Ethnic Polarization and the Duration of Civil Wars[1]

Jose G. Montalvo
Universitat Pompeu Fabra and IVIE

Marta Reynal-Querol
Universitat Pompeu Fabra, CEPR and CESifo

Abstract:

In this paper we analyze the relationship between ethnic polarization and the duration of civil wars. Several recent papers have argued that the uncertainty about the relative power of the contenders in a war will tend to increase its duration. In these models, uncertainty is directly related to the relative size of the contenders. We argue that the duration of civil wars increases the more polarized a society is. Uncertainty is not necessarily linked to the structure of the population but it could be traced back to the measurement of the size of the different groups in the society. Given a specific level of measurement error or uncertainty, more polarization implies lengthier wars. Our empirical results show that ethnically polarized countries have to endure longer civil wars than ethnically less polarized societies.

Keywords: ethnicity, discrete polarization, civil wars, duration.

---

1. *Introduction*

Only recently civil wars have been recognized as one of the main impediments for economic development. Their effects are not only related to the destruction of infrastructure or human life but also to the elimination of the rule of law, the generation of an uncertain environment for future foreign investment and the destruction of institutions. This generates a conflict trap (Collier et al 2003) that keeps countries in a low per capita income equilibrium.

The study of the causes of civil wars has generated an incipient literature. The analysis has taken three different approaches: the onset of civil wars, their incidence and their duration. Obviously these three analyses are complementary but deal with different sides of the civil war phenomenon. We can create an analogy with the analysis of macroeconomic cycles. Researchers in that field distinguish between shocks and its propagation mechanism as two different and independently interesting issues. For instance, a cycle could be caused by a productivity shock that it is propagated through many alternative mechanisms. One of them, for instance, could be the time to build an argument.

In the case of civil wars the situation is similar although the identification is more difficult. In principle some factors that may affect the onset of a civil war could have no impact on its duration. For instance, in many situations civil wars start by random acts, which trigger, given a particular propagation mechanism, full-fledged conflicts. However, some of the elements that explain the onset of a civil war may also be embedded in the propagation mechanism. A country with a high proportion of mountains or jungles in its territory may have a higher probability of a rebel group forming. Notwithstanding, at the same time, the mountains help rebels to hide and prolong the duration of the war given all the other factors. Per capita income may have a similar effect: low per capita income reduces the opportunity cost of a war which implies not only that the onset of a war is more likely, but also that this conflict lasts longer than otherwise. Other factors may be more easily identified. A government in a weakly democratic state may engage into a conflict with some victimized rebel group. It is less likely, however, that the level of democracy present at the beginning of the civil war also affects its duration.

Empirical research has used different approaches to deal with the explanation of the basic elements of civil wars. Researchers have analyzed the onset of civil wars (Fearon and Laitin

2003), the incidence[2] (Montalvo and Reynal-Querol 2005a), the hazard or the duration (Collier et al. 2004). These indicators are related but measure different concepts. The probability of onset is the conditional probability of being in state A (war) at time t given that it was in state B (peace) at time t-1. The incidence is the unconditional probability that we observe state A at time t. The hazard function measures the probability of being in state A between period t and t+1 given that it was in state A in period t. The duration measures the units of time in state A for each event. Following these definitions, there is a classical intuitive equation for the aggregate cross sectional variation: incidence = onset x duration.

Conflict onset is a very unpredictable event. In many conflicts the onset is related with some unpredictable shocks like, for instance, the original trigger of the genocide in Rwanda or the recent events in France (the burning of vehicles during November of 2005). But even if the onset was predictable, what if the duration is very short? The empirical study of the onset of civil wars does not take these facts into consideration since it does not consider any measure of duration. In addition, if we are dealing with rare (infrequent) events there is a strong statistical argument in favor of analyzing incidence or duration instead of onset. Moreover, Hegre and Sambanis (2006) perform a sensitivity analysis of the specification for the onset of civil war using the variables typically included in the civil war literature. They conclude that only population exhibits a robust significant effect on onset civil war. These results suggest that onset is not the appropriate variable to study the causes of civil war.

From a policy point of view, knowing what makes civil war longer, and which are the propagation mechanisms that make conflict persistent is important in order to define the policies that need to be implement to stop civil wars, and to ensure peaceful post-conflict periods. Identifying the propagation mechanisms is important to design the adequate policies and institutional environment in order to reduce the propagation effect.

Based on our previous research (Montalvo and Reynal-Querol 2005a) we examine in this article whether ethnic polarization is a possible dimension of the propagation mechanism of civil wars. Therefore, in this paper we investigate empirically the influence of ethnic polarization on the duration of civil wars. We also discuss several theoretical models (Smith and Stam 2004, Slantchev 2003 and 2004) that can explain why ethnic polarization, or measures closely related to it, can affect the length of civil wars.

---

[2] In the epidemiological literature this is referred to as prevalence.

Our empirical estimation shows that the hazard function for the event of the end of a civil war is much lower for countries with a level of ethnic polarization over the median than for the rest of the countries. In fact, an increase of ethnic polarization from 0 to 1 multiplies by 2 the median duration of a civil war. This result suggest that we should pay special attention to countries that are polarized, in order to implement appropriate institutional settings and policies that reduce the feeling of grievance, that seem to be higher in polarized countries.

2. A short guide to the recent literature

Recent literature on the causes of civil wars has postulated several factors as robust determinants. Collier and Hoeffler (2004) use logit regression to estimate the onset of a war breaking the time dimension into five-year episodes. They find that higher primary commodity exports[3], diaspora and higher and more dispersed population increase the risk of a conflict. By contrast, a higher per capita income, time since the last conflict, growth and secondary enrollment reduce the risk. Collier and Hoeffler (2004) find no effect of income inequality, political liberties or ethnic heterogeneity on the risk of a civil war. Fearon and Laitin (2003) consider annual data and a different specification that includes as a regressor the lag of the dummy for civil wars[4]. In their analysis mountainous countries, oil exporters, anocracies, and highly populated country face a larger risk of conflict. The only factor that decreases the probability of war is a high per capita income. Fearon and Laitin (2003) find no role for ethnic heterogeneity or democracy as a risk factor for civil wars.

Montalvo and Reynal-Querol (2005a) analyze the incidence instead of the onset of civil wars. They find that ethnic polarization is a very robust determinant of the incidence of civil wars, while the statistical significance of population and the initial level of per capita income seem to be dependent on a particular specification. In their analysis, neither primary exports, democracy nor ethnic fractionalization had any significant effect on the incidence of civil wars. A set of articles on the duration of civil wars are found in a special issue of *Journal of Peace Research* published in 2004. Fearon (2004) classifies the wars into different categories (coups

---

[3] Fearon (2005) disputes this fact and finds that the effect is weak and not robust even if one adopts the Collier and Hoeffler (2004) specification and uses their definition of civil wars.

