

# Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates

Nik Lomax · Paul Norman · Phil Rees · John Stillwell

Published online: 13 September 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** There is no single framework in place in the United Kingdom for collating and harmonizing the data on migration taking place between the subnational administrative units that constitute the home nations, together with the flows of international migration between these areas and the ‘rest of the world’. This paper proposes the construction of complete matrices of subnational migration statistics on an annual basis which can then be used to monitor migration trends throughout the UK in a comprehensive manner. The paper reviews how various ‘known’ flows are estimated by the national statistics agencies and develops estimates to fill in the gaps in the matrices. It then looks at changing patterns of migration in the UK between 2001/2002 and 2010/2011, showing that migration exhibits most activity in the middle of the decade. Flows within the home nations are the most consistent over time, while the patterns of international migration and migration across the borders of the home nations exhibit more substantial change. The latter flows, although relatively small in magnitude, have not previously been studied at the subnational level in any detail. Patterns of subnational internal and international net migration are found to be negatively correlated.

**Keywords** Internal migration · International migration · Local authority districts · United Kingdom · Subnational estimates

## Introduction

Migration statistics are an integral component of population change alongside the natural change components of births and deaths. The UK Statistics Authority (UKSA) has reported that whilst international migration has been the most

---

N. Lomax (✉) · P. Norman · P. Rees · J. Stillwell  
School of Geography, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK  
e-mail: eng4nml@leeds.ac.uk

significant driver of population change in the United Kingdom (UK) in the 2000s, internal migration has a ‘substantial influence on the changing level and composition of the population in local areas’ (UKSA 2009, p. 1). However, each of the UK’s official national statistics agencies (NSAs)—the Office for National Statistics (ONS) in England and Wales, the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA)—recognizes that migration, particularly the subnational dimension of international migration, is the most difficult demographic component to measure or estimate. While death is an event that occurs to a person only once and a birth is experienced by mothers only one to three times on average, a person can experience any number of migrations during a lifetime. Migration measurement is strongly influenced by the temporal and spatial frames used to capture data and there is additional uncertainty caused by the increasing numbers of people who live at more than one residential location.

The need to improve internal and international migration statistics has been widely acknowledged by all three NSAs and by the Interdepartmental Task Force on Migration (National Statistics 2006) that made a series of recommendations for ‘Improving Migration and Population Statistics’ (IMPS), led by the ONS. In 2008, a Parliamentary Committee reviewed the adequacy of official population statistics and its report (House of Commons 2008) resulted in the Migration Statistics Improvement Programme (MSIP), the vehicle through which the Government aimed to deliver the Task Force recommendations by 2012. UKSA (2009) reviews progress on MSIP and the adequacy of co-operation across government to deliver the planned improvements, whilst commissioned research by Rees et al. (2009) published within the UKSA report, provides a comprehensive summary of migration datasets, a critique of MSIP and a review of migration estimation methods. More recently, Raymer et al. (2012) published a conceptual framework for UK population and migration statistics which identifies the concepts and definitions that underpin data coming from a variety of sources, and outlines the methods used to derive estimates.

Statistics on annual subnational migration in the UK are compiled separately by the NSAs and fed through to ONS who assemble an aggregate mid-year estimate (MYE) of the population of each local authority district (LAD) in the UK together with estimates of the components of change using a common methodological approach (ONS 2011b). Each of the NSAs is responsible for producing more detailed MYEs for the subnational LADs within its borders. There are, however, a number of availability and consistency problems associated with the international and internal migration data used in the population estimation process.

Given the contemporary focus on migration statistics noted above, we begin this paper by proposing the construction of a time series of UK-wide annual migration estimates at LAD level that are based on data from administrative sources and the 2001 Census. The creation of a time series from the start of the twenty-first century, coinciding with the matrix of origin–destination migration flows during 2000/2001 that is available from the 2001 Census, would provide evidence of changing migration propensities and patterns during the 2001–2011 period and beyond. This will set the context for the projections of subnational migration and population into

the future. Comparison with data outputs from the 2011 Census will be carried out in due course.

The component parts of the subnational migration system are outlined in “[The subnational migration matrix and its component parts](#)” section and a distinction is drawn between those parts of the matrix that can be filled with ‘known’ estimates generated by the NSAs and those parts that currently remain ‘unknown’ and require further estimation. The various data sources and estimation methods used to create the known flows are reviewed in “[Data available from the NSAs](#)” section and methods for estimating the ‘unknown’ flows are introduced in “[Estimating the missing sections of the matrix](#)” section. Analyses reported in “[Changing patterns of migration in the UK](#)” section show how patterns in the distribution of national and subnational migration in the UK are changing and some conclusions are presented in “[Conclusions](#)” section together with ideas for further research.

### **The subnational migration matrix and its component parts**

The terminology of local government varies across the UK; following the 2001 Census there were various revisions to local government geography. In the work we report here we use the latest geographies (adopted in 2009) used by the NSAs for the publication of their mid-year estimates. In England this comprises 326 local government areas which include the City of London and 32 London Boroughs, 36 Metropolitan Districts, 56 Unitary Authorities (UAs) and 201 Non-Metropolitan Districts (which may variously be referred to as Shire Districts, Borough Councils or District Councils). Wales comprises 22 UAs, Scotland contains 32 Council Areas (CAs) and Northern Ireland is made up of 26 Local Government Districts (LGDs). For simplicity, we will refer to all these geographies as local authority districts (LADs). The schematic UK-wide subnational migration matrix illustrated in Fig. 1 incorporates three types of migration flows between these LADs: (1) inter-LAD flows within each constituent country which can be referred to as ‘internal intranational’ migration; these are flows in the cells labelled A for England, B for Wales, C for Scotland and D for Northern Ireland; (2) inter-LAD flows between each constituent country which we can refer to as ‘internal cross-border’ flows; these are flows in cells labelled E to P; and (3) flows into each LAD in the UK from the ‘rest of the world’ and out of each LAD to the ‘rest of the world’ which we can refer to as ‘international immigration’ and ‘international emigration’ flows; these are flows in cells labelled Q to T and U to X respectively.

Estimation of international and internal migration in the UK has received a good deal of attention in the past, but internal cross-border flows have received much less attention by researchers or planners. Rows of the matrix represent origins and columns are destinations, so the leading diagonal cells (represented as AW, BW, CW and DW in Fig. 1) contain migrations within each LAD that represent a large proportion of the reported migration taking place in the UK system. These within LAD migration flows are excluded from this paper as the focus is on the redistribution of migrants across the UK. This distinction between inter- and intra-LAD migration is important as it is the responsibility of the NSAs in each country to

Origin \ Destination	England				Wales				Scotland				Northern Ireland				Rest of the UK	Outside of the UK	
	LAD 1	LAD ...	LAD 326	Total	LAD 1	LAD ...	LAD 22	Total	LAD 1	LAD ...	LAD 32	Total	LAD 1	LAD ...	LAD 26	Total			
England	LAD 1	AW	A	A	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	LAD ...	A	AW	A	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	LAD 326	A	A	AW	AO	E	E	E	EO	G	G	G	GO	H	H	H	HO	U	EE
	Total	AD	AD	AD	AD	ED	ED	ED	ET	G	G	G	GT	H	H	H	HT	UT	EET
Wales	LAD 1	F	F	F	FO	BW	B	B	BO	I	I	I	IO	J	J	J	JO	V	FF
	LAD ...	F	F	F	FO	B	BW	B	BO	I	I	I	IO	J	J	J	JO	V	FF
	LAD 22	F	F	F	FO	B	B	AW	BO	I	I	I	IO	J	J	J	JO	V	FF
	Total	FD	FD	FD	FT	BD	BD	BD	BT	ID	ID	ID	IT	JD	JD	JD	JT	VT	FFT
Scotland	LAD 1	K	K	K	KO	L	L	L	LO	CW	C	C	CO	M	M	M	MO	W	GG
	LAD ...	K	K	K	KO	L	L	L	LO	C	CW	C	CO	M	M	M	MO	W	GG
	LAD 32	K	K	K	KO	L	L	L	LO	C	C	CW	CO	M	M	M	MO	W	GG
	Total	KD	KD	KD	KT	LD	LD	LD	LT	CD	CD	CD	CT	MD	MD	MD	MT	WT	GGT
Northern Ireland	LAD 1	N	N	N	NO	O	O	O	OO	P	P	P	PO	D	D	D	DO	X	HH
	LAD ...	N	N	N	NO	O	O	O	OO	P	P	P	PO	D	D	D	DO	X	HH
	LAD 26	N	N	N	NO	O	O	O	OO	P	P	P	PO	D	D	D	DO	X	HH
	Total	ND	ND	ND	NT	OD	OD	OD	OT	PD	PD	PD	PT	DD	DD	DD	DT	XT	HHT
Rest of the UK	Q	Q	Q	QT	R	R	R	RT	S	S	S	ST	T	T	T	TT	YT/ZT	JJT	
Outside of the UK	AA	AA	AA	AAT	BB	BB	BB	BBT	CC	CC	CC	CCT	DD	DD	DD	DDT	ITT	KKT	

Data available in migration matrix (since 2006/07)  
 T, O, D Margin total/sub-total available    E, F, O, D Margin total/sub-total not available    A, ..., P LAD to LAD flow available    A, ..., P LAD to LAD flow not available    Notation (A-KK) explained in text

**Fig. 1** Interaction matrix of migration data availability in the UK (since 2006/2007)

provide mid-year population estimates (MYEs) at the LAD scale and therefore it is the inter-LAD flows that are particularly relevant, rather than the intra-LAD flows. The MYEs are essential because they inform resource allocation and policy decisions at national, regional and local levels and considerable importance is attached to the natural change and migration components that are fed into the cohort component model used to produce the MYEs. The data layout presented in Fig. 1 is known as an interaction matrix because it represents the relationship between origins and destinations (Stillwell and Harland 2010).

