

Estimating Global Migration Flow Tables Using Place of Birth Data

Guy J. Abel*

Wittgenstein Centre for Demography and Global Human Capital,
Vienna Institute of Demography, Austria

October 2011

1 Introduction

A methodology to estimate global migration flow tables of movements between all countries is illustrated. The estimated tables are based on sequential place of birth migration stock tables, recently released by the World Bank, see Özden et al. (2011). Using a simple dummy example the basic methodology to derive flows from sequential stock tables is given. This methodology is then applied to the World Bank migration stock data to produce a set of annual global migration flow tables between 1960 and 1999. Selected results are presented. These tables are the first attempt to produce a comparable set of global migration flow estimates, that will ultimately be used as the migration baseline data for the upcoming IIASA-Oxford expert argument-based population projections.

2 Data

International migration flow data provided by national statistical offices are not comparable. Tables of bilateral international migration flow data have severe missing data problems and inconsistent measurements, see for example Nowok et al. (2006). Efforts to estimate European migration tables of comparable flow data have been partially successful, see for example Abel (2010); de Beer et al. (2010); Raymer et al. (2010). However, these methodologies rely on a reasonable percentage of double counted flows, i.e., reported values of a movements from both the sending and receiving counties. The application of these estimation methods to global data does not appear to be possible, as the availability of reported migration flows from non-European countries remains scarce.

The World Bank recently released a global bilateral foreign born migrant stock database for the last five census rounds (Özden et al., 2011). The data is primarily based on place of birth responses to Census questions or details collected from population registers. In order to construct a set of complete bilateral tables, issues of definitions, changes in geography, aggregated

*Contact Email: guy.abel@oeaw.ac.at

data and missing values were addressed by the World Bank. The resulting tables represent the most comparable global data set of past international migration stocks available.

3 Methodology

The idea of estimating migration flows from place of birth data is not new. Work by Rogers and von Rabenau (1971) and Rogers and Raymer (2005) focused on US Census place of birth data to estimate inter-state migration streams. The methodology outlined in this section partly builds on their insights.

Consider two sequential place of birth stock tables, as in Table 1, where the rows represent the countries of birth and the columns represent the countries of residence. These can be

Table 1: Dummy Example of Place of Birth Data

P^t	A	B	C	D	Sum	P^{t+1}	A	B	C	D	Sum
A	900	250	10	30	1190	A	850	250	60	30	1190
B	100	500	30	20	650	B	125	450	55	20	650
C	5	20	200	0	225	C	15	10	200	0	225
D	40	100	25	600	765	D	60	80	5	620	765
Sum	1045	870	265	650	2830	Sum	1050	790	320	670	2830

expressed as set of population stocks, P_{bi}^t and P_{bj}^{t+1} where the b subscript represents the place of birth, the i subscript represents the populations location at time t and the j subscript represents the populations location at time $t + 1$. Note, between these time periods there are no natural changes (from births and deaths) in the population totals. Differences in cells between years are solely driven by migration flows, which can only occur across rows. Thus the row totals representing the sum of nationals, in some location, does not change. We may also consider this data as a set of incomplete birth place specific migration tables, shown in Table 2

The cells in each of these tables may be considered as missing data from a log-linear model, which are commonly used in the estimation of migration flow tables when only marginal totals are known, see for example Willekens (1999). A range of log-linear models can be fitted using a constrained maximisation routine developed by Raymer et al. (2007), to obtain maximum likelihood estimates. As the margins of these tables include populations that will not move, it is assumed that the diagonal elements are composed of the maximum values possible, given the known marginal totals. These diagonal cells and marginal totals are constrained by fitting a quasi-independent log-linear model, with dummy variables to saturate the fit to diagonal terms. The resulting estimates for the non-diagonal terms in the sub-tables may then be derived as in Table 3.

To obtain an origin-destination migration flow table we may then sum over all birth places, ignoring the diagonal elements of those that did not move, displayed in Table 4.

