Forecasting Age Patterns of International Migration: Adapting and Extending the Lee-Carter Model to Different Data Types and Time Series

James Raymer[1,2] and Arkadiusz Wiśniowski[1]

[1] *ESRC Centre for Population Change, University of Southampton*
[2] *Australian Demographic and Social Research Institute, Australian National University*

15 November 2013 (Draft)

1. Introduction

In this paper, we explore the flexibility of Lee-Carter type models to forecast age-specific migration. Lee-Carter (1992) and their variants have been widely used to forecast age-specific mortality and, to a lesser extent, age-specific fertility. Like mortality and fertility, migration exhibits regularities in age profiles over time and across space. Thus, the notion of using a fixed age profile as a basis for modelling is appealing. However, unlike fertility and mortality, the quality of migration data is often considerably worse with differences in measurement across countries.

To illustrate the flexibility of the Lee-Carter modelling framework, we forecast age-specific immigration and emigration for the United Kingdom and Sweden representing passenger survey data and population register data, respectively. The data represent the main sources for each county. The United Kingdom data is the most problematic because it is based on a sample of roughly 200-300 thousand passengers a year with intended migrants comprising about one per cent of the sample. While the age profiles of immigration and emigration are fairly consistent over time, irregularities occur across age due to the relatively small sample size.

2. Background

Migration is considered the most difficult demographic component to forecast in population accounting models. There are three main reasons for this. The first is the vast heterogeneity of migrants. Migration occurs for many reasons including, for example, labour, family reunion, asylum, and retirement. The second reason is that migration is poorly measured. Unlike fertility and mortality, there is no clear observable event. Countries choose different criteria to qualify migrants and there exist many different data collection systems from population or migration registers to censuses and surveys. Often these data are not available or consistent over time, thus providing a weak data base for forecasting. The third reason is that migration levels are influenced strongly by economics and politics. In order to effectively forecast migration, one must also be able to forecast future economic and political conditions.

Because of the reasons outlined above, there has not been much work carried out in migration forecasting. National statistical offices often rely on very simple assumptions regarding net migration based only on a few recent years or assume future net migration to be zero. Age and sex patterns are either ignored or kept fixed (Bijak 2010). This is unfortunate because migration is often the most influential component of population growth, especially in developed societies.

3. Models

For modelling age patterns of migration, we consider an extension of the Lee-Carter (1992) model. In general, we assume that counts of migrants follow a Poisson distribution, with an age- and time-specific mean, which is log-normally distributed. Emigration counts are denoted by $E_{x,t}$ for single age group x and in year t. The corresponding rates of emigration are denoted by $\gamma_{x,t}$. Here,

$$E_{x,t} \sim \text{Poisson}(\gamma_{x,t} R_{x,t}) \quad (1)$$

$$\log(\gamma_{x,t}) = a_x^E + b_x^E k_t^E + \xi_{x,t}^E, \quad (2)$$

where $R_{x,t}$ denotes the population at risk, a_x and b_x are age-specific model parameters, k_t is time specific parameter and $\xi_{x,t}$ is the normally distributed error term with variance σ^2. The superscript E is used to distinguish between the model for emigration from the model (below) for immigration (superscript I). For immigration, we are interested in forecasting counts, $\varphi_{x,t}$. The corresponding equations are

$$I_{x,t} \sim \text{Poisson}(\varphi_{x,t}) \qquad (3)$$

$$\log(\varphi_{x,t}) = a_x^I + b_x^I k_t^I + \xi_{x,t}^I. \qquad (4)$$

We consider two types of specifications of the time-specific parameter k_t for both emigration and immigration: (i) univariate and (ii) bivariate. In the most general univariate case, we assume that k_t for both emigration and immigration follow an autoregressive process of order one, AR(1), with drift and with logarithmic and linear trends. In the bivariate case, we assume that k_t for emigration and immigration follow a vector autoregressive, VAR(1), process with drift and trends. The bivariate model permits exploration of the correlation between the time patterns exhibited by the rates of emigration and the counts of immigration.

