ALTERNATIVE APPROACHES TO FORECASTING MIGRATION: FRAMEWORK AND UK ILLUSTRATIONS

Philip Rees[11], Nikolas Lomax[12], Peter Boden[23]

[1]Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds LS2 9JT
[2]Edge Analytics Ltd, Leeds Innovation Centre, 103 Clarendon Road, Leeds LS2 9DF

February 2015

Summary

This paper is a review of the migration component of population projection models. A general population accounting framework is defined that underpins projection models. The framework identifies migration variables internal to a country, international migration to and from a country and its constituent regions and international migration between other countries. These different types of migration can be represented in projection models as flows (migration numbers), flows projected by a time series model, migration transmission rates multiplied by the origin population at risk, migration admission rates multiplied by the destination population at a risk, or through an explanatory model. The arguments for and against each migration projection model are discussed through an analysis of 16 projection examples linked to published studies. The importance of understanding the forces affecting migration at origin and destination is stressed. Simulation experiments are carried out for a UK population disaggregated by the ethnicity, testing model results against 2001-2011 estimates. The paper shows how a proper understanding of the spatial context is needed for successful population projections.

KEYWORDS: Population Projection Models, Accounting Frameworks, Internal Migration, International Migration

1. INTRODUCTION

Population projections are carried out by international agencies (e.g. UN 2014, World Bank 2014), demographic research centres (e.g. Lutz et al 2014), national statistical offices (e.g. ONS 2013 for the UK and Home countries; ONS 2014a for subnational areas), local authorities (e.g. GLA 2014 for London Boroughs) or academic teams (e.g. Rees et al 2011 for local authorities in England by ethnicity). These projections figure significantly in policy debates on climate change, food crises, demographic ageing, ethnic transitions, regional development, planning and housing.

The research question posed in this paper is "which is the best migration model to use in a population projection?" Of course, this is an impossible question to answer without knowing the objectives of the projection exercise, the nature of the populations being studied and the character of the migration data available. So instead we evaluate the range of alternative methods available. Along the way we point to some ingredients that should be included in the migration models for projection: full understanding of migration data available, the need to embed migration variables within a population accounts framework,

[1] p.h.rees@leeds.ac.uk
[2] n.m.lomax@leeds.ac.uk
[3] pete@edgeanalytics.co.uk

and testing which migration model has performed best against a historical series, with attention to the limits within which we can be confident future migration flows lie.

Except at the very smallest spatial scales, the usual projection method used is the cohort-component model. This model tracks by age and sex the changes to the population due to births, deaths, in-migration flows and out-migration flows. Age is a vital ingredient because fertility, mortality and migration components vary in intensity across the life span, in different ways. Sex is important because there are differences between men and women in their exposure to mortality and migration risks. Most fertility analyses use female populations at risk. It is always useful to distinguish between internal migration between areas within a country and international migration between countries. This is because these two migration flows have different drivers and constraints.

Demography has paid relatively little attention has been paid to the proper measurement and projection of migration. Academics have been good at building theories of migration such as the mobility transition (Zelinsky 1971), the cascading hierarchy of counter-urbanization migration flows (Champion 1998) or spatial interaction models of migration flows (Stillwell and Congdon 1991, Champion et al 2002). Very little of this has been incorporated into projection methodology.

The aim of this paper is to fill this gap in knowledge by reviewing the ways in which migration has been represented and modelled in population projections, assessing their advantages and disadvantages. We first outline the framework of population accounts within which migration must be embedded (section 2). We then consider the choice between representing migration as a flow of people or as a population at risk multiplied by a suitable migration rate (section 3). Section 4 reviews systems of interest in population projection, distinguishing migration flows internal to the system from those that are external. Section 5 shows how the different migration models can be used in single region, bi-regional, multi-regional and multi-country projection models. Section 6 illustrates the impact of migration model choice using a cohort-component model for the four countries of the UK. A final section discusses choice of a "best" migration model and what factors need to be considered.

2. A POPULATION ACCOUNTING FRAMEWORK

Population accounts are frameworks for assembling the components of change within which migration data are located in a consistent way. Migration can be measured in censuses and surveys as *transitions* over (1) a fixed time interval ("Where did you live one year ago?"), (2) over an uncertain interval ("Where did you live before your present residence?") or (3) over a person's lifetime ("Where were you born?). Migration can also be measured as (4) *moves* in registration systems ("What was your previous registered residence?"). Measures based on census or survey questions are subject to survivor bias and estimates must be made of non-surviving migrants. Registration systems capture the migration event as it occurs. It is essential that type of migration measure and design of projection model are matched (Rees 1985). Only measures (1) and (4) are easy to use in projection models.

Migration statistics must be placed within the context of population change along with births and deaths. The essential feature of demographic accounts is that they must count all people or events that leave the population of interest and count all people or events that enter the population of interest. Here we focus only on demographic accounts based on moves, because this is the standard adopted by National Statistics (ONS 2012).

In Table 1 is set out a population accounting framework including migration data of the movement type. Although no projection will be as broad in scope as this framework, we later explain how aggregations or sub-sets underpin projection models in practice. The table rows list the regions in a country of interest and a set of countries in the rest of the world which send migrants to the regions. The set of regions and set of countries together cover the whole world. A final row holds births which add new-born infants to the regional and national populations. The regions and countries appear also in the columns of the table as destinations of the migrations and births. A final column contains the exits to death from the regions and countries.

Entries in the table are migration border-crossing moves and life state moves for births and deaths. An individual can experience only one birth or one death in a time interval but can make many migrations. This feature distinguishes movement accounts from transition, in which only one transition can occur in a time interval. We focus subsequent discussion on migration only. The migration part of Table 1 is divided into four quadrants holding different migration variables, about which decisions must be made in designing population projection models. The *top left quadrant* holds migration variables, *M*, that count migrations between *n* regions within the country of interest. There should be at least two regions – where there is only one region, the population accounts refer to a single country. In the UK projections are routinely carried out for around 400 lowest tier local authorities but separately in each home country[4]. The superscripts *r1r2* in variable M^{r1r2} indicate the migrations that take place between region *r1* and region *r2*. Counts of number of migrations (moves) between regions are found in the cells off the principal diagonal. The variables in the principal diagonal are residual counts derived from the accounting relationships for the table rows (see the Table 1 notes). The migration variables in the *top right quadrant* are labelled *E* and represent emigration, migration out of a country to other countries. The superscripts, e.g. *r1c2*, indicate region of origin and country of destination. The migration variables in the *bottom left quadrant* refer to immigration – migration into a region within the country of interest from another country, for example, from country *m* to region *n*, I^{cmrn}. The *bottom right quadrant* completes the set of migration variables, holding counts of migration between origin and destination countries besides the country of

[4] The Home Countries are England, Wales, Scotland and Northern Ireland.

