WP 13.1
16 October 2013

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint Eurostat/UNECE Work Session on Demographic Projections
organised in cooperation with Istat
(29-31 October 2013, Rome, Italy)

Item 13 – Bayesian approaches

Bayesian Probabilistic Projection of International Migration Rates

Jonathan J. Azose, University of Washington
Adrian E. Raftery, University of Washington / University College Dublin

Bayesian Probabilistic Projection of International Migration Rates

Jonathan J. Azose
University of Washington

Adrian E. Raftery
University of Washington and University College Dublin

September 26, 2013

Abstract

We propose a method for obtaining joint probabilistic projections of migration rates for all countries, broken down by age and sex. Joint trajectories for all countries are constrained to satisfy the requirement of zero global net migration. We evaluate our model using out-of-sample validation and compare point projections to migration rates produced using a persistence model.

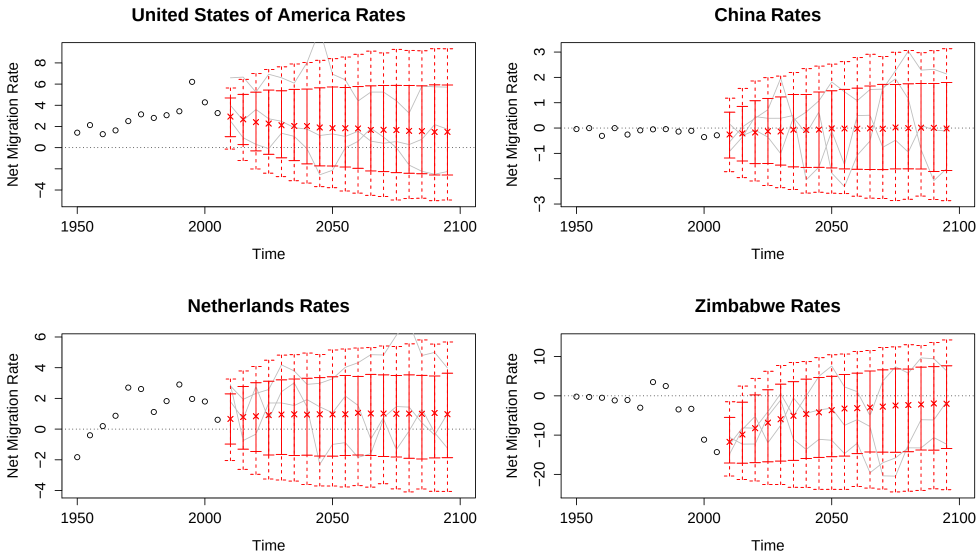

Figure 1: Net International Migration Probabilistic Projections: 80% and 95% predictive intervals for four countries, with example trajectories included in gray.

1 Introduction

In this paper we propose a method for probabilistic projection of net international migration rates. Our technique is a simple one which nonetheless overcomes some of the usual difficulties of migration projection. Firstly, we produce both point and interval estimates, providing a natural quantification of uncertainty. Secondly, since our model uses only demographic variables as input, we can make long-term projections without explosion in the degree of uncertainty. Thirdly, simulated trajectories from our model satisfy the common sense requirement that worldwide net migration sum to zero. Lastly, we sidestep the difficulty in projecting a complete large matrix of pairwise flows by instead working directly with net migration rates. Sample projections from our model for a diverse selection of countries are given in Figure 1.

2 Methods

2.1 Data

We use data from the 2010 revision of the United Nations Population Division's biennial *World Population Prospects* (WPP) report (United Nations Population Division, 2011). WPP reports contain both estimates and projections of countries' age- and sex-specific fertility and mortality rates as well as estimates and point projections of their net migration rates.

The quantity we are interested in forecasting is $r_{c,t}$, the net annual migration rate for country c in time period t, reported in units of migrants per thousand individuals in the WPP data. For calculations, we sometimes convert *rates* $r_{c,t}$ to corresponding *counts* $y_{c,t}$. Our method also requires knowledge of the average population of countries, $n_{c,t}$, indexed by country and time, and projections of $n_{c,t}$ into the future for all countries.

