H i C N Households in Conflict Network
The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE
www.hicn.org

Migration, Forced Displacement and Fertility during Civil War: A Survival Analysis[1]

Philip Verwimp[2], Davide Osti & Gudrun Østby[3]

HiCN Working Paper 246

May 2017

Abstract:

The civil war in Burundi (1993-2005) caused a mass flow of refugees into neighboring countries as well as a large number of internally dis- placed persons. In fact, half of the population was displaced at least once during the course of the conflict. The aim of this study is to explore to what extent migration during the conflict impacted fertility outcomes. Using retrospective data on birth and residential histories at the mother-year level from a nationally representative survey con- ducted in August 2002, we examine the impact of war and migration on the probability of first births and on birth spacing. A parametric survival regression model is adopted to predict the hazard of having an additional child on a sample of about 4,500 Burundian women. Our results suggest that the risk of an additional pregnancy is higher in years of forced displacement of the mother, whereas it is lower in the case of residence in the forced displacement site. We do not find a statistically significant effect different from no migration in the years that the women voluntary migrated. Fertility however sharply increases once the women resided in the migration site.

Keywords: fertility, forced displacement, migration, civil war, Burundi.

JEL Classification: C25, C41, I15, J13, N37, N47.

---

[1] The bulk of the work on the paper was performed in 2016 when the second author was a graduate student at ECARES, Université Libre de Bruxelles. We would like to thank the audience at the PRIO Workshop "Armed conflict, maternal and child health, and the impact of development aid in sub-Saharn Africa" in Oslo in September 2016 as well as Eliana La Ferrara for helpful comments. Funding for this work has been provided by Research Council of Norway (grant 193754).

[2] ECARES, Centre Emile Bernheim, Université Libre de Bruxelles and Households in Conflict Network, corresponding author philip.verwimp@ulb.ac.be.

[3] International Peace Research Institute Oslo (PRIO).

Migration, Forced Displacement and Fertility during Civil War: A Survival Analysis[1]

Philip Verwimp[2], Davide Osti & Gudrun Østby[3]

May 5, 2017

Abstract

The civil war in Burundi (1993-2005) caused a mass flow of refugees into neighboring countries as well as a large number of internally displaced persons. In fact, half of the population was displaced at least once during the course of the conflict. The aim of this study is to explore to what extent migration during the conflict impacted fertility outcomes. Using retrospective data on birth and residential histories at the mother-year level from a nationally representative survey conducted in August 2002, we examine the impact of war and migration on the probability of first births and on birth spacing. A parametric survival regression model is adopted to predict the hazard of having an additional child on a sample of about 4,500 Burundian women. Our results suggest that the risk of an additional pregnancy is higher in years of forced displacement of the mother, whereas it is lower in the case of residence in the forced displacement site. We do not find a statistically significant effect different from no migration in the years that the women voluntary migrated. Fertility however sharply increases once the women resided in the migration site.

Keywords: fertility, forced displacement, migration, civil war, Burundi.

JEL codes: C25, C41, I15, J13, N37, N47.

[1]The bulk of the work on the paper was performed in 2016 when the second author was a graduate student at ECARES, Université Libre de Bruxelles. We would like to thank the audience at the PRIO Workshop "Armed conflict, maternal and child health, and the impact of development aid in sub-Saharn Africa" in Oslo in September 2016 as well as Eliana La Ferrara for helpful comments. Funding for this work has been provided by Research Council of Norway (grant 193754).

[2]ECARES, Centre Emile Bernheim, Université Libre de Bruxelles and Households in Conflict Network, corresponding author `philip.verwimp@ulb.ac.be`.

[3]International Peace Research Institute Oslo (PRIO).

1 Introduction

Armed conflict and the associated displacement of significant groups of the population may upset normality in every aspect of society, including its reproductive regime. In theory, armed conflict can both imply a reduction and increase in fertility, and there is mixed empirical evidence for both. However, it is likely that the relationship between war and fertility is complex and multidimensional, depending on factors such as variations in the *location* and *intensity of warfare*, various *types of displacement*, and the *resilience* of different segments of the population. In this paper, we analyze the impact of war and displacement on reproduction, using the case of Burundi, a country that was plagued by civil war during the years 1993-2005, and of which half the population was displaced at least once during this period (Verwimp and Van Bavel 2014).

Based on unique survey data with individual migration and fertility histories, we analyze how individual fertility outcomes, such as the probability of *first births* and *birth spacing*[4], reflect temporal and geographical variation in terms of exposure to war and associated displacement. We also consider the extent to which different socio-economic and demographic characteristics condition the fertility responses to conflict and displacement.

Our results indicate that, for first births, the effect of *forced* displacement is quite different to *voluntary* migration: in the year of moving, forced displacement increases the probability of a first pregnancy by 28%, whereas in case of voluntary migration it decreases it by 20%. Residence in the forced displacement site, on the other hand, increases the risk of a first pregnancy by 18%, whereas residence in the new migration site increases it by 69%. Being married has the usual high effect and the company of the women while being displaced also has an impact on the risk.

Turning to birth spacing, we find that the risk of an additional pregnancy is higher in a year in which the woman is forcibly displaced, whereas it is lower in the case of residence in the forced displacement site. Voluntary migration does not seem to have a statistically significant impact (compared to no displacement, of course) on the risk of an additional pregnancy. The other exogenous control variables (education, religion, proxy for assets) have the usual effect.

The remainder of the paper is structured as follows: Section 2 provides a brief literature review and a theoretical framework for studying the relationship between war, displacement and fertility, Section 3 outlines our

[4]By which we mean the time distance between subsequent births as measured in years.

micro level study, Section 4 presents the data, Section 5 the econometric model, Section 6 provides the analysis and discusses the results, and Section 7 concludes.

2 The impact of civil war and displacement on fertility

To date, relatively little systematic research has addressed how *armed conflict* and *displacement* jointly affect fertility outcomes. Caldwell (2004) reviews a body of literature demonstrating that economic shocks tend to have a negative short-term effect on fertility. As noted by Urdal and Che (2013), armed conflicts may be expected to have similar short-term effects on fertility.

In general, there are many direct mechanisms throughout which exposure to armed conflict can impact fertility in different manners. First, war may affect fertility *negatively* due to the mobilization of militia and other military reserves and the conscription of new recruits. This in turn implies both *delayed marriages* and *disruption* of marital fertility due to the separation of couples. Violent conflict can lead to an increase in the age at marriage and to an increase in the *proportion of women that never marry*. On the one hand, war may cause increased mortality among men, typically unmarried young men. The females born in the same or slightly younger birth cohorts as these deceased men may find it difficult to find a husband as the younger males usually prefer younger brides. In many developing countries, unmarried women occupy non-enviable positions in the household, often the household of a sibling. Late marriage or *single status* will decrease the fertility of these women. Analogous case is the one of women who lost their husbands in the war.

Additionally, during war or in periods of increased insecurity, it is unlikely that women marry at young age. This may be linked to the need to *provide labour* on the farm or to *generate income*. If, for example, the husband and the oldest son of the household are recruited by the army or a rebel group, the mother/wife faces the difficulty to manage the household, the farm and potential other income generating strategies all by herself. Other children and family members may need to stay on the farm to help her. The household may even attempt to recruit new household members to replace the loss of male labour (Fafchamps and Quisumbing 2006). Consequently, after the end of the war, we may see a spike in marriages.

Third, in war zones, the psychological stress and the strain of carrying out daily activities may reduce the *frequency of marital intercourse*. Furthermore, conflict-related stress can have a negative effect on both semen quality and the menstrual cycle, both of which increase the risk of infertility. Fourth, conflict may also lead to a temporary decline in the number of planned births due to the expected negative impacts of conflict on the economy. A fifth mechanism linking conflict to reduced fertility is related to the *disruption of commerce* and *food supply* that may occur during wartime. Furthermore, military presence may divert resources away from the civilian population, exacerbating existing food shortage. It is well documented that undernutrition significantly hampers female reproductive ability (Abu-Musa et al. 2008). Finally, warfare may generate *migration* and *refugee flows*, often resulting in the separation of couples for longer time periods.

A number of studies have documented significant reductions in fertility during conflict, including Agadjanian and Pratas (2002) on Angola; Blanc (2004) on Eritrea; Caldwell (2004) in general on fertility transition; and Lindstrom and Berhanu (1999) on Ethiopia. However, in some instances, the end of a conflict is associated with a *fertility increase* (e.g. Caldwell 2004).

On the other hand, long-lasting armed conflicts could also be expected to have the opposite effect on fertility behviour (Iqbal 2010; Urdal & Che 2013). Among the more proximate channels linking conflict and increased fertility through temporary migration, are shortages in access to *family planning* and *abortion services* due to the (temporal) shutdown of health clinics. Second, the demand for children may decrease as a result of the *closing of schools*, which in turn implies that the cost of children rearing declines at the same time, as their value as labour participants increases. Hence, parents may prefer short-term income from many children (extensive margin) over long-term return from fewer, educated children (intensive margin)[5].

Finally, a third mechanism linking conflict exposure to increased fertility throughout displacement is the wish (or pressure) to replace lost children and combatants. In other words, such *mortality effects* may either arise when the death of a child causes replacement of that child, or when broader expectation in society about future mortality causes hoarding. Nobles et al. (2015) refer to the former as *replacement fertility* by individual women and to the latter as *population-rebuilding* in the context of conflict or other disasters with high overall death tolls. Women who undergo violence from militia members may be subject to raping and thus increase the number of children they would otherwise have in case of no war.

[5]See also Rosenzweig and Wolpin (1980).

For example, Schindler and Brück (2011) in a study of conflict and fertility in Rwanda, found a strong *replacement effect*. Relatedly, albeit not resulting from conflict as such, Nobles et al. (2015) found that mothers who lost one or more children in the 2004 Indian Ocean tsunami were more likely to bear additional children after the tsunami (a natural disaster). Also, they found support for the so-called *population-rebuilding* mechanism, whereby women without children before the tsunami also initiated family-planning earlier after the tsunami.

As discussed above, there are several potential mechanisms linking conflict and displacement to fertility in different ways, and it is likely that the relationship is complex and multidimensional. For example, it is not unlikely that different population groups and segments of society will react to conflict in different ways. For example, better-educated and more affluent people should be both more willing and able to control their fertility behavior in response to war (Agadjanian and Prata 2002: 218). Further, Verwimp and Van Bavel (2004) in a study on fertility of refugees in Rwanda, found that refugee women had higher fertility than other women.

However, reproductive health in general, and fertility behaviours in particular, may vary a lot in refugee situations depending on the overall conditions in the camps, the length of the stay, the access to health care and so on. In a comparative study of more than 600,000 people living in 52 post-emergency phase camps in six countries (Thailand, Myanmar, Nepal, Ethiopia, Uganda, and Tanzania), Hynes et al. (2002) found better reproductive health outcomes[6] among refugees and internally displaced populations in these camps compared to the populations in their respective host country and country-of-origin. They attribute their findings to better access of camp residents to preventative and curative health care services, and to food and nonfood items, as well as improvements in water supply and sanitation.

We may not necessarily expect conflict to have the same effect on fertility for refugees in camps as compared to refugees outside camps. In general health conditions are likely to be worse for refugees that concentrate outside camps as these may not benefit from public services or international aid. Hence it is also likely that access to family planning will be higher in the camps, leading to potential lower fertility for refugees in camps than other refugees.

A couple of studies have addressed war and fertility at the national level.

[6]Lower fertility, lower neonatal mortality, lower maternal mortality, and higher birth weight.

Iqbal (2010: 82-83) finds no significant effect of major armed conflict on fertility rates. Urdal and Che (2013) find that armed conflicts are associated with higher overall fertility only in developing countries. However, it is seldom the case that an entire country is engulfed by war. Some places are harder hit than others. This calls for a disaggregated approach to the study of conflict and displacement induced fertility. Furthermore, during conflict people migrate for different reasons, voluntary or involuntary, alone or accompanied with their partners, to assigned IDP or refugee camps or outside of camps. All these factors may significantly impact fertility behavior and outcomes. We nevertheless believe that a person who is forcibly displaced is possibly likely to be subject to some violence and therefore should be more prone to higher fertility if compared to someone who deliberately decided to move.

