# CHAPTER 3

# MODELING TERROR ATTACKS: A CROSS-NATIONAL, OUT-OF-SAMPLE STUDY

Ryan Bakker, Daniel W. Hill, Jr. and
Will H. Moore

## ABSTRACT

Purpose – *The purpose of this study is to assess the ability of a theoretically motivated statistical model to accurately forecast annual, national counts of terror attacks out-of-sample.*

Methodology/approach – *Bayesian multi-level models, classification analysis, marginal calibration plots*

Findings – *We find that the model forecasts reasonably well, but conclude that its overall performance suggests that it is not ready for use in policy planning. This is likely due to the coarse temporal and spatial aggregation of the data.*

Research limitations/implications – *The implications of this study are that social scientists should devote more effort into evaluating the predictive power of their statistical models, and that annual, national data on*

**Understanding Terrorism: A Socio-Economic Perspective**
**Contributions to Conflict Management,**
**Peace Economics and Development, Volume 22, 51–69**
**Copyright © 2014 by Emerald Group Publishing Limited**
**All rights of reproduction in any form reserved**
**ISSN: 1572-8323/doi:10.1108/S1572-8323(2014)0000022008**

*violent conflict are probably too coarse to provide useful information for policy planning*.

Originality/value of paper – *The primary value of our modeling effort is to provide a baseline against which to evaluate the performance of more region- and country-specific models to be developed in the future*. AU:1

## INTRODUCTION

There has been a recent surge in interest in forecasting conflict events (e.g., Bagozzi, 2011; Blair, Blattman, & Hartman, 2012; Brandt, Freeman, & Schrodt, 2011a, 2011b; Clauset, Young, & Gleditsch, 2007; Goldstone et al., 2010; Hegre, Karlsen, Nygård, Strand, & Urdal, 2013; Tikuisis, Carment, & Samy, 2013; Ulfelder, 2012; Ward, Greenhill, & Bakke, 2010; Ward et al., 2012; Zammit-Mangiona, Dewar, Kadirkamanat, & Sanguinettia, 2012). Yet there is a past as well: efforts to forecast conflict events such as domestic terror attacks has a considerable history (e.g., Andriole & Hopple, 1984; Andriole & Young, 1977; Gurr & Lichbach, 1979, 1986; Gurr & Moore, 1997),[1] though interest waned rather considerably from the mid-1980s through the 1990s. This recent rise is almost certainly linked to the US government's support of such research (e.g., O'Brien, 2010), and the decline from the 1980s through the early 1990s was likely due to the US government's disinterest in such efforts.

In this chapter we report a forecasting effort that uses a Bayesian, multilevel, Poisson-log normal mixture model of the number of terror attacks that occurred in each country in the world over the period from 1970 to 1996 (Moore, Bakker, & Hill, 2011). To specify the independent variables we developed a theoretical account that, unlike previous efforts, places the competition between dissidents and the state at the center of attention. Krieger and Meierrieks (2011) provide a thorough review of cross-national statistical studies of terror attacks and Gassebner and Luechinger (2011) take a "run a million regressions" approach to determine what independent variables have robust relationships with the number of terror attacks countries experience. Both studies produce the same list of structural characteristics of economies, polities, and societies that populate existing large-N studies. Neither reports variables that measure the behavior of the state or the non-terror behavior of dissident groups. In other words, *all* existing cross-national research studies terror attacks as a function of structure alone. The primary contribution of our work is to bring behavior in. We

argue that terror is a tactic, selected among several, by some dissident groups as they press their claims against the state. If we statistically model the annual number of such attacks per country, then we are modeling the outcome of decisions by dissident groups to use their resources to produce terror attacks rather than use protest demonstration tactics, or guerrilla tactics that seek to degrade the state's coercive capacity (rather than terrorize citizens and state officials).[2] Below we briefly describe the major claims of the theoretical argument, and also sketch both the empirical approach, as well as the findings of our previous study.

There are at least two good reasons to explore a statistical model's ability to forecast.[3] First, the fit of a model to the data is important. In political science we have witnessed a valuable shift away from "maximizing explained variance" that dominated some analyses during the 1980s to a focus on the substantive impact of a variable upon the outcome. Further, hypothesis testing is the primary goal of most academic researchers who estimate statistical models. Thus, attention justifiably focuses upon hypothesis testing and the substantive interpretation of the size of effects. Yet, as Ward et al. (2010) explain, most research tends to overfocus on hypothesis testing to the exclusion of predictive power. That is, while the move away from maximizing measures of in-sample fit (e.g., "explained variance") has been fruitful, a complete lack of attention to how well a model fits the data is lamentable. In particular, few political scientists appear to recognize the value of forecasting to assess the ability of one's model to predict the outcome of interest. Analyzing the individual impact of each variable has value, but is incomplete; the accuracy of the model's predictions remains outside of our view. Further, as shown in Ward et al. (2010), variables that produce statistically significant coefficient estimates do not necessarily improve model fit. In other work we tested our hypotheses and assessed the impact of individual variables (Moore et al., 2011). Had we stopped there we would have left important work undone. Here we extend the assessment of the impact of the variables in our model, shifting away from their individual effects on terror attacks to their ability to collectively predict which countries experience terror attacks.