Data availability across the decade from 2000/2001 varies from year to year and the matrix in Fig. 1 shows the availability of migration flow estimates since 2006/2007. We will briefly discuss each part of the matrix whose components have labels ranging from A to KK. Part A of the matrix shows migration flows between LADs in England, with the total outflow from each LAD as an origin (the margin labelled AO) and inflow to each LAD as a destination (the margin labelled AD). Flows within Wales are represented in the portion of the matrix labelled B. Data availability is good in England and Wales, with both LAD-to-LAD flows and marginal totals present. Part C of the matrix represents migration flows within Scotland, where data availability is also good, with all flows being estimated by NRS. Northern Ireland is represented by the portion of the matrix labelled D. This is the first data gap: the margins representing total inflows (DD) and outflows (DO) for each LAD are available, but no estimates for LAD-to-LAD migration (labelled D) are readily available.

The second data gap can be identified in the parts of the matrix representing within UK cross-border flows, labelled E to P. The only available cross-border origin–destination data are between England and Wales (sections E and F). For the rest of the cross-border sections (labelled G to P) not only is there missing information for LAD-to-LAD flows, but the majority of marginal, country-to-

country totals (except those associated with flows in and out of England and Wales, GO to JO and KD to OD) are also missing. To find a set of consistent marginal totals for the cross-border part of the matrix, we need to look at total flows to each LAD from the rest of the UK (labelled Q to T) and from each LAD to the rest of the UK (labelled U to X). These marginal 'rest of the UK' totals are available for all LADs, and form the basis for estimation of the within UK cross-border flows.

The final parts of the matrix are the total flows from the 'rest of the world' to each LAD (labelled AA to DD) and from each LAD to the 'rest of the world' (labelled EE to HH). Data on immigrants and emigrants are supplied by the NSAs which are used to fill the overseas rows and columns of the matrix. The corners of each subsection of the matrix represent the sum of the column and row, and are labelled with the notation for the subflow followed by a T; so for example, all flows to and from LADs in England to/from the rest of England have the label AT.

### **Data available from the NSAs**

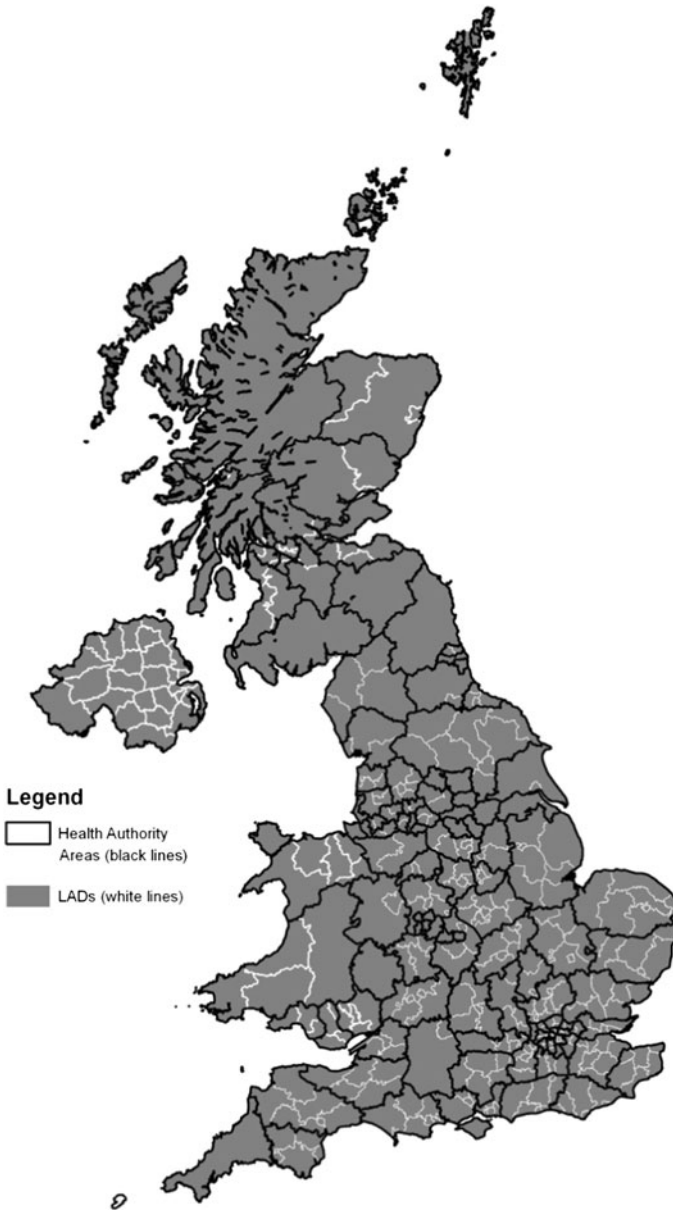
Much of the data used in this paper is taken from estimates supplied by the NSAs. In this section we look at the data and methods used to derive these estimates.

#### **Internal intranational and cross-border migration**

Across the UK, annual internal intranational and internal cross-border migration estimates are derived primarily from National Health Service (NHS) sources by ONS, NRS and NISRA. The National Health Service Central Register (NHSCR) records re-registrations (when people register with a different General Practitioner (GP) doctor) between the 124 former Health Authority (HA) areas in England and Wales and 15 Health Board (HB) areas in Scotland. A database of flows between these health areas, plus Northern Ireland, is collated by ONS. Northern Ireland is split into five Health and Social Care Trust areas, but data from these are not used for the estimation of migration in Northern Ireland. In 2006, England and Wales HAs became redundant health administrative zones but NHSCR-based estimates continue to be published based on their boundaries (ONS 2010c). Figure 2 shows the health geography used for NHSCR reporting (black lines) in relation to LAD boundaries (white lines). The NHSCR provides the framework for within-UK migration estimates, both intranationally in England, Wales and Scotland and for cross-border migration for all four home nations.

#### *Internal intranational migration*

For internal (intranational) migration in England and Wales, ONS uses an additional dataset called the Patient Register Data System (PRDS). NRS uses a dataset called the Community Health Index (CHI) and NISRA uses a dataset which is also known as the CHI (or health card register). These three datasets are more detailed than the NHSCR as they contain the start (origin) and end (destination) postcodes of a migrant. This information allows for reporting at the LAD level, but is considered to



**Fig. 2** Health Authority areas reported in the NHSCR, overlaid on LAD boundaries for the UK

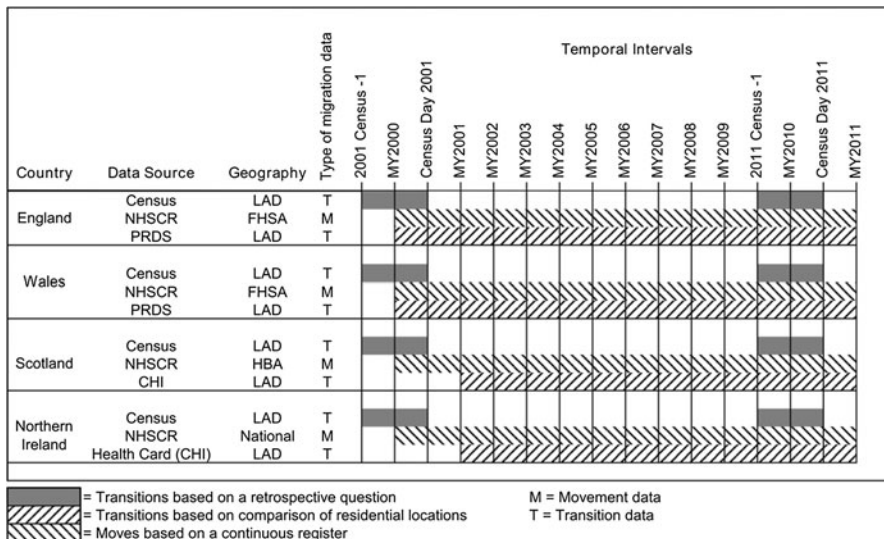
be less complete than the NHSCR data on migration, as they are downloaded only once per year and as such report transitions. In contrast, the NHSCR information is available to the NSAs as a weekly download, and as such is capable of producing movement data.

