Chapter 11 Forecasting Origin-Destination-Age-Sex Migration Flow Tables with Multiplicative Components



James Raymer, Xujing Bai, and Peter W. F. Smith

11.1 Introduction

Estimates of future internal migration are required for making accurate population projections, and for policy development and planning. However, migration forecasting is complicated from a demographic modelling perspective in that it represents a transition from an origin population to a destination population. Andrei Rogers, Frans Willekens, Alan Wilson, and Phil Rees developed the multiregional population projection framework for including such transitions starting in the 1960s (Rogers 1966, 1968, 1975; Wilson and Rees, 1974a, b, 1975; Rogers and Willekens 1976; Willekens and Rogers 1978). However, methods for producing dynamic forecasts of interregional migration flows with measures of uncertainty are still relatively few.

In this chapter, we build from a range of earlier efforts that used multiplicative or log-linear models to forecast counts of migration flows by origin, destination, age and sex (Stillwell 1986; Willekens and Baydar 1986; Van Imhoff et al. 1997; Van der Gaag et al. 2000; Sweeney and Konty 2002; Van Wissen et al. 2008). In particular, this research extends the multiplicative component approach developed by Raymer et al. (2006) for projecting interregional migration in Italy and Raymer et al. (2017) for projecting Indigenous migration in Australia. We illustrate the forecast methodology by using origin-destination-age-sex tables representing

J. Raymer (🖂) · X. Bai

© The Author(s) 2020

School of Demography, Australian National University, Canberra, Australia e-mail: james.raymer@anu.edu.au; xujing.bai@anu.edu.au

P. W. F. Smith

Department of Social Statistics and Demography and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK e-mail: p.w.smith@soton.ac.uk

S. Mazzuco, N. Keilman (eds.), *Developments in Demographic Forecasting*, The Springer Series on Demographic Methods and Population Analysis 49, https://doi.org/10.1007/978-3-030-42472-5_11

internal migration flows amongst Australia's state and territory populations. Our research extends the earlier efforts cited above by forecasting each multiplicative component separately and by integrating them together to provide forecasts of interregional migration by age and sex with measures of uncertainty. Modelling each component separately allows the forecaster more control by being able to specify different models for each component.

The forecasting model for internal migration advocated in this chapter is different from the current approach used by the Australian Bureau of Statistics (ABS), which projects gross flows of in-migration and out-migration to/from each state or territory separately from each other. While simpler to include in demographic accounting models, projections of in-migration and out-migration (or even worse, net migration) totals are not as reliable and are known to result in biased projections (Rogers 1990) and inaccurate uncertainty measures (Raymer et al. 2012). Here, biases refer primarily to projected measures that are systematically above (below) the observed values. Most often, biases in regional projections occur when net migration or in-migration rates are used. They are caused by the use of populations not 'at risk' of migration in the denominators. Thus, by focusing on the underlying structures of migration flows, we argue that more reliable projection models may be produced for both internal migration and the subsequent population totals and agesex compositions. Moreover, when the internal migration projections inevitably fail to predict perfectly the future, we have more detailed information about the potential sources of error.

The structure of this chapter is as follows. We first explore how the internal migration patterns in Australia have changed since 1981. We then explore the stability in the underlying structures of migration flows over time, and identify the most important migration structures required for both estimation and projection. Finally, we illustrate the approach by predicting the observed flows with measures of uncertainty for the 2006–2011 and 2011–2016 periods based on historical migration flow data going back to the 1981–1986 period. We also produce and illustrate the results of forecasts for two time periods beyond the observed data, i.e., 2016–2021 and 2021–2026.

11.2 Multiplicative Component Calculations

Analysing and predicting the counts of migration flows may be considered from a categorical data analysis perspective. The basic categories are origin (O), destination (D), age (A) and sex (S). Migration flow tables typically include two or more of these categories. These tables can be decomposed into various hierarchical structures, not all of which are necessary for understanding or for producing accurate predictions. If certain (important) structures are unavailable, they can be imputed or 'borrowed' from auxiliary data sources. This general modelling framework comes from a sequence of papers on the age and spatial structures of internal migration (Willekens

Table 11.1 Notation for an origin-by-destination migration flow table		Region of destination				
	Region of origin	1	2	3	4	Total
	1	0	<i>n</i> ₁₂	<i>n</i> ₁₃	<i>n</i> ₁₄	<i>n</i> ₁₊
	2	n ₂₁	0	n ₂₃	n ₂₄	n ₂₊
	3	<i>n</i> ₃₁	n ₃₂	0	<i>n</i> ₃₄	<i>n</i> ₃₊
	4	<i>n</i> ₄₁	n ₄₂	n43	0	<i>n</i> ₄₊
	Total	n_{+1}	n_{+2}	<i>n</i> ₊₃	n_{+4}	<i>n</i> ++

1983; Stillwell 1986; Van Imhoff et al. 1997; Rogers et al. 2002, 2003; Sweeney and Konty 2002; Raymer et al. 2006, 2017; Raymer and Rogers 2007; Van Wissen et al. 2008).

