Statistical Modelling of International Migration Flows

Jakub Bijak and Arkadiusz Wiśniowski, on behalf of the IMEM team[1]
Southampton Statistical Sciences Research Institute (S3RI), University of Southampton
Highfield, Southampton SO17 1BJ, United Kingdom
E-mail: J.Bijak@soton.ac.uk and A.Wisniowski@soton.ac.uk

Introduction and Background

Estimation of international migration flows jointly for a system of countries is a difficult and at times very risky task, potentially characterised by very high levels of uncertainty. First of all, many pieces of data on migration, even for developed countries, are missing. Secondly, where statistical information is available, the volume of migration reported by the receiving country of migrants can differ widely from the one reported by the sending country. For this reason, according to Kupiszewska and Nowok (2008: 46), statistics on flows are often dually reported in *double-entry matrices*, following the seminal ideas introduced by the United Nations (1978) and Kelly (1987). Nevertheless, this approach, although useful for analytical purposes, does not answer the ultimate question on the on the magnitude of internationally consistent and harmonised estimates of flows. This has a significant impact on the population estimates of both receiving and sending countries. The size and composition of population stocks, in turn, form a very important basis of policy making at various levels: from local and sub-national, through national, to supra-national, for example of the European Union (EU). Needless to say, the efficiency of resulting policies, based on such estimates, can be compromised by the inadequate information on international migration.

The problems mentioned above have several root causes[2]. First of all, various countries adopt different definitions as to who qualifies as a migrant for statistical purposes. This is despite the presence of standardised international recommendations on migration statistics (United Nations 1998), according to which a *long-term migrant* should be defined as (*idem*: 18):

> "a person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence."

In practice, various criteria on the duration of stay of prospective migrants are applied throughout Europe, usually ranging from three months to one year. These criteria are sometimes different for immigration and emigration, and for various subpopulations of migrants. Besides, two additional criteria can be also used in migration statistics: no time limit, whereby prospective migrants are just required to register with relevant authorities, and permanent stay, only including those who secured a right of permanent residence in a given country. Country-specific details on definitions used in the EU have been comprehensively covered by Poulain et al. (2006), Kupiszewska and Nowok (2008) and Nowok (2010).

In addition to definitional problems, data on migration in Europe are collected through a variety of mechanisms: from relatively accurate interlinked population registers in the Nordic countries, through standalone registers in most of the EU, to sample-based surveys in Cyprus, Ireland and the United Kingdom (*idem*). Furthermore, the coverage of specific subpopulations can also differ between European countries

---

[1] IMEM Team: James Raymer, Jonathan J. Forster, Peter W.F. Smith, Jakub Bijak, Arkadiusz Wiśniowski and Guy Abel, Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, Southampton, United Kingdom; Nico Keilman and Solveig Christiansen, Department of Economics, University of Oslo, Oslo, Norway; Rob van der Erf, Joop de Beer and Jeanette Schoorl, Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands. All calculations by Arkadiusz Wiśniowski. This research is funded from the grant 'IMEM: Integrated Modelling of European Migration' of NORFACE (New Opportunities for Research Funding Co-operation in Europe, www.norface.org). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors, and should not be attributed in any manner to the institutions, with which they are affiliated.

[2] For an overview of issues related to estimation of migration, see e.g. Bilsborrow et al. (1997) and Poulain et al. (2006).

with respect to various groups of foreign nationals, irregular migrants, specific subgroups (e.g. students), etc. On top of that, migration generally tends to be underreported in official statistics, which problem is more serious in the case of emigration than immigration (Poulain et al. 1996).

The problems with European migration data were acknowledged many years ago, which has led to a series of EU-funded research endeavours ultimately aiming to achieve a harmonisation of migration statistics at the level of the European Union. Here, important examples of projects include an inventory of meta-information on European migration statistics (THESIM: Towards Harmonised European Statistics on International Migration), described by Poulain et al. (2006), and the first attempt to harmonise estimates of migration flows and migrant stocks for 31 European countries (MIMOSA: Migration Modelling for Statistical Analyses). The latter project in the context of modelling migration flows is discussed for example by Raymer et al. (2011). At the same time, from the policy perspective, the recent Regulation (EC) No. 862/2007 of the European Parliament and of the Council on *Community Statistics on Migration and International Protection*[3] not only intensified the efforts to harmonise migration statistics across the EU, through requiring the Member States to conform to the United Nations (1998) recommendations, but also explicitly allowed statistical models to be used in the estimation (Article 9).