The distinction between movement and transition data is an important one; as highlighted by Rees and Willekens (1986), movements (which are demographic events equivalent to births and deaths) can occur multiple times within a given time period. A transition compares a person's location at the beginning and end of a given time period, so that only one person transition is measured. Transition data often miss significant migrations; for example, a person who moves from one LAD to another just after the start of a time period, and subsequently moves back just before the end of the time period. To account for this, both ONS and NRS scale their PRDS and CHI estimates to agree with totals at health area available in the NHSCR. NISRA uses the NHSCR to quality-assure the migration estimates derived from the CHI/health card registration data, but do not use the same scaling up procedure. As discussed the "Introduction", the statistical output from NISRA is not consistent with the other NSAs as only the marginal totals for each LAD are published.

Coverage in terms of temporal intervals and subpopulations varies between the datasets, and these differences are shown in Tables 1 and 2.

Table 2 shows that the UK census of population provides migration information for all subpopulations in the 1 year before the census enumeration date (shown in Table 1). These populations are identifiable and subsettable within the data. The data provided in the census are transition data comparable to PRDS, CHI and health card data. However, the temporal time frame differs by 3 months, as the census enumeration year refers to the 12-month period before the census date in April or March, whereas the mid-year NHS data are reported at the end of June. PRDS, CHI and health card data are produced as yearly outputs so changes between 1 year and the next are counted as migrant transitions. The NHSCR is available weekly, but a

Table 1 Temporal intervals reported in the available data



**Table 2** The subpopulations counted in each data source

Migration data source	Subpopulations						
	NHS patients			Armed forces	Prison population	Students	Others
	Infant migrants	Young adult	All others				
Census (transitions)	E	U	Yes	Yes	Yes	Yes	Yes
NHSCR (events)	Yes	U	Yes			U	
PRDS (transitions)	Yes	U	Yes	Yes <sup>a</sup>		U <sup>b</sup>	
CHI (transitions)	Yes	U	Yes			U	
Health card (transitions)	Yes	U	Yes			U <sup>b</sup>	

*E* estimate, *U* undercount

<sup>a</sup> Included with all moves to/from the rest of the UK

<sup>b</sup> Adjustment made for subpopulation

rolling mid-year dataset (consistent with the mid-year download of the other NHS data) is used to provide totals with which the PRDS and CHI are adjusted to agree.

All NHS sources undercount young adults, particularly young men, who are often slow to re-register with a GP when they move (ONS 2010b). For similar reasons, students are undercounted, or counted at their parents' address during term-time. An estimated student adjustment is made by ONS in England and Wales using statistics from the Higher Education Statistics Authority, which gives a term time and parental address for all students in higher education. An adjustment is made in Northern Ireland where, informed by administrative data sources, students are reallocated from most LADs 'to a small number of LGDs with centres of third level education' (NISRA 2007, pp. 3): Belfast, Newtownabbey and Coleraine (NISRA 2006). No adjustment is made to Scottish CHI data for students.

Unlike the census, which aims to enumerate all population subgroups, other migrant populations such as people in prisons and in the armed forces are not, as a whole, dealt with in the NHS datasets. These populations are treated separately in the subnational mid-year estimates produced by the NSAs and, while they can contribute substantially to the resident population of specific areas, are outside the scope of this study. The exception is that armed forces migrants are included in the to/from the 'rest of the UK' figure reported by NRS for Scotland. This has an important impact on consistency as the armed forces population is not reported in the other NHS datasets, meaning that UK-wide, flows to and from the rest of the UK (YT/ZT in Fig. 1) do not sum to the same value. The implication for this in the estimation of missing values is covered in "Calculating internal intranational flows between LADs in Northern Ireland and for UK-wide internal cross-border flows" section.

### *Internal cross-border migration*

The NHSCR gives detail on migration across the borders of the UK at HA level (national for Northern Ireland), but data on cross-border flows between LADs in



each of the constituent countries (labelled D to I in Fig. 1) do not currently exist and have not been estimated by any of the NSAs. This is a major gap in the subnational estimation process and is tackled in our estimations presented in “[Estimating the missing sections of the matrix](#)” section. Flows to and from the ‘rest of the UK’ are reported by the NSAs at LAD level (J to O in Fig. 1), but with no specific origin/destination detail. All of these totals are transition data, and are PRDS, CHI and health-card derived data. The data are slightly more detailed in England and Wales, where the migration of patients into and out of LADs can be broken down by Scotland or Northern Ireland, but there is still no subnational detail for the origin/destination.

### Subnational immigration

Each of the NSAs in the UK has its own method for estimating immigration from the ‘rest of the world’ at the subnational level and these methods have been the subject of substantial revision during the 2000s. This is particularly true for ONS, where the current methodology only applies to statistics for mid-year 2006 onwards. Given the numerous revisions to international migration methodologies, a brief overview of the current situation is presented here, but a detailed assessment of the changing methodologies through time can be found in Lomax et al. (2011, pp. 13–34), in Raymer et al. (2012, pp. 38–53) and in ONS (2011a).

ONS and NRS use a survey source called the International Passenger Survey (IPS) as the basis for both immigration and emigration statistics. From the IPS, a Long Term International Migration estimate (LTIM) is derived for the English regions, Wales and Scotland using the Labour Force Survey (LFS). The LFS is a household survey that covers 60,000 households per quarter and is used to allocate the estimate of immigrants identified in the IPS around the UK (ONS 2007). For data reported at mid-year 2006 onwards, the IPS estimate of immigration is distributed directly to the LAD level in England and Wales by using administrative sources which correspond to the type of migration reported by migrants in the IPS questionnaire (ONS 2011a, based on work by Boden and Rees 2010). The main streams identified are those entering the UK for work, for study, returning migrants and an ‘other’ group who do not state one of the specific reasons for immigration. When migrants state their reason as being for work purposes, the Migrant Worker Scan and the Lifetime Labour Market Database (known as L2) are used to distribute the migrants based on national insurance number (NINo) registrations. For immigrants who state their reason as study, data from HESA and the Department of Business, Innovation and Skills (which records Further Education students) are used. Finally, registrations with a GP (Flag-4 registrations) are used to allocate the ‘other’ migrants. Asylum seeker data taken direct from the Home Office are added to the subnational immigration estimate.

In Scotland, the Scottish share of UK LTIM is distributed to Scottish HBs using overseas inflows recorded on the NHSCR. The distribution of immigrants to LADs uses postcodes reported in the CHI. The majority of asylum seekers are assumed to be supported by the National Asylum Support Service (NASS) and as such are

removed from the LTIM control totals and distributed to Glasgow, which is the only Scottish LAD in contact with the UK Border Agency (UKBA) (GROS 2010b).

The methodology in Northern Ireland differs from the rest of the UK as NISRA does not make use of data from the IPS, instead using health card registration data. Registration with a family doctor requires an immigrant to apply for a health card, at which point he or she must provide information about place of residence and time of stay to the Health and Social Care Business Services Organisation (HSC-BSO) in Northern Ireland (NISRA 2010). Immigration of asylum seekers into Northern Ireland is distributed subnationally using the same Home Office data used by ONS for England and Wales (NISRA 2010).

### Subnational emigration

For emigration estimates, ONS and NRS use the IPS which includes a sample of emigrants interviewed at UK air, sea and Channel Tunnel embarkation points. For England and Wales, a Poisson regression model is used at the LAD level, with the IPS direct estimate as the response variable. The model includes the immigration estimate from the previous year, and uses a number of other variables such as housing type and housing tenure (ONS 2010a). For Scotland, estimation for LADs is based directly on the IPS 'using averaged proportions based on international inflows, outflows to the rest of the UK and the population size of each Health Board' (GROS 2010a, p. 1).

In Northern Ireland, estimates are derived from the health card system which records deregistrations with a family doctor. The reported total is scaled up by 50 % to take into account the low deregistration rate (NISRA 2010) as deregistration is not mandatory and there is little incentive to do it. The deregistration data are combined with the data from the Central Statistics Office (CSO) Irish Quarterly National Household Survey which provides an estimate of numbers moving from Northern Ireland to the Republic of Ireland.