To begin, consider migration from origin *i* to destination *j*, denoted by n_{ij} . These counts may be organised in a two-way table, such as in Table 11.1 for migration between four hypothetical regions. Here, it is important to make a distinction between cell counts (n_{ij}) and marginal totals, i.e., the total number of out-migrants from each region (n_{i+}) , the total number of in-migrants to each region (n_{i+j}) and the overall level of migration (n_{i+1}) . Note, within area movements (i = j) are excluded from the analyses.

For describing, analysing and projecting migration flow patterns over time, consider the following multiplicative decomposition of an origin-destination table:

$$n_{ij} = (T) (O_i) (D_j) (OD_{ij}), \qquad (11.1)$$

where *T* is the total number of migrants (i.e., n_{i+1}), O_i is the proportion of all migrants leaving from area *i* (i.e., n_{i+1}/n_{i+1}) and D_j is the proportion of all migrants moving to area *j* (i.e., n_{i+j}/n_{i+1}). The interaction component OD_{ij} is defined as $n_{ij}/[(T)(O_i)(D_j)]$ or the ratio of observed migration to expected migration (for the case of no interaction). This general type of model is called a multiplicative component model and may be extended to include other categories, such as age or sex.

The data for this research were obtained from the Australian quinquennial censuses from 1981 to 2016 and include following characteristics:

- state or territory of current residence by state or territory of residence 5 years ago,
- five-year age groups (0–4, 5–9, \ldots , 80+ years), and
- sex.

We focus on the migration transitions between the eight states or territories of Australia: New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA), Western Australia (WA), Tasmania (TAS), Northern Territory (NT) and Australian Capital Territory (ACT). Note, in this study, we apply the forecast methodology described in the next section to a particular type of migration flows, namely, transitions between the place of residence 5-years ago and place of residence at the time of the census. However, the methodology may be applied

to any type or category of migration flows so long as they are arranged in a categorical fashion. Other common types of migration flows include population or administrative register data on the number of moves (events) within a 1-year time interval and census or survey data on transitions based on current residence by place of birth, place of residence prior to last move, or place of residence 1 year ago (Bell et al. 2015).

For illustration of the multiplicative component calculations and their interpretations, the Australian interstate migration flow table and the corresponding multiplicative components for the 2011–2016 period are presented in Tables 11.2 and 11.3, respectively. For example, the number of persons who migrated from Australian Capital Territory (ACT) to New South Wales (NSW) ($n_{ACT, NSW}$) was 23,609 persons. The multiplicative components for this migration flow are equal to:

$$n_{ACT,NSW} = (T) (O_{ACT}) (D_{NSW}) (OD_{ACT,NSW})$$

= (824,392) (0.054)(0.232)(2.281)
= 23,609

	Region of destination								
Region of origin	NSW	VIC	QLD	SA	WA	TAS	NT	ACT	Total
NSW	0	61,484	105,703	12,748	19,921	6590	6369	27,567	240,382
VIC	47,925	0	47,070	13,284	18,721	6932	5932	6725	146,589
QLD	77,163	46,585	0	10,611	21,071	7906	9710	7108	180,154
SA	12,904	19,813	15,507	0	7581	1800	3896	2250	63,751
WA	17,863	22,364	20,636	5741	0	4219	4733	2116	77,672
TAS	5356	9816	8944	1660	3645	0	725	794	30,940
NT	6629	6648	14,002	5610	5239	947	0	1253	40,328
ACT	23,609	8035	8295	1541	1818	600	678	0	44,576
Total	191,449	174,745	220,157	51,195	77,996	28,994	32,043	47,813	824,392

 Table 11.2
 Interstate migration in Australia, 2011–2016

Table 11.3 Multiplicative components of interstate migration in Australia, 2011–2016

	O_i	D_j	OD _{i,NSW}	$OD_{i,VIC}$	$OD_{i,QLD}$	$OD_{i,SA}$	$OD_{i,WA}$	OD _{i,TAS}	OD _{i,NT}	OD _{i,ACT}
T = 824,392										
NSW	0.292	0.232	0.000	1.207	1.647	0.854	0.876	0.779	0.682	1.977
VIC	0.178	0.212	1.408	0.000	1.202	1.459	1.350	1.345	1.041	0.791
QLD	0.219	0.267	1.844	1.220	0.000	0.948	1.236	1.248	1.387	0.680
SA	0.077	0.062	0.872	1.466	0.911	0.000	1.257	0.803	1.572	0.609
WA	0.094	0.095	0.990	1.358	0.995	1.190	0.000	1.544	1.568	0.470
TAS	0.038	0.035	0.745	1.497	1.082	0.864	1.245	0.000	0.603	0.442
NT	0.049	0.039	0.708	0.778	1.300	2.240	1.373	0.668	0.000	0.536
ACT	0.054	0.058	2.281	0.850	0.697	0.557	0.431	0.383	0.391	0.000

From these calculations, we see that the overall level of interstate migration was 824,392 persons, the share of all migration from the ACT was 5.4% (i.e., 44,576 / 824,392 * 100), the share of all migration to NSW was 23.2% (i.e., 191,449 / 824,392 * 100), and that there was more than twice the expected value of migration between these two areas (i.e., 23,609 / (824,392 * 0.054071 * 0.232231) = 23,609 / 10,352). In Table 11.2, the largest flows are between the largest population states of NSW, Victoria (VIC) and Queensland (QLD). The smallest flows are between the smallest states or territories (Tasmania (TAS), Northern Territory (NT), and ACT). In Table 11.3, we see that the largest OD_{ij} ratios are between states or territories that are far apart, e.g., TAS and NT.