In this context, the aim of this paper is to present one of the further steps in the harmonisation process, directly in the spirit of Article 9 of Regulation 862/2007. In particular, we discuss a comprehensive statistical model of international migration, applied to an interlinked system of European countries. The exposition is based on the example of a dedicated model 'IMEM' (Integrated Model of European Migration), applied to the system of 27 EU and four EFTA countries for the period 2002–2008. The IMEM model aims to address the data challenges mentioned before, whilst explicitly taking the uncertainty of estimation into account, unlike MIMOSA, which only produced point estimates. The modelling approach adopted in IMEM is Bayesian, which allows for incorporating expert opinion in an explicit and coherent manner.

This paper is structured as follows: after a brief description of the premises and construction of the IMEM model, the discussion focuses on the elicitation of the expert information, which is required for the assumptions on the *a priori* distributions of selected model parameters. Subsequently, selected preliminary results of the application of the model to available European data are presented. The paper concludes with a discussion of the findings and achievements so far, as well as of further steps that would be required for the model to become useful for the users of population estimates.

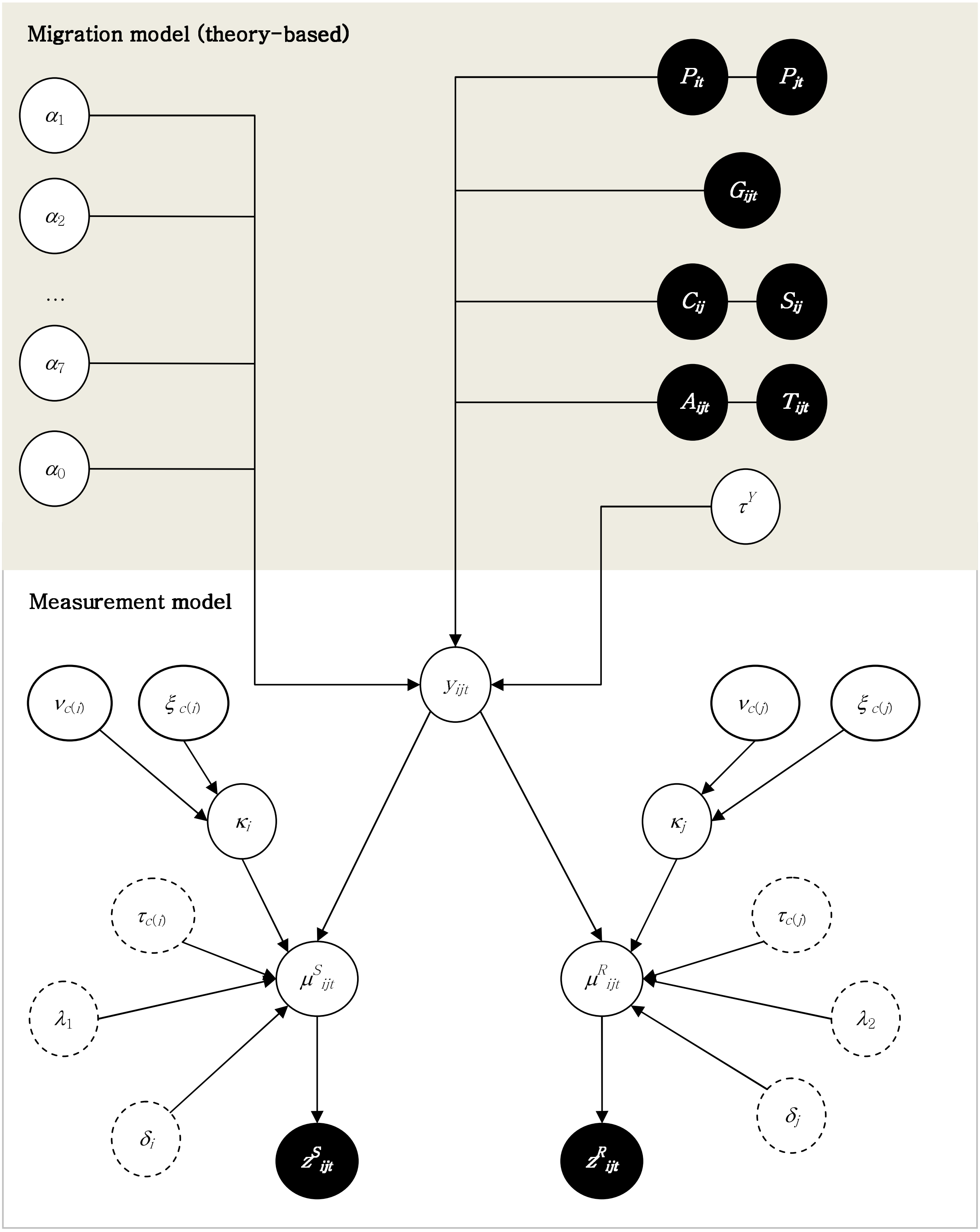
The IMEM Model: Specification

In terms of modelling, the approach to estimating migration undertaken in this study directly extends the ideas developed by Brierley et al. (2008), Abel (2010), and Raymer et al. (2011). In particular, IMEM is a hierarchical Bayesian model, which allows for combining statistical information from different countries with meta-information on definitions and data collection methods. This is further augmenter by inclusion of relevant expert judgement and some hints on possible determinants of population flows offered by migration theories. The Bayesian approach adopted in the model allows for a coherent quantification of uncertainty stemming from different sources (data discrepancies, model parameters, and expert judgement), and allows to supplement deficient data by using other sources of knowledge (e.g. Willekens 1994). The prototype of the current model has been described in more detail in the paper by Raymer et al. (2010).

A concise graphical representation of the model architecture for migration within Europe is presented in Figure 1, together with a list of variables (black nodes) and parameters (white nodes). At the highest level, the hierarchy of IMEM is comprised of two layers: the *migration model*, and the *measurement model*. The former, based on a general gravity framework and a set of quantifiable migration determinants, as suggested by Jennissen (2004) and Abel (2010), utilises insights from migration theory in order to estimate a set of 'true', harmonised migration flows, benchmarked to the United Nations (1998) definition (y_{ijt} in Figure 1).

