DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 28, ARTICLE 18, PAGES 505-546
PUBLISHED 15 MARCH 2013
http://www.demographic-research.org/Volumes/Vol28/18/
DOI: 10.4054/DemRes.2013.28.18

Research Article

Estimating global migration flow tables using place of birth data

Guy J. Abel

Table of Contents

Demographic Research: Volume 28, Article 18
Research Article

Estimating global migration flow tables using place of birth data

Guy J. Abel [1]

Abstract

BACKGROUND
International migration flow data often lack adequate measurements of volume, direction and completeness. These pitfalls limit empirical comparative studies of migration and cross national population projections to use net migration measures or inadequate data.

OBJECTIVE
This paper aims to address these issues at a global level, presenting estimates of bilateral flow tables between 191 countries.

METHODS
A methodology to estimate flow tables of migration transitions for the globe is illustrated in two parts. First, a methodology to derive flows from sequential stock tables is developed. Second, the methodology is applied to recently released World Bank migration stock tables between 1960 and 2000 (Özden et al. 2011) to estimate a set of four decadal global migration flow tables.

RESULTS
The results of the applied methodology are discussed with reference to comparable estimates of global net migration flows of the United Nations and models for international migration flows.

COMMENTS
The proposed methodology adds to the limited existing literature on linking migration flows to stocks. The estimated flow tables represent a first-of-a-kind set of comparable global origin destination flow data.

---

[1] Wittgenstein Centre (IIASA, VID/ÖAW, WU), Vienna Institute of Demography/Austrian Academy of Sciences. E-mail: guy.abel@oeaw.ac.at.

1. Introduction

International moves are typically enumerated in a demographic context using either a measurement of migrant stocks or migration flows. A migrant stock is defined as the total number of international migrants present in a given country at a particular point of time. A migration flow is defined as the number of persons arriving or leaving a given country over the course of a specific period of time. Flow measures reflect the dynamics of the migration process and are typically considered to be less tractable than stock measures (Bilsborrow et al. 1997, p.51).

International migration flow data often lacks adequate measurements of volume, direction and completeness (Kelly 1987; Salt 1993; Willekens 1994; Nowok, Kupiszewska, and Poulain 2006), making cross national comparisons difficult. The lack of comparability in flow data can be traced to a number of causes. First, migration is a multidimensional process (Goldstein 1976) involving a transition between two states. Consequently, movements can be reported by sending or receiving countries. When data collection methods or measurements in countries differ, the reported counts do not match. Second, international migration flow data are typically collected by individual national statistics institutes in each country, where measures have been designed to suit solely domestic priorities. Data are often produced within a legal framework, and hence alterations to their collection are difficult to implement. Finally, in many countries data collection systems for migration flow data do not exist. In other countries, collection methods such as passenger surveys may prove inadequate to report flows at the levels of detail required by some data users.

This paper aims to circumnavigate these difficulties by developing a methodology to derive bilateral migration flows from sequential stock tables. Basing estimated migration flows upon stocks has a number of potential advantages. First, bilateral migration flow tables can be considered as part of a wider account of demographic data. Rees (1980) noted that national account statistics of financial stocks and flows have served economists well in their modelling activities, encouraging users to compare data for consistencies, check for inadequacies, and force analysts to attempt to match available data with a conceptual model. He noted that a similar system of demographic accounts in migration stocks and flows would likely lead to similar improvements. Second, stock data are, in comparison to international migration flow data, far easier to measure. Stock data are more widely available, both across time and countries. This is reflected in the World Bank migration stock data which include bilateral records from over 200 nations and four decades (Özden et al. 2011). In comparison, the 2010 revision of bilateral international migration flow data released by the United Nations (Henning and Hovy 2011) covers only 43 countries, predominately developed nations, from the last two decades. The greater availability of migrant stock data makes it an invaluable source of information on migrant patterns that, as illustrated in the methodological section of this paper, can be used as a basis to estimate global bilateral migration flow tables.

Estimates of global bilateral flow tables can potentially have a number of advantages over presently available international migration data. First, they allow a fuller understanding of population behaviour and change in comparison to other measures of migration. Studying people's movements using both origin and destination dimensions furthers the possibility of deeper insights into migration patterns and migrant behaviours. Such insights can be confounded by more conventional methods of analysis on existing measures of migration flows, such as those to or from only a few select nations, or through the study of net migration. Second, bilateral migration flow tables allow a comparison of migration propensities across multiple countries. Consequently, the contributions made by each nation to the global system of migration can be more easily identified, and comparative summaries of migration flows become more meaningful in a multinational context. Third, estimates of bilateral migration flow tables permit a more comprehensive empirical source for testing migration theories. In addition, the study of public policies towards the control or encouragement of migration flows can be further expanded to incorporate a wider evidence base. Fourth, while stock data can be collected more easily than flow data, information on migration patterns from studying stocks can potentially provide a poor indication of contemporary international migration flows. Furthermore, in countries where there are significant return migrations or mortality among foreign population, migrant stock data can a yield a misleading portrait of the current migration system (Massey et al. 1999, p.200). Fifth, estimates of international migration flow tables can provide a more perceptive base data for global population forecasts. Current global population forecasts, such as KC et al. (2010) or United Nations Population Division (2011) utilise only estimated international net migration totals, where no other comparable migration data exists for population forecasters on such scales (Kupiszewski and Kupiszewska 2008). As such, global forecasting exercises often run the risk of well documented problems of using net migration measures in their models; see for example Rogers (1990) or Rogers (1995). These potential benefits are leading international organisations to call for the promotion and development of methodologies for the collection and processing of internationally comparable statistical data on international migration (United Nations General Assembly 2011).

There is a limited literature on linking migration flows to stocks. Rogers and von Rabenau (1971), Rogers and Raymer (2005) and Rogers and Liu (2005) focused on US Census place of birth data to estimate inter-state migration transitions. The first of these papers relies on a simplistic principal assumption of equal growth in the migrant stock totals in all regions. This assumption is unrealistic when comparing data from multiple countries and is likely to produce many erroneous (including negative) estimates of migration flows. The second and third papers rely on detailed disaggregation of stock data, which are not currently available for international migration data at the global scale.

In this paper a new methodology to estimate global flow tables of migrant transitions between all countries is illustrated in two parts. First, a methodology to derive flows from sequential stock tables is developed. Using a set of simple hypothetical data, rather than a full global table, the methodology is initially demonstrated in a scenario where there are no births and deaths in migrant stock populations between two periods. An extension for natural changes in population totals is then shown. Estimation is undertaken using a spatial interaction model, equivalent to a log-linear model in statistics. Such methods have a developed literature in the indirect estimation of internal migration flows, where marginal totals (immigration and emigration flows into and out of a set of regions) are known, but the table contents are missing, see for example Fotheringham and O'Kelly (1988) or Willekens (1999). Second, the methodology is applied to recently released World Bank migration stock tables between 1960 and 2000 (Özden et al. 2011) to estimate a set of four decadal global migration flow tables. Summary results of the applied methodology are first presented and then discussed with reference to comparable estimates of global net migration flows of the United Nations and models for international migration. These validation exercises give an indication of the performance of the applied methodology. In the final section, potential extensions are outlined and conclusions given.

2. Methodology

A general methodology for the estimation of migration flows from sequential migrant stock tables is derived in this section. The estimation of migration flows between a set of regions is illustrated using a set of simple hypothetical data. Increasing complexity is added to account for additional factors that might also effect changes in stock totals, besides migration flows, such as births, deaths, and movements to and from external regions not considered.

Bilateral migration data are commonly represented in square tables. Values within the table vary, depending on definitions used in data collection or the research question at hand. Values in non-diagonal cells represent some form of movement, for example a migration flow between a specified set of R regions or areas or a foreign born stock. Values in diagonal cells represent some form of non-moving population, or those that move within a region, and are sometimes not presented.

Consider two migrant stock tables in consecutive years (t and $t+1$) in the top panel of Table 1. Regions A to D represent places of birth in the rows, and place of residence in the columns. Hence, non-diagonal entries represent the number of foreign born migrants in each area of residence, while diagonal entries contain the number of native born residents.

Table 1: Dummy example of place of birth migrant stock data

Place of Birth Data in Stock Tables:

Place of Residence (t)

		A	B	C	D	Sum
Place of Birth	A	1000	100	10	0	1110
	B	55	555	50	5	665
	C	80	40	800	40	960
	D	20	25	20	200	265
	Sum	1155	720	880	245	3000

Place of Residence ($t+1$)

		A	B	C	D	Sum
Place of Birth	A	950	100	60	0	1110
	B	80	505	75	5	665
	C	90	30	800	40	960
	D	40	45	0	180	265
	Sum	1160	680	935	225	3000

Place of Birth Data in Flow Tables:

Place of Birth=A

		Destination				
		A	B	C	D	Sum
Origin	A					1000
	B					100
	C					10
	D					0
	Sum	950	100	60	0	1110

Place of Birth=B

		Destination				
		A	B	C	D	Sum
Origin	A					55
	B					555
	C					50
	D					5
	Sum	80	505	75	5	665

Place of Birth=C

		Destination				
		A	B	C	D	Sum
Origin	A					80
	B					40
	C					800
	D					40
	Sum	90	30	800	40	960

Place of Birth=D

		Destination				
		A	B	C	D	Sum
Origin	A					20
	B					25
	C					20
	D					200
	Sum	40	45	0	180	265

In this hypothetical example there are no births or deaths. This results in two noticeable features in the tables. First, the row totals in each time period remain the same, as the number people born in each region cannot increase or decrease. Second, differences in cells must implicitly be driven solely by migration flows. These movements occur when individuals change their place of residence (moving across columns), while their place of birth (row) characteristic remains fixed.

To derive a corresponding set of flows that are constrained to meet the stocks tables, we can alternatively consider the top panel of Table 1 as a set of R birthplace specific migration flow tables where the marginal totals are known, shown in the bottom panel of Table 1. These are formed by considering each row of the two consecutive stock tables as a set of separate margins of a migration flow table. Place of residence totals at time t from the stock data now become origin margin (row) totals for each birth place specific population. Similarly, place of residence totals at time $t + 1$ from the stock data now become destination margin (column) totals for each birth place specific population. As the row totals from the stock tables are equal, the row and column margins in each of the birth place specific migration flow tables in Table 1 are also equal.

Typically migration flow measures can be classified as movement or transition data (Rees and Willekens 1986). Movement data consist of counts migration events across boundaries. Transition data consist of counts of migrants whose location at the end of a specified time periods is different to that at the beginning of the period. Within each birthplace specific table in the bottom panel of Table 1, missing non-diagonal cells must represent the migrant transition flows from origin i to destination j, within time period t to $t + 1$ and categorised by birthplace k. Missing diagonal entries represent the number of people who reside in the same region at t and $t + 1$, which are referred to as stayers throughout the remainder of this paper. In order to estimate the missing migrant transition flows and stayers, model based methods are used to impute values that are constrained to the known marginal totals.