Table 1: Population accounting framework incorporating migration (moves) data

To	Country 1 of interest				Other countries				
From	Region r1	Region r2	…	Regionr n	Country c2	…	Country cm	Deaths	Totals
Region r1	R^{r1}	M^{r1r2}	…	M^{r1rn}	E^{r1c2}	…	E^{r1cm}	D^{r1}	P_t^{r1+}
Region r2	M^{r2r1}	R^{r2}	…	M^{r2rn}	E^{r2c2}	…	E^{r2cm}	D^{r2}	P_t^{r2+}
:	:	:		:	:		:	:	:
Region rn	M^{rnr1}	M^{rnr2}	…	R^{rn}	E^{rnc2}	…	E^{rncm}	D^{rn}	P_t^{rn+}
Country c2	I^{c2r1}	I^{c2r2}	…	I^{c2rn}	R^{c2}	…	M^{c2cm}	D^{c2}	P_t^{c2+}
:	:	:		:	:		:	:	:
Country cm	I^{cmr1}	I^{cmr2}	…	I^{cmrn}	M^{cmc2}	…	R^{cm}	D^{cm}	P_t^{cm+}
Births	B^{r1}	B^{r2}	…	B^{rn}	B^{c2}	…	B^{cm}	0	B^{+}
Totals	P_{t+1}^{+r1}	P_{t+1}^{+r2}	…	P_{t+1}^{+rn}	P_{t+1}^{+c2}	…	P_{t+1}^{+cm}	D^{+}	T^{++}

Notation

P	Population	r1, r2, … , rn	subscripts for particular regions
R	Residual balance	c2, … , cm	subscript for rest of country
M	Migration (moves)	n	number of regions
I	Immigration (moves)	m	number of countries
E	Emigration (moves)	+	summation over subscript replaced
B	Births	t	time at start of interval
D	Deaths	t+1	time at end of interval
T	Total flows	ri, rj, ck, cl	general subscripts for regions and countries

Notes

Accounting relationships for rows in Table 1 (example for region ri) are:

$P_t^{r1+} = R^{r1} + M^{r1r2} + \cdots + M^{r1rn} + E^{r1c2} + \cdots + E^{r1cm} + D^{r1}$ (1)

$R^{r1} = P_t^{r1+} - (M^{r1r2} + \cdots + M^{r1rn}) - (E^{r1c2} + \cdots + E^{r1cm}) - D^{r1}$ (2)

Accounting relationships for columns in Table 1 are:

$P_{t+1}^{+r1} = R^{r1} + M^{r2r1} + \cdots + M^{rnr1} + I^{c2r1} + \cdots + I^{cmr1} + B^{r1}$ (3)

With knowledge of opening and closing populations and natural increase, net migration can be derived as a residual:

$N^{ri} = (P^{+ri} - P^{ri+}) - (B^{ri} - D^{ri})$

This derivation involves summation of both internal and external flow balances:

$$N^{ri} = (\sum_{rj \neq ri} M^{rjri} - \sum_{rj \neq ri} M^{rirj}) + (\sum_{ck} I^{ckri} - \sum_{ck} E^{rick})$$

The term "region" is used to indicate a territorial unit in general rather than a particular territorial unit such as "Standard Region", "Government Office Region" or "English Region".

interest (e.g. M^{c2cm}, migrations from country *2* to country *m*,). As in the top left quadrant, residual terms appear in the diagonal.

It is a requirement for constructing population accounts that the whole world is covered. However, it is usual in projection models to lump places outside the country of interest together as the "Rest of the World". Although this aggregation is forced on the analyst by the poverty of international migration statistics, this is unsatisfactory as both policy makers and citizens have a considerable interest in the origins of immigrants and the destinations of emigrants.

The four quadrants of the accounts differ in terms of the source of the migration data. In the case of the UK, within country migrations are estimated from counts in the National Health Service (NHS) Patient Register(s), supplemented by information from the Higher Education Statistics Agency (HESA) on student entry to and exit from higher education institutions and locations. Some information is also used on migrations between the Armed Forces and prisoner populations and civilian locations. The immigration variables are estimates based on the national Long Term International Migration (LTIM) totals, which use data from the International Passenger Survey (IPS) (migrants intending to stay for 12 months or more), the Home Office (Asylum Seekers), HESA (foreign student immigration), the Labour Force Surveys of the UK and the Republic of Ireland. These national totals are distributed to local areas using data on new NHS registrations with a previous address abroad, new National Insurance registrations with a previous address outside the UK (Department of Work and Pensions), HESA (foreign student registrations) and the Ministry of Defence (migrations from overseas military bases to the UK (ONS 2011). Emigration variables are estimated using a time series model based on a set of determinants and proxies for emigration, such as immigration in the previous year and total internal out-migration, constrained to national total estimates based on the IPS (ONS 2010).

No reliable disaggregation of immigration or emigration by country of origin or destination can be made, beyond broad groupings (European Union, Australia-Canada-New Zealand, South Asia). Recent research combined IPS information with NHS, DWP and HESA administrative data using Bayesian statistical methods (Disney 2014). Immigration visa statistics could also be used. The Home Office has been working to develop a system for recording entry and exit via electronic passport records (Project Semaphore, the e-Borders project) though there have been difficulties which have delayed its implementation (Wikipedia 2014). The system, when developed, will record trips into and out of the UK rather than migrations (which make up about 1% of trips), so further information will be needed to produce comprehensive international migration statistics on origins and destinations for the UK.

Statistics on migration between countries are based on data gathered from national statistical agencies by international statistical agencies such as the United Nations, the World Bank and Eurostat. There are considerable challenges in harmonizing these data, which have been tackled by academic teams for inter-country migration with Europe (see NIDI 2011for the MIMOSA database of migration flows between European countries, see CPC 2013 for the similar IMEM database) and for all countries in the world by the Wittgenstein Centre (Abel 2013, Abel and Sander 2014, Sander, Abel and Bauer 2014a). The European country estimates rely heavily on population register data collected by a majority of European countries. It is necessary to resolve differences in the estimate of flows by origin and destination countries. The global migration flow estimates rely on use of time series of tables of lifetime migration from censuses and surveys. These are tables of country of residence against country of birth/citizenship. An innovative model deduces migration flows for five year intervals from successive censuses, supplemented with a gravity model to fill in missing migration flows.

3. MODELS OF MIGRATION FLOWS FOR POPULATION PROJECTIONS

The Table 1 framework is useful for organising historic statistics on populations and migrations but for future values of the migration flows we need models of migration. Table 2 sets out alternative approaches.

The *first approach* (Table 2, section 2.1) assumes that recent migration flows continue into the future. This is not a good model because it ignores variation in the populations generating the migrations. In addition it fails to consider changes in migration due to economic cycles or secular trends in migration rates. Rates in some developed countries may be in decline, controlling for population size and ageing.

However, if migration flows are subject to government control, then there is a need to project changes in migration. This is the basis of the *second approach* (Table 2, section 2.2) which assumes a trajectory of migration flows. This approach has been used, in particular, to model the migration flows in the emigration and immigration quadrants. Formal time series models may be employed with greater weight given to recent years and lesser weight to distant years. Usually, some plausible limit is set beyond which the first approach is employed. Judgement by statisticians or by a pool of experts is often used to set the trajectory of change in migration and its upper or lower limit. It is claimed, though rarely through assembly of evidence, that immigration policy limits immigration. To a frustrated intending immigrant or a failed asylum seeker it certainly feels like that. The 2010-15 Coalition Government policy of setting a target of less than 100 thousand for net immigration in the UK can be regarded as a failure, given that net international migration has averaged 208 thousand per annum for 2010-2013 (ONS 2014b). However, set against a projection model design which lets the growing population of the Rest of the World determine future migration, the policies of the Coalition Government might be judged successful.

