

1

Introduction and outline

James Raymer[†] and Frans Willekens[‡]

[†] *Division of Social Statistics, School of Social Sciences, University of Southampton*

[‡] *Netherlands Interdisciplinary Demographic Institute, The Hague*

1.1 Introduction

In response to a long history of inadequate, incomparable and missing migration data, a new regulation has been prepared by the European Parliament and the Council on Community Statistics that obliges Member States to produce and provide harmonised estimates of international migration and asylum flows (European Commission 2005). The aim of the new legislation is to increase the availability and comparability of migration statistics. Member States should use the best data sources available to supply Eurostat, the statistical office of the European Union, with detailed metadata. In addition, these data should be explained and assessed in how close they comply with the harmonised definitions. Harmonisation should reduce the impact of different migration definitions and sources on the published migration statistics. However, the guidelines and methodologies for providing harmonised estimates have yet to be developed or implemented.

A major research effort in support of the new regulation was the THESIM (*Towards Harmonised European Statistics on International Migration*) project, funded by the European Commission's Directorate-General for Research under the

Sixth Framework Programme (Poulain *et al.* 2006). One of the objectives of the project was to analyse the current functioning of migration statistics in the 25 European Union (EU) countries and to identify the problems that the countries might encounter in relation to the requirements of the new regulation. During the project, researchers met with experts from the national statistical institutes (NSIs) or national statistical offices (NSOs) and with the authorities responsible for the registration of migrants in EU countries. The meetings were a unique source of information, with explanations provided directly by a range of people involved in the statistics production process. The project pointed to the need for metadata that adequately document rules and procedures used during the processes of data collection and the production of statistics.

The work contained in this book builds on the THESIM study of the migration statistics production process and the resulting data situations. It can be considered an important starting point for the development of methodologies to convert raw data on migration into internationally comparable statistics. Issues, strategies and methods for obtaining consistent estimates of international migration flows are presented in the context of various data situations. Estimates of migration are needed so that national governments and international agencies have the means to improve their policies directed at supplying particular social services or at continuing, increasing or decreasing levels of international migration. Furthermore, our understanding of how or why populations change requires relatively detailed information about migration. Without these, the ability to predict, control or understand that change is limited.

Although all countries and regions of the world face similar challenges measuring international migration, this work focuses on European migration for practical, substantive and policy reasons. First, this work is produced within the context of the European Research Training Network (RTN) programme 'Demographic Sustainability and European Integration' funded by the European Commission (contract HPRN-CT-2001-00234). One of the objectives of the RTN programme was the development of a reliable database on population movements in Europe with internationally comparable data using current work at Eurostat and building on the European Monitoring System, which was initiated in 1992 by the Population Research Centre of the University of Groningen in the Netherlands and the International Institute for Applied Systems Analysis in Laxenburg, Austria.¹

Second, international migration is the major factor contributing to population change in regions of the world and Europe in particular. In 2005, the United Nations estimated that worldwide there were about 191 million persons residing outside their country of birth, or nearly 3% of the world's population (UNFPA 2006; IOM 2005). Excluding the former Soviet Union, Europe hosts about 19% of all foreign-born persons in the world,² which is striking considering that the population of this area only represents 7–8% of the world's population total. Since 1989, net immigration has contributed to at least 50% of the European Union's annual

¹ For details the reader is referred to the RTN proposal and Willekens (1994).

² With the countries that were part of the former Soviet Union included, it is around 36%.

population growth and, in 2002, it represented 85 % compared to 27 % in 1980 (Eurostat 2004). The underlying flows driving this demographic change, however, are largely unknown. The empirical evidence required to permit a truly integrated analysis of migration flows within Europe and between Europe and the rest of the world is not available.

Third, demographic change in Europe has never received the attention it is receiving today. Migration is at the core of the debate. Although there is a great demand for evidence-based policies, the evidence is limited because data are defective. The reason for this is primarily due to differences in data sources, availability, quality and measurement. The advantages to having a consistent and reliable set of migration flows are numerous.

Europe has always been affected by migration. Demographic growth, climatic change, trade, warfare, conquest and formation of nations, states and empires have all initiated population movement. However, since World War II, the levels and sources of migrants have changed considerably (Castles and Miller 2003). Foreign-born populations are growing rapidly as a result of very low or negative levels of natural increase among native-born populations and by increasing levels of immigration. The diversity of foreign-born populations is also expanding as European countries are becoming increasingly more connected with the world.