The resulting estimates represent the minimal number of migrants that are required to transit during the time period (t to $t + 1$) in order to meet the place of birth stock table P_{bj}^{t+1} given in Table 1.

Table 2: Initial Migration Flows for each Place of Birth

$b = A$	A	B	C	D	Sum	$b = B$	A	B	C	D	Sum
A					900	A					100
B					250	B					500
C					10	C					30
D					30	D					20
Sum	850	250	60	30	1190	Sum	125	450	55	20	650
$b = C$	A	B	C	D	Sum	$b = D$	A	B	C	D	Sum
A					5	A					40
B					20	B					100
C					200	C					25
D					0	D					600
Sum	15	10	200	0	225	Sum	60	80	5	620	765

3.1 Extensions for Births, Deaths and Missing

In reality, natural changes from births and deaths in the population occur, causing differences in the row totals between subsequent years. In addition, members of the population who are alive in both time periods may have not been recorded in one of the periods. Each of these problems will be briefly discussed in turn.

Increases caused by births are adjusted for by subtracting the number of births in each country from their native population totals (the diagonal cells) in the second stock table, P_{bj}^{t+1} , (cells where $b = j$). Subtracting the births in the second stock table allows us to constrain the marginal totals for the sub-tables, such as those in Table 2, to ignore natural increase in populations and hence consider only changes from migration in the constrained maximization procedure. Subtracting the births from the native population implies a simplification that new members of the population cannot migrate until their next year of life.

Decreases caused by deaths are adjusted for by creating an extra column in each of the birth place specific matrices shown in Tables 2 and 3. This additional destination (of death) can be constrained to the total reported number of deaths for each country. This corresponds to constraint on the column sum of the final origin-designation migration flow table, rather than individual column sums in the place of birth migration flow tables. This constraint is easily incorporated into the contained maximisation routine using the same quasi-independent model, with an additional level for the destination covariate.

Finally, changes in the population totals between the birth-place stock tables, which have not arrived from births or deaths, can be incorporated with additional row and columns to the rest of the world, in a similar manner as to that previously described for handling deaths. These differences typically occurred for flows to small states, not included in our analysis, or where there have been significant alterations in reporting of stock totals due to measurement issues. The additional categories, of unknown origins or destinations, allows for the identification of changes in stocks that might have occurred due to measurement issues rather than migration.

Table 3: Estimated Migration Flows for each Place of Birth

$b = A$	A	B	C	D	Sum	$b = B$	A	B	C	D	Sum
A	850	0	50	0	900	A	100	0	0	0	100
B	0	250	0	0	250	B	25	450	25	0	500
C	0	0	10	0	10	C	0	0	30	0	30
D	0	0	0	30	30	D	0	0	0	20	20
Sum	850	250	60	30	1190	Sum	125	450	55	20	650
$b = C$	A	B	C	D	Sum	$b = D$	A	B	C	D	Sum
A	5	0	0	0	5	A	40	0	0	0	40
B	10	10	0	0	20	B	10	80	0	10	100
C	0	0	200	0	200	C	10	0	5	10	25
D	0	0	0	0	0	D	0	0	0	600	600
Sum	15	10	200	0	225	Sum	60	80	5	620	765

Table 4: Estimated Migration Flow Table

	A	B	C	D	Sum
A		0	50	0	50
B	45		25	10	80
C	10	0		10	20
D	0	0	0		0
Sum	55	0	75	20	150

4 Results

Place of birth data published by the World Bank were used to provide foreign born migration stock tables at the start of each of the last five decades. As annual flow tables are required for single year population projections, smoothed stock tables were calculated using splines to interpolate between foreign born populations for every cell in the stock table throughout the five Census periods. Diagonal totals in each stock table of the native-born population totals, not provided in the World Bank foreign born migration data, were then derived as a remainder using annual population totals from the World Bank Development Indicators Database. This constrained all column totals of the stock tables to meet those of the reported annual population totals. The conditional maximisation routine was then run to calculate the sequential migration flow tables, accounting for reported births and deaths. The result was a set of 195×195 migration flow tables for each year between 1960 and 2010. An example of the estimates is shown in Figure 1 for 1960 and Figure 2 for 1999. The width of the lines indicates size of the flow. The colouring in the lines indicates the direction of the flow. The more green the line, the greater the inflow, the more red, the greater the outflow. Plots were built using JFlowMap, see Boyandin et al. (2010).

