DYNAMIC CONFLICT FORECAST - IMPROVING CONFLICT PREDICTIONS USING ENSEMBLE BAYESIAN MODEL AVERAGING

MICHAEL D. WARD, JACOB M. MONTGOMERY, AND FLORIAN M. HOLLENBACH

ABSTRACT. Much of the work in political science would greatly benefit if we would be able to make predictions about the future, this is especially true for the field of international relations. Yet, making sensible forecasts about the future is extremely hard. In this paper we propose the use of Ensemble Bayesian Model Averaging (EBMA) modified to be applicable to binary dependent variables. This process combines different statistical component models to increase the accuracy of out-of-sample forecasts. By using EBMA we combine the strengths of different component models to generate predictions with higher accuracy. After explaining our modified approach to EBMA we test the superiority of this process to the individual model out-of-sample forecasts on monthly data on insurgency in 29 Asian countries. We show that compared to the individual component models, EBMA increases the accuracy of the out-of-sample forecast on almost all metrics.

1. INTRODUCTION

Political scientists sometimes lament the invisibility of their research to policy makers. At the same time, the discipline has given insufficient attention to the task of making predictions about important future events that may be most useful to policy-makers on the ground. Rather, the focus has been on conducting *ex-post* analyses to develop and validate theories. However, the models so developed are only rarely applied to out-of-sample data, much less used to make informed guesses about the future. In part, this lack of emphasis on prediction results from the simple fact that it is very hard to be right when making forecasts about the future in many settings. The current wave of protests and uprisings in northern Africa and the Middle East is just the latest example of critical political events that are difficult to see in advance. Nonetheless, forecasts and out-of-sample predictions not only provide the ultimate test of statistical models

Date: March 9, 2011.

This project was undertaken in the framework of an initiative funded by the Information Processing Technology Office of the Defense Advanced Research Projects Agency aimed at producing models to provide an Integrated Crisis Early Warning Systems (ICEWS) for decision makers in the U.S. defense community. The holding grant is to the Lockheed Martin Corporation, Contract FA8650-07-C-7749. All the bad ideas and mistakes are our own.

and theories, they also give political scientists the possibility to increase the relevance of their research to policy makers.

As part of the Integrated Crisis Early Warning Systems (ICEWS) the task was to develop statistical models that would be able to most accurately predict the occurrence of certain political events, such as rebellion, insurgency or international political crises in 29 countries in the Pacific Rim. While several statistical models were very good at forecasting events, an improvement of the accuracy of forecasts seemed possible by combining the strength of each of the individual models. To do so, in this paper we implement the Ensemble Bayesian Model Averaging (EBMA) approach modified for binary outcome data, which is appropriate to the study of international conflict and discrete domestic crises.

The EBMA approach was first introduced by Raftery et al. (2005) as a postprocessing method for producing predictive probability distribution functions (PDF)s for out-of-sample events using multiple weather forecasting models. The original presentation was applied in cases where the PDFs were assumed to be normal, but the method was subsequently extended for applications in precipitation levels (Sloughter et al. 2007) and wind speeds (Sloughter, Gneiting and Raftery 2010). In all of these cases, the EBMA method has been shown to significantly improve out-of-sample predictive power. As a side-effect, the method also provides theoretically motivated way to evaluate the relative contributions of the component forecast models.

The basic intuition of EBMA is to make predictions by pooling across multiple forecast models and calibrating the weight assigned to each component based on a large number of past observations in some training period. Components may not necessarily share covariates, functional forms, or error structures. Indeed, the ensemble components need not even be statistical models. They may be systematic predictions from agent based models, stochastic simulations, or even subject expert predictions recorded over lengthy periods of time. The EBMA method post-processes predictions from each component forecast for both the training period and out-of-sample periods to generate ensemble predictions. The EBMA predictions are the weighted average of the predictions from each component model, where the weights can be interpreted as the posterior probabilities that the model reflects the "true" data generating process. The posterior weight assigned to each model reflects the models' performance during the training period. As we show below, the method provides superior predictive power relative to any single component model.

The rest of this paper will proceed as follows. The next section will provide a brief discussion of future predictions in political science in general and international conflict in particular. We then present the details of the EBMA method and our contribution in developing an approach to handle dichotomous dependent variables. Section 4 first describes the data and the individual statistical models used as input components in the EBMA process. We then show the improvement in the accuracy out-of-sample prediction produced by EBMA over the individual components for the occurrence of insurgencies in 29 Asian countries. We conclude with a brief discussion of future directions for this paper.

2. Dynamic forecasting in political science

While generally seen as the highest validity check of statistical models and theory (De Marchi, Gelpi and Grynaviski 2004; Beck, King and Zeng 2004), out-of-sample predictions are often neglected in political science. However, while many shy away from trying to make predictions about the future, there has been a long history of developing predictive models in specific contexts. Yet, in most cases "forecasts" are conceptualized as a conditional exercise in which values on a dependent variable are calculated based on some estimated model and then compared with the observed values, as illustrated by Hildebrand, Laing and Rosenthal (1976). In some cases, this becomes nothing more than an analysis of residuals, while in others the focus is on randomly selecting subsets of the data that are excluded during model development for testing model validity.

But there is also a tradition of attempting to make political predictions about things that have not yet occurred, in the sense that the *Old Farmer's Almanac*, published continuously since the late 18th Century, predicts the weather for the coming year (as well as fashion trends). An early proponent of using statistical models for making such predictions in the realm of international relations was Stephen Andriole, a research director at ARPA in the late 1970s (Andriole and Young 1977). In 1978, a volume edited by Nazli Choucri and Thomas Robinson provided an overview of the then current work in forecasting in international relations, much of which was done in the context of policy oriented research for the U.S. government during the Vietnam War.[1] Subsequently, there were a variety of efforts to create or evaluate forecasts including Freeman and Job (1979), Singer and Wallace (1979), as well as Vincent (1980). At this time, a few efforts began to generate forecasts of internal conflict (e.g., Gurr and Lichbach 1986).[2]

In the late 1990s, scholars of American electoral politics began making predictions of voting patterns in presidential elections (Campbell 1992). Predicting presidential and party vote shares for US elections has developed into a regular exercise in political science. For the past presidential election a symposium of forecasts was published in *PS: Political Science and Politics* with forecasts of presidential and congressional vote shares developed by Campbell (2008), Norpoth (2008), Lewis-Beck and Tien (2008), Abramowitz (2008), Erikson and Wlezien (2008), Holbrook (2008), Lockerbie (2008) and Cuzàn and Bundrick (2008). Furthermore, responses to the forecast and evaluations were published in a subsequent issue of the journal.

In addition, a few scholars have worked to make predictions (yes, about the future) in other contexts, including Krause (1997), Davies and Gurr (1998), Pevehouse and Goldstein (1999), Schrodt and Gerner (2000), King and Zeng (2001), O'Brien (2002), de Mesquita (2002), Fearon and Laitin (2003), De Marchi, Gelpi and Grynaviski (2004), Enders and Sandler (2005), Leblang and Satyanath (2006), Ward, Siverson and Cao (2007), Brandt, Colaresi and Freeman (2008), Bennett and Stam (2009), and Gleditsch and Ward (2010), among a few others. A

[1] According to Andriole's website `http://www.andriole.com/index.php/About.html`, "He designed and developed one of the first real-time computer-based systems for monitoring and forecasting international events and crises in the 1970s. This system incorporated quantitative indicators, production rules for inferring crisis likelihoods, and an interactive graphic interface. Version 1.0 of the system was developed on Digital Equipment Corporation (DEC) PDP 11/40s; subsequent versions were implemented on PDP 11/45s and 11/70s. A microcomputer-based version was developed on a Tektronix 4054; the first working prototype was fielded in 1976. Output from the system was included in President Reagan's daily briefing book."

[2] Doran (1999) and others have provided some criticism about forecasts in political science and international relations in particular.

summary of classified efforts was declassified and reported in Feder (2002). An overview of some of the historical efforts along with a description of current thinking about forecasting and decision-support is given by O'Brien (2010), a program manager at DARPA who has conceptualized and supported the ICEWS project under which research reported in this manuscript was conducted.