Next, consider the multiplicative components for a four-way table of migration by origin, destination, age and sex. The multiplicative component model that fully explains this table is specified as:

$$n_{ijxy} = (T) (O_i) (D_j) (A_x) (S_y) (OD_{ij}) (OA_{ix}) (OS_{iy}) (DA_{jx}) (DS_{jy}) (AS_{xy})$$
$$(ODA_{ijx}) (ODS_{ijy}) (DAS_{jxy}) (ODAS_{ijxy}), \qquad (11.2)$$

where A_x is the proportion of all migrants in age group x and S_y is the proportion of all migrants in sex group y. This model is a lot more complicated because there are now four main effects, six two-way interaction components, three three-way interaction components and one four-way interaction component between the origin, destination, age and sex variables. However, for the main effects and two-way interaction components, the interpretations of the parameters remain relatively simple. For example, the destination-age interaction (DA_{jx}) component is calculated as $n_{+jx+}/[(T)(D_j)(A_x)]$ and represents the ratio of observed age patterns of in-migration to each region divided by the expected age pattern of in-migration. Fortunately, the three-way and four-way interaction terms do not add much additional information and are rarely needed for estimation or projection (see, e.g., Van Imhoff et al. 1997; Smith et al. 2010). The same is true for the two-way interactions between origin and sex (OS_{iy}) and destination and sex (DS_{jy}) . Thus, for most analyses, estimations and projections, the following reduced model may be used:

$$n_{ijxy} = \left(T\right)\left(O_{i}\right)\left(D_{j}\right)\left(A_{x}\right)\left(S_{y}\right)\left(OD_{ij}\right)\left(OA_{ix}\right)\left(DA_{jx}\right)\left(AS_{xy}\right).$$
(11.3)

To illustrate the effectiveness of this model, consider the migration flows presented in Fig. 11.1. Here, we compare the observed and estimated age patterns of female internal migration between NSW and QLD for the 2006–2011 and 2011–2016 periods using the model specified in Eq. 11.3. Clearly, there are not much differences between the estimated and observed flows of migration in this case.

To assess the goodness-of-fit (g) between the observed and estimated migration flow tables, we focus on the following formula:



Fig. 11.1 Observed and estimated age patterns of female migration between New South Wales and Queensland, 2006–2011 and 2011–2016

$$g = \frac{100}{N} \sum_{i=1}^{N} \frac{\left| n_{ijxy} - \hat{n}_{ijxy} \right|}{\hat{n}_{ijxy}}$$

where *N* denotes the total number of cells in the origin-destination-sex-age table in a single period, which for our tables is equal to 1904, i.e., 8 origins × 8 destinations × 2 sexes × 17 age groups, not including the diagonal elements where i = j. The observed number of interstate migrants by age and sex is denoted by n_{ijxy} and the corresponding estimated flows is denoted by \hat{n}_{ijxy} . The test-statistics for the unsaturated model (Eq. 11.3) applied to the 2006–2011 and 2011–2016 data are 16.3% and 16.1%, respectively. For migration flows, we find this simple goodnessof-fit measure works well due to high likelihood of zeros in the observed data when broken down by origin, destination, age and sex. By placing the estimated values in the denominator, this allows us to provide measures for all predicted cell values.

In summary, multiplicative components are useful for analysing the key structures driving migration patterns. These can then be used for the purpose of estimating migration. Moreover, when particular interaction effects cannot be derived from available data, they may be obtained or calculated using other comparable data sets (e.g., interaction data from historical periods or from other populations). Since Snickars and Weibull (1977) found that historical migration tables provide much better estimates of current accessibility than any distance measure, historical data are often used to capture the spatial patterns of migration (see also Tobler 1995). For projection of internal migration patterns, this means we can effectively utilise trends exhibited by previous migration data sets.

11.3 Trends Over Time

In this section, we calculate and present each of the multiplicative components specified in Eq. 11.3 for the periods 1981–1986 to 2011–2016. The purpose of presenting these patterns is primarily to highlight the consistencies and/or any major deviations found in the trends over time, particularly since extrapolations of these components are combined and then used to predict future counts of migration by origin, destination, age, and sex.

