PREDICTING CIVIL CONFLICTS: ON THE UTILITY OF EMPIRICAL RESEARCH

MICHAEL D. WARD AND KRISTIN BAKKE

ABSTRACT. Large-N studies of conflict produced a large number of statistically significant results, but very little accurate guidance in terms of anticipating the onset of conflict. There are several potential reasons for this. We provide a survey of three major efforts to understand the onset of civil conflict. We use prediction to gauge the effectiveness of these statistical models. We offer some lessons for the use of case control methods and forecasting as *theoretical* endeavors in the study of conflict.

Tis easy to see, hard to foresee, Ben Franklin

1. INTRODUCTION

Following three years of failed negotiations between the separatist government in Chechnya and the central authorities in Moscow, on December 31, 1994, Russian troops stormed Grozny, the republic's capital.[1] During the following months and years, tens (probably hundreds) of thousands of people have lost their lives and homes in this civil war.[2] News reports of deadly attacks and kidnappings remain common even today, and much of the Chechen countryside, as well as villages, towns, and the capital lie in ruins.

The Chechen republic's relationship to the federal capital has been tense at least since November 1, 1991, when the newly elected president and head of the Chechen nationalist movement, Dudayev, issued a declaration of independence. But even before that, the Chechens were among the ethnic groups most resistant to incorporation into tsarist Russia. Stalin's deportation of virtually the entire population of the republic in 1944 did little but exacerbate existing tensions. The Chechen nationalist movement that emerged in Grozny in the late 1980s sought to revive Chechen culture and traditions, as well as oust the republic's communist-led Supreme Soviet, which most Chechens considered corrupt. The failed coup against Gorbachev in August 1991 helped solidify the nationalist movement and even

Date: February 26, 2005.
Authors' Note: Presented at the Conference on Disaggregating the Study of Civil War and Transnational Violence, University of California Institute of Global Conflict and Cooperation, San Diego, CA, USA, 7-8 March 2005. Thanks to Kristian Skrede Gleditsch and Håvard Strand for their guidance on this project. This project was partially supported by by a grant from the Methods, Measurement, and Statistics Program at the National Science Foundation, grant number: SES-0417559 and by a Grant from the Human and Social Dynamics cross-directorate program of the National Science Foundation, HSD-04333927. We would like to thank Jay Ulfelder and James Fearon for their helpful comments.

[1]The following account is based on Tishkov (1997), Dunlop (1998), and Evangelista (2002).

[2]Exact numbers of both deaths and kidnappings vary widely, largely pending on who is counting–the Russian Government, Chechen sympathizers, or international organizations. Estimates very between 40,000 and 200,000 (page 103 Kahn 2002). In 2003, Human Rights Watch reported that more than 100,000 Chechen refugees lived in camps in neighboring Ingushetia. See "Into Harm's Way: Forced Return of Displaced People to Chechnya," Human Rights Watch, 01/29/03, available online at `hrw.org/reports/2003/russia0103/`.

brought Moscow to be briefly supportive of its plans since–unlike the communists–the nationalists took a clear stance against the attempted coup. However, Moscow quickly found the nationalists to be more than they had bargained for. Shortly after Dudayev declared Chechnya independent, Yeltsin issued a state of emergency decree and dispatched 2,500 Interior Ministry troops to the region. Yeltsin did not follow through with the threat of force and withdrew Russian troops the same day they arrived in Chechnya. Dudayev used this opportunity to mobilize additional support for independence among Chechens.

Within Chechnya, discontent with Dudayev's increasingly authoritarian rule grew. In the autumn of 1994, opposition forces briefly seized control of Grozny, but they were soon defeated by Dudayev's troops. Meanwhile, outside Chechnya, hijacking and kidnapping in the border regions intensified Moscow's perception or, arguably, pretext that not only was the republic becoming lawless, the situation was increasingly representing a broader security threat. Thus, on November 30, 1994, Yeltsin sanctioned the use of direct military force. Moscow's use of force seemed to spur mobilization around the Chechen government's demands for independence. As Grozny fell into ruins in early 1995, the Chechen fighters retreated to the mountains, from where they continued guerrilla campaigns that proved far more difficult to combat than what Moscow had counted on. In a humiliating defeat during the winter of 1996 the Russian troops withdrew from Chechnya and a peace settlement was reached.

In 1997, Chechen Army Chief of Staff Aslan Maskhadov was elected president and signed a treaty with Yeltsin. Maskhadov's goal was to establish a nominally independent Chechnya cooperating economically and politically with Russia, but the treaty left the future status of the republic open. The peace treat did not bring with it peace. Spurred by growing anti-Moscow opposition within Chechnya, Moscow apartment bombings attributed to Chechen terrorists, and a Chechen attack on neighboring Dagestan, Russian troops again entered Chechnya in September 1999, resulting in the eruption of the second Chechen war, which continues today.

Given the rapidly growing literature on civil wars, how do we explain the Chechen wars? Why did this center-region relationship turn violent? If we take as our point of departure influential, published studies of recent years (Collier & Hoeffler 2004, Fearon & Laitin 2003*b*), we would expect that factors such as abundance of and reliance on natural resources (oil), significant unemployment and poverty, mountainous terrain, and political instability would contribute to the outbreak of civil war. These factors are certainly present in the Chechen case:

- Prior to the 1990s, oil extraction and processing of crude oil from other regions in Russia constituted Chechnya's main profit source, although the oil industry was in rapid decline in the early 1990s.
- Unemployment was rising in the early 1990s, and Chechnya had never been a particularly prosperous republic.
- Geographically, the republic is located in the North Caucasus, and the southern half of the republic is very mountainous.
- In the early 1990s, the Russian state overall was going through tremendous changes both politically and economically, attempting to democratize and shift from a planned to a market economy at the same time. Likewise, within Chechnya, the first real elections were held in 1991, bringing and end to decades of one-party rule.

The problem, however, is that given these factors, violent conflict could equally well have broken out in any other poor, oil-dependent, and mountainous region in this highly volatile period of regime disintegration. Why Chechnya and not Dagestan?

2. What Causes Civil Wars?

War is God's way of teaching Americans geography, Ambrose Bierce

In the post-Cold War period, scholars have increasingly paid attention to intrastate conflicts. Although not really "discovered" in the international relations literature until the 1990s, the prevalence of such conflicts is part of a longer post-World War II trend. The Correlates of War data show that of all civil wars waged between 1816 and 1997, nearly 45 percent took place after 1950, causing the deaths of more than 10 million people–a number that does not even take into the accounts the deaths in the aftermaths of these wars (Sarkees, Wayman & Singer 2003, Ghobarah, Huth & Russett 2003). Different data bases indicate that, until the 1990s, there was an upward trend in the number, duration, and inherent violence of both major and minor conflicts, including ethnic conflicts (e.g. Sarkees, Wayman & Singer 2003, Fearon & Laitin 2003*b*, Gleditsch, Wallensteen, Eriksson, Sollenberg & Strand 2002, Gurr 2000). In response to these trends, the literature on intrastate conflicts has grown substantially. We provide a brief overview of the major findings in this literature and point out some of its shortcomings. Our main claim is that the relatively poor predictive ability of some of the literature's dominant models can be corrected by incorporating a greater amount of contextual factors.

While international relations scholars began to study intrastate conflicts relatively recently, scholars have long been interested in domestic political violence (e.g., Gurr 1970). The now standard variables in any conflict analysis are legion. One of these "usual suspects" is inequality. Muller & Seligson (1987), for example, argue that income inequality may lead to violent political mobilization on the part of the poorer party. Gurr's theory of relative deprivation (1970) posits that is the discrepancies between peoples' perceptions and what they actually have that leads to violent political mobilization. The mechanism in this line of argument concerns people's grievances, but because such emotions are difficult to measure in any type of large-n analysis, most scholars use Gini coefficients to assess levels of income inequality. Inequality and relative deprivation as causal factors have been challenged both on theoretical and empirical grounds. Theoretically, inequality and deprivation, which are relatively constant factors in many societies, have a hard time explaining variation in conflict patterns. Empirically, major statistical studies in the 1990s have found little support for a relationship between inequality and intrastate conflict (e.g., Fearon & Laitin 2003*b*, Collier & Hoeffler 2004).

