MAFE Migrations between Africa and Europe

MAFE Working Paper 35

Reconstructing Trends in International Migration with Three Questions in Household Surveys

Lessons from the MAFE project

Bruno Schoumaker[1] (UCL), Cris Beauchemin[2] (INED)

July , 2014

Funded under the Socio-economic Sciences & Humanities Theme

[1] bruno.schoumaker@uclouvain.be
[2] cris.beauchemin@ined.fr

The MAFE project is coordinated by INED (C. Beauchemin) and is formed, additionally by the Université catholique de Louvain (B. Schoumaker), Maastricht University (V. Mazzucato), the Université Cheikh Anta Diop (P. Sakho), the Université de Kinshasa (J. Mangalu), the University of Ghana (P. Quartey), the Universitat Pompeu Fabra (P. Baizan), the Consejo Superior de Investigaciones Científicas (A. González-Ferrer), the Forum Internazionale ed Europeo di Ricerche sull'Immigrazione (E. Castagnone), and the University of Sussex (R. Black). The MAFE project received funding from the European Community's Seventh Framework Programme under grant agreement 217206. The MAFE-Senegal survey was conducted with the financial support of INED, the Agence Nationale de la Recherche (France), the Région Ile de France and the FSP programme 'International Migrations, territorial reorganizations and development of the countries of the South'. For more details, see: http://www.mafeproject.com

Le projet MAFE est coordonné par l'INED (C. Beauchemin), en partenariat avec l'Université catholique de Louvain (B. Schoumaker), la Maastricht University (V. Mazzucato), l'Université Cheikh Anta Diop (P. Sakho), l'Université de Kinshasa (J. Mangalu), l'University of Ghana (P. Quartey,) l'Universitat Pompeu Fabra (P. Baizan), le Consejo Superior de Investigaciones Científicas (A. González -Ferrer), le Forum Internazionale ed Europeo di Ricerche sull'Immigrazione (E. Castagnone), et l'University of Sussex (R. Black). Le projet MAFE a reçu un financement du Septième Programme-Cadre de la Communauté européenne (subvention 217206). L'enquête MAFE-Sénégal a été réalisée grâce au soutien financier de l'INED, de l'Agence Nationale de la Recherche, de la région Ile de France, et du programme FSP 'Migrations internationales, recompositions territoriales et développement dans les pays du Sud'. Pour plus d'information, voir : http://www.mafeproject.com.

Abstract

Data on international migration trends are crucially lacking, especially in developing countries. The lack of basic information on migration is in sharp contrast with the increasing importance of migration in the population and development agenda. Demographic surveys offer unique opportunities to collect original data on international migration. A few simple questions in relatively small-scale surveys can be used to reconstruct migration trends with a reasonable degree of precision, and can greatly improve the knowledge of levels, trends, and patterns of international migration. The objectives of this paper are threefold: (1) To describe the data and method used to reconstruct trends in first departure; (2) to reconstruct trends in migration in Senegal, Ghana and DR Congo with data from the MAFE project; and (3) to assess the quality of these estimates.

1 Introduction

Data to study trends of international migration flows are crucially lacking. This is especially true in developing countries but also –to some extent– in developed nations. Census data allow estimating bilateral stocks of migrants for many countries (Parsons et al., 2007), but they give no direct information on migration flows[3]. Administrative statistics on immigration flows are mainly limited to developed countries, and suffer from various imperfections (Poulain et al., 2006)[4]. Statistics on outmigration flows are even less frequent, and are also seriously deficient (OECD, 2008). As a consequence, in most countries, reconstructing trends in departures and return is not possible with existing data. The lack of basic information on migration flows is in sharp contrast with the increasing importance of migration in the policy agenda of both sending and receiving countries.

Demographic surveys offer useful opportunities to collect original data on international migration (Kasnauskienė and Igoševa, 2010; Bilsborrow, 2007). The reconstruction of internal migration trends with survey data is relatively common (Piché, Gregory et al. 1984)Beauchemin, 2011). However, measuring trends in international migration with survey data is less frequent. The Mexican Migration Project (MMP) was a pioneer in this regard (Massey, 1987; Donato, 1998), focusing on flows between Mexico and the USA. The MAFE project[5] also collected data to estimate trends in international migration from three African countries, as well as trends in return migrations. The general approach

[3] See Abel (2013) for an interesting way of estimating migration flows from tables on stocks.

[4] Only legal migrations are recorded in migration statistics. Moreover, data published in some countries only refer to permanent migration (e.g. in the US), or exclude asylum seekers from migration statistics (e.g. in Belgium). Definitions of migration also vary across countries (e.g. 3 months in Belgium, 12 months in France).

[5] The MAFE project is coordinated by INED (C. Beauchemin) and is formed, additionally by the Université catholique de Louvain (B. Schoumaker), Maastricht University (V. Mazzucato), the Université Cheikh Anta Diop (P. Sakho), the Université de Kinshasa (J. Mangalu), the University of Ghana (P. Quartey), the Universitat Pompeu Fabra (P. Baizan), the Consejo Superior de Investigaciones Científicas (A. González-Ferrer), the Forum Internazionale ed Europeo di Ricerche sull'Immigrazione (E. Castagnone), and the University of Sussex (R. Black). The MAFE project received funding from the European Community's Seventh Framework Programme under grant agreement 217206. The MAFE-Senegal survey was conducted with the financial support of INED, the Agence Nationale de la Recherche (France), the Région Ile de France and the FSP programme 'International Migrations, territorial reorganizations and development of the countries of the South'. For more details, see: http://www.mafeproject.com/

followed in these projects is to collect a few simple questions on migration in household surveys, and reconstruct migratory trends with event history models. Such simple data can potentially improve the knowledge of levels, trends, and patterns of international migration. The availability of socio-demographic data (age, gender, education…) may also allow richer descriptions of migration than possible with other data sources. By including such simple questions in existing surveys (Labour force surveys, living standard surveys, DHS), data on international migration could be collected a relatively low cost.

The general objective of this paper is to evaluate the accuracy of trends in departure with only three questions in a household survey. Basically, the method consists in reconstructing trends with retrospective information collected on all surviving household heads' children, including potential information on first departure abroad. One problem of using retrospective information is that the information is only available for and through surviving individuals, i.e. deceased children are (usually) not recorded in household surveys; and the information is lacking for children whose parents are not there to answer at the time of the survey either because they died or moved abroad. In other terms, it might well be that the population for which migratory trends are computed is not representative of the population as a whole. Another limit of the method is that it relies on the information collected only on first departure, which could lead to an underestimation of migration in cases of complex trajectories.

The paper mainly draws on the experience of the MAFE project (Migration between Africa and Europe), that conducted comparable surveys on Congolese, Ghanaian and Senegalese migration. Data were collected both through household questionnaires in origin countries, and through individual biographic questionnaires (including full migration histories of the interviewee and his/her social circle), both in origin and destination countries. Confronting the MAFE household and biographic data allows us to evaluating the above mentioned limitations in three different contexts, where the patterns of migration, and thus the potential biases, are quite different (Schoumaker, Flahaux et al. 2013). Although the method can also be used to compute trends in return, this paper focuses on the computation of trends of departure.

After this introduction, the paper is divided into four parts. In the next part (part 2), we review some discussions about the collection of migration data in surveys carried out in origin countries and we present the MAFE data. Part 3 of the paper presents the methodology used to compute migratory trends of first departure and exposes the baseline results. These results are tested in part 4, where they are confronted to other computation methods in order to assess the limitations and strengths the proposed methodology. In the conclusion, we provide suggestions for future surveys on migration and for further research.

2 Collecting data on migration in surveys at origin

2.1 Advantages and drawbacks of household surveys

The general approach to collecting emigration data in household surveys consists in obtaining information on migrants and a comparison group. Data is collected on people who have migrated (whether they live abroad or have returned) and people who have never migrated. A great advantage of collecting data through a household survey in origin countries is the possibility of obtaining data on migration to all destinations at a relatively low cost[6]. Moreover, given that household surveys are conducted regularly in most countries, adding a few questions on migration can be quite cost-effective.

The collection of migration data in household surveys has also well-known limitations (Bilsborrow, 2007; Beauchemin, 2012). By definition, data on people living abroad are collected from proxy respondents. The information collected on emigrants in this way cannot be as detailed as the information that would be provided directly by the emigrants themselves, and this information is also thought to be less reliable (Bilsborrow, 2007). The questions on migration experience of migrants living abroad are usually few and simple. Another drawback of household survey data is that, if entire households emigrate, information on the migration of its members may not be collected (depending on who the information is collected from). As summarized by Bilsborrow (2007, p.4), "the more people emigrate from a country as entire households, the more a survey in the place of origin will fail to cover emigrants from that country, and the less useful the survey conducted only at the place of origin becomes." Working with survey data also means that the measurements are affected by sampling errors. These will broadly depend on the sample size and on the prevalence of migration. Given that international migration is a relatively rare event, oversampling households/areas with migrants is usually recommended (Bilsborrow, 2007).

Another important question is: who are the migrants on whom information is to be collected? Even though questions about migrants are commonly asked in origin households in surveys on international migration, there is no standardized methodology to register migrants. Each survey adopts its own approach to define the migrants to be included in its household questionnaire. Some define them on the basis of social obligations and expectations, as was the case for the Push-Pull project that registered *"those who are presently residing elsewhere but whose principal commitments and obligations are to that household and who are expected to return to that household in the future or whose family will join them in the future"*[7]. Others use residential criteria, as the NESMUWA surveys that registered individuals who had previously lived in the

[6] Compared to surveys collecting information in origin and destination countries.

