Methodological Issues in Studies of Conflict Processes Misclassifications and Endogenous Institutions*

Thomas Christin† and Simon Hug‡
IPW, Universität St.Gallen and Universität Zürich

Paper prepared for presentation at the
ISA conference March 1-5, 2005, Hawaii
and at the IGCC conference
"Disaggregating the Study of Civil War and Transnational Violence"
San Diego, CA, USA, 7-8 March 2005

Abstract

Empirical work on civil wars and "minorities at risk" is plagued by various intriguing methodological issues. Political institutions thought to mitigate conflict are often not exogenous but nevertheless are treated as such in empirical analyses. Case selection often influences empirical results considerably, especially in nonlinear models, where systematic measurement error in the dependent variable has more complicated consequences than in linear models. In this paper we discuss these issues and illustrate their consequences by reanalyzing several datasets on civil wars. We demonstrate how dealing with the endogenous nature of institutions and the problems of case selection leads to substantive different conclusions for many research questions on civil wars and "minorities at risk."

* This paper builds on research funded by the Swiss National Science Foundation (Grant No. 5004-0487882/1). A previous version of this paper was presented at the Annual Meeting of the American Political Science Association Chicago September 2-5, 2004. Helpful comments from participants at this event and additional stimulating discussions with Frederick Boehmke, Lars-Erik Cederman, and Scott Gates are gratefully acknowledged. Thanks are also due to James Fearon who gave advice on how to update the dataset used in Fearon and Laitin (2003).

† Institut für Politikwissenschaft; Universität St. Gallen; Dufourstrasse 45; 9000 St. Gallen; Switzerland; phone +41 (0)71 224 2600; fax: +41 (0)71 224 2974; email: thomas.christin@unisg.ch

‡ Institut für Politikwissenschaft; Universität Zürich; Hirschengraben 56; 8001 Zürich; Switzerland; phone +41 (0)44 634 50 90/1; fax: +41 (0)71 224 2974; email: hug@pwi.unizh.ch

1 Introduction

Most of the human toll in current conflicts stems from intrastate wars (e.g., Lacina and Gleditsch, 2004; Lacina, 2004). Thus, it hardly surprises that a considerable research effort is devoted to understanding the processes which lead to civil wars or societal conflicts (e.g., Gurr, 1993; Hegre, Ellingsen, Gates and Gleditsch, 2001; Kalyvas, 2002; Collier, Elliott, Hegre, Hoeffler, Reynal-Querol and Sambanis, 2003; Fearon and Laitin, 2003; Fearon, 2004; Sambanis, 2004). Much of the recent progress has resulted from systematic quantitative studies yielding important insights. Work by Fearon and Laitin (2003) suggests that state strength, measured as GDP per capita, influences strongly the onset of civil war, while ethnic and religious fragmentation plays no significant role. Collier, Elliott, Hegre, Hoeffler, Reynal-Querol and Sambanis (2003) report in their work, that "greed" related to the presence of "lootable goods" is of considerable importance in explaining civil wars.[1] Hegre, Ellingsen, Gates and Gleditsch (2001) find an interesting u-shaped relationship between the degree of democracy and the outbreak of civil conflict as do Mansfield and Snyder (1995). Reynal-Querol (2002*a*) reports that proportional representation reduces the likelihood of civil war, while Cohen (1997) shows that federalism makes violent rebellion less likely.[2]

Recent critics have raised questions concerning the soundness of some of these findings. Quinn, Hechter and Wibbels (2003) find that contrary to Fearon and Laitin's (2003) findings ethnic fragmentation all the same influences civil war onset. This results from addressing the issue that at least one of Fearon and Laitin's (2003) independent variable, namely the degree of democracy, is affected by measurement error.[3] Christin and Hug (2003) suggest that the effect of federalism on violent rebellions is hardly a robust finding, since selection effects affect results of empirical studies. Sambanis (2004) argues that many of the mechanisms supposed to link contextual variables to conflicts are found not to operate in case studies. Similar arguments appear in Kalyvas's (2002) painstaking work on the Greek civil war. Cederman (2004) argues that some of the empirical results of

[1] This finding is criticized by Snyder (2001) and Van de Walle (2004).

[2] An excellent survey of the literature and the most important findings appears in Sambanis (2002).

[3] In a revised version of their paper Quinn, Hechter and Wibbels (2003) suggest that taking into account the indirect effects of ethnic fragmentation shows a substantial total effect of this variable in Fearon and Laitin's (2003) dataset.

Fearon and Laitin (2003) might have been generated by processes involving quite clearly ethnic identities.

These critiques either suggest that the level of aggregation is too high to assess the mechanisms assumed to be at work or focus on some methodological problems in the quantitative studies. In this paper we wish to address methodological issues which are at the intersection of these two concerns. First, as most studies on conflict processes employ as primary unit of observation either a societal group or a civil war, much hinges on the precise measure of these elements in quantitative studies. We argue that by the very nature of conflict processes information on both societal groups and civil wars is hard to come by. Neglecting this fact affects quantitative analyses in various ways. We demonstrate how some empirical models may be improved upon by addressing this particular issue.

Second, mostly for policy purposes, empirical models often include as explanatory variables measures of political institutions,[4] ranging from the degree of democracy (e.g., Hegre, Ellingsen, Gates and Gleditsch, 2001; Saideman, Lanoue, Michael and Stanton, 2002; Fearon and Laitin, 2003),[5] electoral institutions (e.g., Cohen, 1997; Reynal-Querol, 2002*a*; Saideman, Lanoue, Michael and Stanton, 2002), all the way to federalism (e.g., Cohen, 1997; Saideman, Lanoue, Michael and Stanton, 2002). These institutions are assumed to affect conflict processes. Especially power-sharing institutions[6] like federalism (e.g., Cohen, 1997; Saideman, Lanoue, Michael and Stanton, 2002) or consociationalism (e.g., Cohen, 1997; Reynal-Querol, 2002*a*) are, however, often put into place in response to societal conflicts. Thus, in the empirical models we wish to estimate, political institutions are hardly exogenous.

In this paper we wish to address these methodological issues in the study of conflict processes. We suggest for both sets of problems ways in which they can be addressed in particular research strategies. We illustrate these proposed solutions by replicating empirical analyses reported in the literature on conflict processes. As we are able to show, addressing these methodological issues may affect substantive results obtained in the literature.

We proceed as follows. In the next section we briefly review the literature on

[4]Collier, Elliott, Hegre, Hoeffler, Reynal-Querol and Sambanis (2003, 123f) discuss this issue related to the World Bank study.

[5]Easterly (2000) also assesses the effect of "good" institutions, which might be endogenous.

[6]On the impact of power-sharing institutions on post-civil war conflict management, see Hartzell and Hoddie (2003).

conflict processes under the angle of what type of empirical information is used and at what level of aggregation the analyses are carried out. Section three is devoted to the two sets of methodological issues we wish to address, namely on the one hand the problems of endogenous variables and on the other the problem of case selection and the induced measurement errors. We then go on to propose ways in which the selection of cases and the potential resulting misclassification of cases may be addressed and illustrate this with an empirical application in section four. In section five we propose ways in which endogeneity problems of specific variables can be addressed. Again, we propose an empirical illustration for this problem, before we conclude in the last section.

2 Studying Conflict Processes

With the increasing human toll stemming from civil conflicts and the dramatic increase of countries embroiled in civil wars (e.g. Fearon, 2004; Hegre, 2004) numerous studies have focused on studying conflict processes, Ranging from societal conflicts over rebellions to civil wars and state failures. A considerable effort has been put into explaining the onset and duration of these conflicts and the ways in which these conflicts may be resolved. In both parts of the literature, quantitative empirical analyses have led to considerable insights, even though they are challenged in part by scholars engaging in case studies (e.g. Snyder, 2001; Kalyvas, 2002; Sambanis, 2004).

Quantitative studies in essence focus on either particular societal groups or civil wars. Empirical models are then devised either to explain why particular societal groups become rebellious and violent or to assess what leads to the onset of civil wars and what explains the latter's duration. Table 1 classifies some quantitative studies according to whether they study societal groups or civil wars and the unit of analysis employed.[7] Studies on societal groups like ethnic groups or "minorities at risk" (MAR) typically use data on a list of such groups and try to assess what variables affect their "rebellious nature." Some studies, instead of using the societal groups as units of analysis, aggregate this information to the level of countries. Thus, the dependent variable becomes whether a particular country is home to a violent societal group or not (possibly in a particular year).

[7]The table is adapted from Christin and Hug (2003).

Table 1: Data and approaches: Some examples

Units of analysis	Ethnic groups, "minorities at risk"	Civil wars
per se	Cohen (1997), Fearon and Laitin (1997), Saideman and Ayres (2000)	Doyle and Sambanis (2000), Collier, Hoeffler and Söderbom (2001), Walter (2002)
per country	Ellingsen (2000), Saideman, Lanoue, Michael and Stanton (2002)	Hegre, Ellingsen, Gates and Gleditsch (2001), Sambanis (2001*a*), Reynal-Querol (2002*a*), Reynal-Querol (2002*b*), Fearon and Laitin (2003)

Studies of civil wars are very similar. On the one hand scholars use civil wars as units of analysis and try to explain, for instance, their duration, their nature etc. On the other hand some studies use the information on civil wars and aggregate it to the level of countries. Here again the research question becomes whether a particular country is embroiled in a civil war or not (possibly in a particular year).

Employing these empirical strategies scholars have come up with many important and interesting insights about the causes of domestic violence and various aspects of civil wars. Many of these insights, however, are also disputed, as noted above. For instance, whether a democracy scale is related to civil wars in an u-shaped functional form is contested by several scholars. Similarly, whether other institutions have the hoped-for effects is also questioned.

3 Case Selection and Endogeneity

Some of these contradictions in the literature have, in our view, something to do with two methodological issues which are hardly addressed in quantitative studies. The first of this problem concerns case selection, while the second relates to the endogenous nature of political institutions.