However, it is no exaggeration to state that most scholars avoided making predictions altogether, perhaps because their models had enough difficulty in describing accurately what *had* happened. The median empirical article in political science (as well as sociology and economics) use predictions only in the sense of in-sample observational studies and residual analysis.

3. Ensemble Bayesian Model Averaging

Uncertainty about the "correct" or "best" model in the social sciences can be high. This is particularly true when the object is to choose a model or method (statistical or otherwise) to predict future events. At a minimum, researchers typically are uncertain about the choice of appropriate co-variates as well as the appropriate statistical methods for handling non-independence in the error structure (e.g., spatial autocorrelation, temporal autocorrelation, etc.). As a result, researchers (or competing research teams) typically fit multiple models using various methodologies and predictor variables with the aim of honing in on a single choice for making predictions.

Rather than searching haphazardly through through the model space, some analysts may apply more formal approaches. Researchers may compare non-nested models using frequentist tests (e.g., the Cox and Vuong test) or select models based on fit statistics that penalize complexity such as the Akaike Information Criterion (AIC) or the Deviance Information Criterion (DIC).

There are numerous dangers to these approaches. First, it fails to incorporate our prior uncertainty about the true data-generating process in our predictions. Second, the approach tends to encourage analysts to search the latent model space for a model that maximizes the within-sample predictive power, which is well known to lead to over-fitting and poor out-of-sample performance (Hastie, Tibshirani and Friedman 2001).

In the end, however, all such techniques are aimed at recovering the "true" model representing the data generating process, which is unlikely to be fully encapsulated under any one predictive approach (statistical or otherwise). Rather, all analysts or statistical models are likely to bring their own substantive insight to the problem at hand, which might be fruitfully combined to generate more accurate predictions.

A promising approach to pooling across various forecasts while incorporating our uncertainty about the "best" model is EBMA, first proposed by Raftery et al. (2005). It is an extension of the Bayesian Model Averaging (BMA) methodology (c.f., Madigan and Raftery 1994; Draper 1995; Raftery 1995; Hoeting et al. 1999; Clyde 2003; Clyde and George 2004), which has been shown to have good performance in a variety of settings (Raftery and Zheng 2003). BMA itself was first introduced to political science by Bartels (1997), and has been applied in a number of contexts (e.g., Bartels and Zaller 2001; Gill 2004; Zaller 2004; Imai and King 2004; Geer and Lau 2006). Montgomery and Nyhan (2010) provide a more in-depth discussion of BMA and its applications in political science.

3.1. **Basic intuition.** The basic BMA approach to forecasting is as follows. Assume we have some quantity of interest in the future to forecast y^* based on previously collected training data y^T that was predicted using K statistical models $M_1, M_2, \dots, M_K$. Each model, M_k, is assumed to come from the prior probability distribution $M_k \sim \pi(M_k)$, the probability distribution function (PDF) for the training data is $p(y^T|M_k)$ and the outcome of interest is distributed $p(y^*|M_k)$. Applying Bayes Rule, we get that

$$p(M_k|y^T) = \frac{p(y^T|M_k)\pi(M_k)}{\sum\limits_{k=1}^{K} p(y^T|M_k)\pi(M_k)}. \tag{1}$$

and the marginal predictive PDF for y^* is

$$p(y^*) = \sum_{k=1}^{K} p(y^*|M_k)p(M_k|y^T). \tag{2}$$

The BMA PDF (2) can be viewed as the weighted average of the component PDFs where the weights are determined by each model's performance within the training data. Likewise, we can make a deterministic estimate from this PDF as the weighted predictions of the components, denoted

$$E(y^*) = \sum_{k=1}^{K} E(y^*|M_k)p(M_k|y^T). \tag{3}$$

3.2. **EBMA for dynamic settings.** We now turn to applying this basic BMA technology to making predictions in a dynamic setting. In generating predictions of significant domestic crises or international disputes, the general task is to first build a statistical model for some set of countries S in time periods T, which we refer to as the training period. Using the same statistical model (or general technique in the case of subject-expert predictions), we then generate a forecast, f_k, for observations in future time periods T^*.[3]

However, no particular model or alternative forecasting method is likely to fully encapsulate the "true" data generating process. Rather, various research teams or statistical techniques will reflect different facets of reality. EBMA aims to collect *all* of the insight from multiple forecasting models in a coherent manner. The aim is not to choose some "best" model, but rather incorporate the insights and knowledge implicit in various forecast efforts via statistical post-processing.

[3] Sloughter et al. (2007) make predictions for only one future time period, and use only a subset of past time-periods (they recommend 30) in their training data. Thus, predictions are made sequentially with the entire EBMA procedure being re-calculated for each future event as observations are moved from the out-of-sample period T^* into the training set T. Due to the more serious data restrictions in social science data, we choose to simply divide *all* the data into discrete training- and test-periods for the entire procedure.

Let us assume that we have K forecasting models predicting outcome events y. Each component forecast, f_k, is associated with a component PDF, $g_k(y|f_k)$, which may be the original predictive PDF from the forecast model or a bias corrected forecast. These components are the conditional PDFs of some outcome y given the kth forecast, f_k, conditional on it being the "best" forecast in the ensemble. For example, the posterior PDF of an outcome y_{st} for some country $s \in S$ in period $t \in T$ given the forecast f_{kst} from model k is $g_k(y_{st}|f_{kst})$. This assumes that $P(M_k|y_{ST}) \equiv w_k = 1$, or that the posterior odds of model k are unity.

The EBMA PDF is then a finite mixture of the K component forecasts, denoted

$$p(y|f_1, \ldots, f_K) = \sum_{k=1}^{K} w_k g_k(y|f_k), \tag{4}$$

where the weight w_k is based on forecast k's relative predictive performance in the training period T. The w_k's $\in [0, 1]$ are probabilities and $\sum_{k=1}^{K} w_k = 1$. The specific PDF of n event y_{st^*} in country $s \in S$ at time $t^* \in T^*$ will then be

$$p(y_{st^*}|f_{1st^*}, \ldots, f_{Kst^*}) = \sum_{k=1}^{K} w_k g_k(y_{st^*}|f_{kst^*}). \tag{5}$$

3.3. **The dichotomous outcome model.** Past work on EBMA does not apply directly to the prediction of conflict or domestic crisis as the assumed PDFs in existing papers are either normal, poisson, or gamma. In most datasets of international crises or conflicts – including the ICEWS data explored below – the data is not sufficiently fine-grained to justify these distributional assumptions. Usually, the outcomes of interest are dichotomous indicator for whether an event (e.g., civil war) has occurred in a given time period in a specified country or region. Thus, none of the distributional assumptions used in past work are appropriate in this context, and it is necessary to adjust EBMA for predicting crisis events. Fortunately, it is a straight forward extension of Sloughter et al. (2007) and Sloughter, Gneiting and Raftery (2010) to deal appropriately with binary outcomes.

We follow Sloughter et al. (2007) and Hamill, Whitaker and Wei (2004) in using logistic regression after a power transformation of the forecast to reduce prediction bias. Let $f_k^{adj} \equiv f_k$ be the forecast from model k. For notational ease, we assume that f_k is the forecast after the adustment for bias reduction. Therefore, let $f_k' \in [0, 1]$ be the forecast on the predicted probability scale and

$$
\begin{aligned}
f_k = \; & \left[(1 + \text{logit}(f_k'))^{1/b} - 1\right] I\left[f_k' > \tfrac{1}{2}\right] \\
& - \left[(1 + \text{logit}(|f_k'|))^{1/b} - 1\right] I\left[f_k' < \tfrac{1}{2}\right],
\end{aligned}
\tag{6}
$$

where $I[.]$ is the general indicator function.

Hamill, Whitaker and Wei (2004) recommend setting $b = 4$, while Sloughter et al. (2007) use $b = 3$. We found that $b = 4$ works best in the example below, but other analysts may try alternative specification. The purpose of this transformation is to “dampen” the effect of extreme observations in the conditional PDF $g_k(y|f_k)$ and reduce over-fitting.