Flowerdew (1991) outlined two main approaches for the use of models to either analyse or estimate flow tables for internal migration data: the gravity model and the spatial interaction model. The gravity model approach derives from movements between regions in a similar manner to particle responses to two gravitational masses, as proposed by Newton in Principia Mathematica. Stewart (1941) and Zipf (1942) framed this approach for migration data, relying on statistical estimation of migration volumes, given information on each origin, destination and a measurement of association between them. The spatial interaction models, associated with Wilson (1970), are based on mathematical algorithms to calibrate a constrained model to origin and destination totals. There are numerous formulations of spatial interaction models such as bi-proportional adjustment, information gain minimizing and entropy maximizing which include various constraints and interaction terms (Willekens 1983).

Poisson regression models have become a popular method for representing migration models as they relate gravity and many spatial interaction models in a single comparative framework. Flowerdew (1982) and Willekens (1983) showed that a Poisson regression model with either row or column dummy covariates is equivalent to an origin or destination constrained spatial interaction model, and when both covariates are present, a doubly constrained spatial interaction is obtained. Such representations, with only categorical covariates, are equivalent to the log-linear regression models of Birch (1963). When row or column dummy covariates are not included, but other origin and destination specific factors are, such as population size, the resulting Poisson regression model is equivalent to the gravity model first proposed by Zipf (Flowerdew 1991).

A simplistic version of the spatial interaction model for the number of migrants in transition n_{ij} from origin i to destination j, during the respective time interval, as in each of the $R = 4$ incomplete data situations of the bottom panel of Table 1, may be considered;

$$y_{ij} = \alpha_i \beta_j m_{ij} \tag{1}$$

where y_{ij} is the expected number of migrants in transition from origin i to destination j and $i, j = 1, 2, \dots, R$ for R origins and destinations. The α_i and β_j parameters represent the background factors that are related to the characteristics of the origin and destination. The m_{ij} factor represents some auxiliary information on migration flows. This is typically additional data related to migration between the same origins and destinations. Willekens (1999) noted, in conventional spatial interaction analysis, $m_{ij} = F(d_{ij})$ where d_{ij} is a measure of distance between i and j and $F(.)$ is a distance deterrence function. Such distance deterrence functions can come in different forms, such as $F(d_{ij}) = d_{ij}^{-\epsilon}$ or $F(d_{ij}) = \exp(-\epsilon d_{ij})$, where $\epsilon > 0$ is a distance sensitivity parameter; see Sen and Smith (1995, p4). Alternative specifications for m_{ij} might be travel costs or past migration flows.

As described by Willekens (1999), the estimation of parameters in a spatial interaction model can be performed by re-expressing the spatial interaction model of (1) in terms of a log-linear model:

$$\log y_{ij} = \log \alpha_i + \log \beta_j + \log m_{ij}, \tag{2}$$

where unlike standard log-linear models, no intercept is included, and the final term is commonly referred to as an offset. The maximum likelihood estimates for the α_i and β_j parameters in (2) can be derived by considering the probability of observing n_{ij} migrant transitions during a unit interval, given by the Poisson distribution function:

$$P(N_{ij} = n_{ij}) = \frac{y_{ij}^{n_{ij}}}{n_{ij}!} \exp(-y_{ij}). \tag{3}$$

The likelihood function for $\mathbf{Y} = \{y_{ij}, i, j, = 1, \dots, R\}$ given $\mathbf{n} = \{n_{ij}, i, j, = 1, \dots, R\}$ migrant transitions, provided that migrant transitions are independent, is

$$L(\mathbf{Y}; \mathbf{n}) = P(N_{11} = n_{11}, N_{12} = n_{12}, \dots, N_{RR} = n_{RR}) = \prod_{ij} \frac{y_{ij}^{n_{ij}}}{n_{ij}!} \exp(-y_{ij}) \tag{4}$$

Inserting the log-linear spatial model of (1) into expression (4) and taking the logarithmic transformation gives the log-likelihood function:

$$\begin{aligned} l(\boldsymbol{\theta}; \mathbf{n}) &= \sum_{ij} \{ n_{ij} \log(\alpha_i \beta_j m_{ij}) - \alpha_i \beta_j m_{ij} - \log(n_{ij}!) \} \\ &= \sum_i n_{i+} \log(\alpha_i) + \sum_j n_{+j} \log(\beta_j) - \sum_{ij} \alpha_i \beta_j m_{ij} + c, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta} = \{\alpha_i, \beta_j, i, j = 1, \ldots, R\}$, $n_{i+} = \sum\limits_j n_{ij}$ and $n_{+j} = \sum\limits_i n_{ij}$ are the marginal totals, and

$$c = \sum_{ij} n_{ij} \log(m_{ij}) - \sum_{ij} \log(n_{ij}!). \tag{6}$$

The maximum likelihood estimates of α_i and β_j are obtained by maximising the log-likelihood function (5). The extra term c, which does not involve the parameters, may be ignored. Thus, conveniently only the marginal totals from the consecutive stock tables are required to estimate the spatial interaction model.

Differentiation of the likelihood function with respect to each parameter gives us the likelihood equations:

$$\frac{\partial l}{\partial \alpha_i} = \frac{n_{i+}}{\alpha_i} - \sum_j \beta_j m_{ij} = 0 \tag{7}$$

and

$$\frac{\partial l}{\partial \beta_j} = \frac{n_{+j}}{\beta_j} - \sum_i \alpha_i m_{ij} = 0. \tag{8}$$

The maximum likelihood estimators for α_i and β_j can then be written as

$$\hat{\alpha}_i = \frac{n_{i+}}{\sum\limits_j \hat{\beta}_j m_{ij}} \tag{9}$$

and

$$\hat{\beta}_j = \frac{n_{+j}}{\sum\limits_i \hat{\alpha}_i m_{ij}}. \tag{10}$$

Direct estimates of α_i and β_j cannot be obtained, since there are no closed-form expressions for the solution of equation (9) and (10). However, as described in Willekens (1999), an iterative procedure can be used to derive indirect estimates. Given initial estimates of

$\hat{\beta}_j^{(0)}$, estimates of $\hat{\alpha}_i^{(1)} = n_{i+} \Big/ \sum_j \hat{\beta}_j^{(0)} m_{ij}$. These estimates are then used to update $\hat{\beta}_j^{(2)} = n_{+j} \Big/ \sum_i \hat{\alpha}_j^{(1)} m_{ij}$. This process repeats until convergence. Maximum likelihood estimates of y_{ij}, the expected number of migrant transitions, are deduced from the converged estimates of $\hat{\alpha}_i$ and $\hat{\beta}_j$ using the spatial interaction model of (1). Willekens (1999) discusses how this procedure is a special case of the iterative proportional fitting algorithm and the Expectation-Maximisation (EM) algorithm. As noted in Raymer, Abel, and Smith (2007), this is also a conditional maximisation, also known as a stepwise ascent.

Using the sufficient statistics of $\mathbf{n}$ shown in Table 1 and the converged estimates of $\hat{\alpha}_i$ and $\hat{\beta}_j$ in the spatial interaction model (12) gives the maximum likelihood estimates of y_{ij}, the expected number of migrant transitions. These values are shown in the top panel of Table 2 for each birthplace specific table. The iterative procedure to estimate $\hat{\alpha}_i$ and $\hat{\beta}_j$ and y_{ij} is undertaken using the `cm2` routine in the `migest` R package (Abel 2012), where all elements of m_{ij} are set to unity ($m_{ij} = 1$). Summing over all birthplaces and deleting stayers in the diagonal elements results in a traditional flow table of migrant transitions from origin i to destination j during the time period t to $t + 1$ shown in the bottom panel of Table 2.

The spatial interaction model of (1) focuses on estimating migrant transitions between two dimensions, origin and destination (rows and columns). This model can be assumed to derive estimates in each of the individual $R = 4$ birthplace specific flow tables presented in the bottom panel of Table 1. Alternatively, the model can be expanded to include a third (table) dimension by adding parameters to consider all birthplace specific tables simultaneously,

$$\log y_{ijk} = \log \alpha_i + \log \beta_j + \log \lambda_k + \log \gamma_{ik} + \log \kappa_{jk} + \log m_{ij} \tag{11}$$

where y_{ijk} is the expected number of migrant transitions from origin i to destination j of people born in birthplace k, during the respective time interval and $i, j, k = 1, 2, \ldots, R$, for R origins, destinations and birthplaces. The α_i, β_j and λ_k parameters represent background factors that relate to the characteristics of the origins, destinations and birthplaces respectively. The γ_{ik} and κ_{jk} parameter sets represent the factors specific to each origin-birthplace and destination-birthplace specific combinations respectively. As in the previous model (1), estimates of all parameter sets can be obtained through solving the extended set of likelihood equations (not shown). Estimates of the fitted values from (11) result in the identical estimates to those in Table 2, so long as elements of the offset are again set to unity. Utilising the solutions to the likelihood equations from (11) allows the simultaneous estimation of parameters to estimate flows in all R tables, rather than individually considering each, one at a time.

Table 2: Estimates of migrant transition flow tables based on stock data in Table 1

Estimates of Origin Destination Place of Birth Flow Tables:

Place of Birth=A

		Destination				
		A	B	C	D	Sum
Origin	A	856	90	54	0	1000
	B	86	9	5	0	100
	C	9	1	1	0	10
	D	0	0	0	0	0
	Sum	950	100	60	0	1110

Place of Birth=B

		Destination				
		A	B	C	D	Sum
Origin	A	7	42	6	0	55
	B	67	421	63	4	555
	C	6	38	6	0	50
	D	1	4	1	0	5
	Sum	80	505	75	5	665

Place of Birth=C

		Destination				
		A	B	C	D	Sum
Origin	A	8	3	67	3	80
	B	4	1	33	2	40
	C	75	25	667	33	800
	D	4	1	33	2	40
	Sum	90	30	800	40	960

Place of Birth=D

		Destination				
		A	B	C	D	Sum
Origin	A	3	3	0	14	20
	B	4	4	0	17	25
	C	3	3	0	14	20
	D	30	34	0	136	20
	Sum	40	45	0	180	265

Estimates of Total Origin Destination Flow Table:

		Destination				
		A	B	C	D	Sum
Origin	A		138	127	17	282
	B	160		101	23	284
	C	93	67		47	207
	D	35	39	34		107
	Sum	287	244	262	87	881

2.1 Extensions for non-movers

Spatial interaction models, such as (1) and (11) have previously been used to estimate unknown cells in migration flow tables with known margins; see for example Raymer, Abel, and Smith (2007). However, the margins under consideration have traditionally been sums of inflows and outflows, to and from, a set of regions. The marginal data in the bottom panel of Table 1 is based on a combination of migrant transitions and stayers. Hence, in order to estimate solely the migrant transitions, an assumption about the number of stayers in the migration system must be taken, and a model adapted accordingly. Not extending a model to account for differences in migrants and stayers would effectively impose the assumption that the cost of migration is the same as the cost not to migrate. One such extension is to fix the diagonal terms in each sub-table to their maximum value, without violating each corresponding marginal constraints, as in Table 3.

