New Lessons Learned? Improving Genocide and Politicide Forecasting

Chad Hazlett

Understanding where and when genocides are likely to occur is central to prevention efforts, yet only one published paper (Harff 2003) has focused primarily on developing a statistical forecasting model.With more than 10 years of new data available, an assessment of this influential model and its performance is well overdue. Here I provide a re-evaluation and an alternative model. I show that without adjustments, the Harff 2003 model does not produce a meaningful or usable forecast. Once adjusted to forecast genocides over the course of a given political instability event, two "warning signs" previously thought to be important - -*prior genocide* and *ethnic character of the ruling elite* - are no longer predictive. An alternative forecasting model emphasizing the strategic motivations of regimes when threatened demonstrates improved specificity and sensitivity, both in-sample and out-of-sample. Finally, I test a time-series model of the likelihood of genocide in the next-year, finding that different variables are useful for determining *when* genocide will occur rather than *if* genocide will occur during a political instability event. These results are of value to analysts employing either quantitative or qualitative methods to make genocide risk assessments and informing short to medium term prevention efforts accordingly.

Preventing and minimizing mass atrocities such as genocide and "politicide" (political mass murder) is of self-evident importance. It is hoped that, with sufficient forewarning, prevention of atrocities could be achievable through targeted diplomacy, mediation, peacebuilding or reconciliation programs, development assistance, threats of punitive measures, or denial of capabilities to commit atrocities. For such an approach to be effective, however, it is critical to determine when and where to deploy finite resources by identifying high-risk situations.

How should policymakers and diplomats determine when and where to invest in preventive activities? Clearly, sudden developments - such as post-election violence, protests that turn to violence, or a surge in the use of ethnically charged hate speech - could spur action by concerned states, international organizations, and NGOs. Developing the capacity to respond to such "triggers" or "late warnings" is likely to be a useful element of genocide prevention (Albright & Cohen, 2008). However, there are also important reasons to improve "early warning" or "risk assessment" capacities that would allow analysts to make longer term assessments regarding the risk of genocide in a particular state. First, resources for engaging in preventive activities are

limited. Agencies such as the UN Office of The Special Adviser on The Prevention of Genocide (OSAPG) have limited resources and need to know what areas to begin paying attention to, develop expertise on, send people to monitor, build relationships with, etc. Second, some projects aimed at preventing genocide or politicide (henceforth simply 'genocide' unless otherwise noted) may require longer time horizons to have any chance of success. This may be true of diplomatic interventions with potential perpetrators, but is more clearly the case when the tools deployed involve indirect means such as targeted economic assistance, human rights and rule of law programs, reconciliation efforts, electoral reform, or grassroots peacebuilding activities (e.g. radio and television programming). Thus, while medium-to-long-term genocide "warnings" will invariably be probabilistic gambles, they are needed nevertheless to inform the effective and efficient deployment of resources.

A small but growing set of cross-national, quantitative analyses has begun to examine the onset or intensity of genocide and politicide (Krain 1997; Colaresi & Carey 2008; Ayden &Gates 2007), "democide" (Rummel1995), or "mass killing" (Valentino et al. 2004; Wayman & Tago 2009; Esteban et al. 2010, Easterly et al. 2006). These studies have examined how the use of mass violence by the state relates to regime type and characteristics (e.g. Rummel 1995; Easterly et al. 2006; Aydin & Gates 2007; Colaresi & Carey 2008; Eck & Hultman 2007; Waymen & Togo 2010), military strength (Colaresi & Carey 2008), openings in political opportunity (Krain 1997), income level (Easterly et al. 2006), natural resources rents and state capacity (Esteban, 2010), or the type of insurgency faced by a state and its degree of public support (Valentino et al., 2004).[1] Limited work (Waymen & Togo 2010) has seriously examined the differences between different definitions and datasets, comparing two commonly used datasets: the Harff (e.g. 2003) data on genocide and the more expansive Rummel (e.g. 1995) dataset on democide.[2]

These studies provide useful theoretical arguments and empirical evidence, and I will return to their substantive results in greater detail to propose variables for a new forecasting model in part

[1] In the present paper, I examine only violence committed by the state, however another strand of literature has focused on mass violence committed by rebels rather than the state, including Azam (2006), Eck & Hultman (2007), Humphreys & Weinstein (2006), Kalyvas (2006).

[2] More recently Eck & Hultman (2007) constructed data on "one-sided killing" from 1984 to 2004, and Easterly et al. (2006) constructed data from 1890 to 1998 on episodes of mass killing. Valentino et al. (2004) provide data for "mass killing."

II. However to my knowledge, with the exception of a brief examination of forecasting power in Colaresi & Carey (2008), Harff 2003 remains the only published attempt specifically to provide an early warning model or to focus on predictive performance of the model as the principle outcome of interest. Given its focus, its publication in the *American Political Science Review*, and the continued work by Harff and others to provide risk assessments to policymakers based on this model, Harff 2003 has also been uniquely understood as a forecasting paper in the policy and advocacy communities. The six "warning signs" identified in the model have thus been widely cited and employed by those interested in genocide forecasting and risk assessment. For example, the Genocide Prevention Task Force (Albright & Cohen, 2008) cites the risk factors identified in Harff 2003 in its description of genocide forecasting, and NGOs such as Minority Rights Group International and the Genocide Prevention Project have formulated watch-lists based at least partly on the warning signs in Harff 2003 or directly on the risk assessments subsequently provided (which have been provided by Harff in most years since the 2003 publication). Given the dominating role it has played in the genocide forecasting community, it is particularly important to understand what that model is estimating, and how well it performs (in-sample and out-of-sample).

This paper's first set of goals is to provide that re-assessment, update, and out-of-sample test. The second set of goals is to then build on this evaluation, together with findings available from recent empirical studies, in order to develop a new set of models with increased forecasting power. Note that this paper is a forecasting project rather than the more traditional theory testing paper. In this regard, both Harff 2003 and this paper differ from much of the literature. The burden here is on showing forecasting power through in-sample and out-of-sample performance, rather than on arguing that marginal effects found in regressions can be given a theoretical or causal interpretation. While the alternative model I propose is informed by theoretical arguments - and shows that the data are consistent with those arguments - stable and accurate forecasting is the goal here so as to produce a useful tool for genocide risk assessment.

I. Revisiting Harff 2003

Definitions of genocide, politicide, or other forms of mass killing are often a point of disagreement and may lead to differing conclusions in empirical studies (see Waymen & Togo, 2010). In the present work I focus exclusively on Harff's definition and data to ensure that any difference in conclusions are not due to differences in data.[3] Harff defines genocide (including politicide) as "the promotion, execution, and/or implied consent of sustained policies by governing elites or their agents – or, in the case of civil war, either of the contending authorities – that are intended to destroy, in whole or part, a communal, political, or politicized ethnic group." This definition is more inclusive than the UN Convention on the Prevention and Punishment of Genocide in that it includes political groups as potential targets, but it is similar regarding the focus on "groups" *as such* and the deliberate effort to destroy them in whole or in part.

Conditioning on Instability

A central feature of the Harff 2003 model is that it examines the probability of genocide *only when a political instability event other than genocide is already taking place*. This decision is justified by Harff's finding that virtually all genocides occur during ongoing instability events.[4] It is also consistent with earlier work by Krain (1997) regarding the importance of instability in providing "political openings" for genocide. The political instability events referred to here are those identified by the Political Instability Task Force (PITF), which include revolutionary wars, ethnic wars, and adverse regime changes (i.e. coups).

[3] Future studies disaggregating these events into potentially unique phenomena with potentially different pathologies may prove productive, akin to the disaggregation of civil wars (e.g. Kalyvas 2001, Sambanis 2001, Fearon 2004) in the civil war literature.

[4] The only genocide identified by Harff that did not fall during an ongoing instability event was the 1981-1982 case in in Syria in which the Muslim Brotherhood was targeted. Harff argues that in fact revolutionary war was ongoing by this time, but below the threshold used by the PITF. Thus, she recodes that case as one that includes revolutionary war. I do the same here.

Major Variables and Hypotheses

Harff's model-building approach calls upon prior comparative and sociological work, particularly by Harff and Gurr (1988, 1998) and Fein (1984, 1993), together with previous empirical work by the PITF (e.g. Goldstone et al. 2002). The first risk-factor proposed is "political upheaval," which is measured as a weighted sum of different types of instability events in the preceding 15 years. Two rationales are offered for this: first, that the more intense and persistent a conflict has been, the more authorities are likely to feel threatened and willing to adopt extreme measures. Second, greater upheaval is argued to greater "opportunity" to employ a genocidal strategy (see also Krain 1997). Relatedly, another risk-factor is the presence of prior genocides in the state in question, argued to be important because it leads to "habituation" to the use of mass violence.

Harff also argues for the role of regime type and several related characteristics. Consistent with other studies, more democratic regimes are expected to have stronger institutional checks on executive power, thereby constraining elites from taking extreme actions such as genocide. In this regard, Harff proxies "executive constraints" (later used more directly by Aydin & Gates 2007 and Colaresi & Carey 2008) with regime type in terms of autocracy versus democracy. Harff also argues for two other regime-related characteristics less commonly seen in the literature. When there is an "exclusionary ideology of the ruling elite," Harff argues, such regimes may call upon an "overriding purpose or principle that justified efforts to restrict, persecute, or eliminate groups, including those that pose no obvious threat to the elite." Additionally, "ethnic character of the ruling elite" is argued to increase the likelihood of genocide because (a) under-represented groups may wish to challenge this unrepresentative elite, and (b) elites may fear this challenge and seek to further weaken or exclude groups that pose potential threats.