[7] Note that this reference to the future household is conceptually problematic. Indeed, the concept of household refers to the group of people who live together in a residence, under the authority of the head, *at the time of the survey*. At another time, the group may be different (with members disappearing and new members arriving), the head might change, as well as the place of residence. The reference to the future is thus not clear at all when talking about a household: does it refer to the group, the place or the head?

household for at least 3 months and who had been living abroad for at least 6 months at the time of the survey (Bocquier 2003). In their guidelines for surveys on international migration, Bilsborrow and his co-authors recommend to define an international migrant as follows: *"a person who used to live in the country in which the interview is being conducted and was a member of the household of the person being interviewed but who left at some point during the five years preceding the interview to live abroad for at least six months"* (Bilsborrow, Hugo et al. 1997). Some censuses that include questions on international migrants focus on the former members of the household that moved to live abroad within the last five years preceding the census. Another, completely different approach consists in referring to family relationships, such as the MMP (Massey 1987) that registers all children of the household head, whatever their place of residence (in Mexico or abroad).

Obviously, the definition of the migrant population has an impact on the analysis potential. When the intention is to reconstruct migratory trends on a period of two or three decades, collecting data only on the more recent migrants (e.g. those who left during the 5 years preceding the survey) is not sufficient. Furthermore, the issue is not only to register long term international migrants but also information on individuals of the relevant comparison group (the comparable persons who could have moved but did not). When analyzing migration trends over long periods, using household members (at the time of the survey) as the reference group is thus not a valid option. Groups of people defined by permanent links (children or sibling of the respondent) are preferable. In this way, a sample of all the people who lived in the past in the origin country can be constituted, and information is collected regardless of their status at the time of the survey (living abroad or not, living in the household or not, alive or not). This is in some ways similar as the data collected on mortality from birth histories or sibling survival histories.

2.2 The MAFE data

The data used in this paper come from the MAFE project (Migration between Africa and Europe). The MAFE project is a multi-site project on international migration. Its objectives and questionnaires were inspired by the Mexican Migration Project (Beauchemin, 2012). The objectives of the MAFE project are to measure trends and patterns of migration, causes of departures and returns, and consequences of international migration on economic and family outcomes. The MAFE project includes both household and individual data, collected in cities of three sub-Saharan countries (Accra and Kumasi in Ghana, Dakar in Senegal and Kinshasa in DR Congo) and in six destination countries (Belgium, France, Italy, Netherlands, Spain, UK). The same questionnaires were used in all the settings, making data entirely comparable across countries.

2.2.1 Household surveys

Household surveys were conducted in sending countries (in 2008/2009) among representative samples of households of selected cities[8] (1,187 in Accra/Kumasi; 1,141 in Dakar; 1,576 in Kinshasa) (Schoumaker and Mezger, 2013).. Data was collected on all members of the household, as well as on a series of people related to the household. These people, who cannot be considered as "household members" [9], are the following:

1. All children of the head living out of the household, whatever their place of residence (including those who are deceased). They may be international migrants or not. This category includes thus domestic migrants (with a possible very short distance between their current residence and the surveyed household);

2. All persons living abroad and who are partner, mother or father of one of the household members[10];

3. All other persons who are living abroad, who are relatives of the household head or his/her partner and who have been in regular contact with the household over the past 12 months.

To collect individual information, the questionnaire includes an introductory module containing socio-demographic variables and a detailed module on migration experience[11]. In order to compute migratory trends, 5 simple questions were asked for all individuals (Figure 1):

- a screening question (A12) indicating whether or not each individual has lived for at least one year out of his/her origin country (whatever the time of departure)
- two questions on the first departure to another country (A13a and b) related to the year and the destination country
- two questions on first return (A13c and d), one indicating whether a return occurred, and if yes the year of first return.

Since trends were not expected to be reconstructed with more details than the year level, questions on migration were restricted to stays (abroad and or at origin in case of return) of at least 12 months.

[8] We later used the name of the countries, although the results apply to migration trends from cities.

[9] In MAFE, the household is classically defined as a group of individuals who live together and share partly or totally their resources to satisfy their essential needs (housing, eating). To be considered as members of a household, individuals must have been living there for at least 6 months or must intend to live there for at least 6 months.

[10] Note that only the living-abroad partners of household members were systematically registered in MAFE-Senegal. Mothers and fathers of household members were registered only if they contributed to the domestic economy (thus entering into the third category). On the contrary, in MAFE-Congo and MAFE-Ghana, living-abroad parents of household members were systematically registered.

[11] The questionnaire is available online at mafeproject.com.

Figure 1. Questions on the first departure and the first return from the MAFE household questionnaire (Ghana).

1st out-migration			1st return	
A12. **Has "Name" already lived outside of Ghana since he/she was born / settled the 1st time in Ghana?** 10. No, never → *Next person* 11. Yes, for less than a year → *Next person* 03. Yes, for one year or longer	**A13a.** **In which year did you/he/she leave Ghana for the first time for at least one year?** Don't Know: **How old were you / was he/she?**	**A13b.** **What was the destination country when you/he/she left Ghana for the first time?** *(i.e. the first country where he/she stayed for at least a year)*	**A13c.** **Did you/he/she return to Ghana for at least a year since you/he/she first left?** 1. Yes → *mark RETURN in the flap* 2. No → *Next person*	**A13d.** **In which year did you/he/she return for the first time?** *(Indicate the date of the first return that lasted one year or longer)* Don't Know: **How old were you / was he/she?** → *Next person*
___	Year ____ ___ years old		❑ Yes ❑ No	Year ____ ___ years old
___	Year ____ ___ years old		❑ Yes ❑ No	Year ____ ___ years old

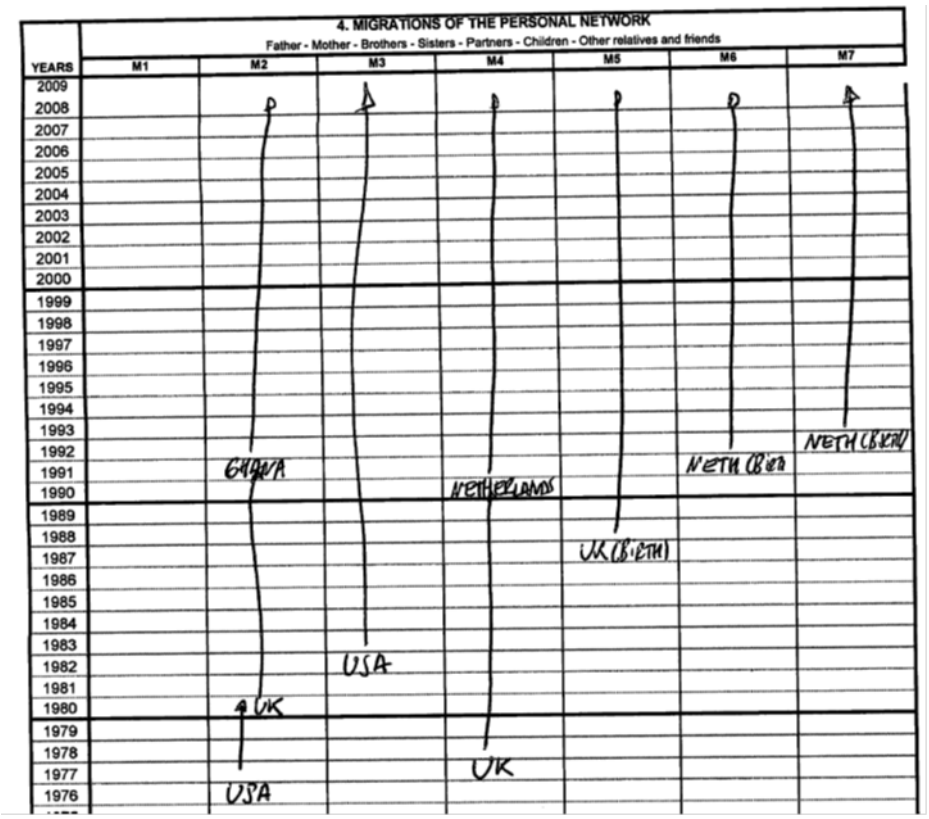
2.2.2 Biographic surveys

The MAFE biographic surveys are also used in this paper, mainly as a way to evaluate some of the assumptions of the household data and to test alternative data collection methods. Biographic data were collected among individuals aged 25 and over both in origin and destination countries: non-migrants and return migrants were interviewed in Africa (around 1500 individuals per country selected in the households), and migrants (at the time of the survey) were interviewed in six European destination countries (200 migrants per destination country: Belgium, Italy, France, Spain, The Netherlands, UK). Life histories were collected, including full migration trajectories of the interviewees[12]. Interestingly, the questionnaire also includes another useful module that can be used to reconstruct migration trends (even if it was not designed for that purpose): the so called *network module*. Each respondent of the biographic survey (regardless of his/her migration status) was asked to reconstruct migration histories (dates and countries of all migrations lasting at least one year) of a series of people who had lived at least one year out of the country of origin. Figure 2 illustrates the way data was collected in the network module. For each individual (children, brothers, sisters, father, mother, and other relatives or close friends the interviewee could have counted on to migrate), the list of all the changes of countries (for at least one year) and the dates of the changes were collected in a grid. For children, siblings and parents, the information covers the period from the first departure until the time of the survey, for the others the information is collected since the time of encounter with the interviewee. By selecting the appropriate

[12] The full questionnaire is available at migrationproject.com.

information registered in the network module, it is possible to mimic the data that would have been obtained in the household survey and thus to test some assumptions made when computing migratory trends with only the three simple questions. More generally, the network module is useful for evaluating the consistency of migration trends with two different tools. In this paper, we only use the data collected in the network module in the origin countries. The data were collected among samples of 1,062 individuals in Senegal, 1,243 in Ghana, and 1,638 in DR Congo (Schoumaker and Mezger, 2013).