To estimate parameters of the models and to produce forecasts, we use Bayesian inference. The Bayesian approach provides a coherent framework for simultaneous estimation as well as direct assessment of the uncertainty of all model parameters and the forecasts of emigration and immigration. The approach also allows the incorporation of expert opinions in form of the informative prior distributions.

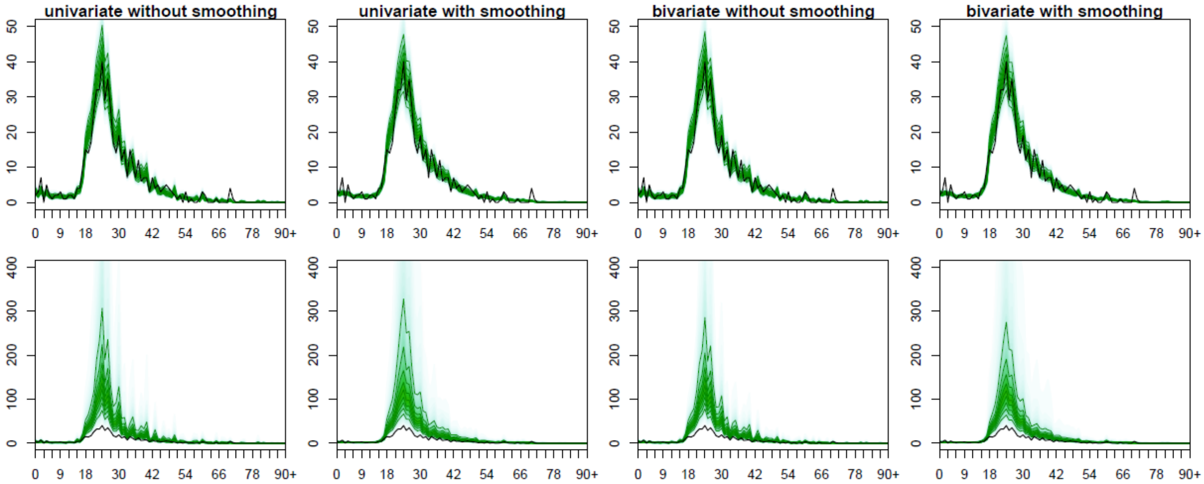
The data on emigration and immigration in the UK are based on sample estimates from an International Passenger Survey, which results in irregularities due to the small sample size. To overcome this problem, we apply age-specific smoothing techniques for model parameters a_x and b_x. Here, the priors for a_x and b_x are normal distributions. For a given age group x (apart from the youngest and oldest), the expected value is the mean of the two neighbouring age groups x-1 and x+1. For the youngest and oldest groups, the expected value is the second youngest and second to the oldest age, respectively. The variances for each of the prior distributions are constructed in a way that the unconditional variance is constant for all age groups. We show that this simple technique effectively removes the large irregularities occurring in the historical data underlying the forecasts.

5. Results

5.1 United Kingdom

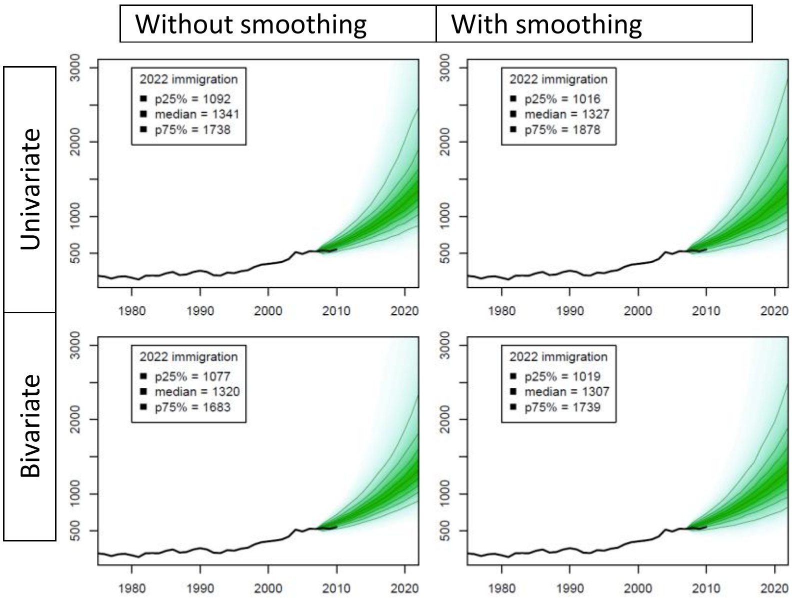
Various specifications of the Lee-Carter models lead to different results in terms of forecasts. Here, we predict immigration of females to the UK for years 2008-2022, based on the data for 1975-2007. As an illustration, consider four Poisson-log normal versions of the Lee-Carter model, with and without smoothing for parameters a_x and b_x and with univariate or bivariate models for the time components k_t for emigration and immigration, as described in previous section. In Figure 1, we present the forecasted age profiles of immigration counts for years 2008 and 2022. We observe that smoothing helps to reduce the irregularities, especially in ages 15-50, which are observed in the reported data (black line). Applying the bivariate model for the time components reduces the uncertainty of the forecasted age profiles in models both with and without smoothing. This leads to a conclusion that there is a correlation between the emigration and immigration patterns over time and including it in the model may reduce the uncertainty.

