

International migration beyond gravity: A statistical model for use in population projections

Joel E. Cohen^{*†}, Marta Roig[‡], Daniel C. Reuman^{*§}, and Cai GoGwilt[¶]

^{*}Laboratory of Populations, Rockefeller and Columbia Universities, 1230 York Avenue, Box 20, New York, NY 10065-6399; [‡]Population Division, Department of Economic and Social Affairs, 2, United Nations Plaza DC2-1984, United Nations, New York, NY 10017; [§]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY Berkshire United Kingdom; and [¶]321 East 54th Street Apartment 4B, New York, NY 10022

Contributed by Joel E. Cohen, August 18, 2008 (sent for review June 18, 2008)

International migration will play an increasing role in the demographic future of most nations if fertility continues to decline globally. We developed an algorithm to project future numbers of international migrants from any country or region to any other. The proposed generalized linear model (GLM) used geographic and demographic independent variables only (the population and area of origins and destinations of migrants, the distance between origin and destination, the calendar year, and indicator variables to quantify nonrandom characteristics of individual countries). The dependent variable, yearly numbers of migrants, was quantified by 43653 reports from 11 countries of migration from 228 origins and to 195 destinations during 1960–2004. The final GLM based on all data was selected by the Bayesian information criterion. The number of migrants per year from origin to destination was proportional to $(\text{population of origin})^{0.86}(\text{area of origin})^{-0.21}(\text{population of destination})^{0.36}(\text{distance})^{-0.97}$, multiplied by functions of year and country-specific indicator variables. The number of emigrants from an origin depended on both its population and its population density. For a variable initial year and a fixed terminal year 2004, the parameter estimates appeared stable. Multiple R^2 , the fraction of variation in log numbers of migrants accounted for by the starting model, improved gradually with recentness of the data: $R^2 = 0.57$ for data from 1960 to 2004, $R^2 = 0.59$ for 1985–2004, $R^2 = 0.61$ for 1995–2004, and $R^2 = 0.64$ for 2000–2004. The migration estimates generated by the model may be embedded in deterministic or stochastic population projections.

generalized linear model | geography | population density | spatial interaction model | stochastic population projection

International migration will play an increasing role in the demographic future of nations if fertility continues to decline in most countries. In projecting international migration, the United Nations Population Division (ref. 1, paragraphs 57–59) identified the need for a demographically plausible, programmable algorithm that automatically projects a zero world balance of net migration and prevents projected net emigration from completely depleting the population of any sending country. To meet this need, we propose an algorithm (based only on demographic and constant geographic variables) for projecting future numbers of international migrants from any country or region to any other. It is comparable in transparency and generality to standard cohort-component methods of projecting births and deaths. The approach presented here is different from methods of projecting migrant flows currently practiced in international demographic institutions, the United States, European countries, and other developed countries (2–6).

Most theories of international migration draw on social, economic and/or political factors to explain migration (4, 7–9), such as differences among countries in gross domestic product, labor markets, migration policies, social networks of prior migrants, and cognitive and behavioral attributes of individuals (3, 10–11). For multidecadal demographic projections, it seems more difficult to project such nondemographic variables than it is to project demographic variables such as fertility and mortality. The proposed model assumes the availability only of constant geographic variables

and of population sizes that can be projected incrementally in time by accepted demographic procedures. The model makes possible both deterministic and stochastic projections of migration and hence of population.

The intellectual antecedents of the proposed model include Zipf's (12, 13) model of intercity migration, which is one of several "gravity" models in the social sciences (6). Zipf (12) aimed to "show with supporting data that the number of persons that move between any two communities in the United States whose respective populations are P_1 and P_2 and which are separated by the shortest transportation distance, D , will be proportionate to the ratio, $P_1 P_2 / D$, subject to the effect of modifying factors." Unlike Zipf, we distinguish the number of people who move from community 1 to community 2 from the number of people who move from 2 to 1. Taking logarithms of $P_1 P_2 / D$ and adding an error term yields a linear model in log-transformed variables, $\log(\text{migrants}) = a_0 + a_1 \log(\text{ppnorig}) + a_2 \log(\text{ppndest}) + a_3 \log(\text{distance}) + \text{error}$, where ppnorig is the population of the origin, ppndest is the population of the destination, and the error term characterizes random deviations. Here and throughout, \log refers to \log_{10} , and \ln refers to the natural logarithm \log_e . Zipf posits that $a_1 = 1$, $a_2 = 1$, and $a_3 = -1$. We estimate all coefficients from data using a generalized linear model (GLM) (14).