Concerning case selection, it is obvious when looking at table 1 that much of the empirical relevance of quantitative studies depends on the accurate identification of societal groups or civil wars. Failing this, quantitative studies are likely to be affected by bias. The easiest way to demonstrate this relates to studies in the first row, namely those who take as their unit of analysis either societal groups or civil wars. If for one reason or another some societal groups or civil wars are missed in the data collection stage, then estimated empirical models are potentially affected by selection bias. Christin and Hug (2003) and Hug (2003), for instance, demonstrate that analyses using the "minorities at risk" (MAR)

dataset (Gurr, 1993) are likely to be affected by selection biases. These biases come about by the fact that the selection mechanisms of groups into the dataset are in some empirical applications related to the phenomenon we wish to explain. Thus, omitted variables in the outcome equation to be estimated are likely to be correlated with the selection mechanism. Neglecting this correlation may lead to serious biases and erroneous inferences in empirical studies (e.g. Christin and Hug, 2003; Hug, 2003). If some information is available on the selection mechanism, however, empirical models exist which allow for the necessary corrections. These models are discussed in detail for continuous dependent variables in Maddala (1983), Muthen and Jöreskog (1983), Bloom and Killingsworth (1985), and King (1989),[8] for dichotomous dependent variables in Sartori (2003), and for duration models in Boehmke, Morey and Shannon (2003).[9]

While the selection issue is very similar in studies of civil wars, some differences exist all the same. Almost systematically scholars employ as criteria the fact that in a country a dispute among at least two parties is going on, one of which is the government, and that there are at least 1000 battle deaths.[10] Such criteria, especially the death threshold, result in a series of important problems discussed in detail by Sambanis (2001*b*) and Gates and Strand (2004). A central problem relates to the possible (and very likely) measurement errors of the number of battle deaths in civil wars. If such measurement error is present, we are again in exactly the same situation as before with the identification of societal groups and "minorities at risk." If the measurement error is related in any possible way with variables included in the empirical model we wish to estimate, biases are very likely. If the error is unrelated to any included variables, linear models provide unbiased estimates for the slope coefficients, while the intercept might be biased. In non-linear models, however, even purely random such errors may affect our slope estimates (e.g. Hausman, Abrevaya and Scott-Morton, 1998).

Since information on battle deaths is probably much more difficult to obtain in poorer, less democratic countries, and in particular regions of the world,[11]

[8]Muthen and Jöreskog (1983) provide suggestive and Hug (2003) more systematic results from Monte-Carlo simulations for this estimator.

[9]Boehmke, Morey and Shannon (2003) also provide results from Monte-Carlo simulations.

[10]Based on the COW dataset Gleditsch, Strand, Eriksson, Sollenberg and Wallensteen (2002) discuss the problem of the death threshold and the resulting classification of wars. They suggest using a threshold of 25 deaths and on the basis of this criterion produce a new dataset of wars.

[11]Illustrative for this is at the time of writing the uncertainty over the number of victims in

biases are very likely to result. First, since many of these elements which explain measurement error also figure as independent variables in many empirical models, and second, since most empirical models employed are non-linear.

A second issue which comes more clearly to the forefront in studies of civil wars is related to the units of analysis employed. Many studies (see table 1) aggregate the information on civil wars to the level of countries. Hence, the dependent variable becomes either whether a country is embroiled in a civil war or how long it has survived without such an embroilment. If measurement error on the number of battle deaths is present, then some countries which are likely to be coded as not being engaged in a civil war are miscoded. Similarly, overestimates of battle deaths will lead to the opposite misclassification.[12] In ordinary regression analyses such measurement error hardly poses any problems as long as it is not correlated with any of the included explanatory variables. Similarly, if the measurement error is systematic, for instance a systematic underestimate of battle deaths, the only problem would appear in the estimated intercept. A quantity most often neglected in empirical studies, except in the context of predictions.

Things look different in most of the empirical models used to analyze data on civil wars. Most of these models, for instance logit, probit or duration models, are nonlinear in their nature. In such nonlinear models measurement error in the dependent variable not only affects the intercept, but the whole vector of estimated coefficients, and this even if the measurement error is uncorrelated with any of the included variables (e.g. Hausman, Abrevaya and Scott-Morton, 1998).[13] Hence, in this case, neglecting the issue of measurement error related to the selection of cases becomes even more of a problem.

Hence, case selection and the resulting measurement error in the dependent variable is likely to put into doubt empirical results in many studies. An additional problem, namely the endogenous nature of institutions, is also likely to put in jeopardy some empirical results. Many recent studies have found various relationships between political institutions violent behavior of societal groups,

the Darfur region in Sudan.

[12]Fearon's (2004, 278) admission that compared to the Fearon and Laitin's (2003) study "a few cases have been dropped or added according to the results of additional research." illustrates this problem nicely. For the analyses presented in this paper we rely on a data file with the most recent updates from Fearon (2004). In the appendix we provide more details on the update and also the results for the original dataset without the new updates).

[13]Carroll, Ruppert and Stefanski (1995) offer a detailed discussion of measurement error in nonlinear models by focusing, however, on errors in the independent variables.

and the onset as well as the duration of civil wars. For many of the institutions considered, however, it is a leap of faith to consider them as exogenous to the research question addressed by the relevant scholars. Most striking is this point with institutions like federalism (e.g., Cohen, 1997; Saideman, Lanoue, Michael and Stanton, 2002), consociationalism (e.g., Cohen, 1997; Reynal-Querol, 2002*a*), and electoral institutions (e.g., Cohen, 1997; Reynal-Querol, 2002*a*; Saideman, Lanoue, Michael and Stanton, 2002). Quite clearly, federal arrangements and particularly consociationalism have been adopted in many countries to overcome societal conflicts.[14] Thus, their presence is part and parcel of the explanation of societal conflicts and potentially civil wars. Similarly, many studies of electoral systems emphasize that the selection of electoral rules often relate to the prevalent cleavage patterns.[15] Again, the latter are clearly related with many of the factors considered to explain ongoing societal conflicts and civil wars.

Neglecting the endogeneity of particular explanatory factors can severely affect their estimated effects. Nevertheless, even though quite easy approaches exist to test for the exogeneity of particular explanatory variables, few scholars dealing with political institutions and their effects on ethnic groups and civil wars check for these problems.[16]

Hence, while case selection is likely to affect most studies of ethnic groups and civil wars, the endogeneity problem in these studies is most apparent in empirical models employing political institutions as independent variables. In what follows, we discuss ways in which these two problems can be addressed and illustrate the respective solutions with empirical examples from the literature. As examples we use two prominent studies, namely the analyses of Fearon and Laitin (2003) and Reynal-Querol (2002*b*). Fearon and Laitin's (2003) article has been very influential and has also attracted much criticism as discussed above. Reynal-Querol's (2002*b*) analyses are interesting for our purposes, since she studies the effect of various institutional features. Hence, these two studies will serve as foil of our discussion of methodological issues in the study of conflict processes.

[14]Lijphart (1999, 189f) as well as Panizza (1999) provide a discussion of factors explaining federal arrangements and decentralization.

[15]Cox (1997) discusses the relevant literature in detail.

[16]Elbadawi and Sambanis (2002) explicitly test for the endogeneity of political institutions and find some weak evidence for one indicator related to the Polity IV democracy score. Similarly, Quinn, Hechter and Wibbels (2003) implicitly find that democracy in Fearon and Laitin's (2003) work is endogenous and that as a consequence the overall effect of ethnic fragmentation on civil war onset is sizeable.

4 Case Selection and Misclassification

The problem of selection bias has attracted considerable attention in the literature over the last decade or so. Work by Achen (1986) Geddes (1991) and King, Keohane and Verba (1994) has alerted scholars to these problems. Nevertheless, in the literature on conflict processes few studies directly address this problem. Cohen (1997), for instance seems aware of potential selection biases in the "minorities at risk" dataset, but his attempt to solve this problem with a time-trend is hardly appropriate. Similarly, Fearon and Laitin (1997) mention the problem of selection bias with respect to the MAR dataset, but come to the conclusion that there is no problem. Christin and Hug (2003) and Hug (2003) show, however, that in many empirical settings, the MAR dataset causes problems related to selection biases. They also illustrate ways in which these problems can be addressed.

Since the solutions to the problems of selection biases are increasingly well known, we refrain from discussing the appropriate empirical models here.[17] Instead we focus here on what happens if a dataset most likely affected by selection bias is aggregated to the country-level (possibly by year). As discussed above this is a strategy often employed in the study of civil wars and leads almost automatically to misclassifications. If for a particular conflict the number of death is measured with error, this conflict may or may not appear as an observation in a dataset on conflicts. Aggregating this information on a country-year level may result in misclassifications, for instance in a code for a particular country-year as conflict-free, even though a deadly conflict has occurred.

Hausman, Abrevaya and Scott-Morton (1998) demonstrate how even small amounts of misclassification affect the estimated coefficients, even if the misclassification is unrelated to any of the independent variables. They also propose a way in which misclassifications can be explicitly modeled in a *probit* setup. In a simple *probit*-model the log-likelihood function is simply

$$L(b|x) = \sum_{i=1}^{n} \{y_i ln\Phi(x_i'b) + (1-y_i) ln(1-\Phi(x_i'b))\} \quad (1)$$

where y is the observed dichotomous outcome, x a vector of explanatory vari-

[17] The references mentioned above give sufficient information on the appropriate solutions.

ables and b the coefficients to be estimated. If a_0 corresponds to the probability that $y_i = 0$ is misclassified as a 1 and if a_1 corresponds to the probability that $y_i = 1$ is misclassified as a 0, Hausman, Abrevaya and Scott-Morton (1998) derive the following log-likelihood function:

$$\begin{aligned} L(a_0, a_1, b|x) &= \sum_{i=1}^{n} \{y_i ln(a_0 + (1 - a_0 - a_1)\Phi(x_i'b)) \\ &\quad +(1 - y_i)ln(1 - a_0 - (1 - a_0 - a_1)\Phi(x_i'b))\} \end{aligned} \tag{2}$$

It is easy to see that equation 2 reduces to equation 1 if $a_0 = a_1 = 0$. Maximizing equation 2 yields estimates for the coefficients b but also for the amount of misclassification in the dataset through the values of a_0 and a_1. While Hausman, Abrevaya and Scott-Morton (1998) report estimates for a model employing this setup, they also suggest that both a_0 and a_1 may depend on some exogenous variables:

$$\begin{aligned} a_0 &= f(z_0) \\ a_1 &= f(z_1) \end{aligned} \tag{3}$$

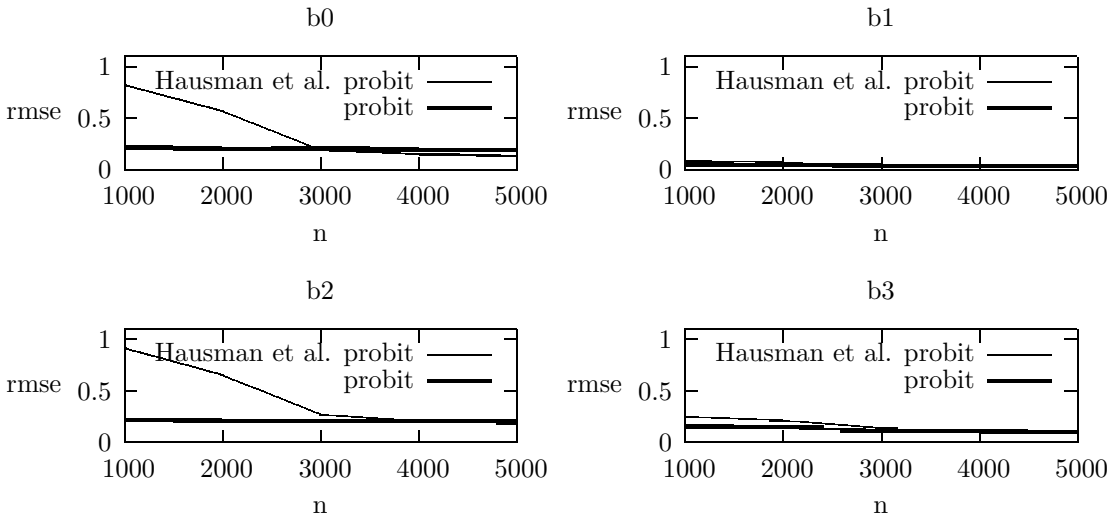
As for the estimates of a_0 and a_1 in Hausman, Abrevaya and Scott-Morton's (1998) original formulation (equation 1), constraints need to be set such that these values remain in the interval $[0, 1]$. As with regression models with dichotomous variables, the most convenient specification is either the logit transformation or the cumulative density function of the normal curve.

