Towards a Strategy:
Challenges and Opportunities for Statistical Forecasting of Genocide

Chad Hazlett
October 2011

In this note I briefly summarize the key findings of a longer, more in depth background paper but primarily focus on a discussion of challenges facing statistical forecasting and ideas for implementing a forecasting program that addresses some of these challenges within reasonable budget constraints.

Summary of Findings in "New Lessons Learned?"

To my knowledge, the best known statistical forecasting model for genocide is Barbara Harff's 2003 paper, "No Lessons Learned from the Holocaust?" This is a highly cited paper not only within Political Science, but is arguably the most authoritative and widely-recognized source of information on genocide forecasting among the network of governmental and non-governmental organizations that comprise the "genocide forecasting community." It is therefore important to understand what the Harff model predicts, and how well it performs. In "New Lessons Learned?" I show the following:

1. The Harff model only aims to determine the likelihood of genocide occurring when a (different) political instability event (as coded by the political instability task force) is already underway. However there is confusion over whether the model estimates a "next-year probability of genocide" or the "probability that genocide occurs sometime during a given instability event." As it turns out, the model does not give either quantity.

2. The model can, however, be easily adjusted to produce the instability-wise probability of genocide. Upon doing so, two variables previously thought to be important – *prior genocide* and *ethnic orientation of the ruling elite* – are no longer predictive.

3. A preliminary alternative model (i.e., the "strategic model") proposed here refocuses attention on the idea, supported by considerable scholarship, that mass killing and genocide are strategies used by regimes to deal with insurgencies and other threats to their power or survival. I build a statistical model by integrating results from recent empirical studies and theorizing as to how regime type, the type of instability event, military power of the state, and executive constraints may alter the probability of genocide during instability events. This approach outperforms the Harff (2003) model by a modest but potentially useful margin, in both in-sample and out-of-sample tests.

4. One problem with the approach summarized above is that even knowing the probability that genocide occurs during an instability event may be of limited use, in that instability events may stretch on for decades. If we instead seek to predict the probability of genocide in the next-year

(but during instability events), this is more difficult and models do not perform as well. However some factors, particularly measures of executive constraints, military power, and the presence of war remain statistically significant. One lesson then is that we can provide policymakers with different types of information – such as instability-wise and next-year probabilities of genocide – in hopes that they can specifically target medium and short-term genocide prevention approaches.

5. An unanticipated finding of this study has been that even conditional on instability occurring, and even accounting for all the variables that appear to be predictive of genocide, the probability of genocide erupting during an instability event seems to be decreasing in recent years. Understanding why this is happening would not only be interesting, but may lead us to information about what prevention approaches are proving most effective, and would further improve our forecasting abilities.

Challenges to Statistical Modeling of Genocide and Mass Killing

The analysis in "New Lessons Learned" served to reveal limitations and challenges that generalize to a large class of statistical forecasting models of genocide and mass killing. The major challenges I believe we might face are outlined below:

1. Retrospective Data Coding

There are two major problems with retrospective data coding. The first problem is that post hoc biases can creep into our data when they are coded or recoded after the fact. For example, if we must determine whether a regime exhibited a trait like "exclusionary ideology" in a previous year, knowledge that genocide later occurred may influence our coding and create a spurious relationship between variables that will lead to mistaken inferences and inaccurate prospective forecasts.

The second problem is arguably more insidious: political instability events – and wars in particular – are sometimes difficult or impossible to code until one or several years after they occur.

2. Conditioning on Instability?

Existing models have benefited greatly from the idea that mass killings or genocides typically occur during political instability events; thus if we limit our universe of cases to instability events, we can (i) raise the baseline probability of genocide, helping with the "rare-events" problem and improving the apparent accuracy of our forecasts, and (ii) limit data collection problems since we only need to collect ~10% as much data.

However this conditioning creates several problems. The first is the retrospective coding problem: we may not even know instability events or wars occur until after the fact. Thus if we must condition on such an event for it to enter our model, we may be too late in forecasting

genocide. Could we estimate the probability of genocide in Libya, Yemen, or Syria right now? Recent violence in these states has not been coded as instability events by the political instability task force, and whether/ how they are coded may depend on fatality counts that are yet unknown or will have to cross some threshold to be retroactively labeled as war. We cannot run our models until we know instability is occurring, but that could take a year or more, and meanwhile genocides often occur early during instability events.