The logistic model for the outcome variables is

$$
\begin{aligned}
\text{logit } P(y = 1|f_k) \; & \equiv \log \tfrac{P(y=1|f_k)}{P(y=0|f_k)} \\
& = a_{k0} + a_{k1} f_k.
\end{aligned}
\tag{7}
$$

and

$$
\text{logit } P(y = 0|f_k) \equiv \log \frac{P(y = 0|f_k)}{P(y = 1|f_k)}
\tag{8}
$$

Putting (7) and (8) together, the conditional PDF of some within-sample event, given the forecast f_{kst} and the assumption that k is the true model, can be written:

$$g_k(y_{st}|f_{kst}) = \begin{array}{l} P(y_{st} = 1|f_{kst})I[y_{st} = 1] \\ +P(y_{st} = 0|f_{kst})I[y_{st} = 0], \end{array} \tag{9}$$

where $I[]$ is the general indicator function. Applying this to (4), the PDF of the final EBMA model for y_{st} is

$$p(y_{st}|f_{1st}, f_{2st}, \ldots, f_{Kst}) = \begin{array}{l} \sum\limits_{k=1}^{K} w_k[P(y_{st} = 1|f_{kst})I[y_{st} = 1] \\ +P(y_{st} = 0|f_{kst})I[y_{st} = 0]]. \end{array} \tag{10}$$

3.4. **Parameter estimation by maximum likelihood and EM algorithm.** Parameter estimation is conducted using only the data from the training period T. The parameters a_{0k} and a_{1k} are specific to each individual component model in the ensemble and require no data from additional components. For model k, these parameters can be estimated using the traditional logistic regression where y is the dependent variable and the covariate list includes only f_k. We emphasize that these parameters should *only* be estimated using forecasts generated from observations contained in the training data.

The difficulty is in estimating the weighting parameters, $w_k \ \forall k \in [1, 2, \ldots, K]$. One approach would be place priors on all parameters, and conduct a fully Bayesian analysis with Markov Chain Monte Carlo techniques. We hope to implement this strategy in future papers. For the moment, however, we follow Raftery et al. (2005) and Sloughter et al. (2007) in using maximum likelihood techniques to estimate model weights.

With the standard assumptions of independence in forecast errors across countries and time-periods, the log-likelihood function for the full EBMA model (10) can be written

$$\ell(w_1, \ldots, w_K|a_{01}, \ldots, a_{0K}; a_{11}, \ldots, a_{1K}) = \sum_{s,t} \log p(y_{st}|f_{1st}, \ldots, f_{Kst}). \tag{11}$$

where the summation is over values of s and t that index all observations in the training time period, and $p(y_{st}|f_{1st},\ldots,f_{Kst})$ is given by (10). Unfortunately, the log-likelihood function cannot be maximized analytically, but Sloughter et al. (2007) suggest using the expectation-maximization (EM) algorithm.

We introduce the unobserved quantities z_{kst}, which represent the posterior probability for model k for observation y_{st}. An alternative interpretation is that z_{kst} represents the probability that model k is the best model for predicting observation y_{st}. The E step for the full EBMA model in (10) involves calculating estimates for these unobserved quantities using the formula

$$(12) \qquad \hat{z}_{kst}^{(j+1)} = \frac{w_k^{(j)} p^{(j)}(y_{st}|f_{kst})}{\sum\limits_{k=1}^{K} w_k^{(j)} p^{(j)}(y_{st}|f_{kst})},$$

where the superscript j refers to the jth iteration of the EM algorithm. It follows that $w_k^{(j)}$ is the estimate of w_k in the jth iteration and $p^{(j)}(.)$ is shown in (10). Assuming these estimates of z_{kst} are correct, it is then straight forward to derive the maximizing value for the model weights. Thus, the M step estimates these w_k using the current estimates of z_{kst} (in this case $\hat{z}_{kst}^{(j+1)}$) as

$$(13) \qquad w_k^{(j+1)} = \frac{1}{n}\sum_{s,t} \hat{z}_{kst}^{(j+1)}.$$

where n represents the number of observations in the training dataset.

The E and M steps are iterated until the improvement in the log-likelihood is no larger than some pre-defined tolerance. Although the log-likelihood will increase after each iteration of the algorithm, convergence is only guaranteed to a local maximum of the likelihood function. Convergence to the global maximum is not assured, and the model is therefore sensitive to initial conditions.[4] We begin with the assumption that all models are equally important, $w_k = \frac{1}{K} \forall k \in [1,\ldots,K]$.

[4] Future versions of this paper will explore these issues more fully.

3.5. **Ensemble prediction.** With these parameter estimates completed, it is now possible to generate ensemble forecasts. If our forecasts, f_k, are generated from a statistical model, we now generate a new set of predictions f_{kst^*} from the previously fitted models. For convenience, let $\hat{\mathbf{a}}_k \equiv (\hat{a}_{k0}, \hat{a}_{k1})$. For some observation in country $s \in S$ in the out-of-sample period $t^* \in T^*$, we can see that

$$\begin{aligned}P(y_{st^*} = 1 | f_{1st^*}, \ldots, f_{Kst^*}; \hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_K; \hat{w}_1, \ldots, \hat{w}_K) = \\ \sum_{k=1}^{K} \hat{w}_k \text{logit}^{-1} \left(\hat{a}_{k0} + \hat{a}_{k1} f_{kst^*}\right).\end{aligned} \tag{14}$$

4. Application to conflict forecasting

The basic task of the ICEWS project is to produce predictions for five dependent variables, for 29 countries, for every month from 1997 through the present plus three months. The variables in question are rebellion, insurgency, ethnic violence, domestic political crises, and international political crises. The twenty-nine countries are Australia, Bangladesh, Bhutan, Cambodia, China, Comoros, Fiji, India, Indonesia, Japan, Laos, Madagascar, Malaysia, Mauritius, Mongolia, Myanmar, Nepal, New Zealand, North Korea, Papua New Guinea, Philippines, Russia, Singapore, Solomon Islands, South Korea, Sri Lanka, Taiwan, Thailand, and Vietnam. This set not a random sample, but rather constitutes the countries of population greater than $500,000$ that are in the Area of Responsibility of the US Pacific Command. The countries in PACOM include about 50% of the total world population, along with five of the largest military powers (China, Russia, India, North & South Korea). The countries range from democratic to authoritarian, from tiny to enormous, from landlocked to archipelago, and vary widely on almost any social or economic indicator you might imagine. One of the main missions of the U.S. Pacific Command is to enhance stability in the Asia-Pacific region by promoting security, cooperation, encouraging peaceful development. As such it is interested in the ebb and flow of events in this region.

The bulk of our data are gleaned from natural language processing of a continuously updated harvest of news stories, primarily taken from Lexus/Nexus archives. These are processed with a version of the TABARI processor for events developed by Philip Schrodt and colleagues in the context of the event data project he has been developing over the past two decades (see `http://eventdata.psu.edu/`). These data are described elsewhere, but essentially each month we receive a drop of two sets of data. The first of these comprises the five dependent variables in this study, known (in a moment of naming hubris) as the ground truth data. In addition, we also receive all of the event data for each event that transpired within or involving each of the 29 countries in the sample. These event data are driven through and expanded version of the CAMEO coding framework, so that they can be recoded and collated to create a variety of more specific indicators.

These data are augmented with a variety of other attribute and network data. In particular we use attributes, coded on a monthly or yearly basis from the Polity, MAR, and World Bank data set. We also include information about the election cycles (if any) in each of the countries. In addition, we use information about relations among the 29 countries, including geography, the length of shared borders, the amount of trade, the movement of people across borders, the number of refugees, as well as the number and types of events between each pair of the 29 countries (plus the US).

In the remainder of this section we will apply the EBMA approach introduced above to make out-of-sample predictions for the occurrence of insurgency in 29 countries in asia and the Pacific. All individual statistical models are built on data from the ICEWS project collected every month from 1997 through the present. The in-sample predicted probabilities from the individual components are used to construct the finale EBMA estimates. We can then compare the performance of the EBMA model to the components on the in-sample as well as the out-of-sample segments of the dataset. For the results presented here the in-sample period ranges from January 1999 to December 2008, while the out-of-sample period is January 2009 to December 2010.