Hausman, Abrevaya and Scott-Morton (1998) report results from Monte-Carlo simulations demonstrating that the proposed estimator performs much better than simple probit estimations in the presence of misclassification. The equation they employ to generate the simulated dataset is the following:

$$\begin{aligned} y &= -1 + 0.2 \times x_1 + 1.5 \times x_2 - 0.6 \times x_3 + e \\ y^o &= 1 \text{ if } y > 0 \\ y^o &= 0 \text{ else} \end{aligned} \tag{4}$$

A certain percentage, namely 2, 5, and 20 percent of the observed y^o (both 0s and 1s) were then randomly selected and their value changed from 0 to 1, respec-

Figure 1: Monte Carlo Simulations: root-mean squared errors of coefficients



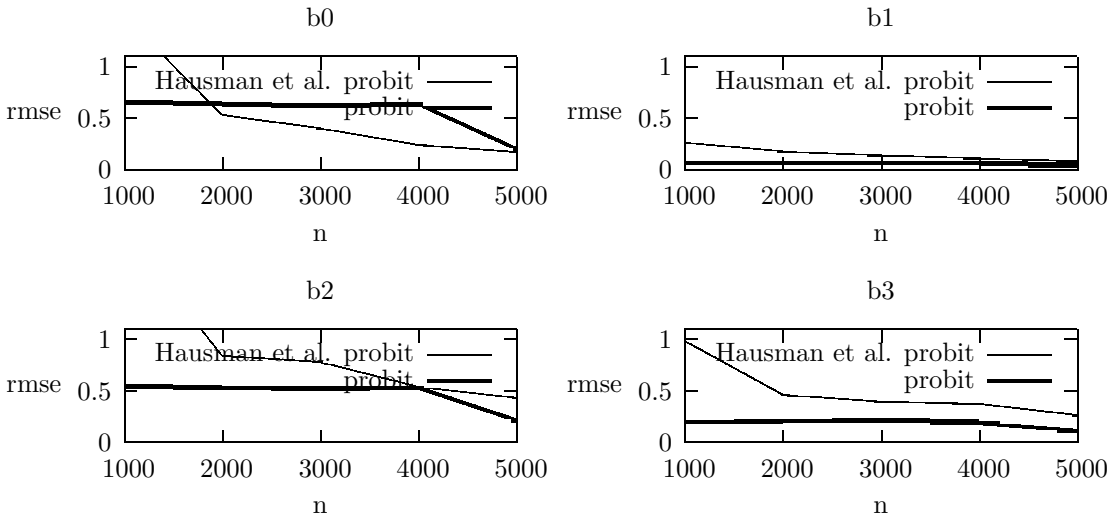
tively from 1 to 0. Monte-Carlo simulations performed by Hausman, Abrevaya and Scott-Morton (1998) with a sample of 5000 observations then clearly show that the estimated coefficients taking into account the problem of misclassification come much closer to the true values.

Since these Monte-Carlo simulations are limited in several ways, we extended these simulations by using exactly the same setup as shown in equation 4. First, we carried out the Monte-Carlo simulations for smaller datasets, namely for samples of 1000, 2000, 3000, 4000, and 5000 observations. Second, while Hausman, Abrevaya and Scott-Morton (1998) in their simulations kept the amount of misclassifications for both types at the same level and only estimated one coefficient, we allow both coefficients in equation 2 to take on the three values reported above and in addition the value 0. For each possible permutation we then estimated the model both under the assumption that $a_0 = a_1$ and under the assumption that $a_0 \neq a_1$. Finally, since the proposed estimator also allows the amount of misclassification to depend on exogenous variables, we also carried out Monte-Carlo simulations with $a_0 = f(x2)$ and $a_1 = f(x2)$.

Our results[18] first of all support Hausman, Abrevaya and Scott-Morton's

[18]We refrain from presenting and discussing in detail our results, since they are of secondary importance for the main arguments presented in this paper. The results of the Monte Carlo simulations are discussed in detail in Hug (2005), shortly available at http://www.ipw.unisg.ch/simonhug.

Figure 2: Monte Carlo Simulations: root-mean squared errors of coefficients



(1998) conclusion that their proposed estimator is preferable for samples of 5000 observations. We find, however, that for smaller samples the probit estimator has smaller root-mean-squared-errors than the estimator for misclassification, provided that the amount of classification is not too large ($a_0, a_1 < 0.05$). We illustrate these findings in figures 1 and 2. Both figures depict the root-mean-squared errors of the probit and the Hausman, Abrevaya and Scott-Morton (1998) estimator for the slope estimates in equation 4 for various sample sizes. Figure 1 reports the root-mean-squared errors for simulations where $a_0 = a_1 = 0.05$ and only one parameter for the misclassification was estimated, i.e., $a_0 = a_1$. Figure 2 reports the same information for simulations where $a_0 = 0.05$ and $a_1 = 0.02$ and two parameters were estimated for the misclassification.

If we allow for different amounts of misclassification for the two types (i.e. $a_0 \neq a_1$) we find that even when estimating a single coefficient for the misclassification Hausman, Abrevaya and Scott-Morton's (1998) estimator is preferable even for smaller samples, provided that neither a_0 nor a_1 are too small. More precisely, they both should clearly exceed 0. If we estimate two separate coefficients for the same datasets, we hardly find instances where the proposed estimator is preferable for the sample sizes considered. Only if of the coefficients is equal 0.2 or both of them clearly exceed 0 do we find Hausman, Abrevaya and Scott-Morton's (1998) estimator preferable.

Table 2: Misclassification: Fearon and Laitin (2003), updated data

	old data			new data		
	(1)	(2)	(3)	(4)	(5)	(6)
	logit	probit	probit	probit	probit	probit
	b	b	b	b	b	b
variables	(s.e)	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Prior war	-0.918	-0.391	-0.339	-0.340	-0.601	-0.804
	(0.223)	(0.117)	(0.126)	(0.126)	(0.255)	(0.358)
Per capita income	-0.345	-0.135	-0.131	-0.131	0.272	0.326
(in 1000's, lagged one year)	(0.066)	(0.026)	(0.028)	(0.028)	(0.135)	(0.180)
log population	0.255	0.108	0.101	0.101	0.201	0.221
(in 1000's, lagged one year)	(0.069)	(0.028)	(0.031)	(0.031)	(0.072)	(0.093)
log % mountainous	0.22	0.091	0.088	0.088	0.210	0.229
	(0.078)	(0.031)	(0.034)	(0.034)	(0.093)	(0.132)
Noncontiguous state	0.364	0.18	0.200	0.200	0.391	0.722
	(0.2)	(0.099)	(0.120)	(0.120)	(0.220)	(0.420)
Oil exporter	0.89	0.352	0.321	0.321	0.605	0.467
	(0.201)	(0.113)	(0.123)	(0.123)	(0.241)	(0.310)
New state	1.733	0.757	0.747	0.748	1.320	1.412
	(0.204)	(0.155)	(0.163)	(0.163)	(0.377)	(0.476)
Instability	0.632	0.259	0.251	0.251	0.482	0.375
	(0.22)	(0.096)	(0.101)	(0.101)	(0.227)	(0.341)
Democracy	0.024	0.008	0.006	0.006	0.012	0.015
Polity IV	(0.015)	(0.006)	(0.007)	(0.007)	(0.013)	(0.018)
Ethnic fractionalization	0.213	0.086	0.129	0.129	0.208	0.346
	(0.122)	(0.068)	(0.156)	(0.156)	(0.327)	(0.531)
Religious fractionalization	0.187	0.124	0.070	0.068	0.453	0.364
	(0.325)	(0.077)	(0.207)	(0.207)	(0.462)	(0.559)
b0	-6.646	-3.224	-3.150	-3.150	-5.235	-5.717
	(0.703)	(0.286)	(0.300)	(0.300)	(1.078)	(1.474)
Per capita income					-0.347	-0.359
(in 1000's, not lagged)					(0.180)	(0.159)
Eastern Europe						-2.157
(Dummy)						(5.437)
Latin America						0.131
(Dummy)						(0.494)
Sub-Saharan Africa						0.015
(Dummy)						(0.477)
Asia						0.063
(Dummy)						(0.500)
North Africa/Middle East						-0.012
(Dummy)						(0.526)
a_0					-1.926	-1.844
					(0.177)	(0.525)
Per capita income					0.406	0.400
(in 1000's, not lagged)					(0.101)	(0.141)
Eastern Europe						-0.592
(Dummy)						(0.560)
Latin America						-0.513
(Dummy)						(0.606)
Sub-Saharan Africa						-0.706
(Dummy)						(1.032)
Asia						-0.390
Dummy						(0.635)
North Africa/Middle East						-1.193
(Dummy)						(0.595)
a_1				-3.234	-0.473	0.157
				(15.483)	(0.717)	(1.195)
llik	-477.330	-481.420	-491.228	-491.229	-478.831	-472.668
n	6327	6327	6327	6327	6327	6327

Based on these Monte-Carlo simulations we clearly find that Hausman, Abre-

vaya and Scott-Morton's (1998) estimator is only advisable for large samples. Hence, while the dataset employed by Fearon and Laitin (2003) fulfills these requirements, Reynal-Querol's (2002*b*) does not. Thus we only illustrate this estimator in table 2 for a replication of Fearon and Laitin's (2003) simplest model explaining the onset of civil wars.[19] In column 1 of table 2 we display a replication of the logit model these authors report (Fearon and Laitin, 2003, 84). In the second column we present the estimated coefficients for the exact same empirical model, except that we employ a probit model. Not surprisingly, the estimated coefficients change, but the substantive insights remain unchanged. Only the coefficient for religious fractionalization becomes almost significant. In column 3 we depict the results based on the simplest way to introduce the probability of misclassification in a probit model, namely through a fixed probability for both types of error.[20] In column 4 we report the results of a model in which the probability of misclassification depends on the GDP per capita. These results appear as the most interesting, since they clearly illustrate that the probability of misclassification is affected by the economic development of the observed country. Compared to the results from the previous model, the estimated coefficients are, however, very similar. Finally, we also let the probability of misclassification be affected by the region to which a country belongs, with Western Europe serving as the base category. The probability of misclassification is not highly dependent on the region to which a country belongs. Only the coefficient for North Africa/Middle East for the probability of misclassification that $y_i = 1$ is classified as a 0 are significant.

