WPS4165

# Quantifying International Migration:
# A Database of Bilateral Migrant Stocks[*]

**Christopher R. Parsons**
**Ronald Skeldon**
**Terrie L. Walmsley**
**and**
**L. Alan Winters**

***Abstract***

*This paper introduces four versions of an international bilateral migration stock database for 226 by 226 countries and territories. The first three versions each consist of two matrices, the first containing migrants defined by country of birth, i.e. the foreign-born population, the second, by nationality, i.e. the foreign population. Wherever possible, the information is collected from the 2000 round of censuses, though older data are included where this information was unavailable. The first version of the matrices contains as much data as could be collated at the time of writing but also contains gaps. The later versions progressively employ a variety of techniques to estimate the missing data. The final matrix, comprising only the foreign-born, attempts to reconcile all of the available information to provide the researcher with a single and complete matrix of international bilateral migrant stocks. The final section of the paper describes some of the patterns evident in the database.*

**Key Words: Migration data, bilateral stock data**

---

1

# 1.    Introduction

Economists have become increasingly cognizant of the importance of international migration over the last few years. In developed countries there has been intense debate about appropriate immigration policies (e.g Borjas, 2003), while among development economists interest has focused on the links between migration and development (e.g. Lucas, 2005, Ozden and Schiff, 2005). Topics include the determinants of migration (e.g. Hatton and Williamson, 2002), the role of migration and remittances in poverty alleviation (Ozden and Schiff, 2005, Part I), migration and human development (e.g. McKenzie and Rapoport, 2005, McKenzie and Hildebrandt, 2005), the brain drain (Stark, 2003, Commander, Kangasniemi and Winters, 2004, Docquier and Marfouk, 2005), South-South migration (e.g. Ratha, 2006) and the links between migration, trade and investment (Rauch, 2001). At a global level, research has explored the benefits from possible agreements to foster South-North migration (Walmsley and Winters, 2005, Pritchett, 2006).

Given this increasing interest in migration, it might seem surprising that few attempts have been made to take a global view of the patterns of international migration between countries and regions. This deficiency essentially reflects the lack of comprehensive and reliable data on bilateral flows of migration. OECD produces detailed stock and flow data on immigrants in OECD and a few other countries (e.g. OECD, 2002), early versions of which have been used for analysis – e.g Mayda (2006). The United Nations adopts a global perspective when summarizing international migrant movements, but its work provides only total migrant stocks for each country. National data sources frequently allow some geographical disaggregation, but attempts to combine them into a broader framework exhibit a number of common failings: for example, data availability only for a particular region and serious challenges of comparability from one country to another. Among the latter is the problem of how to define migrants by place of origin. This can most readily be done either by country of birth or by country of citizenship, but where attempts have been made, typically no distinction is drawn between the two concepts despite the large disparities that exist between them.

This paper makes a start toward rectifying these shortcomings of the data by introducing four versions of two origin-destination matrices for 226 countries and territories using the data from the 2000 round of censuses.[1] The first of these matrices records foreign population by country of birth and the second the population by nationality, which, for the purposes of the paper, is treated as analogous to citizenship.[2] Though countries employ many different methods for collecting and presenting these data, we attempt to reconcile these methods 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 interpolation to compensate for missing data. We report all four because the different versions will be appropriate for different uses, and we discuss

---

[1] For reasons of space, the matrices cannot be reproduced in this paper. The first two versions are available on the website of the Development Research Centre on Migration, Globalisation and Poverty at www.migrationdrc.org/research/typesofmigration/global_migrant_origin_database.html   The   other   two versions are available to researchers by request to the same centre or Version 4c only, aggregated to a 87x87 matrix, is available at: www.gtap.agecon.purdue.edu/models/labor_migration.asp

[2] United Nations (1980) defines a person's citizenship as their legal nationality. Following from this, no distinction is made here between citizenship and nationality, and thus no notion of belonging or of political or legal participation is employed.

them in as transparent a fashion as possible so that readers may choose the version which best suits their needs.

The first version simply contains the raw data with some minor adjustments and will be of more use to a demographer. The later versions contain more bilateral entries, and despite the fact that individual entries are less accurate (some, indeed being pretty crude approximations), they may well prove more useful to the applied economist. The final version (version 4), though covering only the foreign-born, represents as full a picture of bilateral international migrant flows as data permit. By introducing the data in this way, it is hoped that this paper will provide a pathway into a rich resource for further research.[3]

The data in the four matrices are generated by the stock of migrants at 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). At present, no global source exists for flow data in which immigration (or emigration) can be related to specific periods of time, for example, movements over the previous 1 year or 5 years. Flow data of this type are only available for a relatively few, generally more developed, countries. In a way, international migration statistics are not far from where data on internal migration were some 40 years ago.

The paper concludes with a brief description of the patterns that are revealed by this new data source. The most important finding is that patterns do indeed differ substantially across migration corridors. For example, immigration in the United States is dominated by Latin America, whereas Western European immigration draws heavily on Eastern Europe, Central Asia and the Mediterranean region. Over one-third of world migration is from developing to developed countries, with only about a quarter between developing countries. Intra-developed country and intra-FSU (former Soviet Union) flows each account for about 15 percent of the total. Over half of migration is between countries with linguistic ties. Africa accounts for 8 percent of Western Europe's immigration and much less of that in other rich regions.

The remainder of the paper comprises the following. Section 2 describes who qualifies as a migrant under both of the definitions utilized in this paper. Section 3 investigates the sources of migration statistics and their interpretation; it makes clear the heterogeneity that exists among the available statistics and the challenges entailed in making them comparable.[4] Section 4 outlines the different versions of the matrices and the final section provides some summary statistics of the patterns observed.

## 2. Who Counts as a Migrant?

Definitions of a "migrant" are various. International migrants may be recorded in terms of:
- country of birth
- country of citizenship
- last country of previous residence

---

[3] The database has undergone several rounds of improvement and extension. The first public version was described in Parsons et al (2005) and an intermediate version was used and briefly described in Ratha (2006).
[4] Our treatment is far from comprehensive. See Bilsborrow et al (1997).

- duration of time spent away from birthplace or last place of previous residence[5]
- purpose of their stay (visa type)

The above classification comes from Bilsborrow et al (1997), who also discuss other possible definitions such as those based on ethnicity. The United Nations (1998) defines a migrant as "any person who changes his or her country of usual residence", though residence may refer to a change both of residence or of residential status. Tourists and business travellers, among others, are therefore not included in the international migration statistics, as their movements do not involve changing their usual place of residence. Tourists are often recorded in separate statistics and business visitors are often not recorded among the migrant population.

Statistically, the migrant population can be equated directly with the number of foreigners: either those recorded by country of birth, the foreign-born, or that fraction of the population with foreign nationality, the foreign population. The confusion over migrants is exacerbated, however, by the fact that a migrant's nationality need not be the same as their country of birth. Nationality can be changed; birthplace cannot (not legally, at any rate). Thus, one may be born locally but qualify as a foreign national if born to foreign parents, or conversely, one can be born a national but still be characterized as foreign-born if born overseas. Table 1 summarizes the various possibilities of what constitutes the foreign population under each definition, taking the examples of Ireland and the United Kingdom. Only domestic nationals residing in their home country of birth do not qualify as immigrants under one or other of the definitions. A third possibility is that one can adopt the nationality of one's country of residence after some time via naturalization procedures.

Given the variety of nationality criteria and the differences between countries in their willingness to naturalize immigrants, we find the foreign-born criterion more useful for economic analysis, and in our final version of the data, we make use of nationality statistics only where foreign birth data are unavailable. It is also the case that the former 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.[6]

---

[5] Though it is increasingly recognized that distinguishing between shorter and longer term migrants is useful for policy making etc. this division is ignored here since censuses generally fail to make this distinction. Indeed, countries seldom accurately distinguish these alternative series.

[6] The caveat is that sometimes country borders are moved, creating apparent migrants out of static households.

## Table 1. Who qualifies as a migrant?

|  | UK born UK nationality | UK born Irish nationality | Irish born UK nationality | Irish born Irish nationality |
|---|---|---|---|---|
| **Residing in UK** | DOES NOT QUALIFY AS MIGRANT | **Not** foreign population by country of birth | Foreign population in UK by country of birth | Foreign population in UK by country of birth |
|  |  | Foreign population in UK by nationality | **Not** foreign population by nationality | Foreign population in UK by nationality |
| **Residing in Ireland** | Foreign population in Ireland by country of birth | Foreign population in Ireland by country of birth | **Not** foreign population by country of birth | DOES NOT QUALIFY AS MIGRANT |
|  | Foreign population in Ireland by nationality | **Not** foreign population by nationality | Foreign population in Ireland by nationality |  |

One problem arising from these definitional issues is encountered when dealing with former colonies. For example, many Portuguese nationals are born abroad in one of Portugal's former colonies. As outlined in table 2, the Portuguese census reports nearly 50,000 people born in Brazil, but only 32,000 with Brazilian nationality. Care should therefore be exercised since a large proportion of foreign-born migrants will have a domestic nationality.

## Table 2. Disparities between the foreign-born and foreign migrants in Portugal

| Former colony | Migrants from colony defined as foreign-born | Migrants from colony defined as foreign citizens | Migrants born in the colony but with Portuguese nationality |
|---|---|---|---|
| Brazil | 49,891 | 31,869 | 18,022 |
| Mozambique | 76,017 | 4,685 | 71,332 |
| Angola | 174,210 | 37,014 | 137,196 |
| Cape Verde | 44,964 | 33,145 | 11,819 |
| Guinea-Bissau | 21,435 | 15,824 | 5,611 |
| Sao Tome and Principe | 12,490 | 8,517 | 3,973 |
| Macau | 2,882 | 71 | 2,811 |
| Timor Leste | 2,241 | 137 | 2,104 |
| **Total** | **384,130** | **131,262** | **252,868** |

# 3.    About the Data

## *Sources of Migrant Stock Data*

Immigrant stocks[7] are usually recorded by demographic methods that measure the total population, and the most important of these are the population censuses. The census, a

---

[7] This paper omits detailed discussion of immigrant flows. Readers are referred to OECD 2002a and b and 2003.

retrospective system for surveying the entire population at a single point in time, is generally considered the most comprehensive record of the total population. Perhaps the greatest strength of census data for measuring migrant populations, besides their universal coverage, is the limited scope of the questions asked, which facilitates comparability across countries. However, censuses are infrequent and although virtually every country in the world conducts censuses, invariably they are carried out on different dates within census "rounds" that usually span a decade. The latest, 2000, round of censuses that included nearly all countries[8] covers censuses taken between 1995 and 2004.[9]

Two definitions of the total (resident) population are commonly applied in censuses: de facto and de jure. The de facto population refers to all persons physically present in a country at the census moment. 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. Traditionally, censuses define migrants either as the foreign-born, or as the foreign population, although in recent years a growing recognition of the value of obtaining both sets of data has emerged.

An alternative source for calculating migrant stocks is the population register, although these are more often used to measure flows. Population registers 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 this study only where appropriate census data were unavailable.