Table 2: Models of migration flows used in population projections

2.1 Assume flows continue at historic levels (h =time horizon of projections)	
$M_{t=0}^{rirj} = M_{t=1}^{rirj} = \cdots = M_{t=h}^{rirj}$ $E_{t=0}^{ricl} = E_{t=1}^{ricl} = \cdots = E_{t=h}^{ricl}$ $I_{t=0}^{ckrj} = I_{t=1}^{ckrj} = \cdots = I_{t=h}^{ckrj}$ $M_{t=0}^{ckcl} = M_{t=1}^{ckcl} = \cdots = M_{t=h}^{ckcl}$	
2.2 Assume flows vary according to judgment or a time series model	
$M_{t=0}^{rirj}, M_{t=1}^{rirj}, \ldots, M_{t=h}^{rirj}$ $E_{t=0}^{ricl}, E_{t=1}^{ricl}, \ldots, E_{t=h}^{ricl}$ $I_{t=0}^{ckrj}, I_{t=1}^{ckrj}, \ldots, I_{t=h}^{ckrj}$ $M_{t=0}^{ckcl}, M_{t=1}^{ckcl}, \ldots, M_{t=h}^{ckcl}$	
2.3 Assume flows are rates multiplied by a population at risk	
Transmission rates, time constant or varying input to a projection model	
$tm^{rirj} = M^{rirj}/PAR^{ri}$	$M^{rirj} = tm^{rirj} \times PAR^{ri}$
$te^{ricl} = E^{ricl}/PAR^{ri}$	$E^{ricl} = te^{ricl} \times PAR^{ri}$
$ti^{ckrj} = I^{ckrj}/PAR^{ck}$	$I^{ckrj} = ti^{ckrj} \times PAR^{ck}$
$tc^{ckcl} = M^{ckcl}/PAR^{ck}$	$I^{ckcl} = tc^{ckcl} \times PAR^{ck}$
Admission rates, time constant or varying input to a projection model	
$am^{rirj} = M^{rirj}/PAR^{rj}$	$M^{rirj} = am^{rirj} \times PAR^{rj}$
$ae^{ricl} = E^{ricl}/PAR^{rj}$	$E^{ricl} = ae^{ricl} \times PAR^{rj}$
$ai^{ckrj} = I^{ckrj}/PAR^{cl}$	$I^{ckrj} = ai^{ckrj} \times PAR^{cl}$
$ac^{ckcl} = M^{ckcl}/PAR^{cl}$	$M^{ckcl} = ac^{ckcl} \times PAR^{cl}$
Populations at risk	
$PAR^{ri} = ½(P_t^{ri} + P_{t+1}^{ri})$	$PAR^{rj} = ½(P_t^{rj} + P_{t+1}^{rj})$
$PAR^{ri} = ½(P_t^{ri} + P_{t+1}^{ri})$	$PAR^{rj} = ½(P_t^{rj} + P_{t+1}^{rj})$
$PAR^{ck} = ½(P_t^{ck} + P_{t+1}^{ck})$	$PAR^{cl} = ½(P_t^{cl} + P_{t+1}^{cl})$
$PAR^{ck} = ½(P_t^{ck} + P_{t+1}^{ck})$	$PAR^{cl} = ½(P_t^{cl} + P_{t+1}^{cl})$
2.4 Assume an explanatory model for flows (e.g. a spatial interaction model)	
$M^{rirj} = f\left(w_O^{ri}, w_D^{rj}, c^{rirj}, K\right)$	
where w_O^{ri} is a vector of origin determinants of out-migration/emigration w_D^{rj} is a vector of destination determinants of in-migration/immigration c^{rirj} is matrix of variables estimating the "costs" of migration K' is a set of constraints on the model predictions, which vary depending on the context the model is used in	
2.5 Definitions	
tm = transmission migration rate between regions te = transmission emigration rate from regions ti = transmission immigration rate to regions tc = transmission migration rate between countries am = admission migration rate between regions ae = admission emigration rate from regions ai = admission immigration rate to regions ac = admission migration rate between countries	

The *third approach* (Table 2, section 2.3) is to adopt a model that projects migration as the product of migration rate, for a recent period or projected into the future, multiplied by a population at risk, generated by the projection model. Two kinds of rates can be used: the first uses as its denominator the population at risk of the origin region (within the country) or of the origin country in the rest of the world. This we call the *transmission* migration rate. The second kind of rate uses the population at risk of the destination region (within the country) or of the destination country in the rest of the world. This we call the *admission* migration rate. Use of transmission rates of migration implies that the capacity of the destination to absorb in-migration or immigration is not constrained, for example, by housing availability or by visa quotas. The use of admission rates implies there is some control of in-migration or immigration to the capacity of the receiving population. In particular situations, we can observe capacity constraints at work. In the UK Higher Education (HE) sector, quotas of places for home students at each university were set centrally for each university. So, a projection (Rees 1986) using transmission rates over-projected the number of UK students entering HE in the late 1980s and early 1990s on the basis of trends in the 1980s because the system was constrained at University destinations. No quotas were applied to students from outside the UK and between 2004 and 2013 study was the main reason for migration for 28% of immigrants (ONS 2014b). A recurring theme in much projection work is the need to understand migration streams by reason for migration versus the paucity of information about the numbers in each reason group by country. So a careful choice must be made between transmission and admission rates in designing models for projecting migration.

To compute migration rates requires a population at risk (PAR) denominator to match the migration numerator. The ideal PAR is the total person-years of risk in the region or country (per unit of time) but this needs to be sourced from longitudinal population registers which record residence spells in regions in the time interval. An approximation to this PAR measure is the average of start and end populations in the interval (Table 2, section 2.3). In the UK the historic series of rate measures are computed using mid-year populations, which over short intervals are close to the average population. How is the future average population in an interval computed, if the end population is unknown? The most general way is to set up a simple iterative algorithm. In the first iteration the start population is used as the first estimate of the PAR. An end population is then projected. An average PAR can now be computed and the end population is projected for a second time. The computations continue until the PAR converges on a stable value. Normally, this only takes a few iterations. The alternative is to work out an analytical version of the PAR so that the projection computations only need to be carried out once. The disadvantage is that a different analytical specification is needed depending on the migration model used.

In Table 2 , section 2.3 are set out the migration rates used in projection in transmission (t) form (labelled tm, te, ti and tc) and in admission (a) form (labelled am, ae, ai and ac) where m refers to between region migration, e to emigration from regions, i to immigration to regions and c to between country migration.

To obtain projected migration flows the migration rates are multiplied by the matching PAR of origin region for transmission and of destination region for admission.

