EMIG 1.2: A Global Time Series of Annual Emigration Flows

Jonathon W. Moses*

ABSTRACT

EMIG 1.2 is a new, open-source emigration database, based on over 6,500 observations covering 155 countries between 1850 and 2008, which can be used in both aggregate and TSCS analyses. Using comparisons, I show that EMIG 1.2 complements and extends alternative databases. Still, I need to solicit help from area specialists to broaden and deepen its coverage. In its aggregate form, I observe two important trends: (1) the level of emigration today is lower than it was prior to the First World War (when weighted by the number of countries or people); and (2) global emigration rates have been falling since 1994.

INTRODUCTION

Both research and policy communities have become increasingly aware of the importance of migration as a global phenomenon. Initially, this focus was trained on the effects of immigration in developed countries, but the past decade has brought a growing realization of, and interest in, the potential effects of emigration on sending countries.

Unfortunately, this growing interest has been hamstrung by the appalling state of the migration data. Compared to the data used to measure other global flows (e.g., finance and goods), global migration data are notoriously patchy and unreliable. It is not surprising, then, that migration is often left out of more empirically oriented studies of globalization. There are, of course, several good (and familiar) reasons for this, including the fact that migrants often avoid being counted; the status of the migrant is often in dispute; and states have not always prioritized the careful counting of migrants (in contrast to their evident willingness to monitor capital flows, for example).

Even if the reasons for this lack of data are understandable, the status quo is untenable. As a community of scholars interested in the causes and effects of migration, we need a record of migration flows that extends back in time and across the globe – one that is open, accessible and replicable. If we are to say anything useful about the effects of migration, we need to begin with some baseline measure of variations in migration flows, across times and countries. This baseline should build on national statistics in a way that allows aggregation to reveal global patterns. In this paper, I propose such a baseline measure, and appeal to the community of migration scholars for help to fill in the gaps.

In particular, this paper provides a closer look at a global emigration data set employed in my book, Emigration and Political Development (Moses 2011). In doing so, I am motivated by the

* Department of Sociology and Political Science, Norwegian University of Science and Technology (NTNU), Trondheim.
belief that some data is better than none, and by the hope that this data can provide a solid founda-
tion upon which more and better emigration data can be assembled in the future. At the same time,
I believe that this task is bigger than any one researcher, as completing the database will require
the sort of language skills and access to archives that only a collective effort can deliver. For this
reason, I propose an open-source endeavour: I am willing to monitor and maintain the database,
but I need the help of others to hang more meat on the skeleton introduced in the pages that fol-
low.

Towards that end, the remainder of this paper is divided into four sections. In the first, I survey
the state of the existing databases with a focus on emigration-oriented data. In the second section, I
introduce the EMIG 1.2 data set, with all its blemishes. (The full data set, in a number of different
forms and formats, is available to the public for download at http://www.svt.ntnu.no/iss/Jonathon.
Moses/EMIG/index.htm.) This section describes the aggregate picture that results, although the
underlying database is itself constructed with annual, country-level data.

In the third section of the paper, I compare this new data set with a handful of alternatives, so
that readers can get a feel for its strengths and weaknesses. Here, I argue that the trend and scale
of EMIG 1.2 are consistent with those provided by alternative data sets, drawing from different
sources, but which only cover a more limited terrain (either temporally, or in terms of the number
of countries covered).

In the fourth section I consider the utility of three different weighting schemes. The first considers
the possibility that the evident trend is being driven by the number of observations in a given year.
The second measures emigration flows relative to the number of states in the international system
(which increases substantially over the period considered). The third approach measures global emi-
gration relative to the world’s population. The first two comparative measures help us to assess the
sensitivity of the EMIG data set relative to important shortcomings in the data themselves; the third
measure helps us to assess the relative size of emigration (in per capita terms) over time.

The results are both familiar and surprising, and are summarized in the concluding section. The
emigration trend revealed in the unweighted EMIG 1.2 series is camel-shaped, with two prominent
peaks (immediately prior to the First World War and in the mid-1990s), with the more recent peak
surpassing the former. The lowest emigration levels were recorded period to 1870 and in the period
immediately prior to, and after, the Second World War. This is not surprising. What does surprise
is the noticeable decline in global emigration since the mid-1990s.

