
Corrected version
2012-07-07

UN/POP/MIG-10CM/2012/11
3 February 2012

**TENTH COORDINATION MEETING ON
INTERNATIONAL MIGRATION**

Population Division
Department of Economic and Social Affairs
United Nations Secretariat
New York, 9-10 February 2012

PROJECTION OF NET MIGRATION USING A GRAVITY MODEL¹

Laboratory of Populations²

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

² Joel E. Cohen, Laboratory of Populations, The Rockefeller University and Columbia University, New York, NY

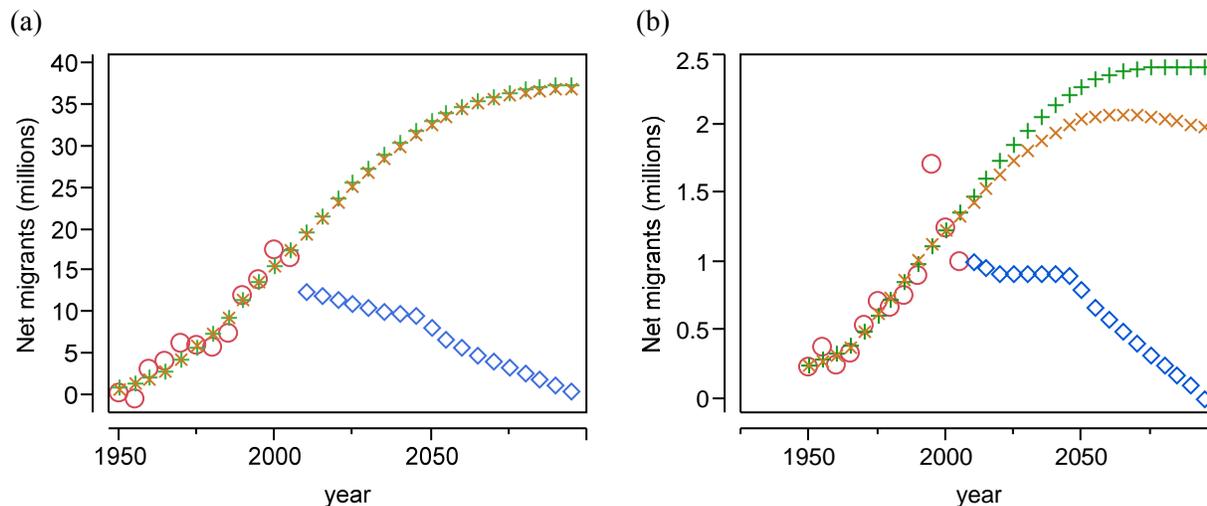
A. INTRODUCTION

International migration remains a frontier area of demography. Countries and demographers differ regarding the definition, estimation, and projection of international human migrant stocks and flows. Despite these difficulties, the United Nations Population Division (henceforth UNPD) prepares, every two years, estimates (referring to past quantities) and projections (referring to future quantities) of net migration (immigration minus emigration) for all countries and regions of the world. This note compares two examples of the most recent UNPD projections (in its *World Population Prospects: The 2010 Revision*) with alternative projections of net migration based on gravity-type models for migrant flows (Cohen et al., 2008; Kim and Cohen, 2010). The two examples are net migration to the more developed regions (henceforth abbreviated M) from the less developed regions (henceforth abbreviated L) and net migration to the United States of America (USA) from the world outside the USA (W-USA).

B. UNITED NATIONS ESTIMATES AND PROJECTIONS

During the 5-year intervals starting from 1950 to 2005, UNPD estimated generally rising net migration to M (open circles in Figure 1a; numbers in Table 1) and to USA (open circles in Figure 1b), apart from a decline in the last two quinquennia. The UNPD assumed declining net migration to M and to USA during the 5-year intervals starting from 2010 to 2095 (open diamonds in Figure 1a,b; numbers for M only in Table 1). According to UNPD *Assumptions Underlying the 2010 Revision* (2011, p. 12, paragraph C.1), “Projected levels of net migration are generally kept constant over the next decades. After 2050, it is assumed that net migration will gradually decline.”

Figure 1. Net migration (millions of net migrants per 5-year interval) from 1950-54 through 2095-2099 to (a) more developed regions and (b) USA. \circ : UNPD estimates (1950-2005). \diamond : UNPD projections (2010-2095) in WPP 2010. $+$: gravity model, \times : linear model. The difference between $+$ and \times is similar in (a) and (b) but the vertical scales differ.



