UN/POP/MIG-9CM/2011/09
12 February 2011

NINTH COORDINATION MEETING ON INTERNATIONAL MIGRATION
Population Division
Department of Economic and Social Affairs
United Nations Secretariat
New York, 17-18 February 2011

INTERNATIONAL MIGRATION FLOWS: STATISTICAL MODELS FOR USE IN POPULATION PROJECTIONS[1]

Laboratory of Populations
The Rockefeller University & Columbia University
New York, New York[2]

[1] The views expressed in the paper do not imply the expression of any opinion on the part of the United Nations Secretariat.

[2] Joel E. Cohen, Ph.D., Dr.P.H., Abby Rockefeller Mauzé Professor, Laboratory of Populations, The Rockefeller University & Columbia University, New York, New York.

.

INTRODUCTION

The Laboratory of Populations at The Rockefeller University and Columbia University, New York City, is an academic research unit devoted to understanding human and non-human populations. This document summarizes the results of recent research on human international migration from the Laboratory of Populationsand its collaborators, and indicates plans for future research. Collaborators in these studies have been colleagues at the United Nations Population Division (Marta Roig) and the Department of Sociology at the University of Wisconsin, Madison (Keuntae Kim) as well asformer members of the Laboratory of Populations (Daniel C. Reuman,CaiGoGwilt).

A. BACKGROUND

International migration will play an increasing role in the demographic future of nations if fertility continues to decline in most countries. Net immigration already accounts for roughly 40% of population growth in the United States of Americaand about 90% in the EU-15 countries (Howe and Jackson, 2006).

In projecting international migration, the United Nations Population Division (2003, paragraphs 57-59) identified the need for a demographically plausible, programmable algorithm that automatically projects a zero world balance of net migration and prevents projected net emigration from completely depleting the population of any sending country. To meet this need, the Laboratory of Populations and collaborators proposedmodels and empirically based equations for projecting future numbers of international migrants from any country or region to any other. These algorithms are comparable in transparency and generality to standard cohort-component methods of projecting births and deaths.

Most theories of international migration draw on social, economic and/or political factors to explain migration (Bijak, 2008; Massey et al., 1993, 1998; Faist, 2000), such as differences among countries in gross domestic product, labor markets, migration policies, social networks of prior migrants, and cognitive and behavioral attributes of individuals (Howe and Jackson, 2006;Ritchey, 1976; Dorigo and Tobler, 1983). Although states and governments influence migration via their laws and regulations, and some past empirical studies attempted to incorporate some form of policy measures, data on this subject are sparse, and predictive models of policy are not available.For multi-decadal demographic projections, it seems more difficult to project such non-demographic variables than it is to project demographic variables such as fertility and mortality.

The intellectual antecedents of the new proposed models include Zipf's (1946, 1949) model of inter-city migration, which is one of several "gravity" models in the social sciences (Rogers, 2008).The proposed models assume the availability only of constant geographic or historical variables and of population sizes which can be projected incrementally in time by accepted demographic procedures.

The models make possible deterministic and stochastic projections of migration and hence of population.The approach presented here is different from methods of projecting migrant flows currently practiced in international demographic institutions, the United States of America, European countries, and other developed countries (Fertig and Schmidt, 2001; Howe and Jackson, 2006; Bijak, 2008; Raymer and Willekens, 2008; Rogers, 2008).

B. MAJOR FINDINGS

The dependent variable of the models is the logarithm (base-10) of the annual number, m_{ijt}, of migrants from origin country i to destination country j in the calendar year t. This dependent variable, called log migrant flow, has advantages and disadvantages. Compared to the choice of net migration as a dependent variable, migrant flow has the advantage of assuring that the world total of net migration will be zero (Cohen, 2008, p. 418). However, migrant flow as a dependent variable has the disadvantages that many countries do not produce reliable data on migrant flows, and that among countries that do publish such data, the definitions of migrant frequently differ from country to country.