Figure 1: Estimated Global Migration Flows, 1960

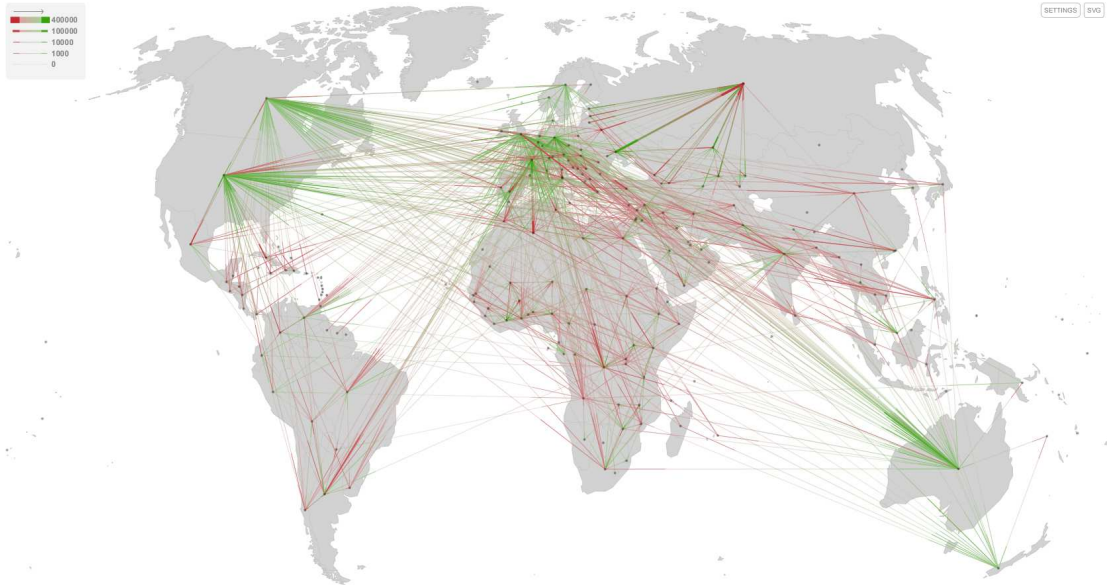
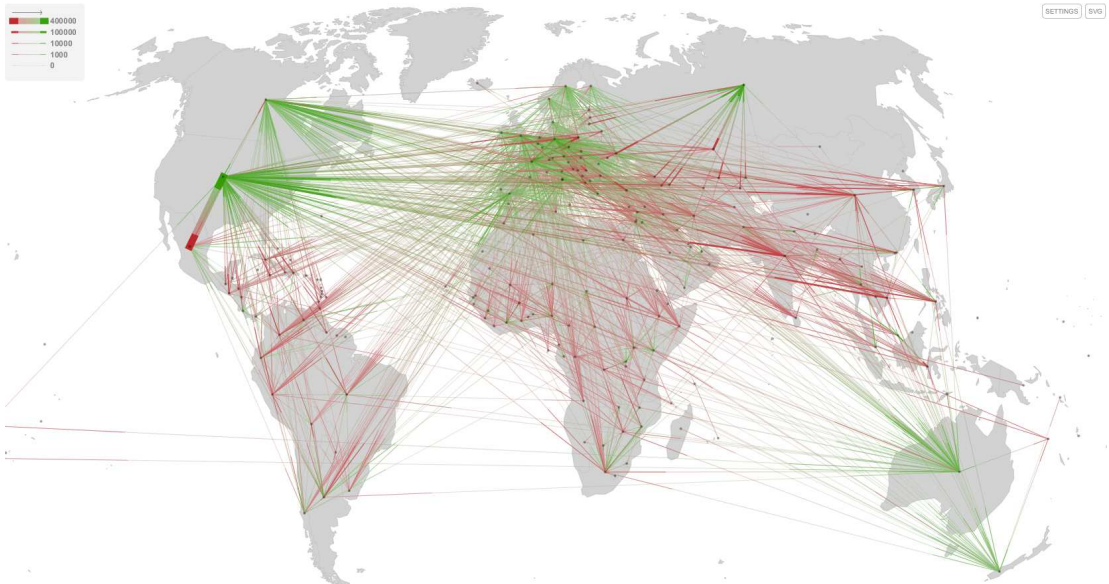