Second, forecasts can speak to policy makers who might use forecast output in decision support (e.g., Andriole & Hopple, 1984; O'Brien, 2010; Shellman & Moore, 2008). Yet, what policy makers might plausibly be interested in forecasts of the number of domestic terror attacks per country-year? To be able to forecast well the number of terror attacks that are expected to occur over a two-week period in a given city could be of obvious use to municipal officials, but at such a gross level of spatial and

temporal aggregation as the country-year it is less obvious what type of decision support value exists. Yet, the US military, for example, has a number of units that engage in country level forecasting at the annual level, indeed using those forecasts to then aggregate further to both regional and semi-decade levels. In the US military those forecasts have primarily been made by expert human judgment without the support of statistical models, but that has begun to change in recent years. On the civilian side of the US military the J5 Strategic Plans and Policy; Deputy Undersecretary of Defense for Strategy Plans and Forces; the Guidance for the Employment of the Force (GEF); and Guidance for the Development of the Force (GDF) might plausibly find value in forecasts of expected terror attacks per country in the coming year. On the uniformed officer side the Combatant Commands, which coordinate planning across all of the services, would be the most likely consumer. Outside of the military the long range planning divisions of intelligence agencies as well as foreign policy bureaucracies could potentially have an interest.

The remainder of the chapter proceeds as follows. We begin by sketching the empirical model in Moore et al. (2011): the theoretical account we developed, the statistical model we employed, and the results we obtained. The second section describes the approach we take to produce forecasts using the model in Moore et al. (2011). We present the results of the forecast in the third section, then wrap up with a conclusion.

## A BRIEF TOUR OF THE EMPIRICAL MODEL

The theory we use to specify the statistical model that we use to forecast out-of-sample below is described in Moore et al. (2011). That paper also provides details of our Bayesian, multi-level statistical model, the measurement models we estimated, and the data sources used. Here we briefly describe those, referring the reader to our other paper for details.

Our knowledge of the set of concepts that influence the number of terror attacks experienced by different countries is rudimentary. Existing work on the incidence of terror focuses upon the structural characteristics of polities, economies, and societies, and fails to place competition between dissidents and states center stage. For example, Krieger and Meierrieks (2011) and Gassebner and Luechinger (2011) each scour the literature for studies that examine the cross-national incidence of terror attacks. Krieger and Meierrieks offer a verbal summary of the various findings, whereas

Gassebner and Luechinger identify all of the independent variables used in those studies, and include them in their own statistical meta-analysis which estimates regressions using every possible combination of variables from published studies. These two articles review effectively the same list of studies and thus consider the same lists of independent variables. As we describe below, we adopt a Bayesian measurement model approach to variable construction, and thus focus our attention on concepts. Details are available in Moore et al. (2011), and here we list the economic, political, and social structures included in our model, along with the expected sign. Only one of these variables is due to us: the coercive bureaucracy of the state (e.g., Gurr, 1988). It is interesting that the state's coercive capacity (i.e., the size of the bureaucratic apparatus responsible for coercion) has gone missing from this literature. Its absence seems to underscore the extent to which the existing literature fails to conceptualize oppositional terror as a tactic selected by dissident groups in response to the structural circumstances they face.

- Political Institutions
  Contestation (+)
  Participation (+)
  Veto players (+)
  Freedom of association, press, and speech (+)
  Coercive bureaucracy of the state (−)
- Structural Characteristics of Economy and Society
  Strength of civil society (−)
  Macro-economic performance (−)
  Ethno-linguistic composition of society (+)
  Physical quality of life (−)
  Population size (+)

As noted above, our project proposes a theoretical framework that places competition between dissident groups and government center stage. More specifically, we argue that the number of terror events one will observe in a given country-year will be a function of the behavior of states and dissidents as they compete for control of the policy agenda, and that the behavior of those actors will be influenced by the structural settings in which they are located. Details of that argument are fully developed in Moore et al. (2011). As we are interested here in whether the statistical model implied by the theory forecasts well out-of-sample, we list below the behavior we expected to have an impact on the annual number of terror attacks in a given country.

- Behavior

  Coercive state behavior (+)
  Nonviolent behavior of dissidents (−)
  Violent (non-terror) behavior of dissidents (+)
  Violent international conflict (+)

We measured most of the concepts above using estimated latent variables (e.g., Treier & Jackman, 2008).[4] That is, for each country-year in our data, we collected a variety of observable indicators across a wide range of latent concepts. Given the variance in both the number of observable indicators and the number of years available per country in our data, the Bayesian strategy allows us to estimate latent variables for country-years that other studies treat as missing.