Modeling Approach

The Harff (2003) model was run on data consisting of one observation per instability event.[5] This includes all the cases that did experience genocide (35) and all the cases that did not (91). That is, there are 126 rows of data, each corresponding to an instability event, of which 35 have a "1"

[5] With the exception of two instability events included twice, corresponding to two genocides.

on the dependent variable, indicating genocide occurred at some point during those events. A logistic regression of this genocide indicator on a set of covariates was then run using this entire set of observations. The final model includes the following variables, and the estimated coefficients shown in Table 1:

Table 1. Original Model

Variable/Description	Coefficient (p-value)	Odds Ratio
Prior upheaval (A measure of the cumulative "prior upheaval" over the past 15 years)	0.048 (0.05)	1.05[a]
Prior genocide (Indicates whether genocide occurred previously in that country)	1.220 (0.09)	3.39
Ideological orientation of the ruling elite	0.937 (0.07)	2.55
Regime type (autocracy dummy)	1.223 (0.03)	3.40
Ethnic character of ruling elite	0.939 (0.09)	2.56
Trade openness Natural log of the share of GDP due to exports and imports	-1.242 (0.01)	0.388[b]
Constant	1.58 (0.27)	
N=126 Pseudo-R^2 =0.28		

[a]Note that this odds-ratio was mistakenly reported to be 1.70 in the original. The odds-ratio for regime type was also mistakenly reported to be 3.50 rather than 3.40.

[b]Odds ratio for movement from 25th percentile to 75th percentile in trade openness

The difficulty, however, lies in uncertainty over what exactly the model predicts. The abstract of Harff 2003 says the model examines "which factors distinguish the 35 episodes [of political instability] that led to geno-/politicides from those that did not." This suggests that the model will estimate the probability that a given *instability event* will experience genocide at some point during the episode. By contrast, elsewhere (p.65) the paper claims to estimate "the conditional

probability that a genocide or politicide will begin *one year later* in a country already experiencing failure" (emphasis mine).

As is, the model does not produce either the "instability-wise probability of genocide" or the "next-year probability of onset," because the choice of time-points from which these observations are drawn is not consistent with either approach. This is clearest to see by first considering how one would go about modeling either of these probabilities:

1. *The instability-wise probability of genocide* could be estimated by taking covariate data from the year of - or the year before - the onset of each political instability event, then using these data to predict whether genocide later occurs at any time during that instability event.

2. *The next-year probability of genocide onset* would most naturally be modeled using time-series methods, however could also be estimated by taking the year prior to each genocide onset and labeling it as a "case," then randomly drawing a set of "control" observations from the pool of <u>all other instability-years</u> not followed by genocide onsets. A version of this in which an integer number of "control" observations are drawn in proportion to each case is called the "case-control" approach.

The approach taken in Harff 2003 does not fit either of these descriptions, or any other evident approach, due to the timing of the covariate data. For state failure episodes that *do* experience genocide, the covariate data were drawn from the year prior to the *genocide*, suggesting a next-year probability of onset model.[6] By contrast the choice of control observations is closer to what would be used for an instability-wise probability model: one "control" observation was taken from each instability event that did not lead to genocide, rather than being randomly drawn from the pool of all remaining instability years.[7]

Drawing data for genocide "cases" from the year before *genocide* (rather than the year before *instability onset*) is not consistent with an "instability-wise probability of genocide" interpretation.[8] However, the model is also inconsistent with the "next-year probability of

[6] With the exception of Yugoslavia 1992, for which data were drawn from two years prior.

[7] These were drawn from a varying set of years ranging from the year before the onset of political instability to three years after it.

[8] The act of choosing the year prior to genocide requires foreknowledge of which year genocide occurs. This problem gives rise to an inability of analysts to actually use the model for prospective forecasting of the *instability-wise* probability of genocide.

genocide onset" approach, for two important reasons. First, drawing one "control" observation from each political instability event not leading to genocide, rather than drawing them randomly from the pool of instability years not followed by genocide, causes a significant problem. This is because no control observations are ever drawn from the instability events that do *eventually* lead to genocide, but not the next year. This model never had to distinguish between two years within a given instability event, one of which is followed by non-genocide and the other that is followed by a genocide onset. More generally, the model never faces a situation in which genocide ultimately occurs in a given state but not in the year following the one being analyzed. This selection bias makes the model appear far more effective than it would actually be if exposed to yearly data from instability events, including some of which are followed by genocide after a delay.

Second, under the "next-year probability of onset" interpretation, the model would vastly over-predict genocides. Across the more than 1000 country-years of political instability, 35 genocide onsets occurred. Thus, less than 3.5% of *instability-years* are followed by genocide onsets. By contrast, since the data are coded at the instability-event level, 35/126 (28%) of the observations seen by the model involve genocide onsets. As the model is fit to data with eight times the rate of genocide onsets that would be expected per instability year, the resulting model will predict approximately *eight times too many genocides* if interpreted as predicting the probability of a given year leading to an onset.[9] Thus, it is vitally important that analysts avoid interpreting the Harff 2003 model as producing estimates of the next-year probability of onset.

Correcting the Harff Model

Fortunately, this model can be adjusted to become interpretable and usable in real-time without *ex post facto* knowledge, allowing us to run it meaningfully on out-of-sample data and to compare its performance to other models. Here I adjust it to achieve the "instability-wise probability of genocide" rather than the "next-year probability of genocide" for two main reasons. First, as discussed, the existing model would widely over-predict the prevalence of

[9] In the absence of other problems, the correction for this is simple: the constant terms in the logistic regression must be adjusted when computing expected probabilities (see King & Zeng, 2001). Harff 2003 makes mention of this paper and follows its advice to interpret the logit coefficients as odds-ratios, but the correction for the differing baseline rate was not applied in Harff 2003. In the present case however, since the choice of "control" observations also was not random, the correction would not solve the problem.

genocide if interpreted as the "next-year probability of genocide." Second, the resulting model would still not be correct due to the non-random selection of "control" observations. The main adjustment needed is that the control observations must come from the year prior to instability onset for *all* instability events, whether genocide later occurs or not. Not coincidentally, this mirrors the data that would be available to analysts at the onset of instability, and this adjustment would allow analysts to actually run the model predictively without foreknowledge of when genocide will occur. A second, minor adjustment must also be made: two instability events – Burundi beginning in 1998 and Indonesia beginning 1949 – are listed as having two genocide events. If the quantity of interest is the probability that genocide occurs during a political instability event conditional on the values of the covariates, then each instability event should be taken as one observation.

Table 2. Adjusting Harff 2003 to estimate "instability-wise probability of genocide"

	I. Original	II. Correct Timing	III. Correct and Restricted
Variable	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)
Years of political upheaval in last 15 years	0.048 (0.05)	0.047 (0.065)	0.048 (0.050)
Prior genocide	1.220 (0.09)	0.57 (0.47)	--
Ideological orientation of the ruling elite	0.937 (0.07)	1.30 (0.022)	1.32 (0.015)
Regime type (autocracy dummy)[10]	1.223 (0.03)	1.52 (0.009)	1.76 (0.001)
Ethnic character of ruling elite	0.939 (0.09)	0.49 (0.41)	--
Trade openness	-1.242 (0.01)	-1.54 (0.001)	-1.53 (0.001)
Constant	1.58 (0.271)	2.92 (0.051)	2.91 (0.052)
N	126	124	124
Adjusted-R^2	0.28	0.28	0.27

[10] The original paper does not specify how the final dummy variable for autocracy was formed, but by inspecting the data and comparing it to the Polity dataset, it appears that autocracy scores of 2 or less were coded as "0" while those of three or more are coded as "1."

Allowing the model to re-estimate its coefficients on the new data, the overall R^2 is unchanged, and as we shall see, the predictive power of the model (with the adjusted coefficients) is not significantly damaged. That said, *prior genocide* drops from a p-value of 0.09 to 0.47. This may seem surprising: surely prior genocide is an important predictor of whether or not a state is likely to engage in genocide. That logic notwithstanding, it is unsurprising that this variable is not very predictive due to the left-censoring of the data.[11] *Ethnic character of the ruling elite* also drops from a p-value of 0.09 to 0.41. Again this may be surprising given the expectation that genocide should be far more likely in an ethnically divided society in which the governments is dominated by one of those ethnic groups. That said, it may be common to see ethnically divided states and ethnically characterized governments that do not experience genocide. I discuss this point at greater length below. Model III above shows that if we drop *prior genocide* and *ethnic character of the ruling elite*, the remaining coefficients are statistically and substantively unchanged, and R^2 drops by only 1%.

The characteristics of these models with which we are most concerned is predictive power. Below I illustrate the predictive performance of these models at two thresholds: one considering any predicted probability at or above 50% to be a "warning", and one using a 25% threshold. The 25% threshold reflects a preference towards detecting more genocides at the cost of potential over-prediction, and is the threshold employed in Harff 2003.