Of course, in an ideal world, we would like to have flow as well as stock data on migration. These can be inferred from census data through the inclusion of questions asking where people used to live in the past. Unfortunately, these data are not available for enough countries to give a broad coverage of origin and destination flows for specific time periods. A second potential source of flow data is immigration records, which provide a continuous register of all legal migrants into a country within specific periods of time and are usually published annually. Again, however, the coverage of such records is patchy and the information collected and processed highly variable, so that in the end they are not a basis for a global dataset. Even developed countries rarely maintain a record of those leaving, irrespective of whether they are natives or foreigners. Looking to a utopian future, immigration and emigration records could provide an ideal source for the study of international migration, especially if they are linked to migrants' domestic records, as is increasingly called for by security services, but there is still a long way to go!

---

[8] 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, http://unstats.un.org/unsd/demographic/sources/census).
[9] It must be emphasized that, although the vast majority of countries conducted censuses during this round, many have still to process and publish the data.

### The Comparability of Migration Statistics

Compilation of comparable data across countries is hampered by the many disparities in collection practices. These include surveying on different dates, omitting different categories of people,[10] defining migration differently, a lack of standardization between the questions asked during the census, and alternative country coding used to record the responses.[11] The United Nations (1980) recommends that migrants should be enumerated according to the borders in existence at the time of census or survey. Over time, however, borders are redefined, and new countries are created; those interviewed may simply be unaware of boundary changes and misclassify their country of birth. Indeed a foreign-born migrant, resident for many years, may classify themselves as native even when they should in fact be otherwise classified, thus biasing results. In other instances, the non-response level of migrants may be very significant as census takers are often poorly trained and many mistakes will go undetected. During the 1969 Kenyan census, for example 78,756 individuals did not respond to the place of birth question, a high figure relative to the 158,692 that stated they were foreign-born (Bilsborrow et al 1997).

The problem of measuring illegal immigration is also considerable and the omission of illegal immigrants will distort official migration figures. The extent to which census data record them will vary across countries, not least with the likelihood that census returns could be used to detect and apprehend people without proper authorization. A pertinent example is Russia, a "migration magnet" in recent years and the country with the longest border, with approximately 450 official border posts. Feasible calculations estimating the number of illegal migrants residing in the Russian Federation range from 3.5 to 6 million persons (Heleniak 2002), though there is no real way of ever accurately knowing. In the United States, which arguably conducts the most comprehensive census, the best estimate of the number of unauthorized persons is calculated using a residual method, i.e. subtracting known authorized persons away from the total number of foreign-born (Passel 2005). Numerous migrants are therefore likely to remain unrecorded, not least because it is not in their interest for the authorities to know that they are residing illegally. In some cases, it is also not in the political interests of the authorities to know or recognize the extent of the problem of illegal entrants.

The rate of naturalization also negatively impinges on the quality of migration statistics. This varies significantly among countries, and differences in a country's propensity to offer citizenship impacts potentially very seriously on migration statistics recorded by nationality. Naturalizations in South-East and Eastern Asia, for example, are rare; so much so, in fact, that in many cases foreign nationality data can be directly equated to the foreign-born. In these countries, first-generation migrants and additionally immigrant births contribute almost solely to the foreign population. Indeed, in the specific case of Japan, migrants of even older generations still need to apply for citizenship, a lengthy and arduous process. Owing to increased marriages between Japanese and foreigners, the number of naturalizations has increased steadily, though the figure is still comparatively very low. Conversely, in countries such as Belgium, where naturalizations are far more common, additional migrants and immigrant births probably contribute

---

[10]The United Nations notes, for example, that developed countries generally include refugees in their censuses whilst these are largely omitted from censuses in developing countries.
[11] For example, the number of Puerto Ricans in the United States will not be recorded since they count as Americans.

significantly to the domestic population as foreign citizens become naturalized. For our purposes, the disparity between two immigration systems will impact on the proportion of the foreign relative to the national population.

Finally, it should be noted that the way in which immigration statistics are measured has caused several discontinuities in the data, such as that resulting from the break-up of the FSU, Yugoslavia and Czechoslovakia. All of these events produced millions of additional international migrants overnight as new boundaries were drawn and new nationalities created. The data have been adjusted accordingly to include these new migrants. On the other hand, the return of Hong Kong to Chinese sovereignty in 1997 did not reduce the number of migrants: the Chinese are still regarded as "migrants" and contribute over 2 million (1 percent) to the United Nations total.

### *The Data Collected*

Fully aware of the challenges of data compilation, we now describe the new dataset. The aim of this study was to include as many of the world's migrants as possible, to assign them *all* to specific countries of origin with the highest degree of accuracy and to produce as full and comparable a bilateral database of international migration stocks as is possible. In its final form, the data should be easy to interpret, suitable for a wide range of academic purposes and easy to update should further material become available. In most cases, data were recorded from their original source. Data from the latest census round were preferred, as these were considered most comparable at the global level, although given the significant lags that often exist between the timing of censuses and their publication "latest" was not always very recent. Data on both foreign-born and foreign nationals were compiled where feasible. Population registers were then drawn upon where censuses were unavailable for the 2000 round of censuses.[12] In some cases where neither source was available, data were obtained from reliable secondary sources that cite the original.[13] Some regions of the world provide significantly better data than others, and some simply do not exist in the public domain or even at all. 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. Annex 1 provides a summary of the raw data collected, showing whether the data are bilateral and the corresponding sources and years.

## *4.* The Versions of the Data Matrix

As noted above, four versions of the data are discussed, each progressively more complete but more speculative than the last. Each of the first three versions of the database contains two matrices covering both emigration and immigration for each of 226 countries or territories,[14] one recording the foreign-born, the other the foreign

---

[12] In such cases, since the majority of censuses refer to the years 2000 and 2001 these years were prioritized; and where data were available for both 2000 and 2001, data for 2001 were preferred since it is more recent.

[13] These include the Organisation for Economic Co-operation and Development, the Migration Policy Institute, the Economic Commission for Latin America and the Caribbean, the Department for International Development, the International Labour Organization, and the United Nations, together with various national statistics bureaux.

[14] In addition to the 226x226 matrices, data aggregated into the 87 GTAP regions corresponding to the GTAP database version 6, may be obtained from the authors. This aggregation will likely smooth over some of the inaccuracies contained in later versions.

nationals. The fourth version covers only the foreign-born. Version 1 contains all the raw data collected with a few adjustments. In the sources drawn upon in this paper it is common for migrants that cannot be assigned to a specific country of birth/nationality to be recorded in remainder categories that typically have the prefix "Other". In version 1, there is no attempt to reassign these migrants and the "Other" category, which often contributes significantly to the overall total, is simply reported below the matrix in the relevant (host nation) column. In a few cases, large "unknown" categories were recorded in the original data, for example the German data collected from the OECD. These "unknowns" were completely removed from the matrices since it is unclear whether they actually refer to migrants, and will often contain members of the domestic population, since domestic populations are normally far larger than foreign populations. Where countries used country coding referring to the USSR, Czechoslovakia or Yugoslavia, and where bilateral information on migrants' destination post break-up was available, these aggregated totals were distributed on the same basis, implicitly assuming therefore that migration proportions were the same after break-up.[15] Zeros were entered where it could be inferred from the data that no migrants were assigned to a specific country. Dependencies and re-coded countries were also aggregated up into one of the 226 countries and territories included in the database.

In version 2 of the matrices, composite or regional origins for which no country-specific migration data were available (including break-up countries) are disaggregated according to their shares in total population.[16] Though a rudimentary method of estimation, as migrants are more likely to come from countries in close proximity at the regional level (Harrison, Britton and Swanson 2003), this method was deemed a reasonable exercise. As an example, the 1996 New Caledonia census recorded one category as "Melanesia", which included 86,788 unassigned migrants from Fiji, Solomon Islands and Papua New Guinea (PNG). Vanuatu was not included under "Melanesia" as it constituted a separate category in the census. Table 3 demonstrates how these migrants for New Caledonia were distributed across these countries in direct proportion to their populations. Table 4 outlines regions and their constituent countries.[17]

**Table 3. New Caledonia migrant split based on population shares for Melanesia**

| Country | Population (1996) | % total Melanesian population | Number migrants assigned |
|---|---|---|---|
| Fiji | 786,603 | 13.5 | 11,710 |
| Solomon Islands | 410,641 | 7.0 | 6,113 |
| PNG | 4,456,781 | 76.4 | 66,346 |
| **Total** | 5,829,952 | 100 | 86,788 |

---

[15] Both of these assumptions are open to challenge but, given the data availability, it is difficult to think of better alternatives than just making them and owning up.

[16] These were, by and large, small remainder categories that overlapped with larger "other" categories, those disaggregated in versions 4a-4c.

[17] Jointly reported countries, those recorded by name, for example "Botswana and Swaziland" are omitted from table 5.

**Table 4. Population splits**

| Region/Country Group | Split into: |
|---|---|
| USSR | Estonia, Latvia, Lithuania, Russian Federation, Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Republic of Moldova, Tajikistan, Turkmenistan, Ukraine and Uzbekistan. |
| Yugoslavia | Serbia and Montenegro, Bosnia and Herzegovina, Croatia, Slovenia, and Former Yugoslavian Republic of Macedonia. |
| CSFR | Czech Republic and Slovakia. |
| Dominicans | Dominica and Dominican Republic |
| Other United States island areas and Puerto Rico | Guam, Marshall Islands, Northern Marianas, Federated States of Micronesia, American Samoa, US Virgin Islands, Puerto Rico |
| Korea | North and South Korea |
| South Asia | India, Bangladesh, Sri Lanka, Pakistan |
| Middle Eastern countries | Turkey, Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Palestinian Territory, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, United Arab Emirates, Yemen, Cyprus |
| West Indies | Anguilla, Antigua and Barbuda, Barbados, British Virgin Islands, Dominica, Grenada, Guadeloupe, Guyana, Jamaica, Martinique, Montserrat St Vincent and the Grenadines, St Lucia, St Kitts and Nevis, Trinidad and Tobago and US Virgin Islands |
| Polynesia | American Samoa, Cook Islands, French Polynesia, Niue, Pitcairn Islands, Samoa, Tokelau, Tonga, Tuvalu, Wallis and Futuna Islands |
| Melanesia | Fiji, New Caledonia, Solomon Islands, Vanuatu, and PNG |
| South East Asian countries | Brunei, Cambodia, East Timor, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam |
| UK dependent territories | Anguilla, Bermuda, British Indian Ocean Territory, British Virgin Islands, Cayman Islands, Falkland Islands, Gibraltar, St Helena, Montserrat, Pitcairn Islands, Turks and Caicos Islands |
| Commonwealth Caribbean | Jamaica, Trinidad and Tobago, Dominica, St Lucia, Grenada, Barbados, Antigua and Barbuda, St Kitts and Nevis, Anguilla, British Virgin Islands, Cayman Islands, Montserrat, Turks and Caicos Islands, Bahamas |
| French Equatorial Africa | Chad, Gabon, Congo, Central African Republic |
| Micronesia | Federated States of Micronesia, Kiribati, Northern Marianas, Guam, Marshall Islands, Palau, Nauru |
| Other Western Asia | Armenia, Azerbaijan, Bahrain, Cyprus, Georgia, Iraq, Israel, Jordon, Kuwait, Lebanon, Palestine, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates and Yemen |
| Southern Africa | South Africa, Botswana, Namibia, Lesotho, and Swaziland. |
| Arab gulf co-op council | Bahrain, Oman, Qatar, Saudi Arabia, Kuwait and United Arab Emirates |
| Other Arab | Algeria, Comoros, Djibouti, Egypt, Eritrea, Iraq, Iran, Jordan, Lebanon, Libya, Mauritania, Morocco, Palestine, Somalia, Sudan, Syria, Tunisia, and Yemen. |
| Trust Territory of the Pacific | Marshall Islands, Federated States of Micronesia, Palau, Northern Marianas, |
| Overseas French Territories | French Polynesia, Mayotte, New Caledonia, St Pierre et Miquelon, Wallis and Futuna |
| Leeward Islands | British Virigin Islands, Anguilla, Guadeloupe, Netherlands Antilles, Antigua, Barbuda, St Kitts and Nevis, Montserrat, Guadeloupe, Dominica |
| Windward Islands | Barbados, Grenada, Martinique, Saint Lucia, St Vincent and the Grenadines Trinidad and Tobago |
| French West Indies | Martinique, Guadeloupe |
| US Indies | Puerto Rico, US Virgin Islands |
| Borneo | Indonesia (Kalimantan), Malaysia (Sabah, Sarawak), Brunei |
| English Caribbean | Anguilla, Antigua and Barbuda, Bahamas, Barbados, British Virgin Islands, Cayman Islands, Dominica, Jamaica, Montserrat, St Vincent and the Grenadines , St Lucia, St Kitts and Nevis, Trinidad and Tobago |
| Nordic Countries/Scandinavia | Norway, Sweden, Denmark, Finland |
| Indian Ocean | Reunion, Mayotte, Madagascar |

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, though the fullest, is arguably the 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. As each of these steps represents a large adjustment, version 4 is

11

split into three sub-components (versions 4a-4c). 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. These three iterations of version 4 are described in greater detail below.