2.1 A micro-level study

Hence, in order to fully capture the relationship between conflict, displacement and fertility, there is a need for temporally and spatially disaggregated data on conflict, combined with detailed data on individual-level migration and fertility histories. This is indeed the main contribution of our paper. In addition to the above discussion, we also address three possible *causal mechanisms* that are assumed to drive fertility among migrants.

(*i*) A *selection* effect refers to the tendency for migrants to self-select for individual characteristics that are associated with lower or higher than average fertility compared to non-migrants at the origin. Migrants indeed often differ from non-migrants on observable socio-economic characteristics such as education, age at marriage and occupation, which have an impact on reproductive choices. Selectivity may also occur on the basis of unobserved heterogeneity in preferences, such as the propensity to postpone childbearing, openness to change or fertility aspirations (on the behavioural side) and unobserved mother-specific fecundity (on the biological side). In the absence of a comparable selection effect into forced migration, one would not expect the same results for the fertility of forcefully displaced women as one would expect for voluntary migrants.

(*ii*) *Disruption* effect in childbearing through spousal separation or a desire to delay childbearing until after the move could also prevail. Such a mechanism would lower the fertility of migrants compared to non-migrants. The impact of disruption therefore, would be found in the timing of a woman's fertility and the impact may last only within a short duration. The disruption effect has been studied most often in the context of temporary migration. Sharma (1992), for example, explored the impact of *temporary spousal separation* on fertility and concluded that any relationship between migration and fertility is reflected only in cumulative fertility and that disruption was not a major factor driving temporary fertility. A high level of disruption could lead couples to make up for lost fertility by spacing births more closely after migration as well as delaying the age at which childbearing is interrupted. It is necessary, therefore to distinguish the potential effects of migration on cumulative fertility versus those on immediate fertility. White et al. (1995) found that a residential move reduced the likelihood of childbearing in the year of occurrence, providing evidence for a disruption effect. However, Goldstein et al. (1997) examined migrant fertility under very restrictive state policy regarding mobility and family planning in a Chinese province. They found, on the one hand, that rural-urban migrants tended to have later first births, which the authors attributed to the disruption, despite it could also be explained invoking a selection effect. On the other hand, they discovered that temporary migrants had a slightly higher chance of (first) birth in a year. Disruption effect may also be modified by gender and the purpose of migration (Lindstrom and Saucedo 2000). If women migrate for marriage then disruption may not be observed, but rather a short-term spike in fertility might be.

(*iii*) *Adaptation* to the fertility regimes of the destination is the third explanatory mechanism linking migration to fertility, which we postulate. The adaptation theory has its roots in both sociological and economic theories explaining determinants of fertility (Findley 1980). Rural women moving to urban areas may adapt to the prevailing norms (is it really a norm or rather better a habitude that of lower number of children per households in the city?) of having children or may find a job thereby increasing the opportunity costs of having children. This may be similar to a situation in an IDP camp where the availability of family planning services may reduce fertility.

2.2 The history of Civil War in Burundi

The latest episode of civil war in Burundi began in October 1993, when the first democratically elected president was assassinated by paratroopers from the Tutsi-dominated army in a failed coup d'état. This was followed by large-scale massacres in the countryside, with peasant supporters of the president killing Tutsi and Hutu who supported the old regime, and the army killing all Hutus in sight in an operation to 'restore order'. In a matter of days, 100,000 people lost their lives in what the UN calls a genocide

(UN 1996). The massacres were followed by the spread of violence and warfare throughout the country, with several Hutu rebel factions opposing the regular government (Tutsi) army. This marked the beginning of one ofthe most brutal and bloody civil wars in recent history (Uvin 1999).

In August 2000, several rebel groups signed the Arusha peace agreements with the still Tutsi dominated Burundian government. This had little effect on the security situation in the field since the two major rebel groups, CNDD FDD and FNL, were not involved in the peace talks. In 2003, the new president (Hutu) announced a one-sided cease fire and allowed the largest rebel group CNDD FDD to descend from the hills and march victoriously on Bujumbura. Rebel leader Nkurunziza was incorporated in the government and rebel combatants were integrated in army and police forces. The intensity of the civil war decreased dramatically and in 2005 Nkurunziza was elected as the new president. One rebel group (FNL) remained outside the peace process and continued murdering and pillaging, as a result of which pockets of insecurity still existed in the country. Human Rights Watch 1998, 2003 describes the Burundian war as a war against civilians.

Civilians were widely used as proxy targets, with both sides (rebel groups and the regular army) targeting civilians deemed supportive of the other group. Direct battles between the army and the rebel forces were relatively rare despite the duration of the war. Both sides of the conflict engaged in massive looting of civilian property and massive human rights violations. Civilians had to flee battle zones, lost wealth and livestock and were put in camps in often deplorable conditions. Displaced individuals and families were prone to attacks, deprivation, bad sanitation and housing conditions and malnutrition. In their strategy to avoid open confrontation with the army, rebel groups were very mobile and obliged villagers to supply food and to carry food and weapons over hilly areas with them. They also requested contributions in cash. Upon return home displaced people would found their land occupied by neighbours or strangers.

3 Data

Data from the *Énquete Sociale et Demographique de Santé de la Reproduction*[7] are employed for the analysis. This nationally representative survey was conducted by the United Nations Population Fund (*UNPFA*) to fill in the information gap generated between the end of the civil war and the previously collected census data in 1990, prior to the onset of the conflict.

[7]Referred hereafter as ESD (2002).

The ESD (2002) data-set is based on a *two stage stratified cluster sample survey*, designed to be representative of the population at the national level, as well as at the rural, urban and refugee camps level. The survey questionnaires were structured in an *individual bulletin*, collecting information for both men and women aged more than 14, as well as children aged less than 14, jointly with a *rugo sheet*, collecting socio-demographic information at the household level.[8]

The time-to-event panel data set used for the analysis is the result of a merge of different STATA 13 data files created from the household survey conducted at the end of August 2002. In particular, a *micro-level right-censored* data set containing fertility histories of 4,783 mothers is merged with a *dynamic panel* data set containing the migration histories of the same group of individuals. The resulting merged data set is shaped in *survival time* format, namely with two time columns allowing to study the length of yearly intervals occurring between subsequent *births* as well as between subsequent *places of residence*. To each of them is associated a dummy variable defining occurrence of partum, with the subsequent health outcomes of the child (still-births, infant survival, distinguished by sex) and a wide variety of (both *time-varying* and *time invariant*) covariates. The next two sub-sections clarify the way the survey was designed in order to guarantee a balanced representation of the various individuals across the different strata (rural, urban and refugee camps).

The ESD (2002) survey was collected on 7,119 households, of which 3,181 were located in 40 refugee camps, 2,820 in 100 rural hills, 1,118 in 28 urban locations, with a total of 32,805 persons interviewed[9]. The general information obtained from the individual bulletins for both men and women pertains to *demographic characteristics*, namely year of birth, gender, marital status, year of marriage, year of separation (if any), nationality, religion; *socio-economic features* such as schooling, educational level, occupational status, assets held (number of cows, sheep, chickens, land tenure) by the household, health status, notably survived to the conflict, or, if not, causes of death (political-military crisis, AIDS/HIV or other), localization of parents or children; *fertility aspirations*, that is, number of children desired by the persons aged more than 14 years old; *residential history*, meaning locations ever resided in, and *migratory history* since the onset of the conflict

[8] A *rugo* is a local Burundian institution, characterized as a group of households sharing the same farming activity, and a common chief, amounting to a familial structure organized in a patriarchal manner.

[9] The overall population inhabiting the country reached 6.8 millions people, at the time of the survey (ESD 2002).

in 1993, regarding simple moves; finally, *reproductive health awareness*, that is, knowledge of the risks of contracting the HIV/AIDS disease.

Table 1 describes the main variables used in the analysis. In particular, we may notice that, on average, women interviewed were born in 1969, about one third of them were forcefully displaced at least once during the nine years of the war and their mean level of education is near to the completion of primary school. Slightly more than half of the women interviewed has resided at least once in a refugee camp between 1993 and 2002 (52.02%). Additionally, by far the majority of women (86%) have been married during the years of the civil war. On average, the women in the sample have moved at least once, regardless the reason for displacement. Among those who were displaced and resided in the new site, those who did so for forceful reasons spent almost five years in the displacement site, whereas those who were displaced by their own will, spent around 4.6 years in the displacement site.

Geographically, the data is collected around each of the 168 primary sampling units, corresponding to the communities where the survey respondents lived at the time of the interview. A *household* is factually defined as group of individuals living under the same *roof* and sharing the same *budget*[10]. The average number of *persons* per household is 4.52 in the whole country, 5.08 in refugee camps, 5.01 in urban areas, and 4.45 in rural areas (ESD 2002). Sampling in rural and urban areas' has been achieved thanks to an enumeration performed by the National Institute of Statistic and Economic Studies (*ISTEEBU*) on the number of households located in rural sub-hills and in urban areas. This enumeration, which was used as a basis for the survey, was grounded on a set of papers filled in by communal officiers based on the information provided by the *hills' chiefs* and the *boroughs' chief*, themselves informed by the *nyumbakumi*[11] and the *roads' chiefs*.

In a *first degree stratification*, a randomization was performed across 30 strata, based on *urban areas*, *rural sub-collines* and *refugee camps*. In a *second degree sampling*, five regions were considered as *rural areas*, notably the North-Western region (provinces of Bubanza, Bujumbura rural and Cibitoke), a Central-Western region (with provinces Kayanza, Muramvya and Mwaro), a North-Eastern region (Kirundo, Muyinga and Ngozi), a Central-Eastern region (Canzuko, Gitega, Karuzi and Ruyigi), and a Southern region (Bururi, Makamba and Rutana). Two strata only were considered

[10]To avoid double counting due to people absent from the household.

[11]In the Great Lakes Region in Africa, the *nyumbakumi* is a local traditional institution, referring in Kishwahili to a social group composed by ten houses. Elected, he takes up the managing responsibility over a broad range of issues affecting the household's every day life such as security of humans, animals and crops.

for the *urban zones* (Bujumbura town and secondary urban agglomerates).

Interviewees in refugee camps were sampled in a two stage procedure. First, 40 camps were selected from the universe of refugee camps and second, 90 household per camp were randomly selected to be interviewed. A theoretical number of 3,600 households translated into an empirical figure of 3,181 household effectively surveyed, among those residing in refugee camps. Concerning *rural areas*, a three stage sampling was carried out. 100 randomly selected sub - collines were assigned to the 17 provinces as a function of the total number of households. In each sub-colline, 28 households were interviewed. To a theoretical number of 2,800 households, corresponded an empirical one 2,820, as it was not always possible to interview 28 households exactly. Nonetheless, whenever necessary, the bias arising from an unequal number of households per *rugo* is corrected by ponderation coefficients.

Concerning sampling in urban areas, 28 urban enumeration zones were attributed to two strata, 26 in the capital and 2 in the towns of Gitega and Ngozi, as a function of the total number of households. 40 households were randomly selected in each of the 28 zones. Weights were assigned to each observation in the survey representing the inverse of the probability of that observation being drawn in each sampled location.

We have data on number of pregnancies, number of kids alive, disaggregated by gender, number of still births, all measured per woman on an annual basis, as well as number of boys and girls deceased, mother's age at the date of birth of the child and mother's age at the time of the survey. The control variables used in the analysis are about *education*, taking values 0 for no education, 1 for at least some primary education and 2 for at least some secondary education; *religion*, where 1 means Catholic, 2 stands for Protestant, 3 for Muslim and 4 for other; an *assets' proxy* (tropical livestock unit as of 1993) which varies continuously; and *married year* where 1 means unmarried and 2 denotes a married woman.

The migration variables are mainly four. The first one, `displ_any` takes value 0 for no displacement and 1 for any type of displacement (both happened for voluntary or forced reasons, both in or out of a displacement camp). The variable `forced_volun` takes values 0 for no displacement, 1 for forced movement and 2 for voluntary movement. `moving_residing` distunguishes between no displacement, forced or voluntary movement or residence, assuming the values of 0,1,2,3,4, respectively. As last the variable `new_camp` is 0 for no displacement, 1 for forcible camp displacement, 2 for voluntary camp displacement, 3 for forcible camp residence and 4 for for voluntary camp residence.

Figure 1 displays the migration pattern within Burundi across time.