The reasons for migration are many. People move to other countries for employment, family reunion, protection or amenity reasons. Statistics on these flows, on the other hand, are relatively confusing or nonexistent. There are two reasons. First, no consensus exists on what exactly is a 'migration'. The theoretical concept of migration as a change of usual residence across administrative (national) boundaries may be generally accepted but implementation of this concept has yet to take place. Comparative analyses suffer from differing national views concerning who is a migrant. Second, the event of migration is rarely measured directly. More often it is inferred by a comparison of places of residence at two points in time that may be three months, six months, one year or several years apart. The challenge is compounded because individual countries use different methods of data collection and varying concepts of migration. International migration statistics may come from administrative data, decennial population censuses or surveys. Migrants may be defined according to previous or next country of residence, nationality or citizenship. Harmonisation of data collection processes and the data they generate is not even close to being realised. So, how does one overcome these obstacles to obtain an overall and consistent picture of the migration patterns occurring within Europe?

Solutions to deal with the above problems are set out in this book. First, we develop a classification of migration data types. Second, we develop models that are able to combine different data types to predict migration flows that are comparable across data types. Migrant stocks and migration flows measure migration differently and are examples of two data types. Additionally, differences in data sources, reliability and validity are accounted for in the models. The key aspect of the models is that they estimate migration flows using available information from one or more sources in combination with realistic assumptions about the relation between various

data types and structures of the missing data. The outcomes are *synthetic data* that represent the best possible estimates of international migration during a given period. The approaches used to construct these synthetic data are rooted in statistical theory and were originally developed in studies of internal migration (Plane 1981, 1982; Rogers and Castro 1981; Rogers and von Rabenau 1971; Rogers *et al.* 2002, 2003; Sweeney 1999; Willekens 1977, 1980, 1982, 1983, 1999; Willekens *et al.* 1981). It was suggested that, to obtain estimates of international migration, one has to focus on data harmonisation and models, but such efforts have not been attempted until recently (Poulain 1994, 1995, 1999, 2001; Poulain *et al.* 2006; Willekens 1994; Raymer 2007).

This is the first volume that draws together modern statistical approaches to incomplete data and migration modelling. Issues dealing with data on migration are linked directly to models of migration in order to produce the best possible estimates of international migration in Europe. This link is essential both for the estimation of missing data and for the harmonisation of available data. International migration in Europe provides a good starting point for learning how to estimate international migration flows because it plays such a large role in its population change and the data are readily available for many of the countries in this region, albeit in various forms and in varying degrees of completeness.

The modelling approaches are rooted in statistics with an interdisciplinary audience in mind. For example, statistical methods are used to harmonise and estimate missing or incomplete data, demographic techniques are used to identify structures in the spatial and age patterns of migration, geographic methods are applied to model movement across space, and economic ideas about imbalances in the demand and supply of labour are used to produce more accurate estimates. The approach relies heavily on empirical regularities, which can be imposed in regression-type models. In the process of estimating migration flows, the population may be stratified by subpopulations that differ in migration behaviour, which can be incorporated in the modelling process. For the examples used in this book, the migrant subpopulations consist of various ethnicities, nativities and migrant types (i.e. labour, family-related, asylum seekers and so on).

This book is written for those carrying out research on topics that are directly or closely connected with migration patterns and processes. The level is written for upper undergraduate students and graduate-level students in the fields of demography, statistics, geography, regional science, sociology, economics and political science. We also hope that statisticians at national and international organisations will find the ideas and solutions presented in this volume useful.

1.2 Outline

1.2.1 Definitions

Migration is a complicated subject. In order to make sense of the various types of flows and definitions, we provide the following key definitions. First, a distinction between ‘migration’ and ‘migrant’ is important. Migration refers to the *event* of

moving from one country to another. A migrant is a *person* who has changed his or her residential status, from one time to another. Population registers capture movements (i.e. migration events). Censuses and surveys capture migrants (i.e. status transitions).

There is also an important difference between migrant populations and migration flows. The foreign-born population contains persons who were born outside the country of current residence from the perspective of the settlement country. Expatriates are foreign-born persons residing in another country from the perspective of the origin country. Native-born persons are those born in the country of current residence. Foreigners are those who have not acquired citizenship in the country of residence but are residing there. Nationals may include both native-born and foreign-born persons.