Coefficients for all models presented in columns 2 to 5, are very stable. The standard errors are slightly inflated when the probability of misclassification is introduced in the models, which could decrease the statistical significance of the coefficients. This is the case of religious fractionalization, which is almost significant in the initial probit model (column 2) and then clearly fails to achieve statistical significance in the subsequent models (column 3-5).

the latter's effect for a subset of the relevant values. More precisely

Substantively it is of considerable interest to identify what type of misclas-

[19]Information on the datasets and variables used appears in the appendix.

[20]Here, the probabilities of misclassification are specified as a standard cumulative normal density curve. Hence, the predicted probability of misclassification is $\Phi(a_i)$, where a_i are the estimated coefficients. It has to be noted, that the estimates of this model are very sensitive to the selection of starting values and the functional forms chosen.

sifications are present in a given dataset, and on what these misclassifications depend. In addition, however, addressing the issue of misclassification, as discussed here, also mildly affects some of the estimated relationships in empirical models. Tests with other empirical applications has to demonstrate whether these corrections are needed or not.

5 Endogenous Institutions

Political institutions appear with increasing frequency in empirical models explaining societal conflicts or civil wars. As discussed above, these institutions are, however, rarely exogenous as the other variables included in empirical models. Neglecting this problem may affect various substantive conclusions reached based on the misspecified models. Hausman (1978) provided the basic ideas for tests of exogeneity and their application to classical linear regressions. Building on Hausman (1978) Rivers and Vuong (1988) provide a test for exogeneity in probit models. Elbadawi and Sambanis (2002) employ this test in a study of civil war onset. They find, however, that the only variable affected by endogeneity is a variable related to the Polity IV democracy score. Correcting for this problem, they hardly find any differences in the substantively interesting part of the model. It has to be noted, however, that their model does not contain, apart the one element from the Polity IV democracy score any variables related to political institutions.

In essence both the Hausman (1978) and Rivers and Vuong (1988) exogeneity tests rely on finding instruments for the variables potentially affected by endogeneity. The residuals from a regression of the latter variables on the instruments are included as additional regressor in the regression of interest. If the estimated coefficient for the residuals is significant, this suggests that the null hypothesis of endogeneity cannot be rejected for the variable concerned.

Ways to correct the endogeneity problem consists of using the predicted values from the auxiliary regression as instrument for the variable affected by endogeneity. In the probit framework, however, the estimated standard errors for the instrumental variable have to be corrected as Rivers and Vuong (1988) illustrate. We employ this estimation strategy first for the analyses discussed in Fearon (2004) and Reynal-Querol (2002*b*). The former example only has a limited set of institutional variables, while the second employs a series of political institutions

as independent variables. In both cases, we test for endogeneity of institutions, by regressing the institutional variables on all other included variables in the model lagged by one period.[21]

Table 3 reproduces the results for Fearon and Laitin (2003)[22] and tables 4, 5 present the analyses for Reynal-Querol (2002*b*).[23] Like in the previous analyses table 3 replicates the simplest model from Fearon and Laitin (2003). Columns 1 (logit) and 2 (probit), replicates the results of the original model.[24] In the third column, we report the results for the same model, but estimated only with the cases for which we could calculate residuals and predicted values based on our auxiliary regression (table 9 in the appendix).[25] Again the estimated coefficients are very similar to the ones presented in column 2. When including as additional regressor the residuals of the auxiliary regression for the democracy score we find an estimated coefficient and a standard error which does not allow us to reject the exogeneity hypothesis (column 4). Nevertheless, we report in column 5 the results of an estimation where the democracy score is replaced by its predicted values from the auxiliary regression. Not surprisingly, the estimated effect of this variable is hardly different from the results reported in column 4.[26] When comparing the other estimated coefficients, we note that the effect of the variable "noncontiguous state" is negative in column 4 and 5, though hardly significant. More importantly, the effect of political instability is not significant when we presume that the democracy score is endogenous.

When we also address for the same problem the issue of misclassification, we find several interesting results. First of all, under most specification for the misclassification, the predicted values of the democracy score is close to reaching statistical significance. Political instability, on the other hand, loses much of its

[21]Tables 9 and 11 in the appendix report the results of these auxiliary regression for Fearon and Laitin (2003) and Reynal-Querol (2002*b*) analyses, respectively.

[22]The results reported here rely on Fearon and Laitin's (2003) datafile updated according to Fearon (2004). More details about these modifications as well as the original results are presented in the appendix.

[23]In the appendix we report an additional analysis in table 12, where the exogeneity of the institutional variables cannot be rejected.

[24]The small differences that appear compared to table 2 are due to the software used (Stata and Gauss).

[25]In these auxiliary regressions we used as predictors all the remaining independent variables lagged by one period.

[26]Following Elbadawi and Sambanis (2002) we also determined the standard errors for the predicted values for this and the subsequent models by bootstrapping The adjustments, however, were substantively so small that we refrain from reporting them in the tables here.

explanatory power. Similarly, as soon as we let per capita income explain the amount of misclassification, this variable loses much of its explanatory power for the onset of civil wars. On the other hand we find again in column 8 that the dummy for North Africa/Middle East has a significant effect.

Table 3: Endogenous institutions: Fearon and Laitin (2003), updated file

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	logit	probit	probit	probit	probit	probit	probit	probit
	b	b	b	b	b	b	b	b
variables	s.e	s.e	s.e	s.e	s.e	s.e	s.e	s.e
Prior war	-0.821	-0.339	-0.347	-0.403	-0.395	-0.395	-0.642	-0.710
	(0.3)	(0.126)	(0.126)	(0.13)	(0.129)	(0.129)	(0.212)	(0.239)
Per capita income	-0.335	-0.131	-0.131	-0.196	-0.189	-0.189	0.060	0.065
(in 1000's. lagged one year)	(0.071)	(0.028)	(0.028)	(0.046)	(0.045)	(0.045)	(0.117)	(0.126)
log population	0.249	0.1	0.104	0.112	0.111	0.111	0.139	0.140
(in 1000's. lagged one year)	(0.073)	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)	(0.045)	(0.047)
log % mountainous	0.21	0.088	0.095	0.103	0.103	0.103	0.120	0.106
	(0.084)	(0.034)	(0.035)	(0.035)	(0.035)	(0.035)	(0.051)	(0.053)
Noncontiguous state	0.477	0.199	0.194	-0.359	-0.337	-0.337	-0.446	-0.312
	(0.271)	(0.12)	(0.12)	(0.331)	(0.329)	(0.329)	(0.504)	(0.500)
Oil exporter	0.789	0.321	0.33	0.873	0.851	0.851	1.173	1.008
	(0.278)	(0.123)	(0.123)	(0.322)	(0.320)	(0.321)	(0.530)	(0.499)
New state	1.692	0.747	0.75	0.575	0.59	0.590	0.649	0.504
	(0.337)	(0.163)	(0.163)	(0.190)	(0.189)	(0.189)	(0.274)	(0.331)
Instability	0.598	0.25	0.255	0.174	0.187	0.187	0.174	0.148
	(0.234)	(0.100)	(0.101)	(0.111)	(0.110)	(0.110)	(0.163)	(0.159)
Democracy	0.017	0.006	0.007	0.121				
(Polity IV, lagged one year)	(0.017)	(0.007)	(0.007)	(0.064)				
Ethnic fractionalization	0.251	0.129	0.144	0.327	0.332	0.332	0.524	0.527
	(0.372)	(0.156)	(0.158)	(0.189)	(0.188)	(0.188)	(0.281)	(0.395)
Religious fractionalization	0.159	0.07	0.089	0.046	0.043	0.043	-0.040	-0.197
	(0.504)	(0.206)	(0.208)	(0.209)	(0.209)	(0.211)	(0.293)	(0.354)
Democracy residual				-0.114				
				(0.064)				
Democracy predicted value					0.115	0.115	0.173	0.186
					(0.063)	(0.063)	(0.103)	(0.098)
b0	-6.575	-3.15	-3.216	-3.031	-3.048	-3.048	-3.338	-3.218
	(0.731)	(0.300)	(0.304)	(0.321)	(0.320)	(0.321)	(0.566)	(0.660)
Per capita income							-1.887	-3.052
(in 1000's, not lagged)							(1.817)	(4.487)
Eastern Europe								-2.961
(Dummy)								(97.325)
Latin America								-0.447
(Dummy)								(27.817)
Sub-Saharan Africa								0.730
(Dummy)								(24.289)
Asia								0.576
(Dummy)								(24.308)
North Africa/Middle East								0.454
(Dummy)								(24.349)
a_0						-7.948	-1.437	-1.583
						(1617.921)	(0.592)	(24.418)
Per capita income							0.390	0.334
(in 1000's, not lagged)							(0.099)	(0.149)
Eastern Europe								-0.784
(Dummy)								(0.503)
Latin America								-0.798
(Dummy)								(0.473)
Sub-Saharan Africa								-1.242
(Dummy)								(0.920)
Asia								-0.831
(Dummy)								(0.522)
North Africa/Middle East								-1.355
(Dummy)								(0.621)
a_1						-5.203	-0.474	0.582
						(1188.738)	(0.684)	(0.993)
llik	-490.281	-491.229	-485.123	-483.624	-484.124	-484.124	-473.473	-444.532
n	6327	6327	6285	6285	6285	6285	6285	6285

While political institutions appear in Fearon and Laitin's (2003) analyses only through the polity-score, Reynal-Querol (2002*a*, 50-51) studies in more detail the role of political variables on civil wars. Table 4 first replicates the second model from table 1 in Reynal-Querol's (2002*a*, 40) article and then tests again for the endogeneity of the institutional variables. Again the logit (column 1) and probit (column 2) models have different coefficients, but remain substantially equivalent. The third column is again the same probit model as the one presented in column 2, but based only on the cases used in the tests for endogeneity in column 4. In column 4, we add the residuals for democracy (democ) and democracy squared (democ2). While the estimated coefficients fail to reach statistical significance, jointly they are significant. Thus, we cannot reject the hypothesis of endogeneity. In the fifth column we add the predicted values for the democracy and the democracy squared variables instead of the original variables. The estimated coefficients for these two variables are individually not significant. This is contrary to the results appearing in columns 1-3.