Contemporary research that looks specifically at separatist conflicts has proposed a different take on the relationship between inequality and conflict: Focusing on the concentration of wealth at the sub-national level, scholars such as Alesina, Spolare & Wacziarg (2000) and Fearon & van Houten (2002) argue that political mobilization for independence erupts among relatively wealthy groups within the state. Such groups, these scholars argue, find that they can "make it" in the international system without the central state, and their access to resources eases the mobilization process.[3] While this line of argument has moved away from a grievance mechanism, it does suggest that inequality–in particular, group inequality at

[3]See resource mobilization theories, such as McCarthy & Zald (1977).

the sub-national level–merits further investigation. Likewise, Østby (2005) suggests that the negative findings linking inequality and intrastate conflict is due to measures that do not capture the group aspect of the argument: While large-n studies use Gini coefficients for individual income inequality, the theoretical arguments are about group inequality.

Influenced by a comparative literature on nationalism (e.g., Gellner 1983, Smith 1986, Hechter 2000), many studies have looked at ethnic diversity as a source of intrastate conflict. Theories about how ethnic identities may encourage intrastate conflicts fall into different categories based on the causal mechanisms they put forth: Some argue that ethnic identity contributes to conflict because of relatively long-standing hatreds and resentment towards ethnic groups different from one's own (e.g., Petersen 2002), which may stem from economic discrepancies among different groups (Gurr 2000). Others suggest that manipulative political leaders stir up hostility among ethnic groups thereby fueling an ethnic security dilemma. Fearon & Laitin (2000) point to the rôle of political entrepreneurs in constructing identities to achieve various ends–be they material or non-material (Gagnon 1997). It is also suggested that inter-ethnic hostility rests on social psychology and favoritism for one's own group. Social identity theorists suggest that because people seek to see themselves in a positive light, they compare their own group with relevant other groups. The implication is that conflicts can arise out of inter-group relations even where there are no apparent material conflicts of interest (e. g., Hewstone & Greenland 2000, Brown 2000). The most well-known study in this tradition argues that the initiators of violence in ethnic conflicts are most often "backward" groups that are driven by both fear and a wish to boost their self-esteem and, thus, seek to catch up with the more developed groups (Horowitz 1985).

Most of these theories share the notion that the presence of territorially concentrated ethnic groups helps solve the collective action problems associated with protest and organized violence (e.g., Toft 2003). In large-n statistical analyses these ideas tend to be addressed by including a measure of ethnic fractionalization. However, in contrast to this long-standing literature and numerous case studies, recent large-n studies of intrastate conflict have found scant statistical support for the hypothesis linking ethnic diversity and intrastate conflict (e.g., Fearon & Laitin 2003*b*, Collier & Hoeffler 2004)

What independent variables are highlighted in recent empirical works? Both Collier & Hoeffler (2004) and Fearon & Laitin (2003*b*) reject theories based on material and ethnic grievances. Instead they propose that intrastate conflicts are largely driven by the opportunities for insurgency.[4] Collier and Hoeffler argue that civil wars and rebellions are explained "not by motive, but by the atypical circumstances that generate profitable opportunities" (2004: 564). In particular, according to this model, civil wars occur where and when rebel groups have the opportunity to raise revenues, most commonly where and when these groups are able to exploit (loot) natural resources; where and when they can take advantage of high levels of unemployment and poverty and, thus, readily available rebel recruits; and where and when they have ethnic diaspora willing to financially support them. In this view, rebels are indistinguishable from criminals and bandits; they are simply rational agents seeking profits, driven by greed rather than grievances. Research in this vein has looked particularly at the

[4]The comparative literature on revolutions and rebellions drew lessons from rational choice theory, (vide Popkin 1988, Lichbach 1994). See Gurr (2000) for opportunity structures of rebellion, and McAdam, John D & Zald (1996) for the development of the concept of opportunity structures.

relationship between conflict and abundance of natural resources such as oil, timber, minerals, and drugs–the so-called resource-predation hypothesis–suggesting that such resources provide both an incentive for rebellion and means to fund rebellions (e.g., de Soysa 2002).

Fearon & Laitin (2003*b*) also find that oil production is positively correlated with intrastate conflict, but they reject the resource-predation hypothesis and suggest that the causal link here is about state weakness: Abundance of oil and dependency on oil exports, they argue, encourage weak state institutions. Sharing Collier and Hoeffler's rejection of grievance-based arguments, Fearon & Laitin (2003*b*) propose a model that focuses on the state. Their central argument is that intrastate conflicts in the post-1945 era are a result of favorable insurgency conditions; namely, circumstances that ease subnational mobilization by limiting the central state's ability to control its territory. These conditions include mountainous terrain, large populations, political instability, the newness of the state, and low levels of economic development. Notably, although Fearon and Laitin's (2003*b*) argument concerns state strength, the study includes no indicators for state institutions besides democracy, which ends up as a statistically insignificant finding. Their main indicator of state strength is GDP per capita, and whereas they interpret this as support for their argument that state weakness encourages civil war onset, Collier and Hoeffler take the same finding to indicate that poverty encourages civil war onset.

The "mismatch" of interpretation is representative of two broader problems with the emerging civil war literature: First of all, as both Sambanis (2004) and King (2004) have argued, the growing literature on civil wars has paid insufficient attention to causal mechanisms. This is largely a result of the analysis often taking place at the national level and across a large number of cases. The result is that the literature establishes correlations, but it is hard pressed to give any type of policy advice, as the causal mechanisms are–as in the case above–either contradictory or unsubstantiated. Sambanis puts this nicely when he points out that case studies reveal that both Collier and Hoeffler and Fearon and Laitin "... are often right for the wrong reasons yet also wrong for the wrong reasons" (2004: 260). Second, near absent from this literature is a consideration of the rôle of the state and its institutions–political variables. The only institutional variable that has been seriously considered in this literature is democracy–whether a state is democratic, authoritarian, or in transition between the two (for example, Snyder 2000, Hegre, Ellingsen, Gates & Gleditsch 2001) Given that the literature on revolutions and political violence long have argued that the state is a key explanatory factor (among others, Skocpol 1979), this is a rather curious omission. Recent research in the comparative politics and sociological vein has suggested that the state's actions, policies, and institutions may help construct sub-national challengers (Rasler 1996, Goldstone & Tilly 2001, Goodwin 2001). A growing comparative literature is focusing specifically on how institutions of decentralization may either exacerbate or defuse intrastate conflicts.[5] Indeed, this research suggests that if we are interested in predicting conflict, we need to understand the rôle of state institutions, as well as how these institutions may have different effects in different societies. Given the spare record of models in predicting specific conflicts, it might be worth trying to"bring the state back in" in conflict analysis through a wider range of variables.

[5]See works by Suny (1993), Bunce (1999), Hechter (2000), Treisman (2001), Bermeo (2002), Saideman, Lanou, Campenni & Stanton (2002), Milanovic & Sambanis (2003), Bakke & Wibbels (2004), and Filippov, Ordeshook & Shvetsova (2004).

We turn now to an examination of how case control designs can reduce slippage and increase traction between theoretical and empirical materials in the study of intrastate conflicts.

