Centre for Economic Policy Research
POLICY INSIGHT No. 17
DECEMBER 2007

Introducing a Global Database on Migrant Stocks

American in Paris – But How Many Are There?[1]

L. ALAN WINTERS
University of Sussex and CEPR

Introduction

Every day in America or Europe there is a story about migration policy, and every other day there is a story about how little information we have on the number of migrants. In fact for most rich countries we know roughly how many there were at the turn of the century, but for most countries, especially the smaller ones, we hardly know at all. This *Policy Insight* introduces a new database that helps to address this issue, by quantifying how many people from each country resided in each other country around the year 2000. This is interesting in itself – how many Americans are there in Paris? – and allows us to test some simple hypotheses. Even more importantly, it provides a factual base on which researchers and policy-makers can start to address questions about global migration and policy. Perhaps surprisingly, it is the first such database at a global level.

Patterns of Migration

Around 2000, there were approximately 176 million international migrants – people living in a country other than that in which they were born. 12 of the 226 countries recognised in UN population statistics account for half of world immigration, starting with the USA, which is home to nearly 35 million. Table 1 reports the top 12 migrant recipient and sending countries in absolute terms along with the percentages of their resident populations accounted for by immigrants and the percentages of the people born in them (and still living) who live abroad. It excludes countries from the Former Soviet Union (FSU), three of which (Russia, Ukraine and Kazakhstan) are within the top 12, because many of their 'migrants' have never moved – rather their countries moved! As a single country, the USSR had consid-

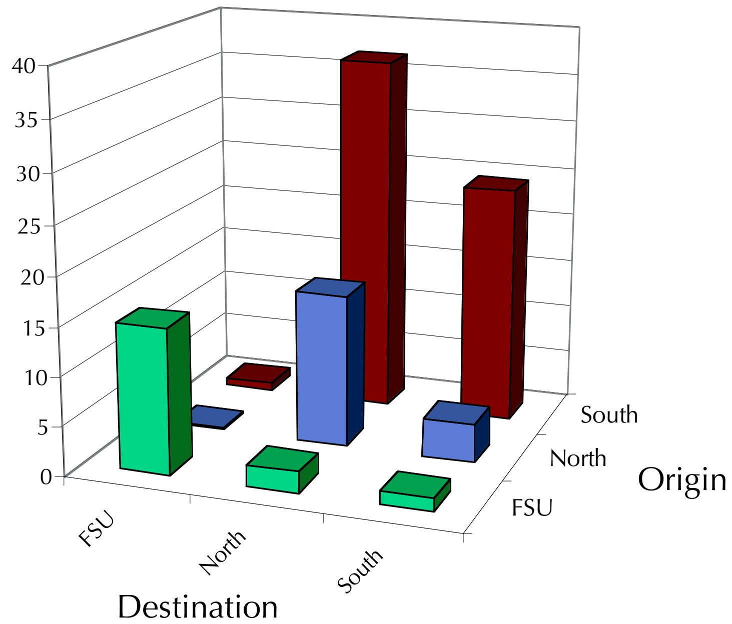
1 This *Policy Insight* is based on 'Quantifying International Migration: A Database of Bilateral Migrant Stocks', by Christopher R. Parsons, Ronald Skeldon, Terrie L. Walmsley, and L. Alan Winters, Chapter 1 of C Ozden and M Schiff (eds.) *International Migration, Economic Develoment and Policy*, Palgrave and The World Bank, 2007, pp. 17-58. The research was initiated and supported by the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex that is supported by the Department for International Development, United Kingdom. The data are available online

Table 1 Principal Recipient and origin countries

	Countries of immigration*			Countries of emigration*	
	Number (million)	as % of pop.		Number (million)	as % of pop.
United States	34.63	12.48	Mexico	10.10	10.01
Germany	9.14	11.15	India	8.96	0.87
France	6.28	10.55	Bangladesh	6.64	5.03
India	6.27	0.61	China	5.79	0.46
Canada	5.72	18.30	UK	4.19	7.08
Saudi Arabia	5.25	23.03	Germany	4.05	4.94
UK	4.87	8.21	Pakistan	3.39	2.39
Pakistan	4.24	3.00	Philippines	3.39	4.26
Australia	4.07	20.97	Italy	3.28	5.71
Hong Kong	2.70	37.74	Turkey	3.00	4.53
Cote d'Ivoire	2.34	15.38	Afghanistan	2.70	9.89
Iran	2.32	3.57	Morocco	2.61	8.96
Total	87.84		Total	58.10	

Note: *The table excludes former USSR countries.