Figure 2. Illustration of the network module in MAFE biographic questionnaire



3 Reconstructing trends in departure with three questions

In this section, we present the methodology used to reconstruct migration trends (departure and return) for the three flows under study in the MAFE project (DR Congo, Ghana, Senegal). Baseline results, i.e. results obtained with the three simple questions included in the household questionnaire, are also presented. To test their accuracy, they will be compared with results obtained through other computation methods in section 4.

3.1 Methodology

3.1.1 Choosing the appropriate population

As suggested before, a big stake to generate trends of migration is to capture the right population at risk of migration. Adopting a retrospective approach that looks back at

several decades forbids considering household members as the reference population because the household composition changes over time. Fixed relationships are more suited to insure that the same categories of individuals are included in the numerator (the people who migrated) and the denominator (the people who were at risk of migrating). Table 1 summarizes the list of persons included in the MAFE household questionnaire and indicates their potential migratory status according to their place of residence at the time of the survey. This table helps to identify which individuals should be selected to compute migratory trends.

Theoretically, we need to include in the analyses all people who have been at risk of moving abroad, whatever their current place of residence. The household heads are not eligible since, by definition, they cannot be migrants at the time of the survey[13]. Actually, household heads' children are the only eligible category of people since they were registered whatever their place of residence at the time of the survey (including the great Beyond). All other groups of people included in the survey are not eligible because we do not have the entire population at risk of migration (i.e. those living in another household within the country or those who deceased). As a result, in this paper migratory trends will be computed using only the information on the heads' children.

Table 1. Information collected in the MAFE household surveys on international migration by relationship to household members, migration status and place of residence (at the time of the survey)

	In the household		In another household			Deceased
	Non migrant	Return migrant	Non migrant	Return migrant	Migrant (abroad)	
Household head	Yes	Yes	No	No	No	No
Children of the HH head	Yes	Yes	Yes	Yes	Yes	Yes
Other HH member	Yes	Yes	No	No	No	No
Spouse of a HH member	Yes	Yes	No	No	Yes	No
Father/mother of a HH member*	Yes	Yes	No	No	Yes	No
Other migrant declared by the respondent	No	No	No	No	Yes	No
* This information was registered only in MAFE-Congo and MAFE-Ghana.						

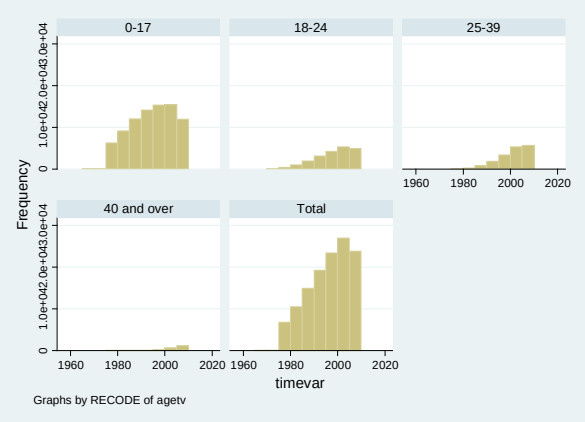
Working on the heads' children has a drawback: it underestimates the population of older adults and thus imposes an upper age limitation for the analyses. Indeed, the older adults can simply not be declared because their parents are dead. This limits the possibilities to go back in the past to describe migratory trends. This is exemplified Figure 3 with the MAFE data. The figure describes the number of person-years (i.e. the cumulative number of years lived by all children from their birth until the time of the survey or until death) and the number of events (1st departure) by age groups and by periods. Figure 3 (a) shows that the number of person-years of people at risk of experiencing migration quickly decreases as one goes back in time. Before the 1980s, the number of person-years is very low. The number of person years above age 40 is also very low before the mid-1990s. Figure 3 (b) shows the number of events by 5-year periods and by age groups. As for exposure, the number of migrations before the 1980s is very

[13] To some extent, the fact that heads declare their spouse abroad may allow us to include the heads in the analyses. It could however lead to an overestimation of return migration.

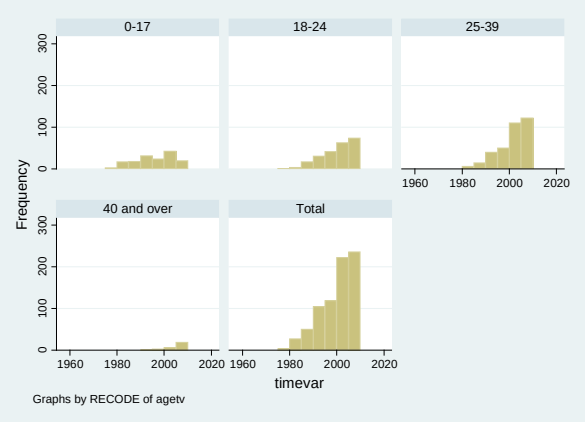
low, and events are very few above age 40. The example presented here is based on MAFE-Congo, but the same issue applies to all contexts. Analyses will thus be limited to age 40, and to periods starting in 1975.

Figure 3. Number of person years and events among children of heads of household, by 5-year periods and by age groups – DR Congo MAFE Household survey

(a) exposure



(b) events



3.2 Computing trends

We describe here how trends can be retrospectively computed using information collected in the household questionnaire on the heads children.

3.2.1 Models

As in the MMP, the trends in departure are computed using a discrete time event history model (Donato, 1998), with only age and period effects (called the age-period model). Data are organized as a person period dataset, in which each individual is represented as many times as the number of years between the time she turns 18[14] and the first migration, or age 40, or the time of the survey if the person never migrated and is under 40. The migration variable (dependent variable) takes the value 0 for all years, except for the year of migration (last year in the person period data file) if the individual migrated (value equal to 1). Age and period effects are estimated using a set of dummy variables. This model relies on the assumption that the age effect is constant over time.

$$log\left(\frac{p}{1-p}\right) = \alpha + f(age) + g(period)$$

[14] The starting time of observation can vary. Here we look only at adult migration. If the individual has migrated before age 18, he/she will not be included in the risk set. As a result, the analysis of first migration will be limited to a subsample of people who have not migrated before age 18.

Age is controlled with two age groups (18-24, 25-39)[15], and different specifications of periods are tested. Based on the age effects and the period effects, a cumulated probability of first migration is computed for each period (Donato, 1998). The indicator measures the probability that someone would do at least one international migration before age 40 if the age-specific probabilities of migration of a given period were observed in a generation. Four models are used: one for all migrations whatever the destination, and three others that distinguish broad destinations (Africa, Europe and other regions).

3.1 Baseline results

The baseline results are those computed using the three simple questions on migration included in the household questionnaire (see Figure 1). The population under study is the group of the surviving children of household heads at the time of the surveys. Migration probabilities are computed between ages 18 and 39.

Temporal frame, sample size and statistical precision

In this section, we compare estimates obtained with different degrees of temporal detail. Even though the reconstruction of yearly estimates may be desirable for a detailed analysis of migration changes (including the effects of changes in policies, impact of events like wars and crises), such estimates are hardly reliable. Apart from the fact that data collected from proxy respondents may be inaccurate, reconstructing migration by single year is affected by large sampling errors. Figure 4 shows the reconstruction of migration trends from DR (Congo) between the mid-1970s and year 2008. As is clear from this figure, confidence intervals are much too large for these probabilities to be interpreted in a meaningful way[16]. This shows a limitation of the MAFE data but it also raises a more general sampling problem because, in almost all contexts, international migration is a rare event and households with migrants a rare population difficult to sample.

As expected, confidence intervals are smaller when one adopts larger intervals (Figure 5 and Figure 6), but the precision of estimates quickly deteriorates as one goes back in time. In Figure 5, the confidence intervals in the late 1970s and early 1980s are as large as (or larger than) the probabilities of migration. These 5-year estimates may be useful to detect important changes (e.g. migration from Ghana significantly decreased between the late 1970s and the early 1980s, as expected because if the deterioration of the migrants' situation in Nigeria in the late 1970s)? But sampling errors are too large (at least before the 1990s) to be used to depict migration trends in a reliable way. In the rest of this paper, we will thus use broader periods, as in Figure 6.

[15] Several functions of age were performed and results vary very little.
[16] Samples in Senegal and Ghana are smaller, and migrations less frequent, so that confidence intervals are event larger in those countries.