2.2 Probabilistic Projection Method

Our technique is to fit a Bayesian hierarchical first-order autoregressive (AR(1)) model to net migration rate data for all countries. We model the migration rate $r_{c,t}$ in country c and time period t as

$$(r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \varepsilon_{c,t}.$$

We put normal priors on each country's theoretical equilibrium migration rate μ_c, and a uniform prior on the autoregressive parameter ϕ_c. Full specifications of priors and hyperpriors are below.

$$\text{Level 1} \begin{cases} (r_{c,t} - \mu_c) = \phi_c(r_{c,t-1} - \mu_c) + \varepsilon_{c,t} \\ \varepsilon_{c,t} \overset{\text{ind}}{\sim} N(0, \sigma_c^2) \end{cases}$$

$$\text{Level 2} \begin{cases} \phi_c \overset{\text{iid}}{\sim} U(0,1) \\ \mu_c \overset{\text{iid}}{\sim} N(\lambda, \tau^2) \\ \sigma_c^2 \overset{\text{iid}}{\sim} IG(a, b) \end{cases}$$

$$\text{Level 3} \begin{cases} a \sim U(1, 10) \\ b|a \sim U(0, 100(a-1)) \\ \lambda \sim U(-100, 100) \\ \tau \sim U(0, 100). \end{cases}$$

We obtain draws from the posterior distributions of all parameters using Markov Chain Monte Carlo methods. In our implementation, we use the Just Another Gibbs Sampler (JAGS) software package for Markov chain simulations (Plummer, 2003).

Having obtained a sample of independent draws from the joint distribution of the parameters, we use these draws to obtain a sample from the joint posterior predictive distribution. For each sampled point θ_k from the joint posterior distribution of the parameters, we simulate a set of joint trajectories $\tilde{r}_{c,t}^{(k)}$ for net migration rates at time points until 2100. However, this procedure generally produces trajectories which are impossible in that they give nonzero global net migration counts. We create corrected net migration rate trajectories $\tilde{r}_{c,t}^{*(k)}$ as follows:

1. On the basis of parameter vector θ_k, project net migration rates for all countries a single time point into the future. Denoting the next time period in the future as t', this allows us to obtain a collection of (uncorrected) projected values $\tilde{r}^{(k)}_{c,t'}$ for all countries c.

2. Convert net migration rate projections $\tilde{r}^{(k)}_{c,t'}$ to net migration count projections $\tilde{y}^{(k)}_{c,t'}$. This is done by multiplying by a fixed point projection of each country's population, $\tilde{n}_{c,t'}$. We obtain these point projections from WPP 2010 (United Nations Population Division, 2011).

3. Further break down migration counts by age a and sex s to obtain estimates of net male and female migration counts for all countries and age groups, $\tilde{y}^{(k)}_{c,t',a,s}$. This is accomplished by applying projected model migration schedules to all countries. We use the naïve projection method which takes each country's age- and sex-specific migration schedule to be the same as the migration schedule in the most recent time point for which detailed data were available for that country.

4. Within each age and sex category, apply a correction to ensure zero worldwide net migration. The correction we apply redistributes any overflow migrants to all countries, proportional to their projected populations. Specifically, take the corrected migration count projection $\tilde{y}^{*(k)}_{c,t',a,s}$ to be

$$\tilde{y}^{*(k)}_{c,t',a,s} = \tilde{y}^{(k)}_{c,t',a,s} - \frac{\tilde{n}_{c,t'}}{\sum_c \tilde{n}_{c,t'}} \sum_c \tilde{y}^{(k)}_{c,t',a,s}.$$

5. Convert the corrected age- and sex-specific net migration counts $\tilde{y}^{*(k)}_{c,t',a,s}$ back to corrected net migration rates $\tilde{r}^{*(k)}_{c,t'}$ by disaggregating and converting counts to rates in the obvious way.

6. Continue projecting trajectories one time step at a time into the future by repeating steps 1-5.

Note that although the uncorrected net migration rates $\tilde{r}_{c,t'}$ come from the desired marginal posterior predictive distributions, the correction in step 4 changes those distributions by projecting them onto a lower dimensional space. Sensitivity analysis suggests that the correction introduces only minor changes in the marginal distributions with and without the correction, as measured by estimating the Kolmogorov-Smirnov distance between the two distributions.