Thus, the effects of democracy in this model of Reynal-Querol's (2002*a*) article seem largely due to the endogenous nature of this political institution. Comparing the remaining coefficients of models 4 and 5, with those in the third column shows that some coefficients change considerably, but their statistical significance (or the absence thereof) remains the same. For example, the effect of GDP is in column 4 and 5 negative, contrary to the third column, but it remains not significant.

Table 4: Endogenous institutions: Reynal-Querol (2002a) 1

	(1)	(2)	(3)	(4)	(5)
	logit	probit	probit	probit	probit
	b	b	b	b	b
variables	(s.e)	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Lpop	0.428	0.245	0.162	0.223	0.139
	(0.254)	(0.138)	(0.143)	(0.347)	(0.326)
Lgdp	1.482	0.816	0.595	-0.324	-0.419
	(0.702)	(0.399)	(0.416)	(0.724)	(0.689)
Educ	-0.288	-0.152	-0.197	-0.684	-0.562
	(0.218)	(0.118)	(0.125)	(0.368)	(0.340)
Ex	0.129	0.082	-0.008	0.073	0.040
	(0.121)	(0.070)	(0.011)	(0.074)	(0.056)
Ex2	-0.005	-0.003	0.000	-0.003	-0.001
	(0.004)	(0.003)	(0.000)	(0.003)	(0.002)
I	-0.144	-0.081	-0.071	-0.078	-0.077
	(0.051)	(0.028)	(0.028)	(0.031)	(0.030)
C	0.015	0.009	0.014	0.002	-0.003
	(0.028)	(0.016)	(0.017)	(0.024)	(0.023)
Trib	7.377	4.105	3.542	4.468	4.853
	(2.369)	(1.275)	(1.326)	(2.214)	(2.169)
Democ	0.679	0.360	0.326	1.355	
	(0.267)	(0.147)	(0.153)	(2.853)	
Democ2	-0.087	-0.046	-0.040	-0.058	
	(0.032)	(0.017)	(0.017)	(0.290)	
Safrica	-1.397	-0.768	-0.986	-1.079	-0.972
	(1.110)	(0.615)	(0.627)	(0.670)	(0.657)
Asiae	3.847	2.158	2.135	2.689	2.815
	(1.417)	(0.775)	(0.822)	(0.917)	(0.892)
Laam	-0.378	-0.206	-0.613	-0.608	-0.308
	(0.956)	(0.519)	(0.517)	(0.570)	(0.525)
cristx	5.870	3.236	3.245	3.319	3.631
	(1.893)	(1.069)	(1.150)	(1.176)	(1.166)
musx	6.257	3.504	3.306	3.459	3.869
	(2.126)	(1.174)	(1.261)	(1.294)	(1.286)
budx	4.778	2.700	2.385	2.384	2.975
	(1.841)	(1.056)	(1.119)	(1.146)	(1.129)
hinx	8.680	4.865	4.607	5.284	5.150
	(2.554)	(1.438)	(1.574)	(1.666)	(1.620)
anix	-0.335	-0.180	0.303	0.128	0.230
	(1.837)	(1.002)	(1.054)	(1.064)	(1.059)
res. (democ)				-1.027	
				(2.847)	
res. (democ2)				0.014	
				(0.290)	
pred. (democ)					1.699
					(2.655)
pred. (democ2)					-0.097
					(0.268)
Constant	-22.008	-12.335	-9.612	-3.906	-2.842
	(7.736)	(4.298)	(4.373)	(6.354)	(6.079)
llik	-78.694	-77.923	-69.726	-66.618	-70.000
Pseudo r^2	0.4075	0.413	0.426	0.452	0.424
LR chi2(3) =				6.210	
Prob > chi2 =				0.045	
n	369	369	313	313	313

Table 5: Endogenous institutions: Reynal-Querol (2002a) 2

variables	(1) logit b (s.e)	(2) probit b (s.e.)	(3) probit b (s.e.)	(4) probit b (s.e.)	(5) probit b (s.e.)
Lpop	0.671	0.332	0.339	-0.337	-0.401
	(0.153)	(0.079)	(0.117)	(0.344)	(0.347)
Lgdp	-0.384	-0.226	0.618	-1.421	-0.744
	(0.400)	(0.216)	(0.392)	(0.859)	(0.762)
Educ	-0.249	-0.142	-0.198	-0.942	-0.657
	(0.137)	(0.073)	(0.107)	(0.296)	(0.256)
I	-0.064	-0.032	-0.082	-0.097	-0.083
	(0.030)	(0.016)	(0.027)	(0.039)	(0.035)
C	-0.033	-0.014	0.016	0.003	0.001
	(0.019)	(0.010)	(0.018)	(0.029)	(0.026)
Trib	6.125	3.627	3.605	6.121	6.397
	(1.920)	(1.050)	(1.305)	(1.753)	(1.689)
Democ	0.150	0.076	0.031	-0.186	
	(0.062)	(0.033)	(0.051)	(0.794)	
Incv	-1.011	-0.474	-0.541	6.417	
	(0.340)	(0.160)	(0.246)	(3.124)	
LtIndem	0.233	0.059	0.689	-25.780	
	(0.469)	(0.285)	(0.585)	(24.151)	
Safrica	-1.809	-1.094	-0.962	-1.307	-0.838
	(0.783)	(0.407)	(0.638)	(0.690)	(0.609)
Asiae	1.840	0.967	2.158	2.480	2.292
	(0.635)	(0.353)	(0.720)	(0.792)	(0.748)
Laam	-0.350	-0.254	-0.375	-0.609	-0.273
	(0.616)	(0.314)	(0.501)	(0.578)	(0.493)
cristx	6.118	3.444	3.664	3.697	3.706
	(1.467)	(0.846)	(1.087)	(1.135)	(1.096)
musx	6.041	3.400	3.595	3.700	4.164
	(1.556)	(0.889)	(1.219)	(1.283)	(1.249)
budx	6.883	3.898	3.072	3.014	3.484
	(1.525)	(0.873)	(1.123)	(1.175)	(1.116)
hinx	3.997	2.308	4.502	4.858	5.088
	(1.716)	(0.981)	(1.513)	(1.626)	(1.536)
anix	0.263	0.001	0.299	0.097	-0.493
	(1.439)	(0.825)	(1.020)	(1.015)	(0.998)
res. (democ)				0.225	
				(0.795)	
res. (Incv)				-7.242	
				(3.146)	
res. (LtIndem)				26.505	
				(24.089)	
pred. (democ)					-0.679
					(1.353)
pred. (Incv)					6.569
					(5.369)
pred. (Ltndem)					-30.125
					(25.585)
Constant	-7.240	-3.893	-11.718	11.152	7.004
	(3.835)	(2.041)	(4.097)	(7.624)	(7.102)
llik	-138.280	-138.702	-73.379	-67.444	-73.229
Pseudo r^2	0.3704	0.369	0.418	0.465	0.419
LR chi2(3)				11.870	
Prob > chi2				0.008	
n	596	596	319	319	319

Table 5 replicates model 7 from table 6 in Reynal-Querol's (2002*a*, 50-51) article. She demonstrates that controlling for the inclusiveness of political institutions cannot reduce the effect of animist diversity (Trib) on ethnic civil wars. This is confirmed by all our models. Again, in column 4 we add the residuals for the political institution variables and find an overall significant effect. Thus, these institutional variables are hardly exogenous to the analysis presented here. In column 5 we replace again the original variables by their predicted values. The reported estimated coefficients suggest that none of the political variables have any significant effect. The effect for the level of inclusiveness of the political system turns out to be positive but fails to reach statistical significance. This contrary to the result reported in Reynal-Querol (2002*a*), who found a significant negative effect (column 1). These results would be enough to take the impact of power-sharing institutions on civil wars with prudence.

6 Conclusion

Studies of conflict processes have made tremendous progress in assessing what causes rebellious and violent behavior, or what affects the onset and duration of civil wars. In the face of the human tragedy these events entail, this is of great value. But several of the key findings in the literature are contested either on methodological grounds or with respect to the research design employed.

In this paper we raised two issues which affect many quantitative studies of conflict processes, namely case selection and the resulting misclassifications, and the endogenous nature of political institutions. By reanalyzing prominent examples of empirical studies and showing ways how to address these two problems, we were able to show that these problems may affect some of the substantive results of the studies discussed.

Appendix

In this appendix we report information on the datasets employed and additional analyses referred to in the main text. The results presented in the paper refer to the models presented in Fearon and Laitin (2003). However, the dataset has been updated according to the modifications applied in the Fearon's (2004) dataset.[27] In the appendix we also present analyses of endogeneity on a revised dataset that treats lagged variables in a different way than what Fearon and Laitin (2003) have done. Contrary to these authors, we do not impute the value of the first country / year for the missing lagged variables. This corrected version leads to a datafile with more missing cases (161 lagged). As a consequence the variable "new state" becomes a constant, since the variable is coded 1 only within the first 2 years of the country series. Therefore the variable "New State" is not included in the model. The results of the analyses on the lagged corrected file are presented in tables 8 and 10. Like in the analyses presented in the table 3, we report in column 1 (logit) and 2 (probit) the results of the original models. We then, in column 3 report the results without the variable "New State". The results of this model are very similar to the ones presented in the second column. In the fourth column we present the results of the same model, but estimated only with the cases for which we could calculate residuals and predicted values based on the auxiliary regression (presented in table 10). Once again, the estimated coefficients are very similar to the ones presented in column 3. The only difference is that political instability is significant in the fourth model. Like in table 3, the inclusion in the model of the residuals of the auxiliary regression for the democracy score does not allow us to reject the exogeneity hypothesis (column 5). Finally, when democracy score is replace by the predicted value of the auxiliary regression (column 6), the rejection of the exogeneity hypothesis is not possible.

We finally recall the results of the analyses we applied previously on the Fearon and Laitin's (2003) datafile in tables 13, 14 and 15.