### Estimating the missing sections of the matrix

Two main parts of the matrix need to be estimated: first, the internal intranational flows within Northern Ireland; and second, all internal cross-border flows. In both cases, an Iterative Proportional Fitting (IPF) routine can be implemented. IPF is a procedure used to adjust flows in contingency tables so that they are consistent with a set of known marginal constraints. A comprehensive study of the history and application of IPF is provided by Založnik (2011), who emphasizes that IPF is a procedure employed across a wide range of disciplines from engineering and transport studies to economics and demography. It is known by different names across the fields, e.g. 'Cross-Fratar' and 'Furness' methods in transport engineering and 'RAS' in economics (Norman 1999, p. 7; Wong 1992, p. 340). Johnston and Pattie (1993, pp. 321) conclude that 'other applications have employed different terminology using the IPF procedure as a means to a well known mathematical goal, the maximisation of entropy'. Entropy maximization retains the structure of the

original contingency table, so the estimated values are the ‘maximum likelihood estimates of the unknown values’ (Johnston and Pattie 1993, p. 317).

In its classical application (as identified by Bishop et al. 1974; Denteneer and Verbeek 1985; Založnik 2011), IPF is used to combine data from two or more sources. The first use of IPF in its classical sense, to fit a contingency table using marginal constraints, is widely credited to Deming and Stephan (1940) who used the procedure on US census data to extrapolate a 5 % sample to the entire population. The idea of using information from ‘different geographical areas, time periods and data sources’ to improve partial or inadequate data is presented by Rogers et al. (2003, p. 68), while Raymer and Rogers (2007, p. 199) update ‘the migration data of a census in order to satisfy the marginal totals obtained or estimated for a later period of interest’ in the United States using a log-linear model.

The initial contingency table is often called the ‘seed’ as it provides a starting value from which to adjust estimates in subsequent iterations. The IPF procedure (after Wong 1992, pp. 340–341; Norman 1999, pp. 4) can be expressed as

$$P_{ij(k+1)} = \left( \frac{P_{ij(k)}}{\sum_j P_{ij(k)}} \right) \times Q_i \quad (1)$$

$$P_{ij(k+2)} = \left( \frac{P_{ij(k+1)}}{\sum_i P_{ij(k+1)}} \right) \times Q_j \quad (2)$$

where  $P_{ij(k)}$  is the contingency table component in row  $i$  and column  $j$  at iteration  $k$ .  $Q_i$  is the row total while  $Q_j$  is the column total. Equations (1) and (2) are employed iteratively and will theoretically stop (‘converge’) at iteration  $m$  where

$$\sum_j P_{ijm} = Q_i \quad \text{and} \quad \sum_i P_{ijm} = Q_j \quad (3)$$

In practice, the process stops at a predefined threshold error (in our model 0.001) or maximum number of iterations (here set at 50), whichever comes first. The 2001 Census provides the initial seed values for  $P_{ij(k)}$  which are then updated using the marginal in/out totals (informed by changes in the larger-area health geography migration data) for the year being estimated.

Using IPF to estimate missing migration data

IPF is a technique that has been widely used in the estimation of missing or incomplete migration data. Previous studies have used the technique to improve existing origin–destination migration flows, to produce estimates for a particular time period where only marginal totals are known and to derive migration estimates for subsections of the population. To improve existing distribution of origin–destination flows, Chilton and Poet (1973) use in and out marginal totals to estimate the small flows masked by disclosure control for the 33 LADs of London in the 1966 Census. Similarly, Rees and Duke-Williams (1997) address suppression of origin–destination flows in the 1991 Census Special Migration Statistics, estimating the missing migration flows using marginal totals and producing a set of revised tables where all subtotals were consistent.

A starting distribution of origin–destination flows can be updated and constrained to marginal totals for a given time period to produce time-series estimates. Nair (1985), in response to the limitation of many Third World countries only reporting lifetime origin–destination migration, uses this distribution in India and Korea to produce 1, 5 and 10-year migration matrices based on the marginal totals available. Nair (1985, p. 140) concludes that IPF is an approach suited to ‘estimating intercensal (usually 10 years) migratory flows.’ Schoen and Jonsson (2003) use IPF to produce new estimates of interregional migration in the US between 1980 and 1990 as a benchmark against which to test their own estimation methodology.

To create origin–destination estimates for subsections of the population, Willekens et al. (1981) use IPF to derive age-specific flows from an aggregate matrix, as does Willekens (1982). Van Imhoff et al. (1997) use IPF to produce a simplified multidimensional migration dataset by age and sex.

So why have we chosen IPF to estimate the missing flows in our dataset? The selection of an appropriate technique for estimating missing data in origin–destination migration tables is largely down to the researcher’s preference: Raymer (2007) highlights that log-linear models, gravity models, spatial interaction models, entropy and information maximization models and IPF are all approaches that have been successfully applied to the estimation of place-to-place migration flows. He cites Willekens (1980, 1983) as two papers that demonstrate the ‘equivalences’ between all of these techniques. A useful case study in the selection of an appropriate method for estimating migration tables is provided by van Imhoff et al. (1997), who favoured IPF for modelling a multidimensional age/sex/origin and age/sex/destination dataset for Europe owing to the efficiency of the technique when producing a range of model results. They first attempted to use a log-linear approach in the software package GLIM, but found that to run a model ‘takes several hours, which is prohibitive for an exploratory analysis’ (p. 139). When comparing methods, they concluded that ‘the fitted rates of IPF and GLIM are the same. Also, IPF is many times faster’ (p. 139). In the estimation presented in this paper, IPF is a suitable approach as consistent marginal totals are available for cross-border and within-Northern Ireland migration, and the speed in which the routine can be implemented in the software package R allows for efficient estimation across the decade. This speed and ease of implementation also provides the potential to model origin/age/sex and destination/age/sex flows in the future.

#### Calculating internal intranational flows between LADs in Northern Ireland and for UK-wide internal cross-border flows

As the marginal inflow and outflow totals are available for each LAD in Northern Ireland for each year, the internal intranational flows in Northern Ireland can readily be estimated using the IPF routine. The process is not so straightforward for UK internal cross-border flows, however.

To use the routine on internal cross-border flows (labelled D to I in Fig. 1), the marginal flow totals to and from the rest of the UK are used (based on a recommendation made by Raymer 2012, personal communication). As we are looking at a closed system where the sum of all moves from one part of the UK to

another part should have an overall net effect of zero, the count in the corner cell of the cross border margin in Fig. 1, labelled YT/ZT, should equal both total inflows (Q to T in Fig. 1) and total outflows (U to X in Fig. 1). This is not the case for two reasons: first, the effect of rounding individual cells to 10 in the ONS data, and second, the inclusion of armed forces moves in the NRS data for Scotland. Moves to and from the armed forces are included in the ‘rest of UK’ figure for Scottish LADs, but it is not possible to distinguish between an armed forces move within Scotland or armed forces moves to/from another part of the UK. It is the inclusion of armed forces which appears to cause a large proportion of the inconsistency between total inflows and total outflows (YT/ZT), as can be seen in Fig. 3. The comparison for Scotland (light grey bars in Fig. 3) has been drawn from national-level NHSCR data (which do not include armed forces moves) and summing the CHI data (which do include armed forces moves). By taking the difference between NHSCR and CHI, we are left with moves to/from the armed forces for Scotland. These armed forces moves account for the majority of the total difference seen for the UK (dark grey bars in Fig. 3).

For the IPF routine to converge, the marginal totals must sum to the same value, so the totals have to be adjusted to ensure consistency. The Scottish data are adjusted to remove the armed forces moves, while the small remaining difference is attributed to the rounding issue in England and Wales. Thus, where  $\sum_j D_j$  is total immigration and  $\sum_i O_i$  is total outmigration, if

$$\sum_j D_j - \sum_i O_i = E \neq 0 \tag{4}$$

where  $E$  is the difference between total inflow and outflow, then an adjustment needs to be made to ensure that the total of all origins and destinations are equal. For all years, total inflow is higher than outflow, so the outflow totals for each LAD in Scotland were adjusted upwards (as were the LADs in England and Wales to account for the small difference in rounding) as follows:

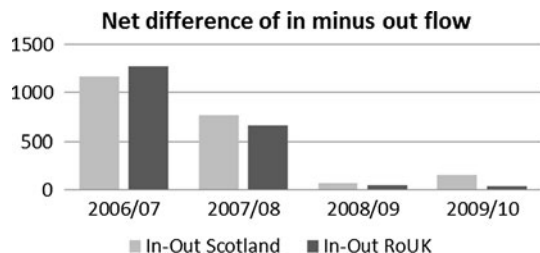
$$\widehat{O}_i = O_i + E \times (O_i/O_+) \tag{5}$$

where:

$$O_+ = \sum_i O_i \tag{6}$$

Any error is distributed across origins in proportion to the estimated outmigration total. The error is distributed across origins rather than destinations as the

**Fig. 3** A comparison of the difference between origin and destination migration totals for the UK and for Scotland



destination totals are more certain in census and survey migration tables because recall bias is avoided. For register-based datasets, although this argument does not apply, only the census gives comprehensive coverage of the population groups, so we go with the census logic.