Immigration refers to the flow of migration into a particular country (receiving country or destination country) and emigration refers to the flow of migration from a particular country (sending country or country of origin). Both immigration and emigration are associated with a particular time period. In this book, we find immigration flows that use anywhere from no specific time criteria (Germany) to permanent (Poland) as the time period for immigration or emigration flows. Mostly, we think of immigration or emigration flows being a one-year flow, in alignment with the United Nations recommended duration. These flows may contain native-born, foreign-born, foreigners and nationals. Finally, immigration and emigration data may be provided by the country of origin or the country of destination.

1.2.2 Frequentist and Bayesian approaches

Both frequentist and Bayesian models of migration are presented in this book. Bayesian methods remain relatively undeveloped for estimating migration patterns, as they are relatively complicated and intensive to implement. However, Bayesian methods are ideally suited to inadequate and missing data problems because they allow a direct combination of prior information or beliefs and empirical observation. This means that one can build a flexible model that incorporates the strengths of theory and empirical data, which is important for situations such as combining data from multiple sources, missing data and inconsistent data. The outcomes of Bayesian analysis include distributions of both parameters and predicted values, meaning that the user has much better measures of the uncertainties. This is very important when the data of interest are inadequate or missing.

1.2.3 Structure of the book

This book contains five parts aside from the ‘Introduction and outline’ and ‘Conclusion’. Part I focuses on migration data and issues and includes four chapters – each addressing a particular migration data type. In Chapter 2, Jean-Christophe Dumont and Georges Lemaître focus on foreign-born and expatriate population stocks in OECD countries. They point out that these migration data are the most comparable and the most readily available, and therefore should be a starting point for

comparative analyses of migration. Dorota Kupiszewska and Beata Nowok follow this in Chapter 3 with a review of international migration flow statistics produced by the 25 countries in the European Union (situation in 2006) and illustrate the many inconsistencies in the flow data caused by differences in concepts, measurements and the processing of migration data. Beata Nowok then extends this analysis in Chapter 4 with a detailed analysis of migration data in Central Europe over time. In Chapter 5, Alessio Cangiano focuses on Southern European countries and the evolution of their foreign populations. In particular, undocumented migrants are approximated using the information from regularisation programmes and other data sources.

The emphasis of Part II is on models for spatial and age patterns of migration. In Chapter 6, Frans Willekens provides an overview of migration models with complete and incomplete data. Matthew Brierley, Jonathan Forster, John McDonald and Peter Smith present a Bayesian approach in Chapter 7 to estimate international flows in Northern Europe in the context of data both missing and coming from different sources. In Chapter 8, James Raymer and Andrei Rogers review model migration schedules and provide a demonstration of how families of these schedules can be used to estimate age patterns of migration in the context of incomplete or missing data. Peter Congdon models age patterns of migration using Bayesian methods in Chapter 9, using migration between England and Scotland as the illustration.

In Part III, the above models and others are applied in two chapters to obtain current estimates of international migration flows in Europe and to estimate the impact of various asylum-seeker policies on other potential receiving countries. In Chapter 10, James Raymer applies the ideas set out in Chapters 6 and 8 in Part II to estimate age-specific international migration flows between countries in the European Union. In Chapter 11, Leo van Wissen and Roel Jennissen present a methodology for identifying potential redirection impacts of asylum seekers caused by restrictions imposed by particular countries of destination. A distinction is made between overall pattern (generation effect) and migration flows that result from redirection (substitution effect).

An overview of methods for forecasting international migration flows from traditional time series and Bayesian perspectives is provided in Part IV. First, in Chapter 12, Jakub Bijak provides a review of previous attempts to forecast migration before arguing for a Bayesian approach. For illustration, he forecasts migration flows from Poland to Germany. Joop de Beer then reviews extrapolation and time series models and forecasts different types of immigration and emigration flows in the Netherlands in Chapter 13.

Part V contains two chapters on population projections that emphasise the migration component. A multiregional model is presented in Chapter 14 by Marek Kupiszewski and Dorota Kupiszewska to estimate subnational and national populations in Central Europe. Then Philip Rees puts forward a projection model in Chapter 15 to estimate future regional ethnic populations in the United Kingdom.

The contributions of the book are many. First, a distinction is made between the different migration data types. This removes the need for *ad hoc* approaches to estimating flows. Second, the conventional methods (i.e. frequentist) for estimating

incomplete migration data are complemented with a Bayesian perspective. The Bayesian perspective is better able to incorporate the additional information on migration and therefore to determine how better or more information results in better estimates of migration. Finally, this book demonstrates how these estimates can be incorporated in the projection of migration flows and the projection of populations.