Figure 4: Estimates of departure from DR Congo, by single year. (90% confidence intervals). Population: Surviving heads' children, aged 18-39. Data: MAFE Household Surveys

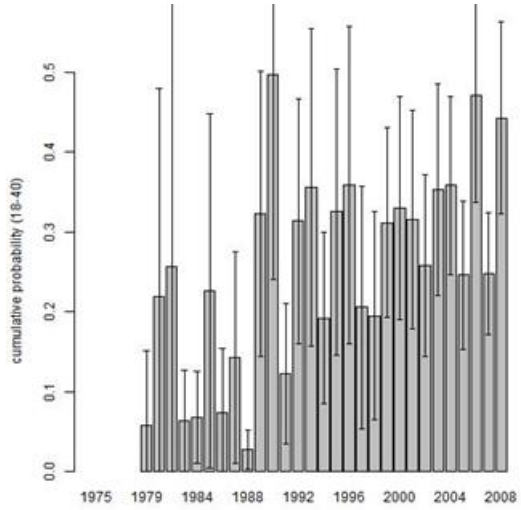


Figure 5: Estimates of departure from DR Congo, Ghana and Senegal by 5-year periods (90% confidence intervals) - Population: Surviving heads' children, aged 18-39. Data: MAFE Household Surveys

DR Congo Ghana Senegal

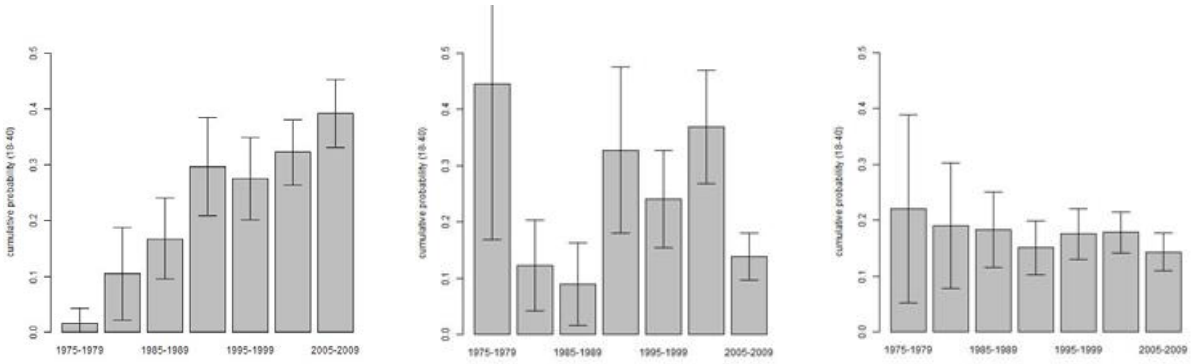
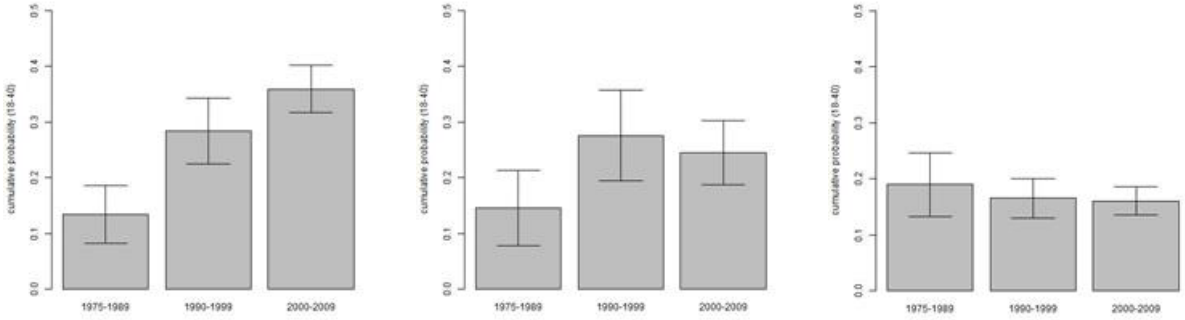


Figure 6: Estimates of departure from DR Congo, Ghana and Senegal by 3 broad periods (90% confidence intervals) - Population: Surviving heads' children, aged 18-39. Data: MAFE Household Surveys

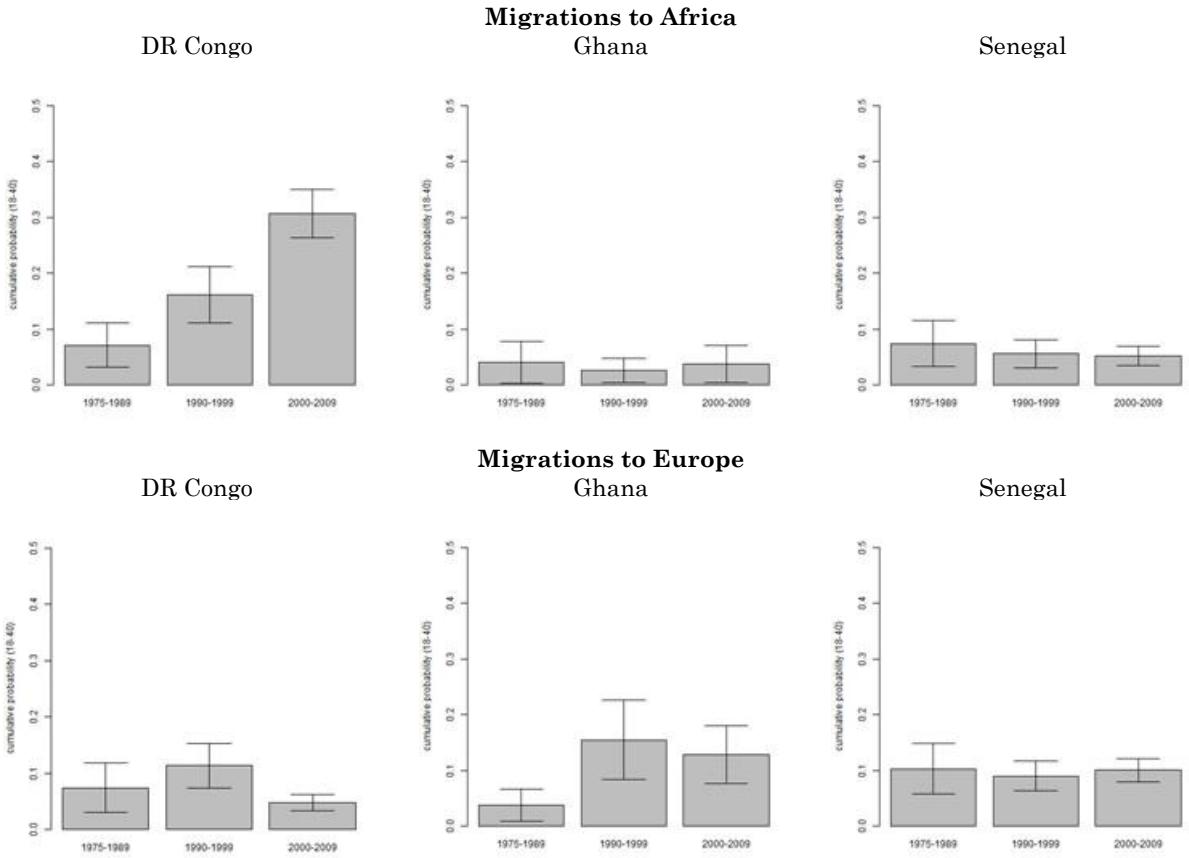
DR Congo Ghana Senegal



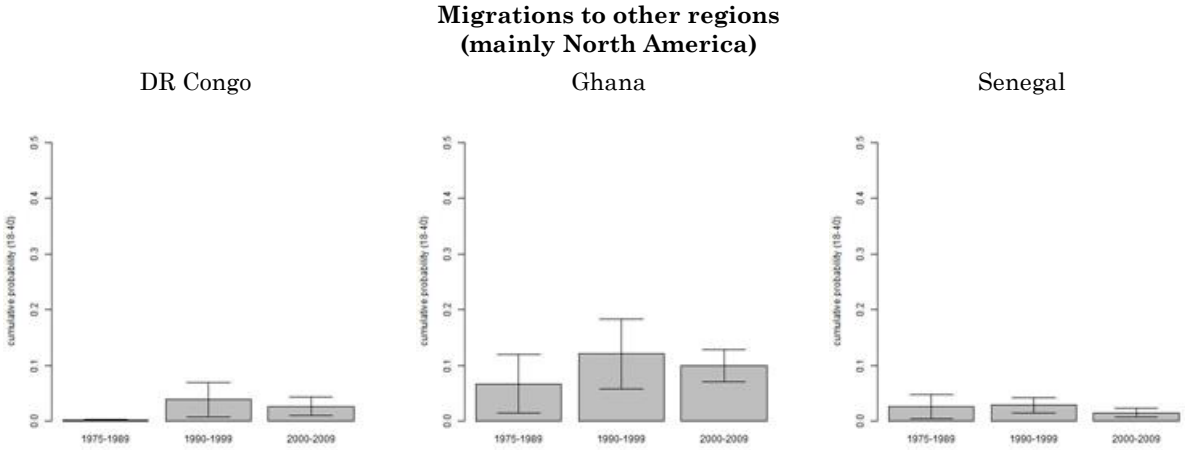
Migration trends by destinations

Reconstructing migration trends by destination is also of major importance[17], both from a policy point of view and from a theoretical point of view. However, because the number of migrations to specific destinations is smaller than migrations to all destinations together, relative sampling errors are larger, especially when results are detailed (results by 5-year periods are presented in Appendix 1).

Figure 7: Estimates of departures from DR Congo, Ghana and Senegal to Europe, Africa and other regions, by 3 broad periods (90% confidence intervals)
Population: Surviving heads' children, aged 18-39. Data: MAFE Household Surveys



[17] Even though the country of destination is collected, we consider destinations as broad regions (Europe, Africa, other regions).



Despite the limitations due to sample size, clear differences are visible by destination (Figure 7). For instance, the increase of Congolese migration results almost exclusively from the increase to African destinations. In contrast, migrations to Europe slightly increased between the 1980s and the 1990s, and then decreased significantly. Migrations to other regions have remained low. Ghanaian migrations to Europe increased significantly between the 1980s and the 1990s, and slightly decreased in the years 2000 (not significant), and a similar trend is observed for migrations to the North America (other regions), but with larger confidence intervals in the 1980s, and less robust conclusions. In contrast, migrations to African destinations remained low throughout the period. Senegalese migrations have not changed in a significant way. In short, in spite of large confidence intervals, these data allow detecting strong changes.