Table 2, section 2.4 proposes a final choice in the menu of migration models: models of migration in which flows are projected using the underpinning determinants, specified in a general way in the table. This choice has only rarely been used in a projection context for two reasons: first, the analyst needs to project the determinants into the future and this presents additional challenges; second, in many cases, a purely demographic model based on recent trends usually performs better than an explanatory model. However, such an explanatory model is needed if the aim is to explore the consequences for migration of regional and country employment, housing and social developments or plans. The elements that enter into such an explanatory model consist of a vector of origin attributes, a vector of destination attributes and a vector of factors which impede or expedite migration between origins and destinations. In some contexts it may be necessary constrain the migration model predictions to external forecasts or controls. A model of migration between former health authority areas developed for the Office of Deputy Prime Minister by a Newcastle-Leeds research team (Champion et al 2002) tests a very large set of determinants, using a two stage model. Stage one predicted the total out-migration from the origins; stage two distributed these projected origin totals across destination regions. The coefficients of both parts of the model are specific to origins, so this is an example of a model where the parameters vary by geographical areas.

4. SYSTEMS OF INTEREST

The population accounts of Table 1 contain both migration flows within countries (domestic migration) and between countries (international migration). However, populations and flows of interest and those not of interest may not coincide with this domestic and international distinction. Figure 1 draws a distinction between internal flows between areas for which the populations are projected and external flows where the populations of source or destination areas for external flows are not projected. The examples in Figure 1 show how the distinction between internal and external flows can overlap international boundaries.

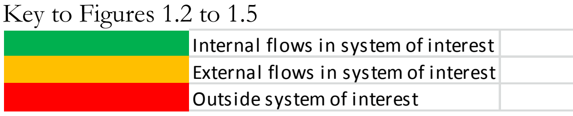
Figure 1.1 provides text labels for the algebraic variables in Table 1. Figure 1.2 shows a system of interest that is just a single region (or a single region repeated a number of times). Internal flows (in green) are confined to the region of interest (birth and death flows). External flows (in orange) are in-migration and immigration streams into the region and out-migration and emigration streams out of the region. Normally these are aggregated if the only migration information available consists of residual net migration. Completely outside the system of interest are flows between other areas (shown in red).

Figure 1: Examples of systems of interest distinguishing internal and external migration flows

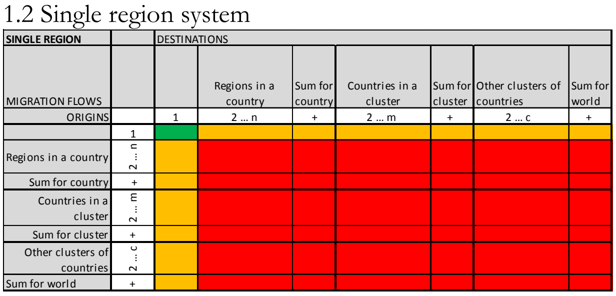
1.1 Types of migration flows between regions and countries

GENERAL		DESTINATIONS						
MIGRATION FLOWS			Regions in a country	Sum for country	Countries in a cluster	Sum for cluster	Other clusters of countries	Sum for world
ORIGINS		1	2 … n	+	2 … m	+	2 … c	+
	1							
Regions in a country	2 … n		Flows between regions in a country		Flows to countries in a cluster		Flows to countries in other clusters	
Sum for country	+							
Countries in a cluster	2 … m		Flows from countries in a cluster		Flows between countries in a cluster		Flows to other clusters to countries in a cluster	
Sum for cluster	+							
Other clusters of countries	2 … c		Flows from countries in other clusters		Flows from other clusters to countries in a cluster		Flows between countries in other clusters	
Sum for world	+							

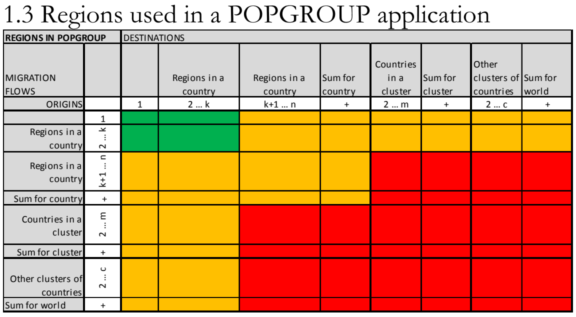
Key to Figures 1.2 to 1.5



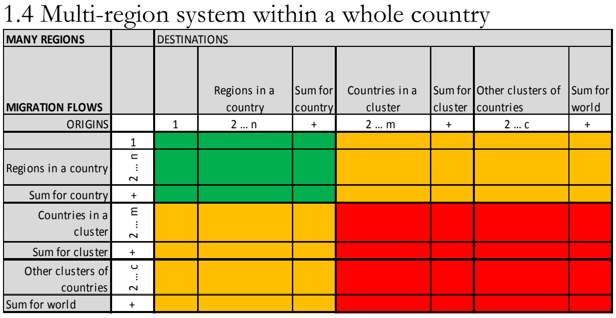
1.2 Single region system



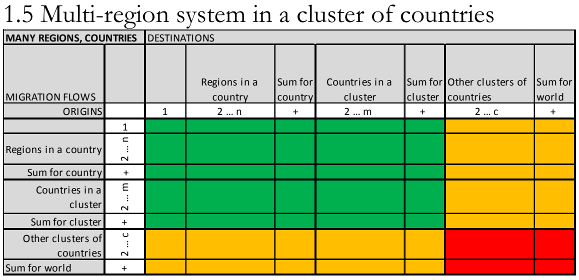
1.3 Regions used in a POPGROUP application



1.4 Multi-region system within a whole country



1.5 Multi-region system in a cluster of countries



1.6 Multi-country system in all clusters of countries

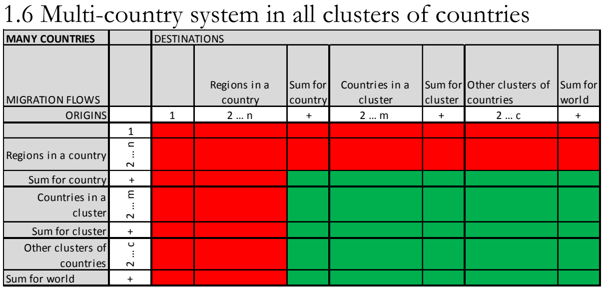


Figure 1.3 shows a system of interest consisting of a set of regions within a country between which internal flows are of interest. However, the set of regions do not cover the whole national territory. This system of interest is often constructed by users of the POPGROUP software for population projection (LGA 2014), where the future population of a local authority of interest will be dominated by flows to and from neighbouring local authorities. The local authorities in the rest of the country can be treated as a single combination. The difficulty of such a system of interest is that it does not automatically generate the populations at risk for areas that strongly or weakly interact with the local authority of interest. Information is often borrowed from a bigger sub-national projection (e.g. ONS 2014a) to supply the populations at risk to be used in migration rate models. However, there will be issues of consistency between the local projections and the official sub-national projections to be resolved.

Figure 1.4 shows a system of interest which divides a country into a number of regions and models the flows between them simultaneously. External flows to or from these regions from other countries complete the system. The alternatives for modelling these external flows are discussed in the next section of the paper. Flows between other countries (highlighted in red) are ignored. The official sub-national projections for local authorities in England adopt such a multi-regional system (ONS 2014a).