The overall level components (T) and proportions of interstate migration in Australia are presented in Fig. 11.2 for the periods 1981–1986 to 2011–2016. During this time, total interstate migration increased from 717 thousand persons in 1981–1986 to 792 thousand persons in 1991–1996, followed by a decline to 774 thousand persons in 2006–2011 and then a sharp increase to 824 thousand persons in 2011–2016. While the total level of interstate migration demonstrated certain amount of fluctuation, its proportion in the total Australian population kept decreasing from around 5.5% in 1981–1986 to 4.4% in 2011–2016. The general decline in the propensity to migrate internally has been observed across Australia by



Fig. 11.2 Total level and proportion of interstate migration in Australian, 1981–1986 to 2011–2016

Bell et al. (2018), as well as in other developed countries (Cooke 2013; Champion et al. 2018). The underlying causes are thought to be population ageing and changing economic structures (i.e., manufacturing to service-based).

For the origin and destination main effect components (O_i and D_j , respectively) presented in Fig. 11.3, we see that the largest states of NSW, VIC and QLD contributed the largest shares of both out-migration and in-migration. While NSW



Fig. 11.3 Relative shares of out-migration (O_i) and in-migration (D_j) by state and territory in Australia, 1981–1986 to 2011–2016

consistently sent out the largest shares of interstate migrants from 1981–1986 to 2011–2016, it never received the largest share of in-migration – the largest share of in-migration was received by Queensland. Indeed, one of the distinctive features of internal migration in Australia over the past several decades is persistent net migration loss from New South Wales to other states in the country. Over 20 years ago, Burnley (1996) attributed this to high levels of immigration to and housing costs in Sydney.

The age and sex main effect components (A_x and S_y , respectively) of interstate migration are presented for the seven time periods in Fig. 11.4. For the age main effects, we find relative increases in shares of migration amongst 30–65 year olds and corresponding declines in the child age groups. These changes are likely caused by the ageing of the population. As for the main effect component for sex, there was a steady (albeit small) decrease in the share of male migrants from 52% in 1981– 1986 to 49% in 2001–2006, which then held constant until the most recent period. This shift towards more female migration is likely caused by the increasing numbers of women seeking tertiary education and employment in Australia.

The values of the origin-destination (OD_{ij}) , origin-age (OA_{ix}) , destination-age (DA_{jx}) and age-sex (AS_{xy}) interaction components, presented in Figs. 11.5, 11.6, 11.7 and 11.8, respectively, represent ratios of observed to expected values. The expected values are calculated based on the multiplication of the overall level component (T) by the main effect components $(O_i, D_j, A_x \text{ or } S_y)$ corresponding to the two variables being interacted. Note, a value of 1.0 implies no difference from the expected value.

For the origin-destination components in Fig. 11.5, there are a couple of things to highlight. First, most of the values are above or below 1.0, which signifies the importance of this component in understanding the migration patterns. Second, there is relative stability in the ratios exhibited over time with all interactions, more or less, remaining the same in terms of being 'higher than expected' or 'lower than expected.' Third, the patterns exhibit clear trends over time, for example, the interaction between SA and NT has been steadily declining since the 1986–1991 period. Fourth, each origin has its own distinct destination patterns with, for example, ACT having more than twice the expected flows to NSW, and nearly half the expected flows to all other states and territories (except VIC which exhibits ratios of around 0.75). The interaction components for migration from VIC, on the other hand, are above 1.0 for state destinations but below 1.0 for territory destinations.

For the origin-age and destination-age components presented in Figs. 11.6 and 11.7, most of the ratios are near the value of 1.0 implying the state /territory age profiles of out-migration and in-migration resemble the overall age profile of migration (A_x). Notable differences in the out-migration age profiles (Fig. 11.5) include higher levels amongst retired age groups for VIC (before 2001) and QLD, relatively low levels of out-migration amongst older persons from WA, TAS (before 2001), NT, and ACT, and a sharp and consistent peak of 15–19 year olds leaving TAS. Notable differences in the in-migration age profiles (Fig. 11.6) include VIC receiving relatively more young adults (in recent periods) and fewer older migrants,



Fig. 11.4 Age (A_x) and sex (S_y) main effect components of interstate migration in Australia, 1981–1986 to 2011–2016

with the opposite occurring for QLD. WA, NT and ACT received considerably fewer older migrants, whereas it was the opposite for TAS. Finally, TAS appears to be growing as a retirement destination while at the same time becoming less attractive to young adults.

Finally, for the female age-sex interaction components (AS_{xy}) presented in Fig. 11.8, we find, for ages above 65 years, there has been a decreasing trend in the



Fig. 11.5 Origin-destination (OD_{ij}) interaction components of interstate migration in Australia, 1981–1986 to 2011–2016

ratios towards 1.0. In general, it can be said that males and females have similar age profiles of migration, except in older age groups where there are more females in the population due to their lower mortality rates.