Two models were developed for different purposes. The purpose of the first model (Cohen et al. 2008) was to find equations useful for prediction of migrant flow from any origin to any destination. The coefficients of the equations were estimated using 43653 recordsfrom 11 industrialized countries of annual migrant flows during 1960-2004. The countries were Australia, Belgium, Canada, Denmark, Germany, Italy, the Netherlands, Spain, Sweden, the United Kingdom [UK] and the United States of America. These countries reported numbers of immigrants annually from 228 origins and to 195 destinations.

The dependent variable log migrant flow was described by a log-linear model with independent variables that took account of calendar year, the populations of origin and destination, the geographical land areas of origin and destination, the great-circle distance from the capital city of the origin to the capital city of the destination, indicator variables for each of the eight countries that provided emigration data, indicator variables for each of the 11 countries that provided immigration data, indicator variables to identify the country that provided the migrant flow data, and a random error term.

An additional variant of this model included a “neighbor” multiplier if source and destination shared a common border. Adding this variable did not greatly improve the overall goodness of fit of the model, probably because few origin-destination pairs were geographical neighbors, but did indicate that being geographically adjacent increasedthe predicted number of migrants by 84.5 per cent when the influence of all other variables wastaken into account.

To clarify the operation of this model, consider migrant flows from Australia to UK as reported by Australia. The indicator variable for the origin Australia gave a multiplier for the proclivity of Australians to emigrate, after accounting for all other variables. The indicator variable for the source of the data (which was also Australia in this example) gave a multiplier for the specific definition and completeness of the enumeration of the Australian statistical system. The indicator variable for the destination UK gave a multiplier for the “receptiveness” to immigrants of UK, after accounting for all other variables.

In a second example, consider migrant flows from Australia to Bolivia. As Bolivia provided no migrant flow data in this study, the only possible source of the data is Australia. Indicator variables are invoked for Australia as in the previous example. No indicator variables for Bolivia are involved in predicting log migrant flow from Australia to Bolivia.

In a third example, consider migrant flows from Bolivia to Kenya. Neither country is a source of the migrant flow data used for this study. Consequently, no indicator variables are involved in predicting log migrant flows from Bolivia to Kenya. The model’s predictions are based entirely on the populations and areas of the origin and destination, the distance between their capitals, and the calendar year. The coefficients are estimated from all the data provided by the 11 reporting countries without data provided by either Bolivia or Kenya. This model makes it possible to estimate migrant flows from any origin to any destination.

A second model (Kim and Cohen, 2010) estimated one equation using 48832 records of migrant inflows to 17 Western countries and a separate equation using 28826 records of migrant outflows from 13 of these countries during 1950-2007. In the first equation, the dependent variable was log migrant inflows. In the second equation, the dependent variable was log migrant outflows.

Among the independent variables considered for inclusion in these equations were demographic, geographic, social and historical, and temporal determinants of migrant flows. The demographic determinants for both origin and destination included population sizes, potential support ratios (the number of people aged 15 to 64 per person aged 65 or over), infant mortality rates, and percentages of urban population. The geographic determinants for both origin and destination included distance between capitals, land area, being landlocked, and sharing a common border (as in “neighbor” above). The social and historical determinants included sharing a common official language, sharing a minority language, and having a prior colonial relationship. The temporal determinants were a term that was linear in the calendar year and another that was quadratic in the calendar year.

The independent variables most influential on log migrant inflows were demographic (log population of origin and destination and log infant mortality rate of origin and destination) and geographic (log distance between capitals and log land area of the destination). Social and historical determinants were less influential. For log outflows, the most influential independent variables were log population of origin and destination, log infant mortality rate of destination, and log distance between capitals. A young age structure in the destination was associated with lower inflows while a young age structure in the origin was associated with higher inflows. Urbanization in destination and origin increased international migration.