Figure 1.5 extends the multi-region system at regional scale to countries which belong to a cluster with special arrangements for international migration, such as the freedom of labour migration which is a key foundation of the European Union. Such a system was used in the DEMIFER project (ESPON 2010, Rees et al 2012) that developed reference and policy scenario projections for European Union regions at NUTS2 level.

Figure 1.6 concentrates on country level populations, dropping the regional level, but for the whole world (Sander et al 2014b, Lutz et al 2014). All migration flows between countries are internal to the system of interest (highlighted in green) and regional flows are outside the system (highlighted in red).

5. PUTTING THE INGREDIENTS TOGETHER

In Table 1 we outlined the general accounting framework for population projection and in Table 2 we specified the various ways in which migration could be modelled/projected. In practice, projection models will only use part of the Table 1 framework which covers the whole world and may use different models to project migration in the different quadrants. Table 3 sets out a range of framework-model combinations that have been used recently. Figure 2 gives graphic representation to this range, as an aid to description. This menu, as in a restaurant, is not exhaustive but sets out a wide range of models, from which a projection analyst can choose. We set out the advantages and disadvantages of each migration model to help in this choice.

Table 3: A classification of selected migration models used in projection

Classes	Territorial units	Representation of migration: Internal	Representation of migration: International
A	**Uni-regional models**		
A1	One region	Net flows (combined)	Net flows (combined)
A2	One region	Net rates (combined)	Net rates (combined)
B	**Bi-regional models**		
B1	Two regions	Out-migration (transmission) rates	Immigration and emigration flows
B2	Two regions	Out-migration (transmission) rates	Immigration flows and emigration (transmission) rates
B3	Two regions	Out-migration (transmission) rates	Immigration (admission) rates and emigration (transmission) rates
B4	Two regions	Out-migration (transmission) rates	Emigration (transmission) rates for rest of world; emigration (transmission) rates for regions
C	**Multi-regional models**		
C1	Many regions	Out-migration (transmission) rates	Immigration flows and emigration flows
C2	Many regions	Out-migration (transmission) rates	Immigration flows and emigration (transmission) rates
C3	Many regions	Out-migration (transmission) rates	Immigration (admission) rates and emigration (transmission) rates
C4	Many regions	Out-migration (transmission) rates	Emigration (transmission) rates for rest of world; emigration (transmission) rates for regions
D	**Multi-country models**		
D1	Many countries	Not applicable	Net migration flows
D2	Many countries	Not applicable	Net migration rates
D3	Many countries	Not applicable	Inter-country migration (transmission) rates
D4	Many regions and countries	Out-migration (transmission) rates	Inter-country migration (transmission) rates; immigration and emigration flows

Figure 2: Graphical representation of the migration models

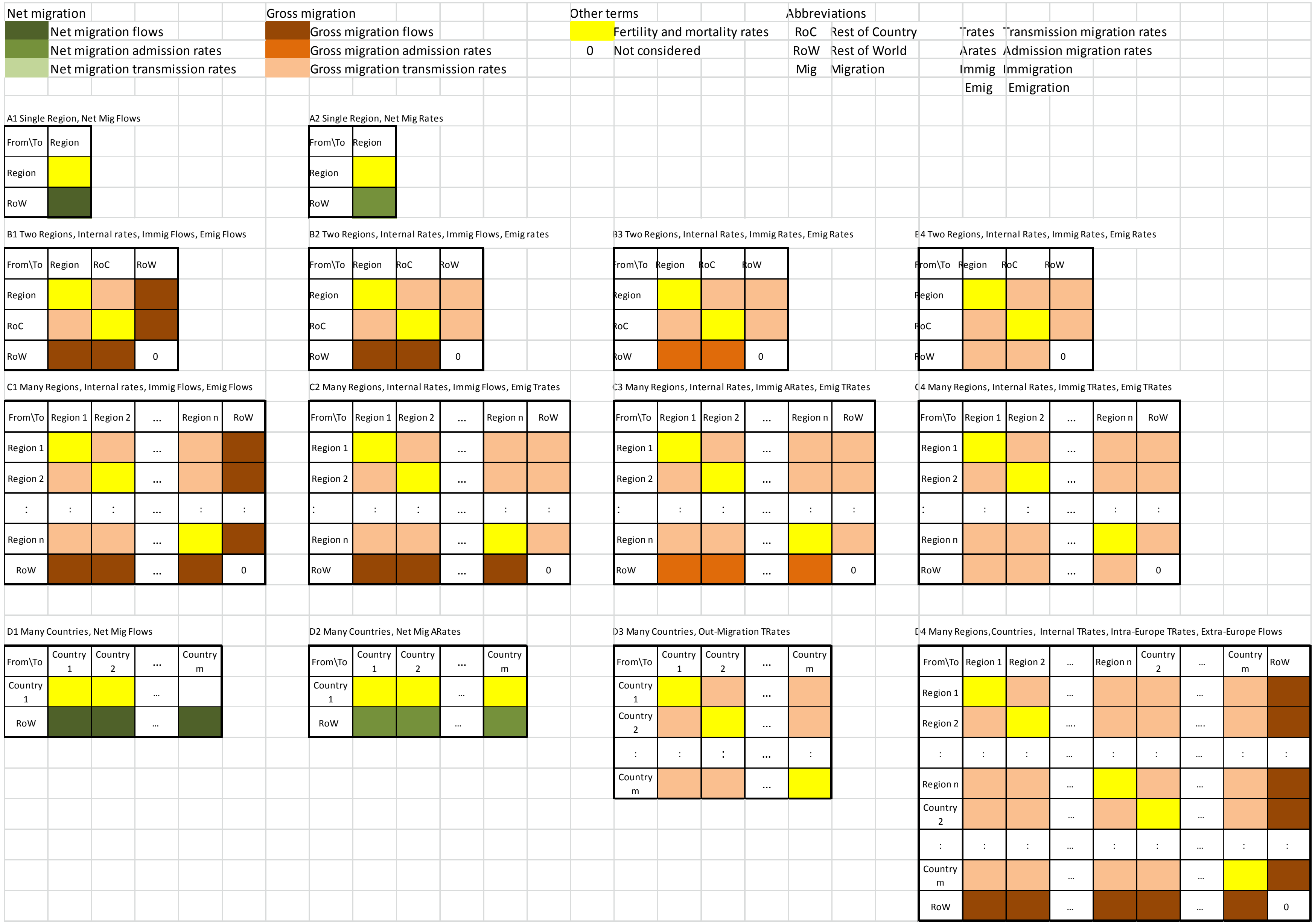


Table 3 is organised into four sets of models: *uni-regional models* (set A), in which only one region is of interest; *bi-regional models* (set B), consisting of a region and the rest of the country; *multi-regional models* (set C) in which all regional populations in a country are projected simultaneously; and *multi-country models* (set D) in which many countries are projected together.

5.1 Uni-regional models

Single region models (A1, A2) use net migration flows or rates in the absence of any better information on migration. These models are appropriate when net migration is derived as a residual (see the notes to Table 1). Net migration is the combination of sets of internal and external inflows and outflows, which could be trending in different directions. It is better to estimate historically and to project into the future the four gross migration flows rather than just one net migration figure. Model A1 assumes a trajectory of net migration flows, while model A2 projects the net migration flow as a net admission migration rate multiplied by the single region population. The trajectory of net migration flows or trajectory of net admission rates may be constant or time varying.

