Ethnic Polarization and the Duration of Civil Wars[1]

Jose G. Montalvo
IVIE and UPF
Marta Reynal-Querol
The World Bank

July 2005

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 the uncertainty is directly related with the relative size of the contenders. We argue that the duration of civil wars increases the more polarized is a society. 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 suffer longer civil wars.

[1] We thank the participants in the Workshop on Polarization and Conflict (PAC Summer Meeting) in Konstanz (June 2004) for many helpful comments and suggestions. This paper is part of the Polarization and Conflict Project CIT-2-CT-2004-506084 funded by the European Commission-DG Research 6th Framework program.

1. Introduction.

Only recently have civil wars being recognized as one of the main impediments for economic development. Their effects are not only related with the destruction of infrastructure or human life but also with 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 keep countries in a low income per capita equilibrium.

The study of the causes of civil wars has generated an incipient literature in recent years. The analysis has taken three different approaches: the onset of civil wars, their incidence and their duration. Obviously these three analysis 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 could be propagated thought many alternative mechanisms. One of them, for instance, could be the time to build 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 a random act that trigger, given a particular propagation mechanism, the full fledge conflict. However some of the elements that explain the onset of a civil war may be embedded also in the propagation mechanism. A country with a high proportion of mountains or jungle in its territory may have a higher probability of a rebel group forming in an area of difficult access but, at the same time, the mountains help rebels to hide and prolong the duration of the war given all the other factors. Income per capita may have a similar effect: low income per capita reduces the opportunity cost of a war which implies that the onset of a war is more likely but also that the duration will most likely be longer than otherwise. Other factors may be more easily identified. A government with a low democratic level may generate a conflict with some victimized rebel group but it is less likely that the democratic level previous to the beginning of the civil war may influence its duration.

We consider that ethnic polarization is one important dimension of the propagation mechanism once a civil war has started. Therefore, in this paper we investigate the influence of ethnic polarization on the duration of civil wars.

2. A short guide to the recent literature.

The 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[2], diaspora and higher and more disperse population increase the risk of a conflict. By the contrary, higher income per capita, time since the last conflict, growth and secondary enrolment 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[3]. In their analysis higher proportion of mountains in the country, being an oil exporter or an anocracy or having a higher population imply an increase in the risk of conflict. The only factor that decrease the risk is a high level of income per capita. Fearon and Laitin (2004) find no role for ethnic heterogeneity or democracy as a risk factor for civil wars.

Montalvo and Reynal-Querol (2005) 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 income per capita seem to be dependent on a particular specification. In their analysis neither primary exports nor democracy nor ethnic fractionalization had any significant effect on the incidence of civil wars. Since the concept of incidence is more related with the duration of a civil war than with its outbreak there seems reasonable to hypothesize that ethnic polarization may have an effect on the duration of civil wars.

The most recent papers on the duration of civil wars are compiled in a monographic issue of the Journal of Peace Research published in 2004. Fearon (2004) first classifies the wars into different categories (coups and popular revolutions, post 91 civil wars in Eastern Europe, decolonization wars, "son of soil" wars and contraband financing) and then 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). Fearon (2004) finds that, controlling for the category of the war the other

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

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

variables (population, GDP per capita, ethnic fractionalization and democracy) have no effect on duration. Obviously, the fact that in the classification of the wars its duration is one of the "informal" determinants implies a clear bias in the estimation 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 definition of COW, monthly data and an piecewise exponential duration model, find that income inequality and population have a positive effect on the duration of civil wars. However, they also find that ethnic fractionalization have also 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 the duration is the initial level of income per capita.

Our empirical approach lies in between Gates and Strand (2004), since we are going to use the definition of civil wars of PRIO, and Collier et al (2004) because we use the standard explanatory variables plus ethnic diversity proxies (we use ethnic polarization). In the rest of the paper we are going to investigate the explanatory power of ethnic polarization on the duration of civil wars. Section 3 presents a discussion on the measurement of ethnic polarization. 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. Ethnic polarization: concept and measurement.