Figure 2: Estimated Global Migration Flows, 1999



5 Conclusion

In this paper, a methodology to estimate global migration flow tables of movements between 195 countries of the world is illustrated. The estimated tables are based on sequential place of birth migration stock tables, recently released by the World Bank, see Özden et al. (2011). Unlike previous methodologies for estimating tables of comparable international migration flow tables, estimation is not reliant on immigration and emigration flow data, which is often unavailable and inconsistent at the global level. The application of the outlined methodology allows for a global view of migration flows to be further explored. In addition, they serve as a initial set of detailed baseline data for global population projections, which have tended to previously rely on simplified assumptions of net migration despite its known weaknesses, see for example Rogers (1990).

References

- Abel, G. (2010). Estimation of international migration flow tables in europe. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Boyandin, I., E. Bertini, and D. Lalanne (2010, May). Using flow maps to explore migrations over time. In *Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science*, Volume 2.
- de Beer, J., J. Raymer, R. van der Erf, and L. van Wissen (2010). Overcoming the problems of inconsistent international migration data: A new method applied to flows in europe. *European Journal of Population/Revue européenne de Démographie*, 1–23.
- Nowok, B., D. Kupiszewska, and M. Poulain (2006). Statistics on international migration flows. In M. Poulain, N. Perrin, and A. Singleton (Eds.), *Towards the Harmonisation of European Statistics on International Migration (THESIM)*, Chapter 8, pp. 203–233. Louvain-La-Neuve, Belgium: UCL–Presses Universitaires de Louvain.
- Özden, Ç., C. Parsons, M. Schiff, and T. Walmsley (2011). Where on earth is everybody? the evolution of global bilateral migration 1960–2000. *The World Bank Economic Review* 25(1), 12.
- Raymer, J., G. Abel, and P. Smith (2007). Combining census and registration data to estimate detailed elderly migration flows in england and wales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170(4), 891–908.
- Raymer, J., J. J. Forster, P. W. F. Smith, J. Bijak, A. Wiśniowski, and G. J. Abel (2010, April). The IMEM model for estimating international migration flows in the European Union. In *Joint UNECE/Eurostat Work Session on Migration Statistics*, Joint UNECE/Eurostat Work Session on Migration Statistics. United Nations Statistical Commission And European Commission Economic Commission For Europe Statistical Office Of The European Communities (EUROSTAT).
- Rogers, A. (1990). Requiem for the net migrant. *Geographical Analysis* 22(4), 283–300.

- Rogers, A. and J. Raymer (2005). Origin dependence, secondary migration, and the indirect estimation of migration flows from population stocks. *Journal of Population Research* 22(1), 1–19.
- Rogers, A. and B. von Rabenau (1971). Estimation of interregional migration streams from place-of-birth-by-residence data. *Demography* 8(2), 185–194.
- Willekens, F. (1999). Modeling approaches to the indirect estimation of migration flows: from entropy to em. *Mathematical population studies* 7(3), 239.