5.2 Bi-regional models

If the uni-regional model is rejected as inadequate or inconsistent, the next alternative is to model the population of the region of interest along with that of the rest of the country, because country level data on births, deaths and international migration are readily available. Internal migration can be modelled by multiplying transmission migration rates by origin region PARs, capturing the linkages between the regions. This method is used in the four types of bi-regional model identified in Table 3.

The B1 model uses trajectories of immigration and emigration flows. If this choice is made, it implies that both immigration and emigration are constrained at the destination, normally at country level. For example, immigration into the UK is controlled by national and international laws. Entry is managed through a visa system which differentiates potential immigrants by nationality (UK, EU and other) and by main reason for immigration (for work, for education, to join family members, for marriage, to seek asylum). Permitted lengths of residence are stipulated. Work permits favour occupations where there is a shortage of labour or which are important for international business. These controls serve to limit immigration and reflect past policy decisions. Policy is also framed in terms of targets but, in the absence of new laws, administrative measures are used such as delays in processing or raising the charges for visas.

The origin country has little influence on emigration, except in times of conflict when affected population groups may be forced to leave. Current concerns about British citizens emigrating to join Jihadi terrorist groups have led to some restrictions. More control on emigration is exerted by the destination countries but there are a large number of possible destinations which weaken the effect of controls in any one. This perspective leads to the second international migration model in which emigration flows are modelled as the product of emigration transmission rates multiplied by the origin region or origin country population. In the ETHPOP population projections (Rees et al. 2011, ETHPOP 2014), this migration model produced increasing emigration flows for ethnic minority populations as these grew rapidly. With immigration flows fixed (to correspond with official net international migration assumptions in the 2008-Based National Population Projections), this meant shrinking net international migration and lower population growth in the B2 model than in the B1 model.

Which model should be preferred? The ETHPOP team gives different advice, recommending model B1, model B2 or an average on different occasions. An issue is the extent to which immigration raises emigration through return migration to a country of origin. ONS (2010) built in a dependence of emigration from England's local authorities on immigration in previous years into its model of sub-national emigration, specific to the main countries of origin. Considerable return migration to the West Indies has occurred at ages around retirement, but it is not yet clear whether this return migration will also occur for South Asian groups. For one immigrant group, students entering higher education, return migration is a condition of their visas, though many are able to gain leave to remain through government programmes such the Fresh Talent initiative in Scotland (The Scottish Government 2008), which has been extended to the whole UK. The argument for the B1 model is that it follows National Statistics practice and projections, while model B2 has yet to be verified as better.

The third two region model, B3, uses immigration admission rates with emigration transmission rates. The argument for immigration admission rates is that, as a country's population increases, so its demand for immigrant labour and its capacity to teach international students expands. However, this argument is untested.

The final two-region model, B4, which treats immigration as a product of emigration rates from the Rest of the World multiplied by Rest of the World populations, is also untested. However, the model is inherently unlikely given the controls applied by destination countries. On the other hand, it is a model that indicates how immigration flows could potentially increase without controls and so provides a benchmark against which the impact of current policy can be measured.

5.3 Multi-regional models

All of the multi-regional models (Table 3, type C; Figure 2, third row of tables) project migration flows for all regions within a country simultaneously, using internal out-migration rates and origin region populations. This methodology has been widely adopted since development between 1968 and 1986 by Rogers and colleagues (see Rogers 1995 for a summary of the methods). The four models differ in the way they handle international migration. The alternatives follow the B1 to B4 sequence of two-region models; the discussion and arguments in section 5.2 apply as well to multi-regional models, so they are not repeated here. Multi-regional models are used by the Office for National Statistics for projecting local authority populations in England (ONS 2014a). The UK National Population Projections (NPP) up to the 2012-based projections consisted of four single region projections for each home country (England, Wales, Scotland, Northern Ireland) linked by a matrix of flows (ONS 2013). Bijak (2012), in a report on how the migration assumptions of the NPP might be improved, suggested that a multi-regional cohort-survival model with explicit inter-country migration rates be adopted. ONS (2014c) reports that this recommendation will be implemented in the 2014-based NPP, though empirical investigation led to the rejection of Bijak's recommendation of model C2 for handling international migration because it did not fit recent experience and did not generate plausible future international migration flows.

5.4 Multi-country models

Population projections for sets of countries are undertaken by international agencies (e.g. UN 2014), national statistical agencies (e.g. US Census Bureau 2014), population research centres (e.g. Kupiszewski 2013; Lutz et al 2014) and multi-national research teams (ESPON 2010, Rees et al 2012). In Table 3 and Figure 2, four illustrative models are presented. D1 is used in UN 2014. D3 is used in the medium term in the world country projections (Lutz et al 2014) after 2050 the UN 2014 model of convergence of net migration flows to zero in all countries is used. Kupiszewksi 2013 uses model D3 for European flows and model D1 for extra-Europe flows. Model D4 is used in ESPON (2010), which projects regional and national populations simultaneously for Europe using a model developed by Kupiszewski and Kupiszewska (2011), though extra-European migration is handled using model C3. This last example corresponds most closely with the general accounting framework of Table 1.

Model D1 is exemplified by the latest UN Population Division projections of the populations of UN Member States (UN 2014). Virtually all attention in the projection methodology is given to the fertility and mortality components with medium, high and low variants. Uncertainty is handled by developing probabilistic projections centred on the medium projection using Bayesian methods applied to the UN estimates time series. Virtually, no attention is given to international migration. Migration is not considered in framing the variants or probabilistic projections: one short-term trajectory of net migration balances is assumed for each country and balances are converged to zero for all countries in the long run, without much justification. The idea is probably that over time all countries will reach similar levels of development and the incentive to move from poorer countries to richer countries will disappear. The international migration literature offers little support for such a convergence hypothesis. In practice, international migration makes little difference to countries in the Global South (less developed states) where natural increase is the dominant component contributing to continuing though diminishing growth. In the Global North (more developed states) where natural increase is low or negative, net immigration makes a significant contribution. So there can be a considerable difference between the UN and national statistical agency projections for more developed countries.

Recently the UN and World Bank have assembled and published time series from 1960 to 2010 of country of birth and country of residence statistics. This time series of migrant stock data has been used by Wittgenstein Centre researchers (Abel and Sander 2014, Sander et al 2014a) to estimate migration flows over decadal intervals, interpolated for five year intervals. Using model D3, Sander et al (2014b) undertake an analysis of the forces affecting country to country flows, drawing on the literature, a meta-expert workshop and a web survey of general expert views. Future gross migrations were generated from a combination of (1) a time series statistical model, (2) meta-expert views about the forces affecting international migration and (3) ordinary expert views with relative weights of four ninths, four ninths and one ninth respectively. The resulting trends were allowed to run until 2050, when a model of convergence to eventual zero net migration was adopted (switching to a D1 model).

