Predicting Terrorism: A Machine Learning Approach

Working Paper · November 2017

DOI: 10.13140/RG.2.2.11401.34408

4 authors, including:

James Thomas Bang
Saint Ambrose University

Atin Basuchoudhary
Virginia Military Institute

Predicting Terrorism: A Machine Learning Approach

James T. Bang
Department of Economics
St. Ambrose University

Atin Basuchoudhary
Department of Economics and Business
Virginia Military Institute

John A. David
Department of Applied Mathematics
Virginia Military Institute

Aniruddha Mitra
Economics Program
Bard College

Abstract

This paper develops and applies new methods for predicting the incidence of terror attacks using machine-learning techniques. Our methodological approach helps identify the variables that best predict terrorist attacks from among the many "usual suspects" identified in the literature. We suggest that any causal explanation of terrorism should start here. Moreover, we notice that predictors of terrorism have non-linear effects. Thus, our methodological approach suggests future terrorism research ought to center on explaining complex non-linear effects of certain kinds of variables (assassinations for example) over others (democracy and economic institutions). Last, we suggest that as a matter of public choice, policy makers can use our methodological approach to understand (a) how to predict terror attacks, and (b) what approaches are predictably most likely to deter terror attacks. This understanding can help allocate terrorism fighting resources more efficiently.

1. Introduction

In this paper, we build machine learning (or, "artificial intelligence") models to *predict* which countries will experience more terror attacks. The extant literature is limited because it has so far failed to evolve a consensus on the key determinants of terrorism (Sandler, 2014), with results being sensitive to the measurement of terrorism, the choice of sample, and the statistical methodology employed. As such, existing models do not predict incidents of terrorism with an acceptable degree of accuracy and therefore are of limited use to the actual design of counterterrorism policy. This paper applies the emerging methodology of machine learning to the study of terrorism that can overcome some of the limitations of classical regression-based methods (Ward et al., 2010) to provide a methodology to evolve a scientifically validated consensus about the key determinants of terrorism and provide policy-makers a clear way to identify the best anti-terrorism levers at their disposal.

There are two methodological strands in the academic research on terrorism, namely psychological/sociological and economics/political science. The former has typically focused on identifying the archetypal terrorist or identifying the social conditions that beget terrorism. Both these attempts have been largely unsuccessful (Lacquer, 1999; Pape, 2005). While we recognize that this approach may be useful for countering terrorism at the operational level through behavioral profiling, for example; nevertheless, we do not take this approach.

The economics/political science approach tends to ask a different set of questions. One strand here tries to identify and deter terrorists through a cost benefit lens. Here deterrence at the operational level involves greater policing/punishment and policies to increase the opportunity costs of terrorism at the tactical level (Frey and Luechinger, 2003; Faria and Arce, 2005). Moreover, research has identified that terrorism is a tactical choice for successful rebellions e.g. in Algeria, Israel, and Cyprus (see Shughart 2002, 2006). Moreover, target choice by terrorists is also a tactical choice (Enders and Sandler, 2006).

Terrorists will substitute away from hard targets suggesting that piecemeal policies that focus on some targets at the expense of others may be unproductive.

Another strand in the literature investigates whether terrorists also choose targets as part of a broader asymmetric strategy where all they need to do is avoid losing while spreading fear and uncertainty within a population. There may be two strategic goals here – to convince a population to induce some desired policy change and/or impose some economic cost on a population as a signal of credibility and as a recruiting tool (Enders and Sandler, 2006, p. 24 and Crain and Crain, 2006 respectively). Last, there is a political economy of terrorism that also straddles the strategic, the tactical, and the operational use of terrorism. At the strategic level, terrorism may be an outcome of a lack of civil liberties or economic liberties. For example, Basuchoudhary and Shughart (2010) find that ethnic tensions may lead to terrorism unless good economic institutions mitigate that link. Thus, fostering economic opportunity in source regions for terrorists may be a strategic element of US national defense.