### *Version 4a*

The 2003 United Nations data set (United Nations 2004), covering migrant stocks for the year 2000, is generally considered the most comparable source on international migrant stocks, despite the fact that some of the data were extrapolated from older sources. Indeed, it is still held in high esteem even though, where no foreign-born data are available, numbers are interpolated from data based on nationality and ethnicity concepts. Where no other data were available it was deemed sensible to use the United Nations data for country totals and our data are essentially calibrated to the United Nations totals.

A high degree of correlation has previously been understood to exist between the foreign-born and foreign-nationality populations (Harrison, Britton and Swanson 2003). If this is indeed the case, then it would be feasible to use information from the nationality matrix to supplement that of the foreign-born. First, it was important to check that this association existed in relation to the data collected. For those 35 countries where data were available for both the foreign-born and the foreign population for comparable years, a technique based on the entropy measure devised by Walmsley and McDougal (2004) was used (Annex 2).

This methodology calculates the differences in the shares between the foreign-born matrix and the nationality matrix as the equally weighted average of the simple share from a particular coefficient in the foreign-born matrix multiplied by the natural log of the ratio of that share, to the equivalent share in the nationality matrix. The two major advantages of this technique are first that it is possible to compare shares between two tables whilst ignoring any zero entries; and second, that the differences between two large shares are given less weight than the same absolute difference in two relatively smaller shares.

On average, the entropy difference between shares for the 35 countries for which data were available for both the foreign-born and the foreign nationality was negligible. Having found that the series were indeed highly correlated, the shares from the nationality matrix were multiplied by the United Nations totals in the foreign-born matrix for those countries where no bilateral information was available. South-East and East Asian countries[18] with detailed bilateral data were not subjected to this treatment as naturalizations are so rare that it is fair to assume that these data may be directly equated to the foreign-born population. We calculated the magnitude of the margin of error of treating foreign nationals as equivalent to the foreign-born to be approximately 1.22 percent for Japan, and 0.15 for South Korea.[19] The foreign-born coefficients for those 75 countries until now lacking any bilateral foreign-born information were therefore filled.

---

[18] Namely Japan, Korea, Macau, the Philippines, Thailand and Myanmar.
[19] These are calculated as the number of naturalizations divided by the total number of foreign nationals.

*Version 4b*

Residual categories, labelled with the prefix "Other", are common in the reported data (e.g. "Other Caribbean"). As the study aims to assign every migrant to a specific country, systematically these migrants, who numbered approximately 11.9 million and 9.6 million respectively under the two definitions of foreign-born and nationality, were distributed across the countries within the "other" region. This was done according to each country's propensity to send its nationals abroad relative to those of the other countries in "Other".

$$P_r = \frac{M_r}{\sum_{s \in other} M_s}$$

Where:  $P_{r/other}$ = Country r's propensity to send nationals abroad relative to the other countries located in the "other" region.
$M_r$ = The number of migrants abroad from country r.
$M_s$ = The number of migrants abroad from countries s in the other region.

These propensities are shares, so they sum to one. The remainder categories were therefore disaggregated by multiplying the propensity shares for those countries contained within an umbrella remainder heading by the total number of migrants to be redistributed. This process relies on a specific ordering, and a strict ordering was adhered to, in which, for each successive allocation, the largest number of migrants was used to recalculate the propensity shares. Therefore the smaller, regional remainder categories (e.g. "Other Caribbean") were disaggregated prior to the larger categories (for example "Other"). As most of the remainder headings refer to continental groupings, users who choose to aggregate the data by continent would avoid most of the approximation introduced here.

An illustrative case is that of Portugal. All migrants in the 2001 Portuguese census had been designated to specific countries with the exception of 6 migrants in the heading "Other Oceania". These remaining 6 migrants had to be shared out between the countries of Oceania. Applying the equation above, the propensity to send migrants anywhere abroad was calculated for each of these countries. Table 5 contains these propensity shares and shows how many immigrants were distributed based on these numbers for each country. This was assumed to be the most pragmatic method for distributing the remainder categories, though two weaknesses are evident. First, migrants are not distributed as integers and thus rounding is required. Second, the propensity to send migrants abroad is based on the sum of the relevant row (the total number of migrants abroad from any particular country). If it happens that a specific country sends many people to a country for which no data were collected, its propensity will be underestimated.

**Table 5. Propensity shares for "Other Oceania" split for Portugal**

| Country | Propensity share | Share*No. migrants to be assigned (rounded) |
|---|---|---|
| Australia | 0.25863 | 2 |
| New Zealand | 0.36202 | 2 |
| American Samoa | 0.02795 | 0 |
| Cook Islands | 0.01569 | 0 |
| Fiji | 0.09991 | 1 |
| French Polynesia | 0.00064 | 0 |
| Guam | 0.05765 | 0 |
| Kiribati | 0.00170 | 0 |
| Marshall Islands | 0.00606 | 0 |
| Federated States of Micronesia | 0.01331 | 0 |
| Nauru | 0.00066 | 0 |
| New Caledonia | 0.00112 | 0 |
| Norfolk Island | 0.00026 | 0 |
| Northern Marianas | 0.00518 | 0 |
| Niue | 0.00458 | 0 |
| Palau | 0.00390 | 0 |
| Papua New Guinea | 0.02229 | 0 |
| Samoa | 0.07457 | 1 |
| Solomon Islands | 0.00188 | 0 |
| Tokelau | 0.00170 | 0 |
| Tonga | 0.03565 | 0 |
| Tuvalu | 0.00109 | 0 |
| Vanuatu | 0.00166 | 0 |
| Wallis and Fortuna | 0.00189 | 0 |
| **Total** | 1 | 6 |

### Version 4c

Some countries provide no bilateral information, and so we have only the totals supplied by the United Nations. As the main aim of the project was to produce a full bilateral matrix, it was important to calculate these coefficients. These were calculated on the same basis as the remainder, "Other" categories, i.e. based on the equation in 4b above, the propensity for countries to send people abroad, with the only difference being that rather than restricting the number of countries used to calculate the shares to those countries comprising the remainder category, the shares were calculated for all 226 countries, i.e. global as opposed to regional shares. The 30 countries until now lacking bilateral data therefore had all their coefficients estimated. Lastly, once the foreign-born matrix was filled, all of the data prior to 2000 were scaled to the United Nations (2004) mid-year-totals for 2000 so that a complete bilateral matrix for the foreign-born for the years 2000-2002 resulted (data for 2001 and 2002 were not scaled to the United Nations totals). Although a later United Nations dataset could have been utilized to extrapolate the data further into the future, it was considered more prudent in terms of accuracy to use that year for which most of the data are recorded. This allowed the inclusion of many countries that have neither collected the relevant migration statistics nor released them. The drawback in these circumstances is that the migration history between the date that the older source was conducted and 2000/01 will be lost as the relative shares are maintained over time. Table 6 provides details of the data contained in each version, together with the number of immigrants included.

**Table 6. Database versions**

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

# *5.* **What the Data Show[20]**

The principal purpose of this paper is to introduce a dataset that we believe will be useful for other researchers. We intend to use it for several purposes, such as updating the Walmsley and Winters (2005) analysis of the economic benefits of migration, in which the lack of bilateral migration data severely hampered the estimation of the distribution of those benefits. The aim of this section is not to embark upon such a detailed analysis but briefly to illustrate the patterns that the bilateral data display both for their intrinsic interest and to allow readers to judge whether the whole enterprise is plausible. Most of the value of the bilateral data lies in their country detail, but that is far too great for us to

---

[20] This summary refers to the current version of the dataset. Ratha's (2006) results and data (which focus on the South-South dimension) differ a little: his are scaled up to 2005, contain 14 fewer countries but 15 million more migrants, have treated "other" and residual categories differently, and have combined foreign-born and foreign nationality differently.

summarize here. Thus, once we move beyond simple totals, data are presented only for the standard World Bank regions of developing countries and geographical aggregations of high-income countries.

The final version of the database, version 4c, contains 175.7 million international migrants. This figure corresponds closely to the United Nations figure for 2000 of 175 million, although our data refer to 2000-2002. Table 7 reports the top 15 migrant recipient and sending countries in absolute terms according to the dataset. We exclude three countries from the FSU, Russia, Ukraine and Kazakhstan, which we believe are not representative of "normal" migration behavior. As a single country, the USSR had considerable internal mobility (some forced); when it split up, many of the people involved were recorded as migrants as the result of the changing country definitions rather than of any movement that they chose to undertake. These data contain no bilateral information but are reported merely to demonstrate that they correspond well to the United Nations totals by country.

As is well known, the United States is the largest recipient of migrants and now has around 12.5 percent (some 35 million) of its population 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 in our sample period. 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, Italy and even the United States within the top 15. Because we use an absolute criterion, all the countries in Table 7 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.