C. ALTERNATIVE PROJECTIONS BASED ON GRAVITY AND LINEAR MODELS

Two alternative projections of net migration use (1) a gravity model and (2) a linear model. After calibration (a) using UNPD estimates; the projection (b) uses the calibrated models. The gravity and linear models both yield estimates and projections of migrant flows in each direction.

1. Gravity model

a. Calibration

Let $L(t)$, $M(t)$, and $N(t)$ denote, respectively, the UNPD estimates of population of L (“less” in Table 1), the population of M (“more” in Table 1), and the net migration (“netmig” in Table 1) to M, for $t = 1950, 1955, \dots, 2005$. In this example, we treat L as a single country and M as a single country, as in the biregional projection model of Rogers (1995, pp. 10ff).

Ignoring all predictor variables other than population sizes of origin and destination in Kim and Cohen’s (2010) gravity model, the number of immigrants from L to M in the 5-year interval starting in year t is expected to be proportional to $in(t) = L(t)^\alpha M(t)^\beta$ and the number of emigrants from M to L in the 5-year interval starting in year t is expected to be proportional to $out(t) = L(t)^\gamma M(t)^\delta$. The values $\alpha = 0.728$, $\beta = 0.602$ from Kim and Cohen (2010, p. 912, Table 2, Model M2) put more weight on L(t) than on M(t). The values $\gamma = 0.373$, $\delta = 0.948$ from Kim and Cohen (2010, p. 914, Table 3, Model M2) put more weight on M(t) than on L(t). Table 1 shows the computed values of $in(t)$ and $out(t)$. We used ordinary least squares to choose the numbers a and b that minimized the sum of the squared deviations between the UNPD estimates of $N(t)$ in the past and the net migration predicted by this gravity model (in the past), that is, we minimized $\sum_{t=1950}^{2005} \{N(t) - [a \times in(t) + b \times out(t)]\}^2$. Although we imposed no constraints on the signs of a and b , the estimates had the expected signs, namely, $a = 0.00185647379417 > 0$ and $b = -0.0024830665233 < 0$. (If the vector $in(t)$ were a multiple of the vector $out(t)$, this procedure would fail because values of a and b would not be uniquely defined.) Using these values of a and b , we estimated net migration (in the past) as $\widehat{N}(t) = a \times in(t) + b \times out(t)$ for $t = 1950, 1955, \dots, 2005$ (+ signs in Figure 1a). These values of $\widehat{N}(t)$ are essentially a smoothing of the UNPD estimates based on the simplified gravity model.

Let $USA(t)$ denote the USA population size for t , and let $W-USA(t)$ denote the world population size outside the USA estimated for t . For net migration to the USA, we replaced $M(t)$ with $USA(t)$ and we replaced $L(t)$ with $W-USA(t)$. Using the same values of α , β , γ , and δ as above, we again calculated $in(t)$ and $out(t)$ as above and used ordinary least squares without constraints to estimate $a = 0.000343441164228713$ and $b = -0.000828297065822388$. The resulting estimates $\widehat{N}(t) = a \times in(t) + b \times out(t)$ of past net migration are shown by the + signs in Figure 1b.

b. Projection

We used the UNPD’s projections of $L(t)$ and $M(t)$ for $t = 2010, 2015, \dots, 2095$, to project future net migration to M as $\widehat{N}(t) = a \times in(t) + b \times out(t) = aL(t)^\alpha M(t)^\beta + bL(t)^\gamma M(t)^\delta$. These projections used the values of α , β , γ , and δ from Kim and Cohen (2010) plus the values of $a > 0$ and $b < 0$ from the calibration. The projected net migration rose smoothly from the most recent estimate and leveled off sigmoidally as year 2100 approached (+ signs in Figure 1a). The diminishing rate of increase in projected net migration resulted from the diminishing rates of increase in the UNPD’s projections of $L(t)$ and $M(t)$.

Similarly, for net migration to USA, we used the same values of α , β , γ , and δ and the values of $a > 0$ and $b < 0$ obtained from the calibration of estimated net migration to USA. We projected $\widehat{N}(t)$ as above with $L(t)$ replaced by $W-USA(t)$ and $M(t)$ replaced by $USA(t)$ (+ signs in Figure 1b).

2. Linear model

The linear model assumes that the number of migrants per time interval from origin to destination is proportional to the population size of the origin, independent of the population size of the destination. The linear model is a special case of the gravity model obtained by setting $\alpha = 1$, $\beta = 0$, $\gamma = 0$, $\delta = 1$. For net migration to M, unconstrained ordinary least-squares fitting of this linear model to UNPD estimates $N(t)$ yielded the estimates $a = 0.00600808381247244$, $b = -0.0118084850702329$. For net migration to USA, the parameter estimates were $a = 0.000492367745402185$, $b = -0.00583460454851918$. We then calculated $\widehat{N}(t)$ for $t = 1950, 1955, \dots, 2095$ (\times signs in Figure 1).