Unlike the bulk of existing cross-national analyses of the incidence of terror, we study the sum of transnational and domestic terror attacks rather than only transnational ones. An absence of data has hampered the cross-national statistical study of domestic terror attacks: in their review of the large-N, cross-national literature Krieger and Meierrieks (2011, p. 19) could find only three studies that pool both domestic and transnational attacks, and two studies that examine only domestic terror attacks. An estimated 85% of terror attacks are domestic (LaFree & Dugan, 2007b), which means that what we know about the covariates of terror attacks is based on roughly 15% of the attacks that have taken place. The existence of the Global Terrorism Database (GTD) (LaFree & Dugan, 2007a) makes an analysis of the covariates of all terror attacks (i.e., both domestic and transnational) possible, and we use the total number of domestic and transnational terror attacks as the dependent variable in our study.

Turning to estimation of the model's parameters, it is conventional in cross-national statistical analyses of this type to assume a single baseline number of terror attacks for each country in the sample (i.e., estimate a single intercept for all countries). Of course, it is not necessary to impose this restrictive assumption, so we relaxed it. More specifically, we estimated a Poisson-lognormal mixture model with random effects for country-specific overdispersion (Cameron & Trivedi, 1998; Winkelmann, 2008). The Poisson-lognormal model allows us to estimate the expected number of counts per country-year as a multiplicative function of the covariates described above and a country-specific normally distributed random effect. We estimated the model's parameters via Monte Carlo Markov Chain (MCMC) in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000).[5] As noted above, details can be found in Moore et al. (2011).

An analysis of the deviance information criterion statistic for (1) a null model, (2) a model of only the structural covariates, (3) a model of only the behavioral variables, and (4) the full model demonstrated that each provided a better in-sample fit than the model(s) listed before it. That provided strong evidence in favor of our claim that existing models – which include *only* structural variables – are missing most of the action by failing to theoretically and empirically model the impact of the behavior of both dissidents and the state upon the number of terror attacks a country experiences.

The specific hypothesis tests provided further support for our theoretical approach. Three of the four behavior variables had the expected impact: violent (non-terror) dissent, government coercion, and international war each had a strong, positive association with the number of terror attacks. All five of the political institutions variables exhibited a substantial impact on the use of terror, though both Veto and Freedom of Association/Press/Religion/Speech had considerably smaller impacts than Contestation, Participation, and the Coercive Capacity of the state. Finally, we found that three of the five socioeconomic structure variables exhibited a systematic relationship with the number of terror attacks occurring within a given country-year: strength of civil society and macroeconomic performance were negatively related to the number of terror attacks, and population had a positive relationship. Having briefly described the model and results reported in Moore et al. (2011), we turn our attention to assessing the model's forecasting performance.

## DESIGNING THE FORECAST

Political scientists have yet to broadly embrace forecasting, and as a result much of what is done suffers weaknesses that are less prevalent in other fields (Brandt et al., 2011a). We wish to determine whether our model's estimates are capable of being put to use to predict the number of terror attacks a given country will experience in a given year. We take as given that we will know the values for our $X$ variables. This is frequently described as an out-of-sample forecast because the analyst first estimates the model parameters on a given sample (e.g., 1970–1996) and then plugs in the values of $X$ for a year outside of the given sample (e.g., 1997), and multiplies those values against the coefficient estimates to obtain forecast values of the dependent variable for that year. A different, and

considerably more challenging, forecasting task requires the analyst to also predict the values of $X$ (e.g., Hegre et al., 2013). We do not undertake such an effort here.

Some researchers use a model's fit statistics, such as the root mean square error (RMSE), when conducting an out-of-sample forecast analysis. Brandt et al. (2011a), among others, criticize this. It is particularly unhelpful in our case because our dependent variable is an event count, which is an integer with a lower bound of zero. Due to this an error-based statistic such as RMSE is unhelpful because residuals are asymmetric: due to the presence of the lower bound under-predictions will be systematically smaller than overpredictions. Thus, measures like RMSE are biased in such settings.

Researchers who assess the fit of binary models such as logit or probit are familiar with the type of approach that we adopt here. In those studies it is common to identify the percentage of correct predictions, and to further break down the analysis and examine the percentage of false negatives and false positives. This is sometimes called a classification analysis, and we employ it below.

We used data from 1970 to 1996 to estimate the model parameter posterior densities, and then multiplied the values for the $X_{i1997}$ vector by those parameters to produce forecasts. We saved 1,000 draws from our estimated posterior densities, thus permitting us to construct credible intervals as a measure of the uncertainty of the forecast. That is, for each of the 153 countries we produced 1,000 forecasts, rounded them to their nearest integer value, and record both the median as well as the values at the .25 and .975 percentiles, thus recording both the central tendency of the forecast and our uncertainty about that value.[6] That uncertainty is called the credible interval, and it is literally the range over which a given percentage (in our case, 95%) of the forecasts are distributed.[7]

We are interested in assessing the resulting 153 predictions in two ways: a binary prediction of whether there would be zero or more than zero attacks, and a count prediction which is the integer value. There are 153 countries in our dataset for 1997. Fig. 1 is a frequency distribution: 70 countries had zero attacks; and 83 had one or more, ranging from a single attack to over 350.