The arrows show a movement starting from the South Western part of the country directed towards North North East, avoiding the Bujumbura surrounding area. Most of the location types lying within the area of internal migration are of rural type, as suggested by the green dots in the centre of the country.

Figures 2 and 3 show the patterns of some statistics related to survival shape of our data. In particular, those figures present the Kaplan-Meier survival and failure estimates, the Nelson-Aelen cumulative hazard and the smoothed hazard estimates. For single failures (i.e. for the onset of fertility), survival probabilities are lower for urban residents than for residents in rural or refugee areas. Nonetheless, this trend seems to invert for higher values of the analysis time, with rural and refugee areas residents showing a lower probability of survival, i.e. a higher risk of first births. For multiple failures, the survival probability is always lower for inhabitants of towns than for rural citizens.

4 Econometric Methods and Estimation Strategy

Hazard rate of occurrence of first birth (*starting* behaviour) and risk of higher order births (*spacing* between subsequent births) are assumed to have an underlying proportional hazards form with spatially correlated random effects. Observations are censored, meaning that some of the mothers in the sample exit the risk set of fertility prior to the end of the observation period (the year 2002), while others still remain fertile even after the end of the survey. This has to be accounted for while formulation the likelihood function whose maximization leads to the estimated parameters of the models. In other words, some *birth intervals* observed in the survey are not *closed*, since the mother is indeed likely to eventually experience another birth after the end-line of the enumeration period[12].

The regressors, both time varying and time constant, affect the waiting time (expressed in *woman-age* metric) from zero to one birth, from one to two, and so on, and they represent the dependent variables in the various declinations of the model. Two possible metric can be chosen to perform Cox (1972) type regressions, that are proportional hazards (PH) and accelerated failure time (AFT). We chose to adopt the proportional hazard metric as is more easily adapted to interpreting results of a survival model characterized

[12]Each woman is assumed to be in the *risk set* of fecund age she is older than 14 and younger than 46. This is a somewhat stringent assumption in that does not allow for randomness in the age at menarche. See also Newman (1983).

by relatively constant or monotone hazards patterns. Therefore, we choose a multiplicative specification of the hazard of the event occurring at a given time is given the form and the covariates affecting waiting time to event:

$$\lambda(t_{ij};\mathbf{x}_{ij}) = \lambda_0(t_{ij})\exp\{\beta^T \times \begin{pmatrix} age_i \\ agesq_i \\ any\_displ_{ij} \\ religion_i \\ educ_i \\ tlu'93_i \\ married\_year_{ij} \\ parti\_part_{ij} \end{pmatrix}\} \tag{1}$$

where t_{ij} is time-to-failure event (first birth or higher order births) or censoring time for individual $i = 1, ..., 4, 783$, in village $j = 1, ..., 168$[13] and for all $t = 1967, ..., 2002$[14]; $\lambda_0(.)$ represents the baseline hazard (or *systematic* part of the hazard rate, coming from the expalantory variables), assumed to have a Weibull form, due to the flexibility peculiar of such a distribution to adapting to many possible functional forms of the "true" data generating process[15]; and the regressors' vector $\mathbf{x}_{ij}$ has a multiplicative effect on the hazard through the term $\exp\{\beta^T\mathbf{x}_{ij}\}$.

One could also consider an Aelen survival regression where the covariates enter additively in the regression model. The hazard ratio of a marginal increase in one unit in a variable $\mathbf{x}$ can be easily obtained by taking the exponential of its associated β coefficient. To equation (1), a frailty term peculiar for each community (or *unsystematic* part, since it comes from unobserved randomness within the population) is again introduced in a multiplicative way to the hazard rate[16]

[13] Also referred as secondary sampling unit (Deaton 1997). Primary sampling units are considered to be rural, urban or internally displaced refugee camp areas.

[14] At a secondary stage of the analysis, a narrower estimation window, coinciding with the decade of the civil war, will temporally restrict the sample.

[15] Recall that a continuous, positive random variable X has the Weibull distribution with *parameters* $\alpha > 0$ and $\beta > 0$ if and only if has pdf $f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp(-[x/\beta]^{\alpha})I(x > 0)$. By varying the values for α and β one can generate interesting shapes for the associated pdf.

[16] An assumption is made to allow for the distribution of the frailty terms not being independent across the sites/communities on which they are assumed to vary. This assumption is induced both by the literature (Chin et al. 2011 and by computational needs).

$$
\begin{aligned}
\lambda(t_{ij};\mathbf{x}_{ij}) &= \lambda_0(t_{ij}) \underbrace{\omega_j}_{\text{frailty}} \exp\{\beta^T\mathbf{x}_{ij}\} \\
&= \lambda_0(t_{ij}) \exp\{\beta^T\mathbf{x}_{ij} + W_j\} \qquad (2)
\end{aligned}
$$

where W_j collects the differences in the hazard of the event typical of each stratum as in Chin et al. (2011), and it is let varying freely across clusters[17]. Including W_j permits the estimation of a different hazard rate (or derived quantities, such as cumulative survivor function) for each cluster in the sample, therefore controlling for *spatial autocorrelation*, that is controlling for risk factors determining the outcome of fertility which may be community specific.

The second peculiarity of the model is the assumption of non-independence of the frailty term W_j's across strata, contrarily to the usual frailty model. In fact, frailties are hypothesized to be autocorrelated across clusters. This implies that the PSUs in proximity to each others are characterized by a more similar fertility hazard than those farther away. As in Chin et al. (2011), the vector $\mathbf{W} \in \mathbf{R}^J$ is assumed to be distributed as

$$\mathbf{W}|(\sigma^2, \psi) \sim N(\mathbf{0}, H(\sigma^2, \psi)),$$

a multivariate inverse Normal distribution with mean zero and exponential variance covariance matrix, where

$$H(\sigma^2, \psi) = \sigma^2 \exp(-\sigma d_{ii'}), \sigma^2 > 0, \psi > 0$$

and $d_{ii'}$ is the distance between PSUs i and i'. It turns out that the correlation between the two primary sampling units depends on their distance, and it reduces with the increase in distance between them.

Firstly, a parametric Weibull hazards model to explain first births (*starting*) is estimated in STATA 13, via partial maximum likelihood methods to account for right censoring, including both time constant and time varying regressors[18]. The former include a dummy variable for religious beliefs, one for educational attainment as of 1993 and an indicator of household asset

[17] Bhalotra and Van Soest (2008), in studying the determinants of infant and child mortality in Uttar Pradesh, allowed for a *subject* specific frailty term.

[18] For a derivation of the partial likelihood contribution, as the product of density functions ruling the length of birth intervals, see the online appendix.

holding (tropical livestock unit in 1993)[19]. The latter constraint mother's age (with a rescaled value of 0 representing 14 years old) and current marital status.

Secondly, to explain the distance between higher order births (*spacing*), a model analogous to the previously described one is estimated, but including also age of the mother at previous child birth, the length in years of the ending birth spell, the current duration of marriage in years, as well as three indicators for the survival of previous children. A first dummy variable indicates whether the previous child birth was a stillbirth or not. A second dummy indicates whether the previous child, if born alive, died in the same calendar year as the year of birth (disentangling among sexes of the kid)[20].

Lastly, the spacing model includes a cluster-specific term, ω_j for all $j = 1, ..., 168$ sites, capturing unobserved heterogeneity (biological fecundity as a function of surrounding natural resources and socio-cultural norms) across women residing in different geographical clusters. The impact of conflict is assessed by searching for the fertility response to forced displacement.

5 Results

5.1 The effect of any displacement on fertility

We start in table 5 with one binary variable indicating whether or not the woman was ever displaced, thereby not distinguishing between forced displacement and voluntary migration. The first column is concerned with the study of single failure events (*first pregnancies*) and the second column with multiple failures (*spacing*) to study starting fertility behavior. The table displays the results of running a parametric survival regression with an underlying Weibull distribution for the systematic component of the hazard

[19]Tropical livestock unit is a convenient measure of caloric intake developed by the Food and Agriculture Organization for quantifying a wide range of different livestock types and sized in a standardized manner. "Exchange ratios" are established with a number of common livestock varieties: 1 TLU = 1.0 camels, 0.7 cattle, 0.1 sheep/goats. The measure is based on basal metabolic rate: energy expenditure per unit of time, i.e. kcal/weight per day, varying as a function of a fractional power of body weight. Under resource driven grazing conditions, the average voluntary feed intake amongst species is remarkably similar, about 1.25 maintenance requirement (1.0 for maintenance, .25 for production). Source: <http://www.fao.org/Wairdocs/ILRI/x5443E/x5443e04.html>.

[20]This dummy is a proxy for neonatal mortality known to have a major role in determining birth spacing (Van Bavel 2004). It is only a proxy because, for example, a child who was born in December 1999 and died in January 2000 should also be counted as a case of infant mortality. Unfortunately, however, the month of death was not recorded in the survey.

(footnote 16). The dependent variable $\lambda(t_{ij};\mathbf{x}_{ij})$ in such survival models is the hazard rate of a first birth for each individual in the sample, allowing intermediate characterizations of intermediate levels of risk (Newman 1983). A hazard rate higher than one should be interpreted in this direction: a marginal increase in the underlying value of an explanatory variable has the effect of augmenting the risk of occurrence of the event under consideration by a given amount.

By inspecting the first column of Table 5, we notice that any form of displacement increases the risk of having a pregnancy by 28%. As for the covariates directly related to the displacement we notice the importance of the company of the women during her displacement. If she is alone, she has an increased risk of 13%, while if she is with her entire household she runs an increased risk of 25%. Importantly, the fact of being married or not in a given year has a very high effect on the hazard of having a first pregnancy. In the Weibull, the age of the mother (a time varying covariate) has a significantly negative impact on the hazard rate. This might be explained by the fact that the Weibull's hazard one is not constant $\lambda(t) = \gamma\alpha t^{\alpha-1}$, for further details, please see the appendix). As for the other covariates, being of Muslim faith doubles the probability of having a first pregnancy, while finishing primary school decreases the probability by 7%. Pre-war household wealth, proxied by the number of tropical livestock units, does not seem to have a statistically significant effect on the probability of having a first birth.

As for multiple pregnancies is concerned, we present the results in the second column of table 5. Here, any type of displacement does not seem to have an effect on having pregnancies, while being alone while displaced increased the risk by 11%. Primary education and in particular secondary education strongly diminish the risk, while being of muslin faith still has an increasing effect, all be it smaller compared to the first pregnancy. Being married increases fertility behavior after the first child. The number of children that died prior to the pregnancy increased the risk of having a subsequent child by 57%, a factor we could of course not include in the single model. These results call for a more profound analysis of the type of displacement, which we turn to now.

5.2 The effect of forced versus no displacement

During civil war as well as peace, women and men make decisions about where they will live. Such choices have to be distinguished from forced displacement, which unfortunately is frequently observed during civil war. We are lucky to have a survey which registered the two types of displacement and can hence distinguish their effect on fertility. As before, the first column does so for the first pregnancy, the second column for multiple pregnancies. As voluntary displacement may be endogenous to the desire to become pregnant, often linked to marriage in Burundi (see Verwimp and Van Bavel 2004), we first exclude all voluntary displacement from the analysis and compare the effect of civil war induced forced displacement on fertility with women who were never displaced. Results in table 3 show a 24% increase of having a first pregnancy when the women is forcibly displaced. When the forced displacement takes part with her entire household, it increases the risk with 22%, while being married has by far the largest effect. Adhering to Muslim faith has a large positive effect as before while the effect of primary education is on the margin of statistical significance.

Continuing to multiple births, we notice in the second column of table 6, no effect of being forcibly displaced as well as no effect of the company during displacement. As before, education has a negative effect, in particular secondary education. The death of children as well as Muslim faith keep the above effects. Our data allow us to distinguish between the year in which the displacement occurred (*moving*) versus the other years in which the women resided in her displaced residence (*residence*) we will separate the two on a year-by-year basis to find out if the actual movement has a different impact compared to residence.

5.3 Forced displacement: moving versus residing in the displacement site

Results in the first column of table 6 point out that the year in which the women actually moved carries a higher risk for first pregnancy compared to the year in which she resides in the displacement site (31% vs 21%) , while both are higher than no displacement. We also find her that the company matters: being forcibly displaced with one's entire household increase the probability for a first pregnancy by 21%. Other variables as before. In the case of multiple births, we present results in column 2 of table 7. It turns out that the risk of having an additional child increases by 11% in the year in which the forced displacement takes places (*moving*), whereas the risk decreases with 6% for the years in which the women resides in the forced displacement site (residence). Both effects are statistically significant at the 1% level. We do not observe an effect of the company of the women during displacement. The interpretation/discussion of the observed effect will follow after we deal with voluntary displacement.