Table 3: Migrant transition flow tables for each place of birth with assumed non-movers

Place of Birth=A

		Destination				
		A	B	C	D	Sum
Origin	A	950				1000
	B		100			100
	C			10		10
	D				0	0
	Sum	950	100	60	0	1110

Place of Birth=B

		Destination				
		A	B	C	D	Sum
Origin	A	55				55
	B		505			555
	C			50		50
	D				5	5
	Sum	80	505	75	5	665

Place of Birth=C

		Destination				
		A	B	C	D	Sum
Origin	A	80				80
	B		30			40
	C			800		800
	D				40	40
	Sum	90	30	800	40	960

Place of Birth=D

		Destination				
		A	B	C	D	Sum
Origin	A	20				20
	B		25			25
	C			0		20
	D				180	200
	Sum	40	45	0	180	265

Consequently, the non-diagonal estimates will represent the minimum number of migration transitions from origin i to destination j between time t and $t+1$ while maintaining marginal constraints.

In order to account for known diagonal elements in each of the sub-tables, an additional parameter can be added to (11),

$$\log y_{ijk} = \log \alpha_i + \log \beta_j + \log \lambda_k + \log \gamma_{ik} + \log \kappa_{jk} + \log \delta_{ijk} I(i = j) + \log m_{ij}, \quad (12)$$

where $I(\cdot)$ is the indicator function,

$$I(i = j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

and the corresponding δ_{ijk} parameter set represents the factors specific to each set of stayers.

The log-likelihood function corresponding to the spatial interaction model in (12), where, for simplicity, δ_{ijk} is now referred to as δ_{iik}, is

$$\begin{aligned} l(\boldsymbol{\theta}; \mathbf{n}) &= \sum_{ijk} \{ n_{ijk} \log(\alpha_i \beta_j \lambda_k \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij}) - \alpha_i \beta_j \lambda_k \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij} - \log(n_{ijk}!) \} \\ &= \sum_i n_{i++} \log(\alpha_i) + \sum_j n_{+j+} \log(\beta_j) + \sum_k n_{++k} \log(\lambda_j) \\ &\quad + \sum_{ik} n_{i+k} \log(\gamma_{ik}) + \sum_{jk} n_{+jk} \log(\kappa_{jk}) + \sum_{ijk} n_{ijk} \log(\delta_{iik}) \\ &\quad - \sum_{ijk} \alpha_i \beta_j \lambda_k \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij} + c, \end{aligned} \quad (13)$$

where $\boldsymbol{\theta} = \{\alpha_i, \beta_j, \lambda_k, \gamma_{ik}, \kappa_{jk}, \delta_{iik}, i, j, k = 1, \ldots, R\}$, $\mathbf{n} = \{n_{ijk}, i, j, k = 1, \ldots, R\}$ and

$$c = \sum_{ijk} n_{ijk} \log(m_{ij}) - \sum_{ijk} \log(n_{ijk}!). \quad (14)$$

Differentiation of the likelihood function with respect to each parameter gives the likelihood equations:

$$\begin{aligned}
\frac{\partial l}{\partial \alpha_i} &= \frac{n_{i++}}{\alpha_i} - \sum_{jk} \beta_j \lambda_k \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij} = 0, \\
\frac{\partial l}{\partial \beta_j} &= \frac{n_{+j+}}{\beta_j} - \sum_{ik} \alpha_i \lambda_k \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij} = 0, \\
\frac{\partial l}{\partial \lambda_k} &= \frac{n_{++k}}{\lambda_k} - \sum_{ij} \alpha_i \beta_j \gamma_{ik} \kappa_{jk} \delta_{iik} m_{ij} = 0, \\
\frac{\partial l}{\partial \gamma_{ik}} &= \frac{n_{i+k}}{\gamma_{ik}} - \sum_{j} \alpha_i \beta_j \lambda_k \kappa_{jk} \delta_{iik} m_{ij} = 0, \\
\frac{\partial l}{\partial \kappa_{jk}} &= \frac{n_{+jk}}{\kappa_{jk}} - \sum_{i} \alpha_i \beta_j \lambda_k \gamma_{ik} \delta_{iik} m_{ij} = 0, \\
\frac{\partial l}{\partial \delta_{iik}} &= \frac{n_{ijk}}{\delta_{iik}} - \alpha_i \beta_j \lambda_k \kappa_{jk} \gamma_{ik} m_{ij} = 0,
\end{aligned} \tag{15}$$

which require only the marginal totals displayed in Table 3, (n_{i++}, n_{+j+}, n_{++k}, n_{i+k} and n_{+jk}) and the diagonal values (n_{ijk}, where $i = j$). The likelihood equations can be used to derive maximum likelihood estimators for $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_i, \hat{\beta}_j, \hat{\lambda}_k, \hat{\gamma}_{ik}, \hat{\kappa}_{jk}, \hat{\delta}_{iik})$;

$$\begin{aligned}
\hat{\alpha}_i &= \frac{n_{i++}}{\sum\limits_{jk} \hat{\beta}_j \hat{\lambda}_k \hat{\gamma}_{ik} \hat{\kappa}_{jk} \hat{\delta}_{iik} m_{ij}}, \\
\hat{\beta}_j &= \frac{n_{+j+}}{\sum\limits_{ik} \hat{\alpha}_i \hat{\lambda}_k \hat{\gamma}_{ik} \hat{\kappa}_{jk} \hat{\delta}_{iik} m_{ij}}, \\
\hat{\lambda}_k &= \frac{n_{++k}}{\sum\limits_{ij} \hat{\alpha}_i \hat{\beta}_j \hat{\gamma}_{ik} \hat{\kappa}_{jk} \hat{\delta}_{iik} m_{ij}}, \\
\hat{\gamma}_{ik} &= \frac{n_{i+k}}{\sum\limits_{j} \hat{\alpha}_i \hat{\beta}_j \hat{\lambda}_k \hat{\kappa}_{jk} \hat{\delta}_{iik} m_{ij}}, \\
\hat{\kappa}_{jk} &= \frac{n_{+jk}}{\sum\limits_{i} \hat{\alpha}_i \hat{\beta}_j \hat{\lambda}_k \hat{\gamma}_{ik} \hat{\delta}_{iik} m_{ij}}, \\
\hat{\delta}_{iik} &= \frac{n_{ijk}}{\hat{\alpha}_i \hat{\beta}_j \hat{\lambda}_k \hat{\gamma}_{iik} \hat{\kappa}_{jk} m_{ij}},
\end{aligned} \tag{16}$$

and can be solved by iteration using six steps for each parameter set:

$$
\begin{aligned}
\hat{\alpha}_i^{(1)} &= \frac{n_{i++}}{\sum\limits_{jk} \hat{\beta}_j^{(0)} \hat{\lambda}_k^{(0)} \hat{\gamma}_{ik}^{(0)} \hat{\kappa}_{jk}^{(0)} \hat{\delta}_{iik}^{(0)} m_{ij}}, \\
\hat{\beta}_j^{(2)} &= \frac{n_{+j+}}{\sum\limits_{ik} \hat{\alpha}_i^{(1)} \hat{\lambda}_k^{(0)} \hat{\gamma}_{ik}^{(0)} \hat{\kappa}_{jk}^{(0)} \hat{\delta}_{iik}^{(0)} m_{ij}}, \\
\hat{\lambda}_k^{(3)} &= \frac{n_{++k}}{\sum\limits_{ij} \hat{\alpha}_i^{(1)} \hat{\beta}_j^{(2)} \hat{\gamma}_{ik}^{(0)} \hat{\kappa}_{jk}^{(0)} \hat{\delta}_{iik}^{(0)} m_{ij}}, \\
\hat{\gamma}_{ik}^{(4)} &= \frac{n_{i+k}}{\sum\limits_{j} \hat{\alpha}_i^{(1)} \hat{\beta}_j^{(2)} \hat{\lambda}_k^{(3)} \hat{\kappa}_{jk}^{(0)} \hat{\delta}_{iik}^{(0)} m_{ij}}, \\
\hat{\kappa}_{jk}^{(5)} &= \frac{n_{+jk}}{\sum\limits_{i} \hat{\alpha}_i^{(1)} \hat{\beta}_j^{(2)} \hat{\lambda}_k^{(3)} \hat{\gamma}_{ik}^{(4)} \hat{\delta}_{iik}^{(0)} m_{ij}}, \\
\hat{\delta}_{iik}^{(6)} &= \frac{n_{ijk}}{\hat{\alpha}_i^{(1)} \hat{\beta}_j^{(2)} \hat{\lambda}_k^{(3)} \hat{\gamma}_{ik}^{(4)} \hat{\kappa}_{jk}^{(5)} m_{ij}}.
\end{aligned}
\tag{17}
$$

Once one cycle of estimation is complete, a new cycle commences using the last set of parameter estimates, $\hat{\alpha}_i^{(7)} = n_{i++} \Big/ \sum\limits_{jk} \hat{\beta}_j^{(2)} \hat{\lambda}_k^{(3)} \hat{\gamma}_{ik}^{(4)} \hat{\kappa}_{jk}^{(5)} \hat{\delta}_{iik}^{(6)} m_{ij}$, and so on. This is a conditional maximization of the likelihood function and converges to give estimates of all the parameters in $\boldsymbol{\theta}$. Note that the choice of initial values of $\beta_j^{(0)}, \lambda_k^{(0)}, \gamma_{ik}^{(0)}, \kappa_{jk}^{(0)}, \delta_{iik}^{(0)}$ in each of (17) implicitly specifies the constraint that is required for parameter identification.

Using the sufficient statistics of $\mathbf{n}$ shown in Table 3 to obtain the converged estimates of $\hat{\boldsymbol{\theta}}$ in the spatial interaction model (12), the maximum likelihood estimates of y_{ijk}, the expected number of migrant transitions can be derived. These values are shown in the top panel of Table 4. The iterative procedure to estimate $\hat{\boldsymbol{\theta}}$ and y_{ijk} is undertaken using the `ipf3.qi` routine in the `migest` R package (Abel 2012), where all elements of m_{ij} are set to unity ($m_{ij} = 1$). Summing over all birthplaces and deleting stayers in the diagonal elements gives us a traditional flow table of migrant transitions from origin i to destination j during the time period t to $t + 1$ shown in the bottom panel of Table 4. These migrant transitions are considerably smaller than those of Table 2. This is due to the extension in model (12) to incorporate information, via additional parameters for assumed diagonal elements, that the distribution of non-moving population and migrant transition flows between the t and $t + 1$ is no longer independent.