The IPF procedure requires an entire origin–destination matrix, so while we are not interested in estimating intracountry flows, all cells (A–P in Fig. 1) need to be included in the table. These internal migration cell values (A–D in Fig. 1) are set to 0.001 (the lowest value possible for the IPF routine to work) so that no value is assigned to them in the rest of the UK estimation model.

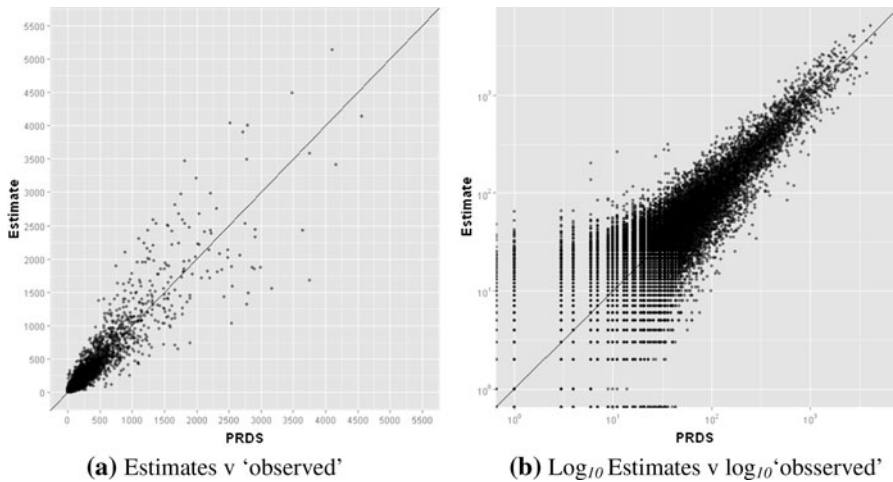
### Testing the IPF routine on observed data

IPF can be applied to generate flows for England and Wales where an official estimate already exists; this estimate is derived from PRDS data. For the purposes of the following comparison this will be referred to as ‘observed’ data. The observed data are the PRDS-derived estimates used in sections A–F of the matrix in Fig. 1. The IPF-derived estimates can be compared with these observed data in order to ascertain the robustness of the method. The IPF estimate is derived by combining marginal totals for each LAD in England and Wales (from the PRDS) with the internal cell structure found in the 2001 Census as the seed value. The procedure is repeated for annual data between 2000/2001 and 2006/2007 to provide a number of years from which comparisons can be drawn.

The coefficient of correlation between observed and estimated flows of 0.94 ( $p < 0.01$ ) shown in Fig. 4a is for all pairs of LAD-to-LAD flows, averaged across 2001–2007. The correlation for each individual year does not drop below 0.910 ( $p < 0.01$ ) although there are some outliers, both where the estimate exceeds the PRDS observed flow and vice versa. Although the correlations between the observed and estimated flows appear to be strong throughout the time period, the distribution exhibits heteroscedasticity when the larger values are considered, so it is necessary to be cautious when interpreting the results as the smaller variance for low and mid-range values may bias the correlation. Figure 4b shows the comparison between estimated and observed data where both scales have been logged, and shows that variations do exist in the lower values that are not evident from the pattern seen in Fig. 4a. Our experiments showed that the estimate based on a prior census, PRDS marginals and IPF is very close to the observed data, while not a perfect match.

### Changing patterns of migration in the UK

By constructing a UK-wide matrix of migration at the LAD level, we are able to interrogate the three different migration flows (internal intranational, internal cross-border and international) in more detail. It is our intention in this section to provide some indications of changes in the pattern of migration flows over the decade by comparing our estimates of the full matrix of flows for the first and last years of the time-series (2001/2002 and 2010/2011) and the mid-decade results from 2006/2007.



**Fig. 4** A comparison of **a** values and **b**  $\log_{10}$  values of ‘observed’ (PRDS) and estimated flows, averaged for 2001–2007

This mid-decade time interval marks the end of the long boom (1992–2006) and exhibits migration activity higher than any other in the decade. The next year 2007/2008 marks the start of the financial and economic crisis for Western countries, which, at the time of writing, has lasted 6 years. We present the national-level picture for each of the countries of the UK, followed by some subnational results at LAD level.

To aid our understanding of migration activity at the subnational level, Table 3 gives an overview of total flows by country, comparing total inflow, total outflow and the net result of each type of migration in 2001/2002, 2006/2007 and 2010/2011. The majority of the total migration is clearly composed of internal (intranational) moves for which the net effect is zero and England accounts for a large proportion of the migration in each midyear to midyear period.

When all migrants are considered, the magnitude of intranational migration is over 150,000 higher in 2006/2007 than in 2001/2002 before it falls back by roughly the same amount between 2006/2007 and 2010/2011. This pattern is true for England, Wales and Northern Ireland. In Scotland, the first two time periods are relatively consistent but the pattern of decline between 2006/2007 and 2010/2011 is evident.

Total UK international immigration and emigration follows the same pattern, with substantial increase in the number of both immigrants and emigrants between 2001/2002 and 2006/2007 (inflow is 117,172 higher in 2006/2007 than in 2001/2002 while outflow is 49,030 higher). The number of immigrants is 32,290 lower in 2010/2011 than in 2006/2007 while the number of emigrants falls by 117,172. This pattern of a mid-time period spike is evident for migrant numbers in England, Scotland and Northern Ireland. Overall, cross-border migration falls throughout the decade.

**Table 3** Total in, out and net flows for each type of migration by country, 2001/2002, 2006/2007 and 2010/2011

Country	Years	Internal Total	Cross-border			International		
			In	Out	Net	In	Out	Net
England	01/02	2,422,040	107,062	-122,423	-15,360	450,747	-302,409	148,338
	06/07	2,566,904	100,540	-118,170	-17,630	530,085	-351,786	178,299
	10/11	2,432,865	98,088	-102,727	-4,638	506,261	-279,049	227,212
Wales	01/02	49,708	64,567	-54,848	9,719	10,533	-8,520	2,013
	06/07	54,010	62,784	-55,756	7,028	18,346	-9,854	8,492
	10/11	53,261	57,034	-54,500	2,534	14,635	-10,278	4,357
Scotland	01/02	118,818	54,408	-49,690	4,717	18,357	-24,400	-6,043
	06/07	117,747	51,542	-42,701	8,840	37,800	-21,000	16,800
	10/11	108,059	43,684	-40,779	2,905	41,000	-16,400	24,600
Northern Ireland	01/02	38,344	12,514	-11,589	924	8,791	-9,613	-822
	06/07	43,251	12,894	-11,131	1,762	19,369	-11,332	8,037
	10/11	36,292	10,322	-11,122	-801	11,414	-13,824	-2,410

A link between economic conditions and migration propensities is well established in the literature, at least for internal migration, with periods of economic growth coinciding with relatively high migration intensities. Stillwell et al. (1992, p. 31) highlight the fluctuation in migration propensity between 1971 and 1991, attributing the reduced rate of migration activity in the 1970s to the decline in economic activity in terms of 'changes in the economy on employment, incomes and housing' where, during the 1979–1983 recession, 'migration activity was at its lowest ebb'. The subsequent increase in migration rate from 1981/1982 onwards correlated closely with a decreasing unemployment rate and improving economic conditions. These findings are echoed by Owen and Green (1992), Ogilvy (1982) and by Champion (1987).

UK per-capita gross domestic product (GDP) is higher in 2006/2007 than any other midyear to midyear period in the decade, having risen steadily since 2001/2002. It then falls back dramatically in 2008/2009 and stagnates to the end of the time series, 2010/2011 (ONS 2013b). The unemployment rate is also higher in 2010/2011 than in 2006/2007 (ONS 2013a). These economic trends appear to be closely related to the pattern of internal and international migration seen in each of our three midyear to midyear time periods.