Current empirical research has tended towards identifying the "correlates" of terrorism and have largely failed to identify a consistent set of such correlates that can predict terrorism. Even the EBA of Gassebner and Luechinger's (2011) efforts at identifying consistent correlates of terrorism are not predictive out of sample. This uncertainty about whether the correlates of terrorism are merely artifacts of, for example, overfitting within a sample, or have real teeth in explaining terrorism can only be tested if they also predict terrorism. A good theory of terrorism should be able to predict terrorism just like a good theory of gravity should be able to predict variations in gravity. Predicting terrorist attacks have so far been largely speculative. Machine learning algorithms can provide scientifically validated predictions of the likelihood of a terrorist attack. Because of our research, national security agencies would have a validated, and parsimonious, list of variables (i.e. policy levers) that can best predict terrorism. We can rank this list of variables in order of predictive importance. If amenable to manipulation in a predictable way policy makers can use these to deter terrorism. Moreover, the ranking can help the policy-maker choose the most effective policy tool to make public policy to deter terrorism.

To our knowledge, no study has closely examined a comprehensive list of "usual suspects" surveyed in the literature as potential predictors of terror. To achieve the our aim, we collect cross-country data on terror events from the Global Terror Database and from several cross-national data sources (World Development Indicators, International Country Risk Guide, Cross-National Time Series, World Values Survey, for example) to compare the accuracy of several standard regression models with various ML models for predicting terror attacks. This will allow us to cull the potential list of covariates (Gassebner and Luechinger, 2011) into a relatively parsimonious list of variables that are strong predictors of terrorism. We describe various approaches to this culling in section 2 and describe the data in section 3. In section 4, we show how to identify the most predictive models by choosing the model with the least out of sample predictive error. These tie down our idea that any claim that there is a universal basis to terrorism must be validated out of sample. Moreover, we suggest that this model can be used to help predict terrorism before it happens, but only to the extent it predicts out of sample. This places a meaningfully validated limitation on both predictive accuracy and the extent to which current approaches in economics actually explain terrorism. We then identify and rank the most important predictors of terrorism. We suggest that this ranking, within the limitations we refer to above, can help policy makers choose the most effective tool to deter terrorism.

2. Machine Learning

Machine learning (ML) methods are a growing, yet still far under-utilized set of methods for predicting and classifying various outcomes. While they have gained significant credibility in the business world (they are used to target online advertising, predict movie preferences, and prioritize online news feeds, for example), they have been slow to gain traction in the academic literature for predicting economic variables (such as growth or recessions), political events (such as regime change, conflict, or terrorism), or health outcomes (such as hospital readmissions or death). One of the main innovations of this paper is to introduce this approach to the analysis and prediction of terrorism.

We will build an empirical model using several parametric and non-parametric ML techniques (classical regression, Poisson regression, artificial neural network, regression tree, bootstrap aggregating, boosting, and random forest) to measure how and how well publicly-available economic, geographic, and institutional variables *predict* the frequency and severity of terror attacks (Hand, Mannila, & Smyth, 2001). The first step in this process will be to identify the Machine Learning approach that best predicts terrorism. Next, using the best technique, we will identify the most important variables for predicting terrorism. Finally, we will plot the partial dependency for terrorism with respect to each variable to show how each variable impacts terrorism across the distribution of its values. This is important because there may be reason to believe that many of the correlates of terrorism have non-linear impacts.

ML techniques identify tipping points in the range of a particular variable that may place a country at a lower or higher risk of terrorism. We illustrate these tipping points using partial dependence plots which show how the incidence and severity of terror attacks fluctuate across the possible values for each of the variables. Further, by identifying the variables that have the *most* predictive power we could help develop a framework to distinguish between competing theoretical explanations of terrorism. Suppose, for instance, political models of terrorism may suggest that terrorism may be a tactic employed by disenfranchised groups that have little or no voice in government, whereas economic models may suggest that groups employ terrorism as a signal of credibility in order to gain a seat at the negotiating table against the regime when it divvies up rents from resource wealth. If ML methodologies rank democracy as a better predictor of terrorism than primary commodities exports, for example, we can assume that the political model may be a better explanation of terrorism than the economic model, or vice versa. Moreover, this ML approach can help eliminate correlates of conflict that do not predict terrorism well. Presumably, correlates that do not predict well cannot really be considered as variables that cause terrorism.