[4] Since the specification is not properly transformed to deal with the inclusion of a lagged dependent variable, the results are econometrically questionable.

and popular revolutions, post 1991 civil wars in Eastern Europe, decolonization wars, "son of soil" wars and contraband financing) and runs a multivariate Weibull regression of the duration on these categories and the usual determinants (GDP per capita, population, democracy and the ethnic war dummy). When he controls for the type of war, the other variables (population, GDP per capita, ethnic fractionalization and democracy) have no effect on duration. Obviously, since duration is one of the "informal" determinants of the type of war, the estimation is biased because of endogeneity.

Gates and Strand (2004) also discuss the estimation of duration models for civil wars. They find a robust effect for the intensity of the conflict, political instability, autocratic governments and the existence of parallel conflicts. They also find that using precise dating (duration measured in days) eliminates the significance of GDP per capita on duration. However, they do not include any control for the ethnic diversity of the countries. Collier et al. (2004), using the COW definition of war, monthly data, and, a piecewise exponential duration model, establish that income inequality and population have a positive effect on the duration of civil wars. However, they also find that ethnic fractionalization also has a nonlinear effect on the duration since its square is also significant. As in the case of the onset of a civil war, the only variable that seems to reduce conflict duration is the initial level of per capita income.

Our empirical approach lies in-between Gates and Strand (2004), since we use the Uppsala/PRIO definition of civil war (see Gleditsch et al. 2002) and Collier et al. 2004, since we use the standard explanatory variables plus ethnic diversity proxies such as ethnic polarization. In the rest of the paper we investigate the explanatory power of ethnic polarization on the duration of civil wars. Section 3 presents a discussion on the measurement of ethnic polarization. The next section describes some theoretical considerations that justify the relationship between ethnic polarization and the duration of civil wars. In Section 5 we discuss the importance of the definition of civil wars and some econometric issues related with the estimation of these models. Section 6 includes the estimation of different duration models. Finally, Section 7 presents the preliminary conclusions and a summary of future lines of research.

3. On the measurement of ethnic polarization

Ethnic polarization has been shown to be an important determinant of the incidence of civil wars (Montalvo and Reynal-Querol 2005a). For this reason, in principle it is also a natural candidate to explain its duration. The most important aspect in the measurement of ethnic polarization is the kind of indicator we should use, and be aware of the theoretical implications of using one index or another.

The number of papers dealing with the effects of ethnic diversity on issues of economic interest is growing at a fast rate, mostly with an empirical content[5]. By now it is common place to include, in growth regressions, an index of ethnic diversity. All these papers use the index of ethnolinguistic fractionalization (ELF) as the indicator of ethnic diversity. The raw data come from the Atlas Narodov Mira (1964), compiled in the former Soviet Union, which refers to the situation in 1960. The criteria for group formation were based on the historical linguistic origin. The measure ELF was calculated by Taylor and Hudson (1972), which summarizes the data of the Atlas using the Herfindahl index. In particular the index takes the form,

$$FRAG = 1 - \sum_{i=1}^{N} \pi_i^2$$

where if we consider religious (or ethnic) diversity, π_i is the proportion of people who belongs to ethnic group i. The broad popularity of the ELF index is based on its intuitive appeal. Hence, the index captures the probability that two randomly selected individuals from a given country will not belong to the same ethnolinguistic group.

However, many authors have found that, even though ethnic fractionalization seems to be a powerful explanatory variable for economic growth, it is not significant in the explanation of civil wars and other kinds of conflicts. These results have led many authors to disregard ethnicity as a source of conflict and civil wars. Fearon and Laitin (2003), and Collier and Hoeffler (2004), find that neither ethnic fractionalization nor religious fractionalization has any statistically significant effect on the probability of civil wars. Nevertheless, it is not clear to what extent an index of diversity could capture potential ethnic conflict. In principle, claiming a positive relationship between an index of fractionalization and conflicts implies that the more ethnic

---

[5] For instance Bluedorn (2001), Vigdor (2002) or Caselli and Coleman (2002).

groups there are the higher is the probability of a conflict. Many authors would dispute such an argument.

The issue of how to construct an index, which is appropriate to measure the relevant aspects of ethnic diversity, is the basic point discussed in Montalvo and Reynal-Querol (2002, 2005a). Let's imagine that there are two countries, A and B, with three ethnic groups each. In country A the distribution of the groups is (0,49, 0,49, 0,02) while in the second country, B, is (0,33, 0,33, 0,34). Which country will have a higher probability of social conflicts? Using the index of fractionalization the answer is B. However, Montalvo and Reynal-Querol (2002, and 2005a) and Reynal-Querol (2002) have argued that the answer is A. They use the index of ethnic polarization RQ, originally constructed by Reynal-Querol (1998, 2001), which takes the form

$$RQ = 1 - \sum_{i=1}^{N} \left[\frac{0.5 - \pi_i}{0.5} \right]^2 \pi_i$$

where π_i is the proportion of each ethnic group and N is the number of ethnic groups. This measure was originally constructed in Reynal-Querol (1998), trying to capture intuitively how far the distribution of the groups is from a bipolar distribution. The RQ index considers, implicitly, that the distances are 0 (an individual belongs to the group) or 1 (it does not belong to the group). The same assumption is made by the FRAG measure. This is very important since it is the basic feature that distinguishes this measure (FRAG) from the GINI index of income inequality. The measurement of polarization in a one-dimensional set-up was initiated by Esteban and Ray (1994) (ER) and Wolfson (1994). Both the income polarization (a particularly important case of one-dimensional polarization) and the Gini index, assume that distances among groups are measured on the real line.

Going from the real line to a discrete metric has important implications. A discrete metric (belong/do not belong) takes the following form

$$\delta(y_i, y_j) = \begin{array}{ll} 0 & if \quad i = j \\ 1 & if \quad i \neq j \end{array}$$

In a context with discrete distances (belong/do not belong) groups cannot be ordered on the real line as in the income case. Montalvo and Reynal-Querol (2002, 2005) show that the index of fractionalization can be interpreted as a GINI index with discrete distance. Moreover, they also

show that the measure of ethnic polarization, RQ, can be interpreted as the index of polarization of ER with discrete distances, by analogy to the relationship between the GINI index and the index of fractionalization.