Recently economists have connected ethnic diversity with important economic phenomena like investment (Mauro 1995), growth (Easterly and Levine 1997) or the quality of government (La Porta et al. 1999). 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[4].

In this respect is common place in recent work to include as a regressor in empirical growth estimations 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 refer 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 (1974), 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 profess religion i (or belongs to ethnic group i). The broad popularity of the ELF index is based on its simple interpretation as the probability that two randomly selected individuals from a given country will not belong to the same ethnolinguistic group. The fact that it could be used without the need to start it from scratch also helped its popularity.

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 has 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 have any statistically significant effect on the probability of civil wars. 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 groups there are the higher is the probability of a conflict. Many authors would dispute such an argument.

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

The issue of how to construct an index which is appropriate to measure the relevant aspects of ethnic diversity is the basic issue discussed in Montalvo and Reynal-Querol (2005). 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, 0,01) while in the second country, B, is (0,33, 0,33, 0,34). Which country will have a higher probability of social conflicts and, therefore, less growth?. Using the index of fractionalization the answer is B. However, Montalvo and Reynal-Querol (2002, and 2005) and Reynal-Querol (2002a) have argued that the answer is A. They use the index of ethnic polarization RQ, originally constructed by Reynal-Querol (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. In the context of income the measurement of polarization was initiated by Esteban and Ray (1994) and Wolfson (1994). Montalvo and Reynal Querol (2005) show the connection between the measure of ER income and this measure of ethnic polarization.

We argue that any index of ethnic heterogeneity should be theoretically oriented. It should accommodate the interpretations and mechanism that different authors have proposed in the explanation of the effect of ethnic diversity on, for instance, growth. Montalvo and Reynal-Querol (2002, 2005) show that the index of fractionalization is at odds with the basic explanations and, therefore, cannot capture the relevant dimensions of ethnic divisions. The main reason why the index has been widely use is it simple interpretation as the probability of being matched with an individual of a different ethnic group. However in the context of conflict and rent seeking models this measure is not a relevant indicator of the intensity of the conflict while the RQ indicator can be easily justified. Horowitz (1985), which is a seminal reference on the issue 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 so then an index of polarization should capture better the likelihood of conflicts, or the intensity of potential conflict, than an index of fractionalization.

The original data for the construction of ELF come from the Atlas Narodov Mira.

Posner (1999) identifies two general kinds of problem with the ELF index: those that stem from the underlying ethnographic data from which its is calculated, and those that arise from attempting to summarize a country's ethnic diversity with a single measure. The first problem with the ethnographic data that Posner (1999) identifies is what he so called "grouping problem". He observes that the ELF measure sometimes subsume groups that in most ethnographic and political accounts are distinct, and occasionally highly antagonistic. The second data problem is the so called "problem of inclusion". It is caused by the enumeration of dozens of groups that may be culturally or linguistically distinct but are irrelevant as political actors in their own right. We are not so sure that this second criticism is that important. The political relevant ethnic groups depend on the particular political systems of the country. The disaggregation of ethnic groups in terms of their political relevance does not allow the identification of two different effects: the impact of the original social structure, and the impact of the political system.
The World Christian Encyclopedia (WCE) contains also an ethnolinguistic classification that is neither purely "ethnic", nor "racial" nor "linguistic" nor "cultural", but ethnolinguistic. Their definition of "ethnolinguistic people" implies that an ethnic, ethno-cultural or racial group speaks its own language or mother tongue[5]. The term ethnolingusitic reefers to that group which speaks the language shown as its first, primary, cultural or official language.
In a few cases, it is difficult to be consistent in the classification on a global scale because a particular class has ethnic, cultural and religious meanings, all closely related, but applied to different countries and with different emphases. The main criteria adopted by the WCE in such ambiguous situations is their answer to the question: "What is the first, or main, or primary ethnic or ethnolinguistic term by which persons identify themselves, or are identified by peoples around them?". The data was in all cases worked out "de novo" for the latest survey, utilizing population census and a host of other sources.
Another source is the Encyclopedia Britannica (EB), in particular the Britannica World Data (BWD). What are the differences between this data set and the data provided by the WCE? Both data sources provide disaggregate information on ethnic groups, but many times, mainly in Africa, the groups differ from one source to another. The WCE