To begin we evaluate the accuracy of binary predictions. For those we recode the predicted value of the count such that any predicted value greater than 0 is reassigned the value of 1. To elaborate, we first multiply the $X_{i1997}$ vector by the posterior parameter densities reported in Moore et al. (2011). We then calculate, for each observation in the 1997 data, the

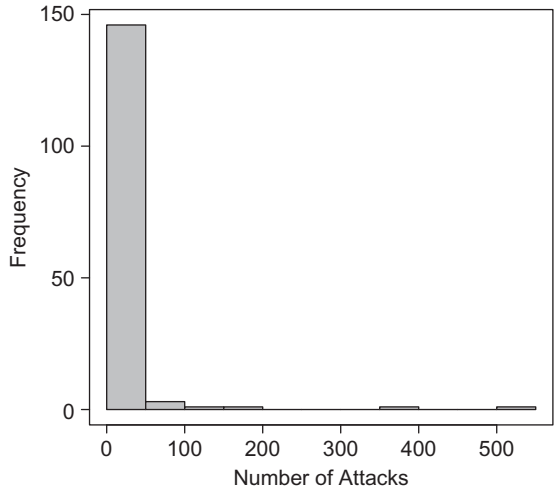


*Fig. 1.* Terror Attacks in 153 Countries, 1997.

probability that the country will experience at least one terror attack.[8] Using these probabilities we generate a binary prediction for each observation that is equal to one if the probability of experiencing at least one terror attack is >.5. We then report the following for our binary forecast: a classification analysis and a receiver operating characteristic (ROC) curve assessment.

We also make use of the predicted count densities that we produced for each observation: to offer an intuitive visual account, we calculate (for each country) probabilities for each count in the empirical distribution, and produce a marginal calibration plot that records both the observed frequencies for each count and the predicted frequencies for each count. Finally, we examine both the mean and the 95% credible intervals of the density for each prediction, and evaluate whether the observed value from 1997 falls within the credible interval, as well as whether it equals the (rounded) median.

## RESULTS

We begin with the integer count forecasts, which are considerably more demanding than the binary ones. To begin, we produce a marginal

calibration plot (Czado, Gneiting, & Held, 2009) to assess how well our count predictions map onto the observed distribution of terror attacks. The concept of marginal calibration is discussed in technical detail in Gneiting, Balabdaoui, and Raftery (2007), but the intuition behind the plot is straightforward: the frequency distribution of the predictions a count model produces should match fairly closely the observed frequency distribution. Fig. 1 presents a marginal calibration plot for our model's count predictions. The plot displays the observed frequencies for 1997 and the predicted frequencies for each count value (to ease presentation, we bin higher counts).[9] These predicted frequencies can then be compared to the actual number of cases that fall into each value/bin. Fig. 1 provides a rather positive assessment of the forecast, but as we see below, figures such as these can lead one to view a forecast more positively than one should.

Table 1 reports both the proportion of observations where the predicted count fell within the 95% credible intervals,[10] and the proportion of observations where the predicted count (rounded to the nearest integer) was equal to the observed count (precisely correct). The scores are 39% and 22%, respectively. This puts the forecast performance in a very different light. To be sure, a precisely correct integer count is a hyper-exacting standard, but note that the forecast does not improve terribly much when we relax it to include the 95% credible interval.

That said, this is a pooled cross-national time-series model of the number of terror attacks that would occur in each country in the world, in 1997. Though the model is not ready for use in a policy making setting, we are frankly surprised that the model performs as well as it does. Note, for example, that we do not include a lagged value of the dependent variable in our model. Further, a different approach that would undoubtedly produce superior out-of-sample forecasts, would estimate each country individually, or perhaps estimate small groups of neighboring countries. That is, the purpose of our effort is not to identify the best out-of-sample forecasts of the number of terror attacks countries observed in 1997. Our goals are instead to (1) for the purpose of theory building add an assessment of the out-of-sample forecasting performance to the work reported in Moore et al.

***Table 1.*** Count Prediction Performance for 1997.

| Metric | Proportion |
|---|---|
| Observed within CI | .39 |
| Precisely Correct | .22 |

(2011), and (2) offer those forecasts as a baseline against which other efforts might be judged. In other words, both theoretical efforts and efforts that simply wish to produce useful forecasts of country's expected number of terror attacks will be able to compare their efforts to ours, and will want to be sure that their models outperform this one.