4.1. **Statistical Models.** We begin by briefly providing short descriptions of the individual component models. To provide considerable variation in the complexity as well as accuracy of predictions of the different statistical models, we included several hierarchical as well as some simple generalized linear models.

4.1.1. *Model LMER Politics.* The Lmer Politics 1 model is a hierarchical model which includes several political variables as predictors for insurgency. In addition to the general intercept, it includes a country specific intercept as the only random effects term. As fixed effect terms this model includes the *executive constrain* variable from the Polity IV dataset (Marshall, Jaggers and Gurr 2009), *a count variable of the number of cooperative events* between the country in question and the USA, and *a count of violent events directed against the government.*

4.1.2. *Model GLM 1.* This model is a simple generalized linear model for binary dependent variables. As independent variables we included a *three months lag of population size*, *gdp growth* (also lagged three months), *the number of minority groups at risk*, and again a count of cooperative events with the United States.

4.1.3. *Model GLM 2.* This second GLM model is the simplest model included in the data analysis. It only includes two predictors and is a simpler version of model GLM 1. The GLM 2 model only includes population size and gdp growth as independent variables, both are lagged three months.

4.1.4. *Model LMER Politics 2.* The second generalized linear mixed effects model includes five fixed effect terms as well as random effects intercept for each country. As fixed effects variables we include the executive constrain variable, as well as the *competitiveness of participation* variable from the Polity IV dataset (Marshall, Jaggers and Gurr 2009). In addition, we include the size of population and *proximity to election* as fixed effects terms. The proximity to election variable is a count of the number of days to the next or from the last election, whichever is closer. The last fixed effects term is *a spatial lag that reflects recent occurrences of insurgencies in the countries' geographic neighbors*.[5]

4.1.5. *Model SAE.* Model SAE is one of the models developed to produce forecasts of conflictual events as part of the ICEWS project and was designed by the SAE team. The model is a generalized linear model for binary dependent variables. The SAE team uses a variety of 27 different independent variables in their predictive model, the variables are all taken from basic event stream data provided through the ICEWS project.

4.1.6. *Duke Model.* The Duke model is another statistical model designed to forecast events as part of the ICEWS project. It is a hierarchical model with several fixed effects and random effects terms. A general intercept that represents the global mean probability of insurgency is included. Additionally as fixed effects terms *a count of the number of new onsets of insurgency that took place in each country in the previous two years*, a measure of the time elapsed since the last election, *a count of the number of domestic conflictual events that involved the country's military*, as well as *a count of the number of conflictual events involving the United States* are included. Furthermore, a *spatial lag that reflects the average number of recent domestic political crises that occurred among the country's geographic neighbors* is included as another fixed effects term. In addition to the fixed effect terms, the model contains a country-specific random effect for *gdp per capita*.

As a stand alone statistical model, the Duke model is consitantly the most accurate models in the ICEWS project. It has produced very good results and had a high accuracy in the out-of-sample predictions. As is shown below, it is hard to improve the performance of the ICEWS model. Thus, the EBMA process is done with and with out the Duke model included.

4.2. **Results.** Table 1 shows the results of the individual models, as well as the EBMA results generated from the in-sample component forecasts. The first column shows the weights that

[5] Geographical proximity is measured in terms of the length of the shared border between the two countries.

the EBMA model associates with each component. As can be seen, the two simple GLM models are effectively excluded, while the Lmer Politics 1 component carries the greatest weight, followed closely by the SAE model.

The constant term associated with each component corresponds to the term a_{k0} in Equation 7, while predictor term corresponds to a_{k1} in Equation 7. AUC is the area under the Receiver-Operating Characteristic (ROC) curve. The advantage of using ROC curves in evaluating forecasts is that it produces an evaluation of the correctly predicted events at each possible cutoff point for making a positive prediction. A value of 1 of the AUC score would mean that all observations were predicted correctly at all possible cut off points (King and Zeng 2001).

As one can see in Table 1, three of the individual models that are used in the EBMA process all have relatively high AUC scores, especially Lmer Politics 1. Yet, even though their predictions were very good individually, the EBMA model's accuracy is higher than any of the individual component's. Similarly, the EBMA model has the highest proportional reduction error (PRE) together with the SAE component. The PRE is the percentage increase of correctly predicted observations relative to the base model. The base in our case is a model of predicting no insurgencies at all. This may seem too primitive to some, however one has to consider that insurgencies are rare events, thus predicting "no insurgencies" for all observations leads to 89% correct predictions.

Another indication for the improvement of forecasts through the EBMA process is the Brier score, which is the average squared deviation of the prediction from the true event, thus a lower score corresponds to higher accuracy of forecasts (Brier 1950). With a score of 0.04 the EBMA model and the SAE component have the lowest Brier score in the in-sample predictions. Similarly the SAE and EBMA model have the highest percentages of correct predictions (the cutoff point for positive predictions is 0.5 here).

Figure 1 shows the separation plots for the EBMA model as well as all the individual statistical components. The plots can be interpreted as follows. In each plot, the oder of observations is not determined by time. Rather, observations are ordered from left to right with increasing

TABLE 1. model statistics – in-sample predictions

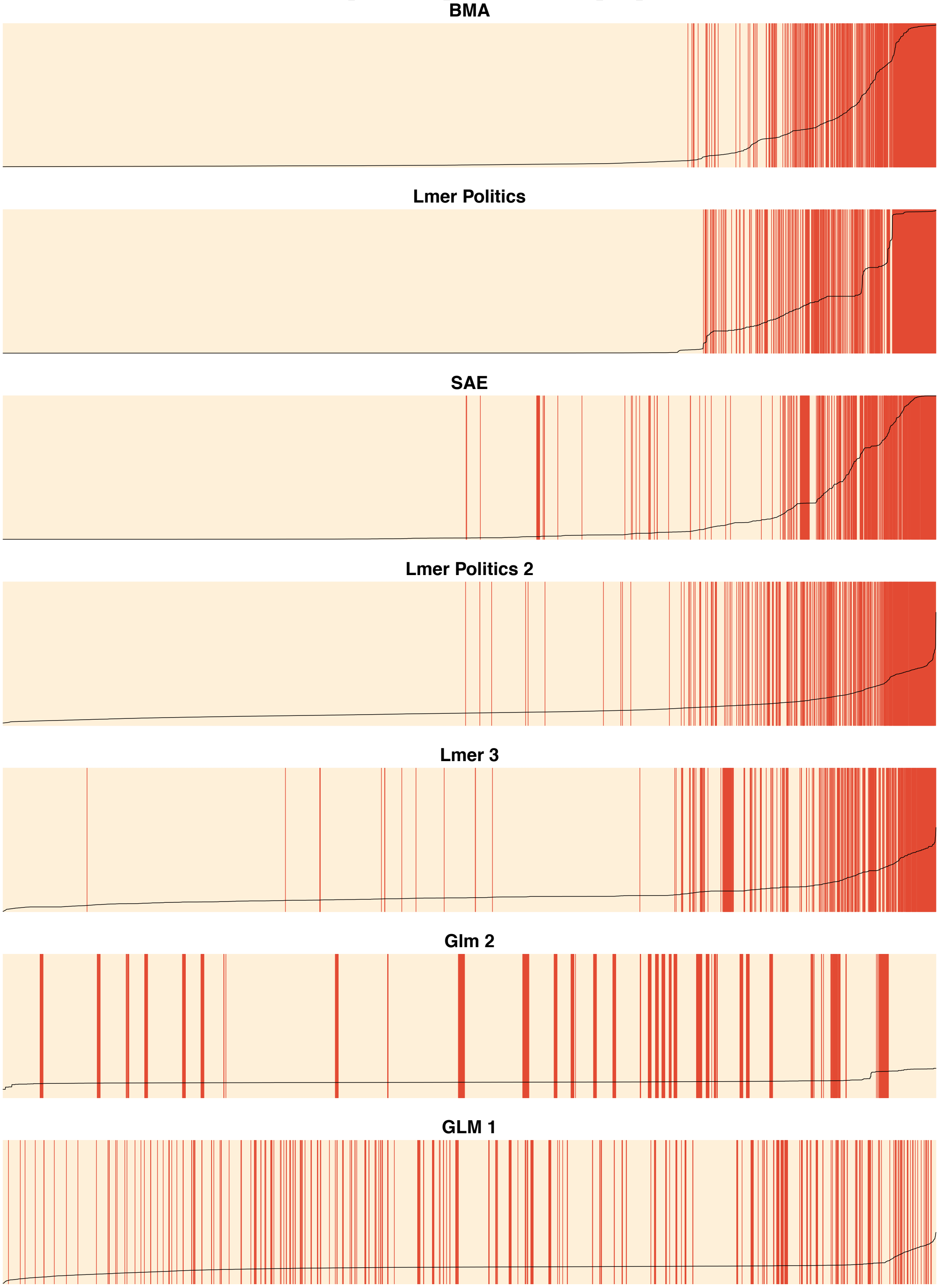
	Weight	Constant	Predictor	AUC	PRE	Brier	% Correct
Lmer Politics	0.48	0.19	7.88	0.97	0.41	0.04	93.28
Glm 1	0.00	0.61	8.32	0.61	0.00	0.10	88.65
Lmer Politics 2	0.15	6.08	28.25	0.96	0.01	0.07	88.79
Glm 2	0.00	0.57	8.16	0.65	0.00	0.10	88.65
Lmer 3	0.01	4.56	23.41	0.93	0.01	0.07	88.74
SAE	0.36	0.04	7.46	0.96	0.48	0.04	94.11
BMA				0.98	0.48	0.04	94.11