## Table 7. Share of migrant population in total and from/to neighboring country

| Top 15 receiving countries* | Number of immigrants | as % of host population | Immigrants from neighboring | % from neighboring | Top 15 sending countries* | Number of emigrants | as % of home population | Emigrants to neighboring | % to neighboring countries |
|---|---|---|---|---|---|---|---|---|---|
| United States | 34,634,798 | 12.5 | 10,281,887 | 29.7 | Mexico | 10,098,858 | 10.0 | 9,347,744 | 92.6 |
| Germany | 9,143,244 | 11.2 | 957,016 | 10.5 | India | 8,958,965 | 0.9 | 1,646,792 | 18.4 |
| France | 6,277,188 | 10.5 | 1,274,652 | 20.3 | Bangladesh | 6,638,008 | 5.0 | 3,807,203 | 57.4 |
| India | 6,270,659 | 0.6 | 5,822,700 | 92.9 | China | 5,793,974 | 0.5 | 2,442,169 | 42.2 |
| Canada | 5,717,003 | 18.3 | 945,091 | 16.5 | UK | 4,193,174 | 7.1 | 248,527 | 5.9 |
| Saudi Arabia | 5,254,812 | 23.0 | 596,912 | 11.4 | Germany | 4,047,061 | 4.9 | 915,277 | 22.6 |
| UK | 4,865,541 | 8.2 | 537,885 | 11.1 | Philippines | 3,405,471 | 4.3 | - | - |
| Pakistan | 4,242,691 | 3.0 | 724,233 | 17.1 | Pakistan | 3,386,516 | 2.4 | 1,360,069 | 40.2 |
| Australia | 4,073,213 | 21.0 | - | - | Italy | 3,283,208 | 5.7 | 704,628 | 21.5 |
| Hong Kong | 2,703,491 | 37.7 | 2,193,425 | 81.1 | Turkey | 3,001,152 | 4.5 | 184,940 | 6.2 |
| Cote d'Ivoire | 2,336,359 | 15.4 | 1,908,976 | 81.7 | Afghanistan | 2,695,589 | 9.9 | 1,898,925 | 70.4 |
| Iran | 2,321,453 | 3.6 | 2,298,835 | 99.0 | Morocco | 2,614,663 | 9.0 | 346,073 | 13.2 |
| Spain | 2,172,201 | 5.5 | 530,358 | 24.4 | USA | 2,269,220 | 0.8 | 620,712 | 27.4 |
| Israel | 1,978,103 | 31.1 | 34,882 | 1.8 | Egypt | 2,248,937 | 3.5 | 216,555 | 9.6 |
| Jordan | 1,945,210 | 39.0 | 218,582 | 11.2 | Algeria | 2,085,260 | 6.7 | 100,668 | 4.8 |
| Total | 93,935,966 | 5.0 | 28,325,434 | 30.2 | Total | 64,720,056 | 1.9 | 23,840,282 | 36.8 |

Note: Neighboring countries include all countries with land borders to the receiving/sending country. According to this definition, island countries do not have any neighboring countries.

More interesting than the aggregates in Table 7 are the columns on migration between neighboring countries. For these major migrant countries, we have used the bilateral dataset to calculate the proportion of migrants who come from or go to countries with a land border with the target country. The striking feature of these columns is the wide variety of experience. For India and Iran, over 90 percent of the flow is from such neighboring countries and for Hong Kong and Côte d'Ivoire over 80 percent. Pakistan, on the other hand, receives most migrants from Bangladesh and has only 17 percent from contiguous countries. At the other extreme, Australia and Israel have hardly any. It is not difficult to explain these patterns intuitively. Australia has no land neighbors and Israel's immigration is mainly religious, but it is useful to have them confirmed statistically and useful also to recognize the difficulty of making generalizations about neighborhood effects on migratory flows.

Turning to the genuinely bilateral data, Table 8 aggregates the data into 12 receiving and sending regions, expressing the recorded stocks as percentages of the total migratory stock: each 0.01 percent corresponds to about 17,500 people. The first feature to note is the strong diagonal: much migration takes place within regions. Nearly 17 percent of recorded migration occurs with Europe and Central Asia (ECA): we do not know precisely how much of this is created by border changes, but we suspect that the majority of it is. The second largest item in this aggregation is the stock of Latin Americans (including Mexicans) in the USA. The third and fourth largest are those within Sub-Saharan Africa (AFR) and South Asia (SAS). The latter two show clearly that South-South migration is quantitatively significant.

**Table 8. Percentage of world migrants recorded as a bilateral movement between pairs of countries/regions (%)**

| Countries/regions of origin | Destination countries/regions | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | USA | Canada | EU15+ EFTA | Australia+ New Zealand | Japan | High income | LAC | ECA | MENA | AFR | EAP | SAS | Total |
| USA | **0** | 0.16 | 0.34 | 0.04 | 0.02 | 0.03 | 0.43 | 0.04 | 0.05 | 0.03 | 0.15 | 0.02 | **1.29** |
| Canada | 0.54 | **0** | 0.10 | 0.02 | 0.00 | 0.01 | 0.02 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | **0.74** |
| EU15+EFTA | 2.22 | 0.98 | **5.59** | 1.13 | 0.01 | 0.14 | 0.68 | 0.78 | 0.16 | 0.39 | 0.20 | 0.19 | **12.47** |
| Australia+New Zealand | 0.06 | 0.02 | 0.16 | **0.23** | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.03 | 0.01 | **0.55** |
| Japan | 0.28 | 0.02 | 0.06 | 0.02 | **0** | 0.00 | 0.04 | 0.01 | 0.00 | 0.00 | 0.05 | 0.01 | **0.50** |
| High income MENA | 0.10 | 0.03 | 0.06 | 0.01 | 0.00 | **0.12** | 0.00 | 0.02 | 0.72 | 0.01 | 0.04 | 0.03 | **1.14** |
| LAC | 10.22 | 0.36 | 1.45 | 0.05 | 0.13 | 0.10 | **2.07** | 0.17 | 0.08 | 0.14 | 0.14 | 0.25 | **15.15** |
| ECA | 1.25 | 0.39 | 4.75 | 0.26 | 0.00 | 0.92 | 0.07 | **16.98** | 0.33 | 0.34 | 0.18 | 0.41 | **25.88** |
| MENA | 0.47 | 0.17 | 2.85 | 0.10 | 0.00 | 1.49 | 0.04 | 0.16 | **1.79** | 0.28 | 0.05 | 0.12 | **7.52** |
| AFR | 0.41 | 0.12 | 1.58 | 0.10 | 0.00 | 0.25 | 0.02 | 0.11 | 0.18 | **7.00** | 0.03 | 0.16 | **9.97** |
| EAP | 3.32 | 0.71 | 1.09 | 0.63 | 0.54 | 0.48 | 0.06 | 0.14 | 0.14 | 0.09 | **3.86** | 0.27 | **11.32** |
| SAS | 0.83 | 0.31 | 1.13 | 0.12 | 0.01 | 2.66 | 0.02 | 0.13 | 2.07 | 0.14 | 0.37 | **5.67** | **13.46** |
| Total | **19.71** | **3.25** | **19.14** | **2.72** | **0.74** | **6.22** | **3.45** | **18.56** | **5.53** | **8.44** | **5.10** | **7.15** | **100** |

Note: "EU15 + EFTA" includes Switzerland.

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 1 summarizes these data graphically, distinguishing North (rich), South (poor) and the FSU countries. The bulk of FSU migration is intra-regional; it is 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 percent 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 percent of total emigration, South-South for 24 percent and North-North 16 percent.



Figure 1:   Bilateral Migration Stocks (% of world total)
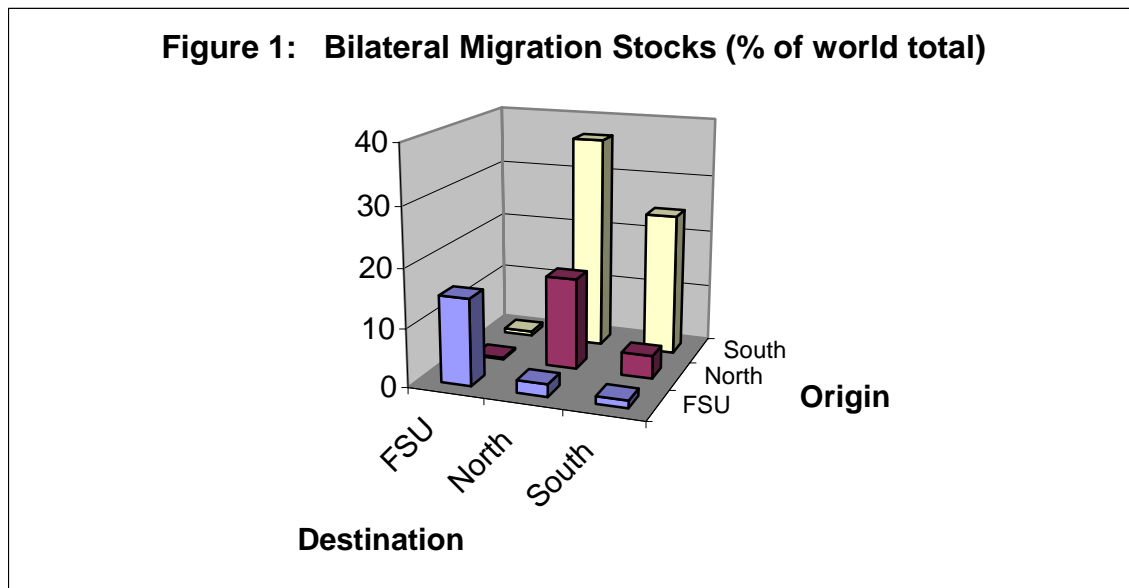
Embedded in table 8 is evidence of the large differences between hosts in the sources of their immigrants and differences between sources in their hosts. Figures 2 and 3 illustrate these differences more obviously.

For each developed country (Northern) group, 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 percent 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 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 (high income MENA) has a high share of South Asians and, perhaps tellingly, no one has a high share of Africans.

20

## Figure 2. Source country shares of certain destination regions
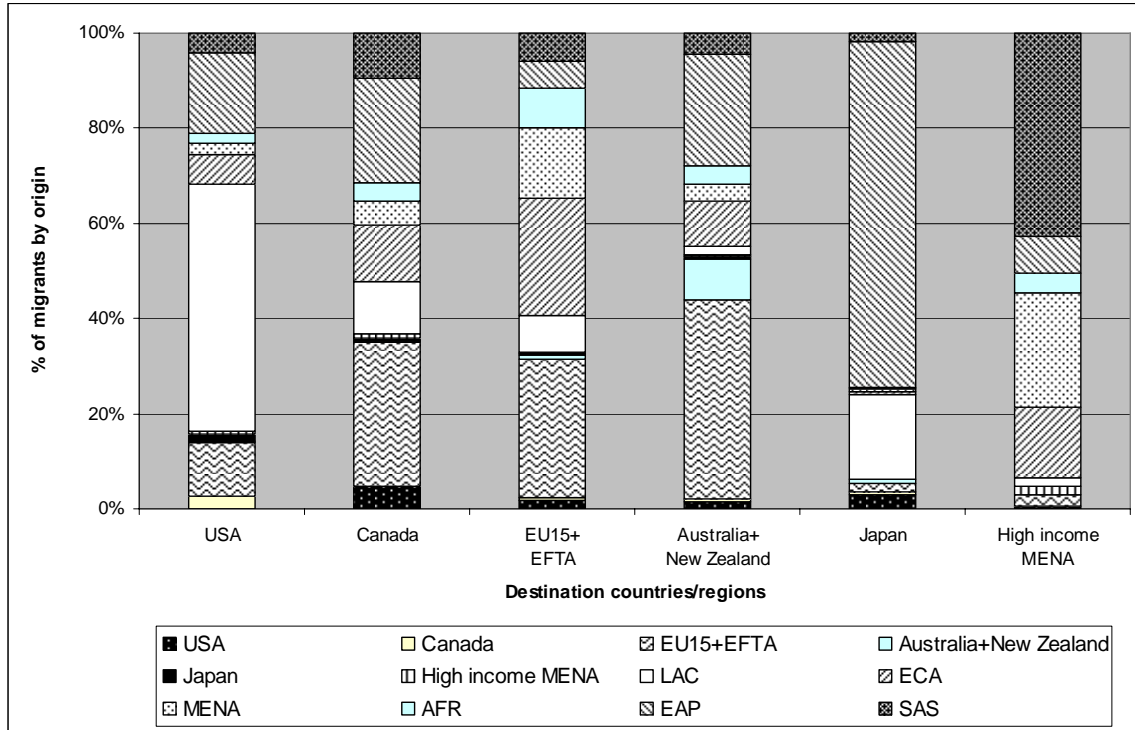


Figure 3 offers the obverse view – the choice of destinations for each source region. For Northern regions, the principal destinations are Northern: fully 80 percent for Europe. On the other hand, because Europe's outflow is so much larger than the other Northern regions, its absolute supply of people to each of the developing regions is larger (Table 8). For developing source countries, 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.