#### Flows at LAD scale

Looking at the net migration balances at LAD scale for each of the three types of migration allows us to decompose the national trends identified in Table 1. Although the use of net migrant balances means that the changes between component inflows and outflows across the time series are not identified, they do provide a good summary measure of the changing pattern of migration across the



decade. To aid our understanding of these changing patterns, the correlations between net migration balances for LADs in each of the years are reported.

Figure 5 shows the pattern of net internal (within each country) migration during each of the three annual periods. The general trend is one of decline in the volume of migrants from the beginning to the end of the decade. Patterns in 2001/2002 and 2006/2007 are similar, with the same areas losing migrants: most London Boroughs, the urban conurbation of the West Midlands, metropolitan LADs in the North West, plus Glasgow, Edinburgh and Belfast. The primary areas of net gain are the LADs in the South West (especially Cornwall), along the south coast and the East of England. Generally the distinction between metropolitan net losses and rural net gains is evident across all three 12-month periods, but is more pronounced in the two earlier midyear-to-midyear periods. The trend of counterurbanization is well researched in the migration literature and has been a longstanding pattern in the UK: in the 1970s and 1980s it is given detailed attention by Cross (1990), Kennett (1980) and Champion (1989), whilst the phenomenon in the 1990s is explored by Kalogirou (2005) and in the 1991 Census by Rees et al. (1996, p. 78). Similar counterurbanization trends are detected from the results of the 2001 Census by Champion (2005), Stillwell and Duke-Williams (2007) and Stillwell (2013) and the pattern seen in Fig. 5 demonstrates a continuation of the trend in the 2000s.

The similarity of the internal migration patterns seen between 2001/2002 and 2006/2007 is confirmed by a strong positive correlation between the net flow for all LADs in the two time periods ( $r = 0.89$ ,  $p < 0.01$ ), suggesting that the same LADs are losing or gaining a similar number of net migrants. A shift in the pattern can be seen to have taken place by 2010/2011, however, which is indicated by a weaker correlation between net flows at the beginning and end of the decade ( $r = 0.79$ ,  $p < 0.01$ ). The pattern of urban loss and rural gain continues, but with a much smaller net balance for most LADs. This shift is particularly apparent in London (where boroughs in the east are now gaining migrants) and Glasgow, Edinburgh and Belfast which now are losing far fewer migrants to the rest of Scotland and Northern Ireland respectively. In Wales, the two predominant LADs for redistribution of migrants in 2001/2002 and 2006/2007, Cardiff (a net gainer) and Swansea (a net loser), show very little net migration activity in 2010/2011. The pattern of net gain in Wales is similar in 2010/2011 to previous years but the number of migrants has reduced dramatically.

Cross-border migration patterns appear to change substantially between the start and end of the time series (Fig. 6). The correlation between net flow for all LADs between 2001/2002 and 2006/2007 is 0.77 ( $p < 0.01$ ) and is lower between 2006/2007 and 2010/2011 ( $r = 0.65$ ,  $p < 0.01$ ). The pattern seen at the beginning and end of the decade shows a positive correlation which is significant but weaker still ( $r = 0.64$ ,  $p < 0.01$ ). The pattern evident in Fig. 6 is one of net gain in rural Wales and Scotland, a phenomenon explored by Jones (1992) who argues that immigration from the rest of the UK to rural Scottish regions is driven by oil-related employment in Highland (especially Aberdeen/Grampian) regions and residential preference for rural areas. Rees et al. (1996) refer to the peripheral gains in northeast Scotland's 'new resource frontiers' resulting from the development of onshore facilities for offshore gas and oil fields. Figure 6 shows that these gains appear to

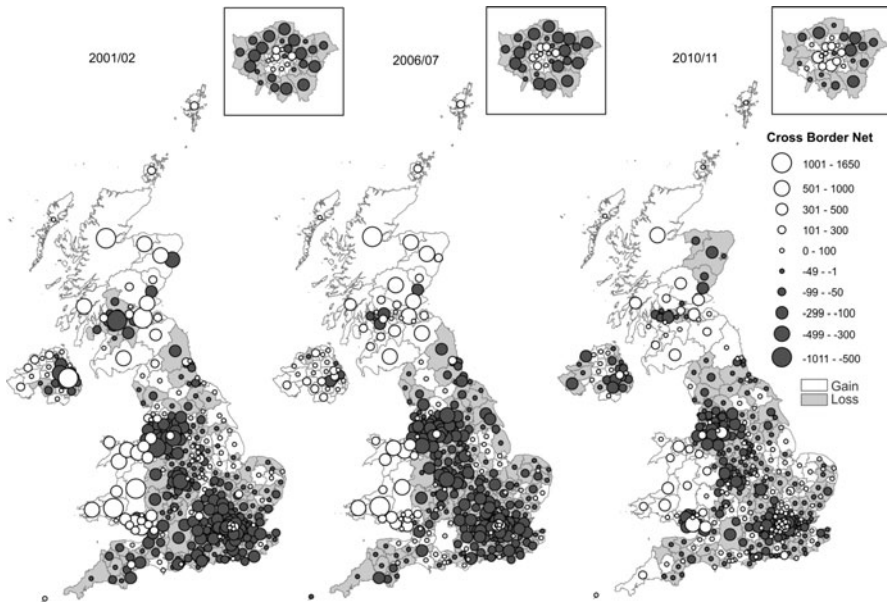


**Fig. 5** Internal net migration balances in 2001/2002, 2006/2007 and 2010/2011

increase in 2006/2007 before a reversal occurs in 2010/2011, where the net migration balance becomes negative.

Northern Ireland exhibits large fluctuations across the time series: the substantial net gain for Belfast in 2001/2002 declines through the decade and LADs in the west of the country move from net gain to net loss. This fluctuation is consistent with the findings of Compton (1992), who, using a time series of migration between Northern Ireland and Great Britain for 1975–1990, finds that the volume of migration varied substantially over time. He attributes this variation to Northern Ireland being very sensitive to economic conditions due to high unemployment, with migrants seeking out labour-deficient regions in Great Britain. Overwhelmingly the pattern of exchanges between LADs in England and the other UK countries is one of net loss. The map for 2010/2011 shows a decline in the size of the net loss in English LADs if not a change in the pattern, although the net gain restricted to central London in the earlier time periods spreads to a number of outer London boroughs.

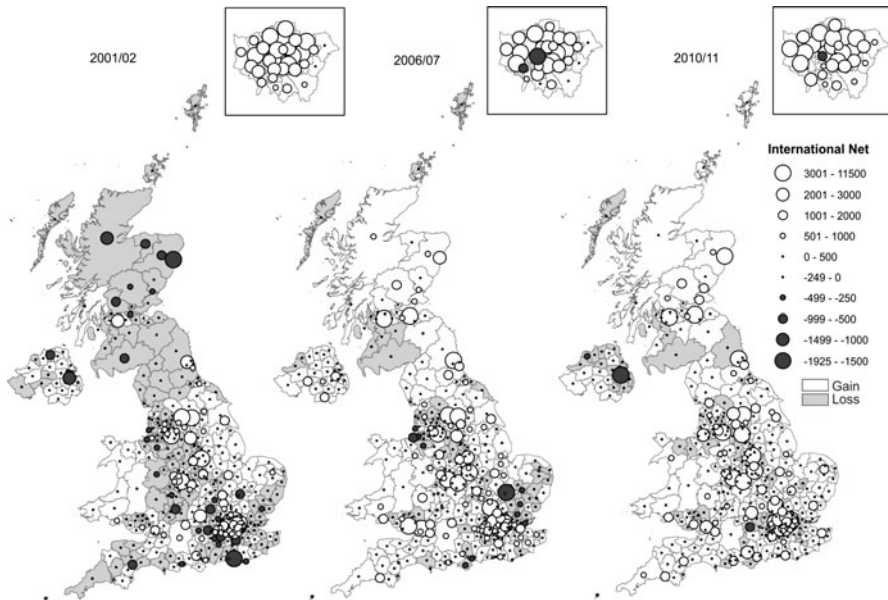
Figure 7 shows that in contrast to internal and cross-border migration, where the largest change is evident in the last year of the time series, international net migration sees the biggest change between 2001/2002 and 2006/2007: the correlation between net flows at the LAD level for these 2 years is 0.73 ( $p < 0.01$ ) whereas the correlation between 2006/2007 and 2010/2011 is stronger at 0.86 ( $p < 0.01$ ). The most striking change between the beginning and end of the decade is the move from net loss to net gain for a large number of LADs in Scotland. This pattern is more striking given the historic trend of high overseas emigration, identified by Jones (1992) as one of the distinctive attributes of



**Fig. 6** Cross-border net migration balances in 2001/2002, 2006/2007 and 2010/2011

Scotland's migration profile. Small net gains in Glasgow and Edinburgh in 2001/2002 become large net gains in 2010/2011 and Aberdeen moves from a position of heavy net loss to having a large positive net migration balance. In England, the pattern changes from one where the majority of LADs were losing net migrants in 2001/2002 to one where most are gaining in 2010/2011, with a clear pattern of net gain that originated in London in the 2001/2002 data beginning to spread across the South East. The effect of London acting as a key destination is identified by Coombes and Charlton (1992), who describe it as a 'transit camp' in terms of a landing point for international immigrants. The wider net migration seen in the South East may be attributed to 'human capital spillovers' (Faggian and McCann 2009, p. 145) where London is the predominant draw. In Northern Ireland, Belfast, after a brief period of net gain in 2006/2007, returns to having a negative balance in 2010/2011.