[5]The ethnic groups in Africa have an additional grouping problem. Many times, ethnic tribes have a common language origin and we can group them following this common origin: bantu, hamites, Nilotic, Pygmies, etc.

provides a deep discussion and explanation for the ethnolinguistic classification. The EB uses a classification for the ethnic composition that divides peoples into ethnic culture-areas. This means that many ethnic groups not based on cultural-areas would not be considered independently in the EB.

Another source of data is the World Factbook (WF) that is totally based on national sources. This book provides information on the religious, ethnic, and language groups of the countries. For the cases of language groups only in very few cases we have percentages of the people speaking the major languages. For the cases of ethnic groups, the World Factbook provides information on the percentage of ethnic groups. However the groups are less disaggregated than in the WCE, and many times the group "others" is quite large. We do not know whether "others" includes many small groups or one big group. This can be problematic because indices of polarization and fractionalization are very sensitive to the size of the groups. Therefore if a small number of groups are measured as one big group the indices will not capture the real ethnic diversity of the country.

Another source of information on ethnic diversity is the Minorities at Risk project (MAR). This database, developed by Ted Gurr, analyzes the status and conflicts of 268 politically active communal groups in 148 different countries. Among the 449 original variables included in the data set, there are assessments of cultural, economic, and political differences between minority and dominant groups, of group grievances and organizational strength, of trans-national support of minority goals, of polity characteristics, and protest, communal violence and rebellion. The MAR database is the first large database on minority groups in conflict with their neighbours and the forces of the state. However this database do not provide information on the percentage of all language groups.

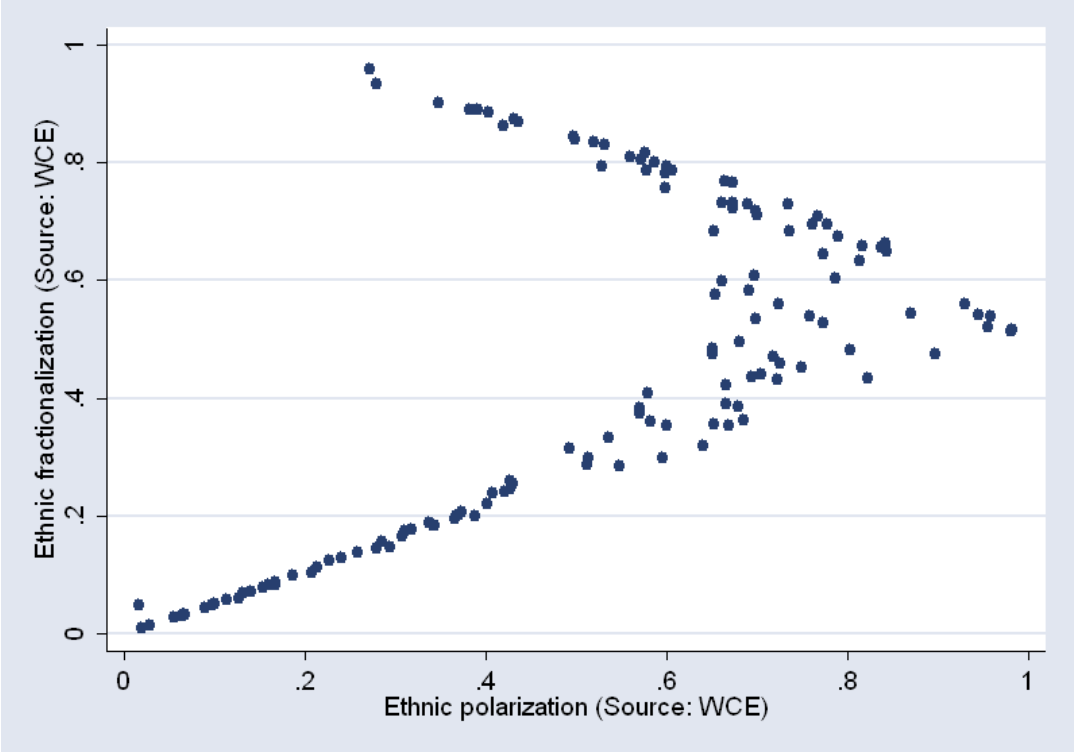
Vanhaven (1999) constructs an ethnic data base for 183 countries around 1990 that captures the major ethnic divisions based on three dimension of ethnicity: race, nationality-tribe language, and old religious communities. Sometimes, the information classified as national/language includes ethnic groups formed as a consequence of immigration. In terms of dealing with empirical issues Vanhaven's data present several shortcomings. First of all, even though he reports the percentage of the biggest ethnic group, it does not provide data on the size of other small groups, which makes impossible the construction of fractionalization and polarization indices. Secondly there