5.4 Voluntary migration versus no displacement

Realizing that voluntary migration is a choice which may be endogenous to fertility choices, we want to compare the effect (not to be interpreted as a causal effect here, but rather as a correlation) here with no displacement, thereby excluding forced displacement. Column 1 in table 5 presents the results for the first pregnancy. We find a strong positive effect of 36% for voluntary migration. In case the migration takes place with the entire household an additional effect of 25% is observed. Education does not seem to matter, while the effect of marriage remains the strongest and Muslim faith also remains strong.

Moving to multiple births in column 2, the effect of voluntary migration disappears, except for the positive effect (9%) of having migrated alone. This shows that voluntary migration and fertility are particularly correlated for the first child. As before in the case of multiple pregnancies, the education variables retain their importance, together with the number of children who died and the fact of being married or not.

5.5 Voluntary migration: moving to and residing in the migration site versus no displacement

As for table 4, table 6 makes the distinction between the year in which the migration took place (*moving*) and the years in which the women resided in the migration site (*residence*). The result is very different from the result obtained for forced displacement: in the year that the voluntary migration takes place, the women has a 28% lower chance to become pregnant (compared to 31% more in the year of forced displacement). While residence in the migration site increase the probability of a first pregnancy by 54% (as compared to 21% for forced displacement). Clearly, only a much higher degree of planning and control over one's fertility can explain these results. While this is confirmed by the 31% increase in case of migration with the entire household, a caveat needs to be made as also migration alone increase the probability by 23%. Turning to multiple births, only primary and secondary education have substantial, statistically significant effects. The effect of *moving* and *residing* is much less outspoken, with the former statistically not significant and at the 10% only. Also, the sign of the effect is the opposite from the forced displacement case: a voluntary move reduces the probability of an additional child, while a voluntary residence increases it. Other variables as before.

5.6 Comparing all: forced displacement and voluntary migration, moving as well as residing compared to no displacement

Table 7 brings all of the above together in one table. We compare the two displacement types (forced vs voluntary) as well as the year of the movement with the years of residence in the new site, while no displacement remains our baseline. Column 1 presents the results for the first pregnancy. In line with the above, the effect of forced displacement is opposite to voluntary migration: in the year of moving, forced displacement increases the probability of a first pregnancy by 28% whereas in case of voluntary migration if decreases by 20%. Residence in the forced displacement site on the other hand increases the risk with 18% whereas residence in the new migration site increases it with 69%. Being married has the usual high effect and the company of the women while being displaced also has an impact on the risk.

Turning to multiple births, the risk of an additional pregnancy is higher in a year in which the women is forcibly displaced whereas it is lower in the case of residence in the forced displacement site. Voluntary migration does not seem to have a statistically significant impact (compared to no displacement of course) on the risk of an additional pregnancy. The other variables have the usual effect.

6 Conclusions

We studied the effect of forced displacement in Burundi on fertility outcomes for a sample of women interviewed in a nationwide survey at the end of the year 2002. The secondary data arising from the survey allowed to construct a panel of retrospective fertility histories with mother-year observations, dating back until the seventies. The panel was merged with information on historical residences for the subjects surveyed. Methods of survival analysis are employed to analyze the data and attempt to draw conclusive evidence on which causal mechanism drives the changing patterns in fertility due to civil conflict through internal migration and village level violence. An important assumption in this sense has been that of exogeneity of war shock on unobserved individual mother's characteristics. This may not be exempt from criticism. Weakening of such an assumption would undermine the causal interpretation which has been given to the model estimates.

Parametric Weibull regression models have been chosen as a suitable functional form to describe an analyze the stochastic process of subsequent births which mothers experienced. In particular it has been distinguished a *starting* fertility behavior (at which age a woman firstly chooses or happens to have a first pregnancy) from a *spacing* behavior (which is the average distance, in years, between subsequent births). An assumption regarding the size of the risk set of the right censored data set was made, namely that women enter their fertile period at a fixed age (14) and exit from it at another fixed age (46). One could argue that such an assumption is simplistic, in that it does not allow for randomness in the beginning of the menarche period, neither in its end. No framework has been here formulated nor applied to the analysis of stopping behavior. There exists theoretical models, such as Perrin and Sheps (1964) that formulate state space models of human reproduction, and let appropriate empirical specification derive from them. But this is beyond the scope of this empirical exercise.

Estimation has been carried out with proportional hazards models, by the Weibull distribution. We have not yet introduced a frailty term (to capture unobserved heterogeneity), but will do so in further work. Covariates have been used that were both assumed to be time invariant, such as educational level, religious belief, and time varying such as age of the mother and whether or not she was married in a given year. The focal effect of interest here, our treatment effect, derived from the displacement questions in the survey. Forced displacement would correspond to exposure to treatment, while absence of displacement is the control group. Voluntary displacement is also studied, not as another form of treatment (because it is most likely endogenous), but as a correlation, and for reason of comparison with forced displacement.

We also distinguish between the year in which the actual displacement took place and the years in which the woman resided in the new site. The findings should be distinguished for *single conception*, allowing to draw conclusions on starting behavior and for *multiple conceptions*, enabling the micro level empirical researcher to say something sensed on spacing between subsequent births. Voluntary displacement is correlated with a higher risk of first birth in the order of 70% while residing in the new site, but with a 20% lower risk in the year moving to the new site.

Forced displacement on the other had increased the risk in the years of movement as well as while in residence, but smaller than in the case of voluntary migration. This may suggest that the mechanism driving migrant fertility is planned family formation: the women does not become pregnant in the year of migration, which always carries other types of risk and uncertainty, but once settled in her new residence, she exhibits an increased probability of pregnancy.

On the side of multiple event study, that is, of time distancing between subsequent births of higher order (second, third, fourth child and so on) there does seem to be any statistically significant correlation with voluntary migration anymore, the women following the fertility trajectory similar to the non-displaced. The forcibly displaced women on the other hand demonstrate higher probability of an additional pregnancy in the years of movement, and lower while in residence in the displacement site. Presumably, this double observation is related to the very nature of forced displacement: it goes hand in hand with insecurity, violence, poverty and so on. This can result in an unwanted pregnancy (in the years of movement) and in a desire to reduce fertility while residing in a hostile environment.

References

Abu-Musa, Antoine, Loulou Kobeissi, Antoine B. Hannoun, and Marcia C. Inhorn (2008). "Effect of War on Fertility: A Review of the Literature", *Reproductive BioMedicine Online*, 3(2), 43-53.

Agadjanian, Victor and Ndola Prata (2002). "War, Peace and Fertility in Angola", *Demography*, 39(2), 215-231.

Akresh, Richard, Philip Verwimp, and Tim Bundervoet (2009). "Health and Civil War in Rural Burundi", *Journal of Human Resources*, 44(2), 536-563.

Bardhan, Pranab, and Christopher Udry (1999). "Development Microeconomics", *Oxford University Press*, New York.

Basu, Kaushik, and Pham Hoang Van (1998). "The Economics of Child Labour", *American Economic Review*, 88.

Bhalotra, Sonia and Arthur Van Soest (2008). "Birth Spacing, Fertility and Neonatal Mortality in India: Dynamics, Frailty and Fecundity", *Journal of Econometrics*, 173, 274-290.

Black, Benjamin O., Paul A. Bouanchaud, Jenine K. Bignall, Emma Simpson, and Manish Gupta (2014). "Reproductive Health During Conflict", *The Obstetrician and Gynaecologist*, 16, 153-160.

Blanc, Ann K. (2004). "The Role of Conflict in the Rapid Fertility Decline in Eritrea and Prospects for the Future", *Studies in Family Planning*, 35, 236-245.

Blayo, Chantal, Christophe Bergouignan, André Etchelecou, Valérie Codazzo, Marine Llopart, Nicodéme Okobo (2002). "Situation démographique et sociale du Burundi: Résultats de l'enquête socio-démographique et de santé de la reproduction ESDSR Burundi 2002", Fonds des Nations Unies pour la Population et Département de la population du Burundi.

Braun, Henry I. (1980). "Regression-Like Analysis of Birth Interval Sequences" *Demography* 17(2), 207-223.

Burda, Martin, Matthew Harding, and Jerry Hausman (2013). "A Bayesian Semi-Parametric Risk Model", *Working Paper MIT Economics Department*.

Butler, J. S., and Robert Moffit (1982). "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model", *Econometrica*, 50(3), 761-764.

Caldwell, John Curtis (2004). "Social Upheaval and Fertility Decline", *Journal of Family History*, 29(4), 382-406.

Cameron, Colin A., and Pravin K. Trivedi (2005). "Microeconometrics", *Oxford University Press*.

Chin, Brian, Livia Montana, and Xavier Basagana (2011). "Spatial Modeling of Geographic Inequalities in Infant and Child Mortality across Nepal", *Health & Place*, 17, 929-936.

Cox, David R. (1972). "Regression Models and Life-Tables", *Journal of the Royal Statistical Society, Series B*, 34(2), 187-220.

Cox, David R., and David Oakes (1984). "Analysis of Survival Data", *Monograph on Statistics and Applied Probability, 21*.

Deaton, Angus (1997). "The Analysis of Household Surveys: A Microeconometric Approach to Development Policy", 17140, *World Bank*, Johns Hopkins University, Baltimore and London.

Eryurt, Mehmet Ali, and Ismet Koç (2012). "Internal Migration and Fertility in Turkey: Kaplan-Meier Survival Analysis", *International Journal of Population Research*,

Fafchamps, Marcel, and Agnes R. Quimsbuing (2007). "Household Formation and Marriage Markets", *Handbook of Development Economics*, 4, 3187-3247.

Fang, Jianquing, and Jiancheng Jiang (2014). "Non- and Semi-Parametric Modeling in Survival Analysis", *Working paper, Universities of North Carolina and Princeton*.

Farewell, V. T., and David R. Cox (1979). "A Note on Multiple Time Scales in Life Testing", *Applied Statistics*, 28(1), 73-75.

Findley, Sally Evans (1980). "A Suggested Framework for Analysis of Urban-Rural Fertility Differentials with an Illustration of the Tanzanian Case", *Population and Environment*, 10(3), 237-261.

Goldstein, Alice, Michael White, and Sidney Goldstein (1997). "Migration, Fertility and State Policy in Hubei Province, China", *Demography*, 34(4), 481-491.

Han, Aaron, and Jerry A. Hausman (1990). "Flexible Parametric Estimation of Duration and Competing Risk Models", *Journal of Applied Econometrics*, 5(1), 1-28.

Hynes, Michelle, Mani Sheik, Hoyt G. Wilson and Paul Spiegel (2002). "Reproductive Health Indicators and Outcomes Among Refugees and Internally Displaced Persons in Post-Emergency Phase Camps", *Journal of the American Medical Association*, 288(5), 595-603.

Kalbfleisch, John D., and Ross L. Prentice (2002). "The Statistical Analysis of Failure Time Data", *Wiley-Interscience*, Hoboken, New Jersey.

Kiefer, Nicholas (1988). "Economic Duration Data and Hazard Functions", *Journal of Economic Literature*, 26(2), 646-679.

Knodel, John (1987). "Starting, Stopping, and Spacing During the Early Stages of Fertility Transition: the Experience of German Village Populations in the 18th and 19th Centuries", *Demography*, 24(2), 143-162.

Iqbal, Zaryab (2010). "War and the Health of Nations". *Stanford, CA: Stanford University Press.*

Lancaster, Tony (1990). "The Econometric Analysis of Transition Data", *Econometric Society Monograph.*

Lindskog, E. Elveborg (2016). "Effect of Violent Conflict on Women and Children", *Dissertation, Stockholm University Demography Unit.*

Lindstrom, David P., and Betemariam Berhanu (1999). "The Impact of War, Famine, and Economic Decline on Marital Fertility in Ethiopia", *Demography*, 36(2), 247-261

Lindstrom, David P., and Giorgiuli S. Saucedo (2002). "The Short and Long - Term Effects of US Migration Experience on Mexican Women's Fertility", *Social Forces*, 80(4), 1341-1368.