Table 4: Estimates of migrant transition flow tables based on stock data in Table 1, with known diagonals

Estimates of Origin Destination Place of Birth Flow Tables:

Place of Birth=A

		Destination				
		A	B	C	D	Sum
Origin	A	950	0	50	0	1000
	B	0	100	0	0	100
	C	0	0	10	0	10
	D	0	0	0	0	0
	Sum	950	100	60	0	1110

Place of Birth=B

		Destination				
		A	B	C	D	Sum
Origin	A	55	0	0	0	55
	B	25	505	25	0	555
	C	0	0	50	0	50
	D	0	0	0	5	5
	Sum	80	505	75	5	665

Place of Birth=C

		Destination				
		A	B	C	D	Sum
Origin	A	80	0	0	0	80
	B	10	30	0	0	40
	C	0	0	800	0	800
	D	0	0	0	40	40
	Sum	90	30	800	40	960

Place of Birth=D

		Destination				
		A	B	C	D	Sum
Origin	A	20	0	0	0	20
	B	0	25	0	0	25
	C	10	10	0	0	20
	D	10	10	0	180	200
	Sum	40	45	0	180	265

Estimates of Total Origin Destination Flow Table:

		Destination				
		A	B	C	D	Sum
Origin	A		0	50	0	50
	B	35		25	0	60
	C	10	10		0	20
	D	10	10	0		20
	Sum	55	20	75	0	150

2.2 Extensions to include births, deaths, and flows to and from outside regions

In reality, natural changes from births and deaths in the population occur, causing differences in the row totals of the stock data between subsequent years. In addition, migrants can move to or from regions outside those under consideration. We will consider each of these three sources of population change in this sub-section using a new set of hypothetical data for $t+1$, displayed in Table 5, where differences between row totals in the two stock tables now exist in comparison to the data in Table 1. In each case, population stocks that form the margins of the birthplace specific flow tables must be adjusted to enable the row and column margins to equal. This is carried out through a four step procedure.

Table 5: Dummy example of place of birth data

		Place of Residence (t)							**Place of Residence** $(t+1)$				
		A	B	C	D	Sum			A	B	C	D	Sum
Place of Birth	A	1000	100	10	0	1110	**Place of Birth**	A	1060	60	10	10	1140
	B	55	555	50	5	665		B	45	540	40	0	625
	C	80	40	800	40	960		C	70	75	770	70	985
	D	20	25	20	200	265		D	30	30	20	230	310
	Sum	1155	720	880	245	3000		Sum	1205	705	840	310	3060

First, alterations to the stock tables to account for sources of natural population change must be made. In order to avoid estimating migrant transitions to meet decreases in stock totals from mortality, the number of deaths in the time interval t to $t+1$ is subtracted from the reported stock data at time t. Typically only the total number of deaths in each place of residence is known, while a decomposition of the numbers of deaths by place of birth is missing. To estimate this breakdown, and hence adjust each native and foreign born population stocks, the total number of deaths is proportionally allocated out to each population stock. This is illustrated on the left hand side of Step 1 in Table 6, where the total number of deaths, given in bold type face in the final sum row, is known. These totals are proportionally split according to the reported population stocks in time t, to provide estimates of the number of deaths by each place of birth.

Table 6: Multi-step correction to stock data

Step 1: Control for Natural Changes

Place of Birth	Place of Death (t)			
	A	B	C	D
A	60.6	4.2	0.6	0.0
B	3.3	23.1	2.8	0.2
C	4.9	1.7	45.5	1.6
D	1.2	1.0	1.1	8.2
Sum	**70**	**30**	**50**	**10**

Place of Birth	Place of Residence $(t+1)$			
	A	B	C	D
A	80	0	0	0
B	0	20	0	0
C	0	0	40	0
D	0	0	0	60
Sum	**80**	**20**	**40**	**60**

Step 2: Estimated Altered Stocks

Place of Birth	Place of Residence (t)				
	A	B	C	D	Sum
A	939.4	95.8	9.4	0.0	1044.7
B	51.7	531.9	47.2	4.8	635.5
C	75.2	38.3	754.6	38.4	906.4
D	18.8	24.0	18.9	191.8	253.4
Sum	1085.0	640.0	830.0	235.0	2840.0

Place of Birth	Place of Residence $(t+1)$				
	A	B	C	D	Sum
A	980	60	10	10	1060
B	45	520	40	0	605
C	70	75	730	70	945
D	30	30	20	170	250
Sum	1125	685	800	250	2860

Step 3: Estimate Flows From (left) and To (right) External Regions

Place of Birth	Future Place of Residence				
	A	B	C	D	Sum
A	0.0	0.0	0.0	0.0	0.0
B	2.5	25.5	2.3	0.2	30.5
C	0.0	0.0	0.0	0.0	0.0
D	0.3	0.3	0.3	2.6	3.4
Sum	2.7	25.8	2.5	2.8	33.9

Place of Birth	Previous Place of Residence				
	A	B	C	D	Sum
A	14.2	0.9	0.1	0.0	15.3
B	0.0	0.0	0.0	0.0	0.0
C	2.9	3.1	29.8	2.9	38.6
D	0.0	0.0	0.0	0.0	0.0
Sum	17.0	3.9	30.0	3.0	53.9

Step 4: Re-estimated Altered Stocks

Place of Birth	Place of Residence (t)				
	A	B	C	D	Sum
A	939.4	95.8	9.4	0.0	1044.7
B	49.2	506.4	44.4	4.6	605.0
C	75.2	38.3	754.6	38.4	906.4
D	18.5	23.6	18.6	189.2	250.0
Sum	1082.3	664.2	827.5	232.2	2806.1

Place of Birth	Place of Residence $(t+1)$				
	A	B	C	D	Sum
A	965.8	59.1	9.9	9.9	1044.7
B	45.0	520.0	40.0	0.0	605.0
C	67.1	71.9	700.2	67.1	906.4
D	30.0	30.0	20.0	170.0	250.0
Sum	1108.0	681.1	770.0	247.0	2806.1

In order to avoid estimating migration flows to meet increases in native born totals from newborns, the number of births between t and $t+1$ is subtracted from the reported stock data at time $t+1$. As with deaths, we tend to only have information on the total number of births, where ideally more detail on the place of residence of newborns at time $t+1$ is desired. In order to adjust stock totals for natural increases, births are assumed to only affect the native born stocks, assuming there is no migration of newborns. This is illustrated on the right hand side of Step 1 in Table 6, where the total number of births, given in bold type face in the final sum row, is known. These totals of newborns are allocated to reside in their place of birth at time $t+1$.

A new set of adjusted stock tables that account for natural population change are shown in Step 2 of Table 6, where both the death and birth estimates of the previous step are subtracted cell-wise from the original data in Table 5. The new altered stock tables still do not have equal row totals. If the estimates (and assumptions) about the changes to population stocks from natural causes are true, the remaining differences between the row totals in the altered stock tables must represent the minimum amount of migrant transitions to or from outside external regions beyond A to D. When this difference is greater than zero, i.e. the row totals for time period $t+1$ are greater than t, migrants have arrived from external regions. When this difference is less than zero, i.e. the row totals for time period $t+1$ are smaller than t, migrants have moved away to external regions. Adjustments can be made for these differences in order to estimate migration solely within regions under consideration (A to D).

To illustrate the estimation of migrant transition flows to and from external regions, consider Step 2 of Table 6 where the adjusted stock of people originally born in region A were 1044.66 and 1060 in time t and $t+1$ respectively. Note, fractions of migrants are given only to fully illustrate the mathematics at work. As the difference is negative (-15.34), the stock of people born in region A has increased, after accounting for the natural population change. This difference provides us with information on the total number of people born in region A that moved from an external region between time t and $t+1$. These immigrants might reside in any place of residence at time $t+1$. In order to estimate this breakdown we proportionally allocate out the 15.34 external migrants to each destination region. This is illustrated on the right hand side of Step 3 in Table 6 for both regions A and C, where proportions are calculated according to the adjusted population stocks in time t (shown in Step 2 of Table 6). Conversely, when the difference in row totals of altered stocks are positive, such as for those in region B in Step 2 of Table 6, we have information on the total number of people born in region B that moved to an external region between time t and $t+1$. These emigrants might have left from any place of residence. The migrant transition flows are estimated by proportional allocation, shown on the left hand side of Step 3 in Table 6 for both region B and D. Proportions are calculated according to the adjusted population stocks in time $t+1$.

A new set of altered stock totals that adjust for migration flows to and from external regions are shown in Step 4 of Table 6. These are calculated by subtracting cell-wise the previously adjusted stock totals in Step 2 from the calculated flows to other regions in Step 3. The resulting stock tables control for both natural population change and moves to and from external regions during the time period t to $t+1$. In addition, they have matching row totals, required to estimate flows using the methodology outlined in the previous subsection.

The sufficient statistics of $\mathbf{n}$ shown in Stage 4 of Table 6 can be considered as a set of $R = 4$ birthplace specific flow tables, shown in the margins in the top panel of Table 7. Using these marginal data and the converged estimates of $\hat{\boldsymbol{\theta}}$ in the spatial interaction model (12) we obtain the maximum likelihood estimates of y_{ijk}, the expected number of migrant transitions, controlling for natural population changes and moves to and from external regions. These values are shown in the cells of the tables in the top panel of Table 7. The iterative procedure to estimate $\hat{\boldsymbol{\theta}}$ and y_{ijk}, controlling for flows to and from outside regions is undertaken using the `ffs` routine in the `migest` R package (Abel 2012). By default, in the `ffs` routine all elements of m_{ij} are set to unity ($m_{ij} = 1$) and the diagonal element are set to their maximum possible values given the known margins. Summing over all birthplaces and deleting stayers in the diagonal elements gives us a traditional flow table of migrant transitions from origin i to destination j during the time period t to $t+1$ shown in the bottom panel of Table 7. Estimates are not directly comparable with previous flow tables as they are formed from a different set of migrant stock data in $t+1$.

Table 7: Estimates of migrant transition flow tables based on stock data derived in Table 6, with known diagonals

Estimates of Origin Destination Place of Birth Flow Tables:

Place of Birth=A

		Destination				
		A	B	C	D	Sum
Origin	A	939.4	0.0	0.0	0.0	939.4
	B	26.4	59.1	0.4	9.9	95.8
	C	0.0	0.0	9.4	0.0	9.4
	D	0.0	0.0	0.0	0.0	0.0
	Sum	965.8	59.1	9.9	10.0	1044.7

Place of Birth=B

		Destination				
		A	B	C	D	Sum
Origin	A	45.0	4.2	0.0	0.0	49.2
	B	0.0	506.4	0.0	0.0	506.4
	C	0.0	4.9	40.0	0.0	44.9
	D	0.0	4.6	0.0	0.0	4.6
	Sum	45.0	520.0	40.0	0.0	605.0

Place of Birth=C

		Destination				
		A	B	C	D	Sum
Origin	A	67.1	4.3	0.0	3.7	75.2
	B	0.0	38.3	0.0	0.0	38.3
	C	0.0	29.3	700.2	25.1	754.5
	D	0.0	0.0	0.0	38.4	38.4
	Sum	67.1	71.9	700.2	67.0	906.4

Place of Birth=D

		Destination				
		A	B	C	D	Sum
Origin	A	18.5	0.0	0.0	0.0	19.0
	B	0.0	23.6	0.0	0.0	23.6
	C	0.0	0.0	18.6	0.0	18.6
	D	11.5	6.4	1.4	170.0	189.2
	Sum	30.0	30.0	20.0	170.0	250.0

Estimates of Total Origin Destination Flow Table:

		Destination				
		A	B	C	D	Sum
Origin	A		8.5	0.0	3.7	12.2
	B	26.4		0.4	9.9	36.7
	C	0.0	34.2		25.1	59.3
	D	11.5	10.9	1.4		23.8
	Sum	37.9	53.6	1.8	38.6	132.0

3. Results

In this section the application of the above methodology to global place of birth migrant stock tables produced by the World Bank (Özden et al. 2011) is outlined, alongside the additional data requirements. A brief overview of the results, focusing on the largest estimated flows is discussed below.