Fig. 11.6 Origin-age (OA_{ix}) interaction components of interstate migration in Australia, 1981–1986 to 2011–2016



Fig. 11.7 Destination-age (DA_{jx}) interaction components of interstate migration in Australia, 1981–1986 to 2011–2016



Fig. 11.8 Age-sex (AS $_{xy}$) interaction components for female interstate migration in Australia, 1981–1986 to 2011–2016

11.4 Forecasts

In this section, we show how the multiplicative component model can be used to produce predictions of internal migration by origin, destination, age and sex. The emphasis is on extrapolating each of the multiplicative components separately and then combining them to derive the forecasts of internal migration. For illustration, we first apply simple linear and log-linear trend extrapolations to each of the components specified in Eq. 11.3 to produce predictions of the 2006–2011 and 2011–2016 flows. For instance, the formulas of the linear and log-linear trend models for OD_{ij} components, respectively, are:

$$OD_{ii}(t) = \alpha + \beta Y(t) + \varepsilon(t)$$
 and (11.4)

$$\ln\left[OD_{ij}(t)\right] = \alpha + \beta Y(t)_t + \varepsilon(t) \tag{11.5}$$

where $OD_{ij}(t)$ denotes the OD_{ij} component at time *t*, Y(t) denotes the corresponding year, and α and β denote the intercept and slope parameters estimated using ordinary least squares regression applied to the training sample data. The extrapolations are based on the 1981–1986 to 2001–2006 multiplicative components. Note, as part of the modelling process, the predicted main effect components are rescaled so that they sum to 1.0 and, when two-way interaction components are included, all predicted values are rescaled to match the estimated overall level (*T*) component. In comparing the goodness-of-fit statistics for the linear and log-linear trend models, we find little difference between the two approaches. The linear model produced slightly lower *g* values of 24.1% and 30.7% for the 2006–2011 period and 2011–2016 period, respectively, compared to 24.4% and 31.2%, respectively for the log-linear model. Note, calculations of the mean squared error (MSE), mean absolute error (MAE) and symmetric mean absolute percentage error (SMAPE) goodness-of-fit measures also resulted in similar values for the linear and log-linear trend models, where:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (n_{ijxy} - \hat{n}_{ijxy})^{2},$$
$$MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{n}_{ijxy} - n_{ijxy}| and$$
$$SMAPE = \frac{100}{N} \sum_{i=1}^{N} \frac{|\hat{n}_{ijxy} - n_{ijxy}|}{(|n_{ijxy}| + |\hat{n}_{ijxy}|)/2}$$

where *N* denotes the total number of cells (i.e., 1904) in the origin [*i*] by destination [*j*] by sex [*y*] by age [*x*] table [$i \neq j$], *n* denotes the observed number of interstate migrants, and \hat{n} denotes the corresponding estimated flows. In the end, because the difference was so small (see also Fig. 11.9), we decided to use the log-linear trend model because it ensured positive predicted values.

11.4.1 Model Selection

To identify the best multiplicative model for forecasting origin-destination-age-sex tables of migration, we predicted a range of unsaturated models starting with the model specified in Eq. 11.3 and used the g measure as a basis for comparison. All models used log-linear trend extrapolation to predict the component values for 2006–2011 and 2011–2016 based on the observed values from 1981–1986 to 2001–2006.

We tested and compared four models. Model 1 includes extrapolations of all components specified in Eq. 11.3. Model 2 replaces the extrapolations for the OA, DA and AS components with the most recent observed component values (i.e., 2001–2006) and held them constant for the holdout sample forecasts. Model 3 only includes extrapolations for the overall level and main effect components and held all two-way interaction components constant at the 2001–2006 values. Finally, Model 4 only extrapolated the overall level component. The remaining components represented the observed 2001–2006 values. These four models are specified as follows:



Fig. 11.9 Observed and predicted female flows of migration between Victoria and South Australia, 2006–2011 and 2011–2016

Model 1	$n_{ijxy} = (\hat{T})(\hat{O}_i)(\hat{D}_j)(\hat{A}_x)(\hat{S}_y)(\widehat{OD}_{ij})(\widehat{OA}_{ix})(\widehat{DA}_{jx})(\widehat{AS}_{xy})$
Model 2	$n_{ijxy} = (\hat{T})(\hat{O}_i)(\hat{D}_j)(\hat{A}_x)(\hat{S}_y)(\widehat{OD}_{ij})(OA_{ix})(DA_{jx})(AS_{xy}),$
Model 3	$n_{ijxy} = (\hat{T})(\hat{O}_i)(\hat{D}_j)(\hat{A}_x)(\hat{S}_y)(OD_{ij})(OA_{ix})(DA_{jx})(AS_{xy}),$
Model 4	$n_{ijxy} = (\hat{T})(O_i)(D_j)(A_x)(S_y)(OD_{ij})(OA_{ix})(DA_{jx})(AS_{xy}),$

where the 'hat' symbol denotes log-linear extrapolation. The goodness-of-fit values, including *g*, MSE, MAE and SMAPE, for these four models are presented in Table 11.4. Surprisingly, there was very little difference between the overall goodness-of-fit tests. The 'best' performing model for both holdout sample prediction periods was the simplest model, Model 4, that only extrapolated the overall level component and held the remaining components fixed at the observed 2001–2006 values. We did not expect Model 4 to perform as well as the other models. Presumably, it did so because the historical trend data used to predict the multiplicative components forced the predicted values further away from the holdout sample than was observed in the most recent period used in the training sample.

