



**HAL**  
open science

# Interethnic and interfaith marriages in sub-Saharan Africa

Juliette Crespin-Boucaud

► **To cite this version:**

Juliette Crespin-Boucaud. Interethnic and interfaith marriages in sub-Saharan Africa. 2019. halshs-01834808v2

**HAL Id: halshs-01834808**

**<https://shs.hal.science/halshs-01834808v2>**

Preprint submitted on 30 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**WORKING PAPER N° 2018 – 37**

**Interethnic and interfaith marriages in sub-Saharan Africa**

**Juliette Crespin-Boucaud**

**JEL Codes: J15, J12, Z12, N37**

**Keywords: Ethnicity, Marriage, Religion, Sub-Saharan Africa**

# Interethnic and interfaith marriages in sub-Saharan Africa

Juliette Crespin-Boucaud <sup>\*†</sup>

## Abstract

This paper documents interethnic and interfaith marriage patterns to better understand which identity-related cleavages matter in sub-Saharan Africa. Using Demographic and Health Surveys (DHS) spanning 15 countries, I build a representative sample of women born between 1955 and 1989. Extrapolating to the population of these countries, I find that 20.4% of marriages are interethnic and 9.7% are interfaith, indicating that ethnic and religious differences are not always barriers. Accounting for diversity levels, both shares are actually similar. Regarding the pooled sample of these 15 countries, the share of interethnic marriages increased, and there is no country where interethnic marriages became less frequent. The share of interfaith marriages decreased in the pooled sample. Only in Cameroon did interfaith marriages become more frequent. The share of Muslim-Christian marriages remained stable in the pooled sample. The increase in the share of interethnic marriages can only partly be explained by increases in urbanization and education levels, suggesting that changes in preferences and in social norms may also be at play. The decrease in the share of interfaith marriages is due to decreasing levels of religious diversity: traditional religions were replaced by Islam and Christianity. These results show that some ethnic boundaries became more porous whereas religious boundaries did not. However, religious boundaries shifted as a result of changes in the religious landscape.

*Keywords:* Ethnicity, Marriage, Religion, Sub-Saharan Africa.

*JEL Classification:* J15, J12, Z12, N37.

---

<sup>\*</sup>Paris School of Economics (PSE); Address: Paris School of Economics, office R6-01, 48 boulevard Jourdan, 75014 Paris ; E-mail: juliette.crespinboucaud@psemail.eu

<sup>†</sup>I extend special thanks to Denis Cogneau for carefully reading my paper several times and for suggesting numerous changes that greatly improved the paper. I am grateful to Sylvie Lambert, Alexander Moradi, Lisa Oberl[Pleaseinsert“PrerenderUnicode~”intopreamble]nder, and Oliver Vanden Eynde for insightful discussions and comments on previous versions of this work. I thank seminar participants at the Paris School of Economics (PSE) and at the University of Sussex for their constructive comments.

# 1 Introduction

Social identity is an individual characteristic that has long been demonstrated to be complex and multi-layered (Posner [2005]). However, a large fraction of the literature on sub-Saharan Africa has relied on a unidimensional view of identity, equating identity with ethnicity. Even though surveys and censuses in most countries now include categories for “mixed race” or “mixed ancestry”, very few large-scale surveys conducted in African countries include such an option, perpetuating the idea that ethnic identity is the allegiance to one group or tribe, and to one homeland. Moreover, while the impact of ethnic fractionalization has been investigated<sup>1</sup>, the manner in which ethnic identities are formed and maintained has received little attention. As ethnicity is transmitted through descent<sup>2</sup>, we would expect interethnic marriages to be rare in societies where ethnic cleavages are rigid: marrying within one’s group is a means for an individual to ensure that her/his identity is passed down to her/his children<sup>3</sup>. Intermarriages have long been used to measure the strength of cleavages within societies (Kalmijn [1998]), as they combine a measure of segregation (who meets whom and where) and a measure of who is thought to be an acceptable spouse. However, there is no quantitative evidence on interethnic marriages in the case of sub-Saharan Africa. This paper aims to fill this gap.

I study interethnic and interfaith marriages in 15 countries in sub-Saharan Africa using data from the Demographic and Health Surveys (DHS). This paper has four aims: providing descriptive statistics on interethnic and interfaith marriages, discussing results at the extensive margin (marrying outside one’s ethnic group) and the intensive margin (how far outside one’s ethnic group?),

---

<sup>1</sup>Since the seminal paper of Easterly and Levine [1997], several works have pointed out the detrimental effects of ethnic diversity on growth and public good provision (Alesina and Ferrara [2005], Churchill and Smyth [2017], de la Cuesta and Wantchekon [2016], Gershman and Rivera [2018], Gören [2014]), with only a few surveys reaching different conclusions (Gisselquist et al. [2016]).

<sup>2</sup>Chandra [2006] defines ethnic membership as “determined by attributes associated with, or believed to be associated with, descent”. Kanbur et al. [2011] summarize debates over definitions of ethnic identity and of related measures. In this paper, I define ethnic groups based on DHS classification, and I do not discuss how intermarriages could have influenced or determined these classifications.

<sup>3</sup>Bisin and Verdier [2000] develop a model of transmission of identity across generations and find that intermarriages result in minority-group parents having access to a weaker socialization technology. In sub-Saharan Africa, ethnic fractionalization does not necessarily result in majority-minority settings. I nonetheless expect the theoretical result put forward to hold: high shares of intermarriage should be associated with weaker ethnic/religious affiliations for parents and their children.

assessing time trends, and analyzing which factors have contributed to the time trends. Contrasting interethnic with interfaith marriages provides a broader picture of intermarriage patterns in African countries, but it must be emphasized that religious identity is more fluid than ethnic identity, as conversion allows individuals to change their religious affiliation.

First, I find that 20.4% of married women are in an interethnic union, contrasting with 9.7% being in an interfaith union. Interethnic unions are hence far from rare events in sub-Saharan countries, and their share ranges from 10.4% in Burkina Faso to 46% in Zambia. Interfaith marriage shares range from 1.8% in Niger to 19.3% in Cote d'Ivoire. Second, using a sample of women born between 1955 and 1989, I find that interethnic marriages became more common for later-born cohorts relative to earlier-born ones, while interfaith marriages became less common. There is no country where the share of interethnic marriages decreased, and interfaith marriages increased only in Cameroon. Third, building on recent research on how to measure ethno-linguistic diversity (Desmet et al. [2016], Gershman and Rivera [2018]), I compute new linguistic distance measures that allow me to take into account diversity within and across countries. In the case of interfaith marriages, I do not use a distance measure but instead study Muslim-Christian marriages separately, as this type of union is arguably the most distant kind of interfaith marriage in sub-Saharan Africa. I find that changes at the extensive margin do not translate into changes at the intensive margin. Interethnic marriage shares increased, but there is no clear pattern regarding variation in linguistic distance. Interfaith marriage shares decreased, but Muslim-Christian marriage shares remained stable. Fourth, I examine whether time trends on intermarriage shares can be explained by increased education and urbanization levels. To do so, I compare time trends across specifications with controls and without controls. The results for interethnic marriages point at the fact that, while education and urbanization play a role in the increase of interethnic marriage shares, part of the increase could come from changes in norms and preferences about interethnic marriages. Likewise, I find that urbanization and education are not the key drivers of the decrease in interfaith marriages: this decrease is mostly due to the decreased levels of religious diversity over time.

This study of intermarriages finds that some – though not all – ethnic boundaries became more porous. Religious boundaries did not become more porous, but the religious landscape changed as traditional religions were replaced by Islam and Christianity. Finally, I confirm that my results are robust to varying definitions of intermarriages. I also test the hypothesis that spouses become more similar as the length of their marriage increases. Ethnic “assimilation” does not drive the results. However, there is evidence of conversion during marriage: my estimate is a lower bound on the decline of interfaith marriages.

This paper contributes to three strands of the literature. First, this paper extends the empirical literature on intermarriages (Kalmijn and Van Tubergen [2006]; Fryer Jr [2007]; Furtado and Theodoropoulos [2011]; Monden and Smits [2005]; Qian and Lichter [2007]; Qian and Lichter [2011]). Second, it contributes to a growing literature that nuances or contests the idea that ethnicity is always the key cleavage in sub-Saharan Africa. Contributions have suggested that which identity category is salient depends on the context (Eifert et al. [2010]; Miles and Rochefort [1991]). Looking at the micro-level literature, Berge et al. [2018] show that there is little evidence of co-ethnic bias in behavior games, contradicting results based on Implicit Association Tests (IAT) (such as Habyarimana et al. [2007] in Uganda; Lowes et al. [2015] in DRC). The complexity of relationships between ethnic groups (at the political level) was emphasized by Francois et al. [2015] and Mozaffar et al. [2003] regarding electoral coalitions and power sharing. Simson [2018] shows that once education is controlled for, public sector jobs in Kenya and in Uganda are rather equitably distributed along ethnic lines. Third, this paper adds to the literature comparing the evolution and salience of ethnic and religious cleavages (McCauley [2014]).

The rest of the paper is organized as follows: Section 2 presents the data. Section 3 lists factors that could explain the prevalence of intermarriages. Section 4 presents the empirical strategy used. Section 5 reports results on the pooled sample and section 6 results at the country-level. Section 7 tests alternative stories and provides robustness checks on the findings. Section 8 concludes.

## 2 Data

In this section, I present the data sources used and explain how the sample is built.

### Data sources: DHS and *Ethnologue*

I use Demographic and Health Surveys (DHS) that were implemented in sub-Saharan Africa (surveys used listed in Table 7, Appendix A). DHS questionnaires procure information on who is married<sup>4</sup> to whom (within the household), and in specific countries and waves, these questionnaires also procure information on respondents' ethnic and religious identity. The descriptive sample includes the 25 countries with information on ethnic identity<sup>5</sup>. The main sample is made up of 15 countries for which there are at least two survey waves that gather information on the ethnic and religious identity of respondents: Benin, Burkina Faso, Cameroon, Cote d'Ivoire, Gabon, Ghana, Guinea, Kenya, Malawi, Mali, Niger, Senegal, Uganda, Togo, and Zambia. The main sample is made up of women born between 1955 and 1989 and of their husbands.

Additionally, I exploit the *Ethnologue* dictionary (Simons and Fennig [2017]) to get information on the classification of each ethnic group's traditional language. I use these classifications to compute the linguistic distance of all of the pairs of ethnic groups. For each pair, I identify the lowest common linguistic node that they share and compute the number of nodes between each group and the common node. The mean of these two distances is the linguistic distance of this pair (detailed methodology in Appendix B).

---

<sup>4</sup>Throughout the paper I use the terms "marriage," "spouse," "husband," and "wife" to refer to married couples as well as to cohabiting couples.

<sup>5</sup>These 25 countries are the 15 countries from the main sample, plus Central Africa Republic, Chad, the Republic of the Congo (Congo-Brazzaville), the Democratic Republic of the Congo (DRC), Ethiopia, Liberia, Mozambique, Namibia, Nigeria, and Sierra Leone. The countries included in this study are not a random sample of African countries. Including a question on the respondent's ethnic identity is not a decision made independently from whether ethnicity matters in a country: the sample above cannot be considered to be representative of countries that did not include such a question in DHS.

### **Comparability over time: Reweighting and recoding**

The main sample includes at least two data waves for each country, thus raising issues about comparability over time. I explain briefly the steps taken to ensure that I can identify time trends using this sample (the online Appendix details the processes used in this study).

The main sample is made up of women born between 1955 and 1989: for each year within this period, the sample includes women from all of the 15 countries of the main sample. I reweight the sample to make it representative of the population of married women in each country. Reweighting and selecting the time period 1955-1989 ensure that the share of each country remains (roughly) constant over time. Changes over time are hence not due to changes in the respective weights of countries in the sample over time.

I recode both ethnic and religious categories to build a classification that fulfills two criteria. First, the classification does not vary within a country. Second, for all of the cohorts and survey waves, all of the groups listed in this classification have a least one member of each gender. Grouping in fewer categories mechanically decreases the number of unions appearing as interethnic/interfaith, so a time-invariant classification is needed to measure changes over time. After recoding, ethnic classifications are specific to each country and include less than 10 categories for most countries<sup>6</sup>. The category “other (ethnicity)” groups together members of ethnic groups that were not listed in all waves, people who did not identify with a specific group, and foreign nationals. Recoding religious classifications makes apparent a key change during the survey period: the surge of Pentecostalism in Africa (Mayrargue [2004], Meyer [2004]). Changes in classification are likely to reflect the agenda of church leaders, as “new Churches” have an interest in being formally recognized in order to proselytize, which is not the behavior of faiths less invested in proselytizing, such as traditional religions. Because new faith groups continue to be listed, harmonizing nomenclatures across waves requires a high level of aggregation: Christian, Muslim, other. The category

---

<sup>6</sup>Depending on the countries, ethnic classifications became more or less detailed. For instance, Akan subgroups are listed separately in Ghana in the older survey waves but are only listed as “Akan” in the recent survey waves. However, the reverse phenomenon happened in Kenya, where groups listed together (Meru/Embu) are listed separately in more recent survey waves.