3.1 Application

Place of birth data published by the World Bank (Özden et al. 2011) provide foreign born migration stock tables at the start of each of the last five decades for 226 countries[2]. To date, this is the most complete set of comparable global data of past international migration stocks available. The data are primarily based on place of birth responses to census questions or details collected from population registers. In order to create a complete and comparable data set, the World Bank undertook a number of adjustment and imputation steps. What follows is a brief description; for full details the reader is referred to Özden et al. (2011). For some nations, where there is no place of birth data available, data on citizenship are taken by the World Bank with the belief that they are a broadly equivalent measure of migrant stock populations. In other countries where neither place of birth or citizenship stock measure was available, missing values were addressed using various propensity and interpolation methods. These were either based on historical or future data when a measure in a specific period was missing, or using available data from countries in the same region when data in all periods were missing. Changes in geography, from countries unifying or partitioning were also accounted on a country by country basis using various imputation measures depending on available related data. For example, historical stocks by place of birth in former USSR countries were not collected in the 1960 to 1980 census rounds, however questions were asked on ethnicity. In the 1989 census both stocks by ethnicity and place of birth data were reported, allowing a proxy measure of past place of birth stocks to be derived.

Of the 226 countries for which stock data was available, 191 also had the demographic data from the United Nations Population Division (2011)[3] throughout the time periods, as required for estimating flow methodology outlined in the previous section. None of the dropped countries had populations in in excess of 100,000 people 2010. Diagonal elements in the stock tables, of the native-born population totals in each place of residence j, (P_j^{NB}), are not provided in the World Bank data. These were derived as a remainder ($P_j^{NB} = P_j - \sum_i P_j^{FB}$) using annual population totals from the United Nations Population Division (2011), (P_j) and the column sums of the foreign born populations in each

---

[2] Data available from http://data.worldbank.org/data-catalog/global-bilateral-migration-database .

[3] Data available from http://esa.un.org/unpd/wpp.

place of residence ($\sum_i P_j^{FB}$). This procedure constrained the column totals of the stock tables to meet those of the reported populations at the start of each decade.

Demographic data on the number of births and deaths in each country, required in the multi-step estimation shown in Table 6, were also taken from United Nations Population Division (2011). Auxiliary data for use in the offset term of the estimation procedure were taken from the Centre d'Etudes Prospective et d'Informations Internationales (CEPII) data base on geographic distance (Mayer and Zignago 2012), which provides a distance measure between all capital cities. The offset term was then calculated as $m_{ij} = d_{ij}^{-1}$. The multi-step accounting method was undertaken to adjust reported stock totals for births, deaths, and flows to external countries beyond the 191 considered. The conditional maximisation routine was then run to calculate the migration flow tables for each decade between the five sets of migration stock tables. Both of these process were undertaken within the `ffs` routine in the `migest` R package (Abel 2012).

3.2 Summary of estimates

The application of the estimation procedure resulted in four 191 × 191 flow tables (see the `flow.xlsx` file in the supplementary material). Compressed tables, aggregated over World Bank regions (used in Özden et al. 2011) are shown in Table 8 to Table 11. Non-diagonal elements in these tables represent rounded thousands of estimated 10-year international migrant transition flows between regions. Diagonal elements represent estimated 10-year international migrant transition flows within regions. As per standard migration flow tables, rows represent origins and columns destinations.

Table 8: **1960s estimated global migrant transition flow (in 1000’s)**
Table aggregated over 12 World Bank regions

	AFR	ANZ	CAN	EAP	EECA	GFNA	JPN	LAC	MENA	SA	USA	WE	Sum
AFR	2392	30	39	10	4	22	0	13	7	76	33	650	3277
ANZ	3	31	0	24	7	0	0	1	0	0	11	29	107
CAN	3	9	0	1	16	1	0	9	0	0	43	49	132
EAP	30	52	119	2152	16	26	84	13	4	132	313	79	3018
EECA	31	97	45	7	6729	143	0	13	39	2	50	1485	8642
GFNA	5	3	5	0	4	45	0	5	124	3	22	6	222
JPN	1	2	1	49	1	0	0	9	0	0	50	4	118
LAC	19	84	165	11	18	9	4	553	2	1	1665	517	3048
MENA	158	30	41	8	42	516	0	28	390	22	45	1535	2813
SA	105	15	61	193	5	232	2	2	70	1896	61	223	2864
USA	17	43	59	19	123	17	8	186	8	1	0	423	904
WE	300	638	241	143	266	78	5	148	27	22	309	4962	7137
Sum	3064	1032	774	2617	7231	1089	104	978	672	2156	2603	9962	32283

Notes: AFR = Africa, ANZ = Australia and New Zealand, CAN= Canada, EAP = East Asia and the Pacific, EECA = Eastern Europe and Central Asia, GFNA = High Income Middle East and North Africa, JPN = Japan, LAC = Latin America and the Caribbean, MENA = Rest of Middle East and North Africa, SA = South Asia, USA = United States, WE = Western Europe. Rows represent origin regions, columns represent destination regions.

Table 9: 1970s estimated global migrant transition flow (in 1000's) Table aggregated over 12 World Bank regions

	AFR	ANZ	CAN	EAP	EECA	GFNA	JPN	LAC	MENA	SA	USA	WE	Sum
AFR	2882	53	40	8	5	148	0	17	11	13	126	742	4046
ANZ	9	122	7	6	17	3	0	3	0	1	30	53	254
CAN	12	14	0	2	17	5	1	5	1	1	117	69	244
EAP	14	186	154	2786	13	251	120	33	4	103	1445	458	5566
EECA	15	59	40	4	4734	159	0	4	38	1	44	2278	7376
GFNA	16	3	11	3	39	100	0	2	165	14	80	93	527
JPN	0	3	1	10	0	0	0	10	1	1	156	34	216
LAC	38	50	172	6	28	10	6	1019	3	0	3109	800	5241
MENA	80	53	44	6	22	1252	0	12	577	8	273	996	3322
SA	33	36	69	40	4	1194	2	3	246	8020	200	345	10193
USA	57	25	142	6	216	42	2	177	11	3	0	498	1178
WE	218	210	169	30	226	65	4	57	16	4	568	2347	3915
Sum	3375	813	847	2908	5321	3230	137	1343	1074	8168	6148	8712	42077

Notes: AFR = Africa, ANZ = Australia and New Zealand, CAN= Canada, EAP = East Asia and the Pacific, EECA = Eastern Europe and Central Asia, GFNA = High Income Middle East and North Africa, JPN = Japan, LAC = Latin America and the Caribbean, MENA = Rest of Middle East and North Africa, SA = South Asia, USA = United States, WE = Western Europe. Rows represent origin regions, columns represent destination regions.

Table 10: 1980s estimated global migrant transition flow (in 1000's) Table aggregated over 12 World Bank regions

	AFR	ANZ	CAN	EAP	EECA	GFNA	JPN	LAC	MENA	SA	USA	WE	Sum
AFR	2636	79	70	14	19	269	6	24	24	34	173	643	3990
ANZ	1	100	24	6	14	4	4	3	1	10	15	67	249
CAN	4	7	0	2	20	9	2	5	3	5	217	57	330
EAP	7	421	226	1264	8	519	182	32	10	93	2143	500	5406
EECA	5	34	58	11	8040	455	1	7	76	11	58	2980	11737
GFNA	5	11	19	4	61	244	1	5	309	30	79	148	916
JPN	0	10	6	15	1	1	0	1	0	0	105	35	173
LAC	20	60	340	13	37	34	56	826	4	3	5055	692	7142
MENA	23	47	38	6	46	1597	1	17	668	13	211	865	3531
SA	23	56	70	96	11	2201	12	4	638	860	383	394	4749
USA	12	18	141	33	94	42	19	97	10	15	0	230	711
WE	120	181	283	47	843	116	13	153	38	57	269	1450	3571
Sum	2858	1025	1275	1509	9192	5492	298	1174	1782	1131	8710	8061	42505

Notes: AFR = Africa, ANZ = Australia and New Zealand, CAN= Canada, EAP = East Asia and the Pacific, EECA = Eastern Europe and Central Asia, GFNA = High Income Middle East and North Africa, JPN = Japan, LAC = Latin America and the Caribbean, MENA = Rest of Middle East and North Africa, SA = South Asia, USA = United States, WE = Western Europe. Rows represent origin regions, columns represent destination regions.

Table 11: 1990s estimated global migrant transition flow (in 1000's) Table aggregated over 12 World Bank regions

	AFR	ANZ	CAN	EAP	EECA	GFNA	JPN	LAC	MENA	SA	USA	WE	Sum
AFR	3941	84	117	19	21	140	5	10	23	20	391	875	5646
ANZ	6	121	10	60	16	7	6	4	1	0	41	145	417
CAN	21	28	0	38	28	21	9	39	5	2	246	236	674
EAP	41	427	678	2818	55	134	394	51	50	8	2096	687	7438
EECA	10	54	90	77	6726	621	5	18	90	1	826	5057	13575
GFNA	8	12	41	28	31	229	0	2	413	63	70	99	997
JPN	1	15	13	43	2	0	0	2	2	0	86	57	221
LAC	34	22	240	42	107	43	240	1043	7	0	8064	1584	11426
MENA	27	46	136	29	74	571	6	10	704	82	267	584	2536
SA	58	106	344	367	13	1136	17	6	70	715	850	891	4572
USA	23	20	13	56	39	51	11	213	27	2	0	271	727
WE	235	40	131	63	719	122	13	86	122	7	410	3928	5876
Sum	4403	977	1811	3639	7832	3076	707	1484	1515	900	13347	14413	54105

Notes: AFR = Africa, ANZ = Australia and New Zealand, CAN= Canada, EAP = East Asia and the Pacific, EECA = Eastern Europe and Central Asia, GFNA = High Income Middle East and North Africa, JPN = Japan, LAC = Latin America and the Caribbean, MENA = Rest of Middle East and North Africa, SA = South Asia, USA = United States, WE = Western Europe. Rows represent origin regions, columns represent destination regions.

Estimates of the total volume of migrant transition flows increase over time, reflecting global population growth. However, in both the 1970s and 1980s estimated flows between all countries were steady at around 42 million. The most popular origins of flows over the entire time period, by estimated volume, were in Eastern Europe and Central Asia (EECA) and, in the later tables, Latin America and the Caribbean (LAC). Most destinations of those in the EECA region tended to be to other countries also within the EECA region, whereas the flows out of LAC tended to go to other regions, most notably the USA. Large migrant transition flows are estimated from Western Europe (WE) and South Asia (SA) regions in the 1960s and 1970s tables respectively. The majority of these large flows are to other nations in the same region. The most popular destinations of flows over the entire time period, by estimated volume, were into the USA and WE. The largest sending regions into the USA were East Asia and the Pacific (EAP) and LAC. The largest flows into WE came from the east (EECA) and other WE nations.