Let us turn our attention to the assessment of the binary forecasts. Table 2 reports a classification matrix that compares these binary predictions to a binary version of the value observed in 1997 (coded 1 if the count is >0), with the row percentage reported below the frequency. Observations in the upper left cell correspond to accurate predictions of zero attacks, while those in the lower right cell correspond to accurate positive predictions. The off diagonals are then false positives (upper right cell) and false negatives (lower left cell). We turn to Table 3 which shows several additional measures of predictive performance for the binary predictions.

Accuracy is the proportion correctly classified; sensitivity is the proportion of ones correctly classified; and specificity the proportion of zeros correctly classified. Positive Predictive Value (PPV, sometimes called "precision") is the proportion of positive predictions that actually experienced at least one attack, and Negative Predictive Value (NPV) is the proportion of 0 predictions that did not experience any attacks. Overall, our model correctly predicts 78% of the country-years in 1997. Note further that it does a better job forecasting countries that do experience an attack than it does those that do not: we correctly forecast 86% of the countries

***Table 2.*** Binary Prediction Classifications for 1997.

| | Pred = 0 | Pred > 0 | Total |
|---|---|---|---|
| Obs = 0 | 25 | 17 | 42 |
| Obs > 0 | 16 | 95 | 111 |
| Total | 41 | 112 | 153 |

***Table 3.*** Binary Prediction Performance for 1997.

| Metric | Proportion |
|---|---|
| Accuracy (correctly classified) | .78 |
| Sensitivity (correct \| obs > 0) | .86 |
| Specificity (correct \| obs = 0) | .60 |
| Positive PV (correct \| pred > 0) | .85 |
| Negative PV (correct \| pred = 0) | .61 |

that experienced at least one terror attack in 1997, but only 60% of the countries that experienced no attacks. In any forecasting endeavor one needs to "choose one's poison," as it were, and decide whether to favor models that have greater sensitivity (i.e., produce relatively more false negatives) or specificity (i.e., those that produce relatively more false positives). Given politicians' interest in retaining office, in a decision support setting we suspect that for the purpose of forecasting terror attacks models with greater sensitivity (correctly predicted ones) would be preferred. Our 86% sensitivity score is certainly respectable (e.g., 80% is sometimes used as a minimum performance threshold). However, in light of the fact that, in the 1997 data, far more countries experience terror attacks (73%) than do not (27%), the model's ability to predict attacks is not as impressive. If one were to simply predict that all countries experience an attack this would produce an accurate prediction 73% of the time. We discuss these issues in detail below.

Fig. 2 reports a ROC curve based on the binary predictions. The ROC curve plots the probability of accurately predicting a 1 (i.e., sensitivity)

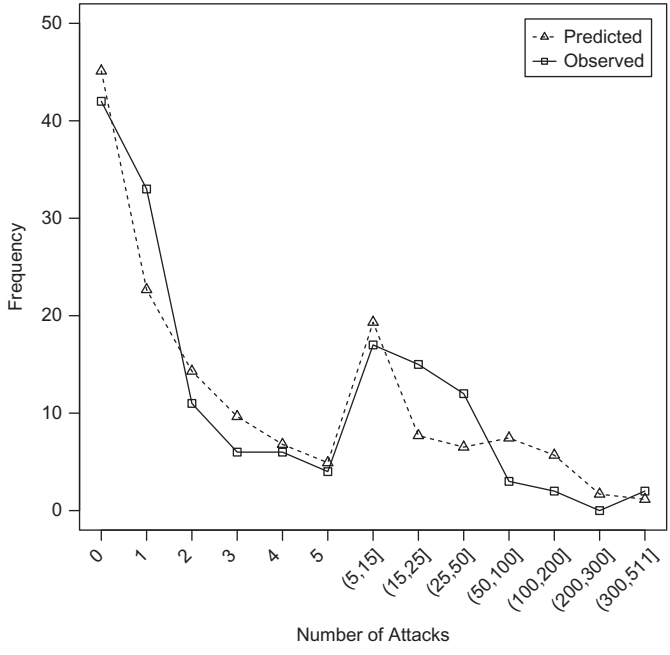


*Fig. 2.* Marginal Calibration Plot for Count Predictions.

against the probability of inaccurately predicting a 0 (the false positive rate, or 1 minus specificity) at different threshold values for a positive prediction. The benefit of the ROC curve is that it does not force one to (arbitrarily) choose a probability threshold for classifying positive predictions (which above, as in most studies, is .5). For a good model this threshold can be varied and the model loses very little predictive power (i.e., the curve stays toward the upper left corner of the graph, meaning the model does a good job of predicting both 0s and 1s). The area under the curve (AUC) is often used as a metric for comparing different models. For our model the AUC is .80, while the AUC for a random guess, represented by the dashed line, is .5 (Fig. 3). AU:2