We argue that any index of ethnic heterogeneity that pretends to capture potential conflict should be compatible with a theoretical model. Montalvo and Reynal-Querol (2002, 2005) show that the index of fractionalization does not capture the relevant conflictive dimension of ethnic divisions. The simple interpretation of the index of fractionalization as a probability is the basic reason why the index has been widely used. However, in the context of conflict and rent seeking models this measure is not the relevant indicator of the intensity of the conflict, while the RQ indicator can be easily justified. Horowitz (1985), in his seminal contribution to the study of ethnic conflict, argues that the relationship between ethnic diversity and civil wars is not monotonic: there is less violence in highly homogeneous and highly heterogeneous societies, and more conflicts in societies where a large ethnic minority faces an ethnic majority. If this is indeed the case, then an index of polarization would depict the likelihood or the intensity of conflict better than an index of fractionalization.

The measures of ethnic polarization (RQ) and ethnic fractionalization (FRAG), assume discrete distances (belong/do not belong). But, why do researchers use indices with discrete distances in the context of ethnic groups? The use of a discrete metric is reasonable since the identity of the groups is clearer than their relative "ethnic distances". There are other reasons that favor the use of indices with discrete distances. First of all, there are no empirical measures of distance across ethnic groups available and generally accepted. The measure of the "distance" across ethnic groups is much more controversial than the identification of the list of ethnic groups and their classification. Obviously, this does not imply that we cannot, imprecisely, get a measure of "closeness". It seems clear that the distance from Catholics to Protestants is "smaller" than the distance from Catholics to Muslims. However, the basic criterion is the "belong/do not belong" distinction in function of some classification. When we say that group A is different from group B and different from group C we mean that the distance across their characteristics have crossed a threshold that makes them "sufficiently different" to be considered different groups. Once the trigger point is achieved, the exact measure of the difference is irrelevant.

Second, for the previous reason, the measurement of distances across groups may generate a larger measurement error than the "belong/do not belong" criterion. Third, if we measure

distance across groups using the strength of the sentiment of identity or political relevance, then there is an important endogeneity problem. At the end we will be explaining conflict using conflict as the explanatory variable, because the sentiment of identity is high when there is conflict. Fourth, as argued by Montalvo and Reynal-Querol (2002) and Duclos, Esteban and Ray (2004), there are many interesting instances in which individuals are interested only in the dichotomous perception Us/They.

Figure 1 shows that the choice of indicator is important. We use the measures of ethnic fractionalization and ethnic polarization from Montalvo and Reynal-Querol (2002, 2005). Ethnic polarization and ethnic fractionalization exhibit a nonlinear relationship. For low levels of fractionalization the correlation between ethnic fractionalization and polarization is positive and high. In particular, we know that when there are only two ethnic groups ethnic polarization is two times ethnic fractionalization. That is the reason why the slope of the line is 1/2 for ethnic polarization up to 0.4. However, for the medium range the correlation is zero, and, for high levels of fractionalization the correlation with polarization is negative.

4. Ethnic polarization and the duration of civil wars: some theoretical considerations

Let's consider a country with two social groups. The size of each group is known with some measurement error[6]. Imagine that something (for instance the discovery of natural resources) ignites a conflict between both groups[7]. Assume that the parties will fight as long as the probability of winning of one of the groups is not significantly higher, in statistical terms, than the probability of the other. Once the conflict begins, troops clash in initial "battles". In each of them some information about the size of each group is revealed and the variance of the posterior distribution is reduced. The theoretical duration is defined as the number of "battles" needed for the conflict to end. Given the size of the groups the higher the initial variance of the measurement error the higher will be the duration conditional on the conflict having started. In addition, the higher the degree of polarization, the longer the duration of the war.

---

[6] We can also assume that the size of the groups is known, but the proportion of fighters that can be attracted by each group is a random variable function of the size of the group. In principle, there is no need to introduce asymmetric information.

[7] In this paper we are not concerned with the onset of civil wars but with the duration conditional on the onset.

We can find an example in the 2000 US presidential election. Although in general the measurement error in an electoral campaign is solved on the election day at latest, in this particular case the degree of polarization among the voters was so high that it took many electoral "battles" to solve the conflict. Obviously, given a particular level of measurement error, the higher the prize of a contested resource is, the lengthier the conflict.

This interpretation differs from the rebellion-as-mistake argument presented in Collier et al (2004). Under that conceptualization, the participants in a war may overestimate their chances to win. These irrational expectations increase the likelihood of the onset of a war but, at the same time, they lengthen the duration of the war. In our interpretation, groups do not have irrational expectations as in the rebellion-as-mistake interpretation. If a country has a high degree of ethnic/religious polarization, then the war will most probably break out along these lines. In this case it is also reasonable to assume that under many circumstances both contenders think rationally they are going to win. Imagine, for instance, a situation where there are two equally sized ethnic groups in a maximally polarized country, one in power and the other starting the rebellion. Since potentially they could form an army as large as their proportion in the population, then the likelihood of winning the war is similar for both parties. This means that their expectations to win are not overoptimistic but rational. In addition, if the mobilization of one of the parties is faster and it starts winning the war, the individuals of the other group could be mobilized endogenously depending on the information released in the different battles, which generates a high degree of violence and a lengthy war.

Our interpretation is closely connected with the recent literature on war as a bargaining process to which Smith and Stam (2004), Powell (2004), Slantchev (2003 and 2004) and others have contributed[8]. If they have any application at all, most of these papers refer to the case of interstate wars. However, these models are very general and could easily be applied to intra-state conflicts. Slantchev (2003) presents a model of simultaneous bargaining and fighting, where both players can make offers and asymmetric information exists about the distribution of power. He argues that learning occurs when strategically manipulable negotiation behavior and nonmanipulative battlefield outcomes reveal information. Slantchev (2003) argues that expectations are in fact central to explanations of rational war termination. War ends when opponents succeed in coordinating their expectations about what each is prepared to concede.

---

8 For an extensive review of this literature, see Powell (2002).

Total victory is a possible outcome. Nonetheless, an important function of wars is to convince the opponent to accept a settlement. This happens when the contenders learn enough about their prospects and understand that continuation is unprofitable. Slantchev (2004) hypothesizes that when observable capabilities are close to parity, wars will tend to be longer. In a way this "closeness to parity" concept is related with polarization. However, Slantchev (2004) interprets the closeness to parity directly as uncertainty. In addition, the measurement of this variable is achieved calculating the absolute difference in army (population) sizes scaled between 0 and 1. Slantchev (2004) shows that the higher the uncertainty and thus the closer to parity the contestants are, the longer is the war.