[11] If a state has not experienced a prior instability event since 1955, the value of *prior genocide* will be zero, even though we have no information about whether that state would have experienced genocide had an instability event occurred. Thus a value of zero can mean either that a state had a prior instability event and did not experience genocide, or that we simply have no information about a state's likelihood of experiencing genocide under political instability.

Table 3: Performance of Harff 2003 and the Adjusted Model

	Harff 2003	**Corrected**	**Reduced**	**Dummy**
$\boldsymbol{Alarm\ if\ P(gen) \geq 0.50}$				
Total correctly classified	105 (83%)	102 (83%)	101 (81%)	72%
Genocides correctly classified	20 (57%)	18 (55%)	16 (48%)	0%
Percent of alarms that are true	77%	72%	73%	NA
$\boldsymbol{Alarm\ if\ P(gen) \geq 0.25}$				
% all cases correctly classified	92 (73%)	96 (77%)	96 (77%)	72%
% genocides correctly classified	26 (74%)	27 (82%)	27 (82%)	0%
% of alarms that are true	51%	55%	55%	NA

Fortunately, the corrected model retains similar performance to the Harff 2003 model, and even improved performance at the 25% threshold. This is important, since the corrected model is now interpretable and can be meaningfully run prospectively. That the performance of the model at the 25% threshold is identical when *prior genocide* and *ethnic character of the ruling elite* are dropped emphasizes that these variables are no longer important on the timing-corrected data. This is worth noting for analysts counting on these findings to determine which "warning signs" are predictive, whether in quantitative or qualitative models.

II. An Alternative: The Strategic Model

In this section I propose an alternative model that considers some of Harff's hypotheses together with findings from recent non-forecasting analyses, integrating them in a model that can be used prospectively as a forecasting tool (i.e. relying only on data drawn prior to or very near the onset of instability events).

Conceptions of Genocide

While an extended discussion of definitional issues is not possible or desirable here, a brief examination of the types of phenomena captured by the genocide/politicide data is necessary. For

many, the term “genocide” immediately - and very appropriately - evokes images of the Holocaust. It was this event that spawned the coining of the term by Raphael Lemkin and its subsequent enshrinement in international law. In that context, it is certainly reasonable to think of genocide as a phenomenon closely related to ethnic hatreds and dehumanization processes that lead some members of one group to attempt to destroy the other group in whole or in part. Accordingly, the vast majority of sociological literature on genocide focuses on ideas of ethnicity, race, or other identities, and/or on the “dehumanization” that becomes possible as a result of out-group formation processes between these identity groups (e.g. Alvarez 1997, Moller 2010, Harris & Fiske 2006, Hagan & Rymond-Richmond 2008).

Whether sufficient or not, this view of genocide as the ultimate “hate crime” may be more accurate if we restricted our attention to genocides that qualify as such under the UN Convention on the Prevention and Punishment of Genocide (henceforth the “Genocide Convention”). However the Harff dataset includes a much wider range of events, most notably “politicides” that would not likely satisfy the Genocide Convention. In particular, to enter this dataset, the target group need not be explicitly ethnically, racially, or nationally defined. Many of the observations in the Harff dataset do not “look like” the Holocaust, Rwanda in 1994, or the Balkans in the 1990s. Indeed, of the 35 geno/politicides listed in Harff 2003, 25 are coded as “politicides” alone and not as genocides. Of the remaining 10, seven are listed as “politicide and genocide” together. Only three events – Rwanda 1994, Bosnia 1992, and Burma 1978 – are listed as “genocide” alone. Thus, political groups or “politicized ethnic groups” are often the target of the events captured in this dataset.

This begins to suggest that taken as a whole, the phenomena we are analyzing here may not fit the conceptions we draw from events such as the Holocaust or Rwanda. Here I adjust the emphasis on logic already hinted at in Harff 2003 to examine a “strategic” view of genocide. As suggested by Fein (1993) and Harff (1987), regimes are more likely to use tools of mass repression, such as genocide, when challengers to a state pose a greater threat. This is also close to the view proposed by Valentino et al. (2004), who find that “mass killing” is more likely to occur when states face threats from guerilla insurgencies – a situation in which attacking whole communities becomes an appealing counter-insurgency strategy. Similarly, Downes (2008) also

argues that civilians are targeted during war by regimes under threat, and that attacks on civilians are particularly likely when a regime is at the greatest risk of losing militarily.

Taking a similar view, I emphasize here a "strategic" model of genocide as a tool sometimes employed by certain types of regimes to fend of significant challengers. I next explore this logic to identify additional possible "warning signs" that, if this theory were true, would be expected to be useful in predicting genocide.

Does Ethnicity Matter?

What role might ethnicity play in this model? Under the strategic view, the presence of ethnic divisions and polarization within a state may matter through several channels. For example, ethnically organized politics could generate potential challenges to ethnically exclusive or unrepresentative governments, as proposed by Harff (2003). Strong ethnic polarization may make it easier for the state to mobilize groups against each other to employ a genocidal strategy against an insurgency or other threat. More simply, any strong identity such as ethnicity or race can become the basis for strong in-group versus out-group distinctions to take root, at which point grievances and animosity between these groups can easily grow.

It is difficult if not impossible to imagine cases such as Rwanda, the Holocaust, or Bosnia without reference to ethnicity. However, it would be un-surprising if measures such as ethnic fractionalization, ethnic character of the ruling elite, and so on turn out to be unpredictive of genocide (conditional on instability). First, identity-based distinctions such as ethnicity may be neither necessary nor sufficient for genocide/politicide to occur. There are many cases of salient ethnic polarization that do not necessarily lead to genocide: the conditions under which a given ethnic group is mobilized to commit mass violence are surely constrained by factors other than the mere presence of ethnic diversity or ethnic minority rule.

Relatedly, even a small group, whether ethnically organized or not, can challenge a weak state, driving a wedge between any country-level measure of ethnic polarization and the "amount" of polarization necessary to achieve mass ethnic violence. Similarly, if a state is threatened, it can use its military (if available) to crush the communities associated with these groups, or find *any* group which can be incentivized to help with the task. Since such a group may again be very small relative to the national population, it may not be hard to find such a group even in a

country that scores low on ethnic fractionalization or other similar measures. In addition, the foot-soldiers needed to commit mass violence can be motivated by economic or other incentives, or their identity-based motives can be manipulated deliberately.

For these reasons, detecting a powerful effect of ethnic fractionalization, polarization, or related variables on the likelihood of genocide is likely to be difficult, even if ethnic fractionalization or ethnic minority rule are mechanistically important. This may be why in Harff 2003, which did find an effect of "ethnic character of the ruling elite" initially, the effect disappears when the data are adjusted to account for timing problems. Moreover, a large number of variables in the PITF dataset related to ethnicity – such as various measures of ethnic fractionalization – fail to ever be predictive.

Predictors under the Strategic View

When and where then would genocides of this sort be more likely to occur, given that a political instability event is underway?

First, much as Harff proposed, states in which the governments need not count on the consent of a broad portion of the population to maintain power may be more able to use genocide or mass killing as a tool to maintain power. As with the Harff model, this view also predicts that autocracies should be more likely to used genocide, but I also expect anocracies to be at greater risk: while anocracies have less consolidated power than autocracies, they can be non-democratic enough that rulers are unaccountable to a large portions of the population. Moreover, anocracies may face even greater fears than autocracies of being ousted than consolidated autocracies. This contrasts with the view of R.J. Rummel (1995), who has argued that increased consolidation of unchecked power - i.e. strong autocracy - leads to increased likelihood of "democide." Under the Rummel thesis, autocracies alone should be more likely to commit such crimes, whereas under the view proposed here, both autocracies and anocracies should be more likely to commit these atrocities.

In addition, for the same reasons that autocracies and anocracies may have a "free hand" to use genocide, a variable measuring "executive constraints" more directly may be useful in addition to or in place of regime type to capture this concept.

Next, I continue using three variables used in Harff 2003. An "ideologically" characterized regime may use its ideological claims to justify and mobilize support against the enemy they seek to defeat by atrocity. Prior upheaval is a potentially useful lagged measure of a state's tendency towards violence or expectations of future challenges to its power. Both past violence and the expectation that more state challengers may soon come along may make the strategy of genocide more palatable and appealing. Less theoretically, prior upheaval may simply capture otherwise unmeasured causes of violence. Finally, trade openness is a useful proxy for the interdependence economic and political interdependence of the state in question. Under a liberal view of international relations, this may be viewed as a proxy for how well the state has been "socialized" into an international community, which generally prohibits the use of genocide. Alternatively, in recent decades the use of genocide and other massive human rights violations is often met with financial and trade sanctions. Trade openness may thus be a proxy for the opportunity cost of using violence that may lead to costly economic damages.

I also expect that even conditional on political instability events, weaker states are more likely to respond to insurgencies with genocidal strategies because they are less likely to have the means of winning such a war otherwise. Genocide can be an inexpensive counter-insurgency tactic, especially (but not only) when there are proxy groups that can readily be mobilized - through ethnic animus, economic incentives, or both - to attack the communities supporting the opposition movement against the state. Violence in Darfur, Sudan in 2003 and 2004 resembles this logic: the so-called "Janjaweed" militias were readily mobilizeable by the government of Sudan, partly by stoking ethnic animus and economic tensions, but also by offering these militias weapons, salaries, land, and even life insurance for their families, while offering government posts to the leadership of these groups. Alex de Waal (2004) has aptly termed the Darfur genocide "counter-insurgency on the cheap."