To summarize, we have shown that while this particular model forecasts binary outcomes well, it nonetheless is not a very good candidate for use in an applied, policy making setting. While its binary sensitivity is respectable for such a setting, this is largely a function of the relatively large number of countries that experienced at least one terror attack during the year for which we produce a forecast. Further, its specificity in inadequate, and its accuracy is not very impressive. Yet the model does have value as a global baseline that can be useful not only to scholars and practitioners developing and testing theory about the incidence of terror attacks,

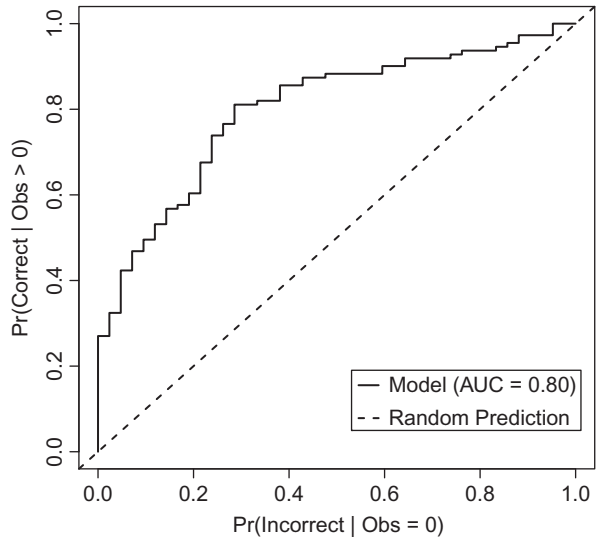


*Fig. 3.* ROC Plot for Binary Predictions.

but also forecasts of that incidence. To this point we have concentrated our attention on annual incidence per country, but we hasten to observe that the baseline value of our model extends beyond the country-year. That is, should one be studying or forecasting using a more fine-grained temporal period, such as the quarter, month, or week, one's model should be able to outperform ours. To the extent that it does not, the variables that we include serve as candidates that one should consider. Some readers might object that data availability may foreclose that possibility, and while that is certainly true, we must note that it would not for data that are only available at annual temporal aggregation: multi-level modeling makes it very attractive/feasible to collect/model data at multiple levels of temporal and/or spatial aggregation (e.g., Gelman & Hill, 2006).

## CONCLUSION

During 2012 there were two high-profile attacks on political science, and each made its case in large part by arguing that political scientists could not predict anything.[11] First, Arizona Congressman Jeff Flake introduced an amendment to the bill that funded the US National Science Foundation that would have eliminated the funding for the political science program.[12] The amendment eventually failed, but it kicked off a buzz well beyond professional political science associations, attracting attention not only in national US magazines and newspapers, but even earning in response an editorial in *Nature*.[13] Then in March 2013 Senator Thomas Coburn successfully attached an amendment to a continuing resolution bill that terminated financial support for the political science program through October of 2013.[14]

Second, in June of 2012 *The New York Times* published an Op-Ed written by a political science Professor at Northwestern University, Jacqueline Stevens.[15] Professor Stevens argues that though she believes that statistical analyses in political science have become almost *de rigueur* in the field, the King is parading naked: unlike in other fields where statistical modelers produce useful forecasts, statistical political science models are so poor that few even try, and those that do fail miserably.

Three things strike us about these episodes. First, they are critiques grounded in ignorance: neither Congressman Flake, who studied political science as an undergraduate at Brigham Young University, nor Professor Stevens has the training to properly evaluate the status of forecasting in

political science. Indeed, in their public statements each points to unexpected events such as the Arab Spring and asserts that political scientists failed to predict them. We do not bother to engage such silliness, other than to speculate that they likely rail against their local weather forecasts as well.

Second, and more importantly, both Flake and Stevens are quite correct to wonder why, given the common use of statistical models in political science, they do not see more forecasts. As we noted in the introduction to this chapter, many researchers do not use *any* goodness-of-fit measures to evaluate their statistical models, and (out-of-sample) forecasts of violent conflict are exceedingly rare, the recent surge notwithstanding. What we did not note there is that the subfield of violent conflict is arguably the vanguard of political science forecasters. Put plainly, political scientists have paid inadequate attention to forecasting for far too long. Indeed, if one scours the World Wide Web for blog posts by political scientists responding to the attacks by Flake and Stevens one will encounter many that argue that the arguments fail because political scientists should not be forecasting as it is not part and parcel to scientific inquiry! Research presentations at recent meetings of the Political Methodology Society, and the articles that follow such gatherings, belie this opinion, but it remains widely enough held to be a concern. Progress is being made, but we need to recognize that there is a long way to go.

Third, and most promisingly, this sort of backlash is the sort of thing one might anticipate as progress is made that changes the status quo. When a field of inquiry is in its youth it will grow in many different directions as it develops a proper identity. Though the unsuccessful Flake amendment of 2012 and the successful Coburn amendment of 2013 are in no small part to be understood through the lens of partisan budget battles, Stevens's broadside was little more than an attempt by a member of a coalition that is losing influence to diminish the influence of one that has recently displaced it.