Zipf (12) treated cities as points of negligible spatial extent. Our communities are countries or regions and our subject is international migration. To let the data reveal whether the area of a country influences its numbers of migrants, we add two terms to the above equation: $\log(\text{areaorig})$, the log area of the origin, and $\log(\text{areadest})$, the log area of the destination. By definition, the population density of the origin is $\text{ppnorig}/\text{areaorig}$, so $\log(\text{density}) = \log(\text{ppnorig}) - \log(\text{areaorig})$. If the number of migrants from origin to destination depends on ppnorig and ppndest and not on their areas, then the estimated coefficients of $\log(\text{areaorig})$ and $\log(\text{areadest})$ should be close to zero. However, if the number of migrants from origin to destination depends on the population density of the origin and the population density of the destination, but not on their respective population numbers *per se*, then the estimated coefficients of $\log(\text{areaorig})$ and $\log(\text{areadest})$ should be nearly the negative of the respective estimated coefficients of $\log(\text{ppnorig})$ and $\log(\text{ppndest})$. The estimated coefficients of the terms for log population and log area of origin and destination reveal the relative importance of population *per se* and population density.

To allow for differences in migratory intensities among origins or destinations, we let the data reveal whether different origins and destinations had numbers of migrants different from the numbers of migrants expected on average from their respective populations, areas, and distances. We introduced four indicator variables for

Author contributions: J.E.C. designed research; J.E.C. performed research; J.E.C., D.C.R., and C.G. contributed new reagents/analytic tools; J.E.C. and M.R. analyzed data; and J.E.C., M.R., and D.C.R. wrote the paper.

The authors declare no conflict of interest.

[†]To whom correspondence should be addressed. E-mail: cohen@rockefeller.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808185105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

each country that provided data. For example, Australia was a source of data on numbers of emigrants and immigrants by year. One indicator variable for Australia, $\text{orig.indicator}\$Australia$, equaled 1 whenever Australia was the origin and equaled 0 if any other country or area was the origin. A second indicator variable, $\text{dest.indicator}\$Australia$, similarly equaled 1 whenever Australia was the destination and equaled 0 if any other country or area was the destination. A third indicator variable, $\text{orig.is.datasources}\$Australia$, equaled 1 if Australia was the origin and was the source of the data and otherwise equaled 0. A fourth indicator variable, $\text{dest.is.datasources}\$Australia$, equaled 1 if Australia was the destination and was the source of the data and otherwise equaled 0. Similar indicator variables were defined for each country that provided data on its flows of immigrants and/or emigrants.

To allow for the possibility that migration rates changed over time and for the simultaneous effects of the other independent variables (populations, areas, distance, and indicator variables), we introduced year as an independent variable. We used the centered variable year minus 1985 to assure the stability of the model's estimated intercept.

To summarize in notation similar to that of Zipf's gravity model, let P be population, A be area, and D be distance. Detailed definitions of these variables are given in *Methods*. Our "starting" model was the logarithmic transformation of model Eq. 1. This logarithmic transformation guaranteed that the number of migrants estimated by the model was positive.

$$\begin{aligned} & \text{migrants from origin to destination} && [1] \\ & = \text{constant} \times P_{\text{orig}}^a P_{\text{dest}}^b A_{\text{orig}}^c A_{\text{dest}}^d D_{\text{orig,dest}}^f \times \\ & [10 \text{ if origin is Australia; } 1 \text{ otherwise}]^g \times \\ & [10 \text{ if destination is Australia; } 1 \text{ otherwise}]^h \times \\ & [\text{similar indicator variables for seven other origins and ten} \\ & \text{other destinations}]^{\text{coef}} \times \\ & [10 \text{ if origin is Australia and Australia is the source of the} \\ & \text{data; } 1 \text{ otherwise}]^i \times \\ & [10 \text{ if destination is Australia and Australia is the source of} \\ & \text{the data; } 1 \text{ otherwise}]^j \times \\ & [\text{similar indicator variables for seven other origins and ten} \\ & \text{other destinations}]^{\text{coef}} \times \\ & 10^{k(\text{year}-1985)} \times \\ & [\text{lognormal error term } \exp(\varepsilon)], \end{aligned}$$

where $\varepsilon \sim N(0, \sigma^2)$ are independent.