[3] Official Journal OJ L 199, 31.07.2007, pp. 23–29; available via http://eur-lex.europa.eu.

Figure 1 Graphical representation of the IMEM model for intra-European migration



Dashed nodes denote parameters, for which the prior distributions were elicited from the experts. Hyper-parameters are not shown for greater clarity of presentation. Indices: i – sending country, j – receiving country, t – time (2002 ... 2008).

A_{ij} **Data:** z^S_{ijt} and z^R_{ijt}: migration observed in <u>S</u>ending and <u>R</u>eceiving countries; P_{it}, P_{jt}: population sizes; G_{ijt}: ratio of GNI per capita; C_{ij}: contiguity dummy; S_{ij}: migrant stocks in 2000; A_{ijt}: EU accession dummy; T_{ijt}: trade volume.

λ_1 **Parameters:** Migration model – α_1 ... α_7: parameters by migration determinants; α_0: constant; τ^Y: precision of the error term. Measurement model – y_{ijt}: 'true' migration flow; μ^S_{ijt}, μ^R_{ijt}: Poisson means; κ_i, κ_j: Normal random effects, with parameters (ν, ξ) specific to groups of countries $c(i)$; ditto $\tau_{c(i)}$: group-specific precision parameters; λ_1, λ_2: undercounts of emigration and immigration, d_i, d_j: duration-of-stay criteria applied in countries i and j.

This part of the model is also used to impute the values of the estimates when the actual data on flows are entirely or partially missing. The measurement model, in turn, distorts the 'true' flow variables by taking into account different definitions used in various countries, varying accuracy of data collection mechanisms, and the overall undercount of migration. Moreover, different coverage of data is modelled by country-specific random effects, which are assumed to be Normal. The distorted values of y_{ijt} are subsequently confronted with the observed migration flows (z_{ijt} in Figure 1), which are used to estimate the model parameters. Both migration and measurement models assume mainly log-linear relationships between the dependent and independent variables, with the measurement model additionally allowing for Poisson variability associated with the 'true', unobserved migration flows.

In addition to the model of intra-European migration, the IMEM has been also equipped with a similar module devoted to migration from and to countries outside the EU and EFTA, which is not shown in Figure 1 for the transparency of presentation. The key difference between the two parts is that 'rest of the World' model relies on single observations from European countries – no external data are used here. The migration model is this time equipped with six covariates for the European countries: population size, Gross National Income (GNI) per capita, a dummy indicating whether the country is a party to the Schengen agreement, stocks of migrants born outside the EU and EFTA, fraction of population aged over 65 years, and female life expectancy at birth. Two last-mentioned variables are proxies for the level of socio-economic development.