predicted probabilities of an occurrence of insurgency by this particular model. The black line corresponds to the predicted probability produced by the model for each observation and is thus increasing from left to right. Vertical red lines are actual occurrences of insurgencies. Thus the separation plots are an easy way to visually evaluate the predictions. The plots are ordered with the EBMA model plot first, and then by decreasing AUC scores from top to bottom.

Corresponding with the results discussed above, it can be seen that the two simple GLM components perform very poorly, whereas of the individual components the SAE model seems to perform the best. The Lmer Politics 1 model has produced almost no false positives, but more false negatives in comparison to the SAE one. Not surprisingly, the overall best performance with very few false positives is associated with the EBMA process. The separation plots also show that the EBMA model produces even less false negatives than the individual SAE component. Further an indication for the improvement of predictions by the EBMA process is that all actual occurrences of insurgency are at the far right of the plot, which corresponds to the increasing predicted probabilities of the model.

The more interesting evaluation of the method, however, is seen in the out-of-sample predictions. Table 2 shows results of using the individual components as well as the EBMA model fit above to produce future predictions. Since the EBMA model was created from the in-sample predictions shown above, the weights associated with each of the individual models, as well as the constant (a_{k0}) and predictor (a_{k1}) terms are the same as for in-sample statistics (see Table 1) and are thus not reported again.

FIGURE 1. separation plots – in-sample prediction



BMA

Lmer Politics

SAE

Lmer Politics 2

Lmer 3

Glm 2

GLM 1

Support for our contention that the EBMA method provides superior predictive power is much stronger in the context of the out-of-sample predictions. While the EBMA model has a marginally smaller area under the ROC curve than the Lmer Politics 2 and Lmer 3 model, it substantially outperforms all other models on any of the other evaluation parameters. The EBMA model has the higher PRE score with 0.11, the lowest Brier score with 0.05 and the highest correct prediction percentage. Again, it has to be emphasized that any improvement of predictions is extremely difficult to achieve, given this dataset, where insurgencies are such a rare event. In the out-of-sample data, only 74 out 696 cases have actual insurgencies. Thus even the base model, which predicts no insurgencies at all, is correct in predicting 89.22% of cases. A 0.11 proportional reduction in error to this base model is quite large.

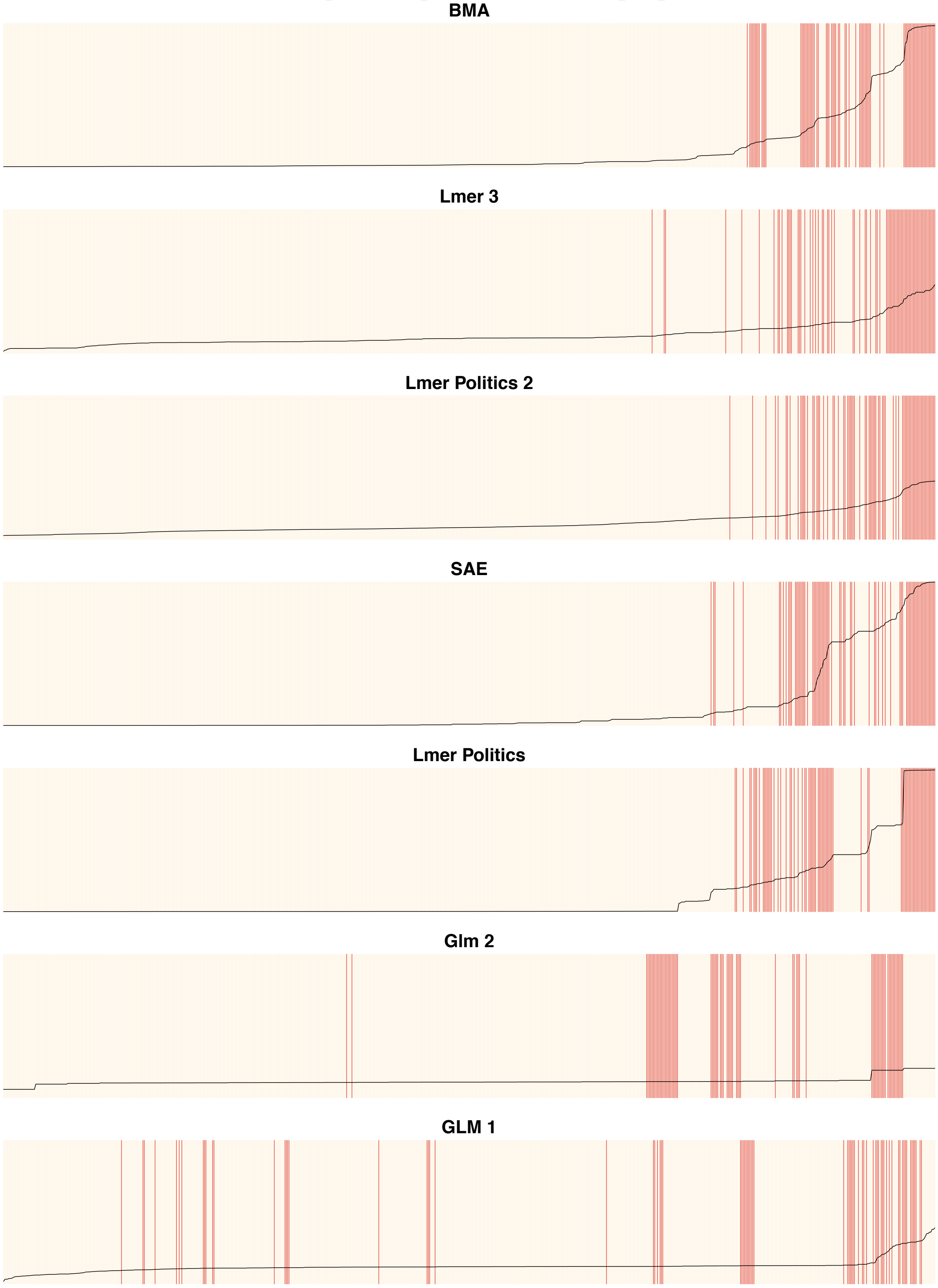
TABLE 2. model statistics – out-of-sample predictions

	AUC	PRE	Brier	% Correct
Lmer Politics	0.94	0.05	0.06	89.94
Glm 1	0.72	0.00	0.09	89.37
Lmer Politics 2	0.97	0.00	0.07	89.37
Glm 2	0.84	0.00	0.09	89.37
Lmer 3	0.97	0.00	0.07	89.37
SAE	0.96	0.04	0.06	89.80
BMA	0.96	0.11	0.05	90.52

Figure 2 shows the separation plots for the components as well as the EBMA model's predictive performance on the out-of-sample data. Not surprisingly, the performance of all models is worse than on the in-sample data. Of the individual components, the Lmer Politics 1 and the SAE model again have the best performance. It can be seen in the separation plots that while the AUC score for the Lmer 3 and Lmer Politics 2 are quite high, they never reach high levels of predicted probabilities for any of the observations.