The infant mortality rate affected inflows and outflows significantly but oppositely. For inflows into the 17 developed countries, a higher infant mortality rate in the destination was associated with more immigration, while a higher infant mortality rate in the origin was associated with lower inflows. For outflows from the 13 developed countries, a higher infant mortality rate in the destination was associated with less emigration from the origin, while a higher infant mortality rate in the origin was associated with higher outflows.

Being landlocked, having a common border, having the same official language, sharing a minority language, and colonial links also had statistically significant but quantitatively smaller effects on international migration.

This second model answered four questions raised by the first model. Should migrant inflows to wealthy Western countries be modeled in the same way as migrant outflows? (No, the coefficients and the included independent variables were different.) Is the assumption (made in the first model) of uncorrelated residuals better than alternative error structures?(Yes, surprisingly.)How much do other demographic features besides population size influence migration? (The infant mortality rate mattered more than the potential-support ratio, but both mattered significantly.) How much do social and historical factors influence migration?(Substantially, but less than demographic and geographic factors.)

This second model provides more refined estimates than those of the first model for migrant flows from Australia to Bolivia and for migrant flows from Bolivia to Australia. It provides two estimates of migrant flows from Australia to UK, one from the outflow equation with Australia as the origin, the other from the inflow equation with UK as the destination. The two equations of this second model provide no predictions for migrant flows from Bolivia to Kenya or vice versa.

C. Future Plans & Conclusions

Within the framework of the two models just described, it would be desirable to make at least three kinds of quantitative comparisons of predicted migrant flows:

- from one industrialized country to another industrialized country given by the first model and the two equations (for inflows and outflows) of the second model,

- from data-source countries to other countries given by the first model and the migrant outflow equation of the second model,
- to data-source countries from other countries given by the first model and the migrant inflow equation of the second model.

It would also be desirable to compare the net migration of every country according to both models with the estimates of the United Nations Population Division.

It would be desirable to obtain additional data for other countries and years, while harmonizing the definitions of migrant flows used by different countries (Raymer, 2008; de Beer, Raymer, van der Erf, and van Wissen, 2010; Raymer, de Beer, and van der Erf, 2010).

From a theoretical point of view, it is desirable to study the asymptotic behavior of projection models with an embedded non-linear migration model such as the two models summarized here. Cohen et al. (2008, p. 15274) gave detailed mathematical questions.

To embed migration models such as these in a conventional population projection model, it will be necessary to distribute the projected total numbers of migrants by age and sex (Rogers et al., 2003;Raymer and Rogers, 2008; Raymer et al., 2010).

An alternative approach, namely, estimating migrant flows from differences in time-series of migrant stocks, is under active development.Ozden et al. (2009) estimated global bilateral migration by sex but not by age at decennial intervals from 1960 to 2000. Cunningham (2010) estimatedhistorical, and projected future, foreign-born age profiles (not by sex). These efforts face missing and incomplete data and inconsistent definitions in different countries. Eventually it would be desirable to find a way to integrate data on migrant stocks with data on migrant flows.

D. References

Bijak,Jakub (2008). Forecasting international migration in Europe: the Bayesian approach.Central European Forum for Migration and Population Research (CEFMR).Warsaw, 29 February.Dissertation.

Cohen, Joel E. (2008). Constant global population with demographic heterogeneity.*Demographic Research* (Rostock, Germany),vol. 18, No. 14 (May), pp. 409-436.

Cohen, Joel E., Marta Roig, Daniel C. Reuman, and CaiGoGwilt (2008). International migration beyond gravity: a statistical model for use in population projections. *Proceedings of the National Academy of Sciences*(Washington, DC), vol. 105, No. 40 (October), pp. 15269-15274.

Cunningham, Kevin (2010). Forecasting foreign-born age profiles: the Box-Cox power exponential model. United Nations Population Division, New York, 4 August.Manuscript.

de Beer, Joop, James Raymer, Rob van der Erfand Leo van Wissen (2010).Overcoming the problems of inconsistent international migration data: a new method applied to flows in Europe.Surmonter les problèmes d'incohérences des données sur les migrations internationales: une nouvelle méthode appliquée aux flux en Europe.*European Journal of Population* (New York), vol. 26, pp. 459–481.