Figure 1 Bilateral migration shocks (% of world total)



erable internal mobility (some forced), so when it split up, many of the people found themselves living in a country other than that of their birth.

At the turn of the century, around 12.5% of the United States' population was foreign-born. Australia, Canada and Saudi Arabia show higher percentages and Western Europe not much smaller. Among developing countries, India and Pakistan have high immigration, as did Côte d'Ivoire and Iran on the census date. Turning to countries of emigration, Mexico, Afghanistan and Morocco show major outflows in proportionate terms and India and Bangladesh in absolute terms. Rich countries are also important sources, with the United Kingdom, Germany and Italy within the top 12. Because we use an absolute criterion, all the countries in Table 1 are large. Large countries generally show proportionately much less migration than small countries, several of which have more people abroad than at home or more immigrants than locals among the resident population.

It is interesting to ask how many of these migrants come from or go to neighbouring countries. Even putting island states (which have no land neighbours) aside, experience is very mixed. For example, in India and Iran, over 90% of immigration is from neighbouring countries, whereas Pakistan receives most of its immigrants from Bangladesh and has only 17% from contiguous countries. The flow from Mexico to the USA accounts for about 93% of Mexican emigration, but only 30% of US immigration.

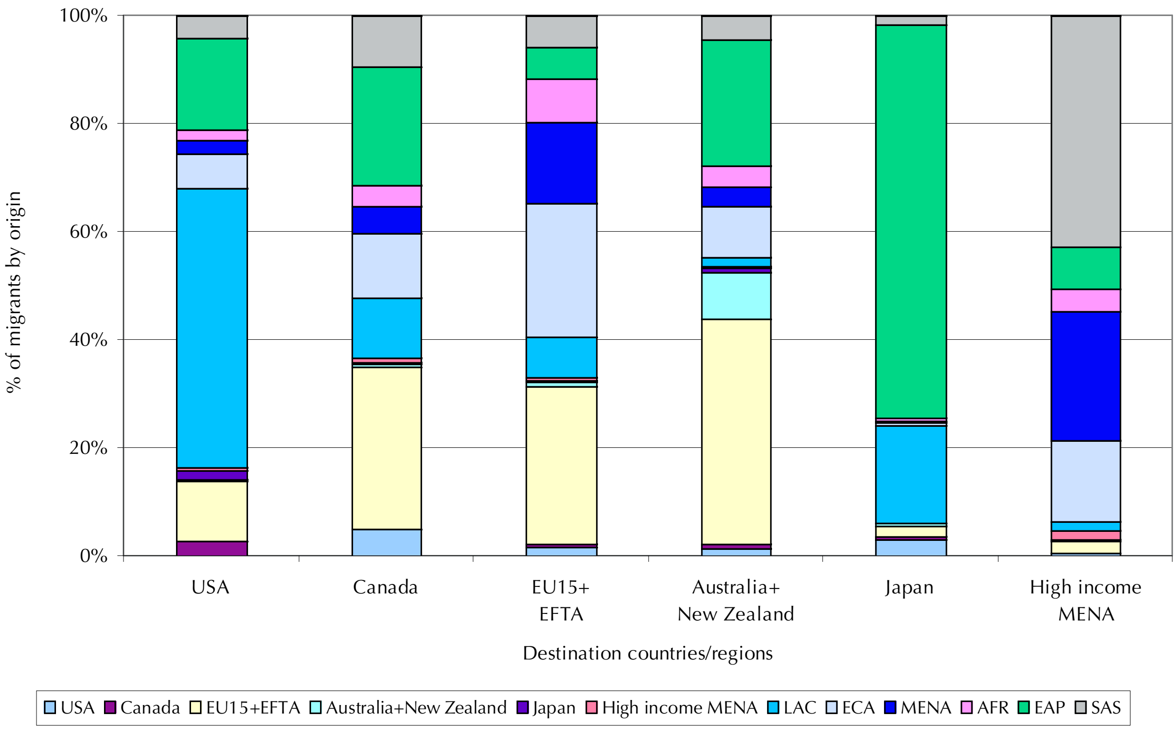
Figure 1 summarizes the regional data, distinguishing North (rich), South (poor) and the FSU countries. The bulk of FSU migration is intra-regional and probably largely spurious. For Northern emigrants, the principal destination is other Northern countries. This reflects mainly the propensity of Europeans to migrate not only within Europe (5.6% of the world total), but also to North America and Oceania. Southern emigrants, on the other hand, go more to Northern destinations than to Southern ones and outnumber Northern migrants in those countries. But as recipients, Southern countries get most of their immigrants from other poor countries. Overall, South-to-North emigration accounts for 37% of total emigration, South-South for 24% and North-North 16%.