Results can also be studied on a country by country basis from the full migrant transition flow tables. Table 12 shows the largest estimated inflows (top) and outflows (bottom) which are taken from the flow table column and row totals respectively. The USA received the highest estimated migrant transition inflows of any country in almost every decade (top of Table 12). Germany was also a large receiver of estimated inflows throughout each decade studied, whilst the United Kingdom and France were also top receivers in the 1990s. France received the largest inflow of all nations in the 1960s. Large migrant inflows to France during this decade were estimated to originate from Algeria, Italy, Spain and Portugal. Russia and the Ukraine also received large numbers of estimated immigrants in all decades with exception of the 1970s. India and Pakistan both received large estimated inflows during the 1970s, as did Saudi Arabia in the 1980s.

Russia was a top origin for estimated migrant transition outflows in each of the first three decades (bottom of Table 12), as was the Ukraine and Kazakhstan in the 1980s and 1990s respectively. The majority of these outflows are to other former USSR countries. For example, the large outflows from Kazakhstan in the 1990s are estimated to meet a large increase in the Kazakhstan-born stocks in Russia (1.8 to 2.5 million). The rise in the number of Kazakhstan-born in Russia alongside large flows from other former USSR countries such as Uzbekistan, the Ukraine and Azerbaijan created a large inflow in the 1990s into Russia. As illustrated in the methodology section, these estimates are ultimately based on changes in the stock data, which in this case originate from World Bank imputation rather than raw data from censuses or registers. Other nations in the former USSR and Indian sub-continent also have large estimated migration flows from big differences in stock imputations over time.

Mexico was a predominant sender in the later decades, with large amounts of estimated flows into the USA. France appears as large sender of estimated outflows in both the 1960s and 1990s. In the 1960s the outflow was driven predominately by migrants born in Poland (their stock falls from 474,000 to 280,000 between the start and end of the decade). In the 1990s the outflow is driven by migrants from Portugal and Spain returning to their place of birth (their stock falls from 609,000 and 429,000 to 141,000 and 151,000 respectively between the start and end of the decade). Germany has a large outflow of migrants in the 1980s. The largest estimated flow from Germany in this decade is to the Czech Republic, to match a large fall in the number of Czech born residents (from around 750,000 in 1980 to 375,000 in 1990). World Bank stock data for Germany, as with the former USSR nations were imputed, as no place of birth data was available. The large outflows from Turkey in the 1970s were also influenced by the stock data in Germany, where the number of Turkish born rose from around 440,000 in 1970 to 1.65 million in 1980. The large outflows from China in the 1990s were to the USA and Hong Kong. The appearance of China and Hong Kong as large senders in earlier decades might be due to peculiarities in the stock data discussed below.

Table 12: Countries with the five largest estimated 10-year migrant transition inflows (top) and outflows (bottom)

Country	1960s	Country	1970s	Country	1980s	Country	1990s
FRA	3,151,489	USA	6,148,134	USA	8,709,721	USA	13,346,816
USA	2,602,511	IND	4,975,291	RUS	3,356,465	RUS	5,082,317
DEU	2,531,999	PAK	2,893,128	SAU	3,065,083	DEU	3,738,591
UKR	1,749,508	DEU	2,708,104	DEU	2,215,946	FRA	2,873,948
RUS	1,741,571	HKG	1,984,883	UKR	1,477,474	GBR	1,819,267

Country	1960s	Country	1970s	Country	1980s	Country	1990s
RUS	3,086,746	BGD	4,800,661	RUS	2,655,499	MEX	5,037,273
HKG	1,500,636	IND	4,066,949	MEX	2,446,505	IND	2,594,894
PAK	1,440,894	RUS	2,117,101	IND	2,403,785	FRA	2,586,883
ITA	1,285,555	TUR	1,774,873	UKR	1,507,850	CHN	1,643,347
FRA	1,243,313	CHN	1,693,328	DEU	1,354,946	KAZ	1,605,811

Notes: FRA = France, DEU = Germany, UKR= Ukraine, RUS = Russia, IND = India, PAK = Pakistan, HKG = Hong Kong, SAU = Saudi Arabia, GBR = Great Britain, ITA = Italy, BGD = Bangladesh, TUR = Turkey, CHN = China, MEX = Mexico, KAZ = Kazakhstan.

Estimated values can be further explored by studying the place of birth dimension, alongside the origin and destination of migrant transition flows. For example, considering the USA as a destination, the largest estimated inflows by origin and place of birth in each decade are shown Table 13.

Table 13: Countries with the five largest estimated 10-year migrant transition inflows to the USA by origin and place of birth

Origin	Destination	Place of Birth	1960s	Origin	Destination	Place of Birth	1970s
CUB	USA	CUB	439,346	MEX	USA	MEX	1,556,421
MEX	USA	MEX	402,118	PHL	USA	PHL	361,425
PRI	USA	PRI	343,996	KOR	USA	KOR	231,073
PHL	USA	PHL	142,902	PRI	USA	PRI	212,880
HKG	USA	CHN	85,868	CUB	USA	CUB	208,657

Origin	Destination	Place of Birth	1980s	Origin	Destination	Place of Birth	1990s
MEX	USA	MEX	2,399,422	MEX	USA	MEX	4,910,358
PHL	USA	PHL	495,993	PHL	USA	PHL	54,1725
SLV	USA	SLV	407,537	IND	USA	IND	500,653
KOR	USA	KOR	383,622	VNM	USA	VNM	437,534
VNM	USA	VNM	329,061	CHN	USA	CHN	411,763

Notes: CUB = Cuba, MEX = Mexico, PRI = Puerto Rico, HKG = Hong Kong, PHL = Philippines, KOR= South Korea, SLV = Slovenia, VNM = Vietnam, IND = India, CHN = China.

Estimates show that throughout the four time periods large flows originate from countries in Latin America (Mexico, Cuba, and Puerto Rico) and Asia (Philippines, Korea, Vietnam, India and China). In addition, we can garner further information on place of birth of these estimated migrant flows, which can also hint at questionable changes in stock data over time. For example, during the 1960s there was a large flow into the USA from Hong Kong of people born in China. This flow is partly a result of a rise in the Chinese born in the USA stock data between 1960 and 1970 (from around 100,000 to 220,000) and partly from a peculiarity of the Hong Kong data. In 1960 there were a reported 1.5 million Chinese born in Hong Kong. This stock drops to 16,823 in 1970 and rises back up to almost 1.9 million in 1980. This dramatic movement in the reported stocks creates a large estimated outflow of Chinese in the 1960s. These emigrants are moving to countries where there are increases in the number of Chinese born, including but not exclusively, China. In turn, during the 1970s there is a large inflow back into Hong Kong of Chinese born, to meet the sudden increase in their migrant stock.

4. Validation of results

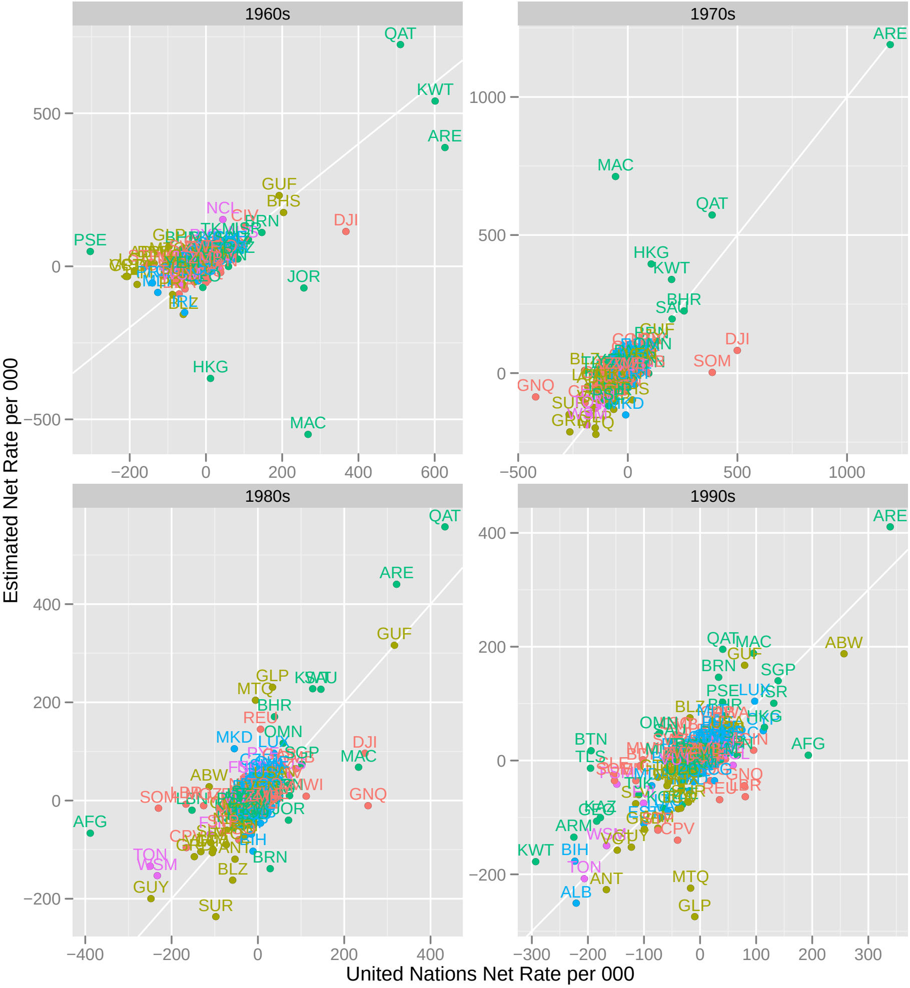
Due to the large size of the estimated migrant flow tables, some form of dimension reduction is required in order to asses estimates beyond looking at compressed tables of regional flows and the top inflows and outflows. In this section, the estimated migration transition flows from the applied methodology are discussed with reference to comparable estimates of global net migration flows of the United Nations and models for international migration.

4.1 Net migration comparison

The United Nations Population Division (2011) published net migration flows for each member country over a five-year period. These data are based predominately on scaled annual flows, derived from either migration records or through demographic accounting. In order to get a 10-year net migration flow, to correspond to the estimated 10-year migrant transitions, the two 5-year net migration flow totals for each decade (one for the first half of a decade and one for the second half) were summed for each country. A net rate was then calculated using the mid-decade population totals and multiplying by 1000. Comparative net migration rates from the estimates were calculated by taking the total inflow away from total outflow in each country, divided by the mid-decade population totals and multiplying by 1000. A scatter plot comparing the estimated net rates (on the y-axis) with the derived United Nations net rate (x-axis) for each decade is shown in Figure 1.