That said, the role of a state's military strength in particular may be conditional on the nature of the regime, or more specifically, on the degree of constraint on the executive (Collars & Carey 2008). Unconstrained executives with larger militaries may be more likely to use these militaries to commit genocide, whereas more constrained executives with larger militaries will refrain from doing so, and may be able to use these militaries instead to protect civilians. To capture this

potential effect, the interaction of state military power with executive constraint must also be included.

Finally, the above influences are much more likely to trigger genocide when the political instability event in question is an ethnic or revolutionary war, rather than an "adverse regime change." Many of the events coded as adverse regime changes in the PITF dataset are discrete and short coups. Certainly coups can trigger genocide, if the new regime attempts to eliminate political threats or the ousted regime attempts to respond by mobilizing a community with genocidal intent. However the majority of these cases involves relatively few players and remains in the realm of elites. Thus I expect political instability events beginning with wars of either type to be more likely to trigger genocidal strategies.[12] While ethnic war may be somewhat more likely to lead to genocide than revolutionary war because it suggests a readily available mechanism for targeting communities, given the difficulty of distinguishing reliably between "ethnic" and "revolutionary" wars, I do not use the distinction here. Note that given the goal of producing practical tools for real-time genocide forecasting, a useful model must be based only on information immediately available at the outset of a political instability event. Thus the only information of this sort that can enter the model is whether a political instability event *began* with regime change or with ethnic or revolutionary war.[13]

Operationalization

In the model below, the variable *war at outset* indicates whether the political instability event began with a war (revolutionary or ethnic). Aside from this variable, which is measured at the very outset of instability, all other covariate values are taken from the year prior to political instability onset. I measure *prior upheaval* and *trade openness* using exactly the same variables as the (corrected) Harff 2003 model above. I measure *autocracy* as a dummy for whether the

[12] It has also been shown (Valentino et al., 2004) that guerilla insurgencies in particular are more likely to result in the use of mass killing. It would likely be useful to include an indicator for guerrilla insurgencies in this model, however, since the use of such tactics is generally not knowable at the outset of an instability event, I do not include it here since that information would not be available to analysts in time.

[13] The greater practical challenge, however, is knowing whether a political instability event has actually begun or not by the end of a coding year. If we must wait one year or more to be able to retrospectively code an event as a political instability, then no model conditioning on political instability events can be fully used in real-time.

polity2 score (Polity IV project, 2009 edition) is -5 or less.[14] I measure *anocracy* as whether polity2 score between -4 and 5 inclusively. "Executive constraint" is measured using the *XCONS* from the same dataset. For simplicity, I use a median-split to produce a binary variable indicating which states have executive constraints at or above the "+4" level, indicating "the emergence of real, albeit limited, constraints on executive authority" (Polity IV codebook).[15]

As a rough measure of state strength, I have opted to use iron and steel production, *IRST*, from the Correlates of War data on National Capabilities (version 4). To examine how the effects of domestic military strength on the likelihood of genocide may be conditional on executive constraints, I use military personnel, *milper*, from the same dataset. Again, I use a median-split of *military personnel* to create a binary variable indicating which states have militaries of over 35,000.[16]

The model I use is a logit at the instability-event level, for consistency with Harff 2003, but with robust standard errors. Below I show the corrected and restricted version of the Harff 2003 model as reported earlier, together with an initial version of the strategic model following the above logic, plus a second (final) strategic model that includes the interaction of military strength with executive constraint.

[14] While the different definition used in this model relative to Harff 2003 is inconvenient, I changed the threshold to accommodate the inclusion of a separate anocracy variable ranging from -5 to 5 on Polity2.

[15] Available at http://www.systemicpeace.org/inscr/p4manualv2009.pdf

[16] Military personnel per capita would also be a sensible measure, and was used by Colaresi and Carey (2008). I find virtually identical results whether using the per capita measure or the absolute measure., so I use the simpler absolute measure.

Table 3. Strategic Model

	I. Corrected, Restricted	II. Strategic Model 1	III. Strategic Model 2
Variable	Coefficient (p-value)	Coefficient (p-value)	Coefficient (p-value)
Years of political upheaval in last 15 years	0.048 (0.050)	0.035 (0.107)	--
Ideological orientation of the ruling elite	1.32 (0.015)	2.88 (0.001)	4.01 (0.001)
Autocracy (polity2<=0)	1.76 (0.001)	--	
Autocracy (polity2<=-5)	--	2.57 (0.001)	6.62 (0.000)
Anocracy (-5<polity2<=5)	--	1.84 (0.018)	3.55 (0.000)
Trade openness	-1.53 (0.001)	-2.92 (0.000)	-4.40 (0.000)
War at outset	--	1.69 (0.016)	2.41 (0.03)
IRST/100	--	-0.0457 (0.000)	-0.069 (0.000)
High Executive Constraints	--	--	5.16 (0.000)
High Military Personnel	--	--	1.66 (0.063)
*High Mil. Per.*Exec.Const*	--	--	-3.38 (0.009)
Constant	2.91 (0.01)	6.29 (0.001)	6.38 (0.009)
N	124	124	124
Adjusted-R^2	0.27	0.41	0.54

The two new models fit the data considerably better than (corrected) Harff 2003 models. While they include more variables, all coefficients are highly significant, and the R^2 is increased substantially from 27% in the corrected Harff model (or 28% in the original) to 41% in the first strategic model and 54% in strategic model 2.

What is not in the Strategic Model?

A few notes are warranted on variables not included in the models shown above. First, the variables originally in Harff 2003 that lost significance upon correcting the data - *ethnic character of the ruling elite* and *prior genocide* - remain widely insignificant if added in the above models, with p-values of 0.5 or higher. Also interesting is that *prior upheaval* begins to lose power in strategic model 1 (p=0.11), and was further weakened (p=0.22) when added to strategic model 2, where it improved R^2 by only 1%, and thus was dropped. Also not shown are interactions of *executive constraint* with polity type, which were never significant here, in contrast to the findings of Aydin & Gates (2008). Instead, *executive constraint* is highly correlated with regime type, allowing the two to nearly proxy for each other.

Finally, while *IRST* is one possible proxy for a general notion of state capacity or strength - and was originally motivated by a focus on measures related to military power - other measures might be considered as well. An alternative measure from the Correlates of War dataset, composite national capabilities (*CINC),* is a summation of *IRST* together with other measures, and was also significant if used instead of *IRST*.[17] However, all other indirect measures of state strength subsequently tested, - such as *GDP*, *GDP per capita,* and *infant mortality*, proved insignificant when added to the above models. This stands in contrast to Easterly et al.(2006) which found that incomes do matter, even conditional on regime type. One reason why variables such as *GDP per capita* may not be significant here despite the expectation that they would be is that we are already conditioning on political instability.

What is in the Strategic Model

As expected, in both strategic models, *ideological orientation of the ruling elite* remains positively associated with the likelihood of genocide, as are *autocracy* and *anocracy*. The presence of war (ethnic or revolutionary) at the onset of instability also strongly predicts increased likelihood of genocide in both strategic models. Increased state capacity (proxied by *IRST*) is associated with decreased likelihood of genocide, again consistent with the strategic theory, in which weaker states are more likely to resort to genocide, though the conditional

[17] Initially, I used *CINC* instead of *IRST,* however, *CINC* is a summation of several variables and upon closer inspection *IRST* was the most important component.

effects of military strength qualifies this finding (see next). Finally, trade openness is again associated with a decreased likelihood of genocide.

In strategic model 2, a conditional relationship between *executive constraints* and *military personnel* appears. Without the interaction term, neither *executive constraint* nor *military personnel* have significant coefficients. Once the interaction term is added, this changes substantially and R^2 is boosted considerably (from 0.41 to 0.54) indicating an important conditionality in the role of these variables. The main effect of *executive constraint* is the opposite of that expected. However the negative effect of the interaction of *executive constraint* and *military personnel* is particularly interesting. Among regimes with unconstrained executives, increasing the size of the military is associated with a greater likelihood of genocide. However, among regimes with high *executive constraints*, increasing the size of the military is associated with a decreased likelihood of genocide. Thus the logic proposed above, in which genocide is a tool used by states with weaker militaries, is true only for more constrained executives. This conditional nature of the size of the military on the probability of genocide is consistent with Colaresi & Carey (2008).

It is worth emphasizing here the caveat that while the models above indicate that the data are broadly consistent with the strategic theory, the aim here is to achieve improved forecasting, not to make theoretical or causal claims. Numerous alternative theoretical interpretations of these findings are possible. *IRST*, for example, could simply be measuring otherwise unobserved features of state or economic strength, which could matter through a number of channels (though neither *GDP* nor *GDP per capita* are significant when added to the model, and their inclusion alter the effect of *IRST*). Similarly, *trade openness* may be a proxy for the presence of successful economic and industrial activities that produce exportable goods and the demand for imports. *Autocracy* and *anocracy*, too, may be picking up on some otherwise unmeasured form of state weakness that relates to genocide for reasons not associated with the lack of democracy itself.