Even with a more detailed regional breakdown, much migration remains intra-regional. [Here we explore the data at a 15-region level – based on World Bank standard regions – but the beauty of the dataset is that one can do it all at a 226-country level if one has the space.] After the huge but largely spurious share within Developing Europe and Central Asia (FSU), the largest regional share is the stock of Latin Americans (including Mexicans) in the USA (10% of the world total), followed by intra-Sub-Saharan African (7%) and intra-South Asian (6%) migration. The latter two show clearly that South-South migration is quantitatively significant.

Correcting for the different sizes of the different regions by calculating migration propensity indices reinforces this view. These express country *i*'s share of *j*'s immigrants relative to *i*'s share of world migration[2]. Every intra-regional propensity exceeds unity, often greatly. Australia and New Zealand have mutual shares 16 times larger than their world size would suggest, perhaps reflecting the integrated labour market created under the Closer Economic Relations Agreement, or

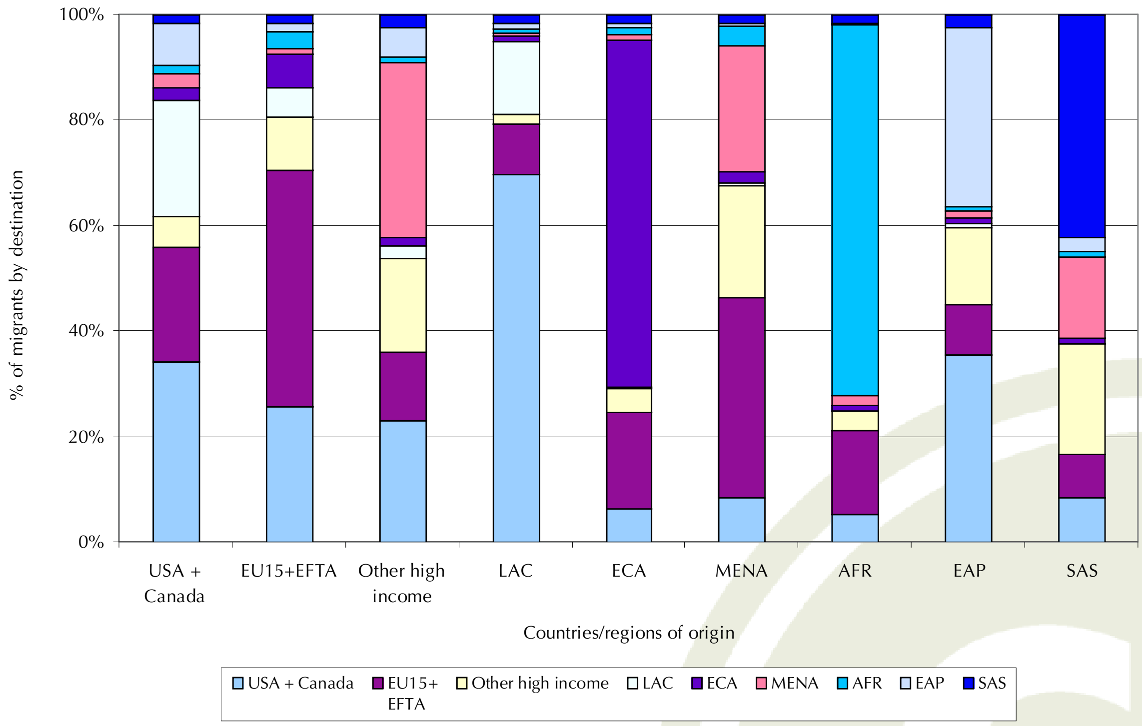
2 Migration intensity indices are calculated as $(X_{i,j}/X_{.,j})/(X_{i,.}/X_{.,.})$ where, $X_{i,j}$ is the stock of migrants from *i* in region *j* and a dot (.) denotes summation across the corresponding subscript. This expresses the share of *i* in *j*'s stock of immigrants relative to *i*'s share of world total immigrants. Thus an intensity index exceeding one implies that *i* is a more important source for *j*'s immigrants than it is for world immigrants on average. (The intensity index also equals *j*'s share of *i*'s stock of emigrants relative to *j*'s share of world emigrants. Hence, a value above 1 indicates that *j* is a more important destination for *i*'s emigrants than for source countries on average.) If bilateral flows were random, so that all emigrants from all countries had an equal chance of ending up in, say, Europe, and an equal (albeit different) chance of ending up in, say, Africa, the intensity indices would be unity. If they exceed unity for a bilateral link, they indicate that there is a bias towards that link.