There is a noticeable general linear trend along the $x = y$ line, indicating a broad conformity of the estimated with the derived UN net rates. This relationship is confirmed in separate regressions for each decade of the estimated rate on the derived United Nations rate and an intercept, shown in Table 14. Parameters are estimated in R using the `rlm` function from the MASS package (Venables and Ripley 2002) to account for outliers (discussed below).

Figure 1: **Scatter Plot of Estimated Net 10-year Migrant Transition Flow Rates vs. Derived UN Rates (per 000). Countries labelled according to their ISO 3166-1 alpha-3 code (ISO, 2006)**



Intercept values are within two standard deviations of zero, suggesting there is no constant effect of the United Nations net rates being significantly different than the estimated net rates in any of the four time periods. The slope coefficients are all positive, but less than unity, indicating the estimated net rates values are lower than the derived United Nations net rate. This result is not unexpected for three reasons. First, the United Nations net figures are based on timing definitions for migration during an interval that is less than 10 years. Subsequently more migrant transitions are recorded from moves over the periods of time. This difference is compounded when a 10-year net was derived. Second, the estimated 10-year migration tables represent transitions and not movements. Hence, multiple movements of a migrant might potentially be recorded by the United Nations over a five-year period, where only one transition comparing the location of the migrant at the start and end of the period is considered in the net rate based on the estimated flow tables. Third, the estimated data from the flow from stocks methodology are the minimal migration flow transitions in each decade required to meet the World Bank stock data. This minimum was derived by setting the diagonal values in each place of the birth table to their maximum value as illustrated in the methodology section.

Table 14: **Regression of derived United Nations (UN) net rates on estimated net rates**

Parameter (Std. Error)	1960s	1970s	1980s	1990s
Intercept	2.19803	-0.33207	-1.05609	-1.52908
	(1.76241)	(1.92035)	(1.90238)	(2.27249)
UN Net Rate	0.53494	0.74672	0.64345	0.66924
	(0.01721)	(0.01508)	(0.02244)	(0.02919)

Outliers with high positive net migration are captured by both the estimated and the United Nations rate, such as the United Arab Emirates (ARE), Kuwait (KWT) and Quatar (QAT). Other outliers such as Hong Kong (HKG) and Macau (MAC) show a disagreement between the United Nations rate and the rate from the estimated bilateral flow table. As discussed previously, these estimated bilateral flows are subject to peculiarities in the World Bank stock data. Similar data peculiarities also occurred over time for the stock of people born in Martinique (MTQ) and Guadalupe (GLP) and then residing in France. Countries such as Jordan (JOR), Djibouti (DJI) and Somalia (SOM) have large derived positive United Nations net migration rates in either the 1960s, 1970s or 1980s, where the net rate from the estimated bilateral flow table is generally closer to zero, having not accounted for the short term refugee movements between neighbouring countries.

4.2 Gravity model

Kim and Cohen (2010) investigated non-economic predictors of reported international migration bilateral flows. Using a log-normal regression model, geographic, demographic and social and historical determinants were estimated twice; once using migration flow data by origin into 17 destinations countries and once using migration flow data by destinations from 13 origin countries. Data were taken from the United Nations Population Division (2009), which published a set of unharmonised time series of sending and receiving flow statistics reported by developed nations. Data for covariates were obtained from United Nations Population Division (2011) for demographic measures on population, potential support ratio (PSR), infant mortality rate (IMR), and urbanisation, and the CEPII for geographic and social data on distances, language, and colonial relationships (Mayer and Zignago 2012).

In this section, the same initial model used in Kim and Cohen (2010) is fitted (once) to the logarithm of estimated bilateral flows between the 191 origin and destination countries. Following the same procedures, estimated migration flows with value zero were excluded and logarithms with base 10 were used throughout. In addition, mid-decade values were taken for time varying covariate measures. The estimated parameters and standard errors are shown in Table 15 alongside the parameters obtained by Kim and Cohen (2010) when the model was fitted to the receiving data (K&C Receiving) and sending data (K&C Sending).

Standard regression diagnostics suggested there were no major problems with the model assumptions. What follows is a brief discussion of the parameter estimates, and comparisons with the results found by the two model fits of Kim and Cohen (2010). For further clarification on the expected direction and justification of variable constructions and selection, the reader is referred to Kim and Cohen (2010).

The constant parameter value from the fit on the estimated data is considerably higher than that found in Kim and Cohen (2010). This reflects the greater average size of the estimated 10-year migration transition flows compared to the collection of reported flow data from the United Nations, which for example has no information on flows within Asia, Africa, Latin America, much of Eastern Europe, in the Indian subcontinent, and the former USSR. The population parameters for both the origin and destination countries are greater than zero, indicating that as the population in an origin or destination increases, the expected logarithm of the estimated migration flow also increases, when all other variables are held constant. These parameters are both smaller in size than those found from the two models fitted to the reported flow data. Parameter estimates for both the effect of the potential support ratio (PSR) and infant mortality rate (IMR) are all negative, indicating that as they increase in value, the expected migration flows decrease. For the PSR parameters this result implies that for countries with lower working age populations (as a fraction of the population totals) there are higher expected migration flows, when all other parameters are held constant. For the destination parameter, this result reflects a high number of estimated migrations from developing to developed countries. This result is of lesser importance for migration from origins where the parameter value is relatively close to zero when the standard error is considered. For destinations, similar results were found in both models based on the receiving and sending data in Kim and Cohen (2010). For the IMR, negative parameters suggest that at higher rates of infant mortality in either origin or destination the number of expected migration flows is lower. For destinations this effect is large, indicating that a destination with a high IMR is expected to have fewer immigrants, having controlled for all other variables. For origins this effect is relatively close to zero, again indicating slightly counter-intuitively that an origin with a high IMR is expected to have fewer emigrants. This can be explained by the inclusion of migration from the poorest countries, such as those in parts of Africa where there are low levels of education, migrant networks, and relatively high costs to migration, leading to very low numbers of movements. The percentages of urban population in the destination and origin increased the expected migration flows in and out of countries respectively, as expected.

For geographic determinants, both distance and an indicator variable for border sharing had large effects on the expected migration flow, especially in comparison with the Kim and Cohen (2010) parameters. Unsurprisingly the distance parameter is very close to negative one, as the same information was used in the offset term of (12) for the estimation of stocks from flows. Other geographic parameters on land area and landlocked follow the intuitive results found by Kim and Cohen (2010); larger countries send and receive more expected migration flows, while landlocked countries have lower expected migration flows due to higher transportation costs.

All social and historical dichotomous variables have positive values. Hence, where there is a shared language, shared minority languages, or a colonial link between origin and destination, the model parameters indicate there was an increase in the expected migration flow. Finally, the linear and quadratic time parameters are relatively close to zero, indicating only small effects of increases in migration when other time varying covariates are controlled for in the model.

Table 15: Estimated parameters of M1 in Kim and Cohen (2010) in a regression on the logarithm of various migration flow counts.

Parameter	**Estimate**	**(Std. Error)**	**K&C Receiving**	**(Std. Error)**	**K&C Sending**	**(Std. Error)**
Constant	2.271	(0.141)	-9.960	(0.231)	-12.408	(0.258)
Demographic determinants						
Log population (destination)	0.206	(0.007)	0.601	(0.009)	0.372	(0.008)
Log population (origin)	0.423	(0.007)	0.728	(0.006)	0.936	(0.011)
Log potential support ratio (destination)	-0.166	(0.055)	-0.811	(0.069)	-0.052	(0.024)
Log potential support ratio (origin)	-0.027	(0.054)	0.045	(0.020)	0.915	(0.079)
Log infant mortality rate (destination)	-0.704	(0.017)	1.007	(0.049)	-0.783	(0.016)
Log infant mortality rate (origin)	-0.083	(0.017)	-0.466	(0.013)	0.359	(0.054)
Log percentage of urban population (destination)	0.239	(0.018)	3.057	(0.072)	0.307	(0.021)
Log percentage of urban population (origin)	0.163	(0.017)	0.332	(0.017)	2.578	(0.077)
Geographical determinants						
Log distance between capitals	-1.026	(0.010)	-0.819	(0.011)	-0.660	(0.012)
Log land area (destination)	0.210	(0.005)	0.234	(0.008)	0.146	(0.007)
Log land area (origin)	0.047	(0.006)	-0.047	(0.005)	0.030	(0.009)
Landlocked (destination)	-0.073	(0.010)	-0.610	(0.040)	-0.086	(0.011)
Landlocked (origin)	-0.104	(0.010)	-0.170	(0.009)	-1.043	(0.038)
Border	0.737	(0.023)	0.077	(0.022)	0.096	(0.024)
Social and historical determinants						
Common official language	0.264	(0.015)	0.138	(0.014)	0.346	(0.027)
9% minority speak same language	0.325	(0.015)	0.266	(0.014)	0.003	(0.027)
Colony	0.686	(0.026)	0.427	(0.017)	0.747	(0.023)
Year - 1985	-0.013	(0.001)	0.008	(0.001)	0.001	(0.001)
(Year - 1985)^2	0.000	(0.000)	0.000	(0.000)	-0.000	(0.000)

5. Summary and discussion

In this paper, a methodology to estimate global migrant transition flow tables is illustrated in two parts. First, a methodology to derive flows from sequential stock tables was outlined. Second, the methodology was applied to recently released World Bank migration stock tables to estimate a set of four decadal global flow tables. The results of the applied methodology were discussed with reference to comparable estimates of global net migration flows of the United Nations and previous models for international migration flows.

The methodology outlined adds to the limited exisitng literature on linking migration flows to stocks. Rogers and von Rabenau (1971), Rogers and Raymer (2005) and Rogers and Liu (2005) focused on US Census place of birth data to estimate inter-state migration transitions. The first of these papers relies on a simplistic principal assumption of equal growth of stock totals in all regions. As a result, the application of the method outlined to the World Bank data produced many strange results, including negative migration flows. The second paper relies on aggregated stock populations by their place of birth, place of residence at time $t-1$, and place of residence at time t. This added dimension of required data negates its application to estimating international migration flows. This point is also valid for the third paper, which relies on information of migrant stocks disaggregated by their place of residence at time $t-1$, place of residence at time t, and age group.

The method for estimating migrant transition flows from stocks demonstrated in this paper constrains results to match those of known sequential stock tables. This approach is beneficial for a number of reasons. First, comparable migration stock data are easier to measure and collect at a global level than flow data. Difficulties in producing comparable international migration flow data have recently lead to a number of competing estimation efforts, all applied to European data, see for example Abel (2010); Beer et al. (2010); Raymer et al. (2012). However, these methodologies rely on a reasonable percentage of double counted flows, i.e., reported values of movements from both the sending and receiving counties. The application of these estimation methods to obtain global migration flow data is hindered by the availability of reported migration flows, which from non-European countries remains scarce. Second, as discovered in the results section of this paper, estimating flows from stocks can provide a good check on the stock data itself. The flow estimation method is partly reliant on the differential change in the stock data. When this change is large, large flows are estimated. Where these flow estimates are unexpected, peculiarities in the data can be quickly identified. Third, estimating flows required to match stock totals results in migrant transitions rather than movement events. Transition measures can be of particular use in global population projection models, such as those outlined by Cohen et al. (2008). Projection models tend to require migration transitions that match the age groups at which the future population is predicted. For example, the results presented in this paper could potentially be incorporated into projection models that age populations at 10-year intervals. If denser intervals are desired, such as five-year or single-year age intervals, new estimates would be required. These could be derived in one of two ways. Stock tables in non-census years could be estimated to create a larger set of sequential stock tables (within the same time frame), allowing flows between closer time periods to be obtained. Alternatively, one could adapt a method such as that outlined by Courgeau (1973) to convert 10-year estimates to shorter periods. If such estimates were derived, they could potentially be compared with flow data from countries that have a Census question on the place of residence one or five years prior to their Census night, forming another conceivable validation exercise.