Returning to the specifics of our study, to conclude their effort to forecast intranational violent crises of interest, Tikuisis et al. (2013, p. 1) write that "while model accuracy … usually exceeded 90 percent, the model did not generate sufficiently high and consistent precision … and [sensitivity] … for practical use." We reach the same conclusion regarding our model's ability to forecast: the precision is too low to recommend its use in a policy setting. Nevertheless, in addition to contributing to what we hope becomes the increasingly common use of out-of-sample forecasting among political scientists who use statistical models to test hypotheses, we contend

that our model can serve as a useful baseline for those who wish to forecast terror incidence both at the country-year level of aggregation and lower levels.

## NOTES

1. Ward et al. (2012) provide a considerably more detailed discussion.

2. We adopt Schelling's classic definition of terror as "violence intended to coerce the enemy rather than to weaken him militarily" (Schelling, 1960, p. 7).

3. See Ward et al. (2012) for a similar argument.

4. We were unable to identify enough variables to identify latent variable models for nonviolent and violent (non-terror) dissent, and thus create additive indices for both. These both come from Banks (2009). We use the number of antigovernment demonstrations and general strikes to measure nonviolent dissent, and the number of riots and acts of guerrilla warfare to measure violent, non-terror dissent. We also use single measures of ethno-linguistic composition of society and violent international conflict.

5. WinBUGS is available online at http://www.mrc-bsu.cam.ac.uk/bugs/.

6. We followed a similar procedure, described below, to generate the binary predictions.

7. A confidence interval is similar in that it is the range over which one expects a given percent of the distribution to lie. Put differently, a credible interval is the Bayesian paradigm's version of a confidence interval.

8. For a Poisson model the probability of any particular count $m$ can be calculated as: $[\exp(-\mu)\times\mu^m]/\text{m!}$, where $\mu$ is $\exp(x\beta)$. For a zero count this is $\exp(-\mu)$. The probability of a count larger than zero is then $1-\exp(-\mu)$.

9. We calculated these predictions by creating a predicted probability for each value of the terror count, from 0 to the maximum observed count, for each observation in the 1997 data, and then calculated the mean of each probability across the observations. The mean probability for each count serves as the predicted proportion of the sample that takes on that particular value. This proportion is multiplied by the sample size to create a predicted frequency.

10. A credible interval is effectively the Bayesian equivalent of a confidence interval in frequentist statistics.

11. Both occurred prior to Nate Silver and others' celebrated success forecasting the electoral college outcome in the 2012 US Presidential election.

12. H.AMDT.1058 (A025), http://thomas.loc.gov/cgi-bin/bdquery/D?d112:44:./temp/ bdeckr::.

13. Available online at http://www.nature.com/nature/journal/v487/n7407/full/487271a.html.

14. *The Economist* "Political Science Research: Singled Out" online at http://www.economist.com/blogs/democracyinamerica/2013/03/political-science-research.

15. Available online at http://www.nytimes.com/2012/06/24/opinion/sunday/political-scientists-are-lousy-forecasters.html.

## ACKNOWLEDGMENTS

The project upon which this chapter draws has received helpful feedback AU:3
from audiences at the Centre for the Study of Civil War Working Group at Peace Research Institute, Oslo; the Departments of Political Science at Essex University, Florida State University, and University of North Texas; and the annual meetings of the International Studies Association (2010, New Orleans, LA) and Peace Science Society (2008, Claremont, CA). We appreciate comments and feedback on the project from John Ahlquist, Dave Armstrong, Jason Barabas, Bethany Barrett, Bill Berry, Halvard Buhaug, David Cunningham, Kathleen Cunningham, Scott Edwards, Mike Findley, James Forest, Scott Gates, Jeff Gill, Havard Hegre, Brian Lai, Jim Piazza, Chris Reenock, Dave Siegel, Havard Strand, and Joe Young.

## REFERENCES

Andriole, S. J., & Hopple, G. W. (1984). The rise and fall of event data: From basic research to applied use in the US department of defense. *International Interactions*, *10*(3–4), 293–309.

Andriole, S. J., & Young, R. A. (1977). Toward the development of an integrated crisis warning system. *International Studies Quarterly*, 107–150. AU:4

Bagozzi, B. (2011). *Forecasting civil conflict with zero-inflated count models*. Working Paper.

Banks, A. S. (2009). *Cross-national time-series data archive*. Binghamton: Center for Social Analysis. Retrieved from http://www.databanks.sitehosting.net/

Blair, R., Blattman, C., & Hartman, A. (2012). *Predicting local-level violence*. Presented at the Northeast Political Methodology meeting, New York, NY.