Model D4 (ESPON 2010, Kupiszewski and Kupiszewska 2012) combines several migration models to generate sets of population projections for the countries of the European Union in 2010, EFTA and Switzerland and for their NUTS 2 regions. Within each country, a multi-regional model is employed, using out-migration transmission rates. Within Europe country to country flows are modelled using a similar model. The projections are connected by sub-models allocating country flows to regions using destination regional populations. Immigration flows from outside Europe and emigration flows to countries outside Europe are handled by developing trajectories of migration numbers. Seven scenarios were developed – three reference and four policy scenarios. The policy scenarios combined assumptions about the rate of change in migration rates or flows linked to economic/environmental performance and assumptions about convergence or divergence between countries and regions within countries linked to the strengthening or weakening of welfare/social solidarity policies (Rees et al 2012b).

5.5 Explanatory models of migration

We briefly referred earlier in the report to the large body of work on this type of migration model, remarking that little of this work had been used for forecasting. However, there is a tradition in the forecasting of local authority populations in the UK of linking the projections to housing plans (Rees 1994). A model used widely by Local Authorities (LAs) in the 1990s and 2000s was the CHELMER Model, developed by the Population and Housing Research Group under the leadership of Professor Dave King, latterly at Anglia Ruskin University. Cambridge Econometrics took responsibility for the model in 2008 (Cambridge Econometrics 2014). The methodology is probably as follows[5]. A housing plan for an LA provides numbers and sizes of planned housing. This is converted using survey or census information into the likely number and size of households and their composition in terms of age and sex. These constitute people who will be added to the local population. They will be found by applying migration admission rates to the planned populations, including from within the same LA. Two problems

[5] Unfortunately, no recent documentation seems to be available.

are faced by the user of such a model: (1) it needs a forecast of housing units to be built, usually based on a judgement about how many housing permissions will result in housing completions and (2) those permissions are useful only in the short term (5 years) beyond which local strategic plans may guide the analyst.

6. ILLUSTRATIONS OF THE IMPACT OF MODEL CHOICES

[This section is to be added. It will illustrate the impact of migration model choice using a cohort-component model for the four countries of the UK.]

7. DISCUSSION AND ADVICE

We now attempt an answer the question posed in the introduction: "which migration models should be used in population projections?"

It is clear that there is no such thing as a "best" model. Rather it is a matter of "horses for courses". That is, the model must be chosen in the light of driving forces of the migration system and of the data available. In particular, different models may be chosen to forecast different migration streams. Usually, a different model will be adopted for domestic migration from that for international migration. So for internal migration the most common choice is a model employing out-migration transmission rates (as in Rogers 1995). Applying such a model to the whole world assumes an absence of destination controls that is rather unlikely. A whole world model therefore indicates the potential for migration flows rather than future reality.

Raymer et al (2012) compare the forecasts of four simple regional population projection models: an overall growth rate model, a component model with net migration, a component model with in-migration and out-migration rates, and a multiregional model with destination-specific out-migration rates. The authors show how both the forecast subpopulation totals differ between the multiregional model and the simpler models, as well as for different assumptions about international migration. These experiments demonstrate that the ways in which migration is handled in projections really matter but offer little in the way of guidance.

We argue, following Rogers (1990), that net migration models hide too much. No net migrants exist. It is difficult to estimate age distributions for net migration. Net migration formulations may hide compositional effects of gross migration. The multi-regional model is therefore preferred.

Why should we therefore consider the bi-regional model? The bi-regional model requires far fewer variables which can be estimated more reliably. Rogers (1976) and Wilson and Bell (2004) both report that the results of a bi-regional model are close to those a full multiregional model when the model is used for all pairs of regions and in-migrations are adjusted so that the total over all regions matches the total for out-migrations.

The experiments of Van Imhoff et al (1997) investigated the impact of adding age disaggregation to a multi-regional projection model. The conclusions were that knowledge of the all age flow matrix plus matrices of origins by age and destinations by age yielded, using iterative proportional fitting, provides good fits to full arrays. This is the "three faces of the migration cube" solution. It is not always ideal because origin-destination structures differ according to life course stages (Rees et al 1996). Van Imhoff et al (1997) focussed on different representations (aggregations) of the full multiregional model. Though useful, the work did not get to the heart of the matter: namely, which migration model is most accurate in producing forecasts. Here we need experiments that use a first part of a time series to calibrate the model and a following part to test the model, assessing the degree and sources of error. To compare the effect of different migration models on future migration requires use of the same overall projection model and assumptions about components in a systematic experimental design (cf Bongaarts and Bullateo 1999, Rees et al 2013).

We conclude with a summary of advice, gleaned from the work reviewed, on which migration model to use in a population projection.

- It is better to use gross migration flows rather than net.
- It is better to use gross migration rates than flows, as long as there are no constraints on the number of migrants who can be received.
- If there are constraints, then a model of future in-migration or immigration totals should be developed. Migration admission rates can be used to source migrants by origin.
- A trade-off needs to be considered between more detail (when input variables are harder to estimate) and less detail (easier to estimate).
- The bi-regional model has less detail than the multi-regional model but gives results that are close.
- The three faces version of the multi-regional model is a reasonable approximation to the full model.
- Different models may be needed for the different sets of flows (internal, international, or intermediate e.g. within Europe).
- The best choice for internal migration (migration transmission rate in a bi-regional or multi-regional model) may not be the best choice for international migration. A world model of inter-country migration may exaggerate migration to richer destinations which may be subject to controls.

We hope that the ideas presented in this report will help in the design of future population projection models incorporating migration.

Acknowledgements

The conceptual analysis reported in this paper was commissioned by Edge Analytics Ltd, which has given permission for its inclusion in this paper. The work producing the illustrative ethnic projections was funded by the ESRC (Grant Ref ES/L013878/1, 2015-2016) as part of project *Evaluation, Revision and Extension of Ethnic Population Projections – NewETHPOP*. We also acknowledge the advice offered by Denise Williams of the Office for National Statistics, who provided access to ONS experimental work in preparation for the UK 2014-based National Population Projections.

Biographies

Philip Rees is Emeritus Professor of Population Geography at the University of Leeds, with interests in ethnic population projections, health outcomes and ageing of the population.

Nik Lomax is a lecturer in Population and Migration. His research focuses on the demographic composition of local areas, which influences policy and resource allocation decisions. This work incorporates measurement and estimation of migration, births and deaths as well as assessment of how these patterns change over time.

Peter Boden is Managing Director of Edge Analytics Ltd, a consultancy that provides demographic analytics to clients in business and government, specializing in demographic forecasting, scenario planning, optimization, planning policy and project management. They maintain and develop the POPGROUP projection software for the Local Government Association.