Table 6 presents the descriptive statistics for Fearon and Laitin (2003), updated and Reynal-Querol (2002*a*) files, respectively. Table 7 displays the variables used as well as their definitions and measures. Tables 9, 14, 10 and 11 report the

[27]These modifications concern the civil war onset variable for Djibouti (1991 and 1993), Ethiopia (1992 and 1997) and Indonesia (1989,1991,1999). Many thanks to James Fearon for having shared the information on these modifications.

results of the auxiliary regression used for the exogeneity tests reported in tables 3, 15, 8, 4, 5 and 12. Finally, table 12 reports a final test of exogeneity, for which the null hypothesis of exogeneity cannot be rejected.

Table 6: Descriptive statistics

Variable names (Fearon and Laitin, 2003), updated file	Variable names (Datafile)	n	Min	Max	Mean	Std. dev.
Civil War onset (dep. var.)	Onset	6610	0	1	0.02	0.13
Prior war	warl	6610	0	1	0.14	0.34
Per capita income	gdpenl	6373	0.05	66.74	3.65	4.54
log(population)	lpopl1	6585	5.40	14.03	9.05	1.46
log(% mountainous)	lmtnest	6610	0	4.56	2.18	1.40
Noncontiguous state	ncontig	6610	0	1	0.17	0.38
Oil exporter	Oil	6610	0	1	0.13	0.34
New state	nwstate	6610	0	1	0.03	0.17
Instability	instab	6596	0	1	0.15	0.35
Democracy	polity2l	6541	-10	10	-0.48	7.51
Ethnic fractionalization	ethfrac	6610	0	0.93	0.39	0.29
Religious fractionalization	relfrac	6610	0	0.78	0.37	0.22
(Reynal-Querol, 2002*a*)	(Datafile)					
Ethnic war (dep. war)	ETHWAR	966	0	1	0.11	0.32
Lpop	LPOP	873	3.74	13.94	8.68	1.75
Lgdp	LGDPCH	867	5.55	10.37	7.70	1.03
Educ	TYR15	725	.09	11.94	4.42	2.75
Ex	EX0	533	0	3092.69	80.19	269.28
Ex2	EX02	533	0	9564710	78804.97	577524.31
I	I	867	1	48	16.37	9.31
C	C	867	10.90	100.50	66.75	12.02
Trib	TRIB	882	0	1.66	0.09	0.21
Democ	DEMOCP3	799	0	10	3.73	4.33
Democ2	DEMOC2	799	0	100	32.62	41.91
Incv	VAR	966	0	3	0.59	1.01
LtIndem	TRVDEM	785	0	22.59	0.09	0.92
MAJOc	MAJOC	771	0	1	0.12	0.33
Ltmaj	TRMAJ	744	0	0.88	0.01	0.06
Safrica	SAFRICA	966	0	1	0.31	0.46
Asiae	ASIAE	966	0	1	0.10	0.30
Laam	LAAM	966	0	1	0.21	0.41
Religious dummies	JUIFSX	966	0	1	0.01	0.08
(not presented)	CRISTX	966	0	1	0.54	0.50
	MUSX	966	0	1	0.20	0.40
	BUDX	966	0	1	0.02	0.15
	HINX	966	0	1	0.01	0.12
	CHIX	966	0	1	0.01	0.08
	ANIX	966	0	1	0.07	0.25
	CULTSX	966	0	1	0.014	0.120
	OTHRELX	966	0	1	0.007	0.085

Table 7: Glossary

Variable names	Definitions and measures
(Fearon and Laitin, 2003)	
Civil War onset	Own categorization close to COW criteria, countries that had a population of at least half a million in 1990, were conflict is between agent of state and non state groups, with a 1000 deaths threshold and at least 100 deaths per year on both sides. They exclude colonial wars. They coded onset as "1" for all country-years in which civil war started and "0" for others.
Prior war	Control variable coded 1 if the country had a distinct civil war ongoing in the previous year.
Per capita income	It is measured as thousands of 1985 U.S: dollars for Penn World Tables and World Bank data. In the general model the lagged one year variable is used.
log(population)	log of country's population (World Bank figures).
log(% mountainous)	Proportion of "rough terrain" according to the coding of the geographer A.J. Gerard.
Noncontiguous state	Dummy variable for countries with noncontiguous territory; territory holding at least 10'000 people and separated from the land area containing the capital city either by land or by 100km of water.
Oil exporter	Dummy variable. County-years are marked "1" if fuel exports exceeded one-third of export revenues, using World Bank data.
New state	Dummy variable, where country in their first and second years of independence are marked with "1".
Instability	Dummy variable indicating whether the country had a three-or-greater change on the Polit0y IV regime index in any of the three years prior to the country-year in question.
Democracy	Polity IV's democracy and autocracy measures (ranging from -10 to +10)
Ethnic fractionalization	The authors use several measure of this concept: (1) Ethnolinguistic fractionalization (ELF) index based on *Atlas Narodov Mira* 1964. (2) a measure of the share of the population belonging to the largest ethnic group (CIA factbook and own data); (3) the number of distinct languages spoken by groups exceeding 1% of the country's population, based on Grimes and Grimes 1996.
Religious fractionalization	An analogue measure of the ELF, constructed using data from the CIA Factbook and other sources.
(Reynal-Querol, 2002*a*)	
Ethnic war (dep. war)	She uses Doyle and Sambanis's (2000) definition of civil war.
Lpop	Log of the population at the beginning of the period.
Lgdp	Log of the real gross domestic product (GDP) per capita of the initial period (1985 international prices).
Educ	Average years of schooling in the total population (Barro and Lee, 1996).
Ex	Share of primary exports in GDP.
Ex2	Square of Ex.
I	Investment share of GDP.
C	Consumption share of GDP.

Variable names	Definitions and measures
Trib	Animist diversity. Dummy variable extracted from a religious polarization variable. She uses several sources (*L'Etat des religions dans le monde*, *World Christian Encyclopedia*, and *The Statesman Year's Book of 1987.*).
Democ	Democracy level from Polity III data source.
Democ2	Square of Democ.
Incv	Level of inclusiveness of the political system (Colomer, 2001)
LtIndem	Interaction term: ltrib x demIncv
MAJOc	Majoritarian system (Colomer, 2001). For nonfree countries, she used data from the Freedom Hose and Polity III data sources.
Ltmaj	Measure of majoritarian system.
Safrica	Dummy variable for sub-Saharian countries.
Asiae	Dummy variable for Asian countries.
Laam	Dummy variable for Latin American countries.
Religious dummies	In all models, she include the religious dummy variables, but don't present the coefficients. The dummies are: JUIFSX, CRISTX, MUSX, BUDX, HINX, CHIX, ANIX, CULTSX, OTHRELX.

Table 8: Endogenous institutions: Fearon and Laitin (2003), with corrected lagged

variables	(1) logit b (s.e)	(2) probit b (s.e)	(3) probit b (s.e)	(4) probit b (s.e)	(5) probit b (s.e)	(6) probit b (s.e)	(7) probit b (s.e)	(8) probit b (s.e)	(9) probit b (s.e)
Prior war	-0.767	-0.314	-0.323	-0.29	-0.321	-0.313	-0.313	-0.446	-0.619
	(0.301)	(0.126)	(0.125)	(0.127)	(0.131)	(0.130)	(0.130)	(0.223)	(0.263)
Per capita income	-0.3	-0.118	-0.122	-0.126	-0.159	-0.15	-0.151	0.018	-0.030
(in 1000's. lagged one year)	(0.07)	(0.027)	(0.027)	(0.029)	(0.046)	(0.045)	(0.045)	(0.115)	(0.131)
log population	0.248	0.099	0.091	0.096	0.102	0.1	0.100	0.126	0.158
(in 1000's. lagged one year)	(0.075)	(0.032)	(0.031)	(0.032)	(0.033)	(0.033)	(0.033)	(0.054)	(0.060)
log% mountainous	0.195	0.082	0.082	0.086	0.088	0.089	0.089	0.101	0.110
	(0.086)	(0.035)	(0.034)	(0.036)	(0.036)	(0.036)	(0.036)	(0.053)	(0.058)
Noncontiguous state	0.358	0.145	0.158	0.112	-0.168	-0.148	-0.148	-0.167	0.068
	(0.288)	(0.126)	(0.125)	(0.131)	(0.335)	(0.333)	(0.334)	(0.453)	(0.516)
Oil exporter	0.732	0.294	0.305	0.304	0.57	0.551	0.552	0.695	0.610
	(0.289)	(0.125)	(0.124)	(0.130)	(0.318)	(0.316)	(0.317)	(0.481)	(0.501)
New state	1.659	0.743							
	(0.462)	(0.223)							
Instability	0.615	0.256	0.229	0.264	0.221	0.236	0.236	0.316	0.345
	(0.235)	(0.101)	(0.100)	(0.101)	(0.112)	(0.111)	(0.112)	(0.177)	(0.192)
Democracy	0.017	0.006	0.007	0.009	0.065				
(Polity IV, lagged one year)	(0.017)	(0.007)	(0.007)	(0.007)	(0.062)				
Ethnic fractionalization	0.364	0.166	0.184	0.14	0.225	0.235	0.235	0.281	0.216
	(0.386)	(0.160)	(0.159)	(0.164)	(0.189)	(0.188)	(0.189)	(0.249)	(0.301)
Religious fractionalization	0.223	0.094	0.106	0.174	0.167	0.162	0.163	0.246	0.165
	(0.522)	(0.211)	(0.210)	(0.218)	(0.218)	(0.218)	(0.218)	(0.275)	(0.349)
Democracy residual					-0.057				
					(0.062)				
Democracy predicted value						0.06	0.060	0.082	0.111
						(0.062)	(0.062)	(0.089)	(0.098)
Constant	-6.647	-3.163	-3.062	-3.152	-3.069	-3.087	-3.087	-3.136	-3.041
	(0.749)	(0.305)	(0.301)	(0.314)	(0.327)	(0.326)	(0.326)	(0.485)	(0.527)
a_1							-3.946	-0.050	1.521
							(176.354)	(1.463)	(0.825)
Per capita income								0.283	0.155
(in 1000's, not lagged)								(0.137)	(0.110)
Eastern Europe									-0.354
(Dummy)									(0.499)
Latin America									-0.903
(Dummy)									(0.515)
Sub-Saharan Africa									-1.229
(Dummy)									(0.641)
Asia									-0.935
(Dummy)									(0.495)
North Africa/Middle East									-1.377
(Dummy)									(0.648)
llik	-467.796	-468.018	-472.676	-437.764	-437.362	-438.036	-438.036	-431.125	-424.369
Pseudo R^2	0.087	0.086	0.077	0.081	0.082	0.081			
n	6212	6212	6212	6055	6055	6055	6055	6055	6055