The IMEM model has been coded and executed in OpenBUGS – software environment specifically devoted to Bayesian computations. In terms of assumptions, all parameters in the migration model, as well as the parameters of Normal random effects in the measurement model, were assigned relatively vague (hardly informative) distributions *a priori*. In turn, for the parameters related to key features of migration measurement systems – accuracy ($\tau_{c(i)}$ and $\tau_{c(j)}$ in Figure 1), duration-of-stay criteria (δ_i and δ_j), and overall undercount of population flows (λ_1 and λ_2) – the prior distributions have been elicited from eleven experts on issues related to European migration statistics. The process and results of expert knowledge elicitation are discussed in the next section.

Elicitation of Expert Opinion

The expert opinion used to construct prior distributions for the key parameters of the measurement model comes from a two-round online Delphi survey carried out amongst eleven European experts on migration statistics. The Delphi approach, despite its known drawbacks as a standalone prediction tool (e.g. Cooke 1991: 12–17), was used here as an auxiliary method of analysis, aimed at supporting the elicitation of prior information (see Bijak and Wiśniowski 2010). In the context of IMEM, the multi-stage design helped achieve the aims of the study not so much by enforcing the convergence of experts' views, but rather through ensuring that common understanding of the underlying concepts is shared by all respondents. This allowed adjusting the formal probabilistic vocabulary used in the questionnaire to become more intuitive and to cater for a heterogeneous group of experts. In addition, the elicitation results were scrutinised during a dedicated expert workshop, where the participants – respondents and other invited migration data specialists – were able to provide feedback on the whole process and its outcomes. This was especially important, since the survey asked about such non-intuitive categories as second-level probabilities, for example uncertainty about the variability of the migration measurement.

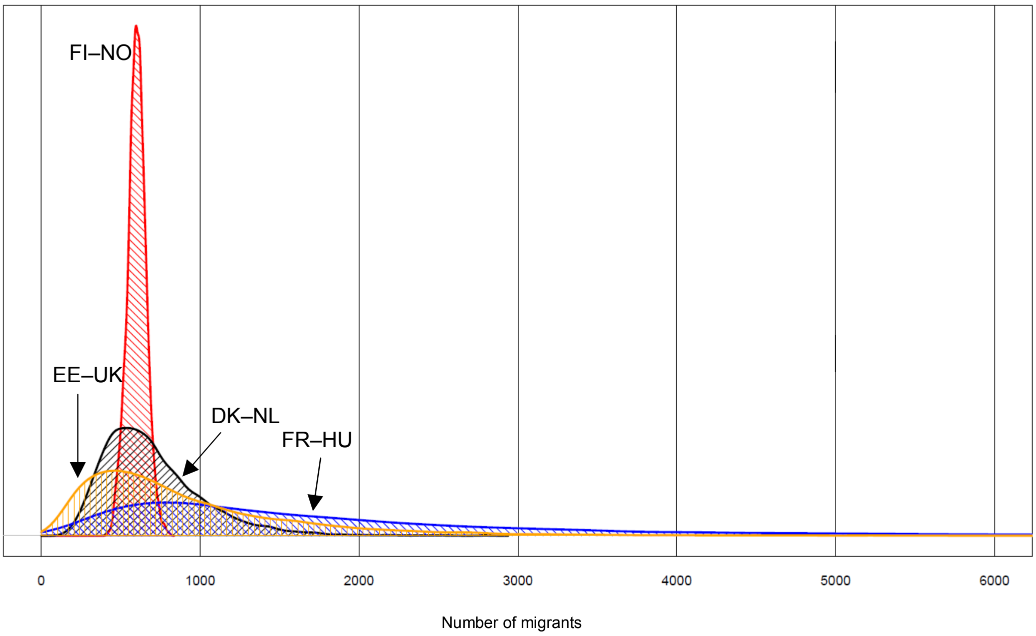
Once elicited, the answers obtained from each expert were translated into appropriate probability distributions: Beta for the undercount parameters λ_1 and λ_2, log-Normal for the duration-of-stay criteria δ_i and δ_j, and Gamma for the accuracy (precision) of migration measurement: $\tau_{c(i)}$ and $\tau_{c(j)}$. The prior distributions used in the IMEM model were ultimately obtained as mixtures of equally-weighted individual, expert-specific densities. Interestingly, there was only slight convergence in expert answers between the two rounds of the Delphi survey. Some of the resulting prior distributions – such as for parameters associated with the accuracy of measurement – were multimodal. This indicates an opposition between two groups of experts: optimists and pessimists with respect to the exactness of statistical reporting on European migration.

It is worth stressing that convergence of the expert answers was not the aim of the Delphi exercise. As mentioned before, given the multitude of problems with the quality of European migration data, the expert opinion forms key input into the IMEM model. This input naturally includes the uncertainty of expert views: for this reason we did not want to artificially suppress it, but rather reflect in the model it in a fully coherent, probabilistic manner. In this way the inevitable heterogeneity of expertise on migration statistics could be incorporated into the model and inform the overall assessment of the errors of the resulting estimates.