This model contains factors of the form $[10 \text{ if origin is Australia; } 1 \text{ otherwise}]^g$. The common logarithm of this expression is $g \times [1 \text{ if origin is Australia; } 0 \text{ otherwise}]$, and the expression in square brackets is an indicator (or dummy) variable. The coefficient g summarizes the nonrandom effects on the expected number of migrants from Australia (in this case) apart from those due to calendar year, the populations and areas of origin and destination, and distance. The indicator variables quantified how numbers of migrants were affected by nonrandom characteristics of individual countries: migratory history, policy, and statistical completeness; economic, geographic, and cultural affinities or disparities; and effects of geographical adjacency not captured by the chosen measure of distance.

We estimated the intercept (log constant) and exponents a, b, c, d, \dots , which were linear coefficients in the GLM. We fitted this and other models to data from 11 countries (Australia, Belgium, Canada, Denmark, Germany, Italy, the Netherlands, Spain, Sweden, the U.K., and the U.S.A.). These countries were selected on the basis of the quality of their data on international migrants by year and places of origin and destination from 1960 to 2004. The data included 228 origins and 195 destinations of migrants. Oceania was the only destination not also an origin, so 229 countries or regions in total were named in these data.

Because of the limited quantity and quality of migration data, this article demonstrates a method and illustrates a modeling approach, rather than specifying numerical parameters definitively. Parameters and models will evolve as more and better data become available. The analysis showed that statistically simple and demographically interpretable modeling accounted for more than half the variation in the migration data. How much more than half could be accounted for by this approach with better data remains to be determined.

Results

Descriptive Bivariate Relationships. On average, but with enormous variability, the log number of migrants increased with increasing log population of origin ($r = 0.43$), increasing log area of origin ($r = 0.18$), increasing log population of destination ($r = 0.27$), and increasing log area of destination ($r = 0.10$) (Fig. 1 *A–D*). The log number of migrants decreased, on average but with enormous variability, with increasing log distance from origin to destination ($r = -0.24$) and increased weakly with year ($r = 0.01$) (Fig. 1 *E* and *F*). Log population and log area were highly correlated (Fig. 1 *G* and *H*) for origins ($r = 0.74$) and destinations ($r = 0.64$). Because these were the two highest-magnitude correlations, collinearity was not a problem in fitting the GLM. Correlations between year and log migrants, and between $\log(\text{ppnorig})$ and $\log(\text{distance})$, were insubstantial.

$\log(\text{ppnorig})$ and $\log(\text{ppndest})$ were negatively correlated ($r = -0.19$) (Fig. 1 *I*). This negative correlation could reflect the absence of data sources among countries with four million or fewer people, which may account for the absence of data points in the lower left quadrant of Fig. 1 *I*. The long horizontal and vertical streaks in Fig. 1 *I* represent the populations of countries that were data sources as origins and destinations, respectively, whereas the short diagonal streaks largely reflect population growth of a given (origin, destination) pair.

Model Selection. The starting linear model was fitted [[supporting information \(SI\) Table S1](#)] with the dependent variable $\log(\text{migrants})$ and with the six independent variables that we call "basic" [year minus 1985, $\log(\text{ppnorig})$, $\log(\text{areaoorig})$, $\log(\text{ppndest})$, $\log(\text{areadest})$, $\log(\text{distance})$] and all indicator variables (orig.indicator , dest.indicator , $\text{orig.is.datasources}$, $\text{dest.is.datasources}$). The Multiple R^2 was 0.5693 and the Adjusted R^2 was 0.5689 (see *Methods*).

When the stepwise algorithm with Bayesian information criterion (15) was applied to this starting model, $\log(\text{areadest})$ was eliminated and all other independent variables were retained in the resulting "final" model (Table 1). To the four significant digits shown, the Multiple R^2 and the Adjusted R^2 were unchanged between the starting and the final models.