**Figure 3. Destinations of emigrants from source regions**
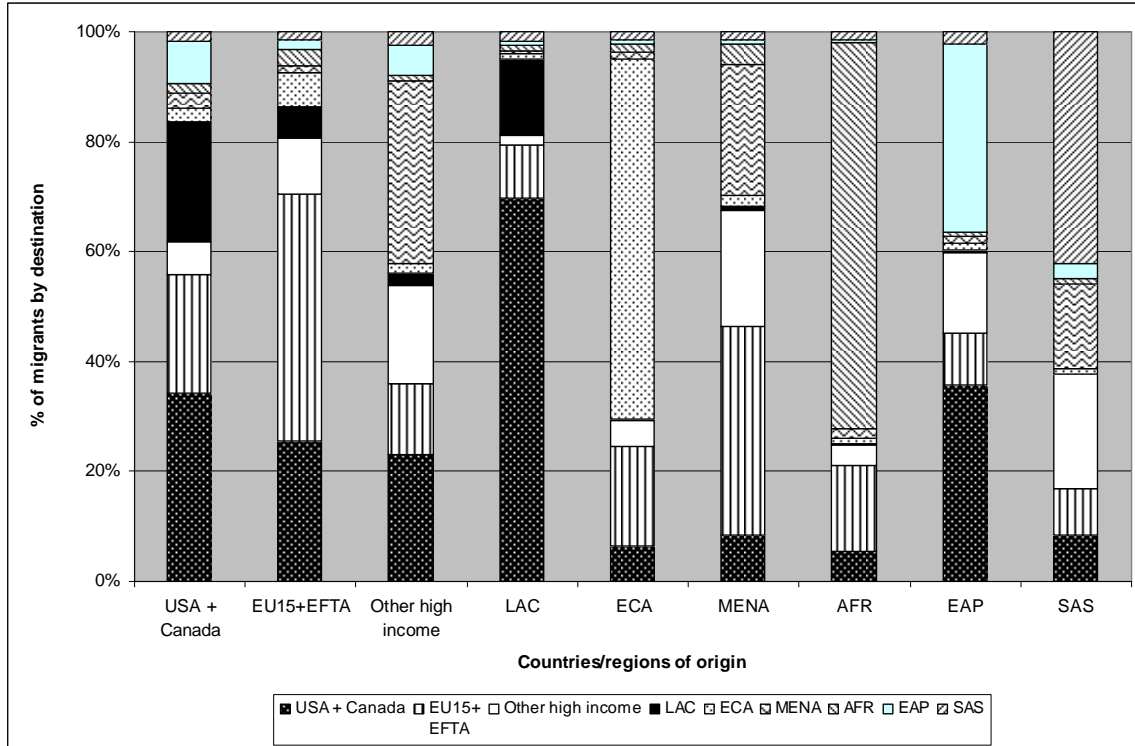


Interesting though table 8 and the figures are, they do not readily show whether migration patterns are biased in the sense that migrants from one source are more likely to go to one place than another. This is because it is not clear whether the importance of, say, North America as a destination for EAP arises because it is a major recipient of migrants from every region or because EAP residents have a particular propensity or ability to migrate there. To answer this question we calculate a series of migration intensity indices. These 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. The indices are reported in table 9.

**Table 9. Migration Intensity Measures**

| Countries/Regions of Origin | Destination Countries/Regions | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | USA | Canada | EU15+ EFTA | Australia+ New Zealand | Japan | High income MENA | LAC | ECA | MENA | AFR | EAP | SAS |
| USA | **0** | 3.77 | 1.36 | 1.09 | 2.32 | 0.37 | 9.56 | 0.16 | 0.70 | 0.23 | 2.21 | 0.23 |
| Canada | 3.69 | **0** | 0.73 | 0.99 | 0.74 | 0.12 | 0.73 | 0.09 | 0.11 | 0.13 | 0.31 | 0.25 |
| EU15+EFTA | 0.90 | 2.41 | **2.34** | 3.34 | 0.15 | 0.18 | 1.58 | 0.34 | 0.23 | 0.37 | 0.31 | 0.22 |
| Australia+New Zealand | 0.54 | 0.96 | 1.50 | **15.76** | 1.21 | 0.13 | 0.14 | 0.12 | 0.13 | 0.17 | 1.17 | 0.30 |
| Japan | 2.87 | 0.95 | 0.65 | 1.42 | **0** | 0.07 | 2.48 | 0.08 | 0.16 | 0.10 | 1.98 | 0.24 |
| High income MENA | 0.47 | 0.72 | 0.28 | 0.28 | 0.03 | **1.63** | 0.13 | 0.08 | 11.45 | 0.09 | 0.65 | 0.42 |
| LAC | 3.42 | 0.73 | 0.50 | 0.11 | 1.19 | 0.11 | **3.96** | 0.06 | 0.10 | 0.11 | 0.18 | 0.23 |
| ECA | 0.25 | 0.46 | 0.96 | 0.36 | 0.02 | 0.57 | 0.08 | **3.54** | 0.23 | 0.16 | 0.13 | 0.22 |
| MENA | 0.32 | 0.68 | 1.98 | 0.49 | 0.06 | 3.20 | 0.15 | 0.11 | **4.31** | 0.44 | 0.13 | 0.21 |
| AFR | 0.21 | 0.38 | 0.83 | 0.38 | 0.03 | 0.41 | 0.05 | 0.06 | 0.33 | **8.31** | 0.06 | 0.22 |
| EAP | 1.49 | 1.94 | 0.50 | 2.06 | 6.43 | 0.68 | 0.16 | 0.07 | 0.22 | 0.09 | **6.68** | 0.33 |
| SAS | 0.31 | 0.71 | 0.44 | 0.33 | 0.13 | 3.18 | 0.04 | 0.05 | 2.78 | 0.13 | 0.54 | **5.89** |

Note: "EU15 + EFTA" includes Switzerland.

Table 9 demonstrates the strong bias towards intra-regional migration. Every diagonal element exceeds unity (except where it is zero because there is only one country in the region), very much so in some cases. Australia and New Zealand have a huge propensity to migrate between themselves, perhaps reflecting the integrated labor market created under the Closer Economic Relations Agreement, or maybe just their isolation. In second place is intra-African migration, more probably a reflection of Africans' relative unwelcomeness elsewhere, than of anything positive. Among the off-diagonal elements of note are the strength of the EAP-Japan link; the close links between the USA and Canada; the bias towards the USA for LAC emigrants, which we have noted before, and the bias of US emigrants towards LAC, which we have not; the European tendency to choose other Northern destinations; and the strong relative movement from high-income MENA (the Gulf) to developing MENA. Again, 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 bilateral data allow us to explore a number of possible factors behind migration, descriptively here, but possibly econometrically in future. One concerns the influence of a common language. For 163 of our countries we have been able to identify a predominant international language: Arabic, Chinese, English, French, Portuguese and Spanish.



**Figure 4: Migration and Common Language**

Legend:
- Arabic
- Chinese
- English
- French
- Portuguese
- Spanish
- different
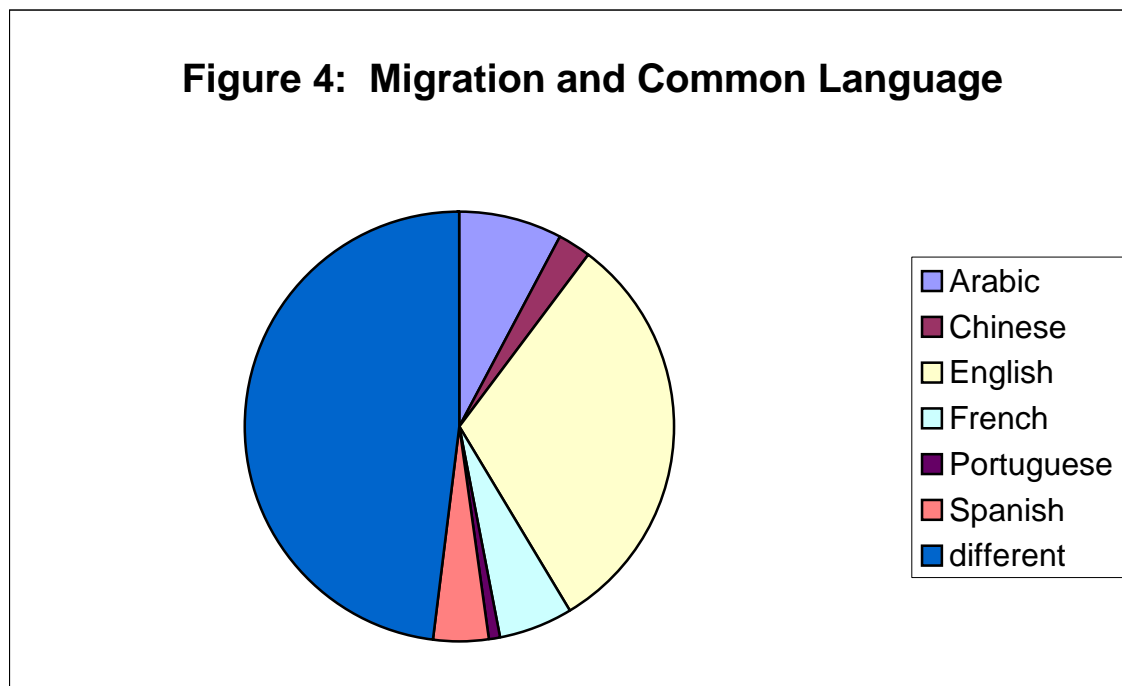
Figure 4 shows that over half of international migration (that on the right hand side of the pie) takes place between countries with a common language, the lion's share of this being between English-speaking countries. Dividing each language group into North and South shows that much of the common language flow is within the North or within the South, but that the North-South dimension is also strong. Every Northern language group draws a large share of its immigrants from regions in the South which share its language and, where this is not the largest share, it is dominated by another well-trodden South-North path. The Arabic-North has a higher percentage from the English-

speaking South (SAS-to-higher income MENA migration); the English North has its greatest share from the Spanish-speaking South (Mexico to US migration); and the French North takes a majority of its immigrants from the Arabic-speaking South (North Africa to France migration).

Table 10, which presents the data as intensity indicators, also shows how important a common language is. The intensities are generally highest within the language families, but often higher for the North-South or South-North links than the others (see the shaded (2x2) matrices down the main diagonal). This is not saying that more migrants travel between income classes than within them, but that given the overall totals of migrants the bias towards "between" is stronger. There are also some strong cross-language intensities, mainly reflecting location. Thus, the French, Spanish and Portuguese North is strongly integrated, the Arabic South feeds the French North, and the Portuguese South feeds the Chinese South.