Measure	Model	2006-2011	2011-2016
g (%)	(T)	24.1	30.3
	(T)(O)(D)(A)(S)	24.9	31.1
	(T)(O)(D)(A)(S)(OD)	24.6	31.5
	(T)(O)(D)(A)(S)(OD)(OA)(DA)(AS)	24.4	31.2
MSE	(T)	39,300	66,981
	(T)(O)(D)(A)(S)	39,537	80,410
	(T)(O)(D)(A)(S)(OD)	43,050	87,791
	(T)(O)(D)(A)(S)(OD)(OA)(DA)(AS)	38,449	72,110
MAE	(T)	71	92
	(T)(O)(D)(A)(S)	73	104
	(T)(O)(D)(A)(S)(OD)	74	108
	(T)(O)(D)(A)(S)(OD)(OA)(DA)(AS)	75	108
SMAPE (%)	(T)	19.0	21.5
	(T)(O)(D)(A)(S)	20.1	23.3
	(T)(O)(D)(A)(S)(OD)	19.8	23.3
	(T)(O)(D)(A)(S)(OD)(OA)(DA)(AS)	20.1	24.3

Table 11.4 Goodness-of-fit statistics of different forecast models for internal migration

11.4.2 Forecasting Internal Migration by Age and Sex with Measures of Uncertainty

In this section, we introduce uncertainty measures to Model 4, which turned out to be both the most effective and simplest model. As stated above, we predict the overall level component for the two most recent periods based on a simple log-linear extrapolation. The estimated total levels of interstate migration are 814,176 persons for the 2006–2011 period and 829,022 persons for the 2011–2016 period. The corresponding observed values were 774,013 persons and 824,392 persons, respectively.

In addition to the point predictions, we include 80% and 95% prediction intervals. These are calculated by simulating predictions of each of the components in the model specified in Eq. 11.3, assuming normal distributions for the logged components. For the components held constant over time, we use a random walk model where the variance of the errors is equal to the observed variance in each of the differenced logged components. For instance, for OD_{ij} components,

$$\ln\left[OD_{ij}\left(t\right)\right] = \ln\left[OD_{ij}\left(t-1\right)\right] + \varepsilon(t).$$

The overall level component is predicted using a linear regression model on the log scale, Eq. 11.5. Here, the variance is equal to the prediction error variance under the model. We used random walk models because they were relatively simple and resulted in good fits for our tests. Note, had the results been less satisfactory, we could have considered other time series models (e.g., AR(1)). Finally, the simulated

Table 11.5 Calibration (%)		2006-2011		2011-2016		
of holdout sample predictions of Australian internal	Model		80% 95%		80% 95%	
migration by age and sex,	Log-linear	85%	96%	88%	98%	
2006–2011 and 2011–2016	Random walk	85%	95%	85%	97%	

components were combined by multiplying them together to provide realisations for the predicted migration flows by origin, destination, age and sex and for each time period. The presented prediction intervals set out below are the empirical quantiles of 1000 simulated predicted flows.

We introduce two models to forecast the inter-state migration flows: (1) log-linear forecasting of the total levels and random walk of the other components around the observed values in the last period (2001–2006), and (2) random walk of both the overall level (T) and the other components around the last observed values. To evaluate the forecasting models, we calculate the coverage of the nominal 80% and 95% prediction intervals as the percentage of the observed origin-destination-sex-age flows that lie within the intervals. These calibration statistics are presented in Table 11.5 as the percentage of the total number of observations, excluding the diagonals, where i = j, in the origin-destination-age-sex tables. While they may not provide accurate estimates of the coverage of the nominal intervals, if there is correlation between the migration flows within and/or between years, they can indicate failures in the measures of uncertainty. However, in general, we find that the calibration statistics for both intervals for both models are reassuringly close to the nominal values.

The predicted and observed levels of out-migration, in-migration and net migration for the 2006–2011 and 2011–2016 periods are presented in Fig. 11.10 for the eight states and territories in Australia. The results were obtained from Model 4 that included log-linear forecasts for the total levels and random walk forecasts for the other components. In general, we find the predicted means are close to the observed values in both periods and that the prediction intervals cover the observed values. There were, however, two notable differences between observed and estimated totals. The first is the results for NSW, where the mean level of out-migration was much higher for both the 2006–2011 and 2011–2016 periods. The other is QLD, where the predicted means of in-migration were higher than the observed values. In both cases, however, the 95% prediction interval covered the observed values. These differences can largely be explained by the unanticipated changes to the O_i and D_j components in the model observed during the 2006–2011 and 2011–2016 periods (see Fig. 11.3).