When the EMIG 1.2 database is weighted, another surprise is revealed. The nature of the trend
changes when the data are weighted for the number of states in the system, and/or for the global
population, in a given year. In particular, the second (more recent) peak becomes less prominent
and smaller in the weighted trends, so that the relative number of global emigrants in recent years
is shown to be smaller than during the pre-war period. Just as interestingly: the apparent drop-off
in global emigration, beginning in the mid-1990s, is evident in each scenario, whether weighted or
not. This evidence seems to support claims that we may be approaching the end of a Third World
emigration life cycle (Hatton and Williamson, 2009).

THE STATE OF THE DATA

The most commonly used data sets count immigrant flows into, and/or stocks in, the developed
world, whether they are collected by the host country itself – for example, in the United States
(USA), the Department of Homeland Security (DHS, 2009) – or by international organizations that
collect and store immigration-oriented data (e.g. EUROSTAT, 2009; MPI, 2009; SOPEMI, 2008).
By their very nature, these data tend to focus on developed-world trends, rely heavily on data from
the developed world, and vary significantly in content and quality from state to state.
While the situation is improving, with several international organizations striving to collect and coordinate international data of both flows and stocks of international migrants, most efforts still aim to capture recent immigrant flows. As far as I am aware, there is not a single database that strives to collect annual migration flows or stocks, for every country, over a significant period of time. The two closest candidates (McKeown, 2004; UN, 2009) are considered in a comparative light below.

Most available data sets are limited to a relatively small number of years in the post-war era. This emphasis is regrettable, if understandable, as we know that the nature of the global migration regime changed significantly after the First World War, and that there was substantial international migration in the late nineteenth and early twentieth centuries.

That is not to say that historical data are not available. Indeed, there are a handful of historical data series, most of which lean heavily on Ferenczi and Willcox (1929). Unfortunately, the Ferenczi and Willcox data (covering 1850–1924) were collected in the USA and were aimed at capturing European (or trans-Atlantic) flows. More global attempts, such as Mitchell’s *International Historical Statistics*, cover the Americas (2007), Africa, Asia and Oceania (2003a) and Europe (2003b), but the migration data included in them are patchy – with only a few countries enjoying complete time series.

For those who want to study emigration flows, the choices are even more meagre. The reasons for this are self-evident: emigrants are often difficult to count, as many do not signal their intent to migrate; emigrants often leave poor countries – countries that do not have the resources to count their emigrants; countries choose different means for measuring and categorizing emigrant/departure flows (when they choose to measure them at all); and emigrants themselves are often unaware of their status when leaving – many may think they are leaving for short stays abroad, and end up settling.

Still, students of emigration are not left without options; it is only that the existing options are not very attractive. For example, the ILO (2009 has attempted to collect emigration figures, but its data is limited to just 56 countries and a relatively short period of time (1986–2007). A better alternative may be the UN’s (2009 net migration data, which provide a glimpse of net-migration stocks in particular territories over time (in 5-year intervals), at a time when a given territory was often reclaimed by different states. These data draw on a wide variety of sources and include refugee and immigrant destination figures. Most of the data are contrived, in the sense that they are constructed from demographic/census data (i.e. the difference between overall population growth and natural increases). Despite their shortcomings, these data at least provide for a consistent measure across time and space. (For a description of sources used for each country included in the UN data set, see http://esa.un.org/wpp/sources/country.aspx.) Both of these data sets, however, are limited to recent years. Finally, various issues of the UN *Demographic Yearbooks* have published emigration data in a rather sporadic fashion (i.e. these data are spread across several years and issues and they vary in form as well as in the willingness of countries to volunteer them). To my knowledge, this time series has not been tapped as a source for cross-national statistical studies of emigration.