“other (faith)” includes followers of traditional religions, atheists, and members of new religious movements that cannot be linked to Christianity or Islam. Among the women who belong to the “other (faith)” group, at least 41.8% identify with a traditional religion. It is a lower bound on their share, as many survey waves do not distinguish traditional religions from other faiths that do not belong to Christianity or Islam.

### **Variables: Intermarriages and individual characteristics**

To study intermarriages, I build variables that measure intermarriages as well as variables that are likely to influence the likelihood of intermarriage.

Ethnic and religious identity are self-declared in the DHS: I hence consider that the respondent’s answer is a measure of their “true identity”. A marriage is interethnic (interfaith) if the spouses’ answers correspond to different ethnic (faith) categories in the recoded classification. Ethnic and religious identity categories may be fluid and change, especially in the case of conversion for marriage. I discuss how religious conversion and ethnic “assimilation” might affect my results in section 7.

I consider two main variables that can lead to intermarrying: education and urban residence. The DHS include little retrospective information, so I cannot reconstruct the individual characteristics at the time that the marriage started. Marriage decisions are taken based on the characteristics of individuals but also on expectations, such as joining a spouse in the city or being able to graduate high school. I use characteristics at survey date to proxy for past characteristics and expectations: current characteristics do not perfectly correspond to past characteristics but allow me to take into account (realized) expectations. I use information on the highest completed level of education at survey date, which should be a good proxy of the level of education at the time that the union was formed<sup>7</sup>. I use urban residence at the survey date. Migration mostly takes place from ru-

---

<sup>7</sup>It is unlikely that women can stay in school after getting married. Considering only women in union who have attended primary school, I find that, under the assumption that girls start school at age 8, 69% of women started their first cohabitant union at least two years after completing their schooling, 21% around the same time as they completed their schooling, and 10% before that.

ral to urban areas<sup>8</sup>: the current place of residence captures some unobserved characteristics of individuals that might be correlated to their propensity to intermarry, such as one's occupation. Using this variable thus results in overestimating the relationship between urban residence and intermarriage.

### **3 Intermarriages and marriage markets: Preferences, norms, and diversity**

I provide in this section a framework for interpretation of the models and of the results presented in the paper.

#### **Types of factors influencing intermarriage shares**

In his seminal paper, Kalmijn [1998] distinguishes three factors that could explain the prevalence of intermarriages: individual preferences, diversity levels within (local) marriage markets, and the influence of norms and of third parties.

The *individual preferences factor* gathers all of the preferences that individuals have concerning their matches on the marriage market. Two main characteristics of matches on the marriage market are socio-economic resources and cultural resources: people are likely to want to marry someone whose economic prospects are good and with whom they share values and preferences.

The *diversity level factor* encompasses all of the channels related to how diverse marriage markets are. Some societies are highly heterogeneous, and others are more homogeneous, for instance if there is a majority group. Moreover, spatial segregation affects how diverse local marriage markets are. Low levels of diversity are associated with low levels of intermarriages (by sheer limitation due to the numbers of potential spouses from other groups).

---

<sup>8</sup>Among women for whom I have information on childhood place of residence, 12.6% of women who live in a rural area at the time of the survey grew up in an urban area, whereas 41.3% of women living in an urban area grew up in a rural area. As this sample consists of women belonging to earlier-born cohorts, these figures may be even higher for later-born cohorts.

The *third parties/norms factor* includes the channel of group identification, the one of group sanctions, and in the case of the setting studied, the fact that members of one's kin may be directly responsible for choosing one's spouse. Field studies have shown transitions from *kin-selected* to *self-selected* marriages, for instance, Bertrand-Dansereau and Clark [2016] (Malawi) and Clark et al. [2010] (Kenya): elders and parents are less involved in the matching process. Third parties influence is likely to work against intermarriages (Spörlein et al. [2014]).

### **DHS variables and the Kalmijn framework**

As indicated in section 2, I use two variables that are likely to influence the likelihood of intermarriage: education and urban residence. These variables capture aspects of the types of factors listed above.

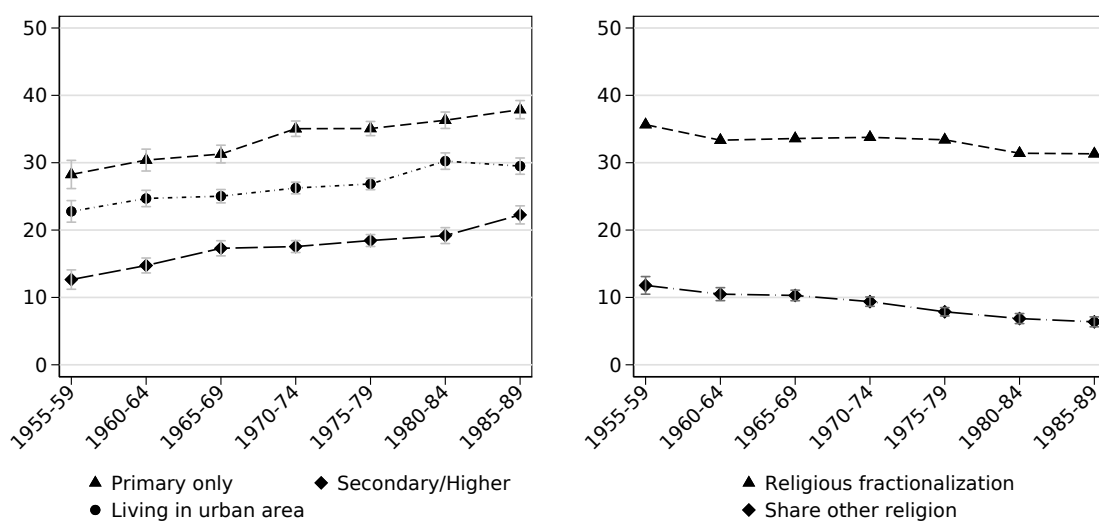
Education could affect *individual preferences* through several channels. Education, especially secondary and higher, is in many countries conducted in a vehicular language, thus helping to remove language barriers in marriage markets. Additionally, by transmitting a common culture, education could switch preferences away from group identification and towards a national identification. Moreover, higher education takes place in (mixed) urban settings (*diversity level factor*). Educated women might have more of a say in the choice of their spouse: *third parties* may be less involved in the matching process.

Urban areas are on average more mixed than rural ones: *diversity levels* are likely to be higher in cities than in the countryside, and marriage markets are likely to be less segregated. *Social norms* may be different in cities, and more accepting of intermarriages.

### **How did marriage markets change?**

My main sample is made up of women born between 1955 and 1989. During this period, several changes linked to the variables listed above took place. Figure 1 shows a visual representation of the changes that I can observe in the data. Its left panel shows that education levels as well as

Figure 1: Changes over time: Education, Urban residence, Diversity levels



Sample & data: Women in union, pooled sample. 95% confidence intervals included (except on the measure of religious fractionalization).

Left panel: Share of women living in urban area, share of women whose highest completed education level is primary/secondary school. The share of women living in urban areas reflects current place of residence: the magnitude of the change would be higher using information on type of place of birth.

Right panel: Weighted average of religious fractionalization at country-level and share of women belonging to the group “other (faith)”

urbanization increased: these changes could lead to higher rates of intermarriages for later-born cohorts than for earlier born cohorts.

Moreover, another key change took place: the decrease in the religious diversity of the population. The right panel of Figure 1 shows the decrease in the share of people identifying with faiths other than Islam and Christianity and the associated decrease in religious diversity. Under the assumption that people meet on a national marriage market and that no other factors affected interfaith marriages, the decrease in religious diversity will mechanically result in lower interfaith marriage shares. Under the same assumptions, the share of interethnic marriages should remain stable, as there is no sizable change in ethnic diversity levels<sup>9</sup>.

## 4 Empirical strategy

In this section, I present measures of ethnic and religious diversity and introduce the specifications used to measure time trends.

<sup>9</sup>A more thorough discussion of ethnic diversity at the country-level and interethnic marriage patterns can be found in the online Appendix.

## 4.1 Descriptive statistics: Comparisons across countries and across identity categories

Cross-country comparisons must be done carefully in order to be meaningful: if intermarriage shares are low in a given country, it could be that, according to the classification used, there is not much diversity in this country to begin with. As discussed in section 2 in the case of Pentecostal churches, the classification chosen in each survey cannot be considered as encompassing the same reality across countries and identity categories. To neutralize this classification effect, I compute the share of intermarriages that we would observe if individuals were matched randomly in a national marriage market. This share measures how diverse each country is. It is equivalent to computing an ethnic fractionalization index (EF) and a religious fractionalization index (RF) for each (national) marriage market. More formally, this fractionalization measure ( $F_c$ ) corresponds to:

$$F_c = 1 - \sum_{i=1}^n p_{wi} * p_{mi}$$

In this equation,  $c$  denotes countries.  $n$  is the number of ethnic (respectively religious) groups in the survey. Subscripts  $w$  denote women, subscripts  $m$  denote men.  $p_{wi}$  is the share of married women who belong to the group  $i$ ,  $p_{mi}$  the share of married men who belong to the group  $i$ . I assume that a person's decision to marry does not depend on the ethnic and religious identity of her/his matches. There are no additional entries on or exits from the marriage market compared to what we observe in the data: polygamous men are thus counted as many times as the number of their cohabiting wives.

Fractionalization indices do not include any information on how (dis)similar groups are<sup>10</sup>:

they are useful to identify changes at the extensive margin but do not allow for identification of

---

<sup>10</sup>Since the seminal paper by Greenberg [1956], adjusting measures of fractionalization by a measure of similarity between groups has been suggested to take into account the depth of cleavages: for instance, some ethnic groups speak the same language (e.g. in Zambia Posner [2005]) while others are very distant according to the linguistic tree (e.g. the Somali and the Kikuyu, in Kenya). These differences are not taken into account by fractionalization measures: the distance between all groups is set to be the same.

changes at the intensive margin.

In the case of ethnicity, I account for these differences by using linguistic distance measures (detailed methodology in Appendix B). Computing the country-specific random linguistic distance ( $\bar{d}_c$ ) is done according to the formula below:

$$\bar{d}_c = \sum_{i=1}^n \sum_{j=1}^n p_{wi} * p_{mj} * d_{ij}$$

where  $d_{ij}$  is the linguistic distance between group  $i$  and group  $j$ ,  $p_{wi}$  the share of married women who belong to the group  $i$ , and  $p_{mj}$  the share of married men who belong to the group  $j$ . The linguistic distance  $d_{ii}$  is set to 0. The linguistic distances  $d_{zi}$  and  $d_{iz}$ , where  $z$  is the group “other (ethnicity)”, are set to the median value of the linguistic distance between pairs of ethnic groups listed in country  $c$ .

In the case of religious differences, there is no standard way to quantify the differences between groups. I study separately interfaith marriages – mixed marriages between Muslims, Christians, and “other (faith)” – and Christian-Muslim marriages, the latter arguably being the most distant kind of intermarriage along the religious dimension in sub-Saharan Africa.

## 4.2 Assessing time trends

I design a simple additive model to estimate trends on intermarriages. My baseline specification is a linear probability model<sup>11</sup> that I run on the pooled sample (including country fixed-effects) and at the country level. I present below the specifications used at the country level. I specify the model at the woman level<sup>12</sup>: individuals ( $w$ ) are women. This decision is due to the fact that 17% of women appear with at least one co-wife in the dataset: a woman appears only one time in the dataset, but a man appears matched with all of his cohabiting wives.

---

<sup>11</sup>The results do not change when using a logit model. Results available from the author.

<sup>12</sup>The results do not change if I study intermarriage by collapsing the dataset to keep one observation per man. In this case, I consider a man to be in an interethnic (interfaith) union if at least one of his wives is not a member of his ethnic (religious) group. Results available from the author.

## Model 1 : Time trends

$$Intermarriage_w = \alpha + \beta_1 BirthYear_w + \beta_2 Age_w + \beta_3 Age_w^2 + \epsilon_w \quad (1)$$

$Intermarriage_w$  is an indicator variable that equals 100 if the union is interethnic (interfaith), and 0 otherwise<sup>13</sup>. I consider unions to be interethnic (interfaith) if spouses do not belong to the same group. When both spouses belong to the group “other,” I consider them to be in an intraethnic (intrafaith) union. In Section 7, I test whether the results are robust to considering these unions as interethnic (interfaith).

$BirthYear_w$  is a continuous variable defined as the year of birth of each woman. It is the main variable of interest: if the coefficient associated to it is positive, it means that the share of intermarriages has increased over time. I use birth year rather than year of first cohabitation to capture time trends for two reasons. First, birth year is available for all women, while using cohabitation year would restrict the sample to women in their first union, as I only have information on the year of first cohabitation. Second, age at marriage or year of cohabitation is endogenous to educational achievements and to the type of place of residence, while year of birth is more exogenous to these individual characteristics. I compare birth year to cohabitation year as a robustness check in Section 7.  $Age_w$  is the age at survey date.