4 Sensitivity of the estimates

In addition of the data used in part 3 to compute baseline results, the MAFE data contains additional information that allows us to evaluate the effects on the estimates of some methodological choices made at the collection or analysis stages. In a first section, using additional information from the household questionnaire, we test to what extent the inclusion (or not) of deceased children in the risk set changes the results. In the second section, we compare the baseline results computed with the household data with results obtained with the same methodology applied to alternative data (biographic surveys). And in the third section, we assess what we call the "filtering effect" of age and destination of first migration (i.e. the fact that some information is lost on migration when we collect only the information on the first departure). Finally, the last section combines several corrections of the computed trends to assess the accuracy of the baseline results.

4.1 Including deceased children or not?

A common issue with retrospective surveys is that data is collected among surviving respondents. For instance, birth histories in fertility surveys are collected among surviving women. Full migration histories are collected among surviving migrants. Using these data to measure trends relies on the assumption that deceased people would

had similar behavior as the surviving people, and/or that their proportion in the total population is small enough to have a minor impact on retrospective estimates. These assumptions are usually thought to be robust, because mortality at adult ages is relatively small, and differential mortality among migrants and non-migrants is not expected to be large (Massey et al, 1994).

It is possible to quantify – to some extent – the impact of this assumption with the MAFE data. As in all the surveys, respondents to the MAFE household surveys were obviously alive at the time of the survey. However, data was collected on both surviving children and deceased children. In addition, the year at death was collected for the children. Even though data on mortality is not perfect, we only use information on mortality above age 18, and we are interested in finding an order of magnitude of the impact of mortality on estimates of migration.

Table 2 shows that values of the cumulative probabilities of migration do vary when deceased children are taken into account, but differences are smaller than the confidence interval. In the three countries, computing outmigration with deceased people included in the data set up to their death leads to lower estimates. This results from the negative correlation between mortality and migration (people who died were less likely to migrate). Not surprisingly, differences are larger in earlier periods (around 10% lower in DR Congo and in Senegal, around 5% in Ghana). Differences in the most recent period are negligible. Excluding deceased children thus tends to underestimate any migration increase.

Table 2: comparisons of cumulative probability of migration (18-40) by period, computed with or without the deceased children (DR Congo, Ghana, Senegal)

Country	Category	Period		
		1975-1989	1990-1999	2000-2009
DR Congo	Survivors only	0.133	0.284	0.359
	Including deceased children	0.120	0.270	0.357
	Relative difference	10%	5%	1%
Ghana	Survivors only	0.146	0.275	0.245
	Including deceased children	0.140	0.271	0.245
	Relative difference	4%	1%	0%
Senegal	Survivors only	0.190	0.165	0.16
	Including deceased children	0.172	0.160	0.16
	Relative difference	9%	3%	0%

4.2 Using network data to reconstruct trends: a comparison with household data

As mentioned before, network data will be used to test some of the assumptions of the household survey data. Before doing that we reconstruct trends in first migration using the network data and the same method as with the household survey data. We select respondents currently living in the origin countries, and full migration histories of the children of the respondent are selected. These migration histories are merged with birth histories, so that the dates of birth of all the children are available. We select the first migration of each child, to obtain similar data as the one collected in the household survey.

Figure 8: Comparisons of cumulative probability of migration to all destinations, computed with the network module and the household survey (DR Congo, Ghana, Senegal) – all destinations (90% confidence intervals)

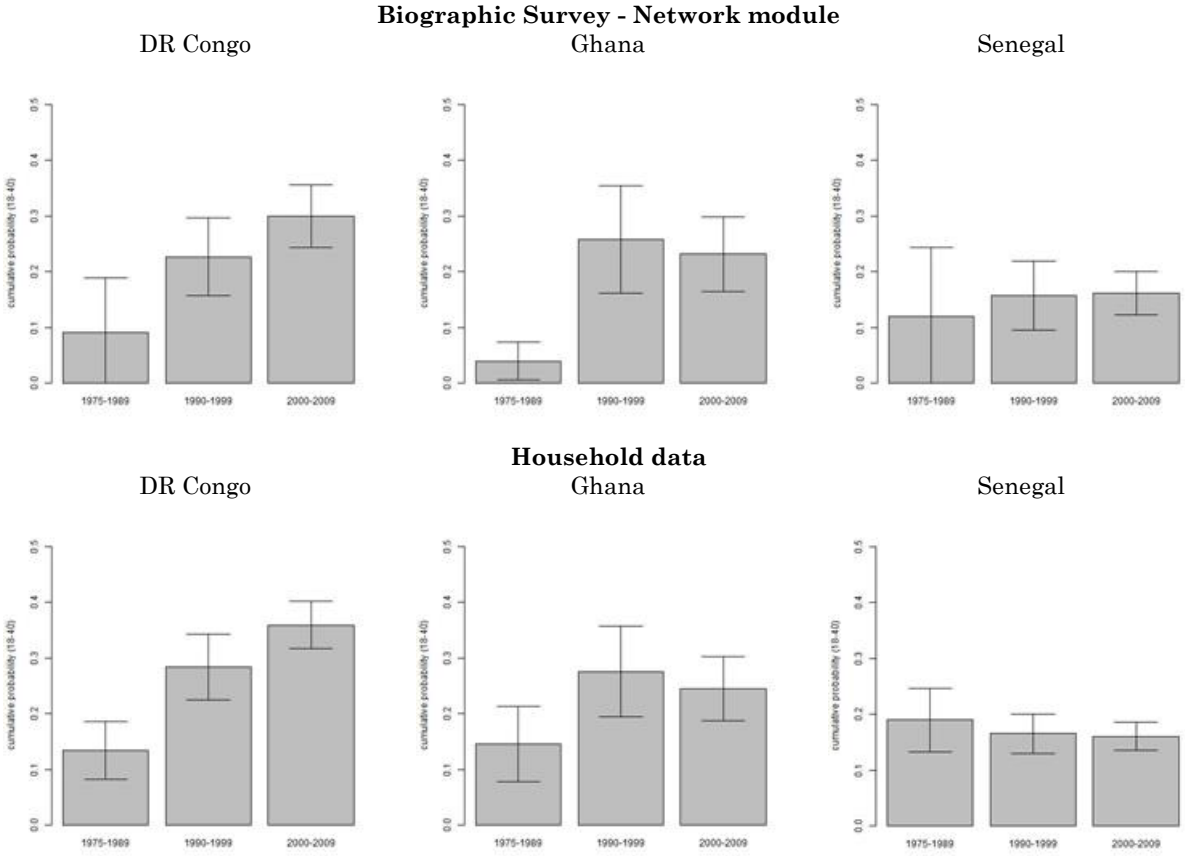


Figure 8 compares trends by broad periods from the two sources. Overall, the general trends are broadly similar, but not equal. In DRC, the level of migration is lower with the network data for all the periods. In Senegal and in Ghana, the estimates are much lower with the network data in the first period, but fairly similar in the two recent periods. The reasons for these discrepancies are not entirely clear, and could stem from the omission of early migrations in the network module, and from the fact that the risk sets are not the same. Even though the results are not dramatically different, these differences are a healthy reminder that results may vary from one tool to another and that they are essentially indicative of levels and trends.

4.3 Filtering effects of first migration: age and destination

4.3.1 Filtering effect of age

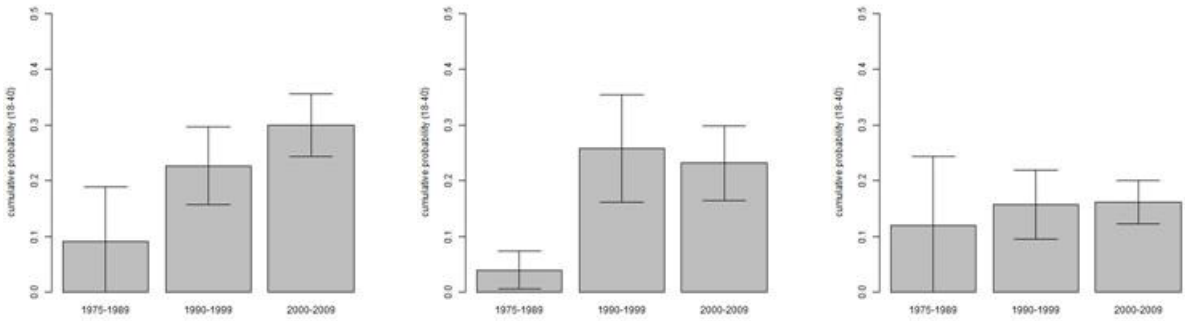
In the household survey, information was collected on the first migration of all the children of the head of the household, regardless of their age at the time of their migration. This is a sensible approach from a data collection point of view, because any additional criteria may complicate data collection and negatively impact data quality. However, one may be interested in the first migration at adult age, e.g. age 18 (as in the MAFE project), and not from birth. Because data is collected on the first migration only,

any migration before age 18 will remove the individuals from the risk set for migrations after 18. This issue is analyzed with the network module of the biographic surveys (only data collected in African countries are retained), in which full migration histories of the children of each respondent are available.