The EBMA model performs better than any of the individual components, with very high predicted probabilities for the majority of actual events. It produces fewer false positives than the SAE component and less false negatives and false positives than the Lmer Politics 1 component.

FIGURE 2. separation plots – out-of-sample prediction



BMA

Lmer 3

Lmer Politics 2

SAE

Lmer Politics

Glm 2

GLM 1

Taking all the evaluation statistics together, as well as the visual evidence, we can conclude that the EBMA model leads to a substantial improvement in out-of-sample forecasts relative to its components, even in datasets with rare events and when individual models are already performing very well. The EBMA model consistently has the best score or is marginally close to the best score of the best component models.

In addition to the five individual components used for the EBMA process above, we also re-ran EBMA with the Duke model included. Recall that the Duke model was excluded from the analysis above because its individual performance is already quite exceptional, and thus it is very influential once it is included in the EBMA process.

Table 3 shows the individual model statistics as well as the statistics of the EBMA process. It can be seen in Table 3 that once we include the Duke model, the EBMA process leads to a ensemble model where the weights are distributed quite unevenly. In particular, the Duke component is assigned a weight of 0.89. Thus, the EBMA model is largely using only the predictions from the Duke component. The Lmer Politics 1 and the SAE model both marginally influence the EBMA process, each being weighted with 0.05, while Lmer Politics 2 is weighted with 0.01.

TABLE 3. model statistics: in-sample predictions with the Duke model

	Weight	Constant	Predictor	AUC	PRE	Brier	% Correct
Lmer Politics	0.05	0.19	7.88	0.97	0.41	0.04	93.28
Glm 1	0.00	0.61	8.32	0.61	0.00	0.10	88.65
Lmer Politics 2	0.01	6.08	28.25	0.96	0.01	0.07	88.79
Glm 2	0.00	0.57	8.16	0.65	0.00	0.10	88.65
Lmer 3	0.00	4.56	23.41	0.93	0.01	0.07	88.74
Duke	0.89	0.12	8.01	0.99	0.71	0.03	96.70
SAE	0.05	0.04	7.46	0.96	0.48	0.04	94.11
BMA				0.99	0.71	0.02	96.72

The statistics associated with the Duke component show why the EBMA process puts so much weight on this model. The area under the ROC curve for the Duke model is .99, while the PRE score is 0.71 and the Brier score is 0.03. Given a cut-off point for positive predictions of 0.5, the Duke model produces 96.70 correct predictions on the in-sample data. It is thus almost impossible to improve the accuracy relative to the Duke model, especially on this dataset with very few actual conflictual events. Not surprisingly the AUC and PRE score of the EBMA model are the same as those of the Duke model. Only on the Brier score and the correct prediction percentage the EBMA model actually improves in comparison to the Duke model.

Figure 3 shows the separation plots for the EBMA model and its components for the predictions on the in-sample data. Again the models are ordered with the largest AUC first and the smallest AUC last. It is easy to see that the EBMA model is very similar to the Duke component, which by far has the highest weight. Both the EBMA and Duke model make very accurate predictions, whereas the Lmer Politics 1 and SAE components have more false negatives than the other two models. Again, not surprisingly, the two simple GLM components are performing the worst.

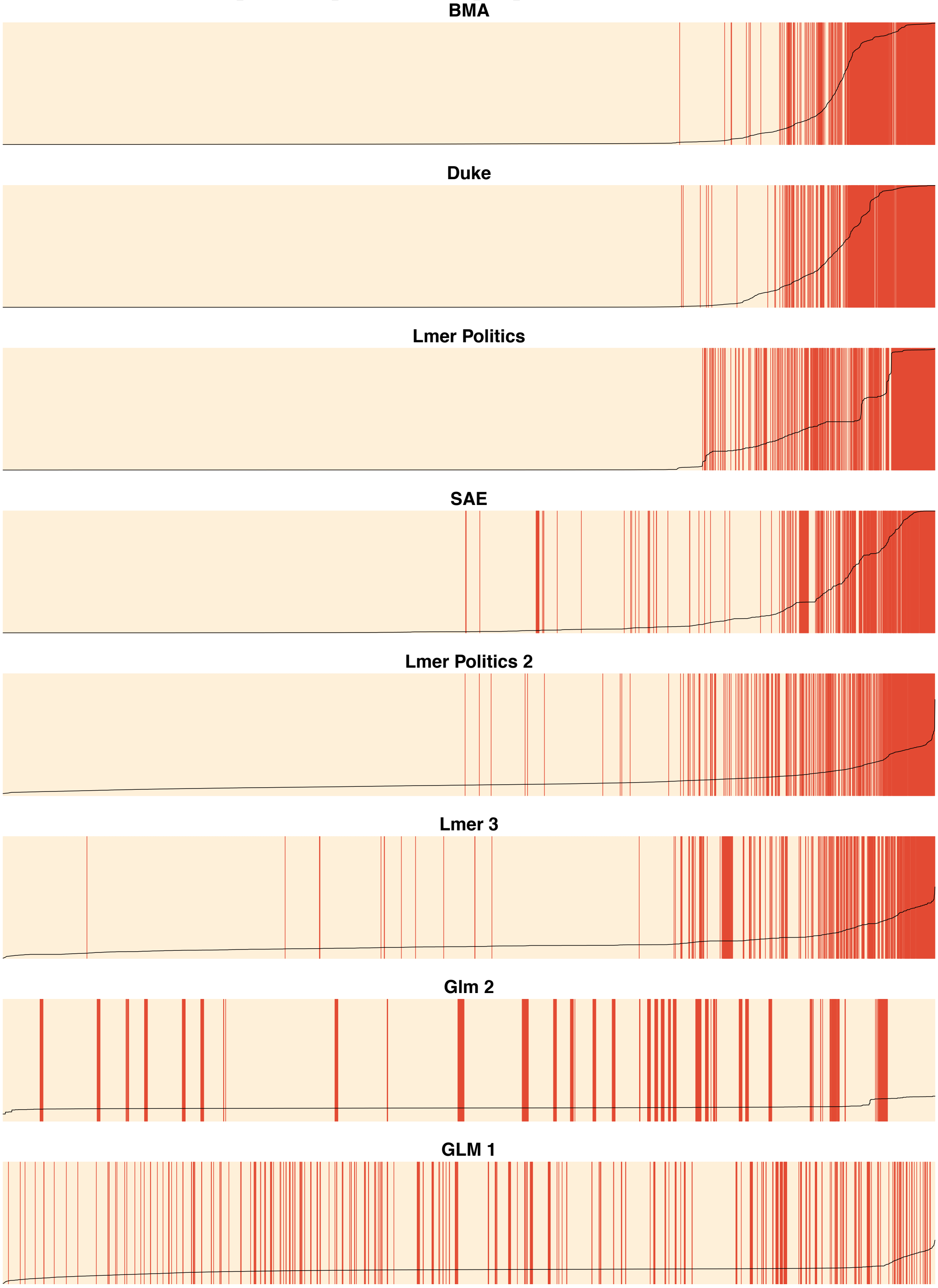
Testing the seven individual components and the EBMA model on the out-of-sample data set leads to similar results. Because of the high weight associated with the Duke component, the results from the EBMA process are almost indistinguishable from the Duke model output. Table 4 shows the results from the out-of-sample test of each of the components and the EBMA model, which here includes the Duke model.