Dorigo, Guido and Waldo Tobler (1983).Push-pull migration laws.*Annals of the Association of American Geographers* (Washington, DC), vol. 73, pp.1-17.

Faist, Thomas (2000).The Volume and Dynamics of International Migration and Transnational Social Spaces.Oxford and New York: Oxford University Press.

Fertig, Michael and Christoph M. Schmidt (2001).Aggregate-level migration studies as a tool for forecasting future migration streams. In *International Migration: Trends, Policy and Economic Impact*, Slobodan Djajic, ed. London: Routledge, pp. 110-136. Previously issued as Institute for the Study of Labor (IZA) Discussion Paper 183 (Bonn, Germany).

Howe, Neil and Richard Jackson (2006). Long-term immigration projection methods: current practice and how to improve it.Center for Retirement ResearchWorking Paper, No. 2006-3.Washington DC: Center for Strategic and International Studies (CSIS).

Kim, Keuntae and Joel E. Cohen (2010). Determinants of international migration flows to and from industrialized countries: a panel data approach beyond gravity. *International Migration Review*(New York), vol. 44, No. 4, pp. 899-932.

Massey, Douglas S., JoaquínArango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor (1993).Theories of international migration: a review and appraisal. *Population Development Review* (New York), vol. 19, No. 3 (September), pp. 431-466.

Massey, Douglas S., JoaquínArango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J. Edward Taylor (1998).*Worlds in Motion: Understanding International Migration at the End of the Millennium*. IUSSP International Studies in Demography. Oxford: Clarendon Press.

Ozden, Caglar, Christopher Parsons, Maurice Schiff and Terrie Walmsley(2009).The evolution of global bilateral migration 1960-2000.2nd Migration & Development conference.World Bank, Washington D.C., September.Paper.

Raymer, James (2008).Obtaining an overall picture of population movement in the European Union. In *International Migration in Europe: Data, Models and Estimates,* James Raymer and FransWillekens, eds.Chichester: John Wiley, pp. 209-234.

Raymer, James and Andrei Rogers (2008). Applying model migration schedules to represent age-specific migration flows. In *International Migration in Europe: Data, Models and Estimates*, James Raymer and FransWillekens, eds.Chichester: John Wiley, pp. 175-192.

Raymer, James and FransWillekens, eds. (2008).*International Migration in Europe: Data, Models and Estimates*. Chichester: Wiley.

Raymer, James, Joop de Beer and Rob van der Erf (2010).Putting the pieces of the puzzle together: age and sex-specific estimates of migration between EU / EFTA countries, 2002-2007.Southampton Statistical Sciences Research Institute, University of Southampton. 27 August 2010. 48 pp.

Ritchey, P. Neal (1976). Explanations of migration.*Annual Review of Sociology* (Palo Alto, California),vol. 2, pp. 363-404.

Rogers, Andrei (2008). Demographic modeling of the geography of migration and population: a multiregional perspective. *Geographical Analysis* (Columbus, Ohio),vol. 40, pp. 276–296.

Rogers, Andrei, FransWillekensand James Raymer (2003).Imposing age and spatial structures on inadequate migration-flow datasets.*The Professional Geographer* (Washington, DC),vol. 55, No. 1, pp. 56–69.

United Nations Population Division (2003).United Nations Population Projections to 2300.ESA/P/WP.184, 30 June 2003**.**UNPD: New York.

Zipf, George K. (1946). The P_1P_2/D hypothesis: on the inter-city movement of persons. *American Sociological Review* (Washington, DC), vol.11, pp. 677-686.

Zipf, George K. (1949). *Human Behavior and the Principle of Least Effort*.Cambridge, Massachusetts: Addison Wesley.