3. Case Controls

> *In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. ... any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long ...,*
> Mark Twain (1884, p. 58) *Life on the Mississippi*

While there is a growing number of experimental studies of social phenomena, most of these experiments are not natural experiment, nor are they strictly controlled. Laboratories and computer compilers provide the main arena in which the experiments are undertaken. As a result, most studies of civil conflict are observational studies of data that are not generated experimentally. Unfortunately, these data are typically examined as if they were generated by experimental methods. Case control techniques have been developed for such situations, but they are rarely used in the conflict literature. Imai & van Dyk (2004) introduced and developed the broad notion of using propensity scores as a means of managing sample matching in parametric studies that are not experimental. The basic idea is to select a subset of the observed data for which a treatment and the covariates are not related. As a result the treatment and control groups will have the same characteristics; i.e., they are *matched* as if they had been assigned randomly from a single population. Ho, Imai, King & Stuart (2004) have developed software, called MatchIt, which implements these procedures to produce matched subsamples. Once the matched subsamples are produced, normal parametric model fitting is undertaken.

Why using matching in studies of civil conflict? First, we know that the data are often produced by looking a the existing list of countries and then including every other plausible, available country-year as additional observations. This means that we typically have observational data in which the dependent variable is highly skewed, if not absolutely rare. But it is not clear that each additional country-year actually adds information. More importantly, many of the important variables can easily be thought of as "treatments." There are many examples, as most empirically oriented studies include many covariates that do not easily change, were pre-existing, and are categorical or binary. Whether a country is an "Oil exporter" for example serves as an important covariate in many studies. In the study by Fearon and Laitin (2003*b*) instability of the government is seen as a pre-existing binary condition based on whether there has been substantial change in the authority patterns of governance over a prior three year interval. Perhaps an additional reason to examine this approach to understanding the model dependency of extant results is that the current standard approaches all produce empirical results which are rife with "statistical significance" but which do not allow one to correctly predict the onset of civil conflicts within the observational data. Leaving aside for the moment, the issue of out-of-sample validation, failure to successfully identify in-sample onsets of civil wars suggests strongly that there is a great deal of model dependence, plausibly even overfitting.

We re-examine three studies of civil conflict, illustrating the dependence of their results on the observational data. We then show that case control methods lead to different results, both in terms of statistical significance and in terms of predictions.

4. Mountains

The man who removes a mountain begins by carrying away small stones.
William Faulkner

One of the most widely cited, recent studies of civil conflict is Fearon & Laitin (2003*b*) which produced two startling results. One of these was that civil conflict was a product the geophysical environment to the extent that countries with mountainous terrain were more likely to witness civil conflicts. The second was that ethnic and religious cleavages did not demonstrate any relationship with the onset of civil conflict. This research was based on using logistic regression to determine the risk factors that historical and contemporary countries face in terms of the onset of civil conflict.[6]

We explored further imputation of the missing information in the data as provided by Fearon and Laitin. This was accomplished by using some sources updated since the collation of data for the 2003 article, as well as some archival work, along with some multiple imputation. These issues are all discussed in gory detail in the Appendix. Our goal was a database without missing data, but one that faithfully replicated the original results in the Fearon and Laitin study (2003). Table 4 illustrates in the first two columns that the original and the imputed data produce virtually identical results.

Beyond the replication and imputation issues, Table 4 illustrates the results of estimating the base model subject to two, separate matching procedures. Matching is explained in Ho et al. (2004). We matched, separately, on two different conditions. First we examined the existence of a prior war as a "treatment." Subsequently, we examined the existence of instability, which measures whether there was prior, recent substantial change in the form of government. Four differences are noteworthy in the results of the parametric model (a logistic regression) after sub-sample matching is undertaken. Mountainous terrain does not appear to be important once matching is undertaken. Democratic governance does seem to gain importance in the matched samples, unlike in the full sample. The prior instability of the governance structure is not shown to be important in the matching sub-samples. Finally, whereas the baseline model is unable to predict any in-sample civil war onsets, the models generated under matching assumptions are able to correctly predict a few onsets.

What is the predictive success of this basic model? The basic results from the Fearon and Laitin specification contains eight variables deemed "statistically significant" in the original article. However, when the *in-sample* predictions are examined, it turns out that the model never produces a single predicted probability that exceeds the standard 0.5 cut point.[7] Stated differently, based on this model one would always predict the absence of civil conflict; however, there are actually more than one-hundred such conflicts in the database. If we simply took the modal value in the observed distribution, we would generate the exact same *in-sample* predictions as the basic model estimated in the Fearon and Laitin study. Even in the model which is run on replicated and imputed data, the estimated model always predicts the absence of the onset of civil war, despite the presence of 107 such onsets in the

[6]Actually, there is an appendix available on James Fearon's Web site which illustrates the stability of their findings with a variety of other techniques, including Poisson, Cox regression, conditional fixed effects Logit, as well as numerous different specifications (Fearon & Laitin 2003*a*).

[7]Choosing an appropriate cutpoint *prior* to analysis is an important task for which there is no recipe. Should it be the base rate? Should it maximize correct predictions? Models that always predict onsets will be more successful in correctly identifying civil war onsets, using this approach, but they will almost always be wrong. Should it balance false positives against false negatives? We do not delve deeply into this issue herein.

Table 1: *Replication of Fearon and Laitin (2003: Table 1, column 1, page 84). These results are identical to those published, with the exception of the coefficient for the "new state" variable which is different by 0.001. Variables deemed statistically significant (at a level of 0.05 or smaller) by Fearon and Laitin are shown with a trailing ✓ in column one; similar notation has been applied to the three other sets of estimates for comparison purposes. A more complete database was generated via imputation. Estimates for the base model using these data are given in column two. The complete, imputed data was used in the two matching subsamples.*

			Matched on:	
	Replication	Imputations	Prior War	Instability
Prior war	−0.954✓	−0.926✓	−1.024✓	−1.045✓
Per Capita income	−0.344✓	−0.34✓	−0.423✓	−0.368✓
Population, logged	.263✓	.267✓	.229✓	.224✓
Mountainous, logged	.219✓	.23✓	.272	.184
Noncontiguous state	.443	.406	.147	.064
Oil exporter	.858✓	.881✓	1.386	1.244✓
New state	1.71✓	1.58✓	2.309✓	2.599✓
Instability	.618✓	.596✓	.232	.22
Democracy	.021	.019	.051✓	.05✓
Ethnic fractionalization	.166	.088	−0.246	−0.176
Religious fractionalization	.285	.398	.644	.383
Constant	−6.73✓	−6.81✓	−6.15✓	−5.77✓
Correctly Predicted Onsets	0 (110)	0 (110)	2 (53)	3 (53)

sample. To be sure, the base mode of civil war onset in the entire, available country-year sample *is* in fact zero. However, the matching process produces somewhat different results, correctly predicting two (of 53) civil war onsets for the model matched on prior wars, and correctly predicting three civil war onsets (of 53) in the case of the model matched on the prior instability of the regime. These successes may seem modest, but can also be viewed as huge *marginal* increases in our ability to understand and predict civil war onsets. At least writing down zero is no longer the best predictive model: matching helps generate statistical models that have greater predictive validity without giving up a concern with "statistical significance" in situations that neither involve samples nor (quasi)experiments.

5. Greed and Grievance

> *The problem of social organization is how to set up an arrangement under which greed will do the least harm, capitalism is that kind of a system.* Milton Friedman

Collier & Hoeffler (2004) have also made fundamental contributions to the growing literature on civil war. Fearon (2005 forthcoming) provides a nice summary of coverage of these contributions in the popular press. Collier and Hoeffler have posed two basic explanations

for the onset of civil wars. One is based on greed; the other on grievance. The greed explanation focuses on economics and the opportunity that exists to create a sustainable rebellion, in particular the extraction/extortion of natural resources via primary commodity exports, donations from diaspora, subventions from hostile governments via proxy during the Cold war, government weakness, the extent of alternative opportunities besides rebellion, the presence of geographical havens, and the extent of social and ethnic cohesion. The grievance hypothesis takes form in terms of measures of ethnic and religious hatred, political repression, political exclusion, and economic inequality. The basic conclusion they reach is that greed is a better explanation of civil war than grievance, though they acknowledge that greed may overlap in some ways with grievance.