Tentative Results[4]

The main results of IMEM are posterior distributions of the estimates of 'true' flows, y_{ijt}, benchmarked to the United Nations (1998) definition. The distributions vary widely, depending on the characteristics of the underlying data and features of their collection systems. Four examples of distributions for 2006 are offered in Figure 2. It can be observed that where the data of both sending and receiving counties are available, and are in agreement, uncertainty is low. This is the case of flows from Finland to Norway, with migration reported by interlinked population registers. On the other hand, where both data items are unavailable (migration from France to Hungary), or one is unavailable and the other based on a less accurate source (Estonia to the United Kingdom, based on the UK International Passenger Survey), uncertainty is higher.

Figure 2 Posterior densities of estimated migration for four selected intra-European flows, 2006



DK–NL: Migration from Denmark to the Netherlands
FI–NO: Migration from Finland to Norway
FR–HU: Migration from France to Hungary
EE–UK:Migration from Estonia to the United Kingdom

Table 1 presents the posterior mean estimates of intra-European migration flows yielded by the IMEM model, averaged over 2002–2008. Thus, in this period, about 1.8 million people migrated every year within the EU-EFTA system. Noteworthy, given that this aggregate includes all the errors of estimation of origin-and-destination-specific flows, it is very uncertain, with 50 per cent credible intervals (CI) ranging from 1.02 to 2.12 million. At the country level, the biggest recipients of migration were Germany (on average, 304,000 migrants annually; 50% CI: 189,000–347,000), France (216,000; 50% CI: 97,000–251,000) and the United Kingdom (207,000; 50% CI: 99,000–242,000), while the most important sending countries were Germany (299,000; 50% CI: 169,000–346,000), Poland (185,000; 50% CI: 109,000–211,000) and the UK (175,000; 50% CI: 91,000–202,000). The single most numerous flow – of 87,000 migrants (50% CI: 55,000–99,000) – was the one from Poland to Germany, retaining a key role in the European migration system despite the EU enlargement and the diversion of Polish flows to the British Isles (Grabowska-Lusińska and Okólski 2009).

[4] The numerical results shown in this section are preliminary. Please, do not cite without the permission of the authors.

Table 1 Mean estimates of intra-European migration flows produced by the IMEM model: averages for 2002–2008