References

Abel G (2013) Estimating global migration flow tables using place of birth data, *Demographic Research*, 28(18): 505-546

Abel G and Sander N (2014) Quantifying global international migration flows, *Science*, 28 March, 2014, 343(6178): 1520-1522

Bijak J (2012) Migration Assumptions in the UK National Population Projections: Methodology Review. Report to the Office for National Statistics. Southampton: University of Southampton, http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/population-projections/npp-migration-assumptions-methodology-review/index.html

Bongaarts J and Bulatao R (1999) Completing the demographic transition. *Population and Development Review* 25, 515–29

Cambridge Econometrics (2014) The Chelmer Population and Housing Model, http://camecon.com/SubNational/SubNationalUK/ModellingCapability/Chelmer_Model.aspx

CEFMR (2014) *Central European Forum for Migration and Population Research*. Warsaw, Poland http://www.cefmr.pan.pl/

Champion A (1998) Population trends of small and medium-sized towns in nonmetro regions / Les tendances démographiques des villes petites et moyennes en régions non métropolitaines. *Revue de Géographie de Lyon*, 73(1), 5-16

Champion, T., Fotheringham, S., Rees, P., Bramley, G. and others (2002) *Development of a Migration Model*. The University of Newcastle upon Tyne, The University of Leeds and The Greater London Authority/London Research Centre. Office of the Deputy Prime Minister, London. http://www.ncl.ac.uk/curds/publications/pdf/DevelopmentofamigrationmodelPDF2404Kb_id1153369.pdf

CPC (2013) Integrated Modelling of European Migration (IMEM) Database. Centre for Population Change, University of Southampton, http://www.imem.cpc.ac.uk/

Disney NG (2014) *Model Based Estimation of UK Immigration*. PhD Thesis, Social Statistics, University of Southampton

ESPON (2010) *DEMIFER - Demographic and Migratory Flows Affecting European Regions and Cities*. http://www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/demifer.html

ETHPOP (2014) *ETHPOP database, publications and presentations*. www.ethpop.org

GLA (2014) *GLA 2013 Round Population and Household Projections*. Greater London Authority, http://data.london.gov.uk/dataset/gla-2013-round-population-and-household-projections

Kupiszewski M (ed.) (2013) *International Migration and the Future of Populations and Labour in Europe*. The Springer Series on Demographic Methods and Population Analysis Vol. 32, Springer

Kupiszewski M and Kupiszewska D (2011) MULTIPOLES: A Revised Multiregional Model for Improved Capture of International Migration. In J Stillwell and M Clarke (ed.) *Population Dynamics and Projection Methods*, Springer, pp.41-60

LGA (2014) *POPGROUP*. Local Government Association, http://www.local.gov.uk/popgroup

Lutz W, Butz W and KC S (ed.) (2014) *World Population and Human Capital in the Twenty-First Century*. Oxford, Oxford University Press

NIDI (2011) Project: MIMOSA: Modelling migration and migrant populations. Netherlands Interdisciplinary Demographic Institute, http://www.nidi.knaw.nl/en/research/mm/230211

ONS (2010) *Estimating International Long-term Emigration by Local Authority: Update*. Office for National Statistics, http://www.ons.gov.uk/ons/guide-method/methodquality/imps/msiprogramme/communication/improvements-mid-2008/methodology-papers/emigrationdetailedmethodology-update.pdf

ONS (2011). *Improved Methodology for Estimating Immigration to Local Authorities in England and Wales*. Web page describing a suite of methodological papers. Office for National Statistics, http://www.ons.gov.uk/ons/guide-method/method-quality/imps/improvements-to-local-authorityimmigration-estimates/index.html

ONS (2012) *A Conceptual Framework for UK Population and Migration Statistics*. Office for National Statistics, http://www.ons.gov.uk/ons/guide-method/method-quality/imps/latest-news/conceptual-framework/index.html

ONS (2013) *National Population Projections, 2012-based: Statistical Bulletin*. Office for National Statistics, http://www.ons.gov.uk/ons/dcp171778_334975.pdf

ONS (2014a) *2012-based Subnational Population Projections for England: Statistical Bulletin*. Office for National Statistics, http://www.ons.gov.uk/ons/dcp171778_363912.pdf

ONS (2014b) *Migration Statistics Quarterly Report, November 2014*. Office for National Statistics, http://www.ons.gov.uk/ons/dcp171778_386531.pdf

ONS (2014c) *An Examination of Options for Setting International Migration Assumptions for the National Population Projections*. NPP Migration assumptions rates-based modelling project 14(02), Office for National Statistics

Raymer J, Abel G, Rogers A (2012) Does specification matter? Experiments with simple multiregional probabilistic population projections. *Environment and Planning A* 44(11) 2664–2686

Rees P (1985) Does it really matter which migration data you use in a population model? In White P and Van der Knaap G (eds.) *Contemporary studies of migration*. Geobooks, Norwich, pp.55-77

Rees P (1986) A geographical forecast of the demand for student places. *Transactions, Institute of British Geographers*, NS11, 5-26

Rees P (1994) Estimating and projecting the populations of urban communities. *Environment and Planning A*, 26, 1671-1697

Rees P, Durham H, Kupiszewski M (1996) Internal migration and regional population dynamics in Europe: United Kingdom case study. *Working Paper 96/20*, School of Geography, University of Leeds, Leeds, UK, http://eprints.whiterose.ac.uk/3734/

Rees P, Wohland P, Norman P and Boden P (2011) A local analysis of ethnic group population trends and projections for the UK. *Journal of Population Research*, 28(2-3), 149-184

Rees P, van der Gaag N, de Beer J, Heins F (2012) European regional populations: current trends, future pathways and policy options. *European Journal of Population,* 28(4), 385-416

Rees P, Wohland P and Norman P (2013) The demographic drivers of future ethnic group populations for UK local areas 2001–2051. *The Geographical Journal,* 179(1), 44–60

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

Rogers A (1990) Requiem for the net migrant. *Geographical Analysis*, 22(4), 283-300

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

Sander N, Abel G and Bauer R (2014a) The global flow of people, the Wittgenstein Centre for Demography and Global Human Capital, Vienna, http://www.global-migration.info/

Sander N, Abel G and Riosmena F (2014b) The future of international migration. Chapter 7, pp.333-396 in Lutz W, Butz W and KC S (ed.) (2014) *World Population and Human Capital in the Twenty-First Century*. Oxford, Oxford University Press

Stillwell J and Congdon P (ed.) (1991) *Migration Models: Macro and Micro Approaches*. London, Belhaven Press

The Scottish Government (2008) *Fresh Talent: Working in Scotland Scheme: An Evidence Review*. Authored by Cavanagh L, Eirich F and McLaren J-G, Scottish Government Social Research, http://www.scotland.gov.uk/resource/doc/235857/0064664.pdf

UN (2014) *World Population Prospects – The 2012 Revision: Methodology of the United Nations Population Estimates and Projections*. Department of Economic and Social Affairs, Population Division, ESA/P/WP.235, New York, United Nations, http://esa.un.org/wpp/documentation/publications.htm

US Census Bureau (2014) *International Data Base – World Population: 1950-2050*. http://www.census.gov/population/international/data/idb/worldpopgraph.php

van Imhoff E, van der Gaag N, van Wissen L and Rees P (1997) The selection of internal migration models for European regions. *International Journal of Population Geography*, 3, 137-159

Wikipedia (2014) *e-Borders*. http://en.wikipedia.org/wiki/E-Borders

Wilson T and Bell M (2004) Comparative empirical evaluations of internal migration models in subnational population projections. *Journal of Population Research*, 21(2), 127-159

World Bank (2014) *Population Estimates and Projections*. http://data.worldbank.org/data-catalog/popuation-projection-tables

Zelinsky W (1971) The hypothesis of the mobility transition. *Geographical Review*, 61(2), 219-249