Mukhopadhyay, Nitis (2000). "Probability and Statistical Inference", *Marcel Dekker*, New York.

Neal, Sarah, Nicole Stone, and Roger Ingham (2016). "The Impact of Armed Conflict on Adolescent Transitions: A Systemic Review of Quantitative Research on Age of Sexual Debut, First Marriage and First Birth in Young Women under the Age of 20 Years", *BMC Public Health*, 16(255), 2-11.

Newman, John L. (1983). "Economic Analysis of Spacing of Births", *The American Economic Review, Papers and Proceedings of the Ninethy-Fifth Annual Meeting of the American Economic Association*, 73(2), 33-37.

Newman, John L., and Charles E. McCulloch (1984). "A Hazard Rate Approach to the Timing of Births", *Econometrica*, 54(4).

Nobles, Jenna, Elizabeth Frankenberg, and Duncan Thomas (2015). "The Effects of Mortality on Fertilty: Population Dynamics after a Natural Disaster", *Demography*, 52(1), 15-38.

Pebley, Anne R. (1980). "The Age at First Birth and Timing of the Second in Costa Rica and Guatemala". *Unpublished Ph.D. dissertation, Cornell University.*

Perrin Edward B., and Midel C. Sheps (1964). "Human Reproduction: A Stochastic Process", *Biometrics*, 20(1), 28-45.

Elbers, Chris, and Geert Ridder (1982). "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, 49(3), 403-409.

Rodriguez, German (2010). "Chapter 7: Survival Models", *Working Paper*, Princeton University.

Rokicki, Slawa, Livia Montana, and Gunter Fink (2014). "Impact of Migration and Fertility on Abortion: Evidence from the Welfare Study of Accra", *Demography*, 51, 2229-2254.

Rosenzweig, R. Mark, and Kenneth I. Wolpin (1980). "Testing the Quantity-Quality Fertility Model: the Use of Twins as a Natural Experiment", *Econometrica*, 48(1), 227-241.

Schindler, Kati, and Tillman Brück (2011). "The Effects of Conflict on Fertility in Rwanda", *World Bank Policy Research Working Paper* 5715.

Sharma, H. L. (1992). "A Study of Relationship between Migration and Fertility", *Demography India*, 21(1), 51-57.

Stata Corp. (2013). "Stata Survival Analysis and Epidemiological Tables Reference Manual: Release 13, Statistical Software", College Station, TX: Stata Corp.

Stock, James, and Motohiro Jogo (2005). "Testing for Weak Instruments in Linear IV Regression", *Andrews DWK Identification and Inference for Econometric Models.* New York: Cambridge University Press, 80-108.

Urdal, Henrik, and Chi Primus Che (2013). "War and Gender Inequalities in Health: The Impact of Armed Conflict on Fertility and Maternal Mortality", *International Interactions*, 39(4), 489-510.

Uvin, Peter (1999). "Ethnicity and Power in Burundi and Rwanda: Different Paths to Mass Violence", *Comparative Politics*, Vol. 31, No. 3, pp. 253-271.

Van Bavel (2004). "Diffusion Effects in the European Fertility Transition: Historical Evidence from within a Belgian Town (1846-1910)", *European Journal of Population*, 20(1), 63-85.

Vaupel, James W., Kenneth Manton, and Eric Stallard (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality", *Demography*, 16, 439-454.

Verwimp, Philip, and Jan Van Bavel (2004). "Child Survival and the Fertility of Refugees in Rwanda after the Genocide", *European Journal of Population*, 21(2-3), 271-290.

Verwimp, Philip, and Jan Van Bavel (2014). "Schooling, Violent Conflict and Gender in Burundi", *World Bank Economic Review*, 28 (2): 384-411.

WoldeMichael, Gebremariam (2010). "Declining Fertility in Eritrea Since the Mid-1990s: A Demographic Response to Military Conflict", *International Journal of conflict and violence*, 4(1), 149-168.

A Definition of some relevant variables

Which are used along the survival analysis and are related to the various forms of migration that are contemplated throughout the paper.

$$\texttt{displ\_any} = \begin{cases} 0 & \text{if no displacement;} \\ 1 & \text{if some displacement of any type} \end{cases}$$

$$\texttt{forced\_volun} = \begin{cases} 0 & \text{if no displacement;} \\ 1 & \text{if forced displacement;} \\ 2 & \text{if voluntary displacement} \end{cases}$$

$$\texttt{moving\_residing} = \begin{cases} 0 & \text{if no displacement;} \\ 1 & \text{if forced displacement;} \\ 2 & \text{if voluntary displacement;} \\ 3 & \text{if residence in the forced displacement site;} \\ 4 & \text{if residence in voluntary migration site} \end{cases}$$

$$\texttt{new\_camp} = \begin{cases} 0 & \text{if no displacement in camp;} \\ 1 & \text{if forced displacement in camp;} \\ 2 & \text{if voluntary displacement in camp;} \\ 3 & \text{if forced residence in camp;} \\ 4 & \text{if voluntary residence in camp} \end{cases}$$

$$\texttt{educ} = \begin{cases} 0 & \text{if no education at all;} \\ 1 & \text{it at least some primary education completed;} \\ 2 & \text{if at least some secondary education completed} \end{cases}$$

$$\texttt{religion} = \begin{cases} 1 & \text{if catholic;} \\ 2 & \text{it protestant;} \\ 3 & \text{if muslim;} \\ 4 & \text{if other} \end{cases}$$

Table 1: Description of the ESD Data Set.

Variable	Meaning	Mean	Std.Dev.	Min	Max
N_gross	Number of pregnancies	4.48	0.04	1	18
N_alive	Number of children alive	3.63	0.03	0	13
N_died	Number of children died	0.85	0.01	1	5
Y_birth	Women's year of birth	1969	0.11	1955	1988
Age	Women's age	32.73	0.11	14	47
Educ	level of education	0.84	0.008	0	2
TLU'93	trop. livestock unit	2.11	0.88	0	214
Religion	religious belief			1	4
Married	% of women ever married	86.9			
Displ_1	% Never displaced	20.95			
Displ_2	% Forced displacement only	35.61			
Displ_3	% Voluntary migration only	12.9			
Displ_4	% Voluntary as well as forced displ.	30.55			
Camp	% of women ever resided in a camp	52.02			
Moving_1	Num. of times women forcibly displ.	1.36	0.63	1	6
Moving_2	Num. of times women volunt. migrated	1.25	0.54	1	6
Residing_1	Num. of yrs. women resided in forc. displ.	4.97	2.74	1	9
Residing_2	Num. of yrs. women resided in vol. migr.	4.62	2.22	1	9

Note: This descriptive table is at the women level, the analysis will be done at the women-year level.

Table 2: Parametric Weibull PH model with `displ_any` as a main explanatory variable.

Variable	**Haz. Ratio**	**(Std. Err.)**	**Haz. Ratio**	**(Std. Err.)**
	Single pregnancy		Multiple pregnancies	
Age	.491***	(.025)	.710***	(.007)
Any displacement	1.280***	(.062)	1.002	(.018)
Religion (catholic baseline)				
Protestant	1.101**	(.040)	1.081***	(.016)
Muslim	2.041***	(.189)	1.139**	(.045)
Other	1.029	(.173)	1.107*	
Education				
At least some primary	.932**	(.035)	.945***	(.014)
At least some secondary	.956*	(.055)	.767***	(.030)
Pre-war wealth	.999	(.003)	.998*	(.003)
Married Year	6.274***	(.392)	2.438***	(.088)
Number of children died			1.578***	(.049)
Migratory pattern				
Forcibly moved with spouse	1.133**	(.061)	1.118***	(.034)
Forcibly resided with spouse	1.044	(.133)	1.001	(.048)
Volunt. moved by herself	1.248**	(.105)	1.007	(.020)
Volunt. resided by herself	1	omitted	1	omitted
Intercept	.0357***	(.009)	.002***	(.000)
$\ln p$	2.005	(.062)	1.713	(.030)
Obs.	40,245	-	91,170	-
No. of subj.	4,391	-	4,400	-
No. of fail.	4,315	-	19,475	-
t at risk	40,245	-	91,170	-
$\chi^2(df)$	2,345	-	1,814	-
df	12	-	13	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,126	-	3013	-
log pseudo-likelihood$_T$	-432	-	8,834	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 3: Parametric Weibull PH model with `forced_volun` as a main explanatory variable.

Variable	**Haz. Ratio**	**(Std. Err.)**	**Haz. Ratio**	**(Std. Err.)**
——	Single pregnancy		Multiple pregnancies	
Age	.506***	(.027)	.713***	(.008)
Reason for dispacement (no displ=baseline)				
Forcibly displaced	1.244***	(.067)	.992	(.020)
Voluntarily dispaced	1	(omitted)	1	(omitted)
Religion (catholic is baseline)				
Protestant	1.095**	(.042)	1.079***	(.015)
Muslim	2.138***	(.211)	1.140**	(.051)
Other	.962	(.170)	1.082*	(.065)
Education				
At least some primary	.937*	(.037)	.945***	(.015)
At least some secondary	.949*	(.066)	.746***	(.036)
Pre-war wealth	.998	(.003)	.999*	(.001)
Married Year	7.586***	(.454)	2.613***	(.100)
Number of children died			1.582***	(.051)
Migratory pattern				
Forcibly moved with spouse	.968	(.150)	.973	(.051)
Forcibly resided with spouse	.909	(.167)	1.031	(.070)
Volunt. moved by herself	1.223**	(.120)	1.019*	(.023)
Volunt. resided by herself	1	(omitted)	1	(omitted)
Intercept	.043***	(.011)	.002***	(.000)
$\ln p$	1.946	(.069)	1.699	(.032)
Obs.	37,957	-	80,427	-
No. of subj.	4,348	-	4,362	-
No. of fail.	3,588	-	16,542	-
t at risk	37,957	-	80,427	-
$\chi^2(df)$	1,793	-	1,703	-
df	12	-	13	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,099	-	2,387	-
log pseudo-likelihood$_T$	-719	-	7,335	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 4: Parametric Weibull PH model with `moving_residing` as a main explanatory variable.

Variable	Haz. Ratio	(Std. Err.)	Haz. Ratio	(Std. Err.)
		Single pregnancy	Multiple pregnancies	
Age	.506***	(.027)	.714***	(.008)
Residing versus moving				
Forcibly displaced (moved)	1.310***	(.105)	1.116***	(.033)
Residence in forced displacement site	1.216***	(.070)	.946**	(.019)
Voluntarily migrated (moved)	1	(omitted)	1	(omitted)
Residence in voluntary migration site	1	(omitted)	1	(omitted)
Religion (catholic is baseline)				
Protestant	1.096**	(.042)	1.078***	(.015)
Muslim	2.137***	(.211)	1.141**	(.051)
Other	.965	(.170)	1.082*	(.065)
Education				
At least some primary	.938*	(.036)	.946***	(.015)
At least some secondary	.950*	(.066)	.746***	(.036)
Pre-war wealth	.998	(.003)	.999*	(.001)
Married Year	7.585***	(.454)	2.612***	(.100)
Number of children died			1.571***	(.050)
Migratory pattern				
Forcibly moved with spouse	.969	(.150)	.987	(.079)
Forcibly resided with spouse	.904	(.167)	1.034	(.069)
Voluntarily moved by herself	1.209**	(.120)	1.017*	(.026)
Voluntarily resided by herself	1	(omitted)	1	(omitted)
Intercept	.043***	(.011)	.002***	(.000)
$\ln p$	1.946	(.069)	1.700	(.032)
Obs.	37,957	-	80,427	-
No. of subj.	4,348	-	4,362	-
No. of fail.	3,588	-	16,542	-
t at risk	37,957	-	80,427	-
$\chi^2(df)$	1,800	-	1,703	-
df	13	-	14	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,099	-	2,387	-
log pseudo-likelihood$_T$	-719	-	7,345	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 5: Parametric Weibull PH model with `voluntary vs no displ` as a main explanatory variable.