From \ To	AT	BE	BG	CH	CY	CZ	DE	DK	EE	ES	FI	FR	GR	HU	IE	IS	IT	LI	LT	LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK	UK	**Total**
AT	-	626	316	3 015	46	1 378	13 261	346	52	1 217	215	2 120	696	2 392	244	44	2 358	130	75	73	63	18	805	182	2 936	335	1 343	637	474	1 525	2 831	**39 755**
BE	435	-	120	1 970	74	362	5 561	582	78	4 894	272	18 830	1 490	544	538	59	3 163	9	84	1 824	58	26	6 689	285	1 730	1 199	360	675	78	220	6 470	**58 677**
BG	1 355	626	-	377	269	1 308	7 168	185	34	10 827	101	1 498	3 067	272	138	18	3 840	2	46	18	94	13	951	169	544	565	1 161	405	83	423	1 763	**37 320**
CH	1 475	1 244	80	-	66	345	9 428	558	58	5 850	312	11 391	1 063	497	406	57	7 956	235	47	119	41	17	1 017	277	808	1 402	306	739	217	311	4 525	**50 848**
CY	33	60	68	47	-	34	364	25	8	49	28	185	925	54	53	2	88	0	9	4	8	3	63	20	112	18	26	84	5	18	2 230	**4 620**
CZ	2 058	696	492	925	87	-	7 497	306	49	898	127	2 380	613	650	322	44	1 346	3	77	43	89	12	720	166	1 815	154	900	384	73	10 815	2 731	**36 469**
DE	19 443	10 476	2 030	21 914	400	5 763	-	4 781	572	22 689	1 777	34 353	14 767	9 351	2 848	411	23 821	158	1 484	1 854	1 088	158	12 701	2 926	47 581	7 001	6 770	5 644	1 255	3 585	30 970	**298 573**
DK	288	713	65	786	35	186	3 390	-	82	1 630	489	2 051	363	207	352	1 329	876	8	254	95	269	21	698	2 907	958	275	156	6 955	23	126	4 307	**29 890**
EE	61	180	16	89	8	35	784	181	-	171	2 391	317	40	71	104	21	183	1	105	8	470	2	139	209	100	70	45	650	4	17	697	**7 168**
ES	789	3 580	601	4 133	52	257	12 531	1 146	73	-	637	18 547	497	293	1 220	62	4 868	37	344	189	73	31	3 326	714	1 857	4 989	1 504	1 431	38	222	11 496	**75 536**
FI	326	636	38	697	53	145	2 423	419	814	1 343	-	1 212	279	252	313	80	658	2	88	60	186	12	537	782	258	162	77	4 154	11	47	2 502	**18 565**
FR	1 463	20 057	374	12 601	248	1 539	21 390	1 723	251	21 331	642	-	3 258	1 823	2 277	188	12 424	37	233	2 644	191	112	4 283	879	7 168	8 030	1 570	1 974	186	999	25 077	**154 973**
GR	521	1 420	531	857	1 062	283	9 709	287	55	810	156	2 731	-	443	191	16	2 055	13	36	46	29	29	1 197	138	1 286	161	1 174	965	17	108	4 875	**31 199**
HU	3 765	1 093	80	1 223	72	523	12 729	355	61	1 140	251	2 977	548	-	241	35	1 750	3	29	39	40	11	960	203	607	211	2 135	868	83	2 359	3 121	**37 512**
IE	228	1 364	76	950	75	160	2 876	338	75	2 269	201	2 629	359	128	-	40	1 239	4	358	62	150	23	789	140	1 798	376	272	477	13	68	21 193	**38 732**
IS	43	66	7	58	4	24	270	1 645	6	178	66	192	32	9	22	-	59	1	21	11	14	4	98	365	167	31	16	623	3	10	313	**4 360**
IT	2 917	9 918	679	19 571	127	881	27 291	1 128	204	12 576	532	37 084	2 570	1 216	1 418	116	-	115	186	611	146	293	2 786	464	4 689	1 730	4 399	1 304	600	720	19 283	**155 554**
LI	53	6	2	169	1	3	76	9	1	29	2	70	8	7	2	1	30	-	1	1	1	1	4	3	9	6	4	3	2	5	17	**527**
LT	162	265	42	129	18	133	2 831	659	197	1 835	138	566	71	42	610	63	570	1	-	13	985	4	373	635	1 006	148	58	654	9	33	2 251	**14 503**
LU	105	1 772	12	306	5	24	2 150	165	11	246	64	2 572	92	42	69	32	387	2	16	-	8	4	226	28	112	457	36	137	15	13	479	**9 587**
LV	98	142	18	103	22	43	1 512	337	388	290	159	386	30	31	400	29	298	1	535	9	-	3	176	218	192	64	33	406	5	24	1 059	**7 010**
MT	19	42	8	36	8	10	173	22	6	40	10	239	32	11	34	2	214	0	3	2	2	-	50	11	14	15	22	37	2	8	1 021	**2 092**
NL	1 161	15 759	194	2 765	116	736	15 339	1 076	88	6 834	476	8 251	1 418	757	1 163	114	2 962	10	117	280	72	54	-	1 024	2 855	1 933	507	1 674	82	275	13 016	**81 105**
NO	155	411	40	350	30	94	1 661	3 141	61	1 909	1 072	1 223	168	114	196	420	429	3	156	24	72	9	603	-	762	223	104	6 664	7	116	3 262	**23 482**
PL	5 607	3 530	192	1 845	189	3 975	86 960	2 661	114	7 858	427	10 826	2 034	858	2 063	649	11 938	8	435	188	253	30	6 471	3 550	-	434	384	5 337	51	1 492	24 418	**184 775**
PT	369	1 671	99	3 913	26	78	7 346	228	73	13 273	98	18 647	216	86	228	56	1 297	15	47	1 287	30	27	1 675	167	212	-	316	331	11	53	6 627	**58 502**
RO	5 410	1 201	115	1 498	283	1 362	21 437	411	32	25 793	177	5 764	2 128	9 582	693	38	39 254	4	21	85	79	13	1 199	334	390	1 157	-	884	53	1 790	2 537	**123 722**
SE	636	1 123	80	1 226	125	240	3 670	3 554	154	2 701	4 636	2 806	1 148	399	529	566	1 211	5	173	121	276	34	980	5 163	1 398	365	217	-	64	123	6 351	**40 076**
SI	1 022	109	30	564	10	102	1 753	49	10	194	21	833	64	178	47	5	979	11	10	25	12	4	104	28	82	37	54	118	-	99	273	**6 827**
SK	2 802	193	55	610	30	17 102	5 687	159	21	666	44	699	110	1 091	122	10	1 110	2	19	22	40	5	425	174	448	53	226	201	61	-	1 344	**33 531**
UK	1 773	6 829	488	5 260	1 877	1 595	16 429	3 426	272	38 339	1 426	24 134	4 342	1 529	18 401	267	10 168	18	1 312	286	701	567	7 472	2 248	13 254	5 635	1 209	4 684	161	747	-	**174 852**
Total	**54 573**	**85 808**	**6 945**	**87 988**	**5 417**	**38 722**	**303 696**	**29 900**	**3 901**	**187 876**	**16 945**	**215 514**	**42 429**	**32 928**	**35 245**	**4 771**	**137 531**	**840**	**6 320**	**10 042**	**5 541**	**1 539**	**57 517**	**24 410**	**95 150**	**37 236**	**25 385**	**49 098**	**3 687**	**26 349**	**207 038**	**1 840 342**