Smith and Stam (2004) explain why countries enter into war in the following way": The act of waging war reveals information about the relative strengths of each side. As a war progresses, each side's beliefs about the likely outcome of continuing the war converge. Once the warring parties' beliefs converged sufficiently, they can find a bargained solution to the conflict". They argue that these beliefs shape nations' (in the case of civil wars instead of nations we have social groups) expectations of the duration of conflict and determine which nation (group) is more likely to be the eventual winner if the war is fought to a decisive conclusion. In their model, as nations (groups) fight battles, both nations learn about the nature of warfare between them. As more and more information is revealed, the nations' (groups') beliefs converge. Their model thus shows how war resolves differences between states' beliefs.

What role does social polarization play in this kind of process? Imagine that a large majority is facing a minority. Imagine that there is a very large measurement error and people in the minority group believe that the anticipated number of fighters is larger than it actually is. After the conflict starts and a battle is fought some additional information is released, reducing the variance of the original measurement error. Therefore, the minority group recalculates its chances and may already concede. Their beliefs about the relative power of each group converge fast to the actual size of the group, which implies a conflict of short duration.

Imagine now that there are two groups of equal size fighting, the extreme polarization case. It is reasonable to think that they both believe that they can win even if the amount of noise is small. In this case, the convergence of beliefs will imply more "battle" rounds (release of information) considering a small amount of measurement error makes their probability of winning statistically equal. In fact, when the probabilities of winning are very similar, given a

11

degree of noise, the same speed of convergence, as in the case of large majority versus small minority, implies a longer war. This may explain why wars in polarized countries, other thing being equal, usually last longer than the disputes fought in less polarized countries[9].

5. Data and econometric issues

The first important issue in the analysis of the duration of civil wars is precisely the definition of the war. In this definition the number of casualties is critical given that most definitions use this criterion (yearly, for the whole period of the war or a combination of both) to set the beginning and the end of the war. Another important choice is the decision on the time span used to measure duration. For instance Fearon (2004) uses years while Collier et al (2004) use months and Gates and Strand (2004) consider days. Gates and Strand (2004) argue that the precision of the duration (using days instead of lower frequency spans) affect the results of the duration model. In particular they show that the effect of GDP per capita on duration, observed using yearly data, disappears when using duration measured in days. However, it is important to realize that the measurement error caused by using years instead of days as unit of duration has to be balanced with the extreme unreliability of casualty figures. If there are problems to know the number of casualties in one year, how could we set the exact date of beginning of a war? Necessarily setting the beginning of a civil war in a particular day is more arbitrary (and generates larger measurement errors) in terms of the conditions needed (which most of the time are different for each conflict and not comparable even when they seem reasonable), than a yearly dating.

A second data-related problem is caused by continuity issues. Depending on the database some wars are counted as one episode while other datasets represent the same war by two or even more episodes. As in the case of the first measurement error, the choice of the moment of the end of the civil wars can, potentially, have important effects on the statistical analysis of the duration of the wars. The previous measurement problems suggest the need to check for the robustness of the statistical analysis using alternative datasets and frequencies.

Another important issue with econometric consequences is the consideration of simultaneous episodes and the existence of competing risks. Gates and Strand (2004) and Fearon (2004) consider the possibility that several conflicts occur in the same country and during the

---

[9] This argument also applies to the case of more than two groups if coalitions are not allowed and one group is

same period of time. The problem with this approach is the need to include conflict-specific variables in the estimation (for instance the type of conflict). However, usually we include country specific characteristics and most of the constructed war-specific variables are endogenous to the duration of the war, which makes them unsuitable for use in the standard duration analysis. The question of competing risks is similar. A civil war may end, for instance, with the victory of the government or a truce. Since these outcomes are not independent, then the possibility of competing risks generates an effect which is similar to the impact that an unobservable variable which is outcome specific has. DeRouen and Sobek accordingly (2004) use a competing hazard analysis to distinguish the characteristics of conflicts that end up with different outcomes (rebel victory, government victory or truce/treaty).

Another sort of unmeasured heterogeneity is country specific. In the context of the onset of civil wars Alexander (2005) argues that the absence of country fixed effects in the estimation of the logit models of Collier and Hoeffler (2004) and Fearon and Laitin (2003) explains why per capita income had a significant effect on the probability of the onset of a civil war. The fact that the same country may experience several civil wars during the period of analysis implies what is known in the duration jargon as multiple failure-time data. This mimics the panel data versus pooled data situation in the linear regression case. The existence of an unobservable country specific effect may violate the condition of independence of failure times conditional on the explanatory variables, which lead to invalid inferences using the standard methods. The econometric literature has proposed several solutions to this problem. The simplest one is to estimate a robust and cluster-consistent estimator for the standard deviation of the parameters. A second possibility is the use of what is called the "frailty approach" which amounts to assume a particular distribution for the unobserved country specific effects (similar to the approach adopted by the so called random effects panel data estimator). Finally, there are also other modelling strategies available for multiple failure-time data like the Andersen-Gill model.

A final econometric issue has to do with the estimation of continuous time models versus their discrete alternative. This difference is not very important when there are single failure-time data but may be important if we deal with multiple failure time data.

---

supposed to win if its size is significantly larger than the sum of the other groups.

6. The effect of ethnic polarization on the duration of civil wars

The arguments exposed above indicate that ethnic polarization may be a powerful propagation mechanism for civil wars and, therefore, have an important influence on the duration of civil wars. We are going to use initially the definition of civil wars of Uppsala/PRIO25 (see the Appendix for a precise definition). This data set includes wars that cause at least 25 yearly deaths. Our database contains 117 civil wars corresponding to 74 countries during the period 1960-99. We consider that a war has finished if there are at least two years of peace before a new civil wars breaks out. Following this definition we have 42 countries with only 1 civil war; 24 countries with 2 civil wars; 5 with 3 wars and 3 with 4 wars. The mean duration is 7.15 years. In 1999 there were 20 ongoing civil wars.

As a preliminary indication we can look at the survivor functions of civil wars in highly polarized countries versus countries with a lower level of ethnic polarization. To separate these two groups we use the median value of ethnic polarization[10]. Figure 2 shows the Kaplan-Meier survivor function for countries with a degree of ethnic polarization above the median versus a degree of polarization below the median. Notice that this only provides a preliminary indication since we have not included any covariate. The Kaplan-Meier estimator is a nonparametric procedure, alternative to the life tables, to obtain survivor functions.

Figure 2 shows that the survivor function of the countries with a high degree of ethnic polarization dominates almost everywhere the function for lower ethnic polarization. Since the failure time in this case is the end of a civil war and the time at risk is the period of civil war the dominance of the survivor function of highly polarized countries implies that the duration of wars is longer in those countries.