## Age and birth year effects

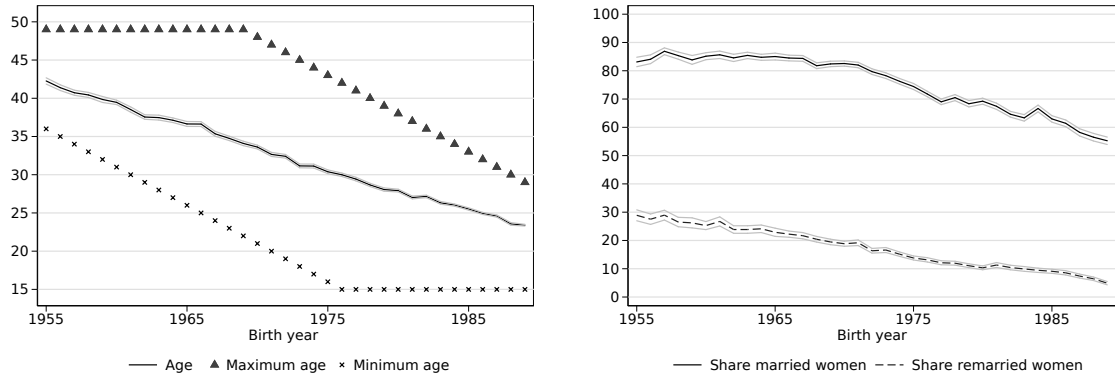
I add quadratic controls for age in the model to control for age effects. This ensures that the patterns that I identify in the data are due to change across cohorts and not to age effects<sup>14</sup>. I use surveys implemented from 1992 to 2018: women born in 1955 were older than 35 in the first DHS survey of each country; women born in 1985 were younger than 35. As shown in Figure 2,

---

<sup>13</sup>The indicator variable is set to 100 so that coefficients can be read as changes in percentage points.

<sup>14</sup>This is possible as my main sample is made of up of countries for which I have at least two survey waves. I hence observe birth cohorts at different ages. Thus, 82.7% of women belong to birth cohorts that were sampled at least twice in their country. Given the quadratic function that I use to estimate age effects, I do not need all of the cohorts to have been sampled twice to estimate age effects and birth years effects separately.

Figure 2: Survey effect: Marital status and age at survey date



*Left panel:* The sample includes only women in union at the time of the survey. 95% confidence intervals. The “maximum age” is the age of the oldest woman surveyed for each birth year. The “minimum age” is the age of the youngest woman surveyed for each birth year. These ages depend on the timing of surveys within each country. Women aged 15 to 49 are surveyed in DHS, hence the flat lines at these two points.  
*Right panel:* These shares are computed using all women surveyed. 95% confidence intervals. “Married women” are the women in union at the time of the survey, not the ever-married women.

the timing of survey waves can be seen in the age composition of cohorts: earlier-born cohorts are older at survey date than later-born cohorts. Whether a woman is married or not is a function of age: differences in age composition of cohorts (left panel) are mirrored by differences in the share of married and remarried women by cohort (right panel). As women in earlier-born cohorts are older at the survey date, they are more likely to have married and more likely to have remarried, either after a divorce or being widowed. The same characteristics are likely to drive both the type of marital status that I observe (married/remarried/never married) and the type of marital outcome that I observe (intermarried or not). For instance, if women who marry young are more likely to marry within their group, then, without age controls, I would estimate time trends that are due to the fact that cohorts differ with respect to their age composition.

### Model 2: Time trends with controls

$$\begin{aligned}
 Inter marriage_w = & \alpha + \beta_1 BirthYear_w + \beta_2 Age_w + \beta_3 Age_w^2 \\
 & + \beta_4 Primary_w + \beta_5 Secondary_w + \beta_6 Urban_w + \beta_7 * Remarried_w + \epsilon_w
 \end{aligned}
 \tag{2}$$

I introduce additional variables in the model to test whether they explain changes in inter-marriage shares. I assume these variables have a constant effect over time. I add dummies for



the highest education level:  $Primary_w$  and  $Secondary_w$ , the reference category being “no education”<sup>15</sup>.  $Urban_w$  is a dummy that takes the value 1 if the respondent lives in an urban area. Moreover, to control further for cohort composition effects, I add a dummy variable,  $Remarried_w$ , which takes the value 1 if the respondent has remarried. I discuss alternative ways to measure the impact of remarriage in section 7.

Throughout the paper, I compare the coefficient associated with birth year between the specification 1 and the specification 2. The birth year coefficient in specification 1 measures time trends. The birth year coefficient in specification 2 measures time trends that cannot be explained by changes in education levels, in urbanization, and in cohort composition due to remarriage. As such, it could capture changes in preferences and in social norms. However, it should be noted that several other variables may contribute to individuals’ likelihood to intermarry, such as parental education, whether one’s parents intermarried, or whether both parents are still alive at the time of the marriage decision: the coefficient in specification 2 may also capture some of these omitted factors.

## 5 Results on pooled sample

This section presents the results on the pooled sample. The country-specific results are detailed in section 6.

### 5.1 Descriptive statistics

Table 1 displays the estimations of observed intermarriage shares and contrasts them with the intermarriage shares that we would have observed under random matching. Interethnic unions are on average more frequent than interfaith unions: 20.4% of women are married to a man who is not from the same ethnic group as them, and 9.7% of women are married to a man who

---

<sup>15</sup>DHS classification distinguishes between secondary and higher education. Only 2.3% of married women in my sample completed university, so I aggregate secondary education and higher education into a single category.

Table 1: Average intermarriage shares and linguistic distance

Interethnic marriages (%)				Linguistic distance (nodes)			
Observed	Random	Ratio	N	Observed	Random	Ratio	N
20.4	80.0	25.5	97111	3.29	3.25	1.01	21704
Interfaith marriages (%)				Muslim/Christian marriages (%)			
Observed	Random	Ratio	N <sup>a</sup>	Observed	Random	Ratio	N
9.7	33.8	28.7	96549	2.4	21.7	11.2	83291

*Data & sample:* Women in union, pooled sample. Weighted data.

*Interethnic, interfaith, and Muslim-Christian marriage shares:* The *observed share* corresponds to the share observed in the population. The *random share* corresponds to the share that we would observe if people currently in union had matched at random, under two assumptions. First, there is no exit or entry into the marriage market compared to what we observe. Second, polygamy decisions are independent from women's ethnicity and religion: Polygamous men appear on the random market the same number of time as in the observed market. Random shares are computed for each country, considering a national marriage market. The random share for the pooled sample is the weighted average of those national random shares. The *ratio* is computed as the ratio of *observed share* to *random share*.

*Linguistic distance:* The random and observed linguistic distances are computed considering a national marriage market, using information on interethnically married couples. Linguistic distance between two spouses when only one of them belongs to the group "other (ethnicity)" is set to the country-specific median linguistic distance, computed on distances for all pairs of ethnolinguistic groups.

*Muslim-Christian marriage shares* (observed and random) are computed using only marriages in which neither spouse is member of the group "other (faith)".

<sup>a</sup> One survey wave in Senegal does not include a question on religion.

is not a member of the same religious group as them. However, the number of categories and the level of diversity differ depending on whether we consider ethnicity or faith: under random matching, we would observe around 80% of interethnic marriages and around 33.8% of interfaith marriages. When we look at the ratio of the observed share of intermarriages to the random share of intermarriages, interfaith marriages and interethnic marriages are roughly as common: between 25% and 30% of the random share of intermarriages is realized.

I find that interethnic unions take place at a linguistic distance that is similar to what we would observe under random matching. In contrast with the fact that 28% of interfaith unions are realized, Muslim-Christian marriages are rare<sup>16</sup>: they make up 2.4% of marriages when considering only Muslim and Christian respondents, and 2.1% otherwise. It is 11.2% of what we would observe under random matching. Most interfaith unions hence involve a spouse who identifies as Muslim or Christian and a spouse who belongs to the group "other (faith)". Individuals who are neither Muslim nor Christian are more likely to be in an interfaith union than Muslims and Christians.

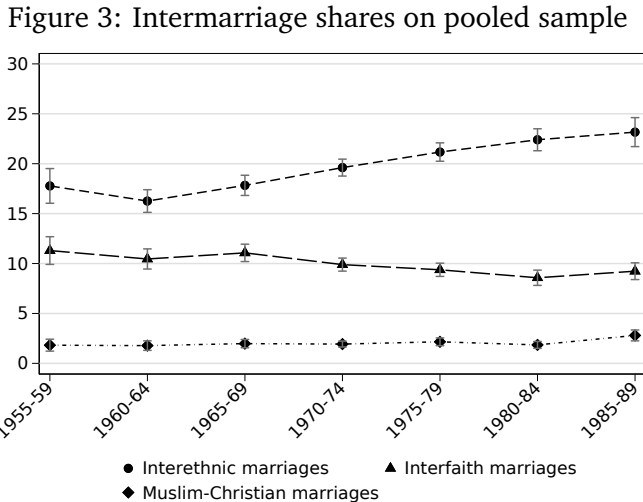
<sup>16</sup>66.5% of Muslim-Christian unions are unions between a Muslim man and a Christian woman. Most Islamic scholars consider that it is forbidden for Muslim women to marry non-Muslim men, but Muslim men can marry women who belong to other monotheist religions. Hence, this imbalance in the types of Muslim-Christian unions indicates that such unions, while rare, are not only counted due to measurement errors.

Couples that include at least one follower of “other (faith)” make up 14% of the sample, but 79% of interfaith couples, most of them taking place between a Christian spouse and an “other (faith)” spouse. The high propensity of “other (faith)” members to intermarry is consistent with the fact that traditional religions are more tolerant of intermarriages. It is also likely that the conversion process from a traditional religion to Islam or to a Christian denomination might not concern both spouses at the same time.

## 5.2 Time trends

### Assessing time trends

Figure 3 shows the shares of each type of intermarriage over birth cohort of women, thus providing visual evidence on the magnitude of changes.



Sample & data: Women in union, pooled sample. 95% confidence intervals included.  
Observed share of intermarriages by birth cohort of women.

Table 2 shows time trends for the three kinds of intermarriages studied: interethnic, interfaith, and Muslim-Christian marriages. Interethnic marriages are more frequent for women in later-born cohorts than women in earlier-born cohorts: in column (1), the coefficient for birth year is positive and significant. After controlling for education, urbanization, and remarriage (column (2)), the magnitude of the coefficient decreases, but the coefficient itself remains positive and significant.

This result indicates that other factors, possibly changes in norms and preferences, contributed to the increase in the share of interethnic marriages.

Table 2: Trends - Intermarriage shares

Dependent variable	(1) Interethnic marriage	(2)	(3) Interfaith marriage	(4)	(5) Muslim-Christian marriage	(6)
Birth year	0.265*** (0.0421)	0.168*** (0.0403)	-0.179*** (0.0289)	-0.149*** (0.0287)	0.0143 (0.0157)	0.00461 (0.0154)
Age	0.428** (0.199)	-0.152 (0.197)	-0.306** (0.141)	-0.305** (0.141)	-0.0307 (0.0777)	-0.120 (0.0770)
Age squared	-0.00515* (0.00294)	0.00202 (0.00292)	0.00131 (0.00221)	0.00134 (0.00220)	0.000225 (0.00117)	0.00134 (0.00115)
Primary		2.925*** (0.637)		0.568 (0.457)		0.568** (0.222)
Secondary/Higher		7.914*** (0.787)		-1.512*** (0.561)		1.339*** (0.306)
Urban		13.12*** (0.661)		-2.523*** (0.414)		1.141*** (0.232)
Remarried		6.443*** (0.643)		3.808*** (0.502)		1.650*** (0.285)
Controls						
Country-fixed effects	✓	✓	✓	✓	✓	✓
Observations <sup>a</sup>	97111	97111	96549	96549	96549	96549
R-squared	0.247	0.272	0.134	0.139	0.029	0.033
Mean dependent variable	20.4		9.7		2.1	

*Data & sample:* Women in union, pooled sample. Weighted data.

*Specification:* OLS regression. Standard errors are clustered at the DHS-cluster level. As the specification includes country-fixed effects, there is no constant in the model.

Columns (1) and (2): the dependent variable is a variable that equals 0 if the union is an intraethnic one, 100 if it is an interethnic one.

Columns (3) and (4): the dependent variable is a variable that equals 0 if the union is an intrafaith one, 100 if it is an interfaith one. Three faith groups are defined: "Muslim", "Christian", and "other (faith)".

Columns (5) and (6): the dependent variable is a variable that equals 100 if one spouse is Christian and the other one Muslim, and 0 otherwise. Results for all columns should be read as changes in percentage points.

Significance levels are denoted as follows: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

<sup>a</sup> One survey wave in Senegal does not include a question on religion.

Interfaith marriages decreased over time (column (3)) as expected from the decrease in religious diversity. The share of Muslim-Christian marriages (column (5)) did not vary over time: it is likely that the norms and preferences concerning interfaith marriages did not change over the period. Norms concerning religious intermarriages remaining the same is consistent with the fact that the decrease in interfaith marriages is due to the decrease in religious diversity documented in section 3 rather than to norms shifting towards less tolerance of interfaith marriages.

Extrapolating the estimations from the model over 25 years (1960-1985), the model (columns (1) and (3)) estimates an increase of 6.6 percentage points in interethnic marriage shares and a decrease in interfaith marriage shares by 4.5 percentage points. Once education and urban residence are controlled for (column (2)), the magnitude of the increase in interethnic marriage

shares is 4.2 percentage points. The fact that women in later-born cohorts are more educated and more likely to live in urban areas than women in earlier-born cohorts explains one-third of the trend on interethnic marriages. Once education and urban residence are controlled for (column (4)), the magnitude of the increase in interfaith marriage shares is 3.75 percentage points: these variables explain little of the trend observed, corroborating the idea that the decrease in interfaith marriages is mainly due to declining levels of religious diversity.