## Table 10. Migration and language: migration intensity measures

| Countries/Regions of Origin | Destination Countries/Regions | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Arabic North | Arabic South | Chinese North | Chinese South | English North | English South | French North | French South | Portuguese North | Portuguese South | Spanish North | Spanish South |
| Arabic North | **1.19** | 8.95 | 0.17 | 0.52 | 0.36 | 0.29 | 0.13 | 0.05 | 0.02 | 0.11 | 0.05 | 0.08 |
| Arabic South | 2.83 | **3.47** | 0.09 | 0.29 | 0.27 | 0.29 | 4.21 | 0.58 | 0.03 | 0.29 | 1.67 | 0.08 |
| Chinese North | 0.05 | 0.07 | **2.12** | 6.30 | 2.09 | 0.14 | 0.11 | 0.07 | 0.38 | 0.36 | 0.00 | 0.20 |
| Chinese South | 0.06 | 0.08 | 16.42 | **0** | 0.87 | 0.73 | 0.16 | 0.09 | 0.08 | 0.22 | 0.31 | 0.17 |
| English North | 0.15 | 0.21 | 0.10 | 1.15 | **1.91** | 0.40 | 0.38 | 0.12 | 0.42 | 0.23 | 0.74 | 1.18 |
| English South | 1.89 | 0.84 | 0.28 | 1.37 | 0.75 | **2.05** | 0.11 | 0.55 | 0.12 | 1.09 | 0.13 | 0.05 |
| French North | 0.26 | 0.35 | 0.04 | 0.46 | 0.74 | 0.29 | **4.03** | 2.22 | 9.74 | 0.78 | 5.98 | 0.50 |
| French South | 0.05 | 0.96 | 0.01 | 0.05 | 0.20 | 1.41 | 1.15 | **7.04** | 0.09 | 0.41 | 0.16 | 0.31 |
| Portuguese North | 0.06 | 0.08 | 0.05 | 0.17 | 0.66 | 0.21 | 7.61 | 0.13 | **0** | 17.23 | 2.19 | 0.95 |
| Portuguese South | 0.04 | 0.08 | 0.03 | 7.85 | 0.30 | 2.72 | 0.43 | 0.92 | 23.83 | **2.01** | 0.83 | 1.52 |
| Spanish North | 0.07 | 0.10 | 0.04 | 0.24 | 0.51 | 0.28 | 6.45 | 0.12 | 2.12 | 3.80 | **1.81** | 6.03 |
| Spanish South | 0.09 | 0.08 | 0.05 | 0.40 | 1.84 | 0.12 | 0.08 | 0.15 | 0.21 | 0.69 | 2.19 | **3.17** |

Note: Calculations based on the sub-sample of 163 countries.

A second factor frequently spoken of is distance. This is clearly related to the question of migration to and from neighbors discussed above, but even for non-contiguous countries, distance is often felt to matter. Calculating distances as great circle distances between the geographical centres of our 226 countries, as given in the CIA World Fact Book, the main body of Table 11 reports the average distance travelled within each cell of the regional bilateral matrix. The distances for the various bilateral country-to-country links are weighted together by the amount of migration for that link. Of course, the distances are exogenous features of geography, so, for example, the distance from South Asia to North America is inevitably about 12,000 kilometers. Hence, there is not much to be read from the numbers per se. It is true that North-North migrants (in the top left of the matrix) generally have to travel smaller distances than South-South migrants (lower right) or North-South or South-North ones, but with Oceania among the Northern set, the difference is not huge. The mean distance for all migrants in the world is 4,225 kilometers.

**Table 11. Average distance between origin and destination countries (km)**

| Countries/Regions of Origin | Destination Countries/Regions | | | | | | | | | Correlation with numbers by origin |
|---|---|---|---|---|---|---|---|---|---|---|
| | USA + Canada | EU15+ EFTA | Other high income | LAC | ECA | MENA | AFR | EAP | SAS | |
| USA + Canada | 2,451 | 7,170 | 12,624 | 3,064 | 8,805 | 10,566 | 11,211 | 12,570 | 11,604 | -0.87 |
| EU15+EFTA | 7,151 | 937 | 14,136 | 9,044 | 2,417 | 3,232 | 6,839 | 11,340 | 6,164 | -0.38 |
| Other high income | 10,859 | 11,938 | 3,627 | 15,886 | 7,411 | 901 | 9,055 | 5,226 | 5,175 | -0.42 |
| LAC | 2,815 | 8,571 | 14,734 | 1,386 | 11,546 | 11,738 | 10,560 | 15,863 | 14,319 | -0.68 |
| ECA | 8,429 | 1,880 | 5,674 | 11,405 | 3,012 | 3,177 | 7,886 | 9,352 | 3,754 | -0.47 |
| MENA | 10,387 | 2,253 | 2,334 | 10,807 | 2,932 | 1,246 | 3,659 | 10,924 | 5,026 | -0.60 |
| AFR | 11,718 | 6,124 | 4,815 | 9,017 | 6,901 | 3,137 | 1,272 | 11,207 | 7,013 | -0.61 |
| EAP | 11,935 | 10,024 | 5,167 | 16,143 | 6,735 | 8,668 | 11,061 | 4,344 | 4,212 | -0.17 |
| SAS | 12,681 | 6,968 | 3,187 | 14,998 | 4,155 | 2,186 | 6,951 | 5,680 | 1,494 | -0.58 |
| Correlation with numbers by destination | -0.51 | -0.88 | -0.47 | -0.75 | -0.38 | -0.59 | -0.72 | -0.54 | -0.48 | |

Note: "EU15 + EFTA" includes Switzerland. "Other high income" includes Australia, New Zealand, Japan and high income MENA countries. Figures are weighted by the number of migrants. Calculations are based on the geographic coordinates of each country, which are rounded latitude and longitude figures of the approximate geographic centre of a country (source: CIA, The World Fact Book).

Migration numbers, on the other hand, are not exogenous and we can test whether they reflect distance at all. First, consider figures 5A and 5B. They report for the developing country regions the average distance migrants have travelled to North America and Western Europe and the number who have migrated in thousands. It is clear that there is a negative relationship between these, especially for Europe. In addition, the last row and column of Table 11 report the correlations (at regional level) between the migration stock reported in table 7 and the distance reported here. They are all negative and half are significantly different from zero (critical value at 5 percent is 0.582). This suggests quite strongly that distance is an important determinant of migration behavior throughout the world.



**Figure 5A  Migrants in North America and Distance**



**Figure 5B.  Migrants in Western Europe and Distance**
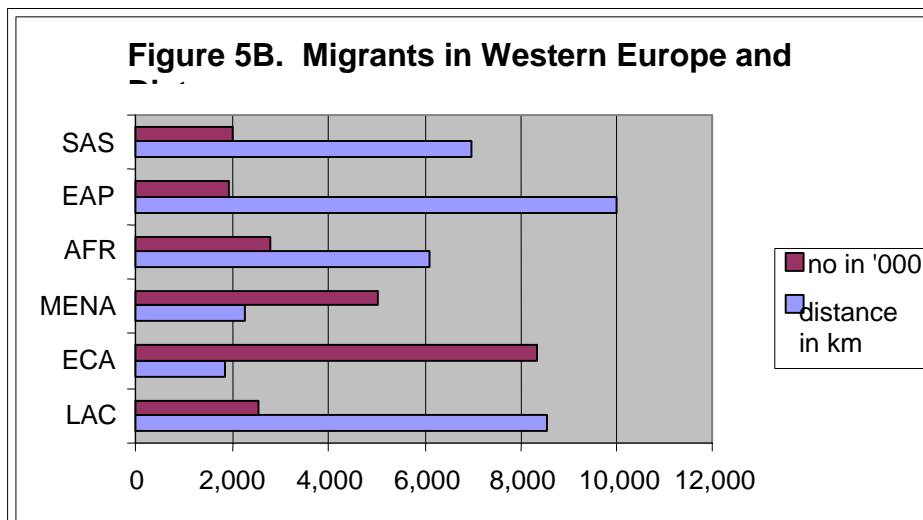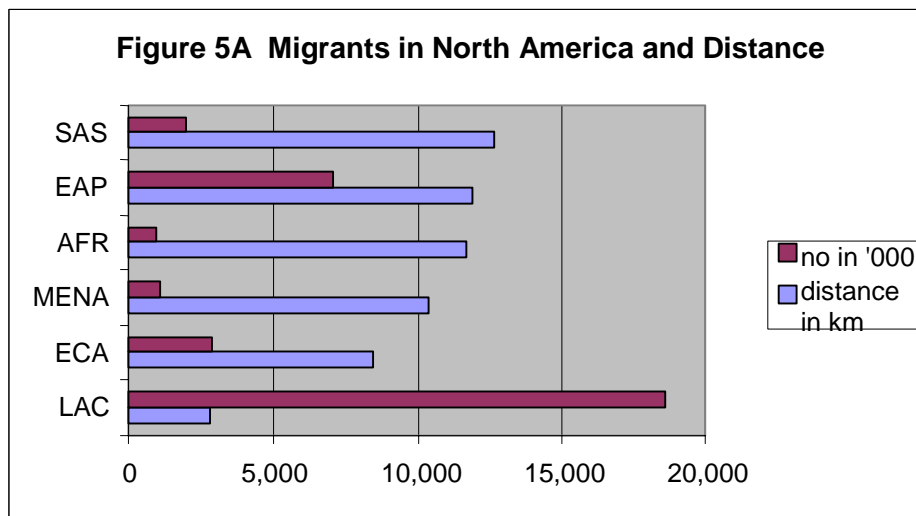
A third influence on migration is income. We expect most migrants to have higher incomes as a result of migration, although in particular instances non-pecuniary benefits may dominate. Income is, of course, an individual or household matter, dependent on the individual's circumstances before and after migration. We cannot gauge that with country-level data, but it is still interesting to ask whether migration is predominantly from

low to high-income countries. Table 12 provides a view of this. For each of our region-to-region links it calculates the percentage change in average gross national income (GNI) per capita for the average migrant. That is, for each bilateral link it calculates the change in current dollar GNI per capita and then weights these together into regional summaries by the number of migrants reported for the link.[21] The GNI is generated in $US using the World Bank Atlas method of conversion and divided by the mid-year population and is taken from the World Development indicators for 2001. The matrix is best thought of in terms of blocks. With the Northern groups reported first, the (3x3) matrix in the top left corner refers to North-North movements. The differentials are small here, but they show the North Americans always "losing", for example. The (3x6) matrix in the top right is North to South and it shows massive "income loses" for Northern people migrating to developing countries. The (6x3) matrix in the bottom left is South-North and shows huge gains. The final sub-matrix – (6x6) in the lower right - is South-South. It has pluses and minuses as one would expect, given the spread of incomes from essentially middle-income regions LAC and EAP to the very poor in AFR and SAS.

---

[21] The GNI refer to 2001, not the (unknown) years in which migrants moved, but the basic argument still holds.