There are several alternatives to deal with the analysis of duration in a parametric set-up. The most important factor is the effect of the parametric distribution on the censured observations. Collier et al (2004) use a piecewise[11] exponential model to overcome the restrictions implied by the simple exponential model (in particular, a constant hazard function). However, most of the literature uses a Weibull model when running a parametric estimation (for instance Fearon (2004), Gates and Strand (2004) in their basic table and DeRouen and Sobek

---

[10] The appendix describes the data on civil wars and the explanatory variables used in the duration models of this section.

[11] The choice of the pieces of the exponential model is somehow arbitrary.

(2004) for one of the outcomes) since this specification is quite flexible in terms of the implied hazard functions, which takes the form

$$\lambda(t) = \lambda \alpha t^{\alpha-1} = e^{X\beta} \alpha t^{\alpha-1}$$

The same specification can be written in an accelerated failure time form instead of a hazard function. This is the form needed to calculate the expected (mean) time of the marginal effect of each explanatory variable. The specification is

$$\ln t_i = X_i \beta + u_i$$

where the error terms are supposed to have an extreme value density function.

Table 1 shows the estimation under different sets of explanatory variables and a Weibull specification. Table 1 demonstrates that the effect of population and ethnic polarization are robust across all the specifications. Both variables have a positive effect on the duration of civil wars as shown by the estimated hazard ratios. In column 6 the oil dummy has a negative effect on duration, opposite to the findings in Fearon (2004)[12]. In any case, this dummy is not very robust. Figure 3 shows the hazard function derived from specification 6, which is very similar to the one obtained from the previous columns. It shows that for the definition of war of Uppsala/PRIO25 the probability of the war ending conditional on having lasted up to t is decreasing in time.

More interesting for our purposes is to look at the hazard function for highly polarized countries versus countries with low polarization. Figure 4 shows the hazard function obtained using a Weibull model. Instead of using the continuous polarization index, we include a dummy variable that takes the value 1 if the value of polarization is over the median[13]. The parameter of this variable is highly significant as it was the index of polarization in its continuous version. Figure 4 shows that the hazard function for the end of a civil war is much lower for highly polarized countries.

Table 2 shows the marginal effect of the explanatory variables on the mean and median duration obtained from the accelerated failure version of the model. Since the inferences on the median are more robust, we are going to concentrate the interpretation on column 2 and, in particular, the effect of ethnic polarization. The main result is that an increase of this variable from 0 to 1 increases the duration of a war by 7.9 years. Evidently, this is not a realistic change because it covers the whole range of the variable. However, it clearly indicates the importance of

---

[12] This effect needs more analysis using alternative proxies for oil like exports, production or dummies related with these variables.

ethnic polarization in explaining the duration of wars. This result is in line with the large difference in the hazard function between low and high polarization countries.

Table 3 shows that the results are robust to the introduction of regional dummies and the use of a definition of civil wars that includes only the most intense (Uppsala/PRIOCW) as in table 4. The only noticeable difference between the results in table 1 and 4 is the fact that, given the estimated α, the hazard function is sloping upward for the estimation of the most intense wars while it is sloping downward when using a more comprehensive definition of civil war.

Another approach for the estimation of the model is to adopt a semi-parametric strategy where part of the model is parametrized and the other is left as a baseline hazard function. In this kind of model the most popular approach is the Cox model. In general the proportional Cox model can be written as

$$\lambda(t) = \lambda_0(t)e^{X\beta}$$

where using pseudo-maximum likelihood estimation we can estimate the parameters β without having to deal with the estimation of the baseline function $\lambda_0(t)$. The baseline function could be thought as group specific generating a hazard function of the form

$$\lambda_i(t) = \lambda_{0i}(t)e^{X\beta}$$

Andersen and Gill (1982) show how to modify the Cox specification to allow for multiple failure-time data. Table 5 shows that the results of this estimation procedure are consistent with our previous results. The effect of ethnic polarization is statistically significant while ethnic fractionalization proves not significant.

Table 6 shows the results of testing the proportionality assumption embedded in the estimation of the Cox model[14]. We test the hypothesis of a zero slope, which would imply that the log hazard ratio function is constant over time. If the test rejects the null hypothesis of equality to 0, this implies a rejection of the proportional hazard assumption. Table 6 shows that we cannot reject the null hypothesis in all except for the initial GDP variable. In any case, this variable was originally not statistically significant[15]. In addition, the global test of the null hypothesis cannot reject the proportionality assumption of the Cox model. Therefore, the results of the Cox

---

[13] Notice that this is the same definition we used previously in the estimation of the Kaplan-Meier function.

[14] Box and Zorn (2003) present models for non-proportional hazards applied to international relationships.

[15] We include the initial GDP in the specification for the sake of correspondence of the models with the results of Table 5. If we eliminate this variable, we cannot reject for any individual variables the null hypothesis.

regression support the previous findings and are not caused by the failure of the proportionality assumption maintained in the estimation of the model.

7. Conclusions

Several recent papers have argued that the uncertainty about the relative power of the contenders in a war will tend to increase its duration through its effect on their expectations. In these models the uncertainty is directly related to the relative size of the contenders. We argue that another link between the degree of polarization of the society and the duration of civil wars can be found if the size of the groups that form the population is measured with error. Given a specific level of measurement error or uncertainty, more polarization implies lengthier wars. Our empirical estimations show that the hazard function for the event of the end of a civil war is much lower for countries with a level of ethnic polarization over the median than for the rest of the countries. In fact, an increase of ethnic polarization from 0 to 1 multiplies by 2 the median duration of a civil war.

Identifying the propagation mechanisms is important for designing the adequate policies and institutional environment in order to mitigate the length of conflicts. The results in this paper suggest that we should pay special attention to countries that are polarized, in order to implement appropriate institutional settings and policies that reduce the feeling of grievance, that seem to be high in polarized countries.

APPENDIX: Definition of civil wars from Uppsala/PRIO.

Uppsala/PRIO25: Uppsala/PRIO definition including armed conflicts that generate more than 25 deaths yearly (minor armed conflicts plus intermediate plus war following Uppsala/PRIO classification). We only consider types 3 and 4 (internal armed conflicts).

Uppsala/PRIOCW: Intermediate and war definition of armed conflict from Uppsala/PRIO. This is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths yearly and a minimum of 1,000 deaths over the course of the civil war. We only consider types 3 and 4 (internal armed conflicts).

REFERENCES

Alexandre, M. 2005. I*s poverty to blame for civil war? New evidence from nonlinear fixed effects estimation*. Mimeo, Harvard University.

Andersen, P. and R. Gill. 1982. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 10 (4): 1100-20.