### **Assessing individual characteristics**

Looking at interethnic marriages and at Muslim-Christian marriages, the coefficients on education and on urban residence are consistent with what we would expect from variables capturing parts of the *individual preferences factor* and of the *diversity level factor*. Completion of primary school rather than having no education is associated with a higher likelihood of intermarrying. Completion of secondary school is associated with an even higher likelihood. Urban residence is also associated with an increase in the likelihood of intermarriage<sup>17</sup>. In the case of interfaith marriages, secondary education and urban residence are negatively correlated to the likelihood of being in an interfaith marriage, but these variables capture the likelihood of belonging to the group “other (faith)”. Members of this group are often followers of traditional religions, and attendance in school and urban residence are negatively correlated with the likelihood of being a member of this group<sup>18</sup>.

Remarried women are more likely to be married outside of their group than women who are still in their first union, whatever the kind of intermarriage considered. Social norms may be different for women who marry for the first time and for women who remarry, as women have more freedom in choosing a spouse when they have already been married (Locoh and Thiriati [1995]). Similarly, earlier-born women may remarry under the same set of (more accepting)

---

<sup>17</sup>Using a sample of earlier-born cohorts for which I have information on childhood place of residence, I run the specifications from Table 2. The coefficient estimated for the current place of residence is slightly higher than the coefficient for the childhood place of residence.

<sup>18</sup> 76.6% of women who belong to the group “other (faith)” did not complete primary school while 44.4% of Muslim and Christian women did not complete primary school. 12.8% of women who belong to the group “other (faith)” live in an urban area while 28.5% of Muslim and Christian women live in an urban area.

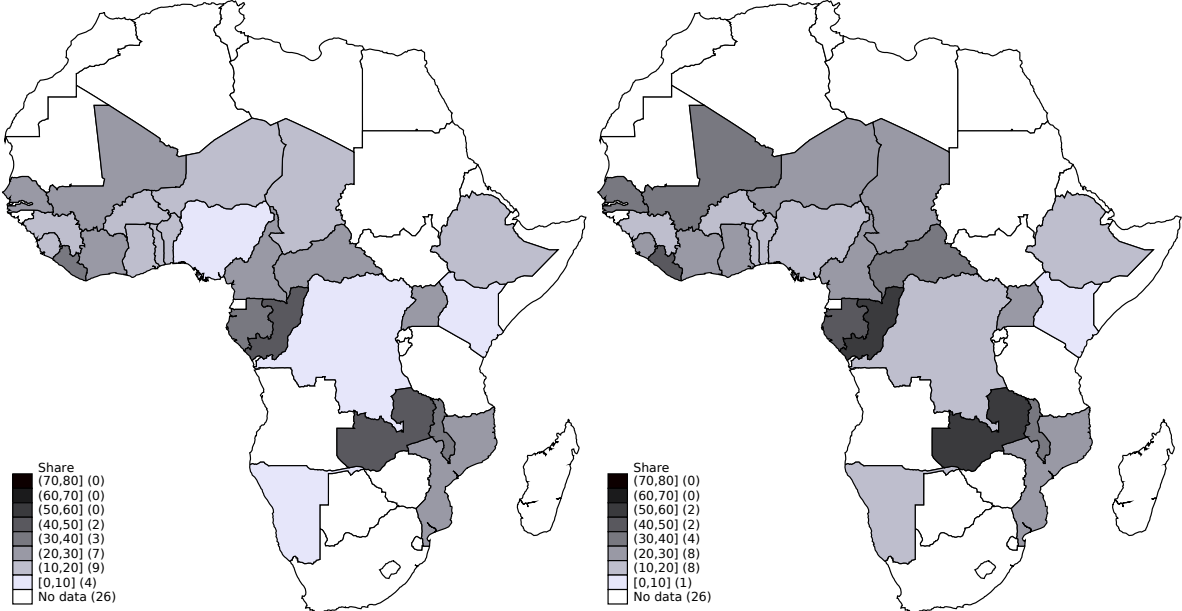
norms as later-born women who enter their first union. Last, women who remarried may have different (unobserved) characteristics that also lead them to marry outside of their group, whether in their first union or in the subsequent ones. I discuss these hypotheses about remarriage in section 7.

## 6 Results at country-level

This section presents the descriptive statistics and results on time trends at the country level. For brevity, tables include only the coefficient associated with the variable *BirthYear* and show results for the two main specifications (with and without controls). The full results at the country level and the results with control variables introduced one by one are available in the online Appendix.

### 6.1 Descriptive statistics

Figure 4: Shares of interethnic marriages

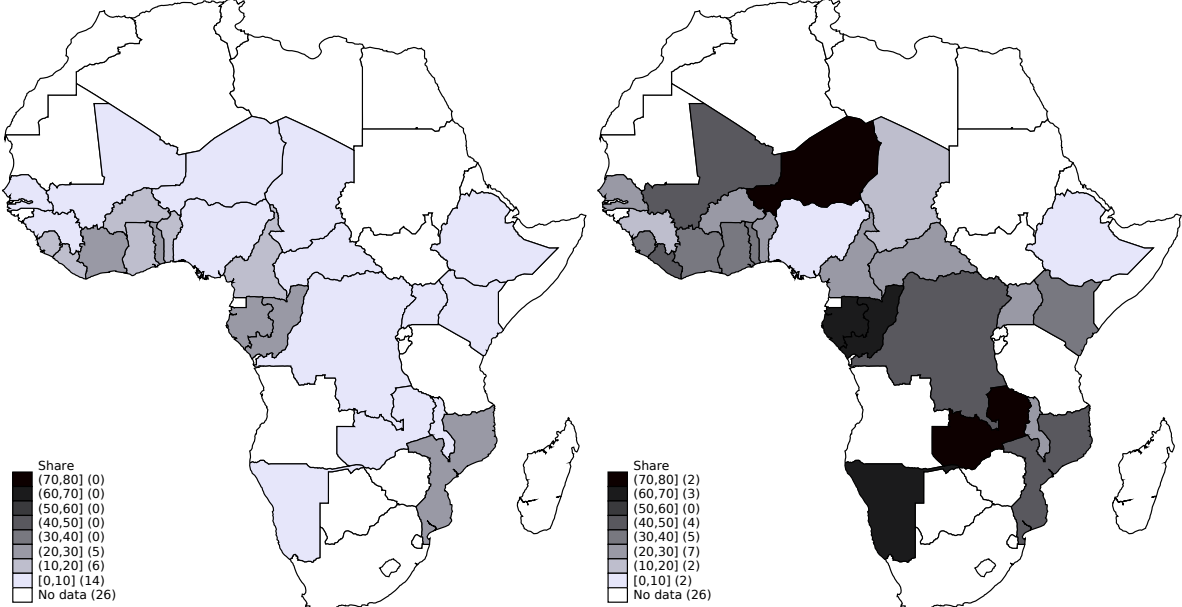


Data: Survey wave (DHS) conducted the closest to 2005. Sample: Women currently in union.  
 Left panel: Share of interethnic marriages. Right panel: Ratio of observed share to random share of interethnic marriages.  
 Higher ratios mean that the share of interethnic marriages is closer to what would be observed under random matching.  
 The corresponding data can be found in the online Appendix.

The maps in figure 4 show the observed share of interethnic marriages and the ratio of the observed to random share of interethnic marriages. Striking differences between countries appear.

In Congo-Brazzaville and in Zambia, more than 40% of married women are in an interethnic marriage, whereas this share is lower than 10% in DRC, Kenya, Namibia, and Nigeria. The observed share and the ratio of observed to random shares are similar. This is because countries have high random shares of interethnic marriages: there are only two countries where this random share is lower than 75%.

Figure 5: Shares of interfaith marriages



Data: Survey wave (DHS) conducted the closest to 2005. Sample: Women currently in union.  
 Left panel: Share of interfaith marriages. Right panel: Ratio of observed share to random share of interfaith marriages.  
 Higher ratios mean that the share of interfaith marriages is closer to what would be observed under random matching.  
 The corresponding data can be found in the online Appendix.

The maps in figure 5 show the observed share of interfaith marriages and the ratio of the observed to random share of interfaith marriages. In stark contrast to interethnic marriage patterns, the share of interfaith marriages is low. The highest share of interfaith marriages is 29.6% (Congo-Brazzaville), while the highest share of interethnic marriages is over 40%. However, the level of religious fractionalization is much lower than the level of ethnic fractionalization, hence the much darker shades of the map on the right panel. Countries are also more heterogeneous with respect religious fractionalization, which ranges from 3.6% (Niger) to 64.1% (Benin).

On the pooled sample, ratios of observed to random shares are similar for interethnic and interfaith marriages shares, but it is not the case when looking at countries separately. Notably, the

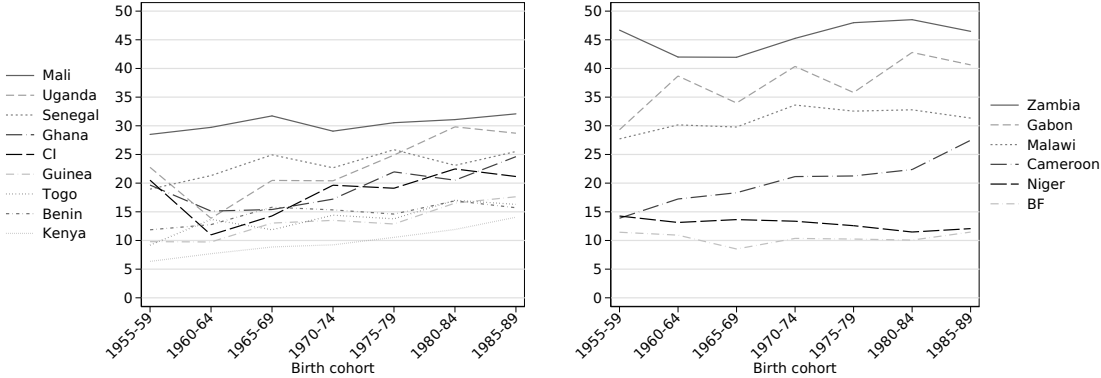
distribution of this ratio is wider when looking at interfaith marriages rather than at interethnic marriages: there is no country for which this ratio is higher than 60% when looking at interethnic marriages, but it is higher than 60% for interfaith marriages in Congo-Brazzaville, Gabon, Namibia, Niger, and Zambia.

## 6.2 Time trends on interethnic marriages

### 6.2.1 Extensive margin

#### Time trends

Figure 6: Observed interethnic marriage shares over birth cohorts



*Sample & data:* Women currently in union, weighted DHS data at country level.  
*Panel A (left):* Countries for which the trend on interethnic marriages is significantly different from 0.  
*Panel B (right):* Countries for which the trend on interethnic marriages is not significantly different from 0.  
 Countries are sorted into these two panels according to regression results from Table 3. Countries appear in the legend in descending order with respect to the share of interethnic marriages in the 1985-1889 cohort.  
 BF: Burkina Faso; CI: Cote d'Ivoire.

Figure 6 presents a visual representation of the changes in the share of interethnic marriages over time. Panel A shows that trajectories of countries where interethnic marriages became more frequent look similar. When looking at panel B, we notice that out of six countries where interethnic marriage shares did not increase, three – Zambia, Gabon, Malawi – already had shares of interethnic marriages higher than 25%. Two countries, Burkina Faso and Niger, are the only countries in the sample where the ethnic fractionalization index is lower than 70%. They both have huge majority groups – the Mossi in Burkina Faso, the Hausa in Niger – which may mean that the context in which unions take place in these two countries is different from what happens in



countries where there is no majority group in the demographic sense. The exception is Cameroon: it has a positive but not significant increase in interethnic marriage shares<sup>19</sup>, while having the same share of interethnic marriages as the average on the pooled sample, and having no majority group.

Turning to regression analysis, Table 3 lists the coefficient associated with birth year for two sets of regressions, without and with the following controls: education, urban place of residence, and remarriage. The share of interethnic marriages significantly increased over time in Benin, Cote d'Ivoire, Ghana, Guinea, Kenya, Mali, Senegal, Togo, and Uganda (Panel A). In terms of magnitudes, estimates of a 0.3 increase in percentage points by birth year translate into an increase of 7.5 percentage points when extrapolating over 25 years. Once I control for the individual characteristics correlated with interethnic marriages, the trends remain positive and significant in six countries out of nine: Benin, Cote d'Ivoire, Guinea, Mali, Senegal, and Uganda. In Ghana, the coefficient turns insignificant when introducing either education or urban residence to the model. In Kenya, the coefficient drops when introducing the type of place of residence to the model but only loses significance when all of the variables are introduced jointly. In the case of Togo, adding education levels to the model explains away the trend. Looking at Panel B, the introduction of control variables does not change the results.