Variable	**Haz. Ratio**	**(Std. Err.)**	**Haz. Ratio**	**(Std. Err.)**
	Single pregnancy		Multiple pregnancies	
Age	.501***	(.026)	.519***	(.027)
Residing versus moving				
Forcibly displaced (moved)	1	(omitted)	1.291***	(.066)
Residence in forced displacement site	1.357***	(.117)	1.497**	(.102)
Religion (catholic is baseline)				
Protestant	1.130**	(.047)	1.119**	(.038)
Muslim	1.981***	(.173)	2.031***	(.182)
Other	1.156	(.174)	1.276**	(.152)
Education				
At least some primary	.939*	(.039)	.899**	(.031)
At least some secondary	.989*	(.066)	.924*	(.055)
Pre-war wealth	.997	(.003)	.999	(.003)
Married Year	7.044***	(.484)	6.265***	(.395)
Number of children died			2.930***	(.211)
Migratory pattern				
Forcibly moved with spouse	1.605*	(.094)	.963	(.065)
Forcibly resided with spouse	1.286*	(.333)	1.075	(.132)
Voluntarily moved by herself	1.250**	(.151)	1.256**	(.106)
Voluntarily resided by herself	1	(omitted)	1	(omitted)
Intercept	.034***	(.010)	.022***	(.005)
$\ln p$	1.946	(.069)	1.699	(.032)
Obs.	37,957	-	39,400	-
No. of subj.	4,348	-	4,391	-
No. of fail.	3,588	-	4.391	-
t at risk	37,957	-	39,400	-
$\chi^2(df)$	1,800	-	2,872	-
df	13	-	14	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,099	-	-2,707	-
log pseudo-likelihood$_T$	-719	-	9	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 6: Parametric Weibull PH model with `voluntary moving and residing versus no displacement` as a main explanatory variable.

Variable	Haz. Ratio	(Std. Err.)	Haz. Ratio	(Std. Err.)
		Single pregnancy	Multiple pregnancies	
Age	.500***	(.026)	.394***	(.043)
Residing versus moving				
Forcibly displaced (moved)	1	(omitted)	.686***	(.053)
Residence in forced displacement site	1	(omitted)	.662***	(.043)
Voluntarily migrated (moved)	.722	.099	.480	(.040)
Residence in voluntary migration site	1.544***	(.136)	1	(omitted)
Religion (catholic is baseline)				
Protestant	1.126**	(.049)	1.182**	(.066)
Muslim	1.985***	(.173)	2.026***	(.297)
Other	1.150*	(.173)	1.625	(.152)
Education				
At least some primary	.940*	(.039)	.913**	(.059)
At least some secondary	1.007*	(.062)	.924*	(.055)
Pre-war wealth	.997*	(.003)	1.012*	(.081)
Married Year	7.089***	(.490)	2.560***	(.200)
Number of children died			2.955***	(.236)
Migratory pattern				
Forcibly moved with spouse	1.235**	(.118)	1.085*	(.075)
Forcibly resided with spouse	1.285*	(.351)	.919*	(.132)
Voluntarily moved by herself	1.314**	(.176)	1.191**	(.104)
Voluntarily resided by herself	1	(omitted)	1	(omitted)
Intercept	.035***	(.010)	.020***	(.007)
$\ln p$	1.975	(.064)	2.338	(.097)
Obs.	37,308	-	4,812	-
No. of subj.	4,345	-	1,291	-
No. of fail.	4,345	-	1,291	-
t at risk	37,308	-	4,812	-
$\chi^2(df)$	2,349	-	731	-
df	13	-	15	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-2,861	-	104	-
log pseudo-likelihood$_T$	-410	-	680	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 7: Parametric Weibull PH model with `forced movement and residing and voluntary movement and residing vs no displacement` as main explanatory variable.

Variable	Haz. Ratio	(Std. Err.)	Haz. Ratio	(Std. Err.)
		Single pregnancy	Multiple pregnancies	
Age	.491***	(.025)	.519***	(.027)
Residing versus moving				
Forcibly displaced (moved)	1.285***	(.098)	1.289***	(.096)
Residence in forced displacement site	1.187**	(.067)	1.265***	(.073)
Voluntarily migrated (moved)	.801*	(.094)	.822*	(.096)
Residence in voluntary migration site	1.693***	(.113)	1.768***	(.117)
Religion (catholic is baseline)				
Protestant	1.099**	(.043)	1.117***	(.038)
Muslim	2.039***	(.190)	2.033***	(.182)
Other	1.033	(.170)	1.272**	(.153)
Education				
At least some primary	.933*	(.035)	.902**	(.031)
At least some secondary	.966	(.056)	.942*	(.056)
Pre-war wealth	1.000	(.003)	.999	(.003)
Married Year	6.292***	(.398)	6.302***	(.400)
Number of children died			2.910***	(.207)
Migratory pattern				
Forcibly moved with spouse	1.133*	(.080)	1.069*	(.075)
Forcibly resided with spouse	1.013	(.138)	1.066	(.137)
Voluntarily moved by herself	1.226**	(.109)	1.276**	(.114)
Voluntarily resided by herself	1	(omitted)	1	(omitted)
Intercept	.037***	(.010)	.022***	(.005)
$\ln p$	2.004	(.062)	1.965	(.065)
Obs.	40,245	-	39,400	-
No. of subj.	4,391	-	4,391	-
No. of fail.	4,315	-	4,391	-
t at risk	40,245	-	39,400	-
$\chi^2(df)$	2,570	-	2,924	-
df	15	-	16	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,127	-	-2,707	-
log pseudo-likelihood$_T$	-389	-	52	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

Table 8: Parametric Weibull PH model with `forcibly in camp, no camp and voluntarily in camp versus no displacement` as main explanatory variable.

Variable	**Haz. Ratio**	**(Std. Err.)**	**Haz. Ratio**	**(Std. Err.)**
		Single pregnancy	Multiple pregnancies	
Age	.491***	(.025)	.519***	(.027)
Camp displacement				
Forcibly displaced in camp (moved)	1.240***	(.071)	1.329***	(.074)
Not forcibly displaced in camp	1.220**	(.086)	1.238**	(.087)
Voluntarily migrated in camp (moved)	1.370***	(.107)	1.460***	(.116)
Not voluntarily migrated in camp	1.482***	(.113)	1.512***	(.109)
Religion (catholic is baseline)				
Protestant	2.032**	(.190)	1.119***	(.038)
Muslim	2.039***	(.190)	2.033***	(.182)
Other	1.034	(.171)	1.275**	(.153)
Education				
At least some primary	.931*	(.034)	.899**	(.031)
At least some secondary	.950	(.056)	.924*	(.056)
Pre-war wealth	.999	(.003)	.999	(.003)
Married Year	6.253***	(.393)	6.262***	(.395)
Number of children died			2.931***	(.210)
Migratory pattern				
Forcibly moved with spouse	1.014*	(.069)	.963*	(.065)
Forcibly resided with spouse	1.025	(.133)	1.076	(.131)
Voluntarily moved by herself	1.224**	(.104)	1.265**	(.109)
Voluntarily resided by herself	1	(omitted)	1	(omitted)
Intercept	.036***	(.09)	.022***	(.005)
$\ln p$	2.006	(.062)	1.967	(.065)
Obs.	40,245	-	39,400	-
No. of subj.	4,391	-	4,391	-
No. of fail.	4,315	-	4,391	-
t at risk	40,245	-	39,400	-
$\chi^2(df)$	2,548	-	2,924	-
df	15	-	16	-
No. of clust.	168		168	
log pseudo-likelihood$_0$	-3,127	-	-2,707	-
log pseudo-likelihood$_T$	-429	-	10	-

Note: * Significant at the 10 percent level
** significant at the 5 percent level
*** Significant at the 1 percent level

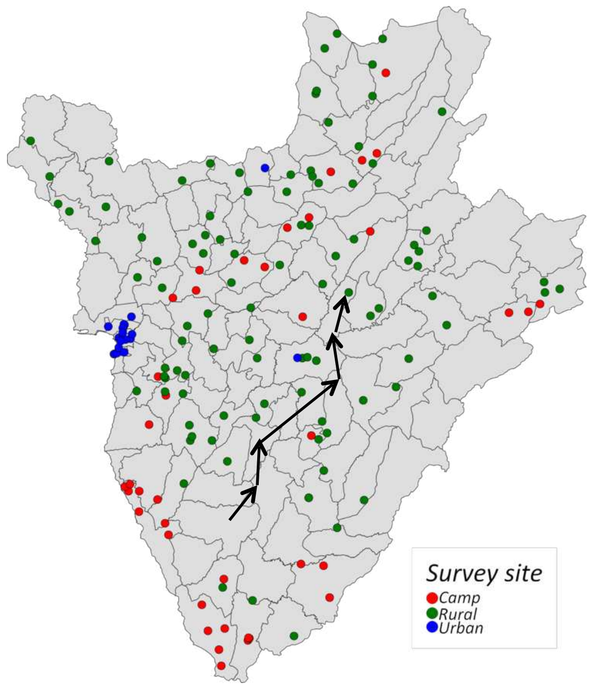

Figure 1: Migration patterns across the country during the war. Thanks to Karim Bhagat for creating the map.

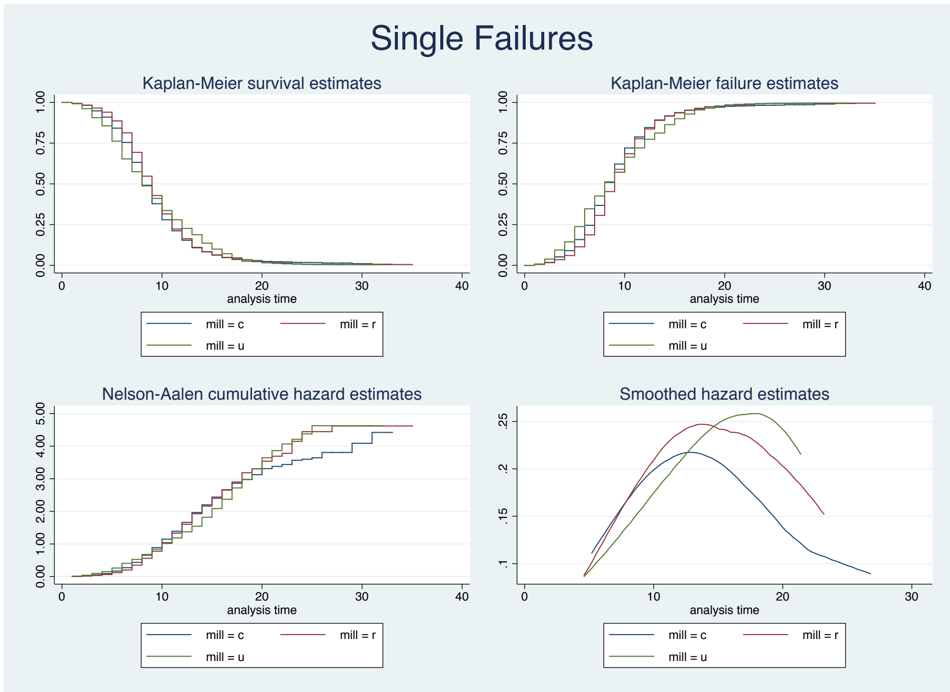

Figure 2: Survival ($S(t)$), failure ($F(t)$), cumulative hazard ($\Lambda(t)$) and hazard rate ($\lambda(t)$) plotted for single failures i.e. for *starting* fertility behaviour.

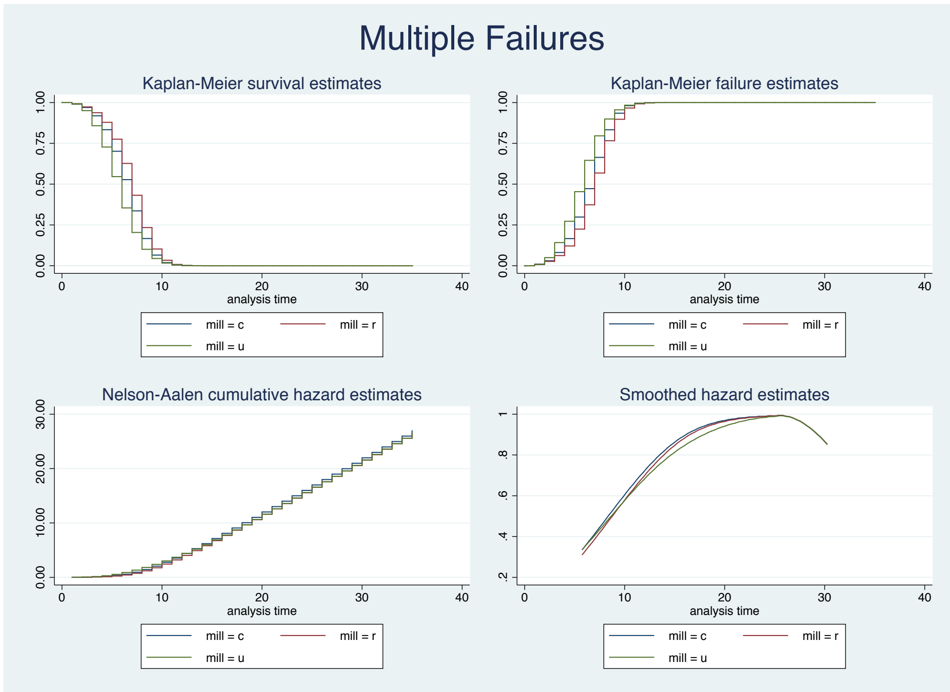

Figure 3: Survival ($S(t)$), failure ($F(t)$), cumulative hazard ($\Lambda(t)$) and hazard rate ($\lambda(t)$) plotted for multiple failures, i.e. for *spacing* fertility behaviour.