Even when data are found, they tend to be shrouded in biases. These include the tendency to focus on intercontinental routes, at the expense of important long-distance migration patterns across individual continents (e.g. within Europe or Asia) and the difficulty of counting return migrants (the number of which can be significant). In the older data, there are class biases that result from only counting ship passengers who travelled in third class or steerage (it was these people that were automatically categorized as “immigrants”, “emigrants” or “labourers”). Worst of all are the idiosyncratic approaches of states with respect to measuring emigration flows: some states choose to count total “departures”; others total emigrants; some choose to distinguish between citizen and alien emigrants; some change the way in which emigration is measured over time, and still others – such as the USA – have chosen not to collect emigration figures at all.
Almost everybody recognizes the need for better migration statistics. The problem is: where do we start?

My approach is simple: *I include only formal (mostly government) figures for the number of emigrant citizens, leaving to a foreign country, annually.*

This is a very conservative approach, on several counts. First, this approach does not try to capture total emigration flows – only those flows that governments are willing and able to count. As a result, this measure will be smaller than competing measures, and the differences generated can provide fodder for analysis. This conservative approach is helpful in so far as it diminishes the likelihood of double counting; but it also provides me with a consistent source of emigration – to ensure a solid trend over time.

Second, this approach embraces a narrow measure of emigration. For example, many states collect the number of annual departures, and this (larger) figure might be used as a surrogate for emigration. But doing so would grossly inflate the number of emigrants, and this inflation would grow over time, as the number of non-migrant departures (e.g. tourists) is likely to rise as the cost of international travel falls. Similarly, by collecting citizen emigrant data, I have intentionally left out alien emigration figures – which many countries collected, especially prior to the First World War.

Third, countries sometimes collect and publish emigrant stock data, generated by national censuses. Including these data would have allowed me to expand the data set in significant ways – but at the cost of mixing flow and stock types of data. For this reason, only flow data were included, unless the survey referred to the year of exit for the emigrant in question.

Finally, in more recent years, I had to decide what to do with foreign guest workers. As the documentation of guest workers tends to be better than for more general emigrants, and because we know that a significant portion of early international emigrants returned home, I decided to include these numbers when they were available. In short, for countries with large formal migrant-worker export programmes (Bangladesh, India, Indonesia, the Philippines, Pakistan, Sri Lanka and Thailand), I used the officially reported outflow of migrant-worker figures, rather than the formal emigration figures, if there was any discrepancy between the two (for elaboration, see Moses, 2011: ch. 4).

When trawling through several successive and overlapping issues of the UN *Demographic Yearbooks*, I had to develop a strategy for addressing changes over time, as countries often update their time series. My solution was to begin entering data from the earliest issues, and then recheck the data at each subsequent issue, up to the last. Going through subsequent publications allowed me to double-check previous entries, ensuring more accuracy. When a discrepancy arose, I included the more recent figures, assuming that countries correct for errors over time.²

In doing this, I drew mostly from the following six data sources: Ferenczi and Willcox (1929), Mitchell (2003a,h, 2007), UN (1958), UN (1979), the UN *Demographic Yearbooks* (various years) and the ILO (2009). The basic components of each of these six data sets (and the resulting EMIG 1.2 data set) are shown in Table 1. What is perhaps most remarkable about these data sets is their limited reach (either temporally or cross-sectionally) and the number of gaps in each. It is not unreasonable to assume that every country in the world, in a given year, can expect at least one person to emigrate. If this is a reasonable assumption, then some of these data sets are reporting a remarkably small coverage share, or amount of emigrant activity. For example, the *Demographic Yearbooks* offers observations for less than 30 per cent of the total possible.³ Other data sets – for example, UN (1979) – do much better, capturing as much as 97 per cent (from an admittedly limited sample of just 19 countries).

Once these data were collected from each of the six data sets, I then aggregated them into one large database. Doing this provided me with an opportunity to cross-check data, looking for entry

---

© 2012 The Author. *International Migration* © 2012 IOM
errors and overlap in the different time series. At this point, I visited the homepage of the national statistical office/bureau for each country listed in the aggregate database (or the inheritors of now extinct states). If a country published a full series for the period under consideration, I stopped collecting data at this point, for that country. If any holes remained in the series, I then e-mailed the statistical bureau in question, described the data I had, and asked if they were aware of where I could find the missing data. I then followed any instructions I received from the responding agency. When there was a conflict in the data received from the statistics office and those found in the published reports, I relied on the former.