### **Individual characteristics**

Countries are strikingly similar with respect to correlates of interethnic marriages, and there are no differences between countries where interethnic marriage shares increased and countries where they did not. The results are consistent with what is found in the pooled sample: primary education, secondary education, urban residence, and remarriage are all positively correlated to the likelihood of being in an interethnic union. The two exceptions are Uganda, where women who attended primary school are less likely than their uneducated counterparts to marry outside of

---

<sup>19</sup>While there seems to be a trend for Cameroon on Figure 6, the trend is insignificant when age controls are added.

Table 3: Trend - Observed interethnic marriage shares

Dependent variable	(1)	(2)	(3)	(4)
	Interethnic marriage		Mean	N
<b>Birth year coefficient</b>	<i>Each cell: coefficient from a separate regression</i>			
<i>Panel A: Increase in interethnic marriage shares</i>				
Benin	0.242*** (0.0741)	0.176** (0.0727)	15.1	10977
Cote d'Ivoire	0.254* (0.136)	0.262* (0.134)	19.2	2677
Ghana	0.262* (0.140)	0.192 (0.140)	19.4	6487
Guinea	0.728*** (0.200)	0.694*** (0.203)	14.0	4732
Kenya	0.257*** (0.0624)	0.0820 (0.0538)	10.5	9169
Mali	0.253* (0.130)	0.305** (0.129)	30.6	8499
Senegal	0.326*** (0.0914)	0.177* (0.0913)	23.5	8339
Togo	0.338*** (0.113)	0.164 (0.114)	14.4	3701
Uganda	0.455*** (0.135)	0.413*** (0.123)	24.3	2465
<i>Panel B: No change in interethnic marriage shares</i>				
Burkina Faso	0.00631 (0.0856)	-0.0680 (0.0849)	10.4	9170
Cameroon	0.545 (0.332)	0.215 (0.317)	20.5	3066
Gabon	0.364 (0.289)	0.410 (0.278)	38.0	2274
Malawi	0.00730 (0.121)	-0.156 (0.120)	31.8	9241
Niger	-0.154 (0.120)	-0.114 (0.123)	12.7	5603
Zambia	0.0675 (0.132)	0.0821 (0.125)	46.0	10711
<b>Controls</b>				
Age & Age <sup>2</sup>	✓	✓		
Education		✓		
Urban		✓		
Remarried		✓		

*Sample & data:* Women currently in union, weighted DHS data at country level. *Specification:* OLS regressions run separately for the 15 countries of the sample. Standard errors are clustered at the DHS-cluster level. The dependent variable is a variable that equals 0 if the union is intraethnic, 100 if the union is interethnic.

Columns (1) and (2) report the coefficient associated to the birth year variable. Each cell corresponds to a separate regression. Column (3) reports the mean number of interethnic marriages in the regression sample. Column (4) reports the number of observations for each country.

Results in columns (1) and (2) can be interpreted as changes in percentage points.

Significance levels are denoted as follows: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

their ethnic group – when urban residence is controlled for –, and Gabon, where urban residence is uncorrelated to interethnic marriage. The share of Gabonese women living in an urban areas is above 80%, while the share of women living in urban areas is lower than 45% in all of the other

countries: urbanization might stop being a mixing factor once urbanization levels are high.

### 6.2.2 Intensive margin: Linguistic distance

#### Descriptive statistics and time trends

Table 4 shows results of the regression of the linguistic distance (conditional on being in an interethnic union) on birth year, and on both sets of controls. Changes at the extensive margin do not necessarily correspond to changes at the intensive margin. The linguistic distance of interethnic marriages increased in three countries, two countries where interethnic marriages became more frequent – Benin and Togo – and one – Cameroon – where interethnic marriages did not increase. The linguistic distance decreased in Cote d’Ivoire, Kenya, and Senegal, all countries where the share of interethnic marriage increased. The linguistic distance did not change in the nine other countries of the sample. Introducing individual characteristics in the model changes the results only in Uganda, where the trend turns negative. The ratio of the observed to random linguistic distance (column (4)) is close to one for all countries: conditional to being in an interethnic union, most of the linguistic distance is realized. Moreover, in seven countries of the sample, this ratio is larger than one, indicating that interethnic marriages are more distant than they would be if they were formed at random (considering only intermarried people). There is only one country, Senegal, where this ratio is lower than one and where the linguistic distance of interethnic marriages decreased.

The evolution of linguistic distances depends on the type of interethnic marriages that are observed in the earlier-born cohorts. For instance, comparing the cases of Benin and Kenya (linguistic trees for these two countries are depicted in Appendix B), the average linguistic distance of interethnic marriages decreased in Kenya and increased in Benin. In Benin, as interethnic marriages became more frequent, all groups started intermarrying more, thus resulting in a decrease of the share of Adja-Fon unions, whose distance is one node, and an increase in the share of unions with a distance larger than four nodes (e.g. Yoruba-Peulh unions). Increasing linguistic distances

Table 4: Trend - Linguistic distance between spouses

Dependent variable	(1)	(2)	(3)	(4)	(5)
	Linguistic distance		Mean	Ratio	N
<b>Birth year coefficient</b> <i>Each cell: coefficient from a separate regression</i>					
<i>Panel A: Increase in interethnic marriage shares</i>					
<i>Increase in linguistic distance</i>					
Benin	0.0151** (0.00692)	0.0145** (0.00686)	3.4	0.8	1650
Togo	0.0155** (0.00746)	0.0148* (0.00790)	3.4	1.5	459
<i>No change in linguistic distance</i>					
Ghana	0.0124 (0.00841)	0.0102 (0.00881)	3.2	1.0	1238
Guinea	0.0230 (0.0199)	0.0151 (0.0209)	5.9	1.4	673
Mali	-0.00813 (0.00819)	-0.0105 (0.00812)	6.6	1.3	2568
Uganda	-0.0170 (0.0117)	-0.0209* (0.0126)	1.7	0.7	605
<i>Decrease in linguistic distance</i>					
Cote d'Ivoire	-0.0197** (0.00941)	-0.0196** (0.00876)	4.0	1.0	515
Kenya	-0.0577*** (0.0205)	-0.0460** (0.0199)	3.9	1.2	922
Senegal	-0.0277** (0.0126)	-0.0283** (0.0123)	3.5	0.8	1813
<i>Panel B: No change in interethnic marriage shares</i>					
<i>Increase in linguistic distance</i>					
Cameroon	0.0779*** (0.0225)	0.0566*** (0.0198)	3.0	0.7	653
<i>No change in linguistic distance</i>					
Burkina Faso	-0.0169 (0.0111)	-0.0125 (0.0112)	4.9	1.2	995
Gabon	0.00515 (0.00538)	0.00530 (0.00542)	1.4	0.6	923
Malawi	0.00144 (0.00181)	0.000873 (0.00182)	1.8	1.1	3023
Niger	0.0108 (0.0102)	0.0120 (0.0102)	4.4	1.3	794
Zambia	-0.00172 (0.00269)	-0.00172 (0.00261)	1.6	0.9	4873
<b>Controls</b>					
Age & Age <sup>2</sup>	✓	✓			
Education		✓			
Urban		✓			
Remarried		✓			

*Sample & data:* Women currently in an *interethnic* union, weighted DHS data at country level. *Specification:* OLS regression run separately for the 15 countries of the sample. Standard errors are clustered at the DHS-cluster level. Dependent variable is the linguistic distance (measure defined in Appendix B) associated to each interethnic union.

Columns (1) and (2) report the coefficient associated to the birth year variable. Each cell corresponds to a separate regression. Column (3) reports the mean linguistic distance for intermarried couples. Column (4) reports the ratio of the mean observed linguistic distance to the random linguistic distance, computed by randomly matching individuals who married outside of their ethnic group. Column (5) reports the number of observations for each country.

Significance levels are denoted as follows: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

indicate that women in later-born cohorts marry further away from their group, and hence that some ethnic cleavages may have lost salience. In Kenya, the decrease in linguistic distance stems mostly from the fact that in earlier-born cohorts there are more interethnic couples in which at least one spouse is Kalenjin or Luo – the only two groups that belong to the Nilo-Saharan branch – than in later-born cohorts<sup>20</sup>. As interethnic marriages became more common, the share of such unions among intermarried people decreased, resulting in the decrease of the average linguistic distance of interethnic unions. This result is still consistent with the fact that some ethnic barriers – not captured by linguistic distances – are becoming less salient.

### **Individual characteristics**

Correlates of the linguistic distance of interethnic marriages are not the same as correlates of interethnic marriages: primary and secondary education are negatively correlated to the linguistic distance of marriage in most countries. Such a reversal between the extensive and the intensive margin can be explained by ‘over-selection’ of individuals. Higher education levels are correlated with a higher likelihood of marrying outside of one’s ethnic group, so individuals who marry outside of their ethnic group and who have also not attended school are likely to have unobserved characteristics, such as being strong-willed, that also make them marry further away from their group or marry without any consideration of group differences.

## **6.3 Time trends on interfaith and Muslim-Christian marriages**

### **Time trends**

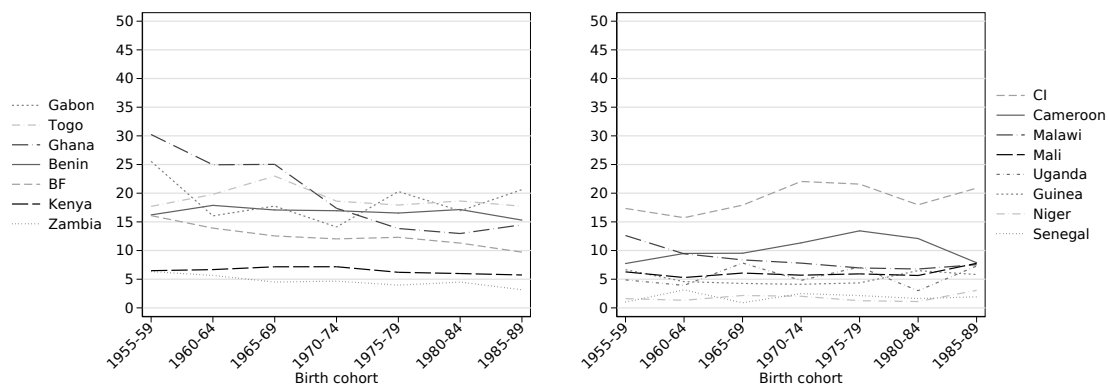
Figure 7 presents a visual representation of the change in interfaith marriage shares over time.

Comparing the two panels, it appears that countries seem to converge towards low levels of in-

---

<sup>20</sup>Linguistic distance does not capture perfectly the cleavages between groups and, contrary to share of intermarriages, is sensitive to extreme values. For instance, in Kenya, the distance between, on the one hand, the Luo and Kalenjin groups (Nilo-Saharan branch) and, on the other hand, all other ethnic groups but the Somali is high (over 7.5 nodes). Interethnic unions in Kenya have an average linguistic distance of 7.2 nodes if at least one of the spouses is Luo or Kalenjin, and a distance of 3.3 nodes otherwise. Luo-Kalenjin unions themselves make up 1.6% of interethnic unions in Kenya, despite the fact the linguistic distance of the pair is lower than with other groups, indicating that other factors than linguistic distance are also at play.

Figure 7: Observed interfaith marriage shares over birth cohorts



Sample & data: Women currently in union, weighted DHS data at country level.

Panels A1 and B1 (left): Countries for which the trend on interfaith marriages is negative and significantly different from 0.

Panels A2, A3 and B2 (right): Countries for which the trend on interfaith marriages is not significantly different from 0; Cameroon (A3) is the only country for which the share of interfaith marriages increased.

Countries are sorted into two panels according to the regression results from Table 5. Countries appear in the legend in descending order with respect to the share of interfaith marriages in the 1985-1889 cohort.

BF: Burkina Faso; CI: Cote d'Ivoire.

terfaith marriages. Apart from Cameroon and Cote d'Ivoire, all of the countries where interfaith marriages did not become less frequent are countries where the share of interfaith marriages is lower than 10% for all of the cohorts.

Table 5 shows the coefficient associated with birth year for country-specific regressions of the likelihood of being in an interfaith union on birth year and on the two sets of additional variables. Without individual controls other than age, the share of interfaith marriages increased only in Cameroon. The share of interfaith marriages decreased in Benin, Burkina Faso, Gabon, Ghana, Kenya, Togo, and Zambia. Controlling for education levels, urban place of residence, and remarriage explains the trend in Benin, Gabon, and Togo. In Benin, the trend turns insignificant when introducing an indicator variable for remarried status. In Togo, the introduction of education variables as well as of urban residence explains the trend. In Gabon, it is the joint effect of the three variables.

Muslim-Christian union shares only changed in three countries, decreasing in Ghana and increasing (although to a small extent) in Togo and Uganda. The coefficients are not significant in all of the other countries, a finding that is consistent with the fact that there is no trend when looking at the pooled sample (Table 2).

Crossing these two sets results with results on variation of the share of “other (faith)” over time (the full results can be found in the online Appendix), I find that the share of “other (faith)” decreased in all of the countries where the share of interfaith marriages decreased, with the exception of Zambia. Such share increased only in Senegal and Niger, where there was no decrease in the share of interfaith marriages. In keeping with results on the pooled sample, decreasing interfaith marriage shares are likely to be driven by the decline in the share of the group “other (faith)” and the resulting decrease in the level of religious diversity. However, at the country-level, education and urbanization suffice to explain the trend in Benin, Gabon, and Togo. In Ghana, where both interfaith marriages and Muslim-Christian marriages became less common, it is likely that social norms or preferences are moving away from tolerating interfaith marriages.