The second issue is that conditioning on instability events seriously limits the time horizons of our forecasts since we cannot make meaningful forecasts for states not currently in instability. We could use a two-stage model to first predict the probability of instability, and then the second model to predict the probability of genocide conditional on instability. The product of these two gives the probability of genocide (and instability). However, this compounds uncertainty on uncertainty. One alternative is to include political instability events as a covariate, rather than restricting our universe of analysis to those country-years. This may work well, especially with non-linear modeling techniques, but we again run into the problem that these data are not always available until one or more years following the onset of instability or war.

Another alternative, which is perhaps the most promising, is to look for types of data that approximate the factors underlying the effect of instability on heightened risk of genocide, but which become available quickly. For example, a variable like "the number of fatalities reported due to political violence" can be collected almost continuously. This variable could be used in a more timely way, and as a (potentially continuous) covariate. Moreover, since it is not nearly as discrete as political instability events, it would allow us to see gradual increases in violence, which may be more useful in predicting future genocide than waiting for a given event to cross the political instability threshold in order to enter our model at all.

3. Forecasting Windows and "Structure vs. Triggers"

There is a general question of what "forecasting window" is most useful. Should we estimate the probability of genocide in the next year, in the next 3 years, or during an entire instability event? This is partly a technical issue and partly a matter to be resolved by collaboration with those who would use our forecasts. However I also raise a broader question about two fundamentally different approaches to forecasting, with very different consequences for the forecasting window:

(a) Structural risk models: Our current models focus on slow-changing, macro-level variables, producing estimates corresponding to something like the "structural risk" of genocide. The usefulness of these estimates is probably greatest when they can be used to inform the deployment of genocide prevention strategies that are also very broad and do not require much temporal specificity to be useful.

(b) Trigger models: So far we have not seen much development in modeling approaches that examine fast-changing "triggers" to update our expectation about the likelihood of mass killing. This is problematic in that it cannot inform the use of genocide prevention measures that involve

high temporal specificity, i.e. urgent interventions of any kind to head off violence. Analysts hold numerous ideas what types of events may trigger mass violence (elections in particular), but little work has been done to test these hypotheses. Identifying what triggers truly predict genocide is perhaps one of the most promising untapped areas for further statistical analysis, and would serve both to inform statistical forecasting models and qualitative forecasting efforts.

4. Need for non-linear, non-additive modeling approaches

There is no reason to expect Generalized Linear Models to work well for predicting events that almost certainly do not follow a linear, additive process. Instead, we may anticipate that the marginal effects of some variables depend heavily on the level of that variable, and on the level of other variables, both of which imply violations of the linear functional form. A particularly important violation would be the likelihood of multiple pathways that can lead to mass killing, in which case linear models do a poor job at modeling *any* of the pathways. Thus we may need to find less functional-form dependent approaches. One approach I favor is a kernel-based regularized least squares approach, and in tests using this approach on the variables on the "strategic model" in the longer paper, it performs well relative to GLMs in out-of-sample tests. At any rate, this is a technological issue and, I believe, the most manageable of the challenges we face.

Practical Recommendations

If our objective is to determine what the U.S. Holocaust Memorial Museum can do in terms of a forecasting approach, I would begin by asking about what resources and constraints such a system would face. For present purposes, I assume that the type of system the Museum has in mind would necessarily have the following constraints due to logistical and budgetary limitations:

Constraint 1. Data collection would largely cull data from existing datasets but could also include coding of some new variables, so long as sources for this information are available already. Beginning to include data relevant to the "trigger" approach may also be possible. Some "trigger" data may be meaningful at the annual level (e.g. elections, counts of protests or riots, economic shocks), but it may also be possible and necessary to collect data at higher temporal frequencies based on existing data collection projects such as the Worldwide Atrocities Dataset.