Table 9: Auxiliary regression: Fearon and Laitin (2003), updated file

variables	Democracy (lagged)
	b
	(s.e)
Prior war (2 lagged)	0.618
	(0.263)
Per capita income (2 lagged)	0.600
	(0.022)
log (population (2 lagged)	-0.071
	(0.064)
log(% mountainous)	-0.078
	(0.062)
Noncontiguous state	4.900
	(0.239)
Oil exporter (lagged)	-5.197
	(0.256)
New state	1.683
	(0.525)
Instability	0.929
	(0.241)
Ethnic fractionalization	-1.566
	(0.327)
Religious fractionalization	0.346
	(0.412)
Constant	-1.726
	(0.586)
R^2	0.2566
Adj. R^2	0.2554
root MSE	6.5192
n	6302

Table 10: Auxiliary regression: Fearon and Laitin (2003) with corrected lagged

variables	Democracy (lagged) b (s.e)
	Democracy (lagged) Coeff. Std. Err
Prior war (2 lagged)	0.661 (0.262)
Per capita income (2 lagged)	0.625 (0.022)
log (population (2 lagged)	-0.087 (0.066)
log(% mountainous)	-0.038 (0.063)
Noncontiguous state	4.940 (0.244)
Oil exporter (lagged)	-5.147 (0.259)
New state (2 lagged)	1.535 (0.512)
Instability (lagged)	1.012 (0.241)
Ethnic fractionalization	-1.451 (0.331)
Religious fractionalization	0.049 (0.418)
Constant	-1.789 (0.599)
R^2	0.2656
Adj. R^2	0.2644
root MSE	6.4933
n	6072

Table 11: Auxiliary regressions: Reynal-Querol (2002*a*)

	democp3 b (s.e)	democ2 b (s.e.)	var b (s.e.)	trave b (s.e.)	majoc b (s.e.)	trmaj b (s.e.)
Lpop (lagged)	0.010	-0.999	0.086	-0.004	0.016	0.000
	(0.128)	(1.210)	(0.030)	(0.007)	(0.014)	(0.001)
Lgdp (lagged)	1.553	14.563	0.339	-0.009	-0.019	-0.003
	(0.361)	(3.412)	(0.086)	(0.021)	(0.037)	(0.003)
Educ (lagged)	0.504	5.815	0.102	-0.003	0.046	0.000
	(0.123)	(1.161)	(0.029)	(0.007)	(0.013)	(0.001)
Ex (lagged)	0.002	0.001	-0.002	0.000	0.001	0.000
	(0.003)	(0.031)	(0.001)	(0.000)	(0.000)	(0.000)
Ex2 (lagged)	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
I (lagged)	-0.004	-0.055	-0.002	-0.001	-0.002	0.000
	(0.031)	(0.296)	(0.008)	(0.002)	(0.003)	(0.000)
C (lagged)	0.026	0.182	-0.001	-0.001	0.002	0.000
	(0.022)	(0.205)	(0.005)	(0.001)	(0.002)	(0.000)
Trib (lagged)	-0.688	-0.457	-0.352	0.004	-0.120	0.010
	(1.105)	(10.443)	(0.264)	(0.063)	(0.113)	(0.009)
Constant	-11.939	-104.908	-2.953	0.274	-0.106	0.028
	(3.498)	(33.048)	(0.833)	(0.200)	(0.362)	(0.030)
r^2	0.377	0.417	0.362	0.015	0.095	0.017
adj r^2	0.363	0.403	0.347	-0.008	0.073	-0.007
root MSE	3.496	33.028	0.858	0.200	0.357	0.030
n	353	353	367	353	335	335

Table 12: Endogenous institutions: Reynal-Querol (2002*a*) 3

variables	(1) logit b (s.e)	(2) probit b (s.e.)	(3) probit b (s.e.)	(4) probit b (s.e.)	(5) probit b (s.e.)
Lpop	0.760	0.402	0.257	0.675	0.521
	(0.186)	(0.099)	(0.131)	(0.495)	(0.467)
Lgdp	0.247	0.151	0.805	-1.149	-1.407
	(0.464)	(0.253)	(0.402)	(0.955)	(0.892)
Educ	-0.392	-0.212	-0.233	-0.411	-0.348
	(0.165)	(0.087)	(0.124)	(0.400)	(0.344)
I	-0.069	-0.042	-0.074	-0.070	-0.067
	(0.033)	(0.018)	(0.029)	(0.042)	(0.039)
C	-0.029	-0.012	0.013	0.019	0.009
	(0.021)	(0.011)	(0.018)	(0.027)	(0.025)
Trib	6.084	3.612	4.007	3.262	3.912
	(1.933)	(1.077)	(1.294)	(2.475)	(2.305)
Democ	0.645	0.346	0.454	-0.237	
	(0.216)	(0.115)	(0.170)	(3.563)	
Democ2	-0.069	-0.038	-0.056	0.120	
	(0.025)	(0.013)	(0.019)	(0.340)	
MAJOc	-0.132	0.001	0.084	-12.964	
	(0.793)	(0.435)	(0.703)	(8.866)	
Ltmaj	-142.207	-73.183	-3864.441	-4414.857	
	(62.321)	(35.331)	(4050033)	(3576016)	
Safrica	-0.640	-0.481	-0.864	-0.891	-0.746
	(0.835)	(0.451)	(0.640)	(0.688)	(0.660)
Asiae	3.350	1.728	2.265	2.429	2.189
	(0.991)	(0.486)	(0.814)	(0.893)	(0.777)
Laam	-0.418	-0.245	-0.650	-0.823	-0.308
	(0.712)	(0.370)	(0.535)	(0.575)	(0.515)
cristx	5.947	3.240	3.327	2.919	3.492
	(1.684)	(0.973)	(1.180)	(1.203)	(1.102)
musx	6.569	3.552	3.411	2.925	3.894
	(1.849)	(1.043)	(1.318)	(1.357)	(1.251)
budx	7.945	4.265	104.113	113.439	3.397
	(1.983)	(1.089)	(100634.400)	(88894.590)	(1.118)
hinx	4.905	2.641	4.568	4.106	4.713
	(2.099)	(1.165)	(1.787)	(1.823)	(1.522)
anix	0.495	-0.003	-0.020	-0.376	-0.367
	(1.571)	(0.889)	(1.079)	(1.092)	(1.003)
res. (democ)				0.656	
				(3.571)	
res. (democ2)				-0.173	
				(0.341)	
res. (MAJOc)				13.313	
				(8.845)	
res. (ltmaj)				137.976	
				(293.285)	
pred. (democ)					0.725
					(3.350)
pred. (democ2)					0.040
					(0.326)
pred. (MAJOc)					-11.084
					(7.563)
pred. (ltmaj)					-46.094
					(277.442)
Constant	-13.567	-7.582	-12.095	-0.825	0.663
	(4.956)	(2.624)	(4.205)	(6.586)	(6.236)
llik	-117.872	-117.438	-61.946	-58.594	-69.161
Pseudo r^2	0.429	0.431	0.481	0.509	0.421
LR chi2(3) =				6.700	
Prob > chi2 =				0.152	
n	555	555	298	298	298

Table 13: Misclassification: Fearon and Laitin (2003)

variables	(1) logit b (s.e)	(2) probit b (s.e.)	(3) probit b (s.e.)	(4) probit b (s.e.)	(5) probit b (s.e.)
Prior war	-0.918	-0.391	-0.402	-0.455	-0.468
	(0.223)	(0.117)	(0.112)	(0.015)	(0.121)
Per capita income	-0.345	-0.135	-0.138	-0.133	-0.131
(in 1000's, lagged one year)	(0.066)	(0.026)	(0.032)	(0.013)	(0.025)
log(population)	0.255	0.108	0.109	0.110	0.115
(in 1000's, lagged one year)	(0.069)	(0.028)	(0.028)	(0.011)	(0.017)
log(% mountainous)	0.220	0.091	0.093	0.087	0.087
	(0.078)	(0.031)	(0.037)	(0.033)	(0.032)
Noncontiguous state	0.364	0.180	0.187	0.207	0.202
	(0.200)	(0.099)	(0.117)	(0.023)	(0.101)
Oil exporter	0.890	0.352	0.358	0.372	0.380
	(0.201)	(0.113)	(0.109)	(0.040)	(0.103)
New state	1.733	0.757	0.768	0.760	0.776
	(0.204)	(0.155)	(0.138)	(0.019)	(0.101)
Instability	0.632	0.259	0.268	0.240	0.235
	(0.220)	(0.096)	(0.092)	(0.023)	(0.084)
Democracy	0.024	0.008	0.008	0.009	0.010
(Polity IV)	(0.015)	(0.006)	(0.007)	(0.006)	(0.007)
Ethnic fractionalization	0.213	0.086	0.088	0.102	0.089
	(0.122)	(0.068)	(0.090)	(0.126)	(0.109)
Religious fractionalization	0.187	0.124	0.143	0.081	0.045
	(0.325)	(0.077)	(0.100)	(0.150)	(0.189)
b0	-6.646	-3.224	-3.184	-3.239	-3.282
	(0.703)	(0.286)	(0.254)	(0.025)	(0.152)
Per capita income				-2.962	-3.148
(in 1000's, not lagged)				(0.017)	(0.522)
Eastern Europe					-1.381
(Dummy)					(3.236)
Latin America					0.078
(Dummy)					(0.055)
Sub-Saharan Africa					0.859
(Dummy)					(1.348)
Asia					-0.069
(Dummy)					(0.358)
North Africa/Middle East					0.592
(Dummy)					(1.158)
a_0			-5.742	-1.458	-1.874
			(11.696)	(0.021)	(1.146)
Per capita income				-1.436	-1.454
(in 1000's, not lagged)				(0.013)	(0.046)
Eastern Europe					0.145
(Dummy)					(0.113)
Latin America					0.100
(Dummy)					(0.013)
Sub-Saharan Africa					0.070
(Dummy)					(0.068)
Asia					0.016
(Dummy)					(0.213)
North Africa/Middle East					0.097
(Dummy)					(0.014)
a_1			-1.100	-2.045	-2.120
			(0.352)	(0.015)	(0.179)
llik	-477.330	-481.420	-481.521	-475.389	-474.250
n	6327	6327	6327	6327	6327

Table 14: Auxiliary regression Fearon and Laitin (2003)

variables	Democracy (lagged)
	b
	(s.e)
Prior war (2 lagged)	0.627
	(0.263)
Per capita income (2 lagged)	0.612
	(0.022)
log(population) (2 lagged)	-0.071
	(0.065)
log(% mountainous)	-0.060
	(0.063)
Noncontiguous state	4.913
	(0.242)
Oil exporter (lagged)	-5.171
	(0.257)
New state	1.696
	(0.522)
Instability	0.980
	(0.241)
Ethnic fractionalization	-1.514
	(0.329)
Religious fractionalization	0.203
	(0.415)
Constant	-1.819
	(0.591)
R^2	0.261
Adj R^2	0.259
root MSE =	6.507
n	6187