Brandt, P. T., Freeman, J. R., & Schrodt, P. A. (2011a). *Racing horses: Constructing and evaluating forecasts in political science*. Annual meeting of the Political Methodology Society. Retrieved from http://eventdata.psu.edu/papers.dir/RHMethods20110720.pdf

Brandt, P. T., Freeman, J. R., & Schrodt, P. A. (2011b). Real time, time series forecasting of inter- and intra-state political conflict. *Conflict Management and Peace Science*, *28*(1), 41–64.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge: Cambridge University Press.

Clauset, A., Young, M., & Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, *51*(1), 58–87.

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254–1261.

Gassebner, M., & Luechinger, S. (2011). Lock, stock, and barrel: A comprehensive assessment of the determinants of terror. *Public Choice*, *149*(3), 235–261.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268.

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M., Marshall, M. G., Ulfelder, J., & Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, *54*(1), 190–208.

Gurr, T. R. (1988). War, revolution, and the growth of the coercive state. *Comparative Political Studies*, *21*(1), 45–65.

Gurr, T. R., & Lichbach, M. I. (1979). Forecasting domestic political conflict. In J. David Singer & M. D. Wallace (Eds.), *To augur well: Early warning indicators in world politics* (pp. 153–193). Beverly Hills, CA: Sage.

Gurr, T. R., & Lichbach, M. I. (1986). Forecasting internal conflict. *Comparative Political Studies*, *19*(1), 3–38.

Gurr, T. R., & Moore, W. H. (1997). Ethnopolitical rebellion: A cross-sectional analysis of the 1980s with risk assessment for the 1990s. *American Journal of Political Science*, *41*(4), 1079–1103.

Hegre, H., Karlsen, J., Nygård, H. M., Strand, H., & Urdal, H. (2013). Predicting armed conflict, 2010–2050. *International Studies Quarterly*. AU:5

Krieger, T., & Meierrieks, D. (2011). What causes terrorism? *Public Choice*, *147*, 3–27. Retrieved from http://dx.doi.org/10.1007/s11127-010-9601-1

LaFree, G., & Dugan, L. (2007a). *Global terrorism database, 1971–1999: Codebook and data documentation*. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research. ICPSR Study # 04586.

LaFree, G., & Dugan, L. (2007b). Introducing the global terrorism database. *Terrorism and Political Violence*, *19*(2), 181–204.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337.

Moore, W. H., Bakker, R., & Hill, D. W. (2011). *How much terror? Dissidents, governments, institutions and the cross-national study of terror attacks*. SSRN Working Paper. Retrieved from http://ssrn.com/abstract = 1977262

O'Brien, S. P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, *12*(1), 87–104.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.

Shellman, S. M., & Moore, W. H. (2008). The new school of political conflict and terror: From attributes to behavior and why policymakers will care. *Perspectives on Terrorism*, *2*(12), 3–10.

Tikuisis, P., Carment, D., & Samy, Y. (2013). Prediction of intrastate conflict using state structural factors and events data. *Journal of Conflict Resolution*.

Treier, S., & Jackman, S. (2008). Democracy as a latent variable. *American Journal of Political Science*, *52*(1), 201–217.

Ulfelder, J. (2012). Forecasting onsets of mass killing. Presented at the Northeast Political Methodology Meeting, New York, NY.

Ward, M. D., Greenhill, B. D., & Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, *47*(4), 363–375.

Ward, M. D., Metternich, N. W., Carrington, C., Doref, C., Gallop, M., Hollenbach, F. M., Schultz, A., & Weschle, S. (2012). *Stepping into the future: The next generation of crisis*

*forecasting models*. Working Paper. Retrieved from http://mdwardlab.com/biblio/learning-past-and-stepping-future-next-generation-crisis-prediction

Winkelmann, R. (2008). *Econometric analysis of count data*. Berlin: Springer Verlag.

Zammit-Mangiona, A., Dewar, M., Kadirkamanat, V., & Sanguinettia, G. (2012). Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Sciences*.

# AUTHOR QUERY FORM

| | | |
|---|---|---|
| Emerald | **Book: CCMPED-V022-3610844**<br><br>**Chapter: 3** | **Please e-mail or fax your responses and any corrections to:**<br><br>**E-mail:**<br>**Fax:** |

Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

**Disk use**

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

☐ Scanning the artwork

**Bibliography**
If discrepancies were noted between the literature list and the text references, the following may apply:

☐ **The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.**

☐ ***UNCITED REFERENCES:*** **This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.**

**Queries and/or remarks**

| Location in Article | Query / remark | Response |
|---|---|---|
| AU:1 | Please provide 5-6 keywords. | |
| AU:2 | Please check the citation of Fig. 3 in the text. | |
| AU:3 | Please check the text set under the heading "Acknowledgments." | |
| AU:4 | Please provide the volume no. in the ref. Andriole & Young (1977). | |

| | | |
|---|---|---|
| AU:5 | Please provide the volume no. and page range in the refs. Hegre et al. (2013); Tikuisis et al. (2013); Zammit-Mangiona et al. (2012). | |