### **Individual characteristics**

Countries are heterogeneous with respect to correlates of interfaith marriages, as education levels and urban residence also capture the likelihood of being a member of “other (faith)” in most countries. Consistent with results on the pooled sample, urban residence and secondary education are negatively correlated to the likelihood of being in an interfaith union in most countries, and signs on primary schooling differ across countries. Contrasting with these results, education levels and urban residence are either insignificant or positively correlated with the likelihood of being in a Muslim-Christian marriage, thus mirroring results on individual characteristics associated with marrying outside of one’s ethnic group.

Table 5: Trend - Observed interfaith marriage shares

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Interfaith marriage Muslim/Christian/Others		Mean	Interfaith marriage Muslim-Christian		Mean	N
<b>Birth year coefficient</b>	<i>Each cell: coefficient from a separate regression</i>						
<i>Panel A: Increase in interethnic marriage shares</i>							
<i>Panel A1: Decrease in interfaith marriage shares</i>							
Benin	-0.150** (0.0754)	-0.103 (0.0753)	16.7	0.0430 (0.0332)	0.0396 (0.0341)	2.9	10977
Ghana	-1.520*** (0.133)	-1.301*** (0.131)	18.3	-0.167*** (0.0581)	-0.172*** (0.0585)	2.8	6487
Kenya	-0.189*** (0.0445)	-0.123*** (0.0451)	6.4	0.0112 (0.0174)	-0.00311 (0.0173)	1.0	9169
Togo	-0.214** (0.109)	-0.0945 (0.112)	18.9	0.0585** (0.0248)	0.0512* (0.0284)	1.2	3701
<i>Panel A2: No change in interfaith marriage shares</i>							
Cote d'Ivoire	0.122 (0.134)	0.148 (0.134)	19.3	0.0544 (0.0472)	0.0616 (0.0474)	2.6	2677
Guinea	0.0423 (0.122)	0.0481 (0.122)	5.1	0.0167 (0.0268)	0.00898 (0.0271)	0.9	4732
Mali	0.00585 (0.0660)	-0.000175 (0.0659)	6.2	0.0198 (0.0257)	0.0201 (0.0252)	1.1	8499
Senegal	-0.0399 (0.0594)	-0.0469 (0.0585)	1.9	-0.0222 (0.0516)	-0.0281 (0.0500)	1.3	7777
Uganda	0.0496 (0.0570)	0.0140 (0.0603)	5.7	0.0948* (0.0501)	0.0572 (0.0507)	4.3	2465
<i>Panel A3: Increase in interfaith marriage shares</i>							
Cameroon	0.489** (0.233)	0.523** (0.236)	10.8	0.0439 (0.0840)	0.0467 (0.0870)	1.6	3066
<i>Panel B: No change in interethnic marriage shares</i>							
<i>Panel B1: Decrease in interfaith marriage shares</i>							
Burkina Faso	-0.251*** (0.0898)	-0.237*** (0.0894)	12.1	0.00294 (0.0447)	-0.0213 (0.0451)	3.6	9170
Gabon	-0.570** (0.239)	-0.402 (0.244)	18.6	0.0729 (0.118)	0.0402 (0.128)	4.0	2274
Zambia	-0.161*** (0.0496)	-0.146*** (0.0492)	4.4	-0.0123 (0.0188)	-0.0126 (0.0191)	0.5	10711
<i>Panel B2: No change in interfaith marriage shares</i>							
Malawi	-0.123 (0.0786)	-0.0130 (0.0816)	7.8	0.0250 (0.0370)	0.0432 (0.0383)	2.2	9241
Niger	0.0912 (0.0572)	0.0962* (0.0567)	1.8	-0.00343 (0.0161)	-0.00439 (0.0163)	0.4	5603
<b>Controls</b>							
Age & Age <sup>2</sup>	✓	✓		✓	✓		
Education		✓			✓		
Urban		✓			✓		
Remarried		✓			✓		

*Sample & data:* Women currently in union, weighted DHS data at country level. *Specification:* OLS regression run separately for the 15 countries of the sample. Standard errors are clustered at the DHS-cluster level.

Columns (1) to (2): Dependent variable is a variable that equals 0 if the union is intrafaith, 100 if the union is interfaith, considering three religious groups (Christians, Muslims, Other (faiths)). Columns (1) and (2) report the coefficient associated to the birth year variable. Each cell corresponds to a separate regression. Column (3) reports the mean number of interfaith marriages.

Columns (4) to (5): Dependent variable is a variable that equals 0 if the union is a Muslim-Christian union, 100 if the union is not. Columns (4) and (5) report the coefficient associated to the birth year variable. Each cell corresponds to a separate regression. Column (6) reports the mean number of Muslim-Christian marriages. Column (7) reports the number of observations for each country.

Results in columns (1), (2), (4) and (5) can be interpreted as changes in percentage points.

Significance levels are denoted as follows: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.



## 7 Robustness analysis

I implement four robustness checks on my findings. First, I relax the assumption that a marriage is intraethnic or intrafaith when both spouses belong to the group “other”. Second, I test whether the results are robust to alternative assumptions on remarried women’s first unions and whether the trends are also found when considering separately women in their first union and remarried women. Third, using only women in their first union, I test whether “assimilation” and conversion take place over the length of a marriage. Fourth, using only women in their first union, I compare time trends measured using birth year and using cohabitation year. Table 6 displays results from the main specification and from the regressions when varying the assumptions as mentioned above. For brevity, the results are presented for the pooled sample. A full discussion of the country-level results can be found in the online Appendix.

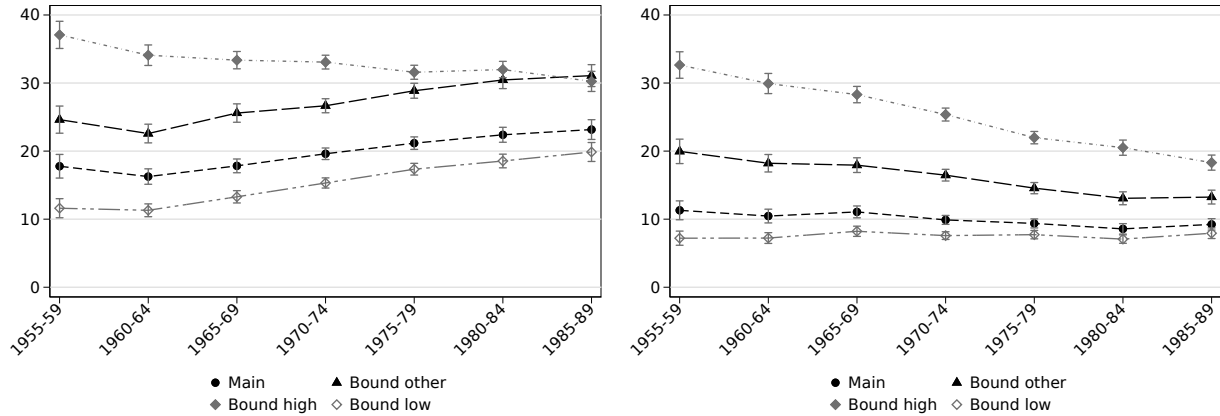
### 7.1 Testing for heterogeneity in the “other” group

The categories “other ethnicity” and “other faith” are categories that are more heterogeneous than other categories. In the main specification, I assume that when both spouses belong to the group “other”, their union is in-group. Assuming that these unions are in fact out-group unions (“*other-other*” assumption), more unions appear as intermarriages. Figure 8 shows the comparison of intermarriage shares using the main assumption and using the “*other-other*” assumption: more unions are now counted as interethnic and as interfaith, thus providing an upper bound on the share of such unions.

In table 6, columns (1) and (2) are similar to columns (3) and (4): the results regarding the pooled sample are robust to counting the “other”-“other” unions as inter-group unions. The absolute magnitude of the coefficient is higher under the “*other-other*” assumption, especially for interfaith marriages, which is consistent with the overall decrease in the share of “other (faith)”. At the country level, the main results carry through. There is no country for which the share of interethnic marriages decreased. The trend on interfaith marriages in Niger turns positive and

significant, which is consistent with the fact that the share of “other (faith)” increased over time in this country.

Figure 8: Shares of interethnic/interfaith marriages over birth cohort



Sample & data: Women in union, pooled sample. 95% confidence intervals included.

Left panel: Interethnic marriages. Right panel: Interfaith marriages.

Bound high: Shares under the assumption that all remarried women were in interethnic/interfaith first unions. Bound low: Shares under the assumption that all remarried women were in intraethnic/intrafaith first unions. Bound other: Shares under the assumption that when both spouses belong to the group “other”, they are in an interethnic/interfaith union.

## 7.2 Testing the remarriage story

Women remarry either after a divorce or after being widowed. As such, remarriage is a function of age – older women are more likely to be widows – and may also be a function of the characteristics of a woman’s first union – intermarriages may be more likely to end by a divorce<sup>21</sup>. The age controls do not capture the fact that some unions may be more likely to end than others: cohort composition effects might bias the estimated time trends. Remarried women are more likely to be in an interethnic (interfaith) marriage than women in their first marriage, but it might be that they have already been in a (first) marriage which was interethnic (interfaith). To understand better how remarriage patterns affect my results, I study bounds on the estimates and use a sub-sample analysis.

First, I bound my estimates by making assumptions on first unions of remarried women. Figure 8 depicts how the different assumptions affect intermarriage shares. The “higher bound” as-

<sup>21</sup> Whether spouses belong to the same group might affect the likelihood of divorce. However, not all remarried women have divorced, some were widowed. Interethnic unions are associated to lower age gaps between spouses, to a higher likelihood to live in an urban area, and to higher education levels: it is likely that intraethnic marriages are more likely to be ended by the husband’s death than interethnic marriages are.

*assumption* assigns an interethnic union to all of the women who have remarried. The “*lower bound*” *assumption* assigns an intraethnic union to all of the women who have remarried. Three points must be noted on these assumptions. First, they are extreme assumptions: either 100% or 0% of remarried women are assumed to have had an interethnic (interfaith) first marriage while 25.6% of them are in an interethnic marriage at survey date and 13.1% in an interfaith one. Second, the “*higher bound*” *assumption* on interethnic marriages is an extreme assumption as the ratio of the share of remarried women who have married outside of their group to the share of not-remarried women who have married outside of their group changes over birth cohorts. Third, there is a trend on remarriage. When regressing the *Remarried* variable on the year of birth, and quadratic age controls, there is a negative and significant trend: the share of remarried women is higher in earlier-born cohorts than in later-born cohorts, even when age is controlled for<sup>22</sup>. It means that the “*higher bound*” *assumption* works against finding a positive trend and that the “*lower bound*” *assumption* works in favor of finding a positive trend.

Table 6 (columns (5) to (8)) shows the coefficient on birth year under these two assumptions. Under the “*lower bound*” *assumption*, the trend on interethnic marriages remains positive and significant. However, under the “*higher bound*” *assumption*, there is no trend on interethnic (first) marriages, and a significant negative trend (of small magnitude) once education and urban residence are controlled for. At the country-level, the time trends turn negative and significant in Benin, Senegal, and Togo, resulting on the insignificant trend on the pooled sample. As the trend on interethnic marriages is barely negative and significant under the “*higher bound*” *assumption* that works in favor of finding a negative trend, it is extremely unlikely that interethnic first marriages became less frequent over time, and unlikely that their share remained constant. The trend on interfaith marriages remains negative and significant under both assumptions. Nonetheless, the magnitude of the coefficient drops under the “*lower bound*” *assumption*, as expected. The results

---

<sup>22</sup>Two factors are likely to explain this trend on the share of remarried women. First, widowhood being a less common experience due to changes in life expectancy. Second, not remarrying may be an option that is accessible to a higher share of later-born women than to their earlier-born counterparts.

on interfaith marriages at the country-level are robust to these bounds.

Second, I test whether the trends that I observe come from remarried women or from women in their first union (columns (9) to (11), and columns (13) and (14), Table 6). There are trends in all of the sub-samples. Such trends are not found for all of the countries, a result that may stem from differences between countries and sub-samples, or from the fact that the sample size is too small in the remarried sub-sample. I find no sub-sample in which the trend on interethnic marriages is negative and significant. Regarding interfaith marriages, the trend is positive and significant only for remarried Nigerian women and for Cameroonian women.

Even if we believe that the results from the “*higher bound*” assumption on interethnic first marriages are correct and hence that the share of interethnic first marriages did not change over time, the results from the sub-sample analysis show that there is a positive and significant trend when considering remarriages, even when education and urban residence are controlled for. These results are consistent with the hypothesis that women who remarry have more of a say on whom they marry, and that social norms around interethnic marriages have relaxed over time. These results may also capture changes in the composition of the remarried sub-sample: among remarried women, the share of widows is likely to be higher in earlier-born cohorts than in later-born cohorts, the share of divorcees is likely to be lower in earlier-born cohorts than in later-born cohorts. Divorced women might have different preferences from widows, and they may also be more likely to choose to whom they remarry.