TABLE 4. model statistics – out-of-sample predictions with the Duke model

	AUC	PRE	Brier	% Correct
Lmer Politics	0.94	0.05	0.06	89.94
Glm 1	0.72	0.00	0.09	89.37
Lmer Politics 2	0.97	0.00	0.07	89.37
Glm 2	0.84	0.00	0.09	89.37
Lmer 3	0.97	0.00	0.07	89.37
Duke	0.97	0.12	0.07	90.66
SAE	0.96	0.04	0.06	89.80
BMA	0.97	0.12	0.07	90.66

The out-of-sample predictions of the EBMA model are roughly equivalent to the Duke component. Furthermore, given the small number of actual occurrences in the data and the high

FIGURE 3. separation plots: in-sample prediction with the Duke model



BMA

Duke

Lmer Politics

SAE

Lmer Politics 2

Lmer 3

Glm 2

GLM 1

number of correctly predicted insurgencies in the Duke model, it is almost impossible to improve the forecast through EBMA. However, the EBMA model certainly does no worse than any of individual model and, in fact, does slightly better by some metrics.

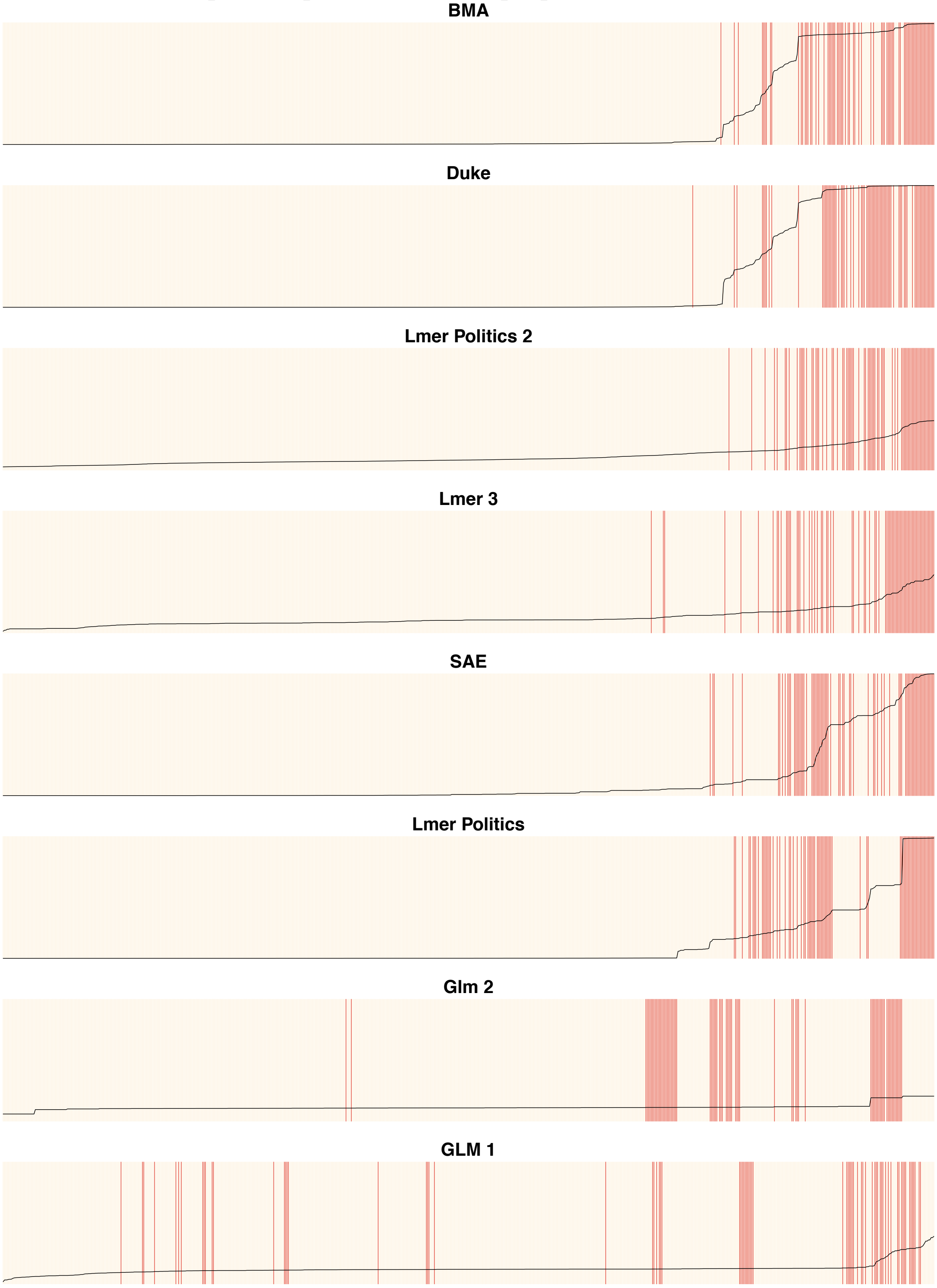
Figure 4 shows the separation plots of the EBMA model and its components. While the EBMA model obviously is very similar to the Duke model, one can see that the EBMA model performs slightly better and has considerable higher predicted probabilities for observations with actual events. However, as emphasized above, it is nearly impossible to improve the forecast by the Duke component on this data set by combining it with any other individual model. Thus even the slight improvement that is achieved here by the EBMA model shows the promise of the EBMA process in improving the accuracy of out-of-sample forecasts.

5. Discussion

In many areas, political scientists rarely make predictions about the future. To a large part this is due to the fact that it is almost impossible to find the true data generating process of events that are interesting to political science. Social events are inherently difficult to predict because of nonlinearities and the "unpredictability" of human behavior. Yet, we believe it should be the ultimate goal of political scientists to make sensible and reliable predictions into the future.

In this paper, we extended previous work on Ensemble Bayesian Model Averaging (EBMA) to make the technique applicable to binary dependent variables. In short, EBMA uses the accuracy of in-sample predictions of individual models to combine the power of multiple forecasts and make more accurate predictions for future events. Moreover, it does so in a transparent manner that allows us to see which component models are most important in informing the broader EBMA model. Thus, EBMA can help us to make reliable forecasts in political science more likely, while also allowing the continued development of multiple theoretical and empirical approaches to study important topics. In this paper we have shown how the method can be adjusted to work for dichotomous dependent variable. The EBMA model developed here, based on previous work Sloughter et al. (2007) and Sloughter, Gneiting and Raftery (2010), is

FIGURE 4. separation plots – out-of-sample prediction with the Duke model



thus applicable to a large fraction of research in political science, but the field of international relations in particular.

We have then shown the contribution of using EBMA to produce out-of-sample forecasts. Using six to seven individual models on the EBMA process and comparing the predictions of the individual components to the predictions of a final EBMA model, we were able to show that the method improves the accuracy of predictions considerably. Even though insurgency is an extremely rare even in our example data and improvements of predictions from the individual models are thus difficult to achieve, the EBMA predictions are always better or at least as good as any of the individual components. Given the difficulty associated with this particular dataset, the improvements of EBMA should be even more distinct when applied to other settings and tested on other datasets. In future versions of this paper we will provide more tests to the superiority of EBMA compared to individual models forecasts. Given the applicability of the framework to event data, one can also use EBMA to increase the accuracy of predicting rare conflictual events such as civil war, revolutions, international conflict.

An additional planned extension of the EBMA method will be greater inclusions of the uncertainty in component estimates in the final ensemble predictions. We anticipate that this may aid us to even further increase the improvement of forecasts achieved by EBMA as this information is currently lost through the sole reliance on point predictions from the component models.

REFERENCES

Abramowitz, Alan I. 2008. “Forecasting the 2008 Presidential Election with the Time-for-Change Model.” *PS: Political Science & Politics* 41(4):691–695.

Andriole, Stephen J. and Robert A. Young. 1977. “Toward the Development of an Integrated Crisis Warning System.” *International Studies Quarterly* 21(1):107–150.

Bartels, Larry. 1997. “Specification Uncertainty and Model Averaging.” *American Journal of Political Science* 41:641–674.

Bartels, Larry M. and John Zaller. 2001. “Presidential Vote Models: A Recount.” *PS: Political Science and Politics* 34(01):9–20.