With this data in hand, now assembled in a large aggregate database, I waded through it – country by country – to piece together the most consistent and reliable time series possible. When there were discrepancies across databases, I examined the component parts to see which data set best reflected my original approach. I then tried to keep the longest stretch of uninterrupted data from the same source (to ensure comparability) or I chose the data set that seemed to be in line with the longest run of, and the most reasonable, numbers. In those countries where there was a need to choose between competing data sets, an individual country file was assembled, where the competing data were listed, so that researchers can see where the final data originated (and why).

EMIG 1.2 is the result of this streamlined data set, and its global aggregation is depicted as Figure 1. Its main descriptive components are listed at the bottom of Table 1, so that the reader can get an idea of how it compares with its component data sets. EMIG 1.2 covers the period from 1850 to 2008, includes 155 countries and 6,537 observations. Given its temporal and spatial spread, we should not be surprised to find that it has the smallest coverage share (about 25%) of all the data sets.

A more detailed table of countries, years covered and sources used, as well as a list of countries not included in the data set, can be found on the data set’s Internet page. These tables reveal that some countries have very full emigration records; others have only a single year (this is necessary to be included in the data set). My hope is that country/area specialists will be able to survey these tables and know whether more or better data actually exist. If better data exist, and fit the main approach described above, then I invite these specialists to share that data, so that I can update and post subsequent improvements of EMIG.

From Figure 1, it is possible to note five characteristics of the general trend, only one of which is particular surprising. First, the level of recorded emigration is not inconsistent with more generalist depictions of global emigration flows. This is not surprising, as we are drawing from the same underlying databases. Second, the cyclical nature of short-term global emigration trends is noteworthy – there is a steady flow and ebb in the aggregate data, as if it were responding to global business cycles or some other patterned event. Third, most of the inter-war period was characterized by
very low levels of international emigration – mirroring the Great Depression era of the 1870s and before. Fourth, it is possible to detect two main peaks in the overall trend: one in 1913 (when about 5.4 million people emigrated); another, higher, peak in 1994 (when almost 8.8 million emigrated). Finally, it may surprise readers to learn that the number of recorded global emigrants has been falling since 1994. This was happening even allowing for the fact that the number of states reporting emigration figures was growing during this time (see below).

COMPARATIVE ANALYSIS

Given many of the caveats associated with these data, as described above, it is reasonable to wonder how reliable they are. As a number of countries are not contributing data in any given year, and because of the wide variance in contributions over time, it is legitimate to wonder whether the global emigration trend evidenced in Figure 1 is a reasonable indicator of real-world events.

While it is difficult to provide a firm answer to this question, it is possible to compare EMIG 1.2 with other collections, to gauge whether these data sets are revealing similar patterns and levels. Should EMIG 1.2 reveal a level and pattern of global emigration that is similar to other, independent, measures, then we can be more confident that its (admittedly patchy) data are providing a reliable indicator.

In this section, I compare EMIG 1.2 with three different data sets, each of which can be used to take a different cut on global emigration trends and levels. The first point of comparison is offered by Adam McKeown’s innovative data set, which builds on three regional tallies (trans-Atlantic, South-East Asia and North Asia) to produce 5-year annual averages for the period from 1846 to 1940. By aggregating these data, I have been able to produce an estimate of global migration levels and trends.
It should be emphasized that these figures are not directly comparable with the EMIG data, as McKeown’s objectives, approach and data requirements are very different from mine. Like EMIG, McKeown draws heavily on Ferenczi and Willcox and Mitchell for documenting the trans-Atlantic flows, but he also includes transcontinental migrations (within countries), such as Russian emigration to Siberia and Chinese emigration to Manchuria. As McKeown is interested in documenting the Asian migration streams that are not well represented in the existing literature, he draws on a wide variety of sources, estimation techniques and anecdotal evidence. For these reasons, much of his data does not fit with my core approach, and thus could not be employed. Still, as a point of comparison, there is no reason to expect the trend or scale of McKeown’s data to differ significantly from those found in EMIG 1.2.