is no distinction between Christians and animist cult followers in the religious dimension. This is problematic in many African countries.

Montalvo and Reynal-Querol (2002, and 2005) use as their basic source the WCE since this is the most detailed primary source based on national sources. Montalvo and Reynal-Querol (2002, 2005) follow Vanhaven (1999) in taking into account only the most important ethnic divisions and not all the possible ethnic differences or groups. Vanhaven (1999) uses a measure of genetic distance to separate different degrees of ethnic cleavages. The proxy for genetic distance is "the period of time that two or more compared groups have been separated from each other, in the sense that inter-group marriage has been very rare. The longer the period of endogamous separation the more groups have had time to differentiate".

Figure 0 shows that the choice of indicator matters very much for the application to data on countries. As shown in figure 0 ethnic polarization and ethnic fractionalization have 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.

Figure 0. Ethnic polarization versus ethnic fractionalization.



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 is under way there are partial "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 is the degree of polarization the longer is the duration of the war.

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

This interpretation is different from the rebellion-as-mistake in Collier et al (2004). Under that conceptualization the participants in the 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 along these lines. In this case it is 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 country (maximum degree of polarization), one in power and the other starting the rebellion. Since potentially they

[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.

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 has a close connection with the recent literature on wars as bargaining processes as Smith and Stam (2003), Ponsati et al. (2005), Powell (2004), Slantchev (2003 and 2004) among many others[8]. All these papers usually refer to the case of wars between nations. However, these models are very general and could easily be applied to conflict among different groups in a country. 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 information is revealed by strategically manipulable negotiation behavior and nonmanipulative battlefield outcomes. He 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. Total victory is a possible outcome but an important function of wars is to convince the opponent to accept a settlement. This happens when the contenders learn enough about their prospects as to 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 value of difference in army (population) sizes scaled between 0 and 1. Slantchev (2004) shows that the higher the uncertainty (closeness to parity) the longer is the war.

Smith and Stam (2003) explain why countries enter into war in the following way :"The act of waging war revels 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 nation's (in the case of civil wars, instead of nations we have social groups) expectations of the duration of

[8] For an extensive review of this literature, see Powell (2002). We thank G. Scheneider for calling our attention about this strand of the theoretical literature.

conflict and which nation (group) is 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 common information about the nature of warfare between them. And as more and more information is reveled, the nations' (groups') beliefs converge. Their model shows how war resolves differences between states's beliefs.

What role does social polarization play in this kind of processes? 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 their size in the potential fighters is larger than it actually is. After the conflict starts and a battle is fought some additional information is released providing a large reduction in the variance of the original measurement error. Therefore, the minority group recalculate 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 with a small amount of noise. In this case the convergence of believes will imply more rounds of "battles" (release of information) since even a small amount of measurement error makes their probability of winning statistically equal. When the probabilities of winning are in fact very similar, given a 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 equals, are usually longer than the ones in less polarized countries.