Collier and Hoeffler use logistic regression to conduct a "horse race" between these two broad explanations. As a result their conclusions are almost entirely influenced by their data and the discovery process they employed to parse through those data. As they note (page 563): they "use econometric tests to discriminate between rival explanations and develop and integrated model which provides a synthesis." We re-examine the results of this careful examination, focusing first on their opportunity model, specifically reported in column 3 of Table 1 (2003, page 573). Using their data, we re-estimated their basic model, the results of which are shown in Table 2. The Brier score is approximately 0.05, representing a reasonably good fit.[8]

Table 2: *Replication of the Collier and Hoeffler (2003) Opportunity model. A ✓ is used to flag variables deemed significant at the* 0.05 *level by Collier and Hoeffler. These replicated results are identical to those published in the original study. Not originally reported, onsets are predicted if the estimated probability exceeded* 0.50

		Predictions Omitting Row Variable	
	Estimate	False Onsets	True Onsets
Primary exports/GDP	16.476✓	5	4
Squared Primary exports/GDP	−23.017✓	5	5
Post-cold war	−.454	4	4
Ln GDP per capita	−.837✓	2	5
GDP growth	−.105✓	1	0
Peace Duration	−.004✓	2	2
Mountainous Terrain	.008	2	4
Geographic dispersion	−.865	2	4
Social fractionalization	−.0002✓	3	3
ln Population	.493✓	3	2
Intercept	−4.068✓	3	4
N	750	3/52	4/52
Brier Score	.054	False Onset	True Onset
Area under the ROC Curve	.836		

[8]The Brier score is a common method for comparing probabilistic predictions with binary outcomes, and is the average squared probabilistic error. It can be thought of as analogous to root mean squared error in the context of such comparisons. A lower score indicates better predictions (Brier 1950).

The second part of Table 2 not only replicates the original analysis of Collier and Hoeffler, but it also utilizes in-sample forecasting to assess the significance of the variables in this specification. What we do is to iteratively re-run the model *without* each variable. Then we use the re-estimated model to generate predictions, focusing mainly on correct predictions of civil war onsets and false predictions of civil war onsets. We focus on these, rather than the correct and incorrect predictions of nonconflict, since this is area in which such models are typically deficient. Only two variables stand out as making a predictive improvement: GDP growth and peace duration. These two variables improve the correct identification of onsets while reducing the incorrect identification of onsets (false positives). All other variables either produce the either better or equivalent results when they are excluded from the equation. Even variables that show statistical significance in a classical sense can produce uniformly worse predictions. If one were looking for variables that would absolutely increase the ability to correctly identify the onset of civil war, regardless of the costs of false positives, then GDP growth, peace duration, social fractionalization, and the logged population would satisfy this criterion. These variables also meet the statistical criteria utilized by Collier and Hoeffler. However, neither of the variables which are used to assess the exploitation of natural resources help to improve the predictive performance of the model. In terms of Brier scores, all the variables have higher Brier scores when they are excluded from the equation, except geographic dispersion, which yields a Brier score of 0.052 when it is *excluded* from the formulation.

We re-estimate the base model using matching on the post-cold war variable as a treatment. These results are presented in Table 3.

Table 3: *Replication of the Collier and Hoeffler (2003) Opportunity model. A ✓ is used to flag variables deemed significant at the* 0.05 *level by Collier and Hoeffler. The second column of results presents the estimates based on the subsample which is created via matching on the post-coldwar treatment.*

Variable	Full Sample	Matched: Post Cold war era
Primary exports/GDP	16.476✓	5.425
Squared Primary exports/GDP	−23.017✓	−1.664
Post-cold war	−.454	−0.154
Ln GDP per capita	−.837✓	−1.555✓
GDP growth	−.105✓	−0.079
Peace Duration	−.004✓	−0.001
Mountainous Terrain	.008	.024
Geographic dispersion	−.865	.956
Social fractionalization	−.0002✓	.0001
ln Population	.493✓	.482✓
Intercept	−4.068✓	−0.740
N	750	420
Brier Score	.054	.04
Area under the ROC Curve	.836	.868
Correctly predicted onsets	4/52	3/22

On the basis of the J-test, Collier and Hoeffler assert (p 577) that the opportunity/greed model is superior to the grievance model. However, they conclude that: "...while the opportunity model is superior, some elements of the grievance model are likely to add to its explanatory power...[and] ...we therefore investigate the combination of the two models as presented in column 3 of Table 5." Table 4 illustrates the results of the combination, along with an assessment of its predictive fit. What is remarkable about this result is that the combination has nine "significant" explanatory variables, but only predicts a single onset correctly, though it does not produce any false positive predictions.

Table 4: *Replication of the combined model of greed and grievance model presented in Collier and Hoeffler (Column 3, Table 5, 2003). A ✓ is used to flag variables deemed significant at the* 0.05 *level by Collier and Hoeffler. These replicated results are identical to those published in the original study. Not originally reported, onsets are predicted if the estimated probability exceeded* 0.50.

"Significant" Variable	Estimate	"Insignificant" Variable	Estimate
Primary exports/GDP	37.072✓	Post-cold war	−.873
Squared Primary exports/GDP	−69.270✓	Peace Duration	.0003
Male secondary Schooling	−.029✓	Mountainous Terrain	.005
GDP growth	−.045✓	Religious fractionalization	.015
Geographic dispersion	−4.032✓	Polarization	−25.276
ln Population	.927✓	Democracy	−.018
Social fractionalization	−.001✓	Income Inequality	.025
Ethnic fractionalization	.041✓	Intercept	−18.246
Ethnic Dominance	2.020✓		
N	479	Brier Score	.05
Area under the ROC Curve	.85	Correctly Predicted Onsets	1

Collier and Hoeffler also examine in some detail a subset of the various combinations of "significant" and "nonsignificant" variables in this combination finally settling on a version they baptize *baseline*. Using a probability cutpoint of 0.5, this model produces three correct onset predictions along with five false positives.[9] Thus, it would appear that the synthesis produced by the datamining is worse at predicting conflicts, in spite of having *nine* variables that are statistically significant.[10]

If we think of this as data mining, then the usual significance tests are valid if and only if each sample is used only once. Without this constraint–and using the same data over and over in myriad regressions–a flawed data mining approach dominates the extant literature and is palpable in many of the canonical articles, for example: " We examined a number of different model specifications. We found that none of the following geographic and demographic characteristics were *significant* [emphasis added]: forest coverage, population density, and the proportion of young men aged 15 to 29. We also investigated the potential endogeneity

[9]A cutpoint of 0.5 implies that the costs of a false positive prediction are equally "costly" to the decision maker as a false negative prediction.

[10]Fearon (2005 forthcoming) illustrates that the significant finding on primary commodity exports is not robust.

of income to civil war. ... To control for this we re-estimated excluding repeat wars. The income variable remained highly significant" (Collier and Hoeffler, 2004, page 587).

The basic problem is that the scholarly literature continues to worry about which variable is significant, while actually predicting none of the actual events correctly. Datamining is intended not to establish statistical significance, but rather to establish predictive performance. Unfortunately, the empirical literature on civil war has mistaken the methods for the goal. As a result there is an enormous mismatch between what we know and what we could know. Not only are our models dependent on our samples and sampling perspective, but they are also conditional on an inability to correctly predict outcomes. We return to this theme subsequent, but the next section re-examines a government sponsored project to examine political instability.