A second point of comparison is offered by the UN’s net migration data (UN, 2009). Like the McKeown data, this data comes packaged in 5-year intervals, but covers only the period after 1950, with projections into the future. Unlike both the McKeown and EMIG data sets, the UN data capture emigrant stocks, not flows. Still, the advantages of the UN data are twofold. First, it is complete, in that it reports data for each country in the world for the entire period under consideration. Second, the UN data do not rely on the willingness of countries to report emigration data. Because these data are mostly generated from underlying demographic data and receiving country statistics of refugee and immigration stocks, the UN can estimate the net migration in and out of a given state/territory, even if its political status is changing (e.g. the region is gaining independence). For these reasons, the UN data should provide the best comparative indicator of both the level and trend of international migration in recent years. These data have the additional benefit of being able to provide a nice control for the sort of political effects that are necessarily inherent to the EMIG data.

A third point of comparison is provided by US immigration figures. As the USA is often recognized to be the world’s premier immigration magnet, and has been for some time; and because the US authorities have maintained good immigration statistics over long periods of time, we might expect that the US immigration data would capture global emigration trends (albeit at a lower level). While the US Department of Homeland Security offers several competing indicators for immigration (including the numbers of refugees, green cards granted, people naturalized and short-term visitors), I have chosen to rely on their data for “Persons Obtaining Legal Permanent Resident Status” (DHS, 2009). Obviously, this is only a rough indicator of annual immigration. A priori, we should expect EMIG 1.2 to mimic the trend found in the American immigration data. At the same time, juxtaposing the two data sets provides for a sort of check on the EMIG data – as we know that global emigration figures must be larger than US immigration numbers.

Figure 2 compares these three very different data sets with the EMIG data, now transformed as 5-year annual averages (to facilitate comparison). The figure suggests that the EMIG 1.2 database offers a reasonable measure of global emigration flows for the 159 years under consideration.

After all, the EMIG data set tends to mimic the overall trend found in the US immigration data, if accentuating the height of the two peaks (the twin US peaks are more evident when a different scale is used on the vertical axis – as the level of US immigration is significantly lower than that in McKeown and in EMIG 1.2). Similarly, the parallel trending in the UN and EMIG 1.2 databases is quite encouraging. In both of these comparisons, the similarities across data sets suggest that EMIG is indeed providing a good indicator for global emigration trends.

But the differences revealed in the comparisons are also interesting, in that they provide food for further thought. On the one hand, I find it remarkable that the EMIG tallies of global emigration exceed those of the UN’s net migration data, given the patchy nature of the EMIG data, and the willingness of states to hide important emigration information. On the other hand, some of this difference might be explained by the fact that the UN is reporting net migration rates (and net return rates can vary significantly). It is also interesting to note the growing gap between the US immigration figures, on the one hand, and the UN and EMIG figures, on the other. This growing gap might say something about the declining role of the USA as a host for global migration, and the rising
importance of other host states (presumably in the European Union, the Middle East etc.). EMIG 1.2 also follows the same general pattern found in the (earlier) McKeown data set. Although we expected some variance, the revealed differences in magnitude between the two data sets are minor.

As EMIG 1.2 tracks both the scale and the pattern of the McKeown data (during the early period) as well as the UN data (for the later period), it should not be controversial to suggest that the EMIG 1.2 data set provides a reasonable indicator of global emigration trends from 1850 to the present.

WEIGHTING CONCERNS

In this section, I will consider three different weighting schemes. The first two consider whether the trends evident in Figure 1 are actually capturing global emigration trends and not the influence of underlying structural factors (such as the number of observations, or states, in a given year). The third approach weighs global emigration by world population estimates, to provide us with a relative (and perhaps more relevant) picture of global emigration flows over time.