Atlas Narodov Mira, Atlas of the People of the World. 1964. Moscow: Glavnoe Upravlenie Geodezii i Kartografii .Bruck, S.I., and V.S. Apenchenko (eds.).

Barret, D. (Ed.).1982. *World Christian Encyclopedia*. Oxford University Press.

Blainey, G. 1988. *The Causes of War*. 3rd edition, London, Macmillan.

Bluedorn, J. C. 2001. Can Democracy Help? Growth and ethnic divisions. *Economics Letters* 70: 121-6.

Box-Steffensmeier, J. and D. Zorn. 2003. Nonproportional hazards and event history analysis in international relationships. *Journal of Conflict Resolution* 47 (1): 33-53.

Caselli, F. and W. J. Coleman. 2002. On the theory of ethnic Conflict. mimeo.

Collier, P., V. Elliot, H. Hegre, A. Hoeffler, M. Reynal-Querol and N. Sambanis. 2003. *Breaking the conflict trap: civil war and development policy*. Washington DC: World Bank and Oxford University Press.

Collier, P., and A. Hoeffler. 2004. Greed and Grievance in Civil War. *Oxford Economic papers* 56: 663-595

Collier, P., A. Hoeffler and M. Söderbom. 2004. On the duration of civil war. *Journal of Peace Research* 41(3): 253-73.

DeRouen, K. and D. Sobek. 2004. The dynamics of civil war duration and outcome. *Journal of Peace Research* 41 (3): 3003-320.

Duclos, J., Esteban, J. and D. Ray. 2004. Polarization: concept, measurement and estimation. *Econometrica* 74: 1737-72.

Easterly, W., and R. Levine. 1997. Africa's growth tragedy: Policies and Ethnic divisions. *Quarterly Journal of Economics* CXII(4): 1203-1250.

Esteban, J. M and D. Ray. 1994. On the measurement of Polarization. *Econometric,a* 62 (4).

Fearon J. and D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1): 75-90.

Fearon, J. 2004. Why do some wars last so much longer than others?. *Journal of Peace Research* 41: 275-301.

___. 2005. Primary commodity exports and civil wars. *Journal of Conflict Resolution* 49 (4): 483-507.

Gates, S. and H. Strand. 2004. Modelling the duration of civil wars: measurement and estimation issues. Mimeo, Centre for the Study of Civil Wars, PRIO.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand. 2002. Armed Conflict 1946–2001: A New Dataset. *Journal of Peace Research* 39 (5): 615-37.

Hegre, H. and N. Sambanis. (2006). "Sensitivity Analysis of Empirical Results on Civil War Onset **".** *Journal of Conflict Resolution* 50: 508-535. [

Horowitz, D. 1985. *Ethnic groups in conflict.* University of California Press.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny. 1999. The Quality of Government. *Journal of Law, Economics and Organization* 15 (1): 222-79.

Mauro, P. 1995. Corruption and Growth. *Quarterly Journal of Economics* CX: 681-712.

Montalvo, J. G. and M. Reynal-Querol. 2002. Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth. UPF Working Paper 660.

____. 2005a. Ethnic polarization, potential conflict and civil wars. *American Economic Review* 95 (3): 796-816.

____. 2005b. Ethnic Diversity and Economic Development. *Journal of Development Economics* 76: 293-323.

Posner, D. N. 1999. Ethnic fractionalization in Africa. mimeo.

Powell, R. 2004. The Inefficient Use of Power: Costly Conflict with Complete Information. *American Political Science Review* 98 (2): 231-41.

Powell, R. 2002. Bargaining Theory and International Conflict. *Annual Review Political Science* 5: 1-30.

Reynal-Querol, M. 1998. Growth and Religious Conflict. Mimeo, Universitat Pompeu Fabra.

____. 2001. Ethnic and Religious Conflict, Political Systems and Growth. PhD Thesis, LSE, London.

____. 2002. Ethnicity, Political Systems, and Civil Wars. *Journal of Conflict Resolution* 46 (1): 29-54.

Slantchev, B.L. 2003. The Principle of Convergence in Wartime Negotiations. *American Political Science Review* 97 (4): 621-32.

____. 2004. How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace. *American Journal of Political Science* 48 (4): 813-29.

Smith., A., and A. Stam. 2004. Bargaining and the Nature of war. *Journal of Conflict Resolution* 48 (6): 783-813.

Taylor, C., and M. C. Hudson. 1972. *The World Handbook of Political and Social Indicators*. 2nd ed. New Haven, CT: Yale University Press.

Vigdor, J. L. 2002. Interpreting ethnic fragmentation effects. *Economic Letters* 75: 271-6.

Wolfson, M. C. 1994. When inequalities diverge. *American Economic Review*, Papers and Proceedings 84 (2): 353-8.

Table 1: Determinants of duration of Civil War using Uppsala/PRIO25 definition. Weibull

	(1)	(2)	(3)	(4)	(5)	(6)
Initial Lngdp	1.25	1.17	1.06	0.97	0.83	1.00
	(1.22)	(0.81)	(0.33)	(-0.13)	(-0.89)	(0.00)
Initial Lpopl	0.67	0.65	0.66	0.63	0.67	0.62
	(-4.97)	(-4.75)	(-4.77)	(-5.59)	(-4.39)	(-5.42)
Oil		1.259	1.65	2.07	1.97	2.05
		(0.76)	(1.76)	(2.39)	(2.06)	(2.35)
Mount			1.003	1.003	1.00	1.004
			(0.73)	(0.65)	(0.00)	(0.75)
Ncontig			0.506	0.58	0.62	0.58
			(-1.35)	(-1.13)	(-0.97)	(-1.15)
Democ				1.32	1.25	1.33
				(1.12)	(0.88)	(1.13)
Ethnic	0.211	0.202	0.18	0.188		0.165
Polarization	(-2.95)	(-3.19)	(-3.15)	(-3.14)		(-2.74)
Ethnic					0.504	1.199
Fractionalization					(-1.50)	(0.38)
Alpha	0.85	0.86	0.88	0.902	0.843	0.905
N	98	98	98	89	89	89

Coefficients are hazard ratios.

* z-statistics (between parenthesis) calculated using a robust and cluster-adjusted estimator.