Our ML approach, and the econometric tests arising out of this approach, will help us better understand causal patterns explaining terrorism. Moreover, we offer a better understanding of how terrorism can be *predicted*, which will be of particular help to policy makers as they design policies of economic terrorism. The remainder of this section outlines the prediction algorithms we use to predict the aggregate terror risk for a country. Readers who are familiar with these algorithms – or will be bored by a technical description of them! – may skip to the results section. Those looking for a more detailed description of the algorithms may consult the coverage of them by Hand, et al. (2001).

2.1 Classical and Other Regression Analysis

In general, using given data from a learning sample, $\mathcal{L} = \{(y_1, \mathbf{x}_1), \ldots(y_N, \mathbf{x}_N)\}$, any prediction function, $d(\mathbf{x}_i)$, maps the vector of input variables, $\mathbf{x}$, into the output variable (the number of terror attacks), y. An effective prediction algorithm seeks to define parameters that minimize some error function over the predictions. Common error functions that many algorithms use include the mean of the absolute deviations (MAD) of the observed values from the predicted values or the mean of the squared errors (MSE). In linear regression models $d(\mathbf{x}_i)$ is simply a linear function of the inputs and their respective slope coefficients, plus a constant, $d(\mathbf{x}_i) = \mathbf{x}_i\beta$. A linear model and an MSE error function yields the ordinary least squares (OLS) regression model:

$$R_{OLS}(d) = \frac{1}{n}\sum_{i=1}^{N}(y_i - d(\boldsymbol{x}_i))^2,$$

where $d(\mathbf{x}_i) = \mathbf{x}_i\beta$ is a linear function of the inputs.

Although OLS can sometimes yield good predictions (on average the *best* prediction among all linear models, in fact), it has some undesirable properties in the case of predicting terror attacks. Specifically, since a large number of cases in our sample experience no terror attacks at all, while some of them experience very large numbers of attacks, we will expect the OLS model to predict *negative* numbers of terror attacks for some observations – which is non-sensical!

As an alternative, one correct this problem by estimating a Poisson regression, which will estimate the average number of terror attacks conditional on the inputs, $\mathbf{x}$, to be an exponential function of a linear combination of the inputs expressed as:

$$\lambda = E(y|\mathbf{x}) = e^{\beta\mathbf{x}}.$$

This means that the probability of observing a specific number of terror attacks will be:

$$p(y|\mathbf{x}) = \frac{e^{y\mathbf{x}\beta}e^{-e^{\mathbf{x}\beta}}}{y!}.$$

The Poisson model then proceeds by estimating the parameters to maximize the likelihood function for this Poisson probability distribution.

While these more sophisticated regression methods successfully purge the bias from the individual parameter estimates that might result from over-dispersion, they do so to the detriment of the model's overall predictive accuracy. Alternative approaches, which will ensure a relatively high degree of accuracy, while also avoiding nonsensical predications, use non-parametric tree methods, or combinations of trees to predict the number of terror attacks.

2.2 Artificial Neural Networks (ANNs)

A feedforward artificial neural network is a series of logistic regression models[1] connecting each of the *K* input variables to *M* hidden nodes, over which, in the case of a regression problem like ours (as opposed to a classification problem in the case of a binary target variable), a linear regression connects

[1] The logistic function is the usual activation (or link) function in the first layer; in general, any sigmoid function (e.g. *tanh*) will have the desired properties.

the hidden nodes to the output we hope to predict in the final layer (Murphy, 2012).[2] Hence, with one output node, an ANN estimates *K·M* parameters.

We present a diagram of a simple ANN for predicting terror attacks in Figure 1. In the figure, each connection corresponds to a weight for each *input* variable (*I*1, …, *I*5) and bias (constant) terms (*B*1 and *B*2) into the hidden nodes (*H*1 and *H*2), or into the output node (O1). In the diagram, we show a two-layer neural network (the inputs do not count as a layer) with five inputs, two hidden nodes, and a constant. The link function connecting the hidden layer to the outputs, and which is not explicitly shown, is linear.