Figure 2 Source country shares of certain destination regions



Notes: Regions: EU European Union; EFTA European Free Trade Area; MENA Middle East and North Africa; LAC Latin America and the Caribbean; ECA Europe and Central Asia; AFR Sub-Saharan Africa; EAP East Asia and the Pacific; SAS South Asia.

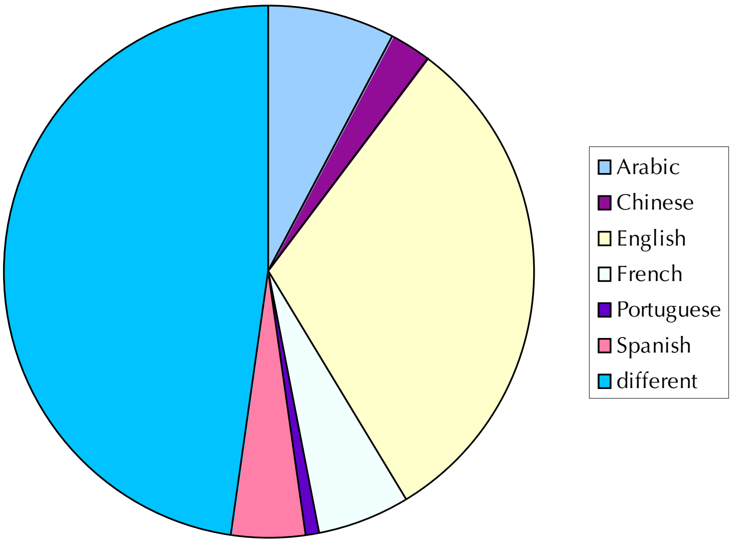
Figure 3 Destinations of emigrants from source regions



maybe just their isolation. In second place is intra-African migration (propensity index of 8), probably a reflection of the facts that (a) much of African migration is forced (refugee flows) and by poor people, so proximity is crucial; and (b) Africans generally find great difficulty in entering other countries. Among the inter-regional propensities of note are the strength of the East Asian-Japan link; the close links between the USA and Canada; the close links between the USA and Latin America (both ways); the Old World-New World flows within OECD, and the strong flow from South Asia to the Gulf. These patterns are not all surprising, but it is useful to have them explicit and their quantification will, we hope, allow a good deal of subsequent research to explain them.

The differences in patterns between regions are illus-

Figure 4 Migration and common language



trated in Figures 2 and 3. For each developed (Northern) region, Figure 2 gives the breakdown of its immigration by origin. Thus we see Europeans accounting for significant shares of immigration in Europe, Canada and Australia, but nowhere else. For the USA, the dominant origin is Latin America, accounting for 52% of the total, and that region is also important in Japan's (admittedly small) total. The latter reflects the special treatment offered to Latin Americans of Japanese ethnic origin. Even more striking is the share of immigrants from East Asia and the Pacific in Japan's total immigrant stock, although in absolute numbers Australia and New Zealand are more important destinations for that region. Among high-income regions, only the Gulf has a high share of South Asians and, as already hinted, nowhere at all has a high share of Africans.