Table 2: Marginal effects

	Mean		Median	
	Marginal	z-stat	Marginal	z-stat
Initial Lngdp	0.20	(0.13)	0.12	(0.13)
Initial Lpop	3.46	(4.27)	2.19	(4.08)
Oil	-5.28	(2.29)	-3.35	(-2.21)
Mount	-0.02	(-0.66)	-0.01	(-0.65)
Ncontig	5.00	(0.89)	3.17	(0.89)
Democ	-1.97	(-1.16)	-1.25	(-1.15)
Ethnic Polarization	12.51	(2.55)	7.93	(2.50)

Table 3: Robustness to regional effects. Weibull specification

	(1)	(2)	(3)
Initial Lngdp	0.90	0.79	0.89
	(-0.40)	(-0.79)	(-0.42)
Initial Lpop	0.65	0.70	0.644
	(-4.57)	(-2.95)	(-4.39)
Oil	2.53	2.16	2.54
	(2.61)	(2.09)	(2.65)
Mount	1.00	0.99	1.00
	(0.76)	(-0.06)	(0.85)
Ncontig	0.30	0.41	0.29
	(-1.41)	(-0.91)	(-1.43)
Democ	1.37	1.28	1.39
	(1.32)	(0.97)	(1.35)
Ethnic Polarization	0.14		0.123
	(-2.77)		(-2.65)
Ethnic		0.43	1.25
Fractionalization		(-1.42)	(0.41)
Regional dummies	Yes	Yes	Yes
Alpha	0.91	0.849	0.92
N	89	89	89

Coefficients are hazard ratios.

* z-statistics (between parenthesis) calculated using a robust and cluster-adjusted estimator.

Table 4: Robustness to other definitions of civil wars: Uppsala/PRIOCW definition. Weibull

	(1)	(2)	(3)	(4)	(5)	(6)
Initial Lngdp	1.04	0.98	0.85	0.95	0.658	0.73
	(0.18)	(-0.06)	(-0.64)	(-0.15)	(-0.85)	(-0.70)
Initial Lpop	0.77	0.75	0.798	0.64	0.858	0.638
	(-1.55)	(-1.54)	(-0.96)	(-1.45)	(-0.56)	(-1.72)
Oil		1.28	1.93	1.91	1.60	2.12
		(0.61)	(1.39)	(1.31)	(1.01)	(1.58)
Mount			1.004	1.005	0.99	1.00
			(0.55)	(0.66)	(-0.95)	(0.07)
Ncontig			0.25	0.339	0.44	0.39
			(-2.65)	(-1.23)	(-0.74)	(-1.16)
Democ				1.199	1.14	1.36
				(0.38)	(0.23)	(0.54)
Ethnic	0.14	0.127	0.074	0.03		0.04
Polarization	(-2.17)	(-2.10)	(-2.11)	(-2.35)		(-2.62)
Ethnic					0.14	0.27
Fractionalization					(-1.22)	(-1.18)
Alpha	1.16	1.17	1.25	1.29	1.17	1.33
N	49	49	49	43	43	43

Coefficients are hazard ratios.

* z-statistics (between parenthesis) calculated using a cluster-adjusted estimator.

Table 5: Determinants of duration of Civil War. Multiple failure time.

Cox model using Andersen and Gill specification

	(1)	(2)	(3)	(4)	(5)	(6)
Initial Lngdp	1.47	1.48	1.35	1.21	1.04	1.30
	(2.44)	(2.09)	(1.70)	(1.02)	(0.21)	(1.31)
Initial Lpop	0.63	0.63	0.66	0.66	0.66	0.63
	(-4.50)	(-4.34)	(-3.92)	(-4.12)	(-3.45)	(-3.89)
Oil		0.97	1.12	1.436	1.51	1.46
		(-0.08)	(0.38)	(1.10)	(1.20)	(1.17)
Mount			0.99	0.99	0.99	0.99
			(-1.39)	(-1.34)	(-1.46)	(-0.86)
Ncontig			0.61	0.746	0.79	0.737
			(-1.19)	(-0.74)	(-0.59)	(-0.77)
Democ				0.85	0.838	0.86
				(-0.60)	(-0.66)	(-0.54)
Ethnic	0.21	0.21	0.21	0.25		0.18
Polarization	(-2.61)	(-2.59)	(-2.58)	(-2.24)		(-2.35)
Ethnic					0.66	1.65
Fractionalization					(-0.71)	(0.83)
Wald	29.71	29.81	32.71	34.35	26.93	36.31
P>chi2	0.00	0.00	0.00	0.00	0.00	0.00
N	98	98	98	89	89	89

* z-statistics (between parenthesis) calculated using a robust and cluster-adjusted estimator.

Table 6. Test of the proportional hazard model

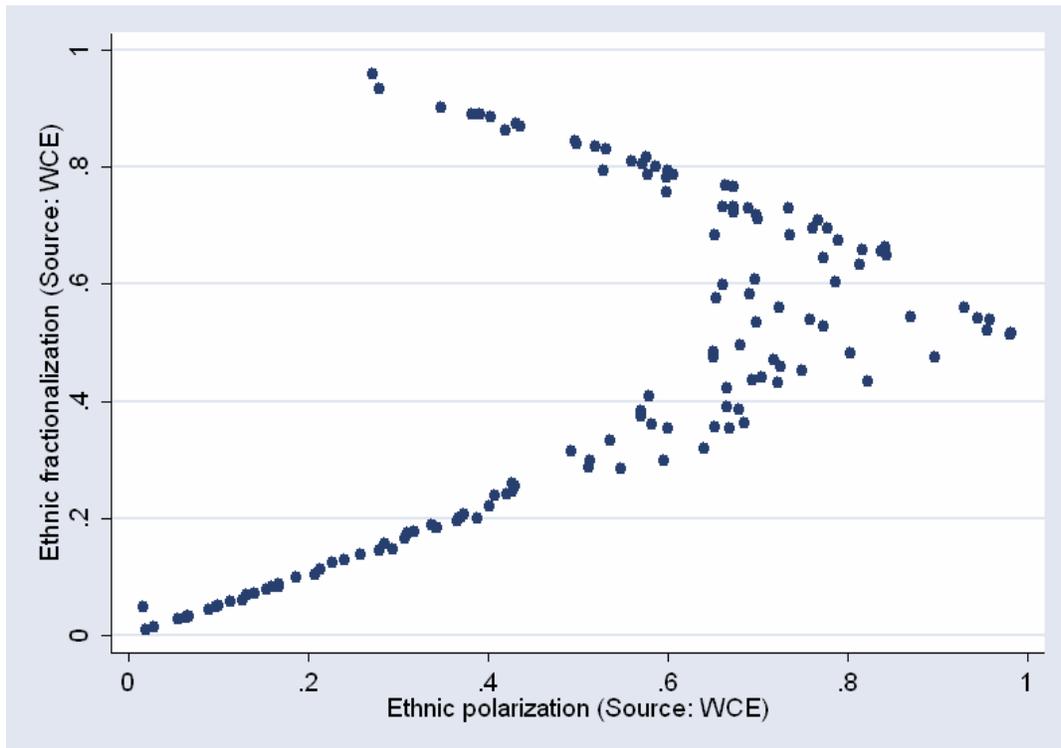|	ρ	χ^2	DF	Prob>χ^2
Initial Lngdp	0.23	4.66	1	0.03
Initial Lpop	0.10	0.84	1	0.36
Oil	-0.11	0.98	1	0.32
Mount	0.04	0.12	1	0.72
Ncontig	0.09	0.68	1	0.40
Democ	-0.10	0.85	1	0.35
Ethnic Polarization	0.02	0.03	1	0.86
Global test		5.93	5	0.54