Using a least-squares objective, the estimation of the ANN minimizes:

$$R(\alpha, \beta; x) = \sum_{i=1}^{N} \left(y_i - f(\alpha, \beta; x_i)\right)^2,$$

where $f(\alpha, \beta; x_i) = Z\beta$ connects the hidden layer to the output, and $Z = \frac{1}{1+\exp(X\alpha)}$ is the logit function connecting the inputs to the hidden layer. Using the first order conditions with respect to the parameters for the hidden layer, α, and the parameters to the output layer, β, the estimation finds the solution according to a gradient descent rule:

$$\beta_m^{r+1} = \beta_m^r - \sum_{i=1}^{N} \frac{\partial R}{\partial \beta_m^r} - \lambda \beta_m^r,$$

where *l* is called the "weight decay" and acts like a penalty on the parameter, and effectively restricts the parameters towards zero to avoid "overfitting" the model to the learning sample.

ANNs often perform well in situations where the interplay between components of the input is more important than any of their individual values, and as such they are often used in image and pattern recognition problems. We estimate the network using the *nnet* package implemented in *R* (Ripley and

---

[2] In a regression problem like ours, the final layer usually contains only one output; the same is true for classification problems involving a binary output. For classification problems involving multinomial outputs, there can be any munber

Venables, 2015). This implementation uses a single hidden layer (in which we used 100 nodes and 100 iterations). This work used all default options, save for specifying that the final layer should be linear. The initial weights were chosen randomly and the goal function was the sum of the squared errors.

2.3 Regression Trees

Classification and regression trees (CART)[3] diagnose and predict outcomes by finding binary splits in the input variables to optimally divide the sample into subsamples with successively higher levels of accuracy in the output variable, *y*. So, unlike linear models, where the parameters are linear coefficients on each input variable, the parameters of the tree models are "if-then" statements that split the dataset according to the observed values of the inputs.

More specifically, a tree, *T*, has four main parts:

1. Binary splits to splits in the inputs that divide the subsample at each node, *t*;
2. Criteria for splitting each node into additional "child" nodes, or including it in the set of terminal nodes, T^*;
3. A decision rule, $d(\mathbf{x})$, for assigning a predicted output value to each terminal node;
4. An estimate of the predictive quality of the decision rule, *d*.

The first step is achieved at each node by minimizing a measure of impurity. The most common measure of node impurity, and the one we use for our tree algorithms, is the mean square error, denoted $\hat{R}(d)=\frac{1}{n}\sum_{i=1}^{N}(y_i - d(\mathbf{x}_i))^2$. Intuitively, this method searches for the cutoff in each input that minimize errors, then selecting which input yields the greatest improvement in node impurity using its optimal splitting point.

Then, a node is declared to be terminal if one of the following conditions is met: (1) that the best split fails to improve the node impurity by more than a predetermined minimum improvement criterion; or (2)

[3] We provide only a brief summary of tree construction as it pertains to our objectives. For a full description of the CART algorithm, see Breiman, et al. (1984).

the split creates a "child" node that contains fewer observations than the minimum allowed.[4] At each terminal node, the decision rule assigns observations with a predicted outcome based on some measure of centrality. In the case of count (number of terror attacks or fatalities) or continuous (amount of property damage) outcomes centrality is usually the mean of the observations conditional on reaching that node.

The predictive quality of the rule is also evaluated using the *mean square error*, $\hat{R}(d)=\frac{1}{n}\sum_{i=1}^{N}(y_i - d(\boldsymbol{x}_i))^2$. This misclassification rate is often cross-validated by splitting the sample several times and re-estimating the misclassification rate each time to get an average misclassification of all of the cross-validated trees.

2.3.1 Boosting Algorithms

Iteratively re-estimating or combining ensembles of trees by averaging their predictions can often improve the accuracy of a tree algorithm. Boosting algorithms, bootstrap aggregating (bagging), and random forests all predict outcomes using ensembles of classification trees. The basic idea of these algorithms is to improve the predictive strength of a "weak learner" by iterating the tree algorithm many times by either modifying the distribution by re-weighting the observations (boosting), randomly resampling a subset of the learning sample (bagging), or randomly sampling subsets of the input variables (random forest). Then either classify the outcomes according to the outcome of the "strongest" learner once the algorithm achieves the desired error rate (boosting), or according