When the values of $\log(\text{migrants})$ were independently and randomly permuted in each of 100 simulations, the maximum of the 100 simulated multiple R^2 values was 0.00147, far smaller than the multiple R^2 value of 0.5693 for the data. Hence, the latter multiple R^2 value could not have been an artifact of the fitting procedure alone.

When all indicator variables were suppressed and only the six "basic" independent variables were retained, the multiple R^2 value dropped to 0.4345 (Table S3). The only notable change in the coefficients of the six "basic" independent variables was the increase in the coefficient of $\log(\text{areadest})$ from 0.0239 to 0.1604. A

Table 1. The “final” model of log migrants as a function of year, log population of origin, log area of origin, log population of destination, log distance, and indicator variables; specification of the model in R and resulting coefficients and statistics

Call:

lm (formula = logmigrants ~ I (year–1985) + logppnorig + logareaorig + logppndest + logdistance + orig.indicator + dest.indicator + orig.is.datasourc + dest.is.datasourc)

Coefficients	Estimate	SE	t value
(Intercept)	–2.5135	0.0886	–28.3730
I (year–1985)	0.0016	0.0003	5.1670
Logppnorig	0.8631	0.0083	103.6400
Logareaorig	–0.2103	0.0066	–31.9050
Logppndest	0.3604	0.0089	40.7010
Logdistance	–0.9685	0.0102	–94.5470
orig.indicatorAustralia	1.1295	0.0436	25.8900
orig.indicatorBelgium	–0.2557	0.0404	–6.3300
orig.indicatorDenmark	–0.0441	0.0409	–1.0760
orig.indicatorGermany	0.0699	0.0409	1.7080
orig.indicatorItaly	0.1844	0.0401	4.5960
orig.indicatorNetherlands	0.0250	0.0408	0.6120
orig.indicatorSweden	0.1602	0.0473	3.3840
orig.indicatorUnited Kingdom	0.2486	0.0397	6.2580
dest.indicatorAustralia	1.4362	0.0558	25.7360
dest.indicatorBelgium	0.1314	0.0524	2.5090
dest.indicatorCanada	0.8557	0.0457	18.7160
dest.indicatorDenmark	0.2560	0.0523	4.8980
dest.indicatorGermany	0.5875	0.0502	11.7020
dest.indicatorItaly	0.7551	0.0493	15.3090
dest.indicatorNetherlands	0.4805	0.0509	9.4480
dest.indicatorSpain	0.6400	0.0470	13.6210
dest.indicatorSweden	0.2528	0.0696	3.6340
dest.indicatorUnited Kingdom	0.6284	0.0491	12.7910
dest.indicatorUnited States of America	1.1444	0.0457	25.0180
orig.is.datasourcAustralia	–0.3038	0.0633	–4.8010
orig.is.datasourcBelgium	0.4595	0.0626	7.3410
orig.is.datasourcDenmark	0.2340	0.0642	3.6450
orig.is.datasourcGermany	0.4872	0.0613	7.9500
orig.is.datasourcItaly	–0.4761	0.0630	–7.5620
orig.is.datasourcNetherlands	0.2112	0.0645	3.2750
orig.is.datasourcSweden	–0.0733	0.0658	–1.1140
orig.is.datasourcUnited Kingdom	1.3529	0.0682	19.8440
dest.is.datasourcAustralia	–0.0317	0.0719	–0.4420
dest.is.datasourcBelgium	0.5479	0.0716	7.6510
dest.is.datasourcCanada	0.1462	0.0638	2.2930
dest.is.datasourcDenmark	0.2687	0.0721	3.7270
dest.is.datasourcGermany	0.5646	0.0674	8.3730
dest.is.datasourcItaly	–0.2356	0.0687	–3.4280
dest.is.datasourcNetherlands	0.4550	0.0718	6.3360
dest.is.datasourcSpain	–0.2294	0.0678	–3.3850
dest.is.datasourcSweden	0.1258	0.0831	1.5140
dest.is.datasourcUnited Kingdom	1.4979	0.0762	19.6680
dest.is.datasourcUnited States of America	NA	NA	NA