The observed and estimated female age-specific patterns of in-migration and out-migration are presented in Figs. 11.11 and 11.12, respectively, for the 2011–2016 period. During this time period, the mean number of female migrants were overestimated by around 0.7%, while the corresponding number of male migrants were overestimated by around 0.4%. We also find that the interstate migration of younger age groups, especially the 20–24 year old age group, are underestimated,



Fig. 11.10 Observed and forecasted in-migration, out-migration and net migration by state and territory in Australia, 2006–2011 and 2011–2016

Note: Values shown on the y-axis represent the counts of interstate migration measured in thousands. Error bars represent the 95% prediction intervals for the forecasted flows.

while the middle age groups are overestimated. These differences can be partially attributed to unanticipated increases in the proportions of migrants aged 20–25 years in 2006–2011 and 2011–2016 (see Fig. 11.4).

In summary, we found the multiplicative component model did well in predicting the observed patterns of migration by origin, destination, age and sex, particularly when the uncertainty in the predictions is taken into account. If this model were to be put into practice, more attention could be placed on the extrapolation of the agespecific components, especially if the aim was to reduce uncertainty in the forecasts.



Fig. 11.11 Observed and estimated age-specific female in-migration with 80% and 95% prediction intervals (in thousands) by state and territory in Australia, 2011–2016

In our illustration, we found some of the predicted age patterns differed considerably from the observed values.

In addition to the holdout sample forecasts, we applied the method described above to the whole time series of data from 1981–1986 to 2011–2016 and forecasted the internal migration tables forward for the periods 2016–2021 and 2021–2026.



Fig. 11.12 Observed and predicted age-specific female out-migration with 80% and 95% prediction intervals (in thousands) by state and territory in Australia, 2011–2016

The forecasted total number of interstate migrants is 825,915 persons in 2016–2021 with 95% prediction interval ranging between 757,295 and 894,984 persons. For the 2021–2026 period, the forecasted total number of interstate migrants increased to 835,248 persons with the 95% prediction interval ranging between 762,748 and 916,675 persons. In Fig. 11.13, we present the forecasted in-migration, out-



Fig. 11.13 In-migration, out-migration and net migration by state and territory in Australia for observed 2011–2016 flows and forecasted 2016–2021 and 2021-2026 flows

Note: Years at the horizontal axis denotes the starting years of each 5-year periods from 2011–2016 to 2021–2026. Values shown on the vertical axis represent the counts of interstate migration measured in thousands. Error bars represent the 95% prediction intervals for the forecasted flows. migration and net internal migration for each state and territory for the two periods with comparisons to the observed levels in 2011–2016. In general, we find the levels of internal migration very stable over time. NSW and QLD are forecasted to keep contributing the largest amounts of out-migration and in-migration. Finally, to illustrate the performance of the model on forecasting age-sex-specific migration flows between pairs of origins and destinations, we present the age profiles for female migrants moving between NSW and QLD, representing a major internal migration flow in the system, and SA and TAS, representing a relatively small flow, in Fig. 11.14.



Fig. 11.14 Age-specific female migration flows (in thousands) between selected states in Australia for the observed 2011–2016 flows and forecasted 2016–2021 and 2021–2026 flows Note: The areas with dark and light grey colours represent the 95% prediction intervals for the forecasted flows in 2016–2021 and 2021–2026 respectively.

11.5 Conclusion

In this chapter, we have shown how the multiplicative component projection model may be used to provide future estimates of internal migration by origin, destination, age and sex with measures of uncertainty. It extends earlier research using multiplicative or log-linear models to forecast internal migration (Stillwell 1986; Willekens and Baydar 1986; Van Imhoff et al. 1997; Van der Gaag et al. 2000; Sweeney and Konty 2002; Raymer et al. 2006; Van Wissen et al. 2008; Raymer et al. 2017) by modelling each component separately and integrating uncertainty. The methodology is relatively simple and robust. It directly provides the forecasted sizes of migration flows that can be used to construct transition probabilities for use in multiregional cohort component projection models, assuming one could also infer the probability of staying or not migrating, or aggregate them for use in standard 'single region' cohort component projection models.

Further research is needed to examine the appropriateness of the simple extrapolation method for each multiplicative component before being used in practice. In particular, it would be useful to assess the forward forecasted results with future measured values as good holdout sample results do not always ensure good outof-sample predictions. The underlying assumptions presented in this chapter are admittedly simple but our aim was to illustrate the method. Further research should investigate differently forecasting assumptions and experimentations with different data and longer time series.

In conclusion, we hope that the methodology presented in this chapter will inspire improving methods for forecasting internal migration. Internal migration has become increasingly important as a component of population change, in both developing and developed societies (White 2016). Also, many countries have internal migration flow data by origin, destination, age and sex – our research has shown how one can make better use of these data to make future predictions of internal migration. The basic argument is that migration processes evolve over time in predictable ways. By modelling the underlying structures of migration flow tables, we are able to both simplify the process of estimation as well as improve the accuracy of the forecasts.