**Table 12. Average difference in GNI per capita between origin and destination countries (current US$)**

| Countries/Regions of Origin | Destination Countries/Regions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | USA + Canada | EU15+ EFTA | Other high income | LAC | ECA | MENA | AFR | EAP | SAS |
| USA + Canada | 6,918 | -8,026 | -10,907 | -27,428 | -29,265 | -32,099 | -31,287 | -30,939 | -29,570 |
| EU15+EFTA | 8,694 | 3,334 | -4,332 | -12,129 | -20,061 | -20,297 | -20,041 | -21,571 | -21,367 |
| Other high income | 5,419 | 2,743 | 1,152 | -29,321 | -17,856 | -14,407 | -19,881 | -18,316 | -19,147 |
| LAC | 29,523 | 17,560 | 22,511 | 2,290 | -2,944 | -2,626 | -3,890 | 510 | -3,933 |
| ECA | 29,018 | 20,755 | 15,123 | 2,586 | 269 | 487 | -839 | 10,862 | -1,268 |
| MENA | 29,244 | 21,298 | 9,853 | 1,577 | -107 | 168 | -1,220 | 9,209 | -1,207 |
| AFR | 31,290 | 21,395 | 12,700 | 4,667 | 819 | 1,825 | 199 | 2,226 | -50 |
| EAP | 27,820 | 19,959 | 16,827 | -977 | -1,432 | 465 | -2,843 | 11,320 | -3,191 |
| SAS | 30,879 | 24,034 | 13,046 | 5,681 | 989 | 4,036 | -39 | 9,204 | 62 |

Note: "EU15 + EFTA" includes Switzerland. "Other high income" includes Australia, New Zealand, Japan and high income MENA countries. Figures are weighted by the number of migrants. Note also that the calculations are based on 180 countries excluding those for which we were not able to obtain data on GNI per capita for 2001
Source: World Development Indicators.

31

Two statistics, however, show the general tendency for migrants to move up the GNI per capita ladder. First, every element on the diagonal, i.e. every intra-regional migrant link, shows gains. On the diagonal, the sets of receiving countries and the sets of sending countries are identical, so the unweighted difference in GNI per capita between them is zero. Thus, the positive signs reported must reflect the fact that more people move along the paths with increasing GNI per capita than along those with falling GNI per capita. The second statistic is to collapse the whole world into a single entity and calculate the differential. Again, the sets of recipients and senders are identical so the average gain of $9,479 must again reflect a bias in the movement toward those that increase GNI per capita. The increase is large and while GNI per capita will clearly not reflect only returns to labor and while migrants will typically have above-average earning power in their home countries and, these days, probably below-average earning power in their new countries, it suggests large returns to migration.

## *6.* **Conclusion**

In spite of the inherent inaccuracy due to a lack of bilateral data for some countries, the assumptions made and the rough-and-ready methodologies used, the database developed here should provide a reasonably realistic view of current migration patterns. Considering the heterogeneity that exists between the available sources and the various missing components, any undertaking on this subject will inevitably be rather crude. We see no way of improving upon it, however, given current data availability and so we present our series "as is". We would be delighted to hear of potential extensions or improvements, or, indeed, to see them made by other researchers. To that end, we have tried to highlight the problems encountered and to provide alternative series so that other researchers can build upon our data as they see fit. Despite their shortcomings, we believe that the data will prove valuable in a wide range of applications throughout the social sciences, enhancing both our understanding of global migration and the policy debate surrounding it.

## Annex 1

## Summary of Data Sources

| Country | Data type | Nationality or Birth | Source | Year |
|---|---|---|---|---|
| Australia | B | B | C | 2001 |
| New Zealand | B | B | C | 2001 |
| American Samoa | B | B | C | 2000 |
| Cook Islands | B | N | C | 2001 |
| Fiji | B | B | C | 1986 |
| French Polynesia | B/B | B/N | C/C | 2002/2002 |
| Guam | B | B | C | 2000 |
| Kiribati | B | E | C | 2000 |
| Marshall Islands | B | N | C | 1999 |
| Micronesia, Federated States of | B | B | C | 2000 |
| Nauru | UN | N | U | 2000 |
| New Caledonia | B | N | C | 1996 |
| Norfolk Island | B/B | B/N | C/C | 2001/2001 |
| Northern Mariana Islands | B | B | C | 2000 |
| Niue | B | N | C | 2001 |
| Palau | B | B | C | 2000 |
| Papua New Guinea | B | B | C | 1971 |
| Samoa | B | B | C | 2001 |
| Solomon Islands | B/B | N/B | C | 1999 |
| Tokelau | B | E | C | 1996 |
| Tonga | B | E | C | 1996 |
| Tuvalu | T | N | C | 2002 |
| Vanuatu | B | N | C | 1999 |
| Wallis and Futuna | B | B | C | 1996 |
| China | UN | E | U | 2000 |
| Hong Kong | B/B | B/N | C/C | 2001/2001 |
| Japan | B | N | C | 2000 |
| Korea, Republic of | B | N | C | 2000 |
| Taiwan | B | N | S | 2000 |
| Macau | B/B | B/O | C/S | 1991/2001 |
| Mongolia | B | N | C | 2000 |
| Korea, Democratic People's Republic of | UN | E | U | 2000 |
| Indonesia | UN | N | U | 2000 |
| Malaysia | B | B | C | 1991 |
| Philippines | B | N | C | 2000 |
| Singapore | B | B | C | 2000 |
| Thailand | B | N | S | 2000 |
| Viet Nam | T | N | C | 1999 |
| Brunei Darussalam | B | B | C | 1991 |
| Cambodia | B | N | C | 1998 |
| Lao People's Democratic Republic | B | N | C | 1995 |
| Myanmar | B | N | S | 2000 |
| Timor Leste | UN | E | U | 2000 |
| Bangladesh | B | B | C | 1974 |
| India | B | B | C | 2001 |
| Sri Lanka | B | N | C | 1981 |
| Afghanistan | UN | E | U | 2000 |
| Bhutan | UN | N | U | 2000 |

| Country | Data type | Nationality or Birth | Source | Year |
|---|---|---|---|---|
| Maldives | UN | N | U | 2000 |
| Nepal | B/B | B/N | C/C | 2001/2001 |
| Pakistan | T | B | C | 1998 |
| Canada | B | B | C | 2001 |
| United States of America | B | B | C | 2000 |
| Mexico | B | B | C | 2000 |
| Bermuda | B | B | C | 2000 |
| Greenland | UN | B | U | 2000 |
| Saint Pierre and Miquelon | B | B | C | 1974 |
| Colombia | B | B | C | 1993 |
| Peru | B | B | C | 1993 |
| Venezuela | B | B | C | 2001 |
| Bolivia | B | B | C | 2002 |
| Ecuador | B | B | C | 2001 |
| Argentina | B | B | C | 2001 |
| Brazil | B/B | B/B | C/C | 1991/2000 |
| Chile | B | B | C | 2002 |
| Uruguay | B | B | C | 1996 |
| Falkland Islands (Malvinas) | B/B | B/N | C/C | 2001/2001 |
| French Guiana | B | N | C | 1990 |
| Guyana | B | B | C | 2002 |
| Paraguay | B | B | C | 2002 |
| Suriname | UN | N | U | 2000 |
| Belize | B | B | C | 2000 |
| Costa Rica | B/B | B/N | C/C | 2002/2002 |
| El Salvador | B | B | C | 1990 |
| Guatemala | B | B | C | 2002 |
| Honduras | B | B | C | 2001 |
| Nicaragua | B | B | C | 1995 |
| Panama | B | B | C | 2000 |
| Antigua & Barbuda | B | B | C | 2001 |
| Bahamas | B | N | C | 2001 |
| Barbados | B | B | C | 1990 |
| Dominica | B | B | C | 1981 |
| Dominican Republic | B | B | C | 2002 |
| Grenada | UN | B | U | 2000 |
| Haiti | B | B | C | 1971 |
| Jamaica | B | B | C | 1960 |
| Puerto Rico | T | B | C | 2000 |
| Saint Kitts and Nevis | UN | B | U | 2000 |
| Saint Lucia | B | B | C | 2001 |
| Saint Vincent and the Grenadines | B | B | C | 1991 |
| Trinidad and Tobago | B | B | C | 2000 |
| Virgin Islands, U.S. | B | B | C | 2000 |
| Anguilla | B | N | C | 2001 |
| Aruba | B | N | C | 2000 |
| Cayman Islands | B | B | C | 1989 |
| Cuba | B | B | C | 1970 |
| Guadeloupe | B | N | C | 1990 |
| Martinique | B | N | C | 1990 |
| Montserrat | T | B | C | 2001 |
| Netherlands Antilles | B/B | B/N | C/C | 2001/2001 |
| Turks and Caicos | B\B | B/N | C/C | 1990/2001 |

| Country | Data type | Nationality or Birth | Source | Year |
|---|---|---|---|---|
| Virgin Islands, British | B | B | C | 1991 |
| Austria | B/B | B/N | C/C | 2001/2001 |
| Belgium | B/B | B/N | O/S | 2001/2000 |
| Denmark | B/B | B/N | P/S | 2001/2001 |
| Finland | B/B | B/N | P/P | 2001/2002 |
| France | B/B | B/N | C/C | 1999/1999 |
| Germany | B/B | B/N | C/S | 2001/2001 |
| United Kingdom | B | B | C | 2001 |
| Greece | B/B | B/N | C/C | 2001/2001 |
| Ireland | B/B | B/N | C/C | 2002/2002 |
| Italy | B | N | S | 2000 |
| Luxembourg | B/B | B/N | C/C | 2001/2001 |
| Netherlands | B/B | B/N | P/P | 2001/2001 |
| Portugal | B/B | B/N | C/C | 2001/2001 |
| Spain | B/B | B/N | C/C | 2001/2001 |
| Sweden | B/B | B/N | P/P | 2001/2001 |
| Switzerland | B/B | B/N | C/P | 2000/2000 |
| Iceland | B/B | B/N | P/P | 2001/2001 |
| Liechtenstein | B | B | S | 2000 |
| Norway | B/B | B/N | P/P | 2001/2001 |
| Andorra | B | N | S | 2002 |
| Bosnia and Herzegovina | B | E | C | 1991 |
| Faeroe Islands | UN | B | U | 2000 |
| Gibraltar | B/B | B/N | C/C | 2001/2001 |
| Macedonia, the former Yugoslav Republic of | B | N | C | 1994 |
| Monaco | B | N | C | 2000 |
| San Marino | UN | B | U | 2000 |
| Serbia and Montenegro | UN | B | U | 2000 |
| Albania | T/B | B/N | C/C | 2001/1989 |
| Bulgaria | B | N | C | 2001 |
| Croatia | B/B | B/E | C/S | 2001/2001 |
| Cyprus | B | B | C | 2002 |
| Czech Republic | B/B | B/N | C/R | 2001/2000 |
| Hungary | B/B | B/N | C/S | 2001/2001 |
| Malta | B/B | B/N | C/C | 1995/1995 |
| Poland | B | B | C | 2001 |
| Romania | B/B | B/E | C/C | 2002/2002 |
| Slovakia | B/B | B/E | C/S | 2001/2001 |
| Slovenia | B | N | C | 2001 |
| Estonia | B/B | B/N | C/C | 2001/2001 |
| Latvia | B | N | S | 2001 |
| Lithuania | B/B | B/N | C/C | 2001/2001 |
| Russian Federation | B/B | B/N | C/C | 2002/2002 |
| Armenia | B/T | B/N | C/S | 2001/2001 |
| Azerbaijan | B | N | S | 2001 |
| Belarus | B | E | S | 1999 |
| Georgia | B | B | C | 2002 |
| Kazakhstan | B | E | S | 1999 |
| Kyrgyzstan | B | E | S | 1999 |
| Moldova, Republic of | B | E | S | 2001 |
| Tajikistan | B | E | S | 1989 |
| Turkmenistan | B | E | C | 1995 |
| Ukraine | B/B | B/E | C/C | 2001/2001 |

| Country | Data type | Nationality or Birth | Source | Year |
|---|---|---|---|---|
| Uzbekistan | B | E | S | 1989 |
| Turkey | B/B | B/N | C/S | 2001/1998 |
| Bahrain | B | N | C | 2001 |
| Iran, Islamic Republic of | B | N | C | 1996 |
| Iraq | UN | N | U | 2000 |
| Israel | B | B | S | 2001 |
| Jordan | T | N | C | 1994 |
| Kuwait | T | N | S | 2001 |
| Lebanon | UN | B | U | 2000 |
| Palestinian Territory, Occupied | B/B | B/N | C/C | 1997/1997 |
| Oman | B | N | C | 1993 |
| Qatar | UN | N | U | 2000 |
| Saudi Arabia | B | N | S | 1995 |
| Syrian Arab Republic | B | N | C | 1981 |
| United Arab Emirates | T | N | S | 1993 |
| Yemen | UN | N | U | 2000 |
| Morocco | UN | N | C | 2000 |
| Tunisia | B | N | C | 1994 |
| Algeria | UN | N | U | 2000 |
| Egypt | B | N | C | 1996 |
| Libyan Arab Jamahiriya | B/B | O/N | C/C | 1964/1973 |
| Botswana | B | N | C | 2001 |
| South Africa | B/B | B/N | C/C | 2001/2001 |
| Lesotho | B | N | C | 1996 |
| Namibia | B | N | C | 1991 |
| Swaziland | B | N | C | 1997 |
| Malawi | B | N | C | 1998 |
| Mozambique | B | N | C | 1997 |
| Tanzania, United Republic of | B | B | C | 2002 |
| Zambia | B | N | C | 2000 |
| Zimbabwe | UN | B | U | 2000 |
| Angola | UN | B | U | 2000 |
| Congo, the Democratic Republic of the | B | N | C | 1984 |
| Mauritius | B | N | C | 2000 |
| Seychelles | B | N | C | 1997 |
| Madagascar | B | B | C | 1996 |
| Uganda | B | N | C | 2002 |
| Benin | B | N | C | 1992 |
| Burkina Faso | B | N | S | 1996 |
| Burundi | UN | B | U | 2000 |
| Cameroon | B | B | C | 1976 |
| Cape Verde | B | B | C | 1990 |
| Central African Republic | B | N | C | 1988 |
| Chad | B | N | C | 1993 |
| Comoros | B | N | C | 1980 |
| Congo | B | N | C | 1984 |
| Cote d'Ivoire | B | N | C | 1992 |
| Djibouti | UN | E | U | 2000 |
| Equatorial Guinea | B | N | C | 1994 |
| Eritrea | UN | E | U | 2000 |
| Ethiopia | B | N | C | 1994 |

| Country | Data type | Nationality or Birth | Source | Year |
|---|---|---|---|---|
| Gabon | B | N | C | 1993 |
| Gambia | B | N | C | 1983 |
| Ghana | B | N | C | 2000 |
| Guinea | UN | N | U | 2000 |
| Guinea-Bissau | B | N | C | 1991 |
| Kenya | T | N | C | 1999 |
| Liberia | B | B | C | 1974 |
| Mali | UN | N | U | 2000 |
| Mauritania | B | N | C | 1988 |
| Mayotte | -- | - | - | - |
| Niger | B | B | C | 1988 |
| Nigeria | B | N | C | 1991 |
| Reunion | B | N | C | 1999 |
| Rwanda | B | N | C | 2002 |
| Saint Helena | UN | B | U | 2000 |
| Sao Tome and Principe | B/B | B/N | C/C | 1991/1991 |
| Senegal | B/B | B/N | C/C | 1976/1976 |
| Sierra Leone | B | N | C | 1985 |
| Somalia | UN | E | U | 2000 |
| Sudan | B/N | B/B | C/C | 1953/1993 |
| Togo | B | N | C | 1981 |

Where two letters are recorded and separated by a forward slash, the first letter refers to data collected by country of birth and the second to information regarding nationality.

T = Total or very limited bilateral entries only
B = Bilateral (may not be for all 226 countries but at least bilateral for main partners).
B/B = Bilateral/Bilateral
B/T = Bilateral/Total
UN = United Nations total only

N = Nationality
B = Birth
E = Ethnicity
B/N = Birth and nationality
O = Other but equivalent

C= Census
PR = Population register
S = Source unclear or not stated but obtained from National Statistics Bureau, either directly or from published yearly handbooks.
U = Unknown, need check with United Nations
R = Register of foreigners
O = Other i.e. survey/permit data

## Annex 2

The entropy measure used to compare the shares of the foreign born and nationality matrices are based on the entropy measure (3) that devised by Walmsley and McDougall (2004):

$$(3)\ E_{r,s} = 0.5 \left[ S_{r,s}^{*A} \left( Log_e \left( S_{r,s}^{*A} / S_{r,s}^{*B} \right) \right) \right] + 0.5 \left[ S_{r,s}^{*B} \left( Log_e \left( S_{r,s}^{*B} / S_{r,s}^{*A} \right) \right) \right]$$

Where:  $E_{r,s}$ = the entropy measure of the difference between $S_{r,s}^{*A}$ and $S_{r,s}^{*B}$

$S_{r,s}^{*A}$ = the adjusted share of migrants from country r in country s to use in foreign born matrix.

$S_{r,s}^{*B}$ = the adjusted share of migrants from country r in country s to use in nationality matrix.

The adjustment to the shares being:

$$S_{r,s}^{*A} = \left( S_{r,s}^{A} \left( 1 - TINY \right) \right) + \left( S_{r,s}^{B} \left( TINY \right) \right)$$

$$S_{r,s}^{*B} = \left( S_{r,s}^{B} \left( 1 - TINY \right) \right) + \left( S_{r,s}^{A} \left( TINY \right) \right)$$

Where:  $S_{r,s}^{A}$ = The proportion of migrants from country r to the total in country s in the foreign born matrix.

$S_{r,s}^{B}$ = The proportion of migrants from country r to the total in country s in the nationality matrix.

$TINY$ = Small number

# References

Bilsborrow, R.E., G. Hugo, A. S. Oberai and H. Zlotnik 1997, *International Migration Statistics, Guidelines for Improving Data Collection Systems*, Geneva, International Labour Office.

Borjas, G. 2003, "The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market," *Quarterly Journal of Economics,* November, pp. 1335-1374.

Commander, S., M. Kangasniemi and L. A. Winters 2004, "The brain drain: curse or boon? A survey of the literature", in R. E. Baldwin and L. A. Winters (eds.) *Challenges to Globalization: Analyzing the Economics*, Chicago, University of Chicago Press, pp. 235-278.

Docquier, F. and A. Marfouk 2005, "International migration by education attainment 1990-2000", in C. Ozden, and M. Schiff (eds.) *International Migration, Remittances and the Brain Drain*, Washington, World Bank, pp. 151-199.

Harrison, A., T. Britton, and A. Swanson 2003, *Working Abroad: the Benefits from Nationals Working in other Economies,* Paris, Organisation for Economic Co-operation and Development.

Hatton, T. and J. Williamson 2002, "What Fundamentals Drive World Migration?" Working Paper No. 9159, Cambridge, MA, National Bureau of Economic Research.

Heleniak, T. 2002, "Migration Dilemmas Haunt Post-Soviet Russia", Washington, Migration Policy Institute, at [www.migrationinformation.org](www.migrationinformation.org)

Lucas, R. E. B. 2005, *International Migration and Economic Development: Lessons from Low-Income Countries*, Cheltenham, Edward Elgar.

McKenzie, D. and N. Hildebrandt 2005, "The effects of migration on child health in Mexico." Washington, World Bank Policy Working Paper No. 3573.

McKenzie, D. and H. Rapoport 2005, "Migration and education inequality in rural Mexico," Stanford Center for International Development, Working Paper No. 258.

OECD 2002a, *Migration and the Labour Market in Asia*, Paris, Organisation for Economic Co-operation and Development.

OECD 2002b, *Trends in International Migration, (SOPEMI 2002)*, Paris, Organisation for Economic Co-operation and Development.

OECD 2003, *Another Look at the International Comparability of Migration Statistics,* Paris, Organisation for Economic Co-operation and Development, Working Party on Migration, Directorate for Employment, Labour and Social Affairs.

Ozden, C. and M. Schiff 2005, *International Migration, Remittances and the Brain Drain.* Washington, World Bank.

Parsons, C. R., R. Skeldon, T. L. Walmsley and L. A. Winters 2005, "Quantifying the international bilateral movements of migrants", Brighton, Development Research Centre on Globalisation, Migration and Poverty, University of Sussex, Working Paper No 13.

Passel, J.S. 2005, "Unauthorized migrants: numbers and characteristics", Background Briefing Prepared for Task Force on Immigration and America's Future, Washington, Pew Hispanic Centre.

Pritchett L. 2006, *Let Their People Come: Beating the Gridlock on Global Labor Mobility*, Washington, Center for Global Development.

Ratha, D. 2006, 'South-South migration and remittances', mimeo, Washington, World Bank.

Rauch J. 2001, "Business and social networks in international trade", *Journal of Economic Literature* I39 (4): 1177-203.

Srivastava, R. and S. K. Sasikumar 2005, Migration and the poor in India: an overview of recent trends, issues and policies, in T. Siddiqui (ed.), *Migration and Development: Pro-poor Policy Choices*, Dhaka, University Press, pp. 157-216.

Stark, O. 2003, "Rethinking the brain drain', *World Development,* 32:15-22.

United Nations 1980, *National Migration Surveys: The Core Questionnaire* (New York, United Nations), Comparative Study on Migration, Urbanization and Development in the ESCAP region, Survey Manuals, volume II.

United Nations 1998, *Recommendations on Statistics of International Migration*, Revision 1, New York.

United Nations 2002, *International Migration Report, 2002*; New York, Population Division, Department of Economic and Social Affairs, ST/ESA/SER.A/220, 2002.

United Nations 2004, *Trends in Total Migrant Stock 1960-2000, 2003 revision,* Department of Economic and Social Affairs, POP/DB/MIG/Rev.2003.

Walmsley, T. L. and L. A. Winters 2005, "Relaxing the restrictions on the temporary movements of natural persons: a simulation analysis" *Journal of Economic Integration*, vol. 20 (4): 688-726.

Walmsley, T. L and R. McDougall 2004, "Using entropy to compare IO tables", GTAP Research Memorandum 9, Center for Global Trade Analysis, Purdue University.

**Authors**

Christopher Parsons is a Visiting Fellow at the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex, Brighton, United Kingdom. Email: C.Parsons@sussex.ac.uk

Ronald Skeldon is a Professorial Fellow at the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex, Brighton, United Kingdom. Ph: +44 1273 877 238. Fax: +44 (0)1273 673563 Email: R.Skeldon@sussex.ac.uk.

Terrie Walmsley is Assistant Professor and Director of the Center for Global Trade Analysis, Purdue University, 403 W. State St, West Lafayette IN 47907. Ph: +1 765 494 5837. Fax: +1 765 496 1224. Email: twalmsle@purdue.edu

L. Alan Winters is the Director of the Development Research Group, The World Bank, 1818 H Street N.W.Washington, D.C. 20433, USA. Ph: +1 202 458-8208. Fax: +1 202 522-1150. Email: Lwinters@worldbank.org