Residual standard error: 0.6957 on 43610 degrees of freedom
Multiple R²: 0.5693, Adjusted R²: 0.5689
F statistic: 1,372 on 42 and 43,610 DF, nominal P value: < 2.2e–16

Because the assumption of independence among observations was implausible, conventional measures of statistical significance were probably inapplicable. The destination.is.data.source parameter for the U.S. is labeled “NA” because one of the destination.is.data.source variables must be eliminated to prevent a singularity in estimating the coefficients. See [S1 Text](#) for details.

mainly from a survey of arriving and departing international passengers; U.K.’s recording of emigration and immigration flows is equally complete. Australia, by contrast, focuses on entries. These differences between countries are detected by the statistical analysis and reflected in the estimated coefficients of the model.

According to its coefficient for dest.is.datasourc, when the U.K. reported the number of immigrants to the U.K., its reports were on average 31.47 times the number of immigrants that would have been reported by the average reporting system over all countries in the study, given the U.K.’s population and area, the destination and year, and the propensity to immigrate to the U.K. estimated without regard to the source of the data. At the opposite extreme, Spain and Italy reported 59% and 58%, respectively, of the numbers of immigrants expected from other factors.

In general, immigration is better recorded at destinations than emigration is recorded at origins, in part because migrants often have more formal incentives to register at their destination than to deregister at their origin. The estimated coefficients of the indicator variables (Table 1) are consistent with this belief, although other interpretations are possible. For example, dest.indicator for Australia had coefficient 1.4362, greater than the coefficient 1.1295 of orig.indicator for Australia. The same inequality, dest.indicator coefficient > orig.indicator coefficient, held for all eight countries that supplied both immigration and emigration data. Similarly, the comparable inequality, dest.is.datasourc coefficient > orig.is.datasourc coefficient, held for all eight countries that supplied both immigration and emigration data. These inequalities indicated greater detection (or greater intensity, an alternative interpretation) of immigration than emigration in every country for which the comparison could be made.

Parameter Stability: How Much of the Past Is Relevant to the Future?

Estimated coefficients varied systematically as a function of the time interval from which data were drawn and as a function of the subset of variables selected from the starting model Eq. 1 (Table S3). The starting model rather than the final model was used for this analysis to allow for the possibility that the log area of destination might become dramatically important for some subset of 1960–2004. As it turned out, this possibility did not occur.

For each set of independent variables, the most recent five years (2000–2004) gave the highest multiple R². Using all variables, all data from 1960 to 2004 gave a multiple R² value of 0.57, whereas the 2000–2004 data gave a multiple R² value of 0.64 (Table S3). This higher value of the multiple R² may be partially due to improved quality of data in more recent years but may also be due to fewer external perturbations to migratory flows during 2000–2004 than during 1960–2004. For example, during the 45-year period of 1960–2004, the Berlin wall and the Soviet Union fell, and Germany was reunified, whereas no such events marked the 5-year period of 2000–2004, an interval only one-ninth as long. For each set of independent variables, as the initial year moved forward while the terminal year was 2004, the multiple R² value increased. Each set of independent variables accounted for more variation in log(migrants) when using the data from 1985 to 2004 than by adding additional data from earlier years.

All estimated coefficients were stabler when using a moving initial year than when using a moving terminal year. They were least stable when using moving tranches of 5 or 10 years’ duration (Table S3). For the starting model with all variables, the standard deviation 0.3877 of the estimated intercept for the five time intervals with moving terminal year 1960–1984, 1960–1989, 1960–1994, 1960–1999, and 1960–2004 was larger than the standard deviation 0.2489 of the estimated intercept for the five time intervals with moving initial year 1965–2004, 1970–2004, 1975–2004, 1980–2004, and 1985–2004. The same inequality held for the standard deviations of the estimates of the coefficients of all six basic variables when the five time intervals with moving terminal year were compared with the five time intervals with moving initial year.

examine trends over time in the coefficients of the indicator variables and to extrapolate those trends forward to obtain updated coefficients for the indicator variables of future migration matrices.

To obtain an age-specific net migration vector, one could apply age-specific models of migration patterns for different countries and regions to $E(i,t)$ and to $I(i,t)$ (19–21) and combine that with the projected age-specific fertility and mortality vectors as in Eq. 2.