6. Fragile States & the State Failure Project

> *Mieux vaut prévoir sans certitude que de ne pas prévoir du tout*, Henri Poincaré, *La Science et l'hypothèse, IV Partie, La Nature, Chapitre IX*

In 1994 the intelligence and warning community in the U.S. government commissioned a quantitative study of "state failures," with an eye to being able to identify states likely to encounter serious political instability.[11] This process was conceptualized to include events ranging from Islamic revolutions and ethnic wars, through genocide, the collapse of the Soviet Union, on to revolutionary conflicts in Africa. Four types of state failure–revolutionary wars, ethnic wars, adverse regime changes, and genocide and politicide–comprise a superset of political instabilities, and include most, if not all, civil wars (King & Zeng 2001) that have served as the focus of other research in this area. Several interim reports generated a considerable amount of discussion and criticism. This resulted in a refinement of the project methods, and resulted in an unusual re-design of the project (Goldstone, Gurr, Harff, Levy, Marshall, Bates, Kahl, Surko, Ulfelder & Unger 2000).[12]

Two major, unique aspects of this project merit emphasis. One of the main goals is to identify the factors that are most closely associated with being able to correctly classify (predict) state failure.[13] Second the most recent (available) project used random case control methodology: "The core of the Task Forces method is random case control comparisons. This technique examines conditions in "failed" countries two years before the onset of failure and compares them with conditions in a randomly selected set of "control" countries, matched by year and region, that did not fail in the ensuing years. This comparison allows

[11]From the project's Web site: "The information was compiled as part of an unclassified study that was commissioned by the Central Intelligence Agency's Directorate of Intelligence in response to a request from senior US policy makers to design and carry out a data-driven study on the correlates of state failure since the mid-1950s. The study was carried out by a Task Force consisting of academic experts, data collection and management specialists from the Consortium for International Earth Science Information Network (CIESIN), and analytic methods professionals from Science Applications International Corporation (SAIC). It must be noted that, although the work of the Task Force was funded by the CIA's Directorate of Intelligence, none of the information contained in this Web site, associated data resources, or Task Force reports is based on intelligence reporting or classified material. The information presented also does not represent the official view of the U.S. Government, the U.S. Intelligence Community, or the Central Intelligence Agency, but rather the views of the individual authors and researchers themselves."

[12]This project is ongoing as of 2005; declassified and open documents and data are available at `www.cidcm.umd.edu/inscr/stfail/` accessed on Monday, February 14, 2005 at 8:53 am.

[13]Though there was some confusion about this initially (see, for example Esty, Goldstone, Gurr, Surko & Unger 1995, Esty, Goldstone, Gurr, Harff, Levy, Dabelko, Surko & Unger 1998).

Table 5: *Comparison of Replicated and Estimated Global Model Coefficients. The original results are taken from Table A-7, page 91 of the Phase III State Failure Task Force Report (2000). A ✓ is used to flag coefficients that have associated p-values of* 0.05 *or smaller.*

Variable	Replicated	Original
(Intercept)	−2.002	NA
Infant mortality rate, logged	.795✓	.799✓
Partial democracy indicator	2.022✓	2.012✓
Democracy indicator	.04	.052
Trade openness, dichotomized at median (55)	.563✓	.519
Total population, logged & normalized	.182	.188
Population density, logged & normalized	.218✓	.217✓
Bordering conflict states, dichotomized at 2	.648✓	.645✓
number of cases	456	456
Brier Score	.15	NA
Area under ROC Curve	.78	NA
Correctly Predicted Onsets	82/114	83/114
Incorrectly Predicted Onsets	32/114	31/114

us to identify those conditions most closely associated with state failure and to estimate the impact of differences in those conditions on the risk of failure. This technique is considered particularly appropriate for analysis of rare events, where analysis of time-series, cross-sectional data tends to produce biased estimates and may overstate the significance of variables that fluctuate over time" (2000, page iv).

The basic results of the project are exceptional in terms of the extant literature in the field of conflict analysis. First and foremost, the project was able to correctly predict about two-thirds of state failures in-sample as well as in the out-of-sample control data. This is an order of magnitude higher than published studies of similar phenomena, such as those examined in detail above.[14] The State Failure Task Force also found that the strongest factor was the character of the regime. In particular, the odds of failure in partial democracies was *seven times higher* than in stable democracies and autocracies.

Using data from the State Failure Phase III report (described in the Appendix A.3) we replicated the State Failure global model. The estimated coefficients from the replication and the original study are presented in Table 6. Infant mortality, an indicator of the partial nature of democratic rule, population density, and the extent of civil conflict in bordering states are flagged as statistically significant factors in the original study as well as in the replicated analysis, which also finds trade openness to be important. These analysis are able to predict over 80 of the extant 114 state failures correctly using the 0.25 cutoff for predicted probabilities that is employed by the State Failure Task Force.[15]

Even relaxing the cutoff on predicted probabilities to 0.5, which was used above for the Fearon and Laitin as well as the Collier and Hoeffler replications, the State Failure Project

[14]The cutpoint chosen by the State Failure Task Force was .25, which implies that the cost of a falsely predicted state failure is three times as costly as a correctly predicted state failure.

[15]It may actually be better to use conditional logit in this kind of application, but that as they say "is another show."

Table 6: *A comparison of the correctly number of state failures shows that exclusion of democracy and openness makes little difference in the ability of the modeling approach to correctly identify state failures. Bordering conflict states, infant mortality, and especially the existence of a partial democracy have substantial influence on correct predictions. Original model values for* β *and* σ_β *are reproduced from Table 6. Predicted probabilities greater that* 0.25 *are classified as a predicted state failure to maintain comparability with the Fearon and Laitin and Collier and Hoeffler studies; the actual cutpoint used by the State Failure Task Force was* 0.25*.*

Replicated Model	Brier Score	Correctly Predicted Onsets	β	σ_β
Full Model	.15	37/114		
Minus Democracy	.15	37	.04	.38
Minus Openness	.15	35	.54	.28
Minus Bordering Conflict States	.16	32	.65	.28
Minus Infant Mortality	.16	29	.79	.21
Minus Partial Democracy	.17	6	2.02	.32

Report Global Model produces a fit that is substantial in terms of predictions. Table 6 illustrates that the full model correctly predicts 37 state failures even with the cut point at 0.5. We have also re-estimated the model deleting each of the important variables, and assessing how many correctly predicted state failures are produced in each scenario. As seen in Table 6 each of the statistically important factors in this model actually result in a reduced level of predictive power, a result not obtained with either of the other replications undertaken in this study.

Indeed, deleting the partial democracy indicator from the analysis reduces the number of correct predictions over five-fold to 6. This is different than just having a large $t - statistic$ for this variable, since many models discussed herein have significant variables with large $t - statistics$ that do not greatly affect their (absence of) predictive success. This kind of information is much more meaningful, we assert, than the level of significance reported for estimated coefficients, particularly in the absence of experimental data. Figure 1 illustrates that there are very few democracies in the countries that experienced state failures and very few partial democracies among the matched group of countries that did not have state failures. Thus, the indicator of partial democracy is almost a perfect predictor of state failure in the control group.

To what extent are the predictions of the State Failure Task Force generated by an analysis of the complete sample? We undertake a brief cross-validation experiment in which the regression and prediction are run $n - 1$ times, leaving out one observation each time. This is known as "leave one out cross validation." We compute the mean Brier score over these iterations. Ideally, the average cross-validation error should be the same as reported for the regression using all observations. In the case of the State Failure Task Force, Phase III model, the Brier score from the cross-validation experiment is substantially higher: .23 compared with .15. This suggests that there is some bias introduced into the forecasts as a result of the use of the complete sample.