In aggregating the national emigration tallies over time, I became increasingly aware of the patchiness of the national time-series data. For example, during periods of world war, many countries were not regularly reporting emigration numbers, even though many people may have been crossing (not to mention redrawing) international borders. Even in more peaceful times, there is remarkably little consistency over time in the data collections of many countries. While developed countries report fairly consistent emigration tallies, other countries – such as Angola – reveal a very inconsistent record: emigration figures appear, then disappear, then appear again, with no rhyme or reason. Consequently, we might find that small global emigration tallies are as much a function of a small number of national reports, rather than the result of fewer people actually emigrating.

This problem is not easily resolved, as there is little natural trending in the emigration data (as one might expect in other forms of demographic data). For this reason, it is risky to develop a

Sources: McKeown (2004), UN (2009), DHS (2009) and EMIG 1.2.
formula to fill in the missing cells. The other alternative – eliminating series with empty data cells – would only leave us with a data set that was mostly confined to Europe.

It is possible to control for the number of observations in a given year, and Figure 3 provides one simple attempt to do so. This weighting scheme tends to emphasize the effect of large numbers of observations (and dampens the effects of years with fewer observations), so it is important that we ignore the actual values on the vertical axis and focus on the trend.

When this is done, we see that the trend flattens out considerably, but that the same general pattern remains. There are two main peaks, the second peak being larger than the first, and emigration figures have been dropping since the mid-1990s. In other words, it would appear that the trend is not being driven by the number of observations in a given year.

A similar problem can result from the effects of a system of states that is growing in number over time. At first glance, the number of states might not seem to matter, as we are counting the number of emigrants (not states) and how it changes over time. But on some reflection we find that the number of states in the world can influence the emigration numbers in so far as the possibilities for international migration grow with the number of states. After all, if there was only one state, there could be no international migration. The demise of the Soviet Union provides a good example. The USSR was not particularly interested in publishing emigration figures, and internal Soviet migrations are not counted as part of my approach. But after 1991 we find several newly independent states on the periphery of Russia, eagerly collecting and publishing emigration figures. Consequently, the number of states in the international system (and their nature) can influence the overall trend.

The simplest way to control for the number of states is to divide the annual emigration figures by the number of states in the world system (for that year). To do this, I used statistics from The Cross-National Time-Series Data Archive (CNTSDA, 2009), which show a remarkable increase in the number of recognized states over the period under consideration (from 54 to 202!). Figure 4 then shows how the emigration pattern looks, relative to the number of states in the system.

Unlike Figure 3, the nature of the emigration pattern changes when weighted by the number of states in the system. While the trend reveals two peaks (pre–First World War and mid-1990s), the relative magnitude of the peaks is reversed. Now the largest emigrant stream is found before the First World War, and the more recent peak is dwarfed by it! This would seem to suggest that the magnitude of the recent flows is being driven by the increased number of states in the world. Holding the number of states constant over time, we would find that the current level of global emigration is actually lower than it was prior to the First World War. At the same time, the number of emigrants has been falling ever since the mid-1990s, even when we take the (increased) number of states into consideration.

Finally, many of us might be interested in the size of the global emigrant stream, relative to estimates of the world’s population, over time. This is because we know that the world’s population is growing (like the number of states!), and that the political and economic effects of migration can turn on its relative (rather than its absolute) size (Moses, 2011: ch. 3). After all, this is the reason why so much ink is spilled in the USA (and elsewhere) about whether the appropriate measure of immigration is the absolute or per capita numbers.

As in the previous example, it is relatively easy to control for the growing world population, by dividing the global emigration numbers by the world population, and multiplying by 1,000. This is done in Figure 5, with population estimates provided by Manning (2007) and the Population Reference Bureau (2008). The usual caveats about population estimates apply here: Figure 5 relies on global estimates, but the national estimates employed in the long form of the data set include somewhat questionable (non-census-based) demographic data from the less developed countries.

As in Figure 4, we find that the relative size of global emigration flows did increase before the mid-1990s, but this recent flow was much smaller (as a percentage of world population) than the flows that we saw before the First World War. As in the raw data, we find that the pre-1870 and inter-war flows were very low, and that the overall emigration trend has been dropping for the past
decade or so. But these population-weighted figures suggest that recent migration flows are much smaller than the global emigration flows prior to the First World War.