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 (generates also 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 of several conflicts in the same country and 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 a similar effect to an unobservable variable which is outcome specific. DeRouen and Sobek (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[9]. The fact that the same country may experience several civil wars during the period of analysis implies what is know in the duration jargon as multiple failure-time data. This mimics the panel data

[9] 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 income per capita had a significant effect on the probability of the onset of a civil war.

versus pooled data situation in the linear regression case. The existence of an unobservable country specific effect may cause the violation of 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 maybe important if we deal with multiple failure time data.

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

The arguments exposed above indicate that ethnic polarization maybe a powerful propagation mechanism for civil wars and, therefore, had an important influence in the duration of civil wars. We are going to use initially the definition of civil wars of PRIO25 (see Appendix for precise definition). In short this definition includes wars that cause at least 25 yearly deaths. Our database includes 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 up. Following this definition we have 42 countries with only 1 civil war; 24 countries did have 2 civil wars; 5 had 3 wars and 3 had 4 wars. The mean duration is 7.15 years. In 1999 there were 20 civil wars still ongoing in the sample.

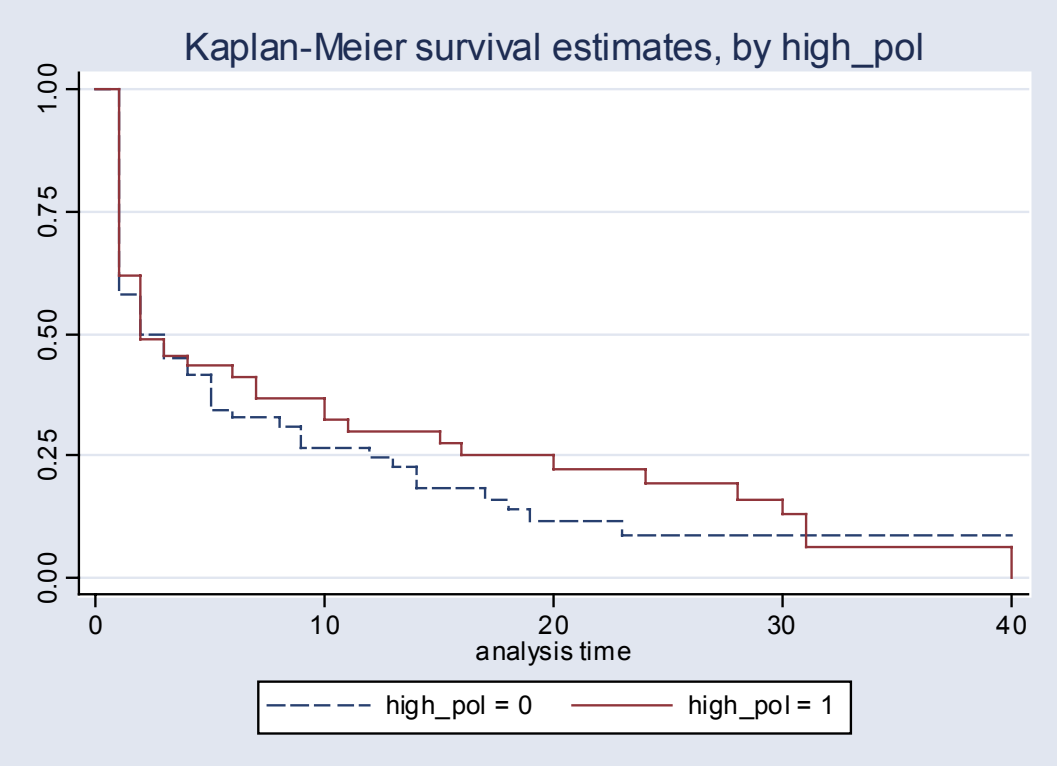
As a preliminary indication we can look at the survivor functions of civil wars in highly polarized countries versus countries with lower level of ethnic polarization. To separate these two groups we use the median value of ethnic polarization[10]. Figure 1 shows the

[10] A previous section explain the data and methods for the construction of the ethnic polarization measure. The appendix describes the data on civil wars and the explanatory variables used in the duration models of this section.