The estimated flow tables are a first-of-a-kind set of comparable global origin destination flow data. Estimates represent the minimum number of migration transition flows required to meet the foreign born stocks. As discussed in the introduction, bilateral flow tables have a number of advantages over currently available international migration data. Stock data are collected in the country of residence via census questions or details collected from population registers which capture a single transition between birthplace and country of current residence. However, the methodology outlined allows flows to be estimated on three dimensions; origin, destination and place of birth. Consequently, multiple typologies of movements such as primary migration, return migration or onward migration flows of non-natives to a third country are estimated. Results from both validation exercises appeared promising. The validation exercises allowed, as did the brief overview of the main results, the identification of unexpected flows. For most cases, further investigation of these unexpected results was driven by large changes in the input stock data. In other cases, the methodology could be further expanded in a number of different ways to address the causes of these unexpected flows.

Estimated flows were minimised by making an assumption on the number of stayers. Empirical evidence, where available, could be utilised to reduce the diagonal element in each birthplace specific table, producing more migration flows to and from a selected country. Discrepancies in the total number of two sequential stock tables were first altered to account for natural population change. Remaining differences were assumed to be due to moves to or from external countries not studied. In reality, in the application of the methodology the majority of these differences were most likely due to measurement issues in the stock data. The methodology could be expanded to incorporate measures of under or over counting of stocks, where known, to reduce the amount of external moves that are proportionally allocated. The World Bank stock data were taken to represent the foreign born stock totals at the start of each decade. In reality, data was actually collected over a number of years around each census period. Corrections for stock data that allow estimates of foreign born stocks at one point of time for all countries might further reduce some of the unexpected flow estimates. Changes in stocks from natural population change were controlled for using rather crude methods. More detailed data are likely to exist in some counties on the mortality by place of birth of foreign born migrants that may negate the need to distribute the total number of deaths proportionally. In addition, the estimation of the place of residence of newborns could be further developed to allow for moves outside their country of birth in these early years. Finally, the log-linear models with an offset have been used to estimate missing internal migration flow data. Alternative auxiliary data, to the distance measure used in this paper, could be utilised within the methodology outlined.

In conclusion, comparable international migration flow data are needed by researchers to better understand people's movements and identify patterns. Policy makers can also use comparable international migration flow data to help forecast populations better, where migration can often play an important role. The methodology outlined in this paper provides a relatively simple yet powerful technique to estimate global migration flow tables, exploiting newly available global stock data.

6. Acknowledgements

I would like to thank James Raymer, Nikola Sander, and Jack DeWaard for their comments on an earlier version of this paper and Joel Cohen for his recommendations concerning the validation exercises. I would also like to thank the reviewers and the Associate Editor for further suggestions on improving the paper.

Corrections:

On June 3, 2013 several typing mistakes were corrected on pages 521, 522, and on page 524.
On April 23, 2014 one mistake was corrected in equation (13) on page 516.

References

Abel, G.J. (2010). Estimation of international migration flow tables in Europe. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173(4): 797–825. doi:10.1111/j.1467-985X.2009.00636.x,.

Abel, G.J. (2012). migest: Useful R code for the estimation of migration. [electronic resource]. The CRAN Project. Retrieved 7 November, 2012, from http://cran.r-project.org/web/packages/migest/.

Beer, J., Raymer, J., van der Erf, R., and van Wissen, L. (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* 26(4): 459–481. doi:10.1007/s10680-010-9220-z,.

Bilsborrow, R.E., Hugo, G., Oberai, A.S., and Zlotnik, H. (1997). *International Migration Statistics: Guidelines for Improvement of Data Collection Systems*. Geneva: International Labour Office, 1st ed.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society, Series B* 25(1): 220–233.

Cohen, J.E., Roig, M., Reuman, D.C., and GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences* 105(40): 15269–15274. doi:10.1073/pnas.0808185105,.

Courgeau, D. (1973). Migrants et migrations. *Population (French Edition)* 28e(1): 95–129.

Flowerdew, R. (1982). Fitting the lognormal gravity model to heteroscedastic data. *Geographical Analysis* 14(3): 263–267. doi:10.1111/j.1538-4632.1982.tb00075.x,.

Flowerdew, R. (1991). Poisson Regression Models Of Migration. In: Stillwell, J. and Congdon, P. (eds.). *Migration Models: Macro and Micro Approaches*. London: Belhaven Press: 92–113.

Fotheringham, A.S. and O'Kelly, M.E. (1988). *Spatial Interaction Models:Formulations and Applications*. New York: Springer, 1st ed.

Goldstein, S. (1976). Facets of redistribution: Research challenges and opportunities. *Demography* 13(4): 423–434. doi:10.2307/2060500,.

Henning, S. and Hovy, B. (2011). Data sets on international migration. *International Migration Review* 45(4): 980–985. doi:10.1111/j.1747-7379.2011.00874_2.x,.

ISO (2006). International Organization on Standardization ISO 3166-1, Codes for the representation of names of countries and their subdivisions. Geneva: International Organization on Standardization. (-Part 1: Country codes, ISO 3166-1: 2006 (E/F)).

KC, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W., and Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050. *Demographic Research* 22(15): 383–472. doi:10.4054/DemRes.2010.22.15,.

Kelly, J.J. (1987). Improving the comparability of international migration statistics: Contributions by the conference of European statisticians from 1971 to date. *International Migration Review* 21(4): 1017–1037. doi:10.2307/2546502,.

Kim, K. and Cohen, J.E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International Migration Review* 44(4): 899–932. doi:10.1111/j.1747-7379.2010.00830.x,.

Kupiszewski, M. and Kupiszewska, D. (2008). International migration component in population dynamics models. In: Raymer, J. and Willekens, F. (eds.). *International Migration in Europe*. Chichester: Wiley: 309–327.

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J.E., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J.E. (1999). *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford: Oxford University Press.

Mayer, T. and Zignago, S. (2012). Notes on CEPII's distances measures: The GeoDist database. [electronic resource]. SSRN. (CEPII Working Paper No. 2011-25). http://ssrn.com/abstract=1994531.

Nowok, B., Kupiszewska, D., and Poulain, M. (2006). Statistics on International Migration Flows. In: Poulain, M., Perrin, N., and Singleton, A. (eds.). *Towards the Harmonisation of European Statistics on International Migration (THESIM)*. Louvain-La-Neuve: UCL–Presses Universitaires de Louvain: 203–233.

Özden, C., Parsons, C.R., Schiff, M., and Walmsley, T.L. (2011). Where on Earth is everybody? The evolution of global bilateral migration 1960–2000. *World Bank Economic Review* 25(1): 12–56. doi:10.1093/wber/lhr024,.

Raymer, J., Abel, G.J., and Smith, P.W.F. (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. doi:10.1111/j.1467-985X.2007.00490.x,.

Raymer, J., Forster, J.J., Smith, P.W.F., Bijak, J., and Wiśniowski, A. (2012). Integrated modelling of European migration: Background, specification and results. London: University College London. (NORFACE MIGRATION Discussion Paper No. 2012-04).

Rees, P.H. (1980). Multistate demographic accounts: Measurement and estimation procedures. *Environment and Planning A* 12(5): 499–531. doi:10.1068/a120499,.

Rees, P.H. and Willekens, F. (1986). Data and Accounts. In: Willekens, F. and Rogers, A. (eds.). *Migration and Settlement: A Multiregional Comparative Study*. Dordrecht: D. Reidel Publishing Company: 19–58.

Rogers, A. (1990). Requiem for the Net Migrant. *Geographical Analysis* 22(4): 283–300. doi:10.1111/j.1538-4632.1990.tb00212.x,.

Rogers, A. (1995). *Multiregional Demography: Principles, Methods and Extensions*. Chichester: John Wiley & Sons, 1st ed.

Rogers, A. and Liu, J. (2005). Estimating directional migration flows from age-specific net migration data. *Review of Urban & Regional Development Studies* 17(3): 177–196. doi:10.1111/j.1467-940X.2005.00102.x,.

Rogers, A. and Raymer, J. (2005). Origin dependence, secondary migration, and the indirect estimation of migration flows from population stocks. *Journal of Population Research* 22(1): 1–19. doi:10.1007/BF03031801,.

Rogers, A. and von Rabenau, B. (1971). Estimation of interregional migration streams from place-of-birth-by-residence data. *Demography* 8(2): 185–194.

Salt, J. (1993). Migration and population change in Europe. New York: United Nations Institute for Disarmament Research, (UNIDIR). (Research Paper No. 19|UNIDIR/93/23).

Sen, A.K. and Smith, T.E. (1995). *Gravity Models of Spatial Interaction Behavior (Advances in Spatial and Network Economics)*. New York: Springer-Verlag.

Stewart, J.Q. (1941). An inverse distance variation for certain social influences. *Science* 93(2404): 89–90.

United Nations General Assembly (2011). Resolution adopted by the General Assembly. In: *65th Session: International migration and development*. Vol.67 170. United Nations.

United Nations Population Division (2009). International migration flows to and from selected countries: The 2008 revision. [electronic resource].

United Nations Department of Economic and Social Affairs/Population Division. http://www.un.org/esa/population/publications/2008ittmigflows/2008_ittmigflows_documentation.pdf.

United Nations Population Division (2011). World population prospects: The 2010 revision, highlights and advance tables. New York: United Nations Department of Economic and Social Affairs/Population Division: 1–164. (Working Paper ESA/P/WP).

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. New York: Springer, 4th ed.

Willekens, F. (1983). Log-linear modelling of spatial interaction. *Papers in Regional Science* 52(1): 187–205. doi:10.1111/j.1435-5597.1983.tb01658.x,.

Willekens, F. (1999). Modeling approaches to the indirect estimation of migration flows: From entropy to EM. *Mathematical Population Studies* 7(3): 239–278. doi:10.1080/08898489909525459,.

Willekens, F. (1994). Monitoring international migration flows in Europe. *European Journal of Population/Revue européenne de Démographie* 10(1): 1–42. doi:10.1007/BF01268210,.

Wilson, A.G. (1970). *Entropy in Urban and Regional Modelling*. London: Pion.

Zipf, G.K. (1942). The unity of nature, least action, and natural social science. *Sociometry* 5(1): 48–62.