Performance of Strategic Model

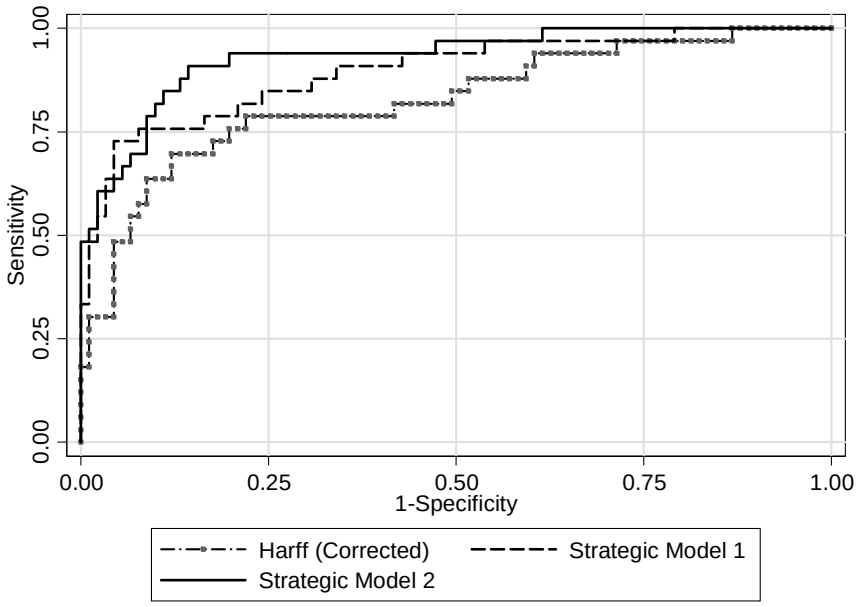
How do these models compare in terms of binary classification success?

Table 4. Comparing Performance of the Models

	Harff 2003 Corrected	**Strategic Model**	**Strategic Model 2**
$\boldsymbol{Alarm\ if\ P(gen) \geq 0.25}$			
% all cases correctly classified	96 (77%)	99(80%)	106 (85%)
% genocides correctly classified	27 (82%)	27 (82%)	30(91%)
% of alarms that are true	55%	59%	67%

At the $P(gen) = .25$ thresholds used in Harff 2003, both strategic models show significant improvements in performance. Rather than choose specific thresholds for comparison, however, it is more informative to compare different models' *sensitivity* (probability of a warning, given that genocide occurs) and *specificity* (probability of a genocide occurring, given that a warning was made) over all thresholds using a Receiver-Operator-Curve. Below I show the results for the (corrected) Harff model compared to the Strategic Model, using the second specification.

Figure 1.ROC of (Corrected) Harff and Strategic Models



Note that both strategic models1 and 2 are preferable to the (corrected) Harff 2003 model at every possible threshold, since for any level of specificity, the sensitivity is as high or higher for

the strategic models. Among the two strategic models, model 2 is clearly preferable at most points, though there are some cross-overs. The total predictive power of each model can be summarized by the area under the curve: strategic model 1 captures 90% and strategic model 2 captures 93%. Both of these are significantly greater than the 83% captured by the Harff model.[18]

[18] χ^2 test of difference in these areas is statistically significant, p=0.01 for strategic model 1 and p=0.003 for strategic model 2.

III. Out-of-Sample Tests

The performance measures above are useful for diagnostic purposes, but it is more important to know how well a model trained on one set of data will predict the outcome on a new data point not before seen but (presumably) drawn from the same joint distribution. In this section I show the out-of-sample performance by two methods.

Out-of-Sample Performance Using Newly Available Data

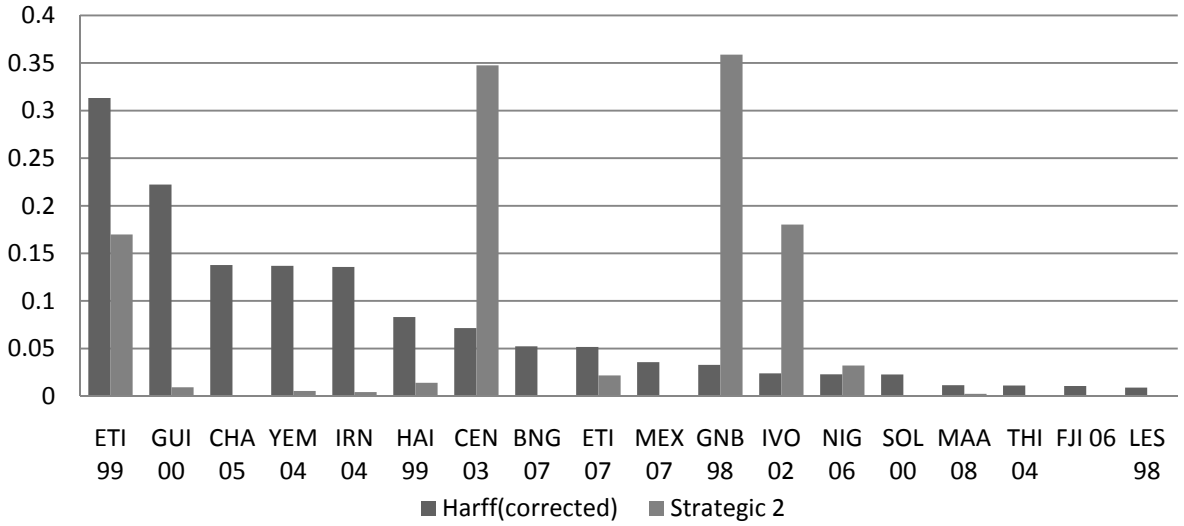
The first test of out-of-sample performance will be based on data that have become available since Harff 2003. The data used for Harff 2003 stop in 1997, with the exception of the 1998 instability event in Yugoslavia. Since the PITF data are currently updated to 2009, we can thus add 11 years of data, in which 18 instability events have begun:

Table 7. Recent Instability Events

Country	**Type(s) and year(s)**
Bangladesh	REG (1/07)
Central African Republic	REG (3/03), ETH(6/05)
Chad	ETH (10/05)
Congo-Brazzaville	REV (6/97), REG (10/97)
Ethiopia	ETH (2/99)
Ethiopia	ETH (1/07)
Fiji	REG(12/06)
Guinea	REV(9/00)
Guinea-Bissau	REV(6/98), REG(6/98), REG(9/03)
Haiti	REG (1/99)
Iran	REG (1/04)
Ivory Coast	ETH (9/02), REG(9/02)
Lesotho	REG (5/98)
Mauritania	REG (8/08)
Mexico	REV (5/07)
Solomon Islands	REG (6/00)
Nigeria	ETH (1/06)
Thailand	ETH (1/04), REG(9/06)
Yemen	REV (6/04)

One may ask how models trained on the original dataset perform on the 18 new instability events. Since no genocides occurred on any of these new events,[19] this test is somewhat limited. Nevertheless, the estimates produced for each of these events using models trained only on the original 124 datapoints are shown in figure 2.

Figure 2. Probability of Genocide for 18 New, Out-of-Sample Instability Events



The estimates from the Harff model tend to be somewhat more spread out, while those from strategic model (2) tend to "focus" more on four cases: Guinea Bissau 1998, Central African Republic 2003, Ivory Coast 2002, and Ethiopia 1999. The case with the greatest disagreement is Guinea Bissau, 1998. Using 1997 data, Guinea Bissau has none of the risk factors appearing in the (corrected, reduced) Harff model, and thus produced a very low probability of genocide by that model (3%). However, that instability event begins with a war, the regime is characterized as an anocracy, and *IRST* is very low, leading to a high estimated probability of genocide under the strategic model.

[19] One genocide – Sudan (Darfur) 2003 – occurred in this period. However Sudan was still listed as having an ongoing instability event that began in 1983 and had already experienced a genocide beginning in 1983. As such, there is no way to model that event. That said, we can ask how the models would have performed if, say, the signing of the Comprehensive Peace Agreement in 2002 had been seen as the end of the ongoing political instability event, which would have allowed the 2003 outbreak of violence in Darfur to be considered a new instability event. In fact, both the corrected Harff model and Strategic model predict that genocide would have occurred in this case with a high probability.

Summing the predicted probabilities produced the "expected" number of genocides based on both models. The (corrected, reduced) Harff model predicts 1.4 genocides, whereas strategic model 2 predicts 1.1. Thus both models "over-predict" somewhat, but this is to be expected since none of the instability event in fact led to genocide.

Hold-Out Tests Using New and Old Events

Given that there were no genocides in the post-1997 data, the out-of-sample test above is of limited usefulness. Another way of using these data for out-of-sample testing is to augment the dataset with them, then split the resulting 142 observationsrandomly into a "training" set and a "hold-out" testing set. Here I segment the data into a training set with 124 observations and a test set with the remaining 18.[20] Iterating this 500 times, I examine the predictive performance on average.

Table 8. Out-of-Sample Performance

	Harff (corrected)	**Strategic Model 2**
Threshold: $\boldsymbol{Pr(gen) \geq 0.25}$		
Ave. correct classification rate	78%	83%
Ave. sensitivity	0.73	0.76
Ave. specificity	0.54	0.62
Threshold: $\boldsymbol{Pr(gen) \geq 0.50}$		
Ave. correct classification rate	82%	86%
Ave. sensitivity	0.46	0.62
Ave. specificity	0.65	0.73
Ave. RMSE	0.084	0.074

[20] I chose to use 124 training observation so that the results would be comparable to the in-sample analyses above, but the findings below are robust to different choices, all the way down to leave-one-out testing.