Figure 9: Estimates of departure from DR Congo, Ghana and Senegal by 3 broad periods (migrations from 18 to 40, with and without filtering effect of age) – network module (90% confidence intervals)

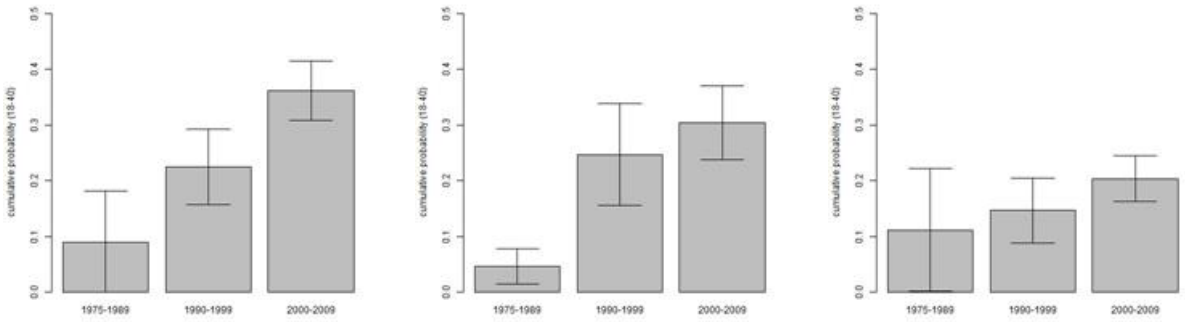
Filtering effect of age 18 (same method as with household)

DR Congo Ghana Senegal



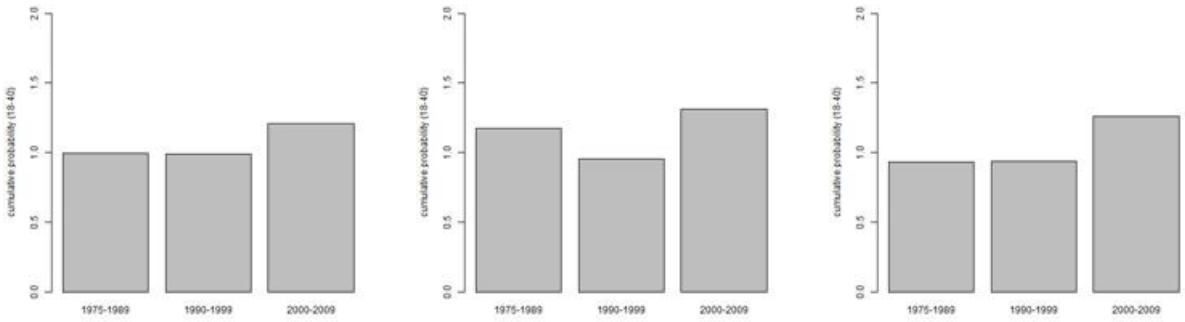
No filtering effect of age (as if question asked about first migration after age 18)

DR Congo Ghana Senegal



Ratios of probabilities with filtering effect / no filtering effect

DR Congo Ghana Senegal



Overall, the trends look broadly similar, but one difference is found in the three countries (Figure 9): the filtering effect tends to underestimate the level of recent migrations. By removing people who had done a migration before age 18 from the risk set, people with a greater propensity of migration in the recent period are removed. As a result, the increase in migration will be underestimated, reinforcing the effect on trends of removing deceased children.

4.3.1 Filtering effect of destination

Analyzing migrations by destination with the household survey data on the first migration may also affected by a filtering effect. Since only the date and destination of the first migration was collected in the household survey, any migration to a destination other than the one under study leads to censoring. For instance, a person who moves from DR Congo to South Africa in 1996 will be in the dataset until the time he/she moves to South Africa, with the value of 0 for the dependent variable in all the years. However the person may come back to DR Congo two years later (1998) and make a move to Europe in 2000. In that case, the first migration to Europe would be 2000, whereas the case is censored with household data. In a similar way, a person may move from South Africa to Europe in 2000. Such moves do not appear in the household data. This issue is potentially a problem if migrations to other destinations are not independent from migration to Europe.

As for the filtering effect of migrations before age 18, the use of network data in the biographic survey allows testing the impact of this factor in the three countries covered by the project. We measure the trends if the question on the date of the first migration was a destination-specific question (i.e. "In what year did [name] moved to [Europe] for the first time for at least one year?"). In this way, any migration to Europe would be identified regardless the person migrated to another region before[18]. This is done for the three broad regions (Africa Europe, Others). This is compared to the trends measured

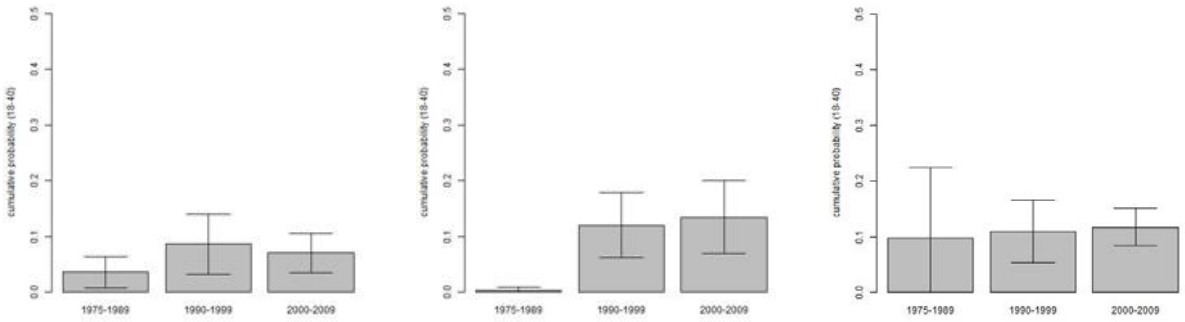
[18] One should note that the person may not have been in the origin country for all the period before the event.

with the network data with data similar as the household survey data (i.e. "In what year did [name] left [country] for the first time for at least one year?")

Figure 10: Comparisons of cumulative probability of first migration with and without the filtering effect of destination. Migrations from DR Congo, Ghana, Senegal to Europe, network module (90% confidence intervals)

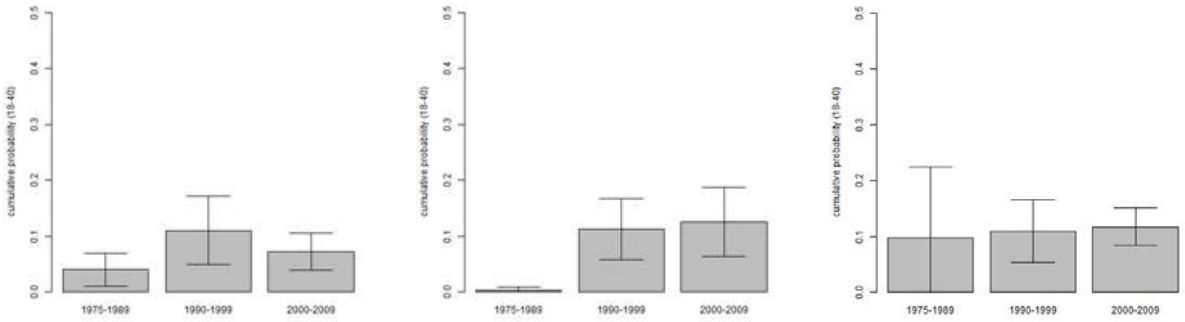
(A) Europe – Network data, filtering effect of destination

DR Congo Ghana Senegal



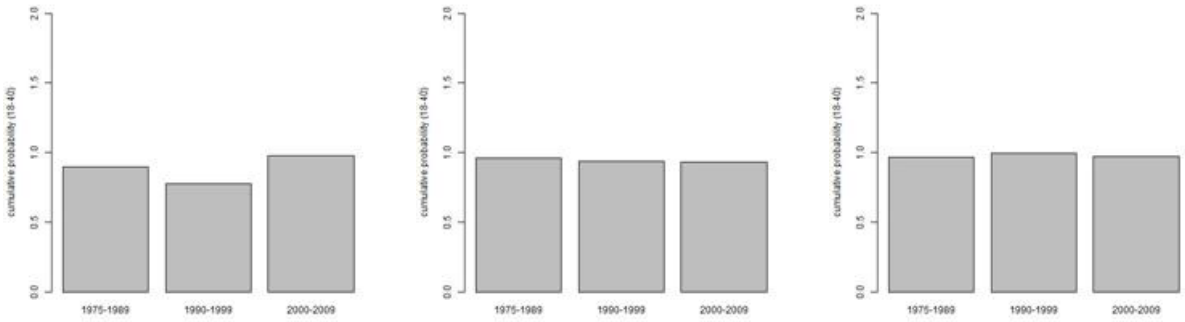
(AB) Europe – Network data, no filtering effect of destination

DR Congo Ghana Senegal



Ratios of probabilities (A)/(B)

DR Congo Ghana Senegal



Migration from Congo to Europe is a little higher in the 1990s without the filtering effect, but results are very close in the 1980s and the years 2000. This suggests that in the 1990s, people who had lived in other African country before were more likely to move to Europe, and that their migration to Europe was not their first migration. Reassuringly, the general trend is broadly similar as the one with the filtering effect and as the one with household survey data. In both Ghana and Senegal, differences are very

small, and chances of migrations to Europe are a little lower when removing the filtering effect of migrations to other destinations. This results from lower propensity of migration to Europe among those who had migrated to Africa before. All in all, results for migration to Europe appear quite robust to this filtering effect for the first migration.

Similar comparisons were done for migrations to Africa and other regions. Overall, migrations trends from DR Congo to Africa are not affected by the filtering effect (people who do their first migration to Europe or North America are not more likely to move to Africa). In contrast, results in Senegal and Ghana show that the filtering effect tends to underestimate migrations to Africa, especially in the most recent periods. This suggests that return migrants from Europe are more likely to do a move to another African destination. However, the probabilities of moving to African countries from Senegal and Ghana are quite small, and actual differences are thus also small.