A Sampling Weights

Sampling weights were calculated to readdress the unequal probability of random assignment of households belonging to different strata, relevant in the three main spatial categories. *Global weights*, *between-zone weights* and *between-province weights* have been considerered in the survey. Global weights are defined as:

$$GW_{jhi} = n/k_{jh} \times n_{jhi} \times (N_{mh}/N)$$

where the indexing $j = u, r, c$ stands for milieu, h for stratum and $i = 1, ..., 168$ for survey site; n being the 7,119 households interviewed, k_{mh} being the # of survey sectors of interview in strate mh, n_{mhi} being the number of households surveyed in the survey site i within strate mh, N_{mh} the total number of households actualized in strate mh and N the total number of households in Burundi at the time of the survey in August 2002, that is 1,181,667. Intra province weights are:

$$IP_{mpj} = \left(\frac{n_p}{k_{mp}} \times n_{mpj}\right) \times \left(\frac{N_{mp}}{N_p}\right)$$

where n_p is the total number of household surveyed, k_{mp} is the number of survey sites, n_{pji} is the total number of household surveyed in province p, site j, N_{mp} the total number of households actualized in municipality m in province p and N_p the total number of households interviewed in province p [see ESD 2002 analyzed by researchers from Bordeaux].

B T-tests for the means of some important variables

Table 9: T-tests on the difference between the means of background variables "Never displaced" versus "Forced or Forced + Voluntary"[21]

Variable	Never displaced	Forcibly displ. or forced + volunt.	t-test
Year of birth	1968.0	1968.6	0.6**
Married before 1993	0.62	0.62	0
Wealth in 1993 (TLU)	1.84	2.36	- 0.52***
Catholic in 2002	74.6	60.1	14.5***
Level of education in 2002	0.94	0.75	-0.19***

The tests show that the background variables of year of birth, marriage status, wealth in 1993, religion and level of education are not *ex ante* different across treated and untreated subjects, i.e. among those women who never moved and those who were forcibly displaced at least once.

C Some important functions in survival analysis

In general and for continuous duration data-sets, and given a continuous random variable T representing a waiting time until a certain event occurs, the following functions are defined

$$\underbrace{\lambda(t)}_{\text{hazard function}} = \lim_{dt \to 0} Pr\{t \leq T \leq t + dt | T \geq t\}/dt^{22}$$

$$\underbrace{\Lambda(t)}_{\text{cum. haz. funct.}} = Pr\{T \leq t\} = \int_0^t \lambda(s)ds$$

$$f(t) = \frac{dF(t)}{dt}$$

$$\underbrace{S(t)}_{\text{survivor function}} = Pr\{T > t\} = \underbrace{1 - \Lambda(t)}_{\text{1-c.h.f.}} = \int_t^\infty f(x)dx$$

$$\underbrace{T_u}_{\text{avg dur. of a spell}} = \frac{1}{\lambda[1 - H(x)]}$$

The conditional probability that T is in $[t, t + dt]$ and $T \geq t$ is

$$\lambda(t) = \frac{f(t)}{S(t)}$$

where the density function $f(t)$ is the same as saying that T is in the interval, and the survivor function $S(t)$ expresses the probability of $T \geq t$. In other words, the probability of an event occurring at duration t is the density of the event at time t divided by the probability of surviving to that duration, without experiencing the event.

Furthermore,

$$-f(t) = \frac{dS(t)}{dt} = \frac{d[1 - F(t)]}{dt}$$

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$

Integrate from 0 to t and introduce $S(0) = 1$, event sure not to have occurred at time 0.

[22] **Numerator**: conditional probability that the event will occur in the interval $[t, t+dt]$ given that it has not occurred before; **denominator**: width of the interval. Ratio between the two $\equiv$ rate of occurrence of the event per unit of time. Taking the limit as the width of the interval goes to zero, we obtain the instantaneous rate of occurrence.

$$S(t) = \exp\{- \underbrace{\int_0^t \lambda(s)ds}_{\text{cum. haz. fct.}} \}$$

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

sum of risks facing when going from 0 to t. To recap

$$\begin{aligned} \text{density} &\longrightarrow f(t) \\ \text{distribution} &\longrightarrow F(t) \\ \text{survivor} &\longrightarrow S(t) \\ \text{hazards} &\longrightarrow \lambda(t) \\ \text{cum. haz.} &\longrightarrow \Lambda(t) \end{aligned}$$

Example: $\lambda(t) = \lambda$, constant over time (hazard), simplest case. $S(t) = \exp\{-\lambda t\}$, an exponential distribution with parameter $\lambda \Rightarrow f(t) = \lambda(t)S(t) = \lambda \exp\{-\lambda t\} = \lambda e^{-\lambda t} = \int \frac{\lambda}{e^{\lambda t}} dt$.

The expectation of life is simply the indefinite integral of the survival function, say μ for random variable T, that is

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t)dt,$$

integrating with $-f(t) = dS(t)/dt$, $S(0) = 1$ and $S(+\infty) = 0$.

D Derivation of the likelihood contribution[23]

To identify the proportional hazards model, a partial likelihood framework is adopted, following Cox (1972, 1979). The question is how to estimate β in the PH model without the need of a simultaneous estimation of all the $\lambda_0(t_j)$, baseline hazards functions, for all $j = 1, ..., N$. The setup is that of an ordered sequence of discrete failure times, $t_1 < t_2 < ... < t_j < ... < t_k$, in a sample of N individuals, $(N \geq k)$. The risk set, $R(t_{ij})$ is defined as the set of individuals located in cluster i who are at risk of failing just prior to the j^{th} ordered failure. Let $D(t_{ij})$ denote the set of subjects having births happening at time t_j, and d_{ij} indicating the number of mothers living in cluster i that give birth to a new life at time t_{ij}. Summing up,

[23] Based on Cameron and Trivedi 2005's, ch. 17.

$$
\begin{gathered}
R(t_j) = \{l : t_l \geq t_j\} \rightsquigarrow \textit{set of spells}^{24} \textit{ at risk at time } t_j \\
D(t_j) = \{l : t_l = t_j\} \rightsquigarrow \textit{set of spells completed at } t_j \\
d_j = \sum_l \mathbf{1}\{t_l = t_j\} \rightsquigarrow \# \textit{ of spells completed at } t_j
\end{gathered}
$$

where **1** stands for a dummy variable taking values 1 if the spell was completed before the end of the survey and 0 otherwise. Tied data are possible at $d_j > 1$. The former of the last three expressions includes spells that are neither completed nor censored. For all individuals $j = 1, ..., 4, 783$ and strata $i = u, r, c$, the probability of a particular at risk spell ending at time t_j is

$$
\begin{aligned}
Pr[T_{ij} = t_{ij} | R(t_{ij})] &= \underbrace{\frac{\overbrace{Pr[T_{ij} = t_{ij} | T_{ij} \geq t_{ij}]}^{\square}}{\displaystyle\sum_{l \in R(t_{ij})} Pr[T_l = t_l | T_l \geq t_{ij}]}}_{\blacksquare} \\
&= \frac{\lambda_j(t_j | \mathbf{x}_j, \beta)}{\sum_{l \in R(t_j)} \lambda_l(t_j | \mathbf{x}_l, \beta)} \\
&= \frac{\phi(x_j, \beta)}{\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta)}
\end{aligned}
$$

where $\square$ is conditional probability that a particular spell at risk is ending at time t_j, that is the probability of failure for individual j in cluster i in spell t_{ij} and $\blacksquare$ expresses the conditional probability that a spell of any individual in the risk set $R(t_{ij})$ fails, i.e. she has an additional child, precisely in time t_{ij}. In the third line, λ_0 has dropped out due to thePH assumption, meaning that the intercept is not identified.

Suppose there are two tied values at time t_j, for individuals j_1 and j_2 with regressors $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$. If j_1 fails[25] and $\mathbf{x}_{i2}$ then the probability of occurrence of a failure of them is:

$$
\frac{\phi(\mathbf{x}_{i1}, \beta)}{\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta)} + \frac{\phi(\mathbf{x}_{i2}, \beta)}{\sum_l R(t_j) \phi(\mathbf{x}_l, \beta)}
$$

[24]Spacing between subsequent births, i.e. this setup may be useful for *spacing* not for *starting* fertility behaviour.

[25]If a mother remains pregnant for starting or if the mother has an additional child for spacing behaviours, let's say. The present exposition is taken from the Cameron and Trivedi 2005 manual, specifically from chapter 17.8.2. Identification of the PH model.

A similar term arises if j_2 fails before j_1 and the likelihood contribution's the sum of the two possibilities. The exact likelihood becomes complicated with many tied values. Cox and Oakes 1984 $\rightsquigarrow Pr[T_j = t_j | j \in R(t_j)] \approx \frac{\prod_{m \in D(t_j)} \phi \mathbf{x}_m, \beta}{[\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta)] d_j}$, approximation working well if the # of failures is small relative to the number of individuals (mothers) at risk (not our case) in the population.

Cox 1972 proposed instead a partial likelihood function, derived as the joint product of $Pr[T_j = t_j | j \in R(t_j)]$ over the k ordered failures. Then:

$$L_p(\beta) = \prod_{j=1}^{k} \frac{\prod_{m \in D(t_j)} \phi(\mathbf{x}_m, \beta)}{\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta) d_j}$$

in which β can be estimated by minimizing the log partial likelihood function

$$L_p = \sum_{j=1}^{k} \left\{ \sum_{m \in D(t_j)} \ln \phi(\mathbf{x}_m, \beta) - d_j \ln \left(\sum_{l \in R(t_j)} \phi(\mathbf{x}_l, \beta) \right) \right\}.$$

only in the second term on the right hand side there appear censored spells, because they do not contribute to the observed births, but until they're censored, they affect the size of the risk set. $d_j = 1$ if an observation is censored and 0 otherwise. Alternatively, one could write the partial log-likelihood contribution, changing the indexation from j to i as:

$$\ln L_p(\beta) = \sum_{i=1}^{k} \delta_i [\sum_{m \in D(t_j)} \ln \phi(\mathbf{x}_m, \beta) - \ln(\sum_{l \in R(t_i)} \phi(\mathbf{x_l}, \beta))].$$

reverting the role of the indicator variable for censoring, such that

$$\delta_i = \begin{cases} 1 & \text{for censored obs.;} \\ 0 & \text{else;} \end{cases}$$

and recalling that it has been imposed along the article, that

$$\phi(\mathbf{x}, \beta) = \exp\{\mathbf{x}'\beta\} \Leftrightarrow$$
$$\Leftrightarrow \ln \phi(\mathbf{x}, \beta) = \mathbf{x}'\beta, \text{ with F.O.C.:}$$
$$\frac{\partial \ln L_p(\beta)}{\partial \beta} = \sum_{i=1}^{N} \delta_i \left[\mathbf{x}_i - \mathbf{x}_i^*(\beta) \right] = 0, \; \mathbf{x}_i^* = \frac{\sum_{l \in R(t_i)} \mathbf{x}_l \exp\{\mathbf{x}_l'\beta\}}{\sum_{l \in R(t_i)} \exp\{\mathbf{x}_l'\beta\}}$$

which is the weighted average of the regressors $\mathbf{x}_l$ for subjects at risk of failure at time t_i. The partial likelihood is equivalent a limited information likelihood, as $\lambda_0(t)$ has dropped out. But it is neither a conditional likelihood nor a marginal likelihood. Then is $L_p(\beta)$ a valid likelihood function? Andersen et al., '93 show that even $\ln L_p(\beta)$ yields a consistent estimator for β. See Lancaster (1990), chapter 9.

$$\mathbf{A}(\beta) = -\mathbf{B}(\beta), \text{ and } \hat{\beta}_{ML} \to_a N\big[\beta, (\mathbb{E}\big[\tfrac{\partial \ln L_(\beta)}{\partial\beta\partial\beta'}\big]\big)^{-1}\big]$$

The indexation p under the likelihood term stands for *partial*. The estimator is inefficient.