The drop-off in the relative number of emigrants after 1994 is curious, in that much contemporary discussion about international migration flows (based mostly on ILO and IOM measures of global migration stocks) suggest that the peak per capita migration rates in the 1990s equalled or
surpassed those from the 1910s. Some of this difference might be explained by the fact that EMIG 1.2 does not include undocumented migrants (which are surely greater now, given the restrictive nature of today’s international migration regime), and that the debate is being driven by figures from the early 1990s. But today’s level of global emigration, as a share of world population, is substantially below what we saw immediately prior to the First World War.

CONCLUSION

In conclusion, EMIG 1.2 provides a reasonable picture of global emigration trends over the long stretch of history under consideration. Just as importantly, EMIG 1.2 offers a solid foundation for further work. At long last, we now have an open empirical indicator for global migration trends over a long stretch of time.

This is not to ignore the many shortcomings inherent to EMIG 1.2. Clearly, the number of international emigrants is much larger than those revealed in the data set, as so many countries are unwilling or unable to count every emigrant. In addition, there will always be undocumented emigrants that manoeuvre under the state’s radar (sometimes literally!).

Also, the EMIG 1.2 data set is limited in that it does not attempt to break down national aggregates. Clearly, it would be both useful and interesting to examine other aspects of international migration flows (e.g. gender, age, end destination etc.), but limited time and resources have made it necessary to limit the query to simple national aggregates. Visitors to the Internet page will find that many data cells still need to be filled in – but doing so need not happen at the expense of additional efforts to collect and present more refined and varied data.

Despite its limited scope, the EMIG database provides a solid foundation upon which we can build. Just as important, the EMIG tallies are not inconsistent with those used in the more generalist literature. For example, in their book The Age of Mass Migration, Hatton and Williamson (1998: 7–8) estimate that on average about 300,000 people emigrated across the Atlantic each
year between 1850 and 1880; about 600,000 per year in 1880–1900, and about a million a year in the new millennium (and before the First World War). The global pattern depicted in Figure 1 is consistent with this description: roughly 700,000 annual emigrants in the first period; 2 million in the second; and almost 4 million annual emigrants, on average, recorded in the period after 1900 and before the First World War. More recently, the OECD (2009) has estimated that the developed world takes in about 5 million immigrants a year: the EMIG figures suggest that there are between 6 and 7 million global emigrants a year. In both periods, the total emigration numbers are surely larger than those documented in EMIG 1.2, but the EMIG numbers are not radically off the mark.

With the help of colleagues who are experts in specific regions and or periods, I hope that EMIG can become larger and even more reliable. My own limited language skills and archive access have been stretched to their limit. I am publishing this database in the hope that I can entice specialists to use EMIG 1.2 as a venue for collecting better and fuller numbers, so that the next generation of migration researchers will have even better data with which to work.

This effort will not be without challenges, and it will be difficult to overcome some of the inherent biases in these sorts of data. Many of these difficulties have already been discussed above, but I want to draw the reader’s attention to two other problems in this concluding section. In collecting these data, I was struck by how apparently unimportant they were to the authorities: countries would sometimes report the numbers, and sometimes not. The same country would mix different measures, and often collect/report the data in willy-nilly ways. It is a sad commentary, but policymakers do not seem to take this task very seriously. As someone who earlier studied global capital flows, I was dumbfounded by how little energy and how few resources states have been willing to invest in the collection of emigration data.

There are also significant variances in the quality of data, and these variances may have political (as well as economic) roots. It should not surprise us that some countries do not have the resources, and these countries will also be undercounted. As these countries are the source of much international emigration, this is not a minor problem. But more affluent countries also choose to avoid publishing (or collecting data) and their motivations are often political. For example, most of the states behind the Iron Curtain refrained from publishing emigration figures, and this tendency is still evident in today’s totalitarian regimes. Similarly, the USA stopped collecting emigration figures in 1957 (although I am not in a position to speculate on its reasons). On the flip side, small island nations are naturally predisposed to counting those who come and go, whereas large, land-locked countries will obviously find it more difficult. In short, the data we do collect will reflect political, economic and geographical realities that will influence the overall trends. As researchers using these numbers, all we can do is be aware of their influence and try to control for them when possible.