This method of projecting international migration in combination with cohort-component methods of projecting fertility and mortality solves the two problems identified by the United Nations Population Division (1). First, the global sum of net migrants is guaranteed to be 0 when the numbers of emigrants, immigrants, and net migrants are derived from a migration matrix (22). Second, net emigration should not completely deplete the population of any sending country for realistic model parameters. In all models considered here, the exponents of the populations of origin and destination are positive and the predicted number of emigrants is a small fraction of the population of origin. Consequently, the projected number of emigrants from an origin declines to 0 as its population declines to 0. Likewise, if the destination's population declines, the models predict that fewer people will migrate there. This feature of the model depends on realistic parameter values and may not hold in all mathematically possible cases. If the coefficients of the model were unrealistic and gave an unrealistically large number of emigrants, then it is possible that the origin population could be depleted.

Open Research Questions. Many open questions remain. Why did the area of the destination influence the numbers of migrants much less than the area of origin? How well would the proposed models work for migration within a country, taking account of international migration? For a given origin-destination pair, were the residuals correlated over time or independent as the error term in the model assumed?

International migrants are mainly younger individuals of working age and their families. Migratory flows of elderly individuals may also be important. Would the fit of the models be improved by replacing total population size with a weighted average that emphasized age groups most prone to migrate, or by a simple index such as the proportion of the population age 20–34 years? Age-structure seems likely to matter to migration increasingly as all countries undergo population aging (23).

What is the long-run behavior of the projection model Eq. 2 assuming constant birth rates and death rates and constant coefficients in Eq. 1? For example, when, if ever, does there exist a fixed vector P of population size by country and a stable growth rate ρ such that $\lim_{t \rightarrow \infty} P(t)/\rho^t = P$? If this case arises, how does ρ depend on the parameters of the basic model Eq. 2? What irreducibility

conditions on the migration matrix assure the uniqueness of P ? When the migration matrix is reducible, could different “ergodic sets” of countries (sets of countries linked through migration flows) have different fixed vectors P of population size by country and different stable growth rates ρ ? Under what conditions on the coefficients of $\log(\text{ppnorig})$ and $\log(\text{ppndest})$ can country populations snowball to infinity (in finite time or with infinite time) or vanish? How sensitive to the initial conditions $P(i,0)$ are each country's proportion of world population $P(i,t)/\sum_i P(i,t)$ for large t , where the summation runs over all countries? How sensitive are country-specific population growth rates $P(i,t+1)/P(i,t)$, for large t , to initial conditions? In short, what ergodic theorems hold (24)?

Methods

Data. All 43,653 data records are provided in [Dataset S1](#). Each record contains 12 variables: a unique serial number, the year in the Western calendar (1960–2004), the name of the country or region of origin of migrants, the log population of the origin in that year, the name of the country or region of destination of migrants, the log population of the destination in that year, the log number of migrants from origin to destination in that year, the log area (square kilometers) of the origin, the log area (square kilometers) of the destination, the log great circle distance (kilometers) from the capital of the origin to the capital of the destination, the source of the migration data, and “neighbor” (see [SI Text](#)). Records for which the value of any variable was missing were excluded.

Each country's definitions of what constituted a migrant, of the origin or destination of a migrant, and of the accounting year were used ([Table S4](#)). Differences among definitions and in the effectiveness of collecting migration data led to hundreds of discrepancies when both the origin and the destination reported migration data in the same year. In [SI Text](#), sources are listed and methods of collecting migration data are discussed.

Data Analysis. A GLM was fitted to a starting model with dependent variable $\log(\text{migrants})$ and with all six basic independent variables [year minus 1985, $\log(\text{ppnorig})$, $\log(\text{areaorig})$, $\log(\text{ppndest})$, $\log(\text{areadest})$, and $\log(\text{distance})$] and all indicator variables (orig.indicator , dest.indicator , $\text{orig.is.datasources}$, $\text{dest.is.datasources}$). The stepwise regression algorithm stepAIC was applied to this linear model to obtain a final model. Details of data management and statistical software are in [SI Text](#).