The extent to which the pattern of net international migration is opposite to that of net internal migration can be seen by comparing Figs. 7 and 9, and is most clear in London and other urban LADs (for example, in the North West) which have net gains of international migrants and net losses of internal migration. The negative correlation in each year shows that the relationships hold across the decade ( $r = -0.67$  in 2001/2002,  $r = -0.74$  in 2006/2007 and  $r = -0.66$  in 2010/2011, all  $p < 0.01$ ). It is clear that London (and to some extent the wider South East region) contributes towards these relationships and plays a key role in redistributing migrants around the UK. The pattern of net gain for international migrants and loss of internal and cross-border migrants suggests that immigrants to the capital quickly become internal outmigrants in favour of other regions. The concept of London and



**Fig. 7** International net migration balances in 2001/2002, 2006/2007 and 2010/2011

the South East as an ‘escalator region’, as set out by Fielding (1992), may go some way towards explaining the large-scale migration activity seen in the region. The ‘escalator region’ attracts a large number of young adults who are mostly well educated and in the early stages of their career, who then subsequently ‘step off’ the escalator to move elsewhere having gained the upward mobility offered by the South East.

## Conclusions

In this paper, we have presented the methodology used in the construction of a set of consistent matrices of estimated origin–destination migration for (1) internal intranational, (2) internal cross-border, and (3) international migration flows in the UK for 2001/2002–2010/2011. The second of these, cross-border flows, have not been estimated across the UK at LAD level before. These matrices draw on estimates and data from administrative sources produced by the three NSAs and an IPF routine has been employed to estimate the gaps. We have sought to highlight the inconsistencies in data and methods used by the three NSAs and suggest that our methodology can be used to produce consistent estimates that will inform the midyear population estimates. Further work will involve the production of age and sex disaggregated estimates in due course and it is envisaged that the results of the estimation procedure can be benchmarked against the 2011 Census when the Special Migration Statistics tables are released.

Using data from three NSAs revealed a number of consistency issues. Different populations are covered by the datasets: for example, students are treated differently in the internal migration datasets, and international migration methodologies vary between the constituent countries. For these two inconsistencies, the data were included unaltered. For the former, the adjustment reallocates students from their parents' address to their term-time address in England, Wales and Northern Ireland but not in Scotland. Although the student adjustment has implications for the allocation of this subpopulation, it does not alter the overall number of migrants in the system. In the latter instance, international estimates are the subject of continuing research at all three NSAs, so any adjustments made in the future could be easily integrated into our matrices without affecting other parts of the system.

The inclusion of armed forces migrants in a 'rest of the UK' group in Scotland had an effect on the IPF routine, meaning that the column and row totals did not sum to the same number of migrants. The solution to this problem was to remove armed forces from this 'rest of the UK' group. All data from the NSAs are transition data reported at midyear-to-midyear (31 June) time intervals which provides for good temporal consistency. Using a census distribution (where the 2001 Census year ran to the end of April) to estimate missing cell values posed the problem of a temporal inconsistency with the marginal totals used in the adjustment (reported at midyear). The census distribution was maintained as it provides the most complete distribution across the decade being estimated.

When the migration patterns estimated in the dataset were examined, it was found that migration propensity corresponded with economic conditions, with the largest flows in 2006/2007 relating to the highest GDP per capita of the time series, alongside a low rate of unemployment. The findings from the analysis of net migration reveal a pattern of counterurbanization for all three time periods, especially for internal migration. Internal migration patterns are most consistent across all time periods and, as with cross-border patterns, the biggest change occurs between the middle (2006/2007) and end (2010/2011) of the series. In contrast, international migration exhibited the largest change between 2001/2002 and 2006/2007, the most notable pattern being a shift from net loss to net gain for a number of LADs in Scotland. Finally the role of London and the South East of England as a region which drives all types of migration can be observed. Net international migration and net internal migration are negatively correlated, and the majority of London boroughs gain international migrants and lose internal migrants. This pattern spreads across the wider South East region in 2006/2007 and 2010/2011.

**Acknowledgments** The authors are grateful to the following for providing information: Michelle Bowen, Alistair Davies, Pete Large and Paul Vickers at the Office for National Statistics; Esta Clarke, Bart Jaron and Ganka Mueller at National Records of Scotland; and Jos Ijpelaar and David Marshall at the Northern Ireland Statistics and Research Agency. The code used for the Iterative Proportional Fitting routine was obtained from the Alaska Department of Labour and Workforce Development (2009). The lead author is supported by an Economic and Social Research Council CASE studentship (ref. ES/H03997X/1ES/H03997X/1) awarded in partnership with the Office for National Statistics. This research used Census data obtained through the MIMAS/CASWEB and CIDER/WICID facilities and GIS boundary data obtained through the EDINA UKBORDERS facility, services supported by ESRC and JISC. The Census and official Mid-Year Estimates for England, Wales, Scotland and Northern Ireland

have been provided by ONS, NRS and NISRA and the digital boundary data by OSGB and OSNI. These data are Crown copyright and are reproduced with permission of OPSI.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Alaska Department of Labour and Workforce Development. (2009). *Iterative proportional fitting R code*. <http://www.demog.berkeley.edu/~eddieh/datafitting.html>. Accessed April 10, 2012.
- Bishop, Y., Fienberg, S., & Holland, P. (1974). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Boden, P., & Rees, P. (2010). Using administrative data to improve the estimation of immigration to local areas in England. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 173, 707–731.
- Champion, T. (1987). Recent changes in the pace of population deconcentration in Britain. *Geoforum*, 18(4), 379–401.
- Champion, T. (1989). *Counterurbanisation: The changing pace and nature of population deconcentration*. London: Edward Arnold.
- Champion, T. (2005). Population movement within the UK. In R. Chappell (Ed.), *Focus on people and migration* (pp. 92–114). Basingstoke: Macmillan.
- Chilton, R., & Poet, R. (1973). An entropy maximizing approach to the recovery of detailed migration patterns from aggregate census data. *Environment and Planning*, 4, 205–233.
- Compton, P. (1992). Migration trends for Northern Ireland: Links with Great Britain. In J. Stillwell, P. Rees, & P. Boden (Eds.), *Migration processes and patterns V2. Population redistribution in the United Kingdom* (pp. 81–99). London: Belhaven Press.
- Cross, D. (1990). *Counterurbanization in England and Wales*. London: Gower Publishing Ltd
- Coombes, M., & Charlton, J. (1992). Flows to and from London: A decade of change? In J. Stillwell, P. Rees, & P. Boden (Eds.), *Migration processes and patterns V2. Population redistribution in the United Kingdom* (pp. 56–77). London: Belhaven Press.
- Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4), 427–444.
- Denteneer, D., & Verbeek, A. (1985). A fast algorithm for iterative proportional fitting in log-linear models. *Computational Statistics & Data Analysis*, 3(1), 251–264.
- Faggian, A., & McCann, P. (2009). Human capital and regional development. In R. Capello & P. Nijkamp (Eds.), *Handbook of regional growth and development theories* (pp. 133–151). Cheltenham: Edward Elgar.
- Fielding, A. J. (1992). Migration and social mobility: South East England as an escalator region. *Regional Studies*, 26(1), 1–15.
- GROS. (2010a). *Methodology used to produce GROS official migration estimates. National records of Scotland*. <http://www.gro-scotland.gov.uk/statistics/theme/migration/methodology.html>. Accessed April 5, 2012.
- GROS. (2010b). *Mid-year population estimates for Scotland: Methodology guide. National Records of Scotland*. <http://www.gro-scotland.gov.uk/files2/stats/population-estimates/mid-year-pop-est-methodology.pdf>. Accessed April 12, 2012.
- House of Commons Select Committee. (2008). *Counting the population: Eleventh report of session 2007–2008*. London: HMSO.
- Johnston, R. J., & Pattie, C. J. (1993). Entropy-maximizing and the iterative proportional fitting procedure. *The Professional Geographer*, 45(3), 317–322.
- Jones, H. (1992). Migration trends for Scotland: Central losses and peripheral gains. In J. Stillwell, P. Rees, & P. Boden (Eds.), *Migration processes and patterns V2. Population redistribution in the United Kingdom* (pp. 100–114). London: Belhaven Press.