AT: Austria, BE: Belgium, BG: Bulgaria, CH: Switzerland, CY: Cyprus, CZ: Czech Republic, DE: Germany, DK: Denmark, EE: Estonia, ES: Spain, FI: Finland, FR: France, GR: Greece, HU: Hungary, IE: Ireland, IS: Iceland, IT: Italy, LI: Liechtenstein, LT: Lithuania, LU: Luxembourg, LV: Latvia, MT: Malta, NL: The Netherlands, NO: Norway, PL: Poland, PT: Portugal, RO: Romania, SE: Sweden, SI: Slovenia, SK: Slovakia, UK: United Kingdom

Discussion and Conclusions

The results presented before suggest that, so far, IMEM has succeeded in producing a coherent set of plausible, harmonised probabilistic estimates for intra-European migration, as well as migration to and from 31 European countries (not shown in this paper). The next steps of the modelling will involve an extension of the analysis to include age and sex. In this way, we are hoping that IMEM will be able to solve the problem of disaggregation of migration data by the main demographic characteristics, besides the countries of origin and destination. So far, harmonisation issues aside, in many countries this information is available solely from sample-based enquiries, such as Labour Force Surveys carried out across Europe, or International Passenger Survey in the United Kingdom. In such cases, the sizes of the subsamples of migrants are usually far too small to allow for detailed disaggregation by origin or destination of migrants, age, and sex.

Although the focus of this paper is mainly conceptual, the main contributions of IMEM are both conceptual and practical. The results are based on whole posterior distributions, and thus any point estimate (e.g. mean or median) can be equipped with the assessment of uncertainty, which at times can be quite wide. This is a direct consequence of the current state of the European data collection systems related to international migration. There are many efforts to harmonise migration statistics at the EU level – Regulation (EC) No. 862/2007 being one of them – but so far the discrepancies in the reported figures are so large, and the data collection mechanisms so prone to bias, that this inevitably becomes reflected in the final estimates. The users of statistics can hope that the concerted effort of European agencies and particular Member States will allow for reducing the uncertainty once harmonisation measures are robustly in place. However, for now, the migration reality is uncertain, which is exactly one of the key messages conveyed by the IMEM results.

A related issue concerns, how to communicate the results of statistical models such as IMEM to the final users of migration and population estimates. Given that interval estimates provide more information than point estimates, measures of central tendencies, such as means or medians, can be reported together with credible intervals, as in the examples presented in the previous section. The additional aim of doing so is to increase the uncertainty awareness of the users. An open question is: what probability should be covered by the reported intervals. Lawrence et al. (2006) noted that overconfidence on the part of users can lead to more extreme policy actions. On the other hand, intervals covering too small probability are largely useless for practical purposes. This constitutes an argument for presenting credible intervals based on 'medium' probabilities (e.g., 50 per cent, as in the examples presented before), in order to avoid the 'illusion of control' amongst the decision makers, and to suggest additional caution. Paraphrasing the caveats of Lawrence et al. (2006) made with respect to forecasting: the ability to minimise the uncertainty assessment should not become a criterion of evaluating the accuracy of the estimation process and of the resulting estimates.

From a statistical point of view, the outcomes produced by the model – whole posterior distributions of the estimated y_{ijt} – can be used for assessing migration at the European level, additionally taking into account relative costs of overestimating or underestimating of flows. Applying a Bayesian decision analysis in this context, however, is not trivial: given that for every year, the output consists of a two-dimensional matrix $\mathbf{Y} = [y_{ij\cdot}]_{31\text{x}31}$, unique solutions to decision problems concerning the system as a whole do not exist. Partial solutions include applying the decision analysis to conditional or marginal distributions of particular flows or to their aggregates. However, more research into possible applications of methods of multi-criteria decision analysis will be needed in order to take full advantage of the possibilities offered by the results of the model.