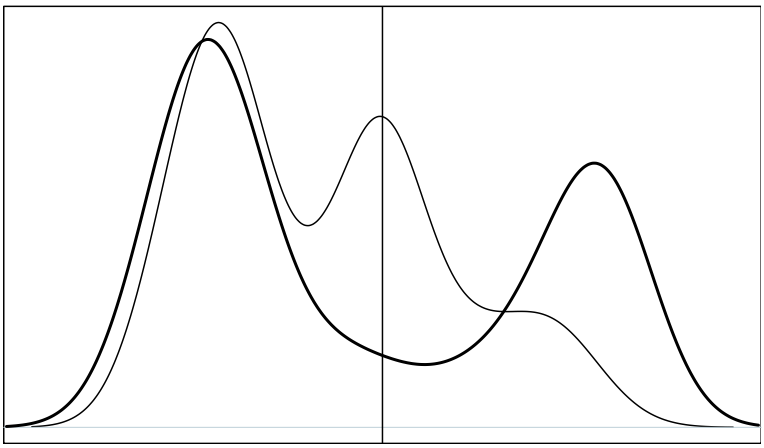

Figure 1: *The empirical distribution of democracies and autocracies, based on the Polity IV measure of democratic rule, combining the extent of democracy as well as autocracies in independent societies. This measure ranges from* -10 *for autocracies to* $+10$ *for fully democractic societies. Shaded line is the density of the quasi-control group, those selected cases without the onset of a state failure.*

7. What Have We Learned?

> *Natural abilities are like natural plants, that need pruning by study; and studies themselves do give forth directions too much at large, except they be bounded in by experience*, Francis Bacon

Using Fearon and Laitin's model, we were able to predict a few civil war onsets by use of matching, treating prior war and instability as treatments. Important variables included prior war, per capita income, new state, and the level of democracy. Matching on the Post-cold war era improved slightly the predictions in the model of Collier and Hoeffler, which showed that per capita income and population were important variables. In the State Failure project based on a matched sample of one onset to three non-onsets, an indicator of partial democracy was responsible for the correct prediction of approximately two-thirds of the 37 correctly predicted state failures using the 0.5 cutoff. Additionally, the absence of trade openness and bordering conflict states in the model specification marginally diminished the number of correct predictions. Infant mortality was also approximately one-third as important as the partial democracy indicator. These brief summaries are, of course, conditional on the extant data and analyses. Nonetheless, our own work has substantially modified–largely by pruning back–the causal claims found in the existing studies.

At a methodological level we have demonstrated several important facts of life in the quantitative study of conflict. Over-fitting–the tuning of results to observed data–needs to be annealed with methods that reduce its bias on the one hand and permit a validation of the inferences drawn on the other. Over-fitting is an important threat to validity in studies which are undertaken on a population–not a sample–of the data. In such studies the estimate standard errors of coefficients which are typically used to for decisions on which variables are important lose much–if not all–of their meaning. Standard errors are estimates of the sampling variability of the results, but if there is no sampling that is undertaken, little is to be gained by confusing estimates of non-existent sampling error with estimates of uncertainty.

Statistical significance is generally a flawed way to prune variables in non-experimental, non-sample regression models, even though we recognize that all of the studies examined here have used statistical significance as a major tool. The estimated standard errors of coefficients tell something about the observed fit of the regression to the data, but do not

reflect uncertainty about the "true parameters." Indeed, prediction appears to be a better way to evaluate models that are constructed without having their findings annealed with out-of-sample validation data. Often, *the most statistically significant variables may actually reduce our ability to make correct predictions.* This seems counter-intuitive, and should serve as a heuristic that something is amiss. This is evident in the international relations literature (Ward & Hoff 2005) and it is true in the studies of the determinants of political instability examined above.

Matching is a technique that provides some leverage on the sampling issue by helping researchers to construct quasi-experiments from observational data in a way to help reduce the selection bias that is inherent in such observational data on important social problems, like civil conflict. We also believe an avoidance of prediction as a goal of research can lead to lot of results, but few powerful guides to policy. Predictive validity, even in-sample, is a useful heuristic. Out-of-sample predictive validity is not the only way to understanding, but it is a very useful heuristic along that path.

Finally, it is worth noting that in terms of forecasting the three studies examined herein are still all producing predictions that are roughly comparable, with the State Failure project being the most successful in absolute numbers. Why is that the case? We can not give a definitive answer to this question, but we do notice several difference between this project and the others. After a little prodding, it seems to have embraced the goal of prediction and paid somewhat less attention to statistical significance. It may well be that they discovered their predictive models as a result of a large amount of flawed datamining that is not transparent. However, the amount of raw space devoted in the Phase III report to the forecasts compared to the attention paid to the reporting of statistical results is substantial. Second, the State Failure project used a case control design that annealed their statistical inferences by damping the impact of selection bias on the one hand and rare events on the other. When similar techniques were applied in the framework of the other two studies, an improvement in the performance of the model was observed. Third, the State Failure project appear to have used normalizations to almost all of their independent variables, almost all of which are dichotomous by the time they actually enter into logistic regressions. This is somewhat unique–and has a direct analogue in the area of neural networks which served in a background way to inform the State Failure Phase III modelling decisions.[16] Almost all of the explanatory variables in the State Failure project have been normalized either to whether or not the country in question is higher or lower than the world average on that variable or whether some other threshold has been met. How these cutpoints were achieved is not exactly transparent, but if you think of the $(0, 1)$ dichotomy as a step-function, the question is where to turn the "switch" on. Neural network analysis provides one way of generating these cutpoints in large data sets, but we have no way of knowing whether this was used by the project. While this reduces variability, it does make linearity assumptions less challenging. The non-dichotomized version of important variables in the the State Failure Project Phase III report, for example, do not generally help in the prediction of state failures. As a result, it seems clear that one path is more highly nonlinear models while another transforms the problem to one that can be dealt with using well understood linear methods.

[16]See Ripley (1996) and Hastie, Tibshirani & Friedman (2001) for general discussions of neural nets and statistical learning in the context of data mining. Examples of neural nets in political science began recently (Beck, King & Zeng 2000, King & Zeng 2001, de Marchi, Gelpi & Grynaviski 2004, Beck, King & Zeng 2004).

If we return to the Chechen example from the introduction, what explicitly do these findings suggest? The main reason which can be attributed to the emergence of this conflict, is most probably the incomplete transition to democracy underway in Russia at the time, coupled with the weakness of the state in economic terms, and even locally in the North Caucuses. Nearby regional conflicts and the public health impacts of collapsing institutions, also appear to play an important rôle. The question we posed initially may be moot by virtue of the continuing difficulties faced in Russia. Those conditions that were supportive of the breakdown of civil authority in Chechnya in the mid-1990 since spread. At present, there is considerable evidence that Dagestan as well as Ingushetia may be headed in a similar direction to that followed by Chechnya (The Economist 2005).

Although several extant studies have simple controls for neighbors, none takes into account the dependencies of the observations. This shortcoming is difficult to remedy, but important for a variety of reasons. Dagestan and Ingushetia are not independent in a formal or substantive way. Nor is Chechnya. All of our contemporary models treat them as if they were. Clearly there is a huge potential for diffusion. In recent years, scholars such as Gleditsch (2002), Ward & Gleditsch (2002) and, Salehyan & Gleditsch (2004) Salehyan and Gleditsch (2004) have demonstrated that understanding intrastate conflicts, including ethnic conflicts, may benefit from taking into account regional and diffusion effects. Recently Raleigh (2005) has shown that instability caused by one country's war may impact the democratic institutions in neighboring countries, which in turn may affect those countries' internal stability. The idea that the neighborhood is important has wide currency, even in popular writings (Diamond 2005) about the reasons for societal as well as institutional collapse.

To the extent that empirical research will be especially informative, it will have to embrace forecasting. To do so will require the restraint of analyzing only some of the available data, and doing so with an eye to not only generating statistical models that are tractable but that also produce valid inferences, even on data that has not (yet) been analyzed. The study of civil conflict has made dramatic advances in recent years. By disaggregating our focus from the nation state, we hope to make substantial further progress. However, considerable progress may actually also be made by disaggregating our large-N studies into matched groupings of cases that help to control for the non-random nature by which the world assigns interesting cases. At the same time, even in more highly disaggregated studies of civil conflict, it will be important to take into account the potential bias introduced by the nature of our non-experimental and non-sampled data. In so doing we have to find more clever ways to recognize the interdependency of these cases over space as well as time.