- Kalogirou, S. (2005). Examining and presenting trends of internal migration flows within England and Wales. *Population, Space and Place*, 11(4), 283–297.
- Kennett, S. (1980). Migration within and between the Metropolitan Economic Labour Areas of Britain, 1966–1971. In J. Hobcraft & P. Rees (Eds.), *Regional demographic development* (pp. 165–185). London: Croom Helm.
- Lomax, N., Stillwell, J., Norman, P., & Rees, P. (2011). *Data sources and estimation methods for sub-national migration in the United Kingdom*. Working Paper 11/4, Leeds: School of Geography, University of Leeds. <http://www.geog.leeds.ac.uk/fileadmin/documents/research/csap/11-04.pdf>. Accessed May 1, 2012.
- Nair, P. S. (1985). Estimation of period-specific gross migration flows from limited data: Bio-proportional adjustment approach. *Demography*, 22, 133–142.
- National Statistics. (2006). *Report of the inter-departmental task force on migration statistics*. London.
- Norman, P. (1999). *Putting iterative proportional fitting on the researcher's desk*. Working Paper 99/3, Leeds: School of Geography, University of Leeds.
- Northern Ireland Statistics and Research Agency (NISRA). (2006). *How are population estimates created? The methodology*. [http://www.nisra.gov.uk/archive/demography/population/midyear/mye\\_methodology.pdf](http://www.nisra.gov.uk/archive/demography/population/midyear/mye_methodology.pdf) Accessed April 6, 2012.
- Northern Ireland Statistics and Research Agency (NISRA). (2007). *Development of methods/sources to estimate population migration in Northern Ireland*. [http://www.nisra.gov.uk/archive/demography/population/migration/dev\\_est\\_mig.pdf](http://www.nisra.gov.uk/archive/demography/population/migration/dev_est_mig.pdf). Accessed February 10, 2012.
- Northern Ireland Statistics and Research Agency (NISRA). (2010). *Migration statistics for Northern Ireland 2009*. <http://www.nisra.gov.uk/archive/demography/population/migration/Migration%20Report%202009.pdf>. Accessed April 18, 2012.
- Office for National Statistics (ONS). (2007). *Improved methods for estimating international migration: Geographical distribution of estimates of in-migration*. [http://www.statistics.gov.uk/downloads/theme\\_population/Geog\\_distn\\_in-migs.pdf](http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_in-migs.pdf). Accessed February 10, 2012.
- Office for National Statistics (ONS). (2010a). *Estimating international long-term emigration by local authority: Update*. <http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/emigration-detailed-methodology-update.pdf>. Accessed April 10, 2012.
- Office for National Statistics (ONS). (2010b). *Improved estimation of student migration within England and Wales*. <http://www.ons.gov.uk/about-statistics/methodology-and-quality/imps/mig-stats-improve-prog/comm-stakeholders/improvements-2008-pop-est/detailed-methodology-papers/student-adjustment-detailed-methodology.pdf>. Accessed April 10, 2012.
- Office for National Statistics (ONS). (2010c). *Estimating internal migration: customer guidance notes*. <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/internal-migration-methodology/estimating-internal-migration-customer-guidance-notes—november-2012.doc>. Accessed April 10, 2012.
- Office for National Statistics (ONS). (2011a). *Summary quality report for long-term international migration releases*. <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/social-statistics/summary-quality-report-for-international-migration.pdf>. Accessed April 10, 2012.
- Office for National Statistics (ONS). (2011b). *Mid-year estimates short methods guide*. <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/pop-ests/mid-year-population-estimates-short-methods-guide.pdf>. Accessed August 14, 2012.
- Office for National Statistics (ONS). (2013a). *Labour market statistics dataset*. <http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/march-2013/dataset-labour-market-statistics.html>. Accessed February 12, 2013.
- Office for National Statistics (ONS). (2013b). *Preliminary estimate of GDP time series dataset 2012 Q4*. <http://www.ons.gov.uk/ons/rel/gva/gross-domestic-product-preliminary-estimate/q4-2012/tsd-preliminary-estimate-of-gdp-q4-2012.html>. Accessed February 12, 2013.
- Ogilvy, A. A. (1982). Population migration between the regions of Great Britain, 1971–1979. *Regional Studies*, 16(1), 65–73.
- Owen, D., & Green, A. (1992). Migration patterns and trends. In T. Champion & T. Fielding (Eds.), *Migration processes and patterns Volume 1: Research progress and prospects* (pp. 17–40). London: Belhaven Press.
- Raymer, J. (2007). The estimation of international migration flows: A general technique focused on the origin-destination association structure. *Environment and Planning A*, 39(4), 985–995.

- Raymer, J., Rees, P., & Blake, A. (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>. Accessed August 12, 2012.
- Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, *44*(2), 199–223.
- Rees, P., & Duke-Williams, O. (1997). Methods for estimating missing data on migrants in the 1991 British Census. *International Journal of Population Geography*, *3*, 323–368.
- Rees, P., Durham, H., & Kupiszewski, M. (1996). *Internal migration and regional population dynamics in Europe: United Kingdom case study*. Working Paper 96/20, Leeds: School of Geography, University of Leeds.
- Rees, P., Stillwell, J., Boden, P., & Dennett, A. (2009). *A review of migration statistics literature. Incorporated in UK Statistics Authority, Migration statistics: The way ahead report 4 (Final)*. London: UKSA. <http://www.statisticsauthority.gov.uk/reports-correspondence/reports/authority-report-4-migration-statistics-the-way-ahead.pdf>. Accessed August 1, 2012.
- Rees, P., & Willekens, F. (1986). Data and accounts. In A. Rogers & F. Willekens (Eds.), *Migration and settlement: A multiregional comparative study* (pp. 19–58). Dordrecht: Reidel.
- Rogers, A., Willekens, F., & Raymer, J. (2003). Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer*, *55*(1), 56–68.
- Schoen, R., & Jonsson, S. H. (2003). Estimating multistate transition rates from population distributions. *Demographic Research*, *9*, 1–24.
- Stillwell, J. (2013). Internal migration, ethnic groups in the UK. In I. Ness (Ed.), *The encyclopedia of global human migration*. London: Blackwell Publishing Ltd.
- Stillwell, J., & Duke-Williams, O. (2007). Understanding the 2001 Census interaction data: The impact of small cell adjustment and the problem of comparison with 1991. *Journal of the Royal Statistical Society Series A (General)*, *170*(2), 1–21.
- Stillwell, J., & Harland, K. (2010). Analysing interaction data. In J. Stillwell, O. Duke-Williams, & A. Dennett (Eds.), *Technologies for migration and population analysis: Spatial interaction data applications* (pp. 69–88). New York: IGI Global.
- Stillwell, J., Rees, P., & Boden, P. (1992). Internal migration trends: An overview. In J. Stillwell, P. Rees, & P. Boden (Eds.), *Migration processes and patterns V2. Population redistribution in the United Kingdom* (pp. 28–55). London: Belhaven Press.
- UK Statistics Authority (UKSA). (2009). *Migration statistics: The way ahead? Report 4 (Final)*, London. <http://www.statisticsauthority.gov.uk/reports-correspondence/reports/authority-report-4-migration-statistics-the-way-ahead.pdf>. Accessed August 1, 2012.
- 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*(2), 137–159.
- Willekens, F. J. (1980). Entropy, multiproportional adjustment and the analysis of contingency tables. *Systemi Urbani*, *2*(3), 171–201.
- Willekens, F. J. (1982). Multidimensional population analysis with incomplete data. In K. Land & A. Rogers (Eds.), *Multidimensional mathematical demography* (pp. 43–111). New York: Academic Press.
- Willekens, F. J. (1983). Log-linear modelling of spatial interaction. *Papers of the Regional Science Association*, *52*, 187–205.
- Willekens, F. J., Por, A., & Raquillet, R. (1981). Entropy, multiproportional, and quadratic techniques for inferring patterns of migration from aggregate data. In A. Rogers (Ed.), *Advances in multiregional demography* (pp. 84–106). Laxenburg: International Institute for Applied Systems Analysis.
- Wong, D. (1992). The reliability of using the iterative proportional fitting procedure. *The Professional Geographer*, *44*(3), 340–348.
- Založnik, M. (2011). *Iterative proportional fitting: Theoretical synthesis and practical limitations*. Ph.D thesis, University of Liverpool.