References

- Castles S and Miller MJ. 2003. *The age of migration: International population movements in the modern world*. New York: The Guilford Press.
- European Commission. 2005. *Proposal for a Regulation of the Parliament and of the Council on Community statistics on migration and international protection*. Brussels. COM (2005) 375 final, 2005/0156 (COD). Available at: http://eurlex.europa.eu/LexUriServ/site/en/com/2005/com2005_0375en01.pdf
- Eurostat. 2004. *Population statistics, 2004 edn*. Luxembourg: Office for Official Publications of the European Communities.
- IOM. 2005. *World migration: Costs and benefits of international migration*. Geneva: International Organisation for Migration.
- Plane DA. 1981. Estimation of place-to-place migration flows from net migration totals: a minimum information approach. *International Regional Science Review* 6(1):33–51.
- Plane DA. 1982. An information theoretic approach to the estimation of migration flows. *Journal of Regional Science* 22(4):441–456.
- Poulain M. 1994. *Internal mobility in Europe: the available statistical data*. Working Paper 17. Conference of European Statisticians, Commission of the European Communities (Eurostat), Mondorf-les-Bains, Luxembourg.
- Poulain M. 1995. Towards a harmonisation of migration statistics within the European Community. In *Demographic consequences of international migration*. Voets S, Schoorl J and de Bruijn B, eds, pp. 11–25. The Hague: Netherlands Interdisciplinary Demographic Institute.
- Poulain M. 1999. *International migration within Europe: towards more complete and reliable data?* Working Paper 37. Joint ECE–Eurostat Work Session on Demographic Projections, Perugia, Italy.
- Poulain M. 2001. *Is the measurement of international migration flows improving in Europe?* Working Paper 12. Joint ECE–Eurostat Work Session on Migration Statistics, Geneva, Switzerland.
- Poulain M, Perrin N and Singleton A, eds. 2006. *THESIM. Towards Harmonised European Statistics on International Migration*. Louvain-la-Neuve: UCL Presses Universitaires de Louvain.
- Raymer J. 2007. The estimation of international migration flows: a general technique focused on the origin–destination association structure. *Environment and Planning A* 39:985–995.
- Rogers A and Castro LJ. 1981. *Model migration schedules*. RR-81-30. Laxenburg: International Institute for Applied Systems Analysis.
- Rogers A and von Rabenau B. 1971. Estimation of interregional migration streams from place-of-birth-by-residence data. *Demography* 8: 185–194.
- Rogers A, Willekens FJ and Raymer J 2002. Capturing the age and spatial structures of migration. *Environment and Planning A* 34:341–359.

- Rogers A, Willekens FJ and Raymer J. 2003. Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer* 55(1):56–69.
- Sweeney SH. 1999. Model-based incomplete data analyses with an application to occupational mobility and migration accounts. *Mathematical Population Studies* 7(3):279–305.
- Willekens FJ. 1977. *The recovery of detailed migration patterns from aggregate data: An entropy maximizing approach*. Laxenburg: International Institute for Applied Systems Analysis.
- Willekens FJ. 1980. Entropy, multiproportional adjustment and the analysis of contingency tables. *Systemi Urbani* 2:171–201.
- Willekens FJ. 1982. Multidimensional population analysis with incomplete data. In *Multidimensional mathematical demography*, Land KC and Rogers A, eds, pp. 43–112. New York: Academic Press. Reprinted in *Readings in population research methodology*, Bogue DJ, Arriaga EE and Anderton DL, eds, pp. 22.49–22.82. Chicago: Social Development Center; and New York: UNFPA.
- Willekens FJ. 1983. Log-linear modelling of spatial interaction. *Chapters of the Regional Science Association* 52:187–205.
- Willekens FJ. 1994. Monitoring international migration in Europe: towards a statistical data base combining data from different sources. *European Journal of Population* 10(1):1–42.
- Willekens FJ. 1999. Modeling approaches to the indirect estimation of migration flows: from entropy to EM. *Mathematical Population Studies* 7(3):239–278.
- Willekens FJ, Pór A and Raquillet R. 1981. Entropy, multiproportional, and quadratic techniques for inferring patterns of migration from aggregate data. In *Advances in multiregional demography*. RR-81–6. Rogers A, ed., pp. 83–124. Laxenburg: International Institute for Applied Systems Analysis.
- UNFPA. 2006. *State of the world population*. New York: United Nations Population Fund. Available at: http://www.unfpa.org/swp/2006/pdf/en_sowp06.pdf