Figure 3 offers the obverse view – the choice of destinations for each source region. For Northern regions, the principal destinations are Northern: fully 80% for Europe. On the other hand, because Europe's outflow is so much larger than that of the other Northern regions, its absolute supply of people to each of the developing regions is larger than theirs. For developing region sources, the principal destinations are intra-regional, especially for ECA and Africa. But beyond that, we see different degrees of reliance on the Gulf (high for SAS and MENA and low for ECA and AFR), Europe (high for ECA and relatively low for EAP and SAS) and North American (low except for LAC and EAP). It is not difficult to see the role of geography in the creation of these corridors, but equally well one can see that geography is not all.

The bilateral data allow one to explore a number of possible factors behind migration, descriptively. For example, for 163 countries we have been able to identify a predominant international language: either Arabic, Chinese, English, French, Portuguese or Spanish. 51% of migration takes place within these language blocks – most strongly the English one, but by no means exclusively so. This is as true of South-South flows as of North-South or North-North.

A second factor frequently spoken of is distance. This is clearly related to the question of migration to and from neighbours and intra-regional flows discussed above, but even for non-contiguous countries, distance is often felt to matter. The mean distance for all migrants in the world is a formidable 4,225 kilometres, but within that there are clear negative correlations between distance and the number of migrants.

A third influence on migration is income. This, of course, is an individual or household matter, dependent on the individual's circumstances before and after migration and so country data are only very crude indications. However, attributing to every migration an income difference equal to the difference in average gross national income (GNI) per capita in 2001 between the source and destination countries, the mean gain in income is $9,479.[3] Similarly for every intra-regional flow the mean is positive.

Where do these data come from?[4]

Describing patterns is easy once you have the data, but where do these come from and how reliable are they? The rest of this note describes them; they are, we believe as good as migration data get, but they are far from perfect.

The Sussex database comprises four versions of two origin-destination matrices for 226 countries and territories. The first matrix records foreign population defined by country of birth and the second by nationality or citizenship. Although countries employ many different methods for collecting and presenting these data, my colleagues and I have attempted to reconcile the different sources to create as full and comparable a dataset as possible. The four versions reflect different points on the trade-off between the comprehensiveness of coverage and the use of assumption and interpola-

3 These figure far exceeds the likely gain for any family: factors other than labour receive some of the GNP pc and in most cases migrants start with above average incomes in their home countries and at least for several years do relatively less well in their new location.

4 This is a brief summary. The original source describes the data and their derivation in excruciating detail.

tion to compensate for missing data. The first simply contains the raw data with some minor adjustments while the later ones gradually add more bilateral entries based on different degrees of approximation. The final version (version 4), which covers only the foreign-born, represents as full a picture of bilateral international migrant flows as data permit. It is the source for the analysis above.

Recall that the data refer to the stock of migrants by destination country and territory disaggregated by country and territory of origin around the year 2000. They make no reference to the time at which a migration has taken place, but provide only an estimate of the cumulative migrations to date into an area (net of re-emigrations).

Compiling comparable data

There are several definitions of "migrant", but we opt for people living in a country other than that of their birth. The common alternative is by nationality, but given that this can be changed on criteria and timescales that vary hugely across countries, it is less reliable. Birth also corresponds more closely to the notion of international movement: defined by country of birth, a person will generally have had to move at some point in their life to be classified as a migrant.[5] The differences are particularly large when dealing with former colonies. For example, many Portuguese nationals are born abroad in one of Portugal's former colonies. The Portuguese census reports nearly 50,000 people born in Brazil, but only 32,000 with Brazilian nationality.

Immigrant stocks are usually recorded by demographic methods that measure the total population, and the most important of these are the population censuses. Virtually every country in the world conducts censuses, but they are infrequent and carried out on different dates within census "rounds" that usually span a decade. The latest (2000) round of censuses that included nearly all countries[6] includes censuses taken between 1995 and 2004.[7] They use one of two definitions of the population: the *de facto* population refers to all persons physically present in a country at the census moment, while the *de jure* population refers to all those persons who either are usually resident, or who qualify as legally resident, at the census moment. They are often defined as those who have been resident in that country for a particular length of time, which can range from between 3 months to 1 year.

An alternative source for migrant stocks is the population register. These are continuous reporting systems for recording births, deaths and changes of residence in a population and so offer data quite different from those in censuses. However, few countries maintain population registers and those that do rarely offer a comprehensive view of either country of birth or nationality. Data from registers are employed in the database only where appropriate census data are unavailable.