These results are promising in that the classification rates are close to the rates found in-sample.The strategic model out-performs the (corrected, reduced) Harff model at both thresholds shown above, with improvements by roughly 5% or more in each category. At the 25% threshold, for example, the strategic model correctly classifies 5% more cases (83% versus 78%); is slightly more sensitive (detecting 76% rather than 73%); and has 8% better specificity (62% versus 54%), indicating greater confidence can be afforded to the estimate. Results at the 50% threshold are similar. The important point about these findings is that they tell us about the expected performance of our model on new data, whereas the in-sample diagnostics always run the risk of being the result of over-fitting to idiosyncrasies.

IV. Within-Instability Time-Series

Predicting the instability-wise probability of genocide imposes several limitations. First, knowing whether genocide is more or less likely during an instability event is not always informative, if that instability event goes on for decades (Indonesia, Burma, and Uganda, for example have experienced instability events lasting over 40 years). Second, the instability-wise approach cannot employ time-varying information that may change the likelihood of genocide during an instability event. Third, these models only tell us the probably of there being *at least one* genocide during a given instability event, and are powerless to predict a "second" or "third" genocide during a single instability. This problem is compounded by the ambiguity over what constitutes a new instability and what is a continuation. Finally, the model can only be meaningfully applied once instability begins, by which time it may be too late.

One alternative approach that would avoid these problems would be to use a model based on country-year data, without conditioning on instability at all. This would allow analysts to determine the probability of genocide in the next year (or an alternative choice of time window), and can be employed whether instability is ongoing or not. While this is very reasonable and should perhaps be our goal, one shortcoming is that since genocide virtually always occurs during instability events - and only about 10% of country-years involve instability - it may not be possible to find a model that is very effective without first conditioning on instability. A model with all country years from 1955 to 2008, for example, would have an unconditional probability of genocide of about 0.4% per country year. Thus a "dummy" model that always predicts "no genocide" would be correct 99.6% of the time, and even with an excellent model, predicted probabilities of genocide are likely to remain very low.[21] Moreover, missing data problems would be far greater for this model than one in which we only have to collect the 10% of country-years in which instability events are occurring.

Given these difficulties, I instead focus on a hybrid approach, a "within-instability time-series" model. Here I continue conditioning on instability events, but run a time-series model within these instability episodes. This still limits the model to periods of instability, but allows analysts

[21] Instability could of course be included as a (time-varying) covariate rather than be used as a conditioning set, however if the model depends strongly on this covariate, then we will have little chance of predicting genocide until an instability event is recorded anyhow. If this covariate is lagged one year, then predicting genocides in the first year of instability would be particularly problematic.

to generate predicted probabilities over a specified period of time (e.g. the next year) during the instability, and would take advantage of time-varying information - especially the expected decreasing likelihood of genocide as an instability event ages. It also allows the analyst to predict second or third genocides that occur during a given instability event, so long as the prior genocide has ended.

I use logistic regression to regress genocide *onsets* on lagged covariate data. Only country years during instability are included. Years of *ongoing* genocide are excluded, since onsets cannot occur during these years (new genocides at different localities within a given country are possible, but I am not able to accommodate this complication at present): modeling years of ongoing genocide would be akin to treating each year of ongoing genocide as a new, independent onset. Following Beck et al. (1998) I include in the model a flexible specification of the variable representing the "time-since-failure," or t, to create a binary time-series cross-sectional (BTSC) model. This both addresses serial correlation concerns and allows for a baseline hazard rate that can vary in either direction or non-monotonically over time. Following Carter & Signorino (2011), I use t, t^2 and t^3 rather than dummies or splines to accomplish this. In this case, the variable "t"is the number of instability years since the last onset of genocide, as this is the "time since the last failure" in the sense of a hazards model, where "failure" is the onset of genocide. Though cases exit the dataset once genocide occurs, if genocide ends while the instability case continues, they re-enter the first year of non-genocide, and the counter t is reset to '1' at that time. In addition, I include $(year - 1955)$ and $(year - 1955)^2$ to capture secular time-trends.

The resulting logit model is difficult to interpret, owing to the number of interactions terms and polynomials, particularly for the temporal variables. Thus in table 9 I show only the results for the non-temporal variables. The effect of the temporal variables will be made more evident in marginal effect plots to follow, though they are intended only for predictive and not causal or theoretical interpretation. All variables below, are lagged by one year and standard errors are clustered on country.

Table 9. Models for time-series within Instability

Variable	Logit $\hat{\beta}$ (p-value)	Pr(gen) when set to low value[a]	Pr(gen) when set to high value
Autocracy	-0.93 (p=0.28)	2.0%	0.8%
Anocracy	0.19 (p=0.70)	0.97%	1.2%
Ideological orientation of the ruling elite	0.96 (p=0.034)	0.79%	2.1%
War	-1.30 (p=0.009)	2.5%	0.79%
Milper/100	0.23 (p=0.013)	0.62%	0.99%
Executive Constraints	-0.34 (0.038)	2.4%	0.63%
*Milper*Executive Constraints/100*	-0.083 (0.013)	2.2%	1.3%
N	1041		
R^2	0.18		

a. These values indicate the estimated probability of genocide with all other variables set to their mean and the variable of interest set to 0 (for binary variables) or the 25th percentile (for continuous/ordinal variables). The next column shows the same but at the higher value ('1' or the 75th percentile). Thus the difference between these two columns is the "first-difference" with all other variables set to their means.

The model fit is generally much weaker than the cross-sectional model, indicating the greater difficulty of estimating the next-year probability of genocide onset compared to the instability-wise probability of genocide. As shown in the last two columns of table 9, even variables with the largest coefficients are associated with an absolute change in the probability of genocide of about 1.5%.

The polity variables *autocracy* and *anocracy,* so useful for predicting instability-wise genocide, show no significant effects in this model. All variables from the original Harff 2003 model which

had lost significance in the (cross-sectional) strategic model (*priorgen*, *ethnic orientation of the ruling elite* and *prior upheaval*) remained widely insignificant here and were thus dropped from the above model. *IRST,* which was strongly predictive in the strategic model, also lost all significance[22] in the time-series model and was subsequently also dropped.

Ideological orientation of the ruling elite, however, remains a useful predictor, with more ideological regimes being more likely to experience a genocide onset in the next year of instability. Most surprising, however, is the reversal of the sign on *war*. Rather than measuring "war at onset", this variable is now a lagged indicator for the presence of either revolutionary or ethnic war. Whereas such wars predicted a strongly increased probability of genocide during an instability event in the instability-wise model, here it predicts a significantly decreased probability of genocide in the next year. This is consistent across all models attempted. While confusing, this is not unreasonable. In a cross-sectional (instability-level) setting, the presence of war at the onset predicts that over the course of that entire instability event, genocide is more likely. However, instability events with ongoing wars are likely to be far longer than those without wars. The average age of an instability event among time points on which war is occurring is 12.1 years, whereas the average age of instability events on time points during which war is not occurring is just 0.25 (difference significant at $p<0.000$). This produces a scenario in which *war* is associated with a lower likelihood of genocide per year, even though it is associated with a higher likelihood of genocide over the course of a whole instability event.

Using ordinal rather than dichotomous measures of *executive constraint* and *military personnel*, these two variables together with their interaction remain relevant in the time-series context. The significance of *executive constraints* is surprising given that the more traditional polity variables were not significant. The interpretation is mostly, but not entirely, consistent with the cross-sectional findings: Among unconstrained executives larger armies are associated with an increased likelihood of genocide, suggesting the willingness to use these forces abusively. By

[22] If the interaction effect between executive constraint and military personnel is left out, *IRST* retains significance with the expected negative sign.

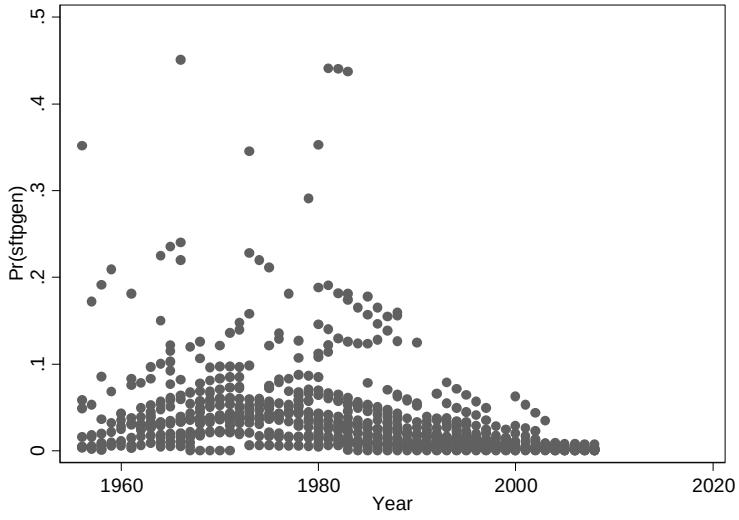
contrast, among regimes with constrained executives, the relationship is reversed, and they are more likely to use genocide when their militaries are smaller.[23]

Effects of Time

Not shown in the above tables is the effect of temporal variables on these probabilities. Figure 3 below shows a simple scatter plot of the estimated probability of genocide over time.