3 Results

3.1 Evaluation

We do not know of any other model that produces *probabilistic* projections of all countries' net international migration rates. However, we can take our model's median projections to be point projections and compare against models that produce point projections only. As a baseline for comparison, we evaluate against the naïve *persistence model* which projects migration rates to continue at the most recently observed levels indefinitely into the future. In the short horizon, the persistence model is similar to the expert knowledge-based projections in the WPP (United Nations Population Division, 2011).

Our historical data consist of a series of migration rates $r_{c,t}$ for 197 countries at 12 time points with five-year resolution, spanning the period from 1950 to 2010. We performed an out-of-sample evaluation by holding out the data from the m most recent time points for all countries and producing posterior predictive distributions on the basis of the remaining $12 - m$ time points. For point forecasts we used the median of the posterior predictive distribution.

We report out-of-sample mean absolute error as a measure of the quality of point forecasts and both interval scores and coverage as measures of quality of our interval predictions. Interval scores provide a negatively oriented score for probabilistic forecasts which rewards sharpness subject to calibration (Gneiting and Raftery, 2007). If we predict a distribution with $(1 - \alpha)$ predictive interval (ℓ, u) and the value x arises, the interval score is given by

$$(u - \ell) + \frac{2}{\alpha}(\ell - x)\mathcal{I}\{x < \ell\} + \frac{2}{\alpha}(x - u)\mathcal{I}\{x > u\}.$$

The intuition is that we are rewarded for providing forecasts with tight bounds and punished if observations fall outside those bounds.

Table 1 contains these evaluation metrics for our Bayesian hierarchical model and the mean absolute errors for the persistence model. Across the board, our point projections outperform the persistence model, and our interval projections achieve reasonably good calibration.

3.2 Projections for the least-developed countries

The United Nations publishes a list of the least-developed countries, with countries classified as least-developed based on assessments of their economic vulnerability, human capital, and gross national income (Committee for Development Policy and United Nations Department of Economic and Social Affairs, 2008). A total of 46 countries in our data fall into the

Table 1: Mean absolute errors, interval scores, and predictive interval coverage for our Bayesian hierarchical model and the persistence model.

# of points held out	Model	MAE	80% I.S.	95% I.S.	80% Cov.	95% Cov.
1 point	Bayesian	3.24	24.5	48.4	91.4%	96.4%
	Persistence	3.57	—	—	—	—
3 points	Bayesian	4.76	27.1	48.1	84.9%	93.4%
	Persistence	6.74	—	—	—	—
6 points	Bayesian	5.12	30.1	60.4	77.2%	89.3%
	Persistence	7.17	—	—	—	—

Table 2: Mean projected change in migration rates among least-developed countries (LDC) versus all other countries (Other).

	LDC	Other
By 2020	+0.02	-1.49
By 2040	+0.29	-2.12
By 2060	+0.34	-2.29

least-developed category. We now consider briefly the projections that our model makes for these least-developed countries in comparison to all other countries.

In the 2005-2010 time period, only 26% of the least-developed countries were net receivers of migration, as compared to 43% of all other countries. Our model projects that this discrepancy will lessen somewhat in the future. By 2050, we project an average of 45% of the countries which currently fall into the least-developed category and 49% of other countries to be net receivers.

Additionally, in 2005-2010, the least-developed countries had an average net migration rate of -0.97 (per thousand), compared with an average of 2.64 (per thousand) in all other countries. Our model again projects that this gap in migration between least-developed and all other countries will narrow over time. Over the next decades, we project growth in net migration rates among the least developed countries and decline in net migration rate on average across all other countries, as summarized in Table 2.

Acknowledgements: This work was supported by the U.S. National Institute of Child Health and Development through grants nos. R01 HD054511 and R01 HD070936, and by a Science Foundation Ireland ETS Walton visitor award, grant reference 11/W.1/I2079.

References

Committee for Development Policy and United Nations Department of Economic and Social Affairs (2008, November). *Handbook on the Least Developed Country Category: Inclusion, Graduation, and Special Support Measures.*

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pp. 20–22.

United Nations Population Division (2011). *World Population Prospects: The 2010 Revision.* United Nations.