There are many disparities in collection practices across countries. These include surveying on different dates, omitting different categories of people, defining migration differently, a lack of standardization between the questions asked during the census, and alternative country coding used to record the responses. There are also bound to be errors – either deliberate or inadvertent. For example, people may not be aware of changes in country borders and hence misreport their country of birth. Illegal immigration also poses a challenge. The extent to which censuses record this will vary across countries, not least with the likelihood that census returns could be used to detect and apprehend people without proper authorization. Even in the best of circumstances, e.g. the USA census, some proportion of migrants is likely to remain unrecorded, not least because it is frequently in neither the migrants' nor the authorities' (political) interests to know that they are there.

Our aim was to include as many of the world's migrants as possible in the dataset and to assign them *all* to specific countries of residence and origin. In most cases, data were recorded from their original source – the latest census wherever possible – and data on both foreign-born and foreign nationals were compiled where feasible. Population registers were then drawn upon where censuses from the 2000 round were unavailable. In some cases where neither source was available, data were obtained from reliable secondary sources that cite the original. Some regions of the world provide good data while for others data simply do not exist in the public domain or even at all. Thus while the data for Europe, the Americas and much of Oceania are of a fair standard, the data for parts of Asia and much of Africa are of more dubious quality. The original paper gives details of sources.

The versions of the data matrix

Version 1 of the matrices contain all the raw data collected with a few adjustments. No attempt is made to assign residual, 'unknown' and "Other" categories to sources. Where censuses used country coding referring to the USSR, Czechoslovakia or Yugoslavia, and where bilateral information on migrants' destinations post break-up was available, these aggregated totals were distributed on the same basis, implicitly assuming that migration proportions were the same before and after the break-up. Dependencies and re-coded countries were also aggregated up into one of the 226 countries and territories included in the database. We manage to get about 163 out of the 176 million migrants into this version of the matrices.

In version 2, composite or regional origins for which no country-specific migration data were available

5 The caveat is that sometimes country borders are moved, creating apparent migrants out of static households, as we argued above for the FSU.

6 Those not taking part include: Afghanistan, Colombia, Peru, North Korea, Myanmar, Bhutan, Taiwan, Uzbekistan, Tajikistan, Moldova, Bosnia and Herzogovina, Western Sahara, Guinea-Bissau, Liberia, Togo, Nigeria, Chad, Cameroon, Gabon, Sudan, Ethiopia, Eritrea, Somalia, Angola, Democratic Republic of Congo, and Madagascar. (Source: United Nations.)

7 Although the vast majority of countries conducted censuses during this round, many have still to process and publish the data, so we do not have comprehensive coverage yet.

Table 2 Database versions

Version	Number of migrants treated	Information contained	Main countries whose status is changed from previous version
1	FB 108.5m Nat 55.3m	• Raw data collected including older primary sources where later information unavailable.· • Meaningless "unknown" totals omitted. • Those countries where totals reported prior to break-up redistributed according to bilateral migration stocks post break-up. • Aggregated in dependences. • Entered zeros where applicable.	
2	FB 111.0m Nat 56.2m	• Separated jointly reported nations, and those prior to break-up where no post break-up migration data available, according to population shares.	**Disaggregated according to subsequent migration stats:** Germany, Italy, Canada, Denmark, Sweden, Finland
3	FB 111.0m Nat 57.6m	• Removed ethnic nationalities with little or no correlation to states or regions. • Added additional DFID figures on the number of Indians residing in Middle Eastern Economies. • Removed ambiguous and "ignored" categories as these most likely account for domestic population and not migrants. • Removed those recorded with dual nationality.	**Removed meaningless ethnicities:** Bulgaria, Romania, Slovakia, Croatia, Belarus, Kazakhstan, Moldova, Turkmenistan, Tajikistan, Uzbekistan **Removed unknown and ignored figures:** Argentina, Netherlands, Sweden, Thailand, Greece, Bulgaria, Hungary **Added DFID figures:** Qatar, Oman, Bahrain, Kuwait **Removed dual nationality:** Liechtenstein
4	FB = Total = 175.7m	• Entered United Nations data for country of birth totals where data missing. • Used entropy measure to compare nationality and country of birth shares. • Having confirmed that the series were highly correlated, used the additional information content in the nationality matrix to supplement the foreign-born matrix with additional coefficients of interest. • Disaggregated remainder categories based on countries' propensity to send migrants abroad. • Used shares based on countries' propensity to send migrants abroad to fill all remaining bilateral coefficients. • Scaled data to United Nations (2004)	**Countries included where no data available previously:** China, Indonesia, Democratic People's Republic of Korea, Morocco, Algeria, Yemen **Countries that had nationality data used to supplement FB Matrix:** Japan, Philippines, Thailand, Vietnam, Italy, Mozambique