References

Alesina, Alberto, Enrico Spolare & Romain Wacziarg. 2000. "Economic Integration and Political Disintegration." *American Economic Review* 90(5):1276–97.

Bakke, Kristin M. & Erik Wibbels. 2004. "Inequality, Ethnicity, and Conflict in Federal States." *Paper presented at the annual meeting of the American Political Science Association, Chicago, Illinois* .

Beck, Nathaniel, Gary King & Langche Zeng. 2004. "American Political Science Review." *Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski* 98(2):379–389.

Beck, Neal, Gary King & Langche Zeng. 2000. "The Problem with Quantitative Studies of Conflict: A Conjecture." *American Political Science Review* 94:21–36.

Bermeo, Nancy. 2002. "The Import of Institutions." *Journal of Democracy* 13(2):96–110.

Brier, Glenn W. 1950. "Verification of forecasts expressed in terms of probabilities." *Bulletin American Meteorological Society* 78:1–3.

Brown, Rupert. 2000. "Social Identity Theory: Past Achievements, Current Problems and Future Challenges." *European Journal of Social Psychology* 30(6):745–78.

Bunce, Valerie. 1999. *Subversive Institutions: The Design and the Destruction of Socialism and the State*. Cambridge, UK: Cambridge University Press.

Collier, Paul & Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56(4):563–595.

de Marchi, Scott, Christopher Gelpi & Jeffrey D. Grynaviski. 2004. "American Political Science Review." *Untangling Neural Nets* 98(2):371–378.

de Soysa, Indra. 2002. "Paradise is a Bazaar? Greed, Creed, and Governance in Civil War, 1989-99." *Journal of Peace Research* 39(4):395–416.

Diamond, Jared. 2005. *Collapse: How Societies Choose to Fail or Succeed*. New York: Viking.

Dunlop, John B. 1998. *Russia Confronts Chechnya: Roots of a Separatist Conflict*. Edinburgh, UK: Cambridge University Press.

Esty, Daniel C., Jack A. Goldstone, Ted Robert Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela T. Surko & Alan N. Unger. 1998. "State Failure Task Force Report: Phase II Findings." *Science Applications International Corporation, McLean, VA* .

Esty, Daniel C., Jack A. Goldstone, Ted Robert Gurr, Pamela T. Surko & Alan N. Unger. 1995. "Working Papers: State Failure Task Force Report: Phase I Findings." *Science Applications International Corporation, McLean, VA* .

Evangelista, Matthew. 2002. *The Chechen Wars: Will Russia Go the Way of the Soviet Union?* Washington, D.C.: Brookings Institution Press.

Fearon, James D. 2005 forthcoming. "Primary Commodities Exports and Civil War." *Journal of Conflict Resolution* 49(2):tba.

Fearon, James D. & David A. Laitin. 2000. "Violence and the Social Construction of Ethnic Identity." *International Organization* 54(4):845–877.

Fearon, James D. & David D. Laitin. 2003*a*. "Additional Tables for 'Ethnicity, Insurgency, and Civil War'." *athttp://www.stanford.edu/~jfearon/papers/addtabs.pdf* .

Fearon, James D. & David D. Laitin. 2003*b*. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97:75–90.

Fearon, James D. & Pieter van Houten. 2002. "The Politization of Cultural and Economic Differences: A Return to the Theory of Regional Autonomy Movements." *Stanford University Working Paper, available at http://www.duke.edu/web/licep/5/JF-PvH/JF-PvH-LiCEP5.pdf* .

Filippov, Mikhail, Peter C. Ordeshook & Olga Shvetsova. 2004. *Designing Federalism: A Theory of Self-Sustainable Federal Institutions*. Cambridge, UK: Cambridge University Press.

Gagnon, Valère Philip. 1997. Ethnic Nationalism and International Conflict: The Case of Serbia. In *Nationalism and Ethnic Conflict*, ed. Michael E. Brown, Jr. Owen R Cote, Sean M. Lynn-Jones & Steven Miller. Cambridge, MA: The MIT Press pp. 132–169.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca, NY: Cornell University Press.

Ghobarah, Hazem Adam, Paul Huth & Bruce Russett. 2003. "Civil Wars Kill and Main People-Long After the Shooting Stops." *American Political Science Review* 97(2):189–202.

Gleditsch, Kristian Skrede. 2002. *All International Politics is Local: The Diffusion of Conflict, Integration, and Democratization*. Ann Arbor, MI: University of Michigan Press.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Goldstone, Jack A. & Charles Tilly. 2001. Threat (and Opportunity): Popular Action and State Repression in the Dynamics of Contentious Action. In *Silence and Voice in the Study of Contentious Politics*, ed. Ronald R. Aminzade, Jack A. Goldstone, Doug McAdam, Elizabeth J. Perry, Jr. William H. Sewell, Sidney Tarrow & Charles Tilly. Cambridge, UK: Cambridge University Press chapter 7, pp. 179–194.

Goldstone, Jack A., Ted Robert Gurr, Barbara Harff, Marc A. Levy, Monty G. Marshall, Robert H. Bates, Colin H. Kahl, Pamela T. Surko, John C. Ulfelder & Alan N. Unger. 2000. "State Failure Task Force Report: Phase III Findings." *Science Applications International Corporation, Arlington, VA* .

Goodwin, Jeff. 2001. *No other Way Out: States and Revolutionary Movements, 1945-1992*. Cambridge, UK: Cambridge University Press.

Gurr, Ted Robert. 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.

Gurr, Ted Robert. 2000. *Peoples versus States*. Washington, D.C.: United States Institute of Peace Press.

Hastie, Trevor, Robert Tibshirani & Jerome Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag.

Hechter, Michael. 2000. *Containing Nationalism*. New York, NY: Oxford University Press.

Hegre, Håvard, Tanja Ellingsen, Scott Gates & Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 18161992." *American Political Science Review* 95(1):33–48.

Hewstone, Miles & Katy Greenland. 2000. "Intergroup Conflict." *International Journal of Psychology* 35(2):136–144.

Ho, Daniel E., Kosuke Imai, Gary King & Elizabeth A. Stuart. 2004. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Working paper at* `Gking.Harvard.Edu/matchit` .

Horowitz, Donald. 1985. *Ethnic Groups in Conflict*. Berkeley, CA: University of California Press.

Imai, Kosuke & David A. van Dyk. 2004. "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association* 99(467):854–866.

Kahn, Jeffrey. 2002. *Federalism, Democratization, and the Rule of Law in Russia*. New York: Oxford University Press.

Kauffmann, Mayeul. 2004. *Les organisations économiques internationales et la sécurité militaire: une analyse économique*. Grenoble, France: Ph. D. Thesis, University of Pierre Mendès France.

King, Charles. 2004. "The Micropolitics of Social Violence." *World Politics* 56(3):431–455.

King, Gary & Langche Zeng. 2001. "Improving Forecasts of State Failure." *World Politics* 53(4):623–658.

Lichbach, Mark I. 1994. "What Makes Rational Peasants Revolutionary? Dilemma, Paradox, and Irony in Peasant Collective Action." *World Politics* 46(3):383–418.

McAdam, Doug, McCarthy John D & Mayer N. Zald. 1996. *Comparative Perspectives on Social Movements: Political Opportunities, Mobilizing Structures, and Cultural Framings*. Cambridge, UK: Cambridge University Press.

McCarthy, John D. & Mayer N. Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212–1241.

Milanovic, Branko & Nicholas Sambanis. 2003. Exploring the Demand for Sovereignty. Working paper World Bank.

Muller, Edward N. & Mitchell A. Seligson. 1987. "Inequality and Insurgency." *American Political Science Review* 82(2):425–452.

