Products/3101.0~Dec+2012~Feature+Article~Recasting+20+Years+of+ERP+(Feature+Article)?OpenDocument.

Australian Bureau of Statistics (2021a). Regional population. Online version accessed March 2022, https://www.abs.gov.au/statistics/people/population/regional-population/2019-20.

Australian Bureau of Statistics (2021b). Regional population by age and sex. Online version accessed March 2022, https://www.abs.gov.au/statistics/people/population/regional-population-age-and-sex/2020.

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, pp. 1341-1354.

Baker, J., Swanson, D. A., Tayman, J. and Tedrow, L. M. (2017). *Cohort Change Ratios and their Applications,* Springer International Publishing, Cham.

Hamilton, C. H. and Perry, J. (1962). A short method for projecting population by age from one decennial census to another. Social Forces, 41, pp. 163-170.

Hauer, M. E. (2019). Population projections for U.S. counties by age, sex, and race controlled to shared socioeconomic pathway. *Scientific Data*, 6, pp. 1-15.

Hauer, M. E. and Schmertmann, C. P. (2020). Population pyramids yield accurate estimates of total fertility rates. *Demography*, 57, pp. 221-241.

Rayer, S. and Smith, S. K. (2010). Factors Affecting the Accuracy of Subcounty Population Forecasts. *Journal of Planning Education and Research*, 30, pp. 147-161.

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

Reinhold, M. and Thomsen, S. L. (2015). Subnational Population Projections by Age: An Evaluation of Combined Forecast Techniques. *Population Research and Policy Review*, 34, pp. 593-613.

Rogers, A. (1976). Shrinking large-scale population-projection models by aggregation and decomposition. *Environment and Planning A*, 8, pp. 515-541.

Rogers, A. (1995). *Multiregional Demography: Principles, Methods and Extensions*, Chichester, New York.

Shapley, L.S. (1953). A Value for n-person Games. In H.W. Kuhn and A. W. Tucker (Eds*) Contributions to the Theory of Games*, Princeton University Press, New Jersey.

Smith, S. K. (1987). Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association*, 82, pp. 991-1003.

Smith, S. K., Tayman, J., and Swanson, D. A. (2013). *A Practitioner's Guide to State and Local Population Projections*, Springer, Dordrecht.

Statistics New Zealand (2008). *How accurate are population projections? An evaluation of Statistics New Zealand population projections, 1991-2006*, Statistics New Zealand, Wellington.

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, pp. 47-63.

Tayman, J. (1996). The accuracy of small-area population forecasts based on a spatial interaction land-use modeling system. *Journal of the American Planning Association*, 62, pp. 85-98.

Tayman, J. (2011). Assessing uncertainty in small area forecasts: State of the practice and implementation strategy. *Population Research and Policy Review*, 30, pp. 781-800.

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, pp. 1355-1383.

Wilson, T. (2016). Evaluation of alternative cohort-component models for local area population forecasts. *Population Research and Policy Review*, 35, pp. 241-261.
Wilson, T. (2022). Preparing local area population forecasts using a bi-regional cohort-component model without the need for local migration data. *Demographic Research*, 46, pp. 919-956.
Wilson, T. and Bell, M. (2004). Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research*, 21, pp. 127-160.
Wilson, T., Brokensha, H., Rowe, F. and Simpson, L. (2018). Insights from the evaluation of past local area population forecasts. *Population Research and Policy Review*, 37, pp. 137-155.
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*, 10, pp 1-31.