4.4 Relaxing several assumptions together

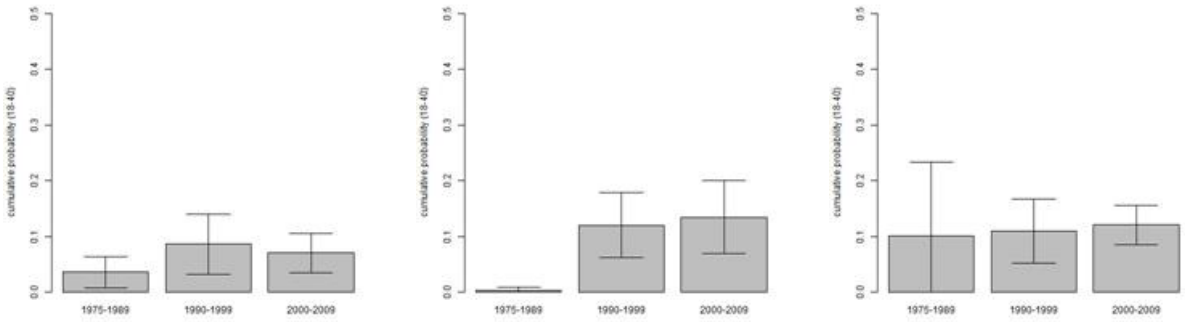
We now turn to a comparison of the trends obtained with three assumptions (deceased children not taken into account, filtering effects of age and destination, as with the household data) with the trends when several assumptions are relaxed together. The comparisons are done using the network data, and focus on migrations to Europe (Figure 11).

In all three countries, using data on the first migration of surviving children tends to overestimate the level of migration in earlier periods and to underestimate it in recent periods and as a result to underestimate any increase in migration probabilities. This results from the reinforcing effects of excluding deceased children, and the filtering effect of age. The qualitative conclusions about migration trends are not affected by these differences, in part because the probabilities of migration are relatively small, and the sampling errors are large. However, they indicate that these estimates of migration are to be taken as orders of magnitude.

Figure 11: Comparisons of cumulative probability of first migration with and without the filtering effect. Migrations from DR Congo, Ghana, Senegal to Europe, network module (90% confidence intervals).

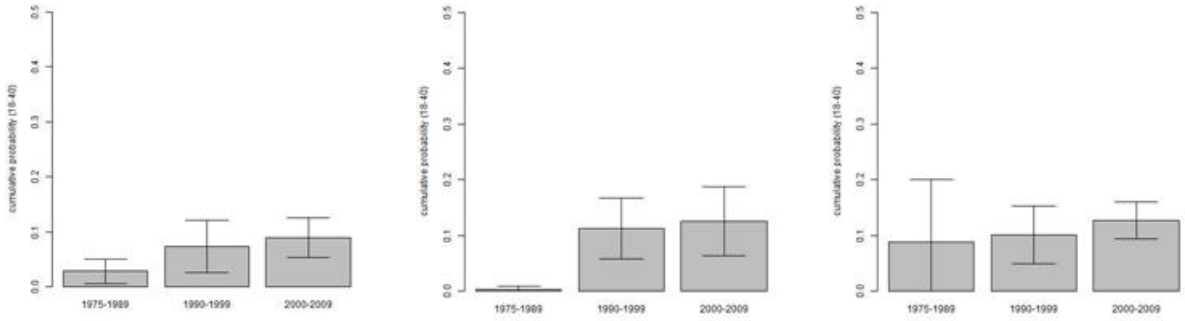
(A) Europe – Network data – same method as with household data

DR Congo Ghana Senegal



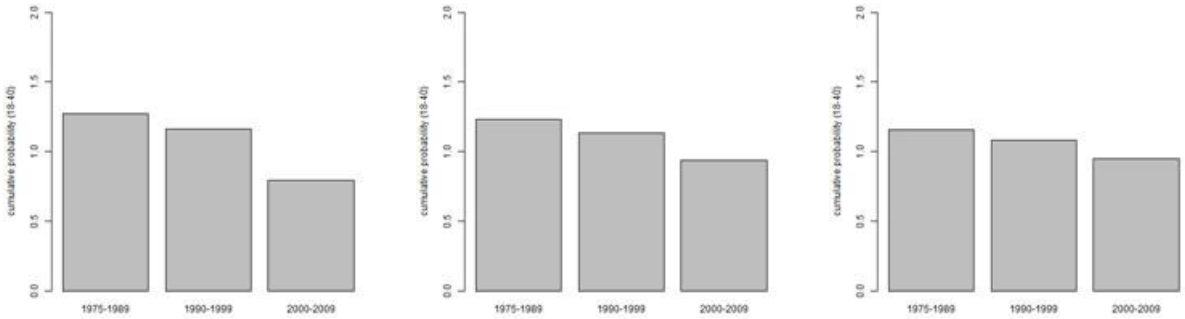
(B) Europe – Network data, no filtering effect (age and destination), deceased children included

DR Congo Ghana Senegal



Ratios of probabilities (A)/(B)

DR Congo Ghana Senegal



5 Conclusions

Simple data collected in household surveys in origin countries provide useful estimates of departures of migrants that are currently crucially lacking. The relative simplicity of the data, combined with the frequent organization of household surveys in most countries, would make it possible to greatly improve the knowledge on the levels and trends of international migration at a relatively low cost. Despite the usefulness of such

data, they are far from perfect. In this paper, we identified some limitations of this approach.

First, sample size is clearly an important issue. Our analyses relied on samples between around 1000-1500 households, and households with migrants were oversampled. The reliability of estimates of departure trends is fairly good for recent periods, but quickly deteriorates in earlier periods, because of small sample sizes. Probabilities by 5-year periods are not very precise, especially for migration trends by destinations. Any analyses on subpopulations (by gender, level of education) are obviously even less precise. Yet, despite the wide confidence intervals, the MAFE data show some clear findings. For instance, migrations from DR Congo to other African countries have greatly increased, while migrations to Europe have not changed in a significant way.

Other issues are related to the fact that the data collected in the MAFE household surveys refer to the first migration only. Although this is sensible from a data collection point of view, it may bias estimates of migration. Analyzing the first migration after a given age (e.g. 18) with such data means some individuals (those who had done a migration before 18) are no longer at risk of a first migration. Tests with data from the network module show that the overall trends in departures are not affected too much by this filtering effect of age, but this effect tends to underestimate recent migration. Using questions limited to the first migration may also have a filtering effect on the trends of migrations by destinations. Tests with the network module show that – in the MAFE countries – the impact is limited, especially for migrations to Europe. Again, it will not necessarily always be negligible, as the example of migrations from DR Congo to Europe in the 1990s show. Finally, we also tested the impact on estimates of taking into account of deceased people or not in the computation of retrospective estimates, i.e. the assumption that survivors and deceased people have the same migration behaviour. Our analyses show that, in all three countries, excluding deceased individuals from the estimates leads to overestimating migration in the past (up to 10% in DR Congo), and as a result to overestimate any increase in migration (or underestimate decreases). Again, the differences are not dramatic, and do not change the qualitative conclusions, but the influence of this assumption is not negligible. Our tests suggest that the combination of these sources of bias (excluding deceased people and filtering effect) leads to overestimating migrations in the past and underestimating it in the recent period, and as a result underestimating increases.

What do these findings mean for future surveys and future research?

The first recommendation is simple: use larger samples. However, how large a sample should be is not straightforward. Samples between 5,000 and 10,000 households, as in many household surveys in developing countries, should clearly allow more precise estimates. Yet, in multi-topic surveys, it may not be possible to oversample areas and households with high migration, so that the improvement of precision with the greater sample size may to some extent be offset by the impossibility of oversampling some areas or households.

Another suggestion is that collecting full migration histories from proxy respondents among a selected sample of people may be a better approach. Instead of collecting data with a few questions in the household roster, it is possible to collect full migration histories on some people (e.g. children of the head of household, or siblings) from a sample of proxy respondents. Such data was collected in the MAFE survey for other purposes, but it proves more flexible than the data from the household survey for analyzing migration trends. The quality of such data needs to be assessed in detail, but our experience and the comparisons with results obtained from few questions in the household roster suggest their quality can be as good as data from the household roster, without several of their limitations. Our experience with the collection of such data in the MAFE surveys is that it does not impose a heavy burden on respondents if the set of people concerned by the module is limited. Of course, additional data could be collected in such a module (e.g. level of education) for analyses for subpopulations.

The analyses in this paper have been limited to data collected among children. One limitation of these data is that, for earlier periods, very few children are exposed to the risk of migration. Moreover, there may be a correlation between the migration of the children (as adults) and the migration of their parents. Because of the potential limitations of the data on children, collecting data on siblings, as in sibling survival histories, may be an alternative approach to cover, or complement data collected on children.

MAFE data allow going further in analyzing some of these questions. For instance, the correlation of migration between parents and children was not explored in this paper, but the MAFE would allow exploring this issue. The MAFE data could also be used (with a few assumptions) to compare results from analyses using migrations of children and migrations of siblings. Microsimulations may be another strategy to evaluate some of these assumptions. Finally, pilot surveys combining different tools could also prove useful in refining data collection methods for reconstructing migration trends.

6 References

Abel, G. (2013), Estimating global migration flow tables using place of birth data, *Demographic Research*, 28(18), pp. 505-546.

Beauchemin, C. (2011), Rural–urban migration in West Africa: towards a reversal? Migration trends and economic situation in Burkina Faso and Côte d'Ivoire, *Population,Space and Place*, 17(1), 47-72.

Beauchemin C. (2012). "Migrations between Africa and Europe: Rationale for a Survey Design", *MAFE Methodological Note*, 5, INED, Paris.

Bilsborrow, R. (2007). Surveys of international migration: Issues and tips – Sixth Coordination Meeting on International Migration United Nations, New York, pp. 1–5.