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

Bennett, D. Scott and Allan C. Stam. 2009. "Revisiting Predictions of War Duration." *Conflict Management and Peace Science* 26(3):256–267.

Brandt, Patrick T., Michael Colaresi and John R. Freeman. 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *The Journal of Conflict Resolution* 52(3):343–374.

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78(1):1–3.

Campbell, James E. 1992. "Forecasting the Presidential Vote in the States." *American Journal of Political Science* 36(2):386–407.

Campbell, James E. 2008. "The Trial-Heat Forecast of the 2008 Presidential Vote: Performance and Value Considerations in an Open-Seat Election." *PS: Political Science & Politics* 41(4):697–701.

Clyde, Merlise. 2003. Model averaging. In *Subjective and Objective Bayesian Statistics*, ed. S. James Press. 2nd ed. Hoboken, NJ: Wiley-Interscience chapter Chap. 13, pp. 320–335.

Clyde, Merlise and Edward I. George. 2004. "Model Uncertainty." *Statistical Science* 19(1):81–94.

Cuzàn, Alfred G. and Charles M. Bundrick. 2008. "Forecasting the 2008 Presidential Election: A Challenge for the Fiscal Model." *PS: Political Science & Politics* 41(4):717–722.

Davies, John L. and Ted Robert Gurr. 1998. *Preventive Measures: Building Risk Assessment and Crisis Early Warning Systems*. Lanham, Md: Rowman & Littlefield Publishers.

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

de Mesquita, Bruce Bueno. 2002. *Predicting Politics*. Columbus, OH: Ohio State University Press.

Doran, Charles F. 1999. "Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics." *International Studies Review* 1(2):11–41.

Draper, David. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):45–97.

Enders, Walter and Todd Sandler. 2005. "After 9/11: Is it All Different Now?" *The Journal of Conflict Resolution* 49(2):259–277.

Erikson, Robert S. and Christopher Wlezien. 2008. "Leading Economic Indicators, the Polls, and the Presidential Vote." *PS: Political Science & Politics* 41(4):703–707.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97(1):75–90.

Feder, Stanley A. 2002. "Forecasting For Policy Making in the Post-Cold War Period." *Annual Review of Political Science* 5:111–125.

Freeman, John R. and Brian L. Job. 1979. "Scientific Forecasts in International Relations: Problems of Definition and Epistemology." *International Studies Quarterly* 23(1):113–143.

Geer, John and Richard R. Lau. 2006. "Filling in the Blanks: A New Method for Estimating Campaign Effects." *British Journal of Political Science* 36(2):269–290.

Gill, Jeff. 2004. "Introduction to the Special Issue." *Politi* 12(4):647–674.

Gleditsch, Kristian Skrede and Michael D. Ward. 2010. "Contentious Issues and Forecasting Interstate Disputes." Presented to the 2010 Annual Meeting of the International Studies Association.

Gurr, Ted Robert and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19(3):3–38.

Hamill, Thomas S., Jeffrey S. Whitaker and X. Wei. 2004. “Ensemble re-forecasting: Improving medium-range forecast skill using retrospective forecasts.” *Monthly Weather Review* (132):1434 – 1447.

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

Hildebrand, David K., James D. Laing and Howard Rosenthal. 1976. “Prediction Analysis in Political Research.” *The American Political Science Review* 70(2):509–535.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery and Christopher T. Volinsky. 1999. “Bayesian Model Averaging: A Tutorial.” *Statistical science* 14(4):382–417.

Holbrook, Thomas M. 2008. “Incumbency, National Conditions, and the 2008 Presidential Election.” *PS: Political Science & Politics* 41(4):709–712.

Imai, Kosuke and Gary King. 2004. “Did Illegal Overseas Absentee Ballots Decide the 2000 US Presidential Election?” *Perspectives on Politics* 2(03):537–549.

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

Krause, George A. 1997. “Voters, Information Heterogeneity, and the Dynamics of Aggregate Economic Expectations.” *American Journal of Political Science* 41(4):1170–1200.

Leblang, David and Shanker Satyanath. 2006. “Institutions, Expectations, and Currency Crises.” *International Organization* 60(1):245–262.

Lewis-Beck, Michael S. and Charles Tien. 2008. “The Job of President and the Jobs Model Forecast: Obama for ’08?” *PS: Political Science & Politics* 41(4):687–690.

Lockerbie, Brad. 2008. “Election Forecasting: The Future of the Presidency and the House.” *PS: Political Science & Politics* 41(4):713–716.

Madigan, David and Adrian E. Raftery. 1994. “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window.” *Journal of the American Statistical Association* 89(428):1535–1546.

Marshall, Monty G., Keith Jaggers and Ted Robert Gurr. 2009. “Polity IV Project: Political Regime Characteristics and Transition 1800-2007.” CIDCM: University of Maryland, MD.

Montgomery, Jacob M. and Brendan Nyhan. 2010. “Bayesian Model Averaging: Theoretical Developments and Practical Applications.” *Political Analysis* 18(2):245–270.

Norpoth, Helmut. 2008. “On the Razor’s Edge: The Forecast of the Primary Model.” *PS: Political Science & Politics* 41(4):683–686.

O’Brien, Sean P. 2002. “Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis.” *Journal of Conflict Resolution* 46(6):791–811.

O’Brien, Sean P. 2010. “Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research.” *International Studies Review* 12(1):87–104.

Pevehouse, Jon C. and Joshua S. Goldstein. 1999. “Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions.” *The Journal of Conflict Resolution* 43(4):538–546.

Raftery, Adrian E. 1995. “Bayesian model selection in social research.” *Sociological Methodology* 25(1):111–163.

Raftery, Adrian E., Tilmann Gneiting, Fadoua Balabdaoui and Michael Polakowski. 2005. “Using Bayesian model averaging to calibrate forecast ensembles.” *Monthly Weather Review* 133(5):1155–1174.

Raftery, Adrian E. and Yingye Zheng. 2003. “Long-run Performance of Bayesian Model Averaging.” *Journal of the American Statistical Association* 98(464):931–938.

Schrodt, Philip A. and Deborah J. Gerner. 2000. “Using Cluster Analysis to Derive Early Warning Indicators for Political Change in the Middle East, 1979-1996.” *American Political Science Review* 94(4):803–818.

Singer, J. David and Michael D. Wallace. 1979. *To Augur Well: Early Warning Indicators in World Politics*. Beverly Hills: Sage Publications.

Sloughter, J. McLean, Adrian E. Raftery, TTilmann Gneiting and Cchris Fraley. 2007. “Probabilistic quantitative precipitation forecasting using Bayesian model averaging.” *Monthly Weather Review* 135(9):3209–3220.

Sloughter, J. McLean, Tilmann Gneiting and Adrian E. Raftery. 2010. “Probabilistic wind speed forecasting using ensembles and Bayesian model averaging.” *Journal of the American Statistical Association* 105(489):25–35.

Vincent, Jack E. 1980. “Scientific Prediction versus Crystal Ball Gazing: Can the Unknown be Known?” *International Studies Quarterly* 24(3):450–454.

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. “Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace.” *American Journal of Political Science* 51(3):583–601.

Zaller, John R. 2004. Floating voters in U.S. presidential elections, 1948-2000. In *Studies in public opinion: Attitudes, Nonattitudes, Measurement Error, and Change*, ed. Willem E. Saris and Paul M. Sniderman. Princeton, NJ: Princeton University Press pp. 166–214.

Contact information for authors:

PROFESSOR M.D. WARD, DEPARTMENT OF POLITICAL SCIENCE, DUKE UNIVERSITY, DURHAM, NC, USA, 27708
E-mail address: michael.d.ward@duke.edu

JACOB M. MONTGOMERY, DEPARTMENT OF POLITICAL SCIENCE, DUKE UNIVERSITY, DURHAM, NC, USA, 27708
E-mail address: jacob.montgomery@duke.edu

FLORIAN M. HOLLENBACH, DEPARTMENT OF POLITICAL SCIENCE, DUKE UNIVERSITY, DURHAM, NC, USA, 27708
E-mail address: florian.hollenbach@duke.edu