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 is simply a preliminary indication since we have not include any covariate. The Kaplan-Meier estimator is a nonparametric procedure, alternative to the life tables, to obtain survivor functions.

Figure 1 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 lengthier in those countries.

Figure 1. Kaplan-Meier survival function estimation for high and low polarization.



There are several alternatives to deal with the analysis of duration in a parametric set-up. The most important fact 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 (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

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

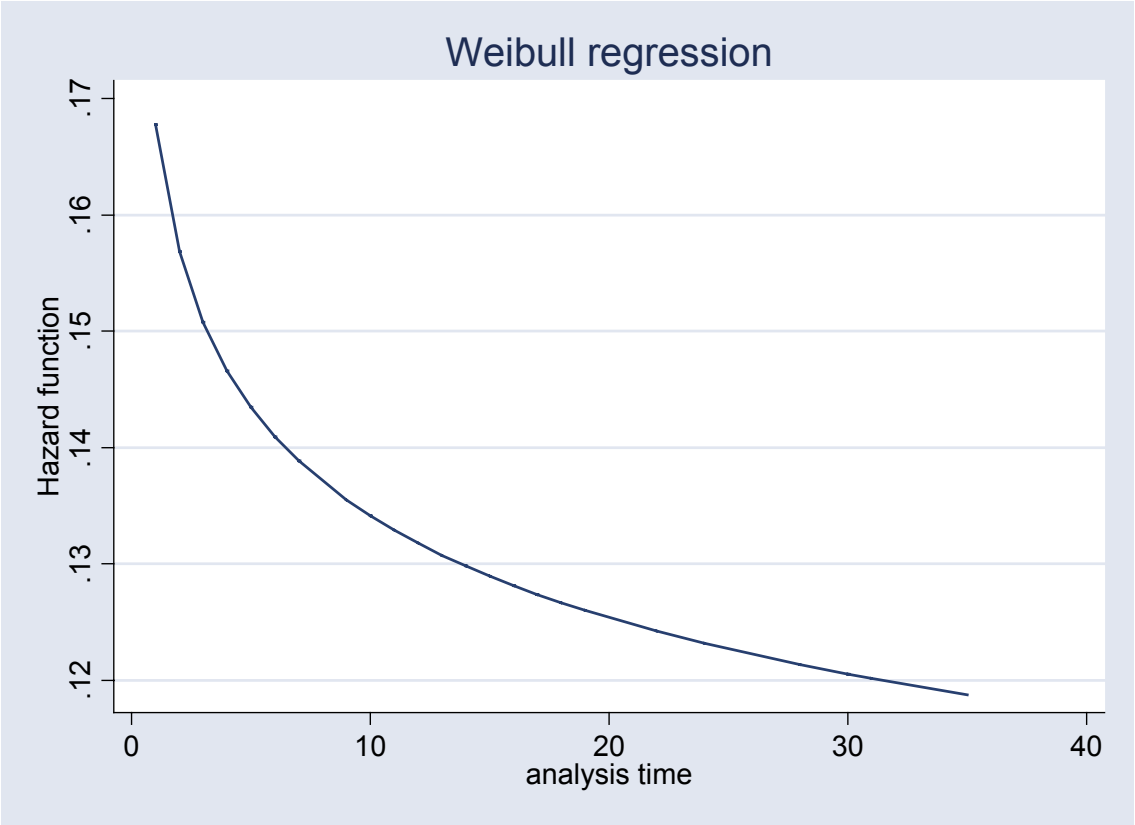
where the error terms is supposed to have and extreme value density function.