Figure 1. Ethnic polarization versus ethnic fractionalization

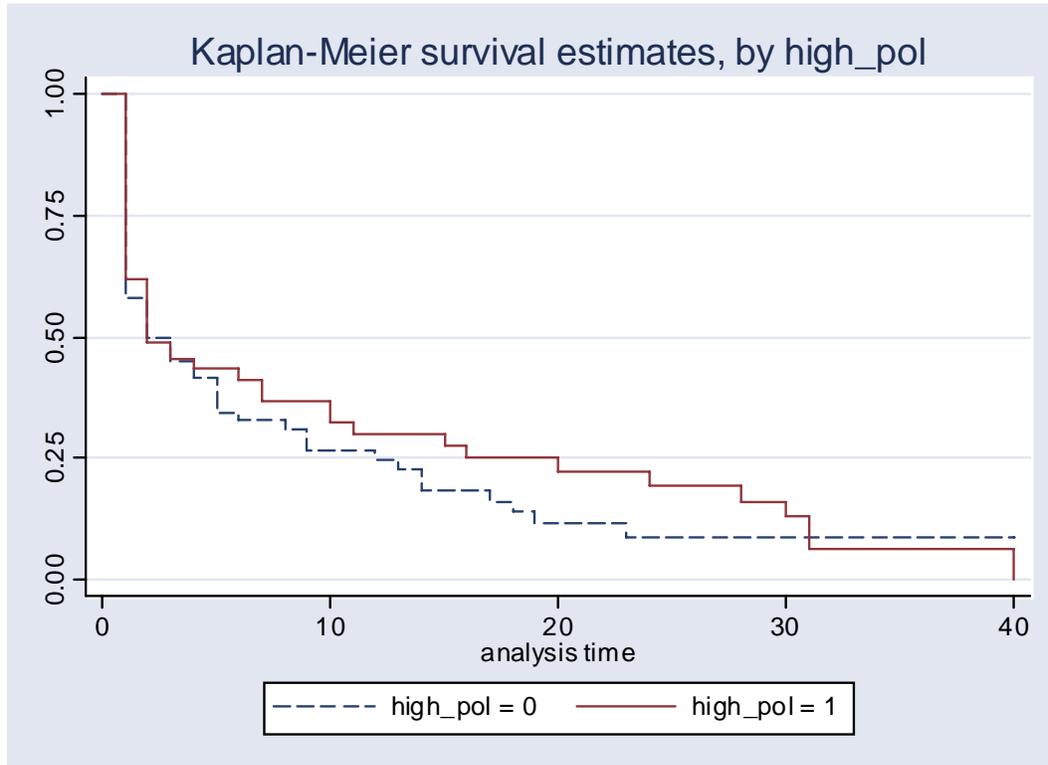*Figure 2. Kaplan-Meier survival function estimation for high and low polarization*
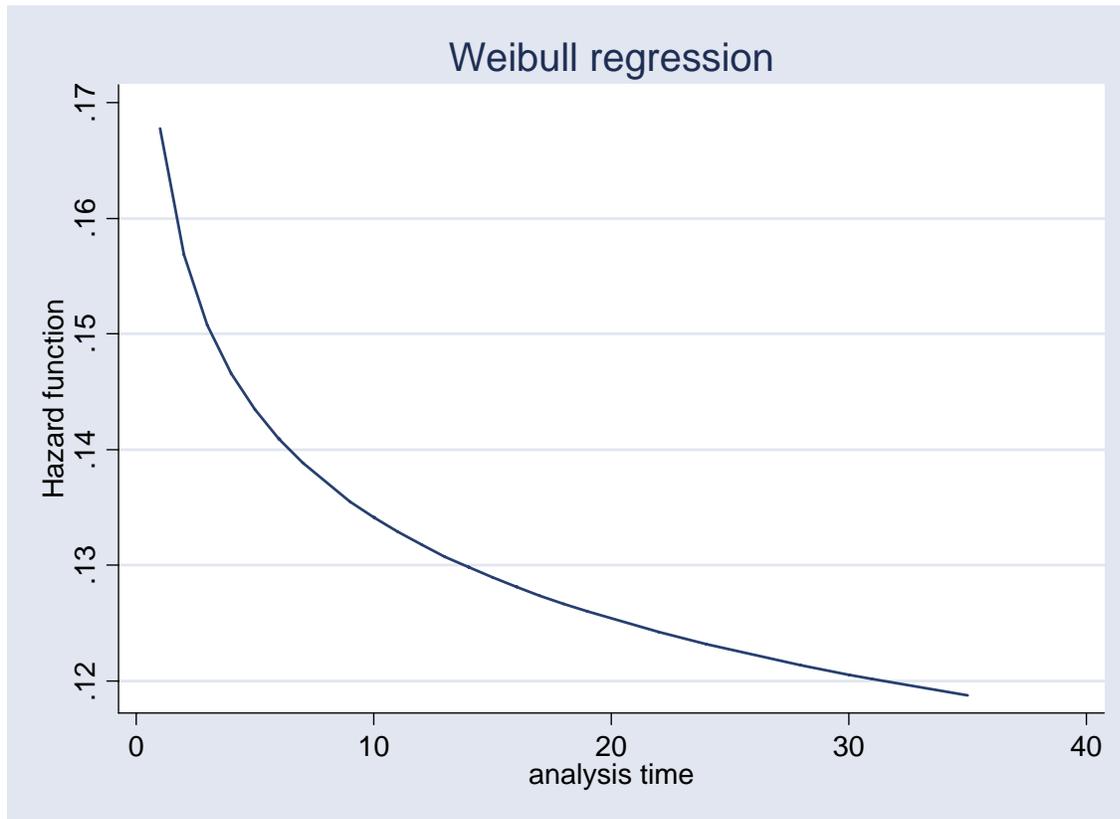
Figure 3. Hazard function

Weibull regression

Figure 4. Hazard function for high versus low polarization countries