In Figure 2, we present the forecasted total female immigration to the UK over time. We observe that the forecasts increase rapidly. This result is most likely due to the recent increase in immigration flows since late 1990s, despite the stabilisation period after 2005. We observe that the bivariate model for the time component leads to lower uncertainty of the forecasts. However, when smoothing techniques are applied, uncertainty of the forecasted total immigration increases, which is clearly observed in the case of the bivariate model (bottom right plot).



Note: Top row plots concern 2008 forecasts, bottom row plots 2022. Black lines are the observed age profiles in 2007. Dark green lines denote deciles of the posterior distribution.

Figure 1: Forecasted age profiles of immigration counts (in thousands) for 2008 and 2022



Note: Black lines denote observed immigration counts from the International Passenger Survey for 1975-2010. Dark green lines are deciles of the posterior distribution of the total immigration in 2008-2022.

Figure 2: Forecasted total immigration to the UK, 2008-2022

5.2 Sweden

For Sweden, we predict emigration and immigration counts of males and females to 2025, based on the data for 1968-2012. In Figure 3, we present the forecasted age profiles for years 2012 and 2025. Here, applying the bivariate model for the time components increases the uncertainty of the forecasted age profiles for female children because of a recent and increased pattern of 15-16 year old male migration, which did not occur in earlier years or for females. This suggests a univariate model may be more appropriate. The forecasted aggregate totals of male and female migration are presented in Figure 4. Swedish immigration levels over time have exhibited considerably more varied patterns resulting in much higher levels of forecasted uncertainty.