### 7.3 Testing the “assimilation”/conversion story

Older women have spent more time in a union than younger women: as spouses spend a longer time in a union, their ethnic or religious identity may change. I cannot test for explanations about conversion or “assimilation” that take place before cohabitation or marriage, but I can test these two channels during the time in union. Exploiting the fact that there are at least two survey waves for each country, I can study whether women who married for the first time the same year and

Table 6: Robustness checks on interethnic marriages and interfaith marriages – Pooled sample

**Regression results** - Dependent variable: Interethnic marriage - Each cell: birth year coefficient from a separate regression

Sample	(1)	(2)	All married women				(7)	(8)	(9)	First union			Remarried	
	Main	Main	Bound	Other	Higher	bound	Lower	bound	Main	Main	Main	Main	Main	Main
<b>Dependent variable: Interethnic marriage</b>														
Birth year	0.265*** (0.0421)	0.168*** (0.0403)	0.296*** (0.0514)	0.228*** (0.0501)	-0.00227 (0.0457)	-0.0808* (0.0445)	0.268*** (0.0369)	0.157*** (0.0351)	0.263*** (0.0437)	0.135*** (0.0415)	0.212*** (0.0433)		0.353*** (0.101)	0.302*** (0.0981)
Year of cohabitation												0.174*** (0.0357)		
Number of years since cohabitation											-0.0597 (0.0458)			
Age at cohabitation											0.821*** (0.0799)			
Share interethnic marriages	20.4		27.9		32.5		16.2		19.4				25.6	
<b>Dependent variable: Interfaith marriage</b>														
Birth year	-0.179*** (0.0289)	-0.149*** (0.0287)	-0.402*** (0.0394)	-0.299*** (0.0388)	-0.354*** (0.0416)	-0.304*** (0.0413)	-0.0960*** (0.0242)	-0.0789*** (0.0241)	-0.146*** (0.0293)	-0.123*** (0.0292)	-0.146*** (0.0292)		-0.273*** (0.0776)	-0.268*** (0.0778)
Year of cohabitation												-0.0796*** (0.0252)		
Number of years since cohabitation											-0.221*** (0.0329)			
Age at cohabitation											-0.258*** (0.0545)			
Share interfaith marriages	9.7		15.5		23.8		7.6		9.1				13.1	
Country fixed-effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age & Age <sup>2</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education & Urban		✓		✓		✓		✓		✓	✓		✓	✓
Remarried		✓												

*Data & sample:* Women in union, pooled sample. Weighted data.  
*Specification:* OLS regression. Standard errors are clustered at the DHS-cluster level. As the specification includes country-fixed effects, there is no constant in the model. Dependent variable is a variable that equals 0 if the union is intraethnic (intrafaith), 100 if the union is interethnic (interfaith). Definition of the dependent variable varies in specifications (1) to (8). Results are estimated under the main specification, but on two sub-samples in columns (9) to (14).  
 Columns (1) to (10) and (13) and (14) report the coefficient associated to the birth year variable.  
 Column (11) reports the coefficient associated to the birth year variable, as well as the number of years since first cohabitation and the age at first cohabitation.  
 Column (12) reports the coefficient associated to the year of first cohabitation variable.  
 (1), (2): *Main specification, All women:* Dependent variable: Interethnic (interfaith) marriages as observed in the data.  
 (3), (4): *Other bound, All women:* Dependent variable: Interethnic (interfaith) marriages, with "other"-other unions counted as interethnic ones.  
 (5), (6): *Higher bound, All women:* Dependent variable: Interethnic (interfaith) marriages, with all women who remarried counted as being in an interethnic (intrafaith) union.  
 (7), (8): *Lower bound, All women:* Dependent variable: Interethnic (interfaith) marriages, with all women who remarried counted as being in an intraethnic (intrafaith) union.  
 (9), (10), (11), (12): *Main specification, First unions:* Only women in their first union.  
 (13), (14): *Main specification, Remarried:* Only women who have remarried.  
 Significance levels are denoted as follows: \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

were born the same year are more (less) likely to report having the same ethnic (religious) group as their husband when the length of union increases. However, the identification ultimately rests on differences across survey waves, so this also captures any effect linked to survey waves.

Table 6 (column (11)) shows results from the regression of the *Intermarriage* variables on birth year, length of union, and age at first union. Comparing column (9) to column (11), I find that the time trends are robust to controlling for length of union and age at first union.

The coefficient of the number of years since cohabitation is not significant for interethnic marriages but is negative and significant for interfaith marriages. Therefore, in the case of interethnic marriages, “assimilation” and selective divorces seem unlikely. In the case of interfaith marriages, the longer the union, the more likely it is that spouses have the same faith. This result is consistent with the fact that the share of “other (faith)” decreased over time: conversion during marriage may be one of the mechanisms that is behind this decrease. Another hypothesis would be that interfaith

unions are more likely to break than intrafaith unions, but this hypothesis does not account for the decline in traditional religions.

Concerning the age at first cohabitation, women who were older when they started cohabiting are more likely to be in an interethnic union, which is consistent with the fact that these women are more educated and more likely to live in an urban area than their counterparts, and that these characteristics are positively correlated to the likelihood of being in an interethnic union. Older women at the time of their first cohabitation are less likely to be in an interfaith union, which is consistent with the fact that they are less likely to belong to a traditional religion.

#### **7.4 Testing Birth year v. Cohabitation year**

The start of cohabitation may be a better measure of norms at the time that the union started: it corresponds to what people may perceive, such as the fact that more (less) people are getting married outside of their group. However, cohabitation year is less exogenous than birth year, and age at marriage is higher for later-born cohorts, leading these cohorts to start cohabiting at even later dates. In table 6, comparing columns (10) to (12), the results are robust to using cohabitation year instead of birth year on the sample of women who are still in their first union. At the country-level, the results are robust to using cohabitation year instead of birth year. Cohabitation year appears to indeed be endogenous to education and urban residence. When using cohabitation year rather than birth year, there is a trend in all of the countries except Burkina Faso and Niger, but after controlling for education and urban residence, this effect remains significant in the countries where there was a trend using birth year and in Gabon. After controlling for education and urban residence, the trends estimated on interfaith marriages are different depending on whether birth year or cohabitation year is used only in Cote d'Ivoire and in Senegal.

## 8 Concluding remarks

This paper documents patterns of interethnic and interfaith marriages in sub-Saharan African countries. I use data from Demographic and Health Surveys that gather information on marital history, education and geographic location to build a sample of women born between 1955 and 1989. I find that the share of interethnic marriages varies between countries but that such unions are not uncommon: 20.4% of women are married to a man who is not from the same ethnic group as them, contrasting with 9.7% of women who are married to someone who does not share their faith, and 2.1% of women in Muslim-Christian marriages.

Studying marital outcomes of women born between 1955 and 1989, I find that interethnic marriages became more common in half of the sample, and that their share remained constant in the other half. This study concludes that higher educational achievements and widespread urbanization contributed to that increase, but that these changes cannot explain all of it, suggesting that ethnic cleavages may lose salience in some parts of sub-Saharan Africa. Linguistic distance for intermarried couples increased in some countries, suggesting that all ethnic boundaries were lessened. In other countries, linguistic distance decreased: boundaries between groups that are close in terms of linguistic distance disappear and boundaries with other groups may be reinforced. As interethnic marriage shares are far from rare and are even increasing in half of the countries of the sample, using ethnicity as a proxy for a strong identification with one single group may be misguided. In contrast, interfaith marriages are becoming less common, a fact that can mostly be attributed to the conversion of former followers of traditional religions to Islam or Christianity, thus resulting in lower levels of religious diversity. The share of Muslim-Christian marriages remains low and has not changed in most countries. However, new religious groups, especially new religious branches within Christianity, are being separately counted in recent survey waves: religious boundaries within faith groups may become more salient as a result of the expansion of Christianity and Islam.

Future research could aim at studying marriages that are intrafaith but interdenominational. An additional strand of research could focus on better understanding the channels by which education and urbanization impact the likelihood of marrying outside one's own ethnic group: Are cities different from rural areas because they are more heterogeneous in terms of ethnicity? Or do people living in cities have more agency to choose a partner? Are educated women more likely to marry outside of their ethnic group because they have more agency in choosing a spouse or because they accessed more mixed markets by attending higher education institutions? Related work could investigate whether marrying outside one's own ethnic group is the result of strategic behavior (Luke and Munshi [2006]). Are urban-dwellers turning away from membership in their ethnic group to benefit from their membership in a religious group? Faith groups or networks related to attending the same church or mosque could also be an opportunity to access jobs and support. Deepening our understanding of whether marriage decisions reinforce or change identity affiliations would bring important contributions to the political economy literature on conflict, as well as to the literature on networks.



## References

- Alberto Alesina and Eliana La Ferrara. Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800, 2005. [1](#)
- Lars Ivar Oppedal Berge, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N Posner, Bertil Tungodden, and Kelly Zhang. Ethnically Biased? Experimental Evidence from Kenya. *Journal of the European Economic Association*, 2018. [3](#)
- Anais Bertrand-Dansereau and Shelley Clark. Pragmatic tradition or romantic aspiration? The causes of impulsive marriage and early divorce among women in rural Malawi. *Demographic Research*, 35:47–80, 2016. [8](#)
- Alberto Bisin and Thierry Verdier. “beyond the melting pot”: cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, 115(3):955–988, 2000. [1](#)
- Cécile Canut. Langues et filiation en Afrique. *Les Temps Modernes*, (4):410–440, 2002. [43](#)
- Kanchan Chandra. What is ethnic identity and does it matter? *Annu. Rev. Polit. Sci.*, 9:397–424, 2006. [1](#)
- Sefa Awaworyi Churchill and Russell Smyth. Ethnic diversity and poverty. *World Development*, 95: 285–302, 2017. [1](#)
- Shelley Clark, Caroline Kabiru, and Rohini Mathur. Relationship transitions among youth in urban Kenya. *Journal of Marriage and Family*, 72(1):73–88, 2010. [8](#)
- Brandon de la Cuesta and Leonard Wantchekon. Is Language Destiny? The Origins and Consequences of Ethnolinguistic Diversity in Sub-Saharan Africa. In *The Palgrave Handbook of Economics and Language*, pages 513–537. Springer, 2016. [1](#)

Klaus Desmet, Ignacio Ortuño-Ortín, and Romain Wacziarg. Linguistic cleavages and economic development. In *The Palgrave handbook of economics and language*, pages 425–446. Springer, 2016. [2](#)

William Easterly and Ross Levine. Africa’s growth tragedy: policies and ethnic divisions. *The quarterly journal of economics*, 112(4):1203–1250, 1997. [1](#)

Benn Eifert, Edward Miguel, and Daniel N Posner. Political competition and ethnic identification in Africa. *American Journal of Political Science*, 54(2):494–510, 2010. [3](#)

Patrick Francois, Ilia Rainer, and Francesco Trebbi. How is power shared in Africa? *Econometrica*, 83(2):465–503, 2015. [3](#)

Roland G Fryer Jr. Guess who’s been coming to dinner? Trends in interracial marriage over the 20th century. *Journal of Economic Perspectives*, 21(2):71–90, 2007. [3](#)

Delia Furtado and Nikolaos Theodoropoulos. Interethnic marriage: a choice between ethnic and educational similarities. *Journal of Population Economics*, 24(4):1257–1279, 2011. [3](#)

Boris Gershman and Diego Rivera. Subnational diversity in Sub-Saharan Africa: Insights from a new dataset. *Journal of Development Economics*, 133:231–263, 2018. [1](#), [2](#), [43](#)

Rachel M Gisselquist, Stefan Leiderer, and Miguel Nino-Zarazua. Ethnic heterogeneity and public goods provision in Zambia: Evidence of a subnational “diversity dividend”. *World Development*, 78:308–323, 2016. [1](#)

Erkan Gören. How ethnic diversity affects economic growth. *World Development*, 59:275–297, 2014. [1](#)

Joseph H Greenberg. The measurement of linguistic diversity. *Language*, 32(1):109–115, 1956.