Table 1 shows the estimation under different sets of explanatory variables and a Weibull specification. Table 1 shows 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 2 shows the hazard function derived from the specification 6, which is very similar to the one obtained from the previous columns. It shows that for the definition of war of PRIO25 the probability of the war ending conditional on having lasted up to t is decreasing on time.

Figure 2. Hazard function.

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

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



More interesting for our purposes it to look at the hazard function for highly polarized countries versus countries with low polarization. Figure 3 shows the hazard function obtained using a Weibull model where instead of 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 3 shows that the hazard function for the end of a civil war is much lower for highly polarized countries.

Figure 3. Hazard function for high versus low polarization countries.

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

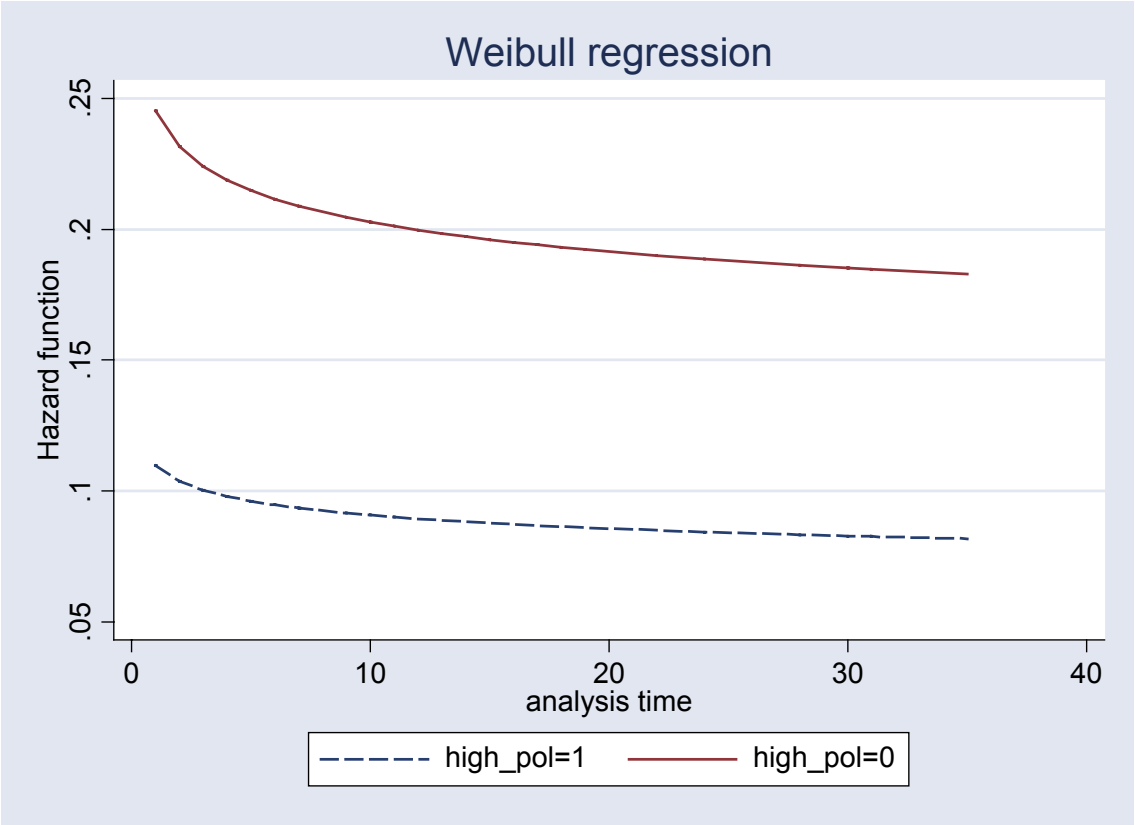


Table 2 shows the marginal effect of the explanatory variables on mean and median duration obtained from the accelerated failure version of the model. Since the inferences on the median are more robust we are going to concentrate the comments on column 2 and, in particular, in the effect of ethnic polarization. An increase of this variable from 0 to 1 increases the duration of a war by 7.9 years!!! Obviously this is not a realistic change since it covers the whole range of the variable, but it gives a clear indication of the importance of ethnic polarization in 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 (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 upward sloping for the estimation of the most intense wars while it is downward sloping 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 parametrize 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 is 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 zero slope, which implies 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 regression support the previous findings and are not caused by the failure of the proportionality assumption maintained in the estimation of the model.

