

Improved Methods for Estimating International Migration

Out-migration - Propensity to Migrate Models for 2002 to 2006

ONS has developed improved methods for the way international out-migrant estimates from the International Passenger Survey (IPS) are distributed between Local Authorities (LAs) in England and Wales. The improved method distributes IPS figures to local areas using factors that reflect the propensity of the population to migrate. This paper provides an update on the models used for both the revisions of the 2002 to 2005 mid-year population estimates and for calculating the 2006 estimates. This is part of a package of changes implemented on 22nd August 2007. For further information see:

http://www.statistics.gov.uk/downloads/theme_population/Geog_distn_out-migs.pdf.

1 The 2006 model

For 2006, the modelling, based on the forward stepwise regression procedure, resulted in the selection of five variables. A summary of the model is given in Table 1.

Table 1: Explanatory variables in the model to predict emigration propensities, 2006

Stepwise Regression for 2006	Coefficients
Constant term in the model	0.0055023045
<i>Standardised variables selected out of 112:</i>	
MIG1: 2006 Estimates of international in-migration from new methodology / 2005 Population	0.0022820893 (Positive Association)
POP25: Percentage of males aged 16-34 with limiting long term illness based on 2001 Census data	0.0010112518 (Positive Association)
SOC2: Percentage people in higher professional SEC classification based on 2001 Census data	0.0023374032 (Positive Association)
POPF13: 2005 proportion of population who are females aged 40-44	-0.0005293653 (Negative Association)
POP4: Percentage of males in 2005 population	-0.0005804334 (Negative Association)

Explanatory power of the model: $R^2 = 91.0\%$

The strongest predictor of emigration in an area is the population of that area. This is reflected in the constant term in the model, the size of which indicates the importance of population. This is the case for all years 2002 to 2006. The results of the model predict the propensity of the population in each area to migrate. The variables are presented in the order they are selected for the model. Thus the first variable, higher professional occupations, is the most powerful predictor of propensity to migrate. Coefficients have been quoted to several decimal places to facilitate onwards work.

The R^2 value of 91.0 per cent indicates how good the model fit is (specifically, it indicates the percentage of the variance that is explained by the model). As a comparison a model based on population alone – broadly equivalent to the current methodology – has an R^2 of 42.2%.

This model provides a plausible model of propensity to migrate, given what is known of migration drivers. For example we know previous migrants have a higher than average propensity to migrate again (a third of all foreign-born immigrants to the UK in the 1990s were estimated to have emigrated again within four years of arrival). The proportion of people in higher professional social economic occupations is a factor related to international mobility. The proportion of females aged 40 to 44 is an indicator of women who are more likely to have families with children and are therefore less likely to migrate.

The proportion of males aged 16-34 with limiting long-term illness has a positive association with out-migration and the percentage of males in the population has a negative association. Neither of these have any easy intuitive explanation. However, regression models select variables to explain the remaining variance in the data and therefore some variables are chosen solely to balance the effects of other variables already selected by the model. It is likely that these variables have been chosen as balancing factors rather than for any direct correlation with out-migration.

2 The 2005 model

For 2005, the modelling, based on the forward stepwise regression procedure, resulted in the selection of four variables. A summary of the model is given in Table 2.

Table 2: Explanatory variables in the model to predict emigration propensities, 2005

Stepwise Regression for 2005	Coefficients
Constant term in the model	0.0052724307
<i>Standardised variables selected out of 112:</i>	
MIG1: 2005 Estimates of international in-migration from new methodology / 2004 Population	0.0016504634 (Positive Association)
HOU4: Proportion of persons per household based on 2001 Census data	-0.0015009081 (Negative Association)
SOC2: Percentage people in higher professional SEC classification based on 2001 Census data	0.0016475483 (Positive Association)
ETH9: 2004 Estimates of people in Asian ethnic group / 2004 Population	0.0006755189 (Positive Association)

Explanatory power of the model: $R^2 = 89.3\%$

The most important variables selected in 2005 are the same as for the 2006 model. Of the other variables, the proportion of persons per household is a factor that correlates highly with indices of deprivation, where international mobility tends to be lower in areas of deprivation. Again, the proportion of persons with Asian ethnicity is likely to be a balancing factor.

3 The 2004 model

The 2004 modelling process resulted in the selection of five variables. A summary of the model is given in Table 3.

Table 3: Explanatory variables in the model to predict emigration propensities, 2004

Stepwise Regression for 2004	Coefficients
Constant term in the model	0.0052998449
<i>Standardised variables selected out of 112:</i>	
MIG1: 2004 Estimates of international in-migration from new methodology / 2003 Population	0.0012226571 (Positive Association)
POPF7: 2003 proportion of population who are females aged 10-14	-0.0007795230 (Negative Association)
ETH10: 2003 Estimates of people in Black ethnic group / 2003 Population	-0.0016651401 (Negative Association)
POP23: Population Density (persons per hectare based on 2003 mid-year population)	0.0017569647 (Positive Association)
EDU1: Percentage people aged 16-74 with higher educational qualifications (levels 4/5) based on 2001 Census data	0.0014435533 (Positive Association)

Explanatory power of the model $R^2 = 89.0\%$

The percentage of people with higher educational qualifications plays a similar role to persons in higher professional occupations selected in 2005. The proportion of people with Black ethnicity is a factor that correlates highly with indices of deprivation, where international mobility tends to be lower in areas of deprivation. The proportion of females aged 10 to 14 is an indicator of families with children who are at a critical stage of their schooling, and therefore are less likely to migrate. Among the possible variables by age and sex the number of females was selected by the model as the best single proxy indicator. The density measure is an indicator of an urban area that is likely to have mobile populations, for example students, professionals and migrant workers.

4 The 2003 model

For 2003, the modelling process resulted in the selection of four variables. A summary of the model is given in Table 4.

Table 4: Explanatory variables in the model to predict emigration propensities, 2003

Stepwise Regression for 2003	Coefficients
Constant term in the model	0.0052444178
<i>Standardised variables selected out of 112:</i>	
MIG1: 2003 Estimates of international in-migration from new methodology / 2002 Population	0.0011334490 (Positive Association)
POPF7: 2002 proportion of population who are females aged 10-14	-0.0011873471 (Negative Association)
SOC2: Percentage people in higher professional SEC classification based on 2001 Census data	0.0022125578 (Positive Association)
POP31: Percentage of females aged 0-15 with limiting long term illness based on 2001 Census data	0.0008886071 (Positive Association)

Explanatory power of the model $R^2 = 85.7\%$

This model contains similar variables to those included in the models for other years.

5 The 2002 model

For 2002, the modelling process resulted in the selection of three variables. A summary of the model is given in Table 5.

Table 5: Explanatory variables in the model to predict emigration propensities, 2002

Stepwise Regression for 2002	Coefficients
Constant term in the model	0.0048320194
<i>Standardised variables selected out of 112:</i>	
EDU1: Percentage people aged 16-74 with higher educational qualifications (levels 4/5) based on 2001 Census data	0.0019075732 (Positive Association)
MIG1: 2002 Estimates of international in-migration from new methodology / 2001 Population	0.0011174181 (Positive Association)
POPF7: 2001 proportion of population who are females aged 10-14	-0.0007782372 (Negative Association)

Explanatory power of the model: $R^2 = 86.5\%$

This model is similar to the model for 2004, with all three variables having been selected in that year and with similar coefficients.