Table 15: Endogenous institutions: Fearon and Laitin (2003)

variables	(1) logit b (s.e)	(2) probit b (s.e)	(3) probit b (s.e)	(4) probit b (s.e)	(5) probit b (s.e)
Prior war	-0.954	-0.391	-0.373	-0.417	-0.408
	(0.314)	(0.130)	(0.130)	(0.134)	(0.133)
Per capita income	-0.344	-0.135	-0.122	-0.173	-0.164
(in 1000's. lagged one year)	(0.072)	(0.028)	(0.027)	(0.045)	(0.044)
log(population)	0.263	0.108	0.110	0.118	0.116
(in 1000's. lagged one year)	(0.073)	(0.031)	(0.032)	(0.033)	(0.032)
log(% mountainous)	0.219	0.091	0.092	0.097	0.097
	(0.085)	(0.034)	(0.035)	(0.036)	(0.036)
Noncontiguous state	0.443	0.179	0.115	-0.317	-0.294
	(0.274)	(0.122)	(0.128)	(0.328)	(0.327)
Oil exporter	0.858	0.352	0.337	0.754	0.733
	(0.279)	(0.123)	(0.127)	(0.315)	(0.313)
New state	1.709	0.757	0.760	0.629	0.644
	(0.339)	(0.163)	(0.223)	(0.241)	(0.241)
Instability	0.618	0.259	0.270	0.204	0.220
	(0.235)	(0.101)	(0.101)	(0.112)	(0.111)
Democracy	0.021	0.008	0.009	0.097	
(Polity IV, lagged one year)	(0.017)	(0.007)	(0.007)	(0.062)	
Ethnic fractionalization	0.166	0.087	0.138	0.276	0.285
	(0.373)	(0.157)	(0.162)	(0.189)	(0.188)
Religious fractionalization	0.285	0.128	0.179	0.154	0.149
	(0.509)	(0.209)	(0.215)	(0.216)	(0.216)
Democracy residual				-0.088	
				(0.062)	
Democracy predicted value					0.091
					(0.063)
Constant	-6.731	-3.224	-3.313	-3.172	-3.188
	(0.736)	(0.303)	(0.313)	(0.328)	(0.327)
llik	-480.401	-481.410	-452.067	-451.192	-451.819
Pseudo R^2	0.108	0.106	0.094	0.096	0.095
n	6327	6327	6170	6170	6170

References

Achen, Christopher H. 1986. *Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.

Barro, Robert J. and Jong-Wha Lee. 1996. "International measures of schooling years and schooling quality." *American Economic Review, Papers and Proceedings* 86(2):218–223.

Bloom, David E. and Mark R. Killingsworth. 1985. "Correcting for Truncation Bias Caused by a Latent Truncation Variable." *Journal of Econometrics* 27:131–135.

Boehmke, Frederick J., Daniel Morey and Megan Shannon. 2003. "Self-Selectivity in Duration Models." Paper prepared for presentation at the Midwest Political Science Association Annual Meeting, April 3-6, Chicago.

Carroll, Raymond J., D. Ruppert and L. A. Stefanski. 1995. *Measurement Error in Nonlinear Models*. London: Capman and Hall.

Cederman, Lars-Erik. 2004. "Articulating the Geo-Cultural Logic of Nationalist Insurgency." Paper prepared for presentation at the Joint Session of Workshops of the ECPR, Uppsala.

Christin, Thomas and Simon Hug. 2003. "Federalism and Conflict Resolution: Considering Selection Biases." Paper prepared for presentation at the Midwest Political Science Association Annual Meeting, April 3-6 Chicago.

Cohen, Frank S. 1997. "Proportional Versus Majoritarian Ethnic Conflict Management in Democracies." *Comparative Political Studies* 30(5):607–630.

Collier, Paul, Anke Hoeffler and Mans Söderbom. 2001. "On the Duration of Civil War." Washington, World Bank Policy Research Working Paper 2861.

Collier, Paul, Lani Elliott, Havard Hegre, Anke Hoeffler, Marta Reynal-Querol and Nicholas Sambanis. 2003. *Breaking the Conflict Trap. Civil War and Development Policy*. New York: World Bank and Oxford University Press.

Colomer, Josep. 2001. *Political Institutions, Democracy and Social Choice*. Oxford: Oxford University Press.

Cox, Gary W. 1997. *Making Votes Count*. Cambridge: Cambridge University Press.

Doyle, Michael W. and Nicholas Sambanis. 2000. "International Peacebuilding: A Theoretical and Quantitative Analysis." *American Political Science Review* 94(4):779–801.

Easterly, William. 2000. "Can Institutions Resolve Ethnic Conflict?" World Bank February 2000.

Elbadawi, Ibrahim and Nicholas Sambanis. 2002. "How Much War Will We See? Estimating the Incidence of Civil War in 161 Countries." *Journal of Conflict Resolution* 46(3):307–334.

Ellingsen, Tanja. 2000. "'Ethnic Witches' Brew: Linguistic, Religious and Ethnic Fragmentation, 1945-94." *Journal of Conflict Resolution* 44(2):228–249.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41(3):275–301.

Fearon, James D. and David D. Laitin. 1997. *A Cross-sectional Study of Large-scale Ethnic Violence in the Postwar Period*. La Jolla: Paper prepared for the conference "Cooperation under Difficult Conditions," University of California, San Diego.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):1–17.

Gates, Scott and Havard Strand. 2004. "Modeling the Duration of Civil Wars: Measurement and Estimation Issues." Paper prepared for presentation at the Joint Session of Workshops of the ECPR, Uppsala.

Geddes, Barbara. 1991. "A Game Theoretic Model of Reform in Latin American Democracies." *American Political Science Review* 85(2):371–392.

Gleditsch, Nils Petter, Hvard Strand, Mikael Eriksson, Margareta Sollenberg and Peter Wallensteen. 2002. "Armed Conflict 1946-99: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Gurr, Ted Robert. 1993. *Minorities at Risk. A Global View of Ethnopolitical Conflict.* Washington: United States Institute of Peace Press.

Hartzell, Caroline and Matthew Hoddie. 2003. "Institutionalizing Peace: Power Sharing and Post-civil Conflict Management." *American Journal of Political Science* 47(2):318–332.

Hausman, Jerry. 1978. "Specification Tests in Econometrics." *Econometrica* 46(6):1251–1271.

Hausman, Jerry, Jason Abrevaya and Fiona Scott-Morton. 1998. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87:239–269.

Hegre, Håvard. 2004. "The Duration and Termination of Civil War." *Journal of Peace Research* 41(3):243–252.

Hegre, Håvard, Tanja Ellingsen, Scott Gates and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992." *American Political Science Review* 95(1 (March)):33–48.

Hug, Simon. 2003. "Selection Bias in Comparative Research. The Case of Incomplete Datasets." *Political Analysis* 11(3):255–274.

Hug, Simon. 2005. "The Effect of Misclassifications in Probit Models. Monte Carlo Simulations and Applications." St. Gallen: IPW Universität St. Gallen.

Kalyvas, Stathis. 2002. *From The Logic of Violence in Civil War.* Chicago: unpublished book manuscript.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge: Cambridge University Press.

King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry. Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Lacina, Bethany. 2004. "Why Civil Wars Come in Different Sizes." Centre for the Study of Civil War (CSCW) International Peace Research Institute, Oslo (PRIO).

Lacina, Bethany and Nils Petter Gleditsch. 2004. "The Human Cost of War? A New Dataset of Battle Deaths." Centre for the Study of Civil War (CSCW) International Peace Research Institute, Oslo (PRIO).

Lijphart, Arend. 1999. *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven: Yale University Press.

Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Mansfield, Edward D. and Jack Snyder. 1995. "Democratization and War." *Foreign Affairs* 74(3):79–97.

Muthen, Bengt and Karl G. Jöreskog. 1983. "Selectivity Problems in Quasi-Experimental Studies." *Evaluation Review* 7(2):139–174.

Panizza, Ugo. 1999. "On the Determinants of Fiscal Centralization: Theory and Evidence." *Journal of Public Economics* 74(1):97–139.

Quinn, Kevin, Michael Hechter and Erik Wibbels. 2003. "Ethnicity, Insurgency, and Civil War Revisited." Laboratory in Comparative Ethnic Processes 7 May 9 - May 11, 2003 Yale University - New Haven.

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

Reynal-Querol, Marta. 2002*b*. "Political Systems, Stability and Civil Wars." *Defence and Peace Economics* 13(6):465–483.

Rivers, Douglas and Quang H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." *Journal of Econometrics* 39(3 NOV):347–366.

Saideman, Stephen M., David J. Lanoue, Campenni Michael and Samuel Stanton. 2002. "Democratization, Political Institutions, and Ethnic Conflict:

A Pooled Time-Series Analysis, 1985-1998." *Comparative Political Studies* 35(1, February):103–129.

Saideman, Stephen M. and R. William Ayres. 2000. "Determining the Causes of Irredentism: Logit Analyses of Minorities at Risk Data from the 1980s and 1990s." *Journal of Politics* 62(4 (Nov)):1126–1145.

Sambanis, Nicholas. 2001*a*. "Do Ethnic and Nonethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Inquiry (Part I)." *Journal of Conflict Resolution* 45(3):259–282.

Sambanis, Nicholas. 2001*b*. "A Note on the Death Threshold in Coding Civil War Events." Paper prepared for presentation at Laboratory in Comparative Ethnic Processes 4 October 26 - 28 Dartmouth College.

Sambanis, Nicholas. 2002. "A Review of Recent Advances and Future Directions in the Literature on Civil War." *Defense and Peace Economics* 13(3 (August)):215–243.

Sambanis, Nicholas. 2004. "Using Case Studies to Expand Economic Models of Civil War." *Perspectives on Politics* 2(2):259–279.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions." *Political Analysis* 11(1):111–138.

Snyder, Richard. 2001. *Does Lootable Wealth Breed Disorder? States, Regimes, and the Political Economy of Extraction.* San Francisco: Paper prepared for presentation at the Annual Meetings of the American Political Science Association, San Francisco.

Van de Walle, Nicolas. 2004. The Economic Correlates of State Failure: Taxes, Foreign Aid, and Policies. In *When States Fail: Causes and Consequences*, ed. Robert I. Rotberg. Princeton: Princeton University Press pp. 94–115.

Walter, Barbara F. 2002. *Committing to Peace. The Successful Settlement of Civil Wars.* Princeton: Princeton University Press.