6. 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 with the relative size of the contenders. We argue that another link between the degree of polarization of the

[14] Following Grambsch and Therneau (1994).

[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.

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 shows that the hazard function for the event of the end of a civil war is much lower for all the time range 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.

TABLES

Table 1: Determinants of duration of Civil War using 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 Fractionalization		0.43	1.25
		(-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: 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

APPENDIX:

Definitions of civil wars from PRIO.

PRIO25: PRIO definition including armed conflicts that generate more than 25 deaths yearly (minor armed conflicts plus intermediate plus war following PRIO classification). We only consider types 3 and 4 (internal armed conflicts).
PRIOCW: Intermediate and war definition of armed conflict from 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), "Is poverty to blame for civil war? New evidence from nonlinear fixed effects estimation," mimeo, Harvard University.

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

Collier, P., A. Hoeffler and M. Söderbom (2004), "On the duration of civil war," *Journal of Peace Research*, 41(3), 253-273.

Collier, P., and A. Hoeffler (2004) "Greed and Grievances in Civil Wars." Oxford Bulletin of Economic and Statistics.

_____ and _____ (1998) "On Economic Causes of Civil Wars." Oxford Economic Papers 50: 563-73.

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.

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

Doyle, Michael W., and Nicholas Sambanis (2000) "International Peacebuilding: A Theoretical and Quantitative Analysis." American Political Science Review 94:4 (December).

Esteban and Ray (1994). "On the measurement of Polarization". Econometrica, 62(4), pp.819-415.

Fearon, J. (2005), "Primary commodity exports and civil wars," forthcoming in the Journal of Conflict Resolution.

Fearon J. and Laitin, D. (2003). "Ethnicity, Insurgency, and Civil War", *American Political Science Review* 97 (February)

Grambsch, P. and T. Therneau (1994), "Proportional hazards test and diagnostic based on weighted residuals," Biometrika, 81, 515-526.

Hegre, H. (2004), "The duration and termination of civil wars," *Journal of Peace Research*, 41(3), 243-252.

Miguel, E., Satyanath, S. and E. Sergenti (2004), "Economic Shocks and Civil Conflicts: an Instrumental Variables Approach," *Journal of Political Economy*.

Montalvo, J. G. and M. Reynal-Querol (2005), "Ethnic polarization, potential conflict and civil wars," *American Economic Review*.

Ponsati, C. Copic, J. (2005) “Mediated bargaining”. Mimeo

Powell, R. (2004) “The Inefficient Use of Power: Costly Conflict with Complete Information”, *American Political Science Review*, Vol.98, No.2. Pag: 231-241.

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

Reynal-Querol, M. (2002) "Ethnicity, Political Systems, and Civil Wars." *Journal of Conflict Resolution* 46(1, February): 29-54.

Slantchev, B.L (2003) “The Principle of Convergente in Wartime Negociations”, *American Political Science Review* Vol.97, No.4 November, Pag: 621-632

_____(2004) “How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace”, *American Journal of Political Science*, Vol.48, No.4, October 2004, Pag: 813-829.

Smith., A., and Stam, A. (2003) “Bargaining and the Nature of war”, *Journal of Conflict Resolution*” Vol. XX, No.X pag. 1-30

Wolfson, Michael C. (1994) “When inequalities diverge”. *American Economic Review* 84(2), Papers and Procedings, pp. 353-358.