Figure 3. Scatter Plot of Estimated Probability of Genocide by Year

Logit model



Moreover this change over time is not merely the result of there being fewer instability events, or of changes in the covariates that seem to drive the likelihood of genocide. Instead the decrease in the likelihood of genocide conditional on instability persists after controlling for covariates, and can be decomposed into a secular decrease in the likelihood of genocide over the years, and a smaller drop in the likelihood of genocide as instability events age. The two relationships are shown separately and then jointly in figure 4, together with 95% confidence intervals.

[23] This differs from the previous finding only in the direction of the coefficient on the main effect of executive constraint – in this case, more constrained executives are less likely to use genocide (in the next year) than unconstrained executives, even when military personnel is very low.

Figure 4A. *Probability of Genocide by Year, Conditional on Instability*

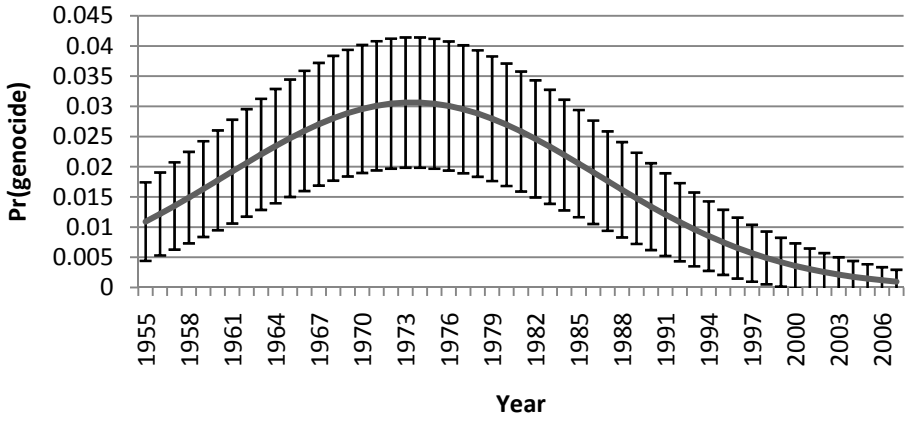


Figure 4B. *Effect of Instability Event Age*

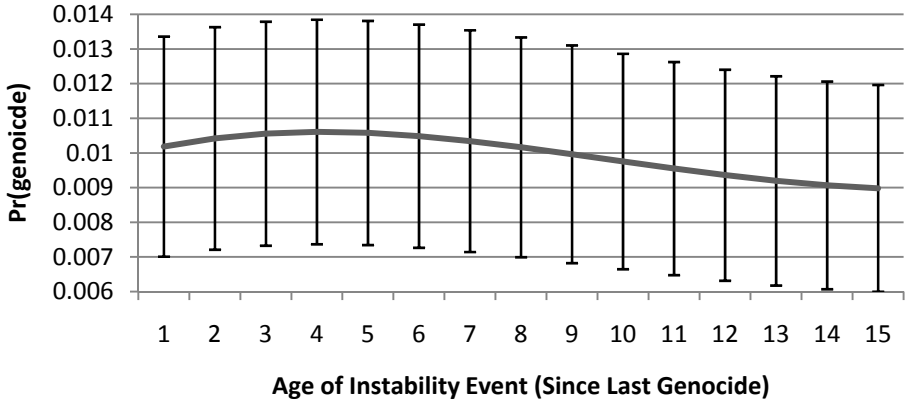
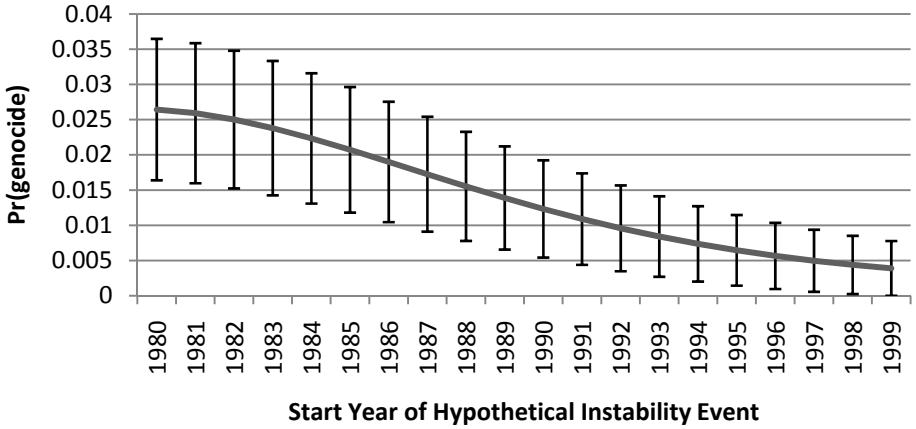


Figure 4C. *Combined Effect for an Instability Event Beginning in 1980*



*Note: The confidence intervals shown above are approximate only for present purposes.

With other variables at their means, the secular time trend shows an increased conditional probability of genocide until the early 1970s, followed by a rapid decline that continues to the present (figure 4A). This decrease, from 3% in the mid-70s to less than 0.5% in the 2000s, is more dramatic than any of the first differences associated with the non-temporal covariates reported earlier. The age of an instability event may also be related with a reduced probability of genocide beginning, though this relationship is quite weak (figure 4B). The combined effect of increasing both calendar time and instability age, as they would increase together for any observed event, is also strong: for a hypothetical instability event beginning in 1980, with all other covariates set to their means, the probability of genocide begins at 2.6%. By 1990 the probability drops to 1.2%, and by 2000 it is less than 0.4%.

Predictive Accuracy of Time-Series Model

Finally, how well does this model predict next-year onsets of genocide? This task is much more difficult and produces lower performance than the instability-wise model. Given that the estimated probabilities are quite low in all cases (the 95th percentile is 0.12), I use a low threshold of $P(gen) \geq 0.10$ to test classification rates.

$\boldsymbol{Alarm\ if\ P(gen) \geq 0.10}$	Time-series, Logit	Dummy Model
Total correctly classified	93%	96%
Genocides correctly classified	56%	0%
Percent of alarms that are true	21%	NA

The low threshold that must be used can be expected to produce relatively low specificity. In this case, detecting even half of genocides comes at the cost of finding roughly five false-positives for every one true-positive. Nevertheless, for such a daunting task, the model is relatively effective, and combined with the instability-wise estimate of the likelihood of genocide, perhaps this information can be useful for prioritizing potential diplomatic and other interventions.

Where are these false-positives coming from, and are many of them simply cases when a warning was given too early, but genocide did occur within a few years? Of the 67 relevant

instability-years[24] for which data are available and the probability of genocide in the next year is estimated to be at least 10%, 14 of them were correct. Of the remaining 53 "false-alarms," a large portion were simply "too-early." Six genocides were predicted too-early at least once, accounting for 17 (32%) of the false-positives. On average these came fewer than 3 years before the actual genocide. However, eight of these false-alarms come from a single event: the model forecasts that the 1980 instability event in Iraq would experience genocide each year until 1988 when it is actually coded as beginning. If we relax the testing procedure slightly, among the country-years predicted to have a genocide with probability 10% or more, 21% experienced genocide in the next year, and 27% experience it within the next 3 years.

V. Conclusions

The ability to understand the risk of genocide in a particular place and time is potentially useful for deploying preventive efforts. Harff 2003 is the only published paper quantitatively endeavoring to forecast genocide, and by far the best-known and most widely-cited forecasting model. That said,it is unclear what quantity is being estimated by that model. Once adjusted to produce an interpretable quantity - the probability that genocide occurs at some point during a political instability event - its performance remains relatively good, but neither *prior genocide* nor *ethnic character of the ruling elite*remain statistically or predictively significant.

The view of genocide as a strategic choice made by regimes to fight off potential challengers is by no means new, and is to some degree considered in the Harff 2003 model itself and in work by Fein (1983) cited therein. Moreover, it is a major theme in much of the more recent empirical work on genocide and mass killing (e.g. Valentino et al. 2004, Esteban 2010). I take this view in order to develop a model that can be used for prospective forecasting. The resulting model indicates that regime type, the presence of war at the onset of instability, low state capacity, ideological character of the ruling elite, and trade-openness are all strong predictors of genocide. In addition, the presence of a large military is associated with greater likelihood of genocide

[24] "Relevant instability-years" are those country-years during an instability even but not during ongoing genocide and for which all data are available. Note that overall, missing data is not a major problem: of the 1115 potential time points, 1041 are able to be modeled.

under weakly constrained executives, but a lower likelihood under highly constrained executives, consistent with other recent findings (Colaresi & Carey, 2008).

This "strategic" model out-performs the prevailing model on all measures tested and at all thresholds, both in-sample and out-of-sample. Using a threshold of a predicted probability of 25% and the same sample as Harff 2003, the final strategic model used here correctly classifies 85% of instability cases according to whether they experience genocide, correctly detecting 91% of genocides. Of warnings issued by this model, 67% were true. As expected, out-of-sample performance is lower, but only slightly (correctly classifying 83% of cases, detecting 76% of genocides, and with 62% of warnings being true).