References

- Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal migration and development: Comparing migration intensities around the world. *Population and Development Review*, 41(1), 33–58.
- Bell, M., Wilson, T., Charles-Edwards, E., & Ueffing, P. (2018). Australia: The long-run decline in internal migration intensities. In T. Champion, T. Cooke, & I. Shuttleworth (Eds.), *Internal migration in the developed world: Are we becoming less mobile?* (pp. 147–172). Abingdon: Routledge.
- Burnley, I. H. (1996). Associations between overseas, intra-urban and internal migration dynamics in Sydney, 1976-91. *Journal of the Australian Population Association*, 13(1), 47–65.

- Champion, T., Cooke, T., & Shuttleworth, I. (Eds.). (2018). Internal migration in the developed world: Are we becoming less mobile? Abingdon: Routledge.
- Cooke, T. (2013). Internal migration in decline. The Professional Geographer, 65, 664-675.
- Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, 44(2), 199–223.
- Raymer, J., Bonaguidi, A., & Valentini, A. (2006). Describing and projecting the age and spatial structures of interregional migration in Italy. *Population Space and Place*, 12(5), 371–388.
- Raymer, J., Abel, G. J., & Rogers, A. (2012). Does specification matter? Experiments with simple multiregional probabilistic population projections. *Environment and Planning A*, 44(11), 2664– 2686.
- Raymer, J., Biddle, N., & Campbell, P. (2017). Analysing and projecting Indigenous migration in Australia. Applied Spatial Analysis and Policy, 10(2), 211–232.
- Rogers, A. (1966). Matrix methods of population analysis. *Journal of the American Institute of Planners*, 32(1), 40–44.
- Rogers, A. (1968). *Matrix analysis of interregional population growth and distribution*. Berkeley: University of California Press.
- Rogers, A. (1975). Introduction to multiregional mathematical demography. New York: Wiley.
- Rogers, A. (1990). Requiem for the net migrant. Geographical Analysis, 22(4), 283-300.
- Rogers, A., & Willekens, F. (1976). Spatial population dynamics. *Papers of the Regional Science Association*, *36*(1), 3–34.
- Rogers, A., Willekens, F., Little, J. S., & Raymer, J. (2002). Describing migration spatial structure. *Papers in Regional Science*, 81, 29–48.
- Rogers, A., Willekens, F., & Raymer, J. (2003). Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer*, 55(1), 56–69.
- Smith, P. W. F., Raymer, J., & Giulietti, C. (2010). Combining available migration data in England to study economic activity flows over time. *Journal of the Royal Statistical Society: Series A* (*Statistics in Society*), 173(4), 733–753.
- Snickars, F., & Weibull, J. W. (1977). A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, 7(1–2), 137–168.
- Stillwell, J. (1986). The analysis and projection of interregional migration in the United Kingdom. In R. Woods & P. Rees (Eds.), *Population structures and models: Developments in spatial demography* (pp. 160–202). London: Allen & Unwin.
- Sweeney, S. H., & Konty, K. J. (2002). Population forecasting with nonstationary multiregional growth matrices. *Geographical Analysis*, 34(4), 289–312.
- Tobler, W. (1995). Migration: Ravenstein, Thornthwaite, and beyond. Urban Geography, 16(4), 327–343.
- Van der Gaag, N., Van Imhoff, E., & Van Wissen, L. (2000). Internal migration scenarios and regional population projections for the European Union. *International Journal of Population Geography*, 6, 1–19.
- Van Imhoff, E., Van der Gaag, N., Van Wissen, L., & Rees, P. (1997). The selection of internal migration models for European regions. *International Journal of Population Geography*, 3, 137–159.
- Van Wissen, L., Van der Gaag, N., Rees, P., & Stillwell, J. (2008). In search of a modelling strategy for projecting internal migration in European countries. In J. Poot, B. Waldorf, & L. Van Wissen (Eds.), *Migration and human capital* (pp. 49–74). Cheltenham: Edward Elgar.
- White, M. J. (Ed.). (2016). International handbook of migration and population distribution. Dordrecht: Springer.
- Willekens, F. (1983). Log-linear modelling of spatial interaction. Papers of the Regional Science Association, 52, 187–205.
- Willekens, F., & Baydar, N. (1986). Forecasting place-to-place migration with generalized linear models. In R. Woods & P. Rees (Eds.), *Population structures and models: Developments in spatial demography* (pp. 203–244). London: Allen & Unwin.
- Willekens, F., & Rogers, A. (1978). Spatial population analysis: Methods and computer programs (Research Report 78-18). Laxenburg: International Institute for Applied Systems Analysis.

- Wilson, A. G., & Rees, P. H. (1974a). Accounts and models for spatial demographic analysis 2: Age-sex disaggregated populations. *Environment and Planning A*, 6(1), 101–116.
- Wilson, A. G., & Rees, P. H. (1974b). Population statistics and spatial demographic accounts. Journal of the Royal Statistical Society, Series D (The Statistician), 23(3–4), 229–257.
- Wilson, A. G., & Rees, P. H. (1975). Accounts and models for spatial demographic analysis 3: Rates and life tables. *Environment and Planning A*, 7(2), 199–231.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