(including break-up countries) are disaggregated according to their shares in total population. Although this a pretty rudimentary approach, it refers only to small numbers of migrants, so it is probably not doing too much violence to the results overall. As table 2 below shows, about 3.4 million migrants were added to the database at this stage.

Version 3 removes nationality headings which use ethnic background to distinguish migrants, for example "Crimean Tatars". These categories proved difficult to assign geographically and were removed entirely from the dataset, and their numbers subtracted from the country totals. These had previously been included to facilitate research on ethnicity. Other headings referring to totals largely incorporating nationals were likewise removed. This included those persons who possessed dual nationality and ambiguous "ignored" totals. Additionally, Srivastava and Sasikumar (2005) was drawn upon to supplement the database with information on the estimated number of migrants from India abroad in the Middle East.

Version 4, is the fullest but least accurate set of data. This version supplements the foreign-born matrix with shares across sending countries derived from the nationality matrix, augments the foreign-born matrix with the United Nations (2004) totals for 2000 where no other data were available, and reconciles all of the remainder categories. Finally, it scales all the data predating 2000 to the United Nations (2004) total for that year. Relative to version 3, information on about 7.1 million new migrants was added at this stage and nationality shares for 56 million translated into birth shares. The decision to produce the final matrix utilizing the foreign-born definition was based on the fact that a greater number of countries report data by place of birth, that this definition is less vulnerable to differences in naturalization policies across countries and that it more readily accords to the actual movement of migrants, which is the economist's principal interest.

Conclusion

In spite of the inaccuracies due to the lack of bilateral data and the rough-and-ready methodologies used to infer them, we believe that the Sussex data provide a reasonably realistic view of migration stocks around 2000. They offer a base from which research into the determinants and consequences of global migration can be launched and a source of factual information that is of direct interest in itself. So, how many Americans are there in Paris? We can't say because we don't have city data, but in 2000 there were slightly over 204,000 Americans in France.

L Alan Winters is Professor of Economics at the University of Sussex. He is a Research Fellow of the Centre for Economic Policy Research (CEPR, London), and a Senior Visiting Fellow of the Centre for Economic Performance, at the London School of Economics. He has previously worked in the Universities of Cambridge, Bristol, Wales and Birmingham and as Research Manager for International Trade in the World Bank. In addition he has advised, inter alia, the World Bank, OECD, the European Commission, the European Parliament, UNCTAD, the WTO, DfID and the Inter-American Development Bank.

He specialises on empirical and policy analysis of international trade. His one hundred and seventy published books and articles cover areas such as regional trading arrangements, non-tariff barriers, European Integration, East-West trade, global warming, agricultural protection, trade and poverty, and the world trading system. He is currently working on trade and the location of industry in the UK and Europe (ESRC), trade liberalisation and poverty (DfID), the brain drain (DfID), temporary migration and the world trading system.

The Centre for Economic Policy Research (www.cepr.org), founded in 1983, is a network of over 700 researchers based mainly in universities throughout Europe, who collaborate through the Centre in research and its dissemination. The Centre's goal is to promote research excellence and policy relevance in European economics. CEPR Research Fellows and Affiliates are based in over 237 different institutions in 28 countries. Because it draws on such a large network of researchers, CEPR is able to produce a wide range of research which not only addresses key policy issues, but also reflects a broad spectrum of individual viewpoints and perspectives. CEPR has made key contributions to a wide range of European and global policy issues for over two decades. CEPR research may include views on policy, but the Executive Committee of the Centre does not give prior review to its publications, and the Centre takes no institutional policy positions. The opinions expressed in this paper are those of the author and not necessarily those of the Centre for Economic Policy Research.