- James Habyarimana, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein. Why does ethnic diversity undermine public goods provision? *American Political Science Review*, 101(4): 709–725, 2007. [3](#)
- Matthijs Kalmijn. Inter-marriage and homogamy: Causes, patterns, trends. *Annual review of sociology*, 24(1):395–421, 1998. [1](#), [7](#)
- Matthijs Kalmijn and Frank Van Tubergen. Ethnic intermarriage in the netherlands: Confirmations and refutations of accepted insights. *European Journal of Population/Revue europeenne de demographie*, 22(4):371–397, 2006. [3](#)
- Ravi Kanbur, Prem Kumar Rajaram, and Ashutosh Varshney. Ethnic diversity and ethnic strife. An interdisciplinary perspective. *World Development*, 39(2):147–158, 2011. [1](#)
- Therese Locoh and Marie-Paule Thiriat. Divorce et remariage des femmes en Afrique de l’Ouest. Le cas du Togo. *Population*, pages 61–93, 1995. [18](#)
- Sara Lowes, Nathan Nunn, James A Robinson, and Jonathan Weigel. Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT). *American Economic Review*, 105(5): 340–45, 2015. [3](#)
- Nancy Luke and Kaivan Munshi. New roles for marriage in urban Africa: Kinship networks and the labor market in Kenya. *The Review of Economics and Statistics*, 88(2):264–282, 2006. [37](#)
- Cédric Mayrargue. Trajectoires et enjeux contemporains du pentecôtisme en Afrique de l’Ouest. *Critique internationale*, (1):95–109, 2004. [5](#)
- John F McCauley. The political mobilization of ethnic and religious identities in Africa. *American Political Science Review*, 108(4):801–816, 2014. [3](#)
- Birgit Meyer. Christianity in Africa: From African independent to Pentecostal-charismatic churches. *Annu. Rev. Anthropol.*, 33:447–474, 2004. [5](#)

- William FS Miles and David A Rochefort. Nationalism versus ethnic identity in sub-Saharan Africa. *American Political Science Review*, 85(2):393–403, 1991. [3](#)
- Christiaan WS Monden and Jeroen Smits. Ethnic intermarriage in times of social change: The case of Latvia. *Demography*, 42(2):323–345, 2005. [3](#)
- Shaheen Mozaffar, James R Scarritt, and Glen Galaich. Electoral institutions, ethnopolitical cleavages, and party systems in Africa’s emerging democracies. *American political science review*, 97(3):379–390, 2003. [3](#)
- Daniel N Posner. *Institutions and ethnic politics in Africa*. Cambridge University Press, 2005. [1](#), [10](#)
- Zhenchao Qian and Daniel T Lichter. Social boundaries and marital assimilation: Interpreting trends in racial and ethnic intermarriage. *American Sociological Review*, 72(1):68–94, 2007. [3](#)
- Zhenchao Qian and Daniel T Lichter. Changing patterns of interracial marriage in a multiracial society. *Journal of Marriage and Family*, 73(5):1065–1084, 2011. [3](#)
- Gary F Simons and Charles D Fennig. Ethnologue: Languages of the world. *SIL International*, 20, 2017. [4](#), [43](#)
- Rebecca Simson. Ethnic (in)equality in the public services of Kenya and Uganda. *African Affairs*, 118(470):75–100, 2018. [3](#)
- Christoph Spörlein, Elmar Schlueter, and Frank van Tubergen. Ethnic intermarriage in longitudinal perspective: Testing structural and cultural explanations in the United States, 1880–2011. *Social science research*, 43:1–15, 2014. [8](#)

## Appendix A Data

Table 7: Data description by survey wave

Survey	Ethnicity			Religion		Remarried	Survey	Ethnicity			Religion		Remarried
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)	(11)	(12)
Benin 1996	9	11	4.4	7	35.9	20.0	Malawi 2000	9	11	7.7	8	2.6	24.6
Benin 2001	9	11	4.4	7	28.2	18.7	Malawi 2004	9	9	7.6	8	1.5	19.9
Benin 2006	9	11	4.3	11	27.5	14.5	Malawi 2010	9	14	6.8	9	1.9	21.2
Benin 2011	9	10	1.8	10	22.0	9.7	Malawi 2015	9	11	5.6	8	0.5	23.6
Benin 2017	9	10	5.1	10	16.7	14.3	Mali 1995	10	12	10.5	4	8.3	12.7
Burkina Faso 1993	10	12	7.5	5	20.1	13.2	Mali 2000	10	11	4.5	5	4.4	12.5
Burkina Faso 1998	10	19	11.2	5	18.3	13.5	Mali 2010	10	14	5.6	6	6.1	14.0
Burkina Faso 2003	10	12	7.8	5	14.0	11.3	Mali 2012	10	13	6.9	8	3.3	6.3
Burkina Faso 2010	10	17	13.2	6	9.6	10.3	Niger 1992	7	12	2.5	4	0.3	29.5
Cameroon 1998	23	43	24.1	6	11.5	19.5	Niger 1998	7	11	1.7	4	0.7	24.9
Cameroon 2004	23	48	21.3	8	10.9	19.2	Niger 2006	7	9	1.1	4	1.3	17.7
Cote d'Ivoire 1994	14	44	44.3	5	29.0	16.8	Senegal 1992 <sup>b</sup>	7	9	5.3			23.5
Cote d'Ivoire 2011	14	56	46.8	9	18.3	14.7	Senegal 2005	7	9	8.3	2	0.0	19.4
Gabon 2000	9	10	26.6	7	12.2	29.1	Senegal 2010	7	8	9.4	4	0.7	12.0
Gabon 2012	9	11	31.7	8	5.1	25.3	Senegal 2014	7	8	5.1	3	0.5	12.7
Ghana 1993	8	13	3.7	7	23.5	23.9	Senegal 2015	7	8	7.7	3	1.0	12.8
Ghana 1998	8	13	4.7	9	15.6	23.9	Senegal 2016	7	8	6.6	3	0.5	12.4
Ghana 2003	8	10	8.6	9	11.0	22.1	Senegal 2017	7	8	9.0	4	0.0	13.5
Ghana 2008	8	10	4.7	11	9.8	19.3	Togo 1998	6	7	8.6	7	48.5	22.1
Ghana 2014	8	9	3.3	10	6.4	21.1	Togo 2013	6	8	6.8	13	28.9	15.7
Guinea 1999	7	8	1.3	5	7.3	15.0	Uganda 1996	20	32	10.4	5	1.4	21.5
Guinea 2005	7	8	1.1	4	6.1	16.6	Uganda 2016	20	47	11.1	11	0.9	22.3
Guinea 2012	7	7	3.8	5	7.4	15.6	Zambia 1996	31	46	6.8	6	1.8	17.8
Kenya 1993	11	12	4.7	6	3.7	5.2	Zambia 2001	31	48	8.1	6	1.5	18.3
Kenya 1998 <sup>a</sup>	10	11	4.8	6	3.0	4.5	Zambia 2007	31	52	6.2	5	1.2	16.0
Kenya 2003	11	15	7.1	6	2.7	5.8	Zambia 2013	31	54	5.3	5	1.0	17.9
Kenya 2008	11	14	4.5	5	2.7	4.9							
Kenya 2014	11	23	6.2	5	1.6	7.2							

*Note:* This table lists all survey waves included in the main sample, and presents results useful to understand recoding choices across waves.

*Sample:* Women in union at the time of the survey.

*Ethnicity:* Columns (1) and (7): Number of ethnic groups in the common classification (at least one married woman and one married man in each survey wave and each birth cohort). Columns (2) and (8): Number of ethnic groups with at least one woman listed in the DHS classification. Columns (3) and (9): Share of women belonging to the “other (ethnicity)” group (includes foreigners) in the common classification.

*Religion:* Columns (4) and (10): Number of religious groups with at least one woman listed in the DHS classification. Common classification is made up of three groups for all waves, except Senegal 2005, where Christian and Others are pulled together. Columns (5) and (11): Share of women belonging to the “other (faith)” group in the common classification.

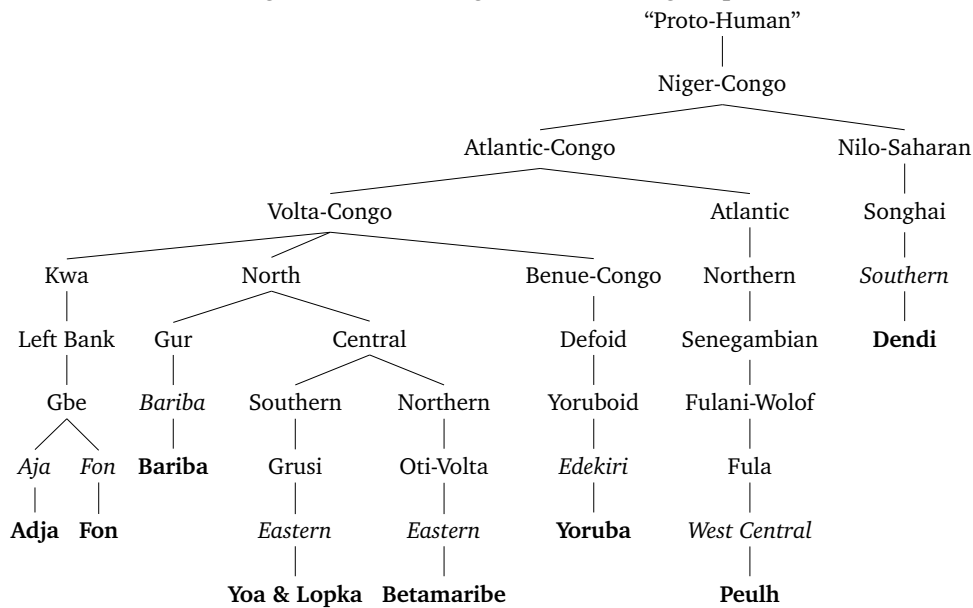
*Remarriage:* Columns (6) and (12): Share of remarried women.

<sup>a</sup> The only Somali people in this wave are two men: this wave is not representative of the north-east of Kenya.

<sup>b</sup> Religious affiliation not included in questionnaire. I use a different set of weights when not using this wave.

## Appendix B Linguistic distance measures

Figure 9: Ethnolinguistic tree for groups listed in DHS Benin



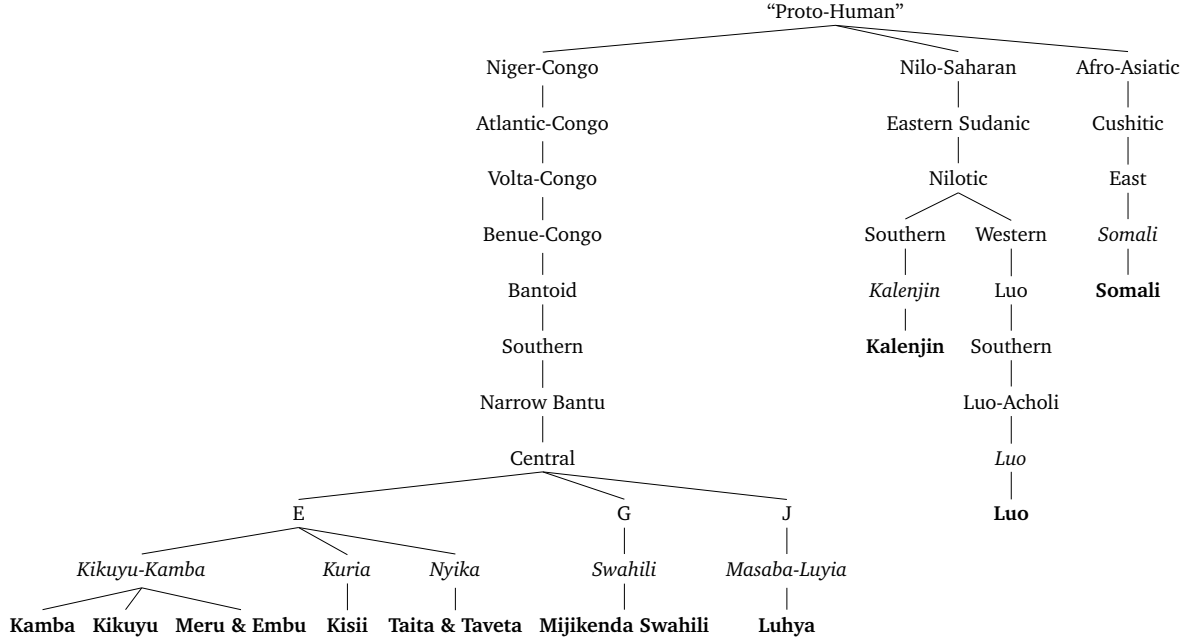
The names in bold are the names of the **ethnic groups** listed in DHS Benin. The names in italic are the names of *linguistic groups* the ethnic groups were matched to.

Gershman and Rivera [2018] describe the specificity of linguistic trees in the case of ethnolinguistic groups in sub-Saharan Africa. While languages have not always been associated with ethnicity (Canut [2002]), there is currently a strong association between languages (or language groups) and ethnic groups. First, I match each group listed the recoded DHS classification to the corresponding linguistic group according to the information listed in the *Ethnologue* dictionary (Simons and Fennig [2017]). Second, I compute the linguistic distance of each pair of linguistic groups within a country.

I define linguistic distance between two groups as the mean number of nodes to their first common subfamily. The number of nodes is computed from the linguistic group level. For instance, the linguistic distance between the ethnic groups Adja and Fon, in Benin, is 1 (one step needed to go to the last common linguistic subfamily: *Aja-Gbe* (1) and *Fon-Gbe* (1)). The linguistic distance between Kamba and Kikuyu is 0 (they share a subfamily: *Kikuyu-Kamba*). The linguistic distance between Kalenjin and Luo is 4.5 (the average of the distance *Kalenjin-Nilotic* (3) and *Luo-Nilotic* (6)).

When computing the value of the observed linguistic distance, if one spouse does not belong to any identified ethnic group, but her/his partner does, I define the linguistic distance between them as the mean linguistic distance of intermarried couples within the country. When computing the value of the random linguistic distance, I define the linguistic distance of couples in which only one spouse is “other (ethnicity)” as the median value of the defined distances at the language pair level.

Figure 10: Ethnolinguistic tree for groups listed in DHS Kenya.



The names in bold are the names of the ethnic groups listed in DHS Kenya. The names in italic are the names of linguistic groups the ethnic groups were matched to.