ACKNOWLEDGMENTS. Earlier versions of this article were presented to the Technical Working Group on Long-Range Population Projections, Population Division, United Nations Headquarters, on June 30, 2003; the United Nations Population Division Seminar on October 21, 2004; the Stanford Workshop on Formal and Quantitative Demography on August 15, 2005; the Policy Research Division of the Population Council on January 23, 2006; and the Estimates and Projection Section of the Population Division of the United Nations on June 6, 2008. We thank for helpful comments the many participants in those presentations and Adam E. Cohen, Jakub Bijak, Joshua R. Goldstein, Ryuichi Kaneko, Ronald D. Lee, Jim Oeppen, and Hania Zlotnik. We thank Mr. and Mrs. William T. Golden and family for their hospitality during this work. Priscilla K. Rogerson assisted in preparing data and the manuscript. J. Bijak and R. Kaneko helpfully reviewed two versions of the article. This work was supported by U.S. National Science Foundation Grants DEB-9981552 and DMS-0443803.

- United Nations Population Division (2003) *United Nations Population Projections to 2300*. ESA/PIWP.184, 30 June 2003. (UNPD, New York).
- Fertig M, Schmidt CM (2001) *Aggregate-Level Migration Studies as a Tool for Forecasting Future Migration Streams*. In *International Migration*, ed Djajic S (Routledge, London), pp 110–136.
- Howe N, Jackson R (2005) *Projecting Immigration: A Survey of the Current State of Practice and Theory. A Report of the CSIS Global Aging Initiative, with Contributions by Rebecca Strauss and Keisuke Nakashima* (Center for Strategic and International Studies, Washington, DC).
- Bijak J (2008) PhD Dissertation. (Central European Forum for Migration and Population Research, Warsaw).
- Raymer J, Willekens F, eds (2008) *International Migration in Europe: Data, Models and Estimates* (Wiley, Chichester).
- Rogers A (2008) Demographic modeling of the geography of migration and population: A multiregional perspective. *Geog Analysis* 40:276–296.
- Massey DS, et al. (1993) Theories of international migration: A review and appraisal. *Popul Dev Rev* 19:431–466.
- Massey DS, et al. (1998) *Worlds in Motion: Understanding International Migration at the End of the Millennium (IUSSP International Studies in Demography)* (Clarendon, Oxford).
- Faist T (2000) *The Volume and Dynamics of International Migration and Transnational Social Spaces* (Oxford Univ Press, Oxford).
- Ritchey PN (1976) Explanations of migration. *Annu Rev Sociol* 2:363–404.
- Dorigo G, Tobler W (1983) Push-pull migration laws. *Ann Assoc Am Geogr* 73:1–17.
- Zipf GK (1946) The P_1P_2/D hypothesis: On the inter-city movement of persons. *Am Sociol Rev* 11:677–686.
- Zipf GK (1949) *Human Behavior and the Principle of Least Effort* (Addison Wesley, Cambridge MA).
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*. 2d ed (Chapman and Hall, London).
- Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464.
- Technical Panel on Assumptions and Methods (2003) *Report to the Social Security Advisory Board* (Social Security Administration, Washington, DC).
- Bijak J, Kupiszewska D, Kupiszewski M (2008) Replacement migration revisited: Simulations of the effects of selected population and labour market strategies for the ageing Europe, 2002–2052. *Popul Res Policy Rev* 27:321–342.
- Raymer J (2008) in *International Migration in Europe: Data, Models and Estimates*, eds Raymer J, Willekens F (Wiley, Chichester), pp 209–234.
- Rogers A, Castro LJ (1981) *Model migration schedules. Research Report-81-30* (International Institute for Applied Systems Analysis, Laxenburg, Austria).
- Rogers A, Willekens F, Raymer J (2003) Imposing age and spatial structures on inadequate migration-flow datasets. *Prof Geogr* 55:56–69.
- Rogers A, Castro LJ, Lea M (2004) Model migration schedules: Three alternative linear parameter estimation methods. *Mathematical Population Studies* 12:17–38.
- Cohen JE (2008) Constant global population with demographic heterogeneity. *Demogr Res* 18:409–436.
- Raymer J, Rogers A (2008) in *International Migration in Europe: Data, Models and Estimates*, eds Raymer J, Willekens F (Wiley, Chichester), pp 175–192.
- Cohen JE (1979) Ergodic theorems of demography. *Bull Amer Math Soc NS* 1:275–295.