Bilsborrow, R. E., G. Hugo, et al. (1997). International migration statistics: guidelines for improving data collection systems. Geneva, International Labour Office, United Nations Population Fund.

Bocquier, P. (2003). Analyzing Urbanization in Africa. New Forms of Urbanisation. G. Hugo and A. Champion. Ashgate, Aldershot U.K., IUSSP Group on Urbanisation: 133-150.

Donato, K. (1998). "Factors that influence migration. Mexican Migration Project Data", Migration Between Mexico and the United States: Binational Study. Vol. 3. Research report and background materials, MEXICO-UNITED STATES BINATIONAL MIGRATION STUDY, Mexican Ministry of Foreign Affairs U.S. Commission on Immigration Reform, Mexico City Washington, D.C, Morgan Printing in Austin, Texas.

Kasnauskienė, Gindra, Igoševa, Rima (2010). "The practice of using household surveys for estimating undeclared migration", Ekonomika, 89(3) : 122-137.

Massey D., Goldring L. and J. Durand (1994), "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities", *American Journal of Sociology*, 99(6), 1492-1533.

Massey, D. S. (1987). "The Ethnosurvey in Theory and Practice." International Migration Review 21 Special Issue Winter: 1498-1522.

OECD (2008). *Perspectives des migrations internationales*, OECD, Paris.

Parsons C., R. Skeldon, T. Walmsley and A. Winters (2007), « Quantifying international migration: a database of bilateral migrant stocks », Policy Research Working Paper, 4165, The World Bank, Washington D.C.

Piché, V., J. Gregory, et al. (1984). "L'analyse historique des migrations: la pertinence de l'analyse longitudinale à partir des histoires rétrospectives." Genus XI(3-4): 25-45.

Poulain M., N. Perrin N. and A. Singleton (eds) (2006), Towards Harmonised European Statistics on International Migration, Louvain-la-Neuve, Presses universitaires de Louvain, 744 p.

Riosmena F. (2006). "Return versus settlement among undocumented Mexican migrants, 1980 to 1996", in J. Durand & D. S. Massey (Eds.), Crossing the border: Research from the Mexican Migration Project, (pp. 265-280). New York: Russell Sage Foundation.

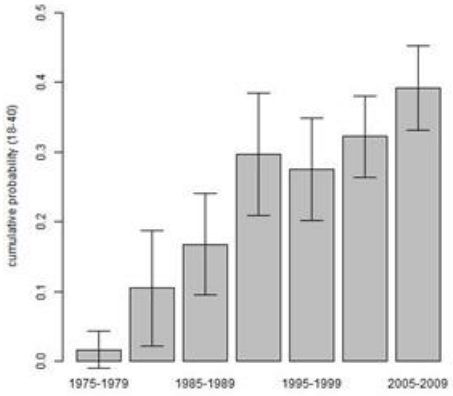
Schoumaker, B., M.-L. Flahaux, et al. (2013). Changing Patterns of African Migration: A Comparative Analysis. MAFE Working Paper. Paris, INED: 43.

Schoumaker, B. and C. Mezger (2013), Sampling and Computation Weights in the MAFE Surveys, *MAFE Methodological note, 6*, Paris, INED, 64 p.

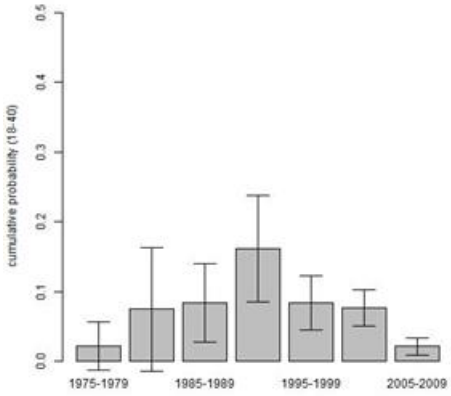
Appendix 1. Migration trends by destination, 5-year periods

Figure 12: Migration trends from DR Congo by 5-year periods (90% confidence intervals).

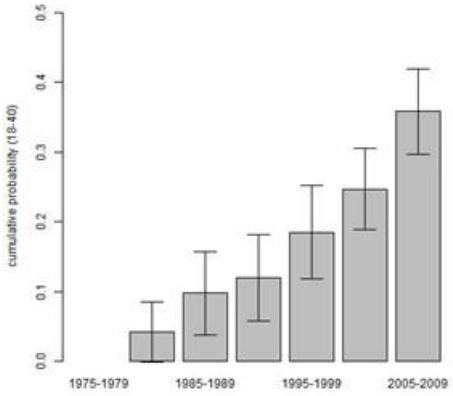
All destinations



Europe



Africa

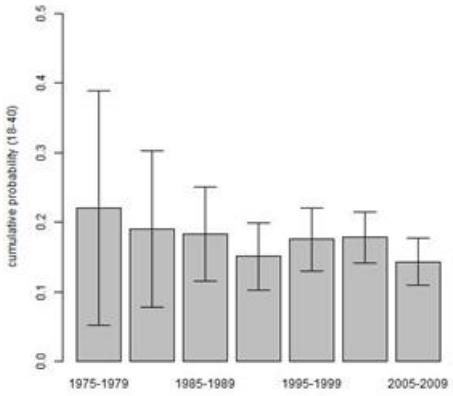


Probabilities of first migration between 18-40, computed using data on first migration from the household survey (children of the head of household).

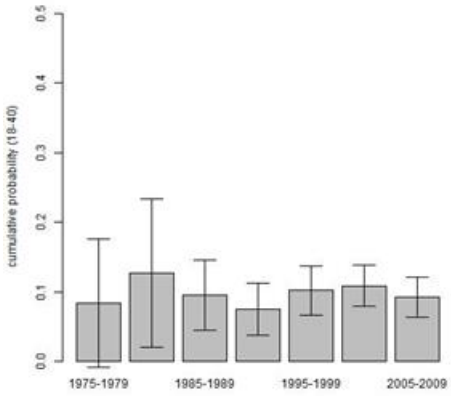
-

Figure 13: Migration trends from Senegal by 5-year periods (90% confidence intervals) (children)

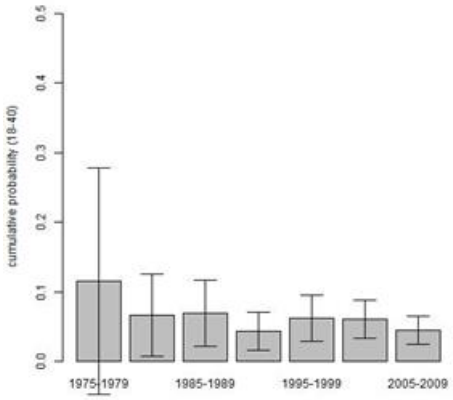
All destinations



Europe



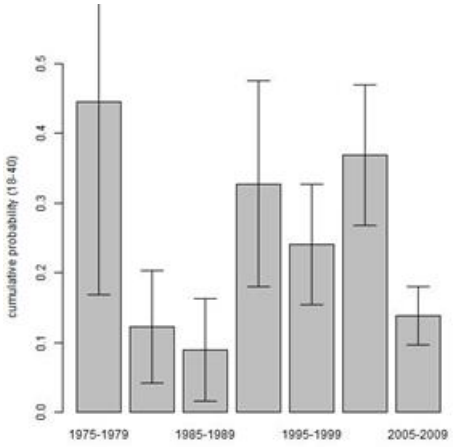
Africa



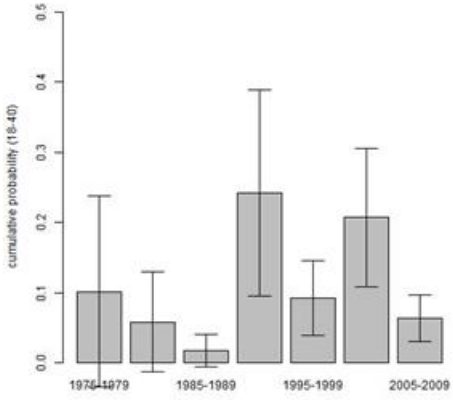
Probabilities of first migration between 18-40, computed using data on first migration from the household survey (children of the head of household).

Figure 14: Migration trends from Ghana by 5-year periods (90% confidence intervals) (children)

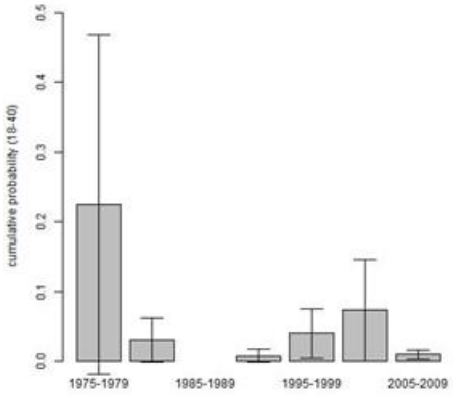
All destinations



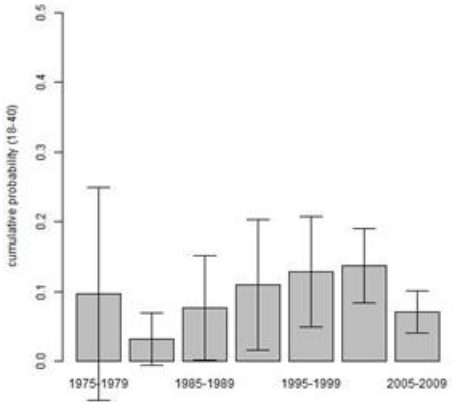
Europe



Africa



Others



Probabilities of first migration between 18-40, computed using data on first migration from the household survey (children of the head of household).