Østby, Gudrun. 2005. "Horizontal Inequality and Civil Conflict." *Paper presented at the annual Norwegian Political Science Conference, Hurdalsjøen, Norway, available at* `http://www.statsvitenskap.uio.no/konferanser/nfkis/cr/` .

Petersen, Roger. D. 2002. *Understanding Ethnic Violence: Fear, Hatred, and Resentment in Twentieth-Century Eastern Europe*. Cambridge, UK: Cambridge University Press.

Popkin, Samuel. 1988. Political Entrepreneurs and Peasant Movements in Vietnam. In *Rationality and Revolution*, ed. Michael Taylor. New York: pp. 9–23.

Raleigh, Clinodah. 2005. "A Discussion of the Role of Neighbors in Conflict, Development, and Democracy." *Paper presented at the annual Norwegian Political Science Conference, Hurdalsjøen, Norway, available at* `http://www.statsvitenskap.uio.no/konferanser/nfkis/cr/Raleigh.pdf` .

Raleigh, Clionadh Anne. 2004. "Neighbours and Neighborhoods: Understanding the Role of Context in Civil War." *Fifth Pan-European International Relations Conference, den Haag, September 9-11* .

Rasler, Karen. 1996. "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociology Review* 61(1):132–152.

Ripley, Brian D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Saideman, Stephen, David Lanou, Michael Campenni & Samuel Stanton. 2002. "Democratization, Political Institutions, and Ethnic Conflict: A Pooled Time-Series Analysis, 1985-1998." *Comparative Political Studies* 35(1):103–129.

Salehyan, Idean & Kristian Skrede Gleditsch. 2004. "Refugees and the Spread of Civil War." *Paper Presented at annual meeting of the American Political Science Association, Chicago, available at* `http://weber.ucsd.edu/~kgledits/projects.html` .

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

Sambanis, Nicholas. 2004. "Using Case Studies to Expand Economic Models of Civil War." *Perspectives on Politics* 2(2):259–280.

Sarkees, Meredith Raid, Frank Whelon Wayman & David J. Singer. 2003. "Inter-State, Intra-State, and Extra-State Wars: A Comprehensive Look at Their Distribution Over Time, 1816-1997." *International Studies Quarterly* 47(1):47–70.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge, UK: Cambridge University Press.

Smith, Anthony D. 1986. *The Ethnic Origins of Nations*. Oxford, UK: Blackwell.

Snyder, Jack. 2000. *From Voting to Violence: Democratization and Nationalist Conflict*. New York: Norton.

Suny, Ronald Grigor. 1993. *The Revenge of the Past: Nationalism, Revolution, and the Collapse of the Soviet Union*. Stanford, CA: Stanford University Press.

The Economist. 2005. "An Empire's Fraying Edge." *The Economist* 374.

Tishkov, Valery. 1997. *Ethnicity, Nationalism, and Conflict In and After the Soviet Union: The Mind Aflame*. London: Sage.

Toft, Monica Duffy. 2003. *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory*. Princeton, NJ: Princeton University Press.

Treisman, Daniel. 2001. *After the Deluge: Regional Crisis and Political Consolidation in Russi*. Ann Arbor: Michigan University Press.

Ward, Michael D. & Kristian Skrede Gleditsch. 2002. "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace." *Political Analysis* 10(3):244–260.

Ward, Michael D. & Peter D. Hoff. 2005. "Persistent Patterns of International Commerce." *University of Washington, Center for Statistics and the Social Sciences,* Working Paper 45: available at `http://www.csss.washington.edu/Papers/`.

Appendix A. Data

A.1. **Fearon and Laitin.** To address missing data in the population series, we examined those observations with missing data. Both the Federal Republic of Germany and the German Democratic Republic were missing data for early years in the 1950s. The missing data in part comes from the splitting up of the Third Reich during the occupation of Germany and annexed Austria after the Second World War. Both East and West Germany were deleted prior to 1951, owing to their post war occupation status. One could make the argument to actually delete them until about 1955. Estimates of post 1950 East and West German populations are, however, available at `www.gesis.org/en/social_monitoring/social_indicators/Data/System/keyindic/population.pdf`. For the Federal Republic, 68,377 and 68,879 are the population estimates in millions. Figures for East Germany were given as 18,338, 18,351, 18,328, and 18,178 in millions annually from 1950 through 1953. Both of these sets of numbers include the population of Berlin, East and West. Austria was excluded before 1951, on the same grounds. Japan is also a similar case, though the Japanese Federal Statistical office `www.stat.go.jp/data/chouki/zukyou/` has long series on Japanese population figures. We excluded Japan prior to 1954. Since Libya was not independent until 1952, it was not included until 1953. Syria is missing data for 1960; we interpolated between adjacent years and utilize 4.307 million as its 1960 population. Finally, Laos was not independent until 1954; we use the figure of 2.103 million for its 1953 population.

Fearon and Laitin (2003, page 81 and footnote 18) point out that instability is assessed on the basis of a change of at least three points on the polity democracy score in any of the prior three years. Foreign occupations are considered to be transitions, but foreign occupation is treated as missing, which causes the observation to be deleted from the analysis. Where

possible, the Polity data were updated from more recent releases and corrections of the Polity project. Countries in transition were coded as -10 per the suggestions in Fearon and Laitin. Thus, codes of -10 were assigned to the following: GDR (1990), Hungary (1957), Bosnia (1996-1999), Uganda (1980), Syria (1959-1961), Lebanon (1991-1999), China (1945-6), Cambodia (1980-1988), Laos(1954), and South Vietnam (1954-5). Ghana (1957-1960) was rated -8; Zimbabwe (1965-1970): 4; Tunisia (1956-9): -9; Kuwait (1961-3): -8; Kuwait (1991): -10; and India (1947-1950): 9. By implication, the following have instability: Ghana (1959-60), Zimbabwe (1967-1970), Tunisia (1958-9), Syria (1961), Kuwait (1963), Japan (1952), and India (1949-50).

Many countries are missing values for the GDP per capita variable. About one-fifth of these can be easily imputed from one period lagged values of GDP, which are almost perfectly correlated ($R^2 = .99$). The equation used was $y_t = -0.075 + 1.0048 \cdot y_{t+1}$. This still left about 200 country years for which there were no data on GDP per capita. After experimenting with multiple imputation based on other available data, no satisfactory set of imputations were available, and these cases were omitted from further analysis. This resulted in a slightly smaller sample of onsets, 106 versus 110 in the original data.

A.2. **Collier & Hoeffler.** These data are provided in Stata® format at `http://users.ox.ac.uk/~ball0144/g&g.zip`. No special handling was necessary to conduct the analyses reported herein.

A.3. **State Failure Project, Phase III .** The State Failure Task Force, Phase III report is quite detailed and the Web site (`http://www.cidcm.umd.edu/inscr/stfail/`) that provides supporting information also is replete with information and data. More is, unfortunately, not always better, though it is always more. Currently, it is not possible to replicate their work with information publicly available on the Web site. The databases therein provided do not include the imputed values necessary to replicate the projects findings. In principle, the actual data that were used by the project, including the imputed data, are available in printed form in the project report (Goldstone et al. 2000), as Appendix Table A-9. Unfortunately, Appendix Table A-9 is missing 25 rows, a mistake that limits its function as a data archive.

To replicate the data analysis of the task force, we "screen scraped" the available data from the PDF report on the web site. We obtained the missing 25 rows of data from Jay Ulfelder of SAIC, Inc. The first author will make these data available upon request, until the State Failure Task Force web materials are corrected.

About the authors:

Michael D. Ward: Department of Political Science, University of Washington, Seattle, Washington, USA, 98195-3530
E-mail address: mdw@u.washington.edu

Kristin Bakke: Department of Political Science, University of Washington, Seattle, Washington, USA, 98195-3530
E-mail address: kmbakke@u.washington.edu