E Survivor function for the Cox proportional hazards model

For the proportional hazards model, it is possible to estimate non-parametrically the survivor function, once β is obtained after having maximized the partial log likelihood. Estimates are analogous to those of Kalpan and Meier. PH survivor function is defined as $S(t|\mathbf{x},\beta) = S_0(t)^{\phi(\mathbf{x},\beta)}$, using $S(t|\mathbf{x},\beta) = \exp\{-\int_0^t \lambda_0(s)\phi(\mathbf{x},\beta)ds\}$ and defining $S_0(t) = \exp\{-\int_0^t \lambda_0(s)ds\}$. Assume a discrete time formulation with baseline hazard $1-\alpha_j$, at discrete failure time t_j, $j = 1,...,k$. $\hat{\alpha}_j$ is the solution to

$$\textstyle\sum_{l\in D(t_j)} \frac{\phi(\mathbf{x}_l,\hat{\beta})}{1-\hat{\alpha}_j^{\phi(\mathbf{x}_l,\beta)}} = \sum_{m\in R(t_j)} \phi(\mathbf{x}_m\hat{\beta}),\ j = 1,...,k.$$

$\hat{\beta}_{PML} \equiv$ partial likelihood estimator of β, $\rightsquigarrow$ $S_0(t) = \prod_{j|t_j\le t} \alpha_j$, the cumulative product of the instantaneous conditional survivor probabilities. Estimated survival function (baseline) is,

with no regressors, $\hat{S}_0(t) = \prod_{j|t_j\le t} \hat{\alpha}_j$.

$\hat{S}_0(t)$ reduces to the Kaplan - Meier estimator, normalize $\phi(\mathbf{x}_l,\beta) = 1$ and the expression yields hazard rate $1-\hat{\alpha}_j = \frac{d_j}{r_j}$.

With regressors but without ties, the baseline hazard rate,

$$1-\hat{\alpha}_j = \phi(\mathbf{x}_j,\hat{\beta}) / \textstyle\sum_{m\in R(t_j)} \phi(\mathbf{x}_j,\hat{\beta}).$$

Survivor function for individuals with regressors $\mathbf{x} = \mathbf{x}^*$ can be estimated via $\hat{S}(t|\mathbf{x}^*,\beta) = \hat{S}_0(t)^{\phi(\mathbf{x}^*,\hat{\beta})}$. Linear transformation of regressors don't change the estimates of β, but they do change the hazard function.

$$\begin{aligned} \lambda(t|\mathbf{x},\beta) &= \lambda_0(t)\exp\{\mathbf{x}\beta\} \\ &= \lambda_0(t)\exp\{\mathbf{x}'\beta\} \qquad \underbrace{\exp\{(x-\bar{x})'\beta\}}_{\text{deviation from the mean of x's}} \\ &= \lambda_0^*(t)\exp\{(x-\bar{x})'\beta\} \end{aligned}$$

new baseline hazard leads to demeaning each regressor will change the baseline hazard and requires care in interpretation.

Following Kalbfleisch and Prentice 2002, α_j is derived.

$$
\begin{aligned}
S(t_j|\mathbf{x},\beta) - S(t_{j+t}|\mathbf{x},\beta) &= S_0(t_j)^{\phi(\mathbf{x},\beta)} - S_0(t_{j+1})^{\phi(\mathbf{x},\beta)} \\
&= [\alpha_j^{-1} S_0(t_{j+1})]^{\phi(\mathbf{x},\beta)} - S_0(t_{j+1})^{\phi(\mathbf{x},\beta)} \\
&= [\alpha_j^{-\phi(\mathbf{x},\beta)} - 1] S_0(t_{j+1})^{\phi(\mathbf{x},\beta)}
\end{aligned}
$$

since $S_0(t_{j+1}) = \prod_{l=1}^{j} \alpha_l = \alpha_j S_0(t_j)$, and the first term means that subject with duration time t_j has likelihood contribution equal to the probability of survival time $t > t_j$.

For these subjects who are censored at time t_j, the likelihood contribution is the propbability of survival $t > t_j$ or $S_0(t_{t+1})^{\phi(\mathbf{x},\beta)}$. So subjects that either exit the risk set or are censored at $[t_j, t_{j-1}]$ contribute probability $S_0(t_{j+1})^{\phi(\mathbf{x},\beta)} = \prod_{l=1}^{j} \alpha^{\phi(\mathbf{x},\beta)}$ with an additional multiplier $\{\alpha_j^{-\phi(\mathbf{x}\beta)} - 1\}$ for subjects that deliver before the end of the survey time. The over all failure times the likelihood $\rightsquigarrow$ $L(\alpha,\beta) = \prod_{j=1}^{k} \big[\prod_{l \in D(t_j)} \big(\alpha_j^{-\phi(\mathbf{x}_l,\beta)} - 1\big) \prod_{m \in R(t_j)} \alpha^{\phi(\mathbf{x},\beta)} \big]$.

$\ln L(\alpha,\beta) = \sum_{j=1}^{k} \big\{ \sum_{l \in D(t_j)} \ln(\alpha_j^{\phi(\mathbf{x},\beta)} - 1) + \sum_{m \in R(t-j)} -\phi(\mathbf{x},\beta) \ln \alpha_j \big\}$.

$\frac{\partial \ln L(\alpha,\beta)}{\partial \alpha_j} = 0 \Leftrightarrow$ as before.

Grouped duration models are appropiate when failure times are aggregated and observed/recorded at aggregate time intervals like a *week*, a *month*, a *year*. A simple method is to form a panel and estimate a stacked logit or probit model of the probability of an individual failure in each period, with separate intercept for period (fixed effects). Discrete time variant of a continuous time PH model considered by several authors such as Kalbfleisch and Prentice (1980), [...][26].

Grouped data with grouping point t_a, where a stands for annum in our case, $a = 1, ..., A$, the discrete hazard function is defined by:

$$\lambda^d(t_a|\mathbf{x}) = Pr[t_{a-1} \leq T < t_a | T \geq t_{a-1}, \mathbf{x}(t_{a-1})],\ a = 1, ..., A$$

Time varying regressors are permitted. The associated discrete time survivor function is:

$$S^d(t_a|\mathbf{x}) = Pr[T \geq t_{a-1}|\mathbf{x}] = \textstyle\prod_{s=1}^{a-1}(1 - \lambda^d(t_s|\mathbf{x}(t_s))).$$

We obtain the first general relation between the discrete and continuous time hazards. Discrete time hazards $\rightsquigarrow$ probability of failure in $[t_{a-1}, t_a)$ divided by the probability of surviving until at least time t_{a-1}:

$$\lambda^d(t_a|\mathbf{x}) = \tfrac{S(t_a|\mathbf{x}) - S(t_a|\mathbf{x})}{S(t_{a-1}|\mathbf{x})}.$$

[26]Following Blake, Lunde and Timmerman (1999).

$S(t|\mathbf{x}) \equiv$ survivor function. $S(t|\mathbf{x}) = \exp\{-\int_0^t \lambda(s)ds\}$, and, after some algebra,

$$\lambda^d(t_a|\mathbf{x}) = 1 - \exp\{-\int_{t_{a-1}}^{t_a} \lambda(s)ds\}.$$

Now, let us specialize to the discrete-time hazard model associated with the continuous PH model $\rightsquigarrow$

$$\lambda(t) = \lambda_0(t)\exp\{\mathbf{x}(t_{a-1})'\beta\}, \text{ for } t \in [t_{a-1}, t_a).$$

Regressors are constant within the interval, but can vary across intervals. $\lambda_0(t)$ instead can vary within the interval.

$$\begin{aligned}\lambda^d(t_a|\mathbf{x}) &= 1 - \exp(-\exp\{\mathbf{x}(t_{a-1})'\beta\})\int_{t_{a-1}}^{a_t} \lambda_0(s)ds \\ &= 1 - \exp\{-\lambda_0(t_a)\exp\{\mathbf{x}(t_{a-1})'\beta\}\}.\end{aligned}$$

Associated discrete $t_i = \int_{t_{a-1}}^{t_a} \lambda(s)ds$. The survivor function:

$$S^d(t_a|\mathbf{x}) = \prod_{s=1}^{a-1} \exp\{-\exp\{\ln \lambda_{0s}\mathbf{x}(t_{s-1})'\beta\}\}.$$

The density for the i^{th} subject is the product of the survivor function in each period that the subject survives (i.e. the mother does not get get pregnant for an additional child) times the hazard at the time of the vital event.

$$\begin{aligned}L(\beta, \lambda_{01}, ..., \lambda_{0A}) = \prod_{i=1}^N \Big[\prod_{s=1}^{a_i-1} \exp\{-\exp\{\ln \lambda_{0s} + \mathbf{x_i}(t_{a-1})'\beta\}\}\Big] \times \\ \times \big[1 - \exp\{\exp\{\ln \lambda_{0a_i} + \mathbf{x}_i(t_{a-1})'\beta\big],\end{aligned}$$

when censoring is ignored for simplicity and failure is assumed to occur at time t_{a_i} for the i^{th} observation. At least one failure is assumed to occur in each interval $[t_{a-1}, t_a)$. The MLE maximizes the latter likelihood function w.r.t. βa dn $\lambda_{01}, ..., \lambda_{0A}$, such as a polynomial in time. $\lambda_0(s) = \int_{t_{a-1}}^{t_a} \alpha s^{\alpha-1}ds$, as in a fully parametric Weibull model.

References

Blayo, Chantal, Christophe Bergouignan, André Etchelecou, Valérie Codazzo, Marine Llopart, Nicodéme Okobo (2002). “Situation démographique et sociale du Burundi: Résultats de l’enquête socio-démographique et de santé de la reproduction ESDSR Burundi 2002”, Fonds

Table 10: The Weibull and Exponential distributions.

FUNCTION	EXPONENTIAL	WEIBULL
$f(t)$	$\gamma \exp\{-\gamma t\}$	$\gamma \alpha t^{\alpha-1} \exp\{-\gamma t^{\alpha-1}\}$
$F(t)$	$1 - \exp\{-\gamma t\}$	$1 - \exp\{-\gamma t^{\alpha}\}$
$S(t)$	$\exp\{-\gamma t\}$	$\exp\{-\gamma t^{\alpha-1}\}$
$\lambda(t)$	γ	$\gamma \alpha t^{\alpha-1}$
$\Lambda(t)$	γt	γt^{α}
$\mathbb{E}[t]$	$\frac{1}{\gamma}$	$\gamma^{-\frac{1}{2}} \Gamma[\alpha^{-1} + 1]$
$Var[t]$	$\frac{1}{\gamma^2}$	$\gamma^{-\frac{2}{\alpha}} [\Gamma(2\alpha^{-1} + 1)] - [\Gamma(\alpha^{-1} + 1)]$ [27]
γ, α	$\gamma > 0$	$\gamma > 0, \alpha > 0$

des Nations Unies pour la Population et Département de la population du Burundi.

Cameron, Colin A. and Pravin K. Trivedi (2005). "Microeconometrics", *Oxford University Press.*

Cox, David R. (1972). "Regression Models and Life-Tables", *Journal of the Royal Statistical Society, Series B*, 34(2), 187-220.

Kalbfleisch, Jack D. and Ross L. Prentice (1980). "The Statistical Analysis of Failure Time Data", *John Wiuley and Sons Inc.*

Lunde, Asger, Allan Timmermann, and David Blake (1999). "The Hazards of Mutual Fund Underperformance: A Cox Regression Analysis", *Journal of Empirical Finance*, Vol. 6, No. 2, p. 121-152.