The ratio $-b/a$ is the ratio of the probability per person of migration from M to L (or from USA to W-USA, respectively) divided by the probability per person of the reverse migration. In both examples, $-b/a > 1$ mainly because L and W-USA are so much more populous than M and USA.

D. CRITICAL COMMENTS

These procedures have shortcomings. One internal contradiction is, fortunately, easily repaired and of small quantitative effect. Although $M(t+5) = M(t) + B(t) - D(t) + N(t)$, where $B(t)$ is the number of births, $D(t)$ is the number of deaths, and $N(t)$ is the net migration in M from t to $t+5$, the procedure above treats the UNPD's projected $M(t)$ and $L(t)$ as if they were independent of $N(t)$. Instead, one should project one 5-year interval at a time. Given the most recent estimate of $L(t)$ and $M(t)$, one should project $\widehat{N}(t) = a \times in(t) + b \times out(t)$ and $B(t)$ and $D(t)$, combine them to get the next $M(\widehat{t+5}) = B(t) - D(t) + \widehat{N}(t) + M(t)$ (and similarly for $L(\widehat{t+5})$) and then iterate.

The quantitative effect of ignoring that $L(t)$ and $M(t)$ depend on prior values of net migration is likely to be small, because the largest projected value of $\widehat{N}(t)$, namely, 37.48 million (for 2095-2099) is less than 3% of $M(2095) = 1329.32$ million and about 0.4% of $L(2095) = 8767.79$ million. The proportionate effect of correcting each future projected $L(t)$ and $M(t)$ for the projected $\widehat{N}(t)$ is likely to be at most a few percent, and the reciprocal effect on the next $N(\widehat{t+1})$ of making those corrections is likely to be orders of magnitude smaller. Nevertheless the projections should be done step by step.

The coefficients a for immigration and b for emigration must satisfy $a \geq 0$, $b \leq 0$. If unconstrained estimation of them using ordinary least squares had violated these constraints, it would have been necessary to impose the constraints. The observation that unconstrained least squares yielded coefficients with sensible signs lent some confidence to the models in the sense that the estimated $L(t)$, $M(t)$, and $N(t)$ (or $USA(t)$, $W-USA(t)$, and $N(t)$) made these signs natural.

These examples treated L, M, and W-USA as if each were a single country when, in fact, each region was a set of countries. Net migration numbers did not specify where immigrants originated and where emigrants went. Using Rogers' biregional approach partially avoided this difficulty because migrants to one region must have come from the other. However, the exponents α , β , γ , δ were based on flows to individual more developed countries from individual less developed countries and vice versa (Kim and Cohen, 2010), while some migrants to or from USA may have come from, or gone to, other more developed countries. To refine this approach, it would be desirable to estimate country-specific flows based on the UNPD's rapidly growing database on migrant flows (UNPD, 2009), or on time series of migrant stocks (Abel, 2012), or on some combination of these approaches. These estimates can be smoothed by using the linear model, which posits $\alpha = 1$, $\beta = 0$, $\gamma = 0$, $\delta = 1$, or by re-estimating a gravity model for each country and the world outside that country. More complex versions of the gravity model can also take account of many other demographic, geographic, and historical attributes of origins and destinations.

If the methods used here for USA were applied to every country, the sum over all countries of each country's net migration might not be zero. To meet the logical requirement that the summed net migration of all countries must be zero, it would be necessary to adjust the initial estimates and projections of net migration to meet that constraint, perhaps by some kind of proportional redistribution.

To use these procedures in UNPD projections, it would be necessary to distribute net migration by sex and age, perhaps by using model schedules. It might then be necessary for some countries, especially small ones, to impose the constraint that every age group of each sex must remain non-negative.

These projections were deterministic. For stochastic projections, one could use the distribution of residuals (differences between observations or point estimates and modeled estimates) from Kim and Cohen (2010) or from the residuals here $\bar{N}(t) - N(t)$.

It would be highly desirable to validate these methods by excluding some recent estimates from the calibration and then comparing the projected net migration with those estimates.

The approaches illustrated here offer practical alternatives, based on explicit and testable analyses of historical estimates, to projections of net migration based on assumption. Future data will reveal whether the assumed future declines in net migration or the projected increases are more realistic.

E. ACKNOWLEDGMENTS

I thank Guy Abel and Patrick Gerland for very helpful comments on earlier drafts. This work was supported by U.S. National Science Foundation grant EF-1038337, the assistance of Priscilla K. Rogerson, and the hospitality of the family of William T. Golden.