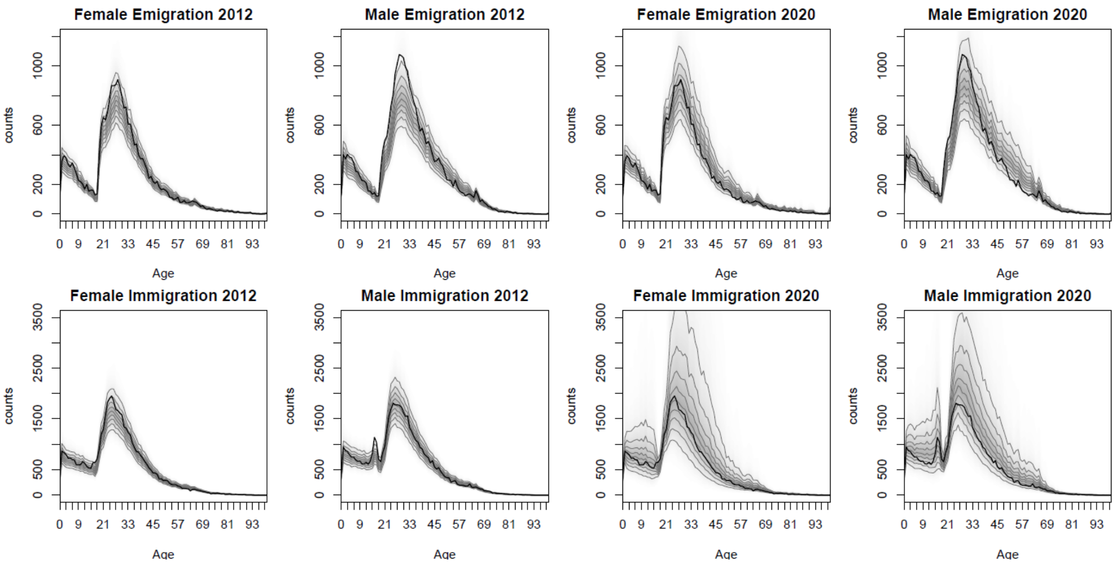


Figure 3: Forecasted age profiles of emigration and immigration counts for Sweden, 2012 and 2020

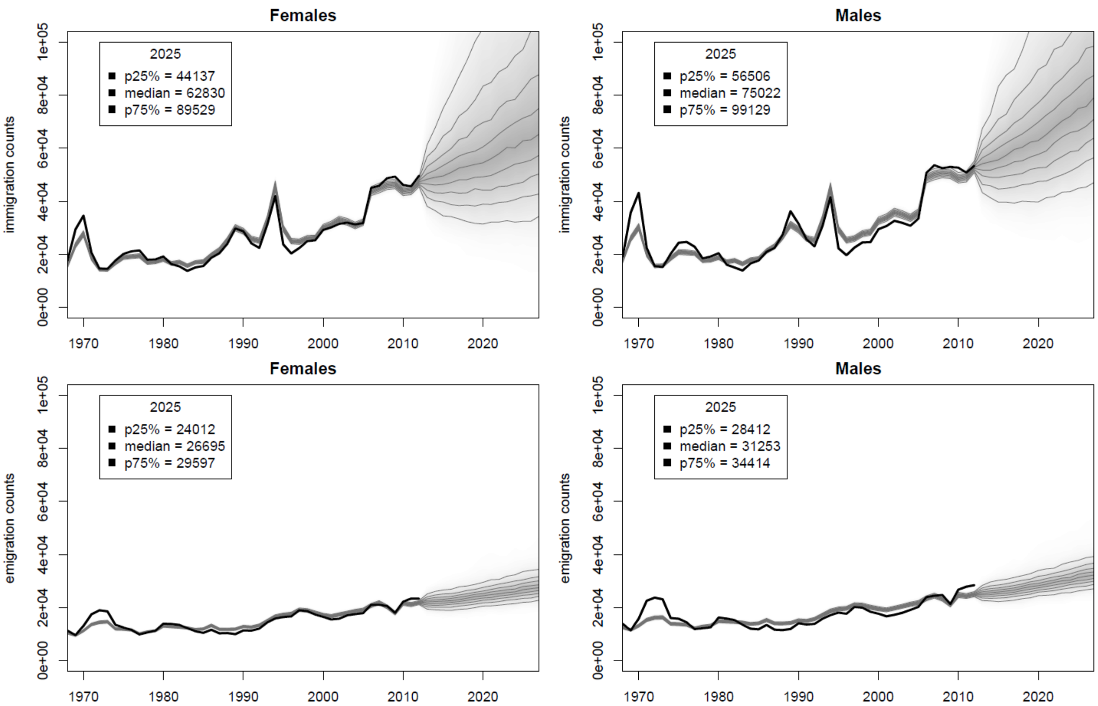


Figure 4: Forecasted total immigration and emigration for Sweden, 2012-2025

6. Conclusion

There are two main contributions of this paper. First we have explored the application of the widely used Lee-Carter framework developed for forecasting mortality to forecast emigration rates and immigration events for the UK and Sweden. Second, we have shown how the model framework can be extended to include bivariate relationships and smoothing. The bivariate model is important for capturing the strong correlations exhibited by patterns of immigration and emigration over time. The smoothing is necessary to limit the effect of the age-specific irregularities contained within the International Passenger Survey data.

References

Bijak J (2011) *Forecasting international migration in Europe: A Bayesian view*. Springer: Dordrecht.

Lee RD and Carter LR (1992) Modeling and forecasting US mortality. *Journal of the American Statistical Association* 659-671.