Finally, I examine a hybrid time-series model estimating the probability of genocide for each year during instability events. This allows the model to use time-varying information during instability, and allows the analyst to determine the probability of genocide over a known time-span rather than an instability event of unknown length. The probability of genocide is found to decrease rapidly in time - both in terms of calendar time and the age of an instability event since its last genocide. State capacity is no longer a significant predictor, and counter-intuitively, the presence of war predicts *decreased* likelihood of genocide, perhaps because it is strongly associated with longer instability events which have a much lower per-year probability of genocide. Ideological character of the ruling elite remains significantly associated with the probability of genocide. The interaction of military personnel and executive constraints is found again here, and adds considerable power to the model. As expected, attempting to predict the next-year probability of genocide onset produces far lower probabilities and weaker performance. However the model still produces surprisingly useful forecasts: at the 10% threshold, it predict 56% of genocide onsets, and 21% of the warnings are followed by genocide then next year, with 27% following genocide within 3 years.

What does this mean for analysts and policymakers? At minimum, analysts responsible for examining the probability of genocide for their own agencies or for distributable products like the Atrocities Watchlist or the Instability Watchlist issued by the National Intelligence Council should be aware that (a) the Harff 2003 model must be adjusted to produce the "instability-wise" probability of genocide, and more importantly, cannot be used to produce the "next-year"

probability of genocide; and (b) once the model is adjusted to produce a meaningful quantity, two seemingly intuitive "warning-signs" (*prior genocide* since 1955 and *ethnic character of the ruling elite*) no longer demonstrate any predictive power. Moreover, the strategic model proposed here, while tentative, shows improved predictive performance. These lessons are hopefully of value not only to quantitatively-oriented analysts but also to qualitative ones who use the warning signs derived from empirical work. Finally, while the time-series model employed here is less powerful than the "instability-wise" model, combining analyses from the two models can assist policymakers in understanding the medium-term and short-term risks of genocide for policy planning purposes.

VI. Future Directions

Much remains for investigators to discover and explain. The present analysis has some serious limitations. First, all models here are conditional on political instability events. This improves accuracy considerably but means that analysts cannot use these models until instability erupts. More troubling, instability events may not be codeable as such until one or several years after the fact, in which case any approach that depends on them can be problematic. Second, while the "next-year" probability of genocide is very specific, such models are inherently less accurate than "instability-wise" models. A hybrid approach that does not require conditioning on instability and which examined the "3-year" or "5-year" probability of genocide may be the most useful but tractable approach for future work.

This analysis relies mainly on "structural" or slow-changing data such as regime type, trade-openness, etc. Another important approach - closer to that frequently undertaken by analysts - is to consider potential "triggers" such as contested elections, the death of a leader, the emergence of hate speech, protests or riots that may begin over economic or other issues, etc. Understanding how these types of events alter the likelihood of genocide or mass killing remains an important area for research. This is an area where statistical inference could potentially shed light on the types of triggers and warning signs that ultimately prove predictive and those that generate more false-alarms than true-alarms.

The choice of dependent variable remains another area for further exploration. The outcome of interest here was genocide/politicide as defined previously by Harff. However related variables such as "mass killing" (Valentino et al., 2004; also see Easterly et al. 2006 for an alternative dataset) or one-sided killing (Eck & Hultman 2007) may also be useful, and preliminary results show that analyses with these variables may produce different results, and potentially greater forecasting abilities. Particularly important is the extension of such a forecasting model to mass killing committed by non-state actors rather than the state.

Finally, one of the most substantively important effects reported here is the apparent effect of time on the likelihood of genocide. Judging by figures 4A and 4C, genocide is simply becoming far less likely. While it has previously been reported that internal wars in general are decreasing in frequency (e.g. Mack 2005), the finding here shows that genocides too are becoming less frequent, and that this decrease cannot be due to the changing rate of instability events or changes in known covariates. This is certainly good news, but it is also an important opportunity to attempt to discover the causes for this major substantial change. A number of explanations are possible, including differences in the willingness of the international community to impose sanctions on such crimes, the increased use of mediation third-party involvement in settlements, or increased media coverage and access to technology by would-be victims. Arbitrating among these and other explanations remains an important area for future research.

References

Albright, Madeleine & William Cohen (co-chairs). 2008. *Preventing Genocide: A Blueprint for U.S. Policymakers*. Genocide Prevention Task Force, Washington, D.C.

Alvarex, Alexander. 1997. Adjusting to Genocide: The Techniques of Neutralization and the Holocaust. Social Science History 21(2): 139:78.

Aydin, Aysegul, and Scott Gates. 2007. Rulers as mass murderers: Political institutions and human insecurity. Paper presented at the annual convention of the International Studies Association, Chicago, February 28 to March 3.

Azam, Jean-Paul. 2006. On Thugs and Heroes: Why Warlords Victimize Their own Civilians. Economics of Governance 7(1):53–73

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker.1998. Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable**.** American Journal of Political Science 42(4):1260-1288.

Carter, David B. and Curtis Signorino. 2010. Back to the Future: Modeling Time Dependence in Binary Data. *Political Analysis* 18(3): 271-292.

Colaresi, Michael and Sabine C. Carey. 2008. To Kill or to Protect: Security Forces, Domestic Institutions, and Genocide. Journal of Conflict Resolution, 52 (1): 39-67.

de Waal, Alex. 2004. Counter-Insurgency on the Cheap.Review of African Political Economy. 31(102):716-725.

Downes, Alexander. 2008. Targeting Civilians in War. New York: Cornell University Press

Easterly, William, Roberta Gatti and Sergio Kuriat. 2006. Development, Democracy and Mass Killings. Journal of Economic Growth, 11 (1): 129-56.

Eck, Kristine, and Lisa Hultman. 2007. One-Sided Violence Against Civilians in War: Insights from New Fatality Data. Journal of Peace Research, 44 (2): 233-46.

Esteban, Joan, Masimo Morelli and Dominic Rohner. 2010. Strategic Mass Killings. Institute for Empirical Research in Economics, University of Zurich Working Paper No. 486.

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

Fein, Helen. 1984. "Scenarios of Genocide: Models of Genocide and Critical Responses." In *Toward the Understanding and Prevention of Genocide: Proceedings of the International Conference on the Holocaust and Genocide*, ed. Israel W. Charny. Boulder, CO: Westview Press, 3–31.

Fein, Helen. 1993. “Accounting for Genocide After 1945: Theories and Some Findings.” International Journal on Group Rights 1: 79–106.

Goldstone, Jack A., Ted Robert Gurr, Barbara Harff, Marc Levy, Monty G. Marshall, Robert Bates, David Epstein, Colin Kahl, Mark Woodward, Pamela T. Surko, and Alan N. Unger. 2002. *State Failure Task Force Report: Phase III Findings*. McLean, VA: Science Applications International Corporation.

Hagan, John, Wenona Rymond-Richmond. 2008. The Collective Dynamics of Racial Dehumanization and Genocidal Victimization in Darfur. American Sociological Review. 73(6): 875-902.

Harris, Lasana & Susana Fiske. 2006. Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. Psychological Science 17(10): 847-853.

Harff, Barbara & Ted Robert Gurr. 1998. Toward an empirical theory of genocides and politicides: Identification and measurement of cases since 1945. International Studies Quarterly 32(3): 359–371.

Harff, Barbara. 1987. The Etiology of Genocide. In *The Age of Genocide*, ed. Michael N. Dobkowski and IsadorWallimann. Westport, CT: Greenwood Press, 41–59.

Harff, Barbara. 2003. "No Lessons Learned from the Holocaust: Assessing the Risks of Genocide and Political Mass Murders since 1955," *American Political Science Review*, 97 (1): 57-73.

Humphreys, Macartan, and Jeremy M. Weinstein. 2006. Handling and Manhandling Civilians in Civil War. American Political Science Review, 100 (3): 429-47.

Krain Matthew. 1997. State-sponsored mass murder: a study of the onset and severity of genocides and politicides. J. Confl. Resolut. 41(3):331–6.

Moller, A. C., & Deci, E. L. 2010. Interpersonal control, dehumanization, and violence: A self-determination theory perspective. Group Processes and Intergroup Relations, 13, 41-53.

Rummel, Rudolph J. 1995. Democracy, power, genocide and mass murder.Journal of Conflict Resolution 39 (1): 3-26.

Sambanis, Nicholas. 2001. Do ethnic and non-ethnic civil wars have the same causes? A theoretical and empirical inquiry.Journal of Conflict Resolution 45 (3): 259-82.

Kalyvas, Stathis. 2001. ‘New’ and ‘Old’ Civil Wars. A Valid Distinction? World Politics 54: 99-118.

Kalyvas, Stathis. 2006. The Logic of Violence in Civil War. Cambridge: Cambridge University Press.

Mack, Andrew.(ed.), 2005. Human Security Report. New York: Oxford University Press, two volumes, in press.

Valentino, Benjamin, Paul Huth and Dylan Balch-Lindsay. 2004. Draining the Sea: Mass Killing and Guerilla Warfare. International Organization, 58 (Spring): 375-407.

Wayman, Frank, & Atsushi Tago. 2010. Explaining the Onset of Mass Killings. Journal of Peace Research, 47 (3): 3-13.