REFERENCES

- Abel, Guy J. (2012). Estimating global migration flow tables using place of birth data. Vienna Institute of Demography, Working Paper 01/2012.
- Cohen, J. E., Roig, Marta, Reuman, Daniel C. and GoGwilt, Cai (2008). International migration beyond gravity: a statistical model for use in population projections. *Proceedings National Academy of Sciences* 105(40):15269-15274, October 7. www.pnas.org/cgi/doi/10.1073/pnas.0808185105
- Kim, K., 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. <http://onlinelibrary.wiley.com/doi/10.1111/j.1747-7379.2010.00830.x/abstract>
- Rogers, Andrei (1995). *Multiregional Demography: Principles, Methods and Extensions*. John Wiley, New York & London.
- United Nations Population Division (2009). *International Migration Flows to and from Selected Countries: The 2008 Revision*. CD-ROM and CD-ROM Documentation. Department of Economic and Social Affairs, New York. POP/DB/MIG/Flow/Rev.2008
- United Nations Population Division (2011). *World Population Prospects, the 2010 Revision*. United Nations, New York.

Table 1. Net migration to more developed regions from less developed regions during 1950-54 to 2095-2099. Columns: year = initial year of quinquennium, e.g., 1950 means 1950-54. netmig: estimates of UNPD WPP 2010 (millions of net migrants per 5-year interval). projected netmig WPP2010: UNPD *World Population Prospects* 2010 projections (millions of net migrants per 5-year interval). less: aggregate population of less developed region (millions). more: aggregate population of more developed region (millions). in: $in(t) = L(t)^\alpha M(t)^\beta$. out: $out(t) = L(t)^\gamma M(t)^\delta$. Gravity predicted netmig and Linear predicted netmig are alternative projections (2010-2095) based on smoothing of UNPD estimates (1950-2005).

year	netmig N(t)	projected netmig WPP2010	less L(t)	more M(t)	in	out	Gravity predicted netmig	Linear predicted netmig
1950	0.32		1721.04	811.19	12792.25	9221.40	0.85	0.76
1955	-0.45		1910.95	861.93	14318.71	10156.25	1.36	1.30
1960	3.00		2125.08	913.33	16018.71	11163.14	2.02	1.98
1965	4.14		2368.86	964.15	17910.95	12236.87	2.87	2.85
1970	6.12		2689.77	1006.42	20160.69	13363.45	4.25	4.28
1975	6.08		3030.16	1046.26	22507.67	14494.60	5.79	5.85
1980	5.64		3371.91	1081.09	24813.03	15559.67	7.43	7.49
1985	7.43		3750.34	1112.95	27283.62	16641.31	9.33	9.39
1990	11.89		4162.02	1144.40	29930.80	17763.71	11.46	11.49
1995	13.82		4556.79	1169.45	32391.33	18755.42	13.56	13.57
2000	17.45		4933.96	1188.81	34662.82	19623.13	15.63	15.61
2005	16.56		5295.75	1210.90	36902.00	20502.68	17.60	17.52
2010		12.52	5659.99	1235.90	39212.43	21428.93	19.59	19.41
2015		12.02	6028.12	1256.17	41457.22	22279.55	21.64	21.38
2020		11.45	6383.09	1273.44	43577.20	23056.63	23.65	23.31
2025		11.04	6716.24	1286.74	45505.21	23730.94	25.55	25.16
2030		10.55	7025.29	1296.09	47225.63	24298.73	27.34	26.90
2035		10.16	7309.47	1302.40	48751.18	24774.64	28.99	28.54
2040		9.8	7567.16	1306.89	50100.02	25178.79	30.49	30.03
2045		9.48	7796.07	1309.96	51271.25	25516.96	31.82	31.37
2050		8.06	7994.40	1311.73	52260.14	25790.27	32.98	32.54
2055		6.76	8163.30	1311.61	53058.72	25989.91	33.97	33.56
2060		5.74	8304.85	1310.35	53695.70	26133.18	34.79	34.42
2065		4.87	8422.01	1309.19	54217.34	26248.13	35.48	35.14
2070		4.08	8517.93	1309.18	54666.04	26359.14	36.03	35.72
2075		3.31	8594.70	1310.77	55064.31	26477.80	36.48	36.16
2080		2.61	8654.55	1313.99	55425.04	26608.28	36.82	36.48
2085		1.92	8701.15	1318.46	55756.14	26747.70	37.09	36.71
2090		1.25	8738.35	1323.74	56064.19	26891.90	37.31	36.87
2095		0.57	8767.79	1329.32	56344.02	27033.18	37.48	36.98