Constraint 2. Updates to risk forecasts based on these new data would likely occur only once or a few times a year, together with an annual review process of the modeling approach itself. I hope it will also be possible, though, to support some discrete research projects that will be necessary to develop the optimal model, discussed below.

Given these constraints and the challenges discussed above, I propose several strategic elements here. I list them roughly in sequential order, though there are components that can be performed in parallel.

1. *First, take stock of the demand for forecasts.*

This task may be achievable by this workshop or through further discussions with some of the people involved. I'm proposing that we need a relatively demand-driven approach: begin with practitioners working to prevent mass killing in order to establish what types of forecasts are useful in combination with different preventive activities. For example, some analysts at the UN or other international or regional bodies may argue that preventing mass killing is a long-term process of engaging with potential genocidaires, identifying potential institutional or political processes that could lead to mass killing in those states, building relationships that can shift norms or provide aid in ways that help avert mass killing, or at least be better positioned to mediate in these states should disaster occur. Such approaches would lead to an interest in something like the 5-year probability of genocide. Alternatively, short-term estimates may be demanded by planners for Mass Atrocity Prevention and Response Operations (MAPRO) in the Office of the Secretary of Defense, or the Combatant Commanders charged with planning for contingencies in their areas of command. My proposition is that a smart approach would require understanding the needs of these different stakeholders and their strategic beliefs in order to tune our analytic approach accordingly and to focus on useful forecasting windows as well as the thresholds used to issue warnings.

2. Consider Some Critical Research Tasks

The challenges mentioned above may have solutions, but we don't know them yet. If possible, it would be very useful to commission a series of targeted studies aimed at addressing these sorts of challenges. For example:

(a) Triggers. We need to know more about what data can be used for "triggers" and which triggers may be predictive. I also argue that it would be useful to work toward an approach in which we can update the output of our "long-models" using information about triggers. That is, first construct one model that gives a best estimate of the probability of mass killing in a given window conditional on the "slow" data alone (i.e., having no information about possible trigger events). Then construct an updating mechanism that allows us to improve that estimate conditional on trigger information as it becomes available.

(b) What to condition on. We need to examine alternatives to the "condition on political instability" approach. In particular we need to examine whether other data – such as the number of recent fatalities in political violence – could (i) address the "retrospective coding problem" associated with relying on political instability events and (ii) allow us to make forecasts in states not currently coded by the PITF as undergoing political instability events.

(c) Set the model. More generally, after conducting the above tasks and others, we will need to decide on the model that we will use going into the future.

(d) Efficacy of prevention activities? Finally, another important area for empirical validation regards the efficacy of policies, practices, and interventions aimed at preventing mass killing. Unless these work, there is little point to forecasting. Despite the importance of this question, however, I worry it would not be a particularly productive area for the Museum to focus on. First, I assume that this question may be somewhat outside the Museum's current aims. Second, inference about interventions of these types -- which are never randomly assigned – is extremely difficult and not well suited to the Museum's comparative advantages.

3. Routine Operation.

The core function of the Museum's early warning system would likely be an annual or twice-annual risk assessment. Each year, estimates could be provided for varying time horizons (e.g. a one-year probability of mass killing and a 5-year probability of mass killing). These would be based on different models (one in which the DV was a binary variable for mass killing in the next year; another in which the DV was a binary variable for mass killing in a 5-year window).

4. Myth-busting?

This may be beyond the Museum's immediate goals but another way in which statistical forecasting approaches could be useful is for "myth-busting." By "myth-busting," I mean that instead of (or in addition to) thinking about, "How do we make the best forecasting model we can?" we utilize statistical methods to ask, "What do people believe predicts genocide, and can we show whether it is likely to be correct or not?"

Analysts working in government, UN, and other offices involved in forecasting activities may hold beliefs about the roles of different state characteristics or triggering events in influencing the likelihood of mass killing. These will often be testable, at least if we set causal inference aside. Eliciting and testing these beliefs could be useful both to our statistical models – producing new ideas for what might be predictive – and to practitioners who may be surprised by the results. It could also generate a nice series of small, accessible publications that help to position the Museum as a source of practical expertise in genocide forecasting.