In summary, statistical modelling of the whole European migration system, as demonstrated by IMEM, offers the users a set of harmonised estimates, with an assessment of their uncertainty – inevitable given the imperfections of the mechanisms of data collection and measurement of population flows. By producing whole distributions rather than the point estimates, which used to be the standard in previous attempts to harmonise migration data (for example in the MIMOSA study), IMEM offers the users more information. The question on how to make the best use of all the insights offered by probabilistic models, however, remains open. To answer it, a proper dialogue between the statistical modelling community and the users of population and migration estimates needs to be established, if such outcomes are to become of practical use.

REFERENCES

Abel GJ (2010) Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society, Series A* 173(4):797-825.

Bijak J and Wiśniowski A (2010) Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society, Series A* 173(4):775-796.

Bilsborrow RE, Graeme H, Amarjit SO, and Zlotnik H (1997) *International migration statistics: Guidelines for improving data collection systems*. Geneva: International Labour Office.

Brierley MJ, Forster JJ, McDonald JW and Smith PWF (2008) Bayesian estimation of migration flows. In J Raymer and F Willekens (eds.), *International migration in Europe: Data, models and estimates*, pp. 149-174. Chichester: Wiley.

Cooke RM (1991) *Experts in Uncertainty. Opinion and Subjective Probability in Science*. Oxford: OUP.

Grabowska-Lusińska I, Okólski M (2009) *Emigracja ostatnia?* [*Last emigration?*] Warsaw: Scholar.

Jennissen R (2004) *Macro-economic determinants of international migration in Europe*. Amsterdam: Dutch Univ. Press.

Kelly JJ (1987) Improving the comparability of international migration statistics: Contributions by the Conference of European Statisticians from 1971 to date. *International Migration Review* 21:1017-1037.

Kupiszewska D and Nowok B (2008) Comparability of statistics on international migration flows in the European Union. In J Raymer and F Willekens (eds.), *Op. cit.*, pp. 41-71. Chichester: Wiley

Lawrence M, Goodwin P, O'Connor M and Önkal D (2006) Judgemental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting* 22(3): 493–518.

Nowok B (2010) *Harmonization by simulation: A contribution to comparable international migration statistics in Europe*. Amsterdam: Rozenberg Publishers.

Poulain M, Perrin N and Singleton A, eds. (2006) *THESIM: Towards Harmonised European Statistics on International Migration*. Louvain: Presse Universitaire de Louvain.

Raymer J, de Beer J, van der Erf R (2011) Putting the pieces of the puzzle together: Age and sex-specific estimates of migration amongst countries in the EU/EFTA, 2002-2007. *European Journal of Population* 27(2):185-215.

Raymer J, Forster JJ, Smith PWF, Bijak J, Wiśniowski A and Abel G (2010) The IMEM model for estimating international migration flows in the European Union. Joint UNECE/Eurostat Work Session on Migration Statistics, Geneva, 14-16 April 2010, Working Paper 14.

United Nations (1978) *Statistics of international migration. Demographic Yearbook 1977*. New York: United Nations.

United Nations (1998) *Recommendations on statistics of international migration*. New York: United Nations.

Willekens F (1994) Monitoring international migration flows in Europe. Towards a statistical data base combining data from different sources. *European Journal of Population* 10(1):1-42.

ABSTRACT

The paper deals with uncertainty in estimating international migration flows for an interlinked system of countries. The related problems are discussed on the example of a dedicated model 'IMEM' (Integrated Model of European Migration). The IMEM is a hierarchical Bayesian model, which allows for combining data from different countries with meta-data on definitions and collection methods, as well as with relevant expert information. The model is applied to 31 EU and EFTA countries for the period 2002–2008. The expert opinion comes from a two-round Delphi survey carried out amongst 11 European experts on issues related to migration statistics. The adopted Bayesian approach allows for a coherent quantification of uncertainty stemming from different sources (data discrepancies, model parameters, and expert judgement). The outcomes produced by the model – whole posterior distributions of estimated flows – can be then used for assessing the true magnitude of flows at the European level, taking into account relative costs of overestimating or underestimating of migration flows. In this context, problems related to application of the decision statistical analysis to multidimensional problems are briefly discussed.

Keywords: European migration, Migration estimates, Bayesian methods, Uncertainty