Still, by comparing EMIG 1.2 with other databases, we can be confident that the picture it paints is a reasonable facsimile of the underlying trends. In many ways, the resulting picture confirms our general view of the rise and fall of global migration flows: I expected the trend to reveal two peaks and valleys (one prior to the First World War and the other more recently), and few of us would be surprised by the cyclical nature of the trend. But the picture that results also contains a couple of surprises – surprises that might draw our attention in the future. The first of these is the recent drop-off in global emigration. This decline is evident in the raw data, but also in each of the weighted variants shown in the preceding section.

The second surprise concerns the relative size (and influence) of contemporary emigration flows. When these flows are weighted (by either the number of states in the international system, or the world’s population), we find that the magnitude of the most recent emigration peak is smaller than the one experienced prior to the First World War. This observation challenges the assertions of the many pundits who refer to the unprecedented levels of today’s international migration flows.

© 2012 The Author. International Migration © 2012 IOM
ACKNOWLEDGEMENTS

I am most grateful to Thomas Halvorsen, Jo Jakobsen, Adam McKeown, Indra de Soysa, the many faceless workers at national statistical offices around the world and the two anonymous reviewers at *International Migration* for their comments and suggestions as to how the database and its presentation can be improved. I alone am responsible for any remaining errors or omissions.

NOTES

1. There are only two exceptions to this general rule. In a few, very unique, cases, I have extended the inclusion criteria in order to capture important flows that were not included, or to replace extreme observations that were having an undue influence on the trends. See Moses (2011: ch. 4) for details.
2. As one of the reviewers pointed out, however, this may not be the case. It may be that errors creep up in the re-copying. Obviously, changes can also occur due to new definitions or categories.
3. The total possible number of observations is equal to the number of years multiplied by the number of countries in the data set. The coverage share is then set to be the number of actual observations divided by the total possible number of observations.
4. The original data were presented and explained in McKeown (2004), but were subsequently elaborated upon in McKeown (2007, 2008). His data include flows between major sending and receiving regions and are not strictly concerned with international migration. For example, in North Asia, most of the movement took place within the boundaries of particular states (Russia and China). When the North Asian data are removed, the total migration level drops, but the general pattern remains consistent with what we see in Figure 2. Professor McKeown generously shared his data with me for the purposes of this comparison.
5. In particular, I used a formula based on the number of emigrants in a given year ($M_t$), the total number of observations in a given year ($T_t$) and the actual number of observations in a given year ($O_t$). Thus, $M_t/(T_t + 1 - O_t)$. The weighting bias is evident if you consider three examples: one in which every country reported an emigration figure ($O_t = 100$); one in which only one country reported ($O_t = 1$); and one in the middle ($O_t = 50$). Thus, $T_t = 100$, and let us assume that $M_t = 500$. In the first scenario, we get $500(1/(100 + 1) - 100) = 500$. In the second scenario, where there is only one observation, we get $500(1/(100 + 1) - 1) = 5$. And in the middle scenario, we get $500(1/(100 + 1) - 50) = 100$.
6. Thus, in 2007, we take the world emigration rate (6,234,263), divide it by the world’s estimated population (6,602,274,812) and get 0.0009443. To get rid of the zeros, we multiply by 1,000, to give 0.944.

REFERENCES


© 2012 The Author. *International Migration* © 2012 IOM
ILO (International Labour Organization)  

Manning, S.  

McKown, A.  
2007 “Global migration and regionalization, 1840–1940”, paper for conference on Mapping Global Inequalities, Santa Cruz, California, 13-14 December.  

Mitchell, B.R.  

Moses, J.W.  

MPI (Migration Policy Institute)  

OECD (Organization for Economic Co-operation and Development)  

Population Reference Bureau  

SOPEMI  

UN  
1979 “Trends and characteristics of international migration since 1950”, Demographic Studies No. 64, United Nations Department of Economic and Social Affairs, New York.  

various years Demographic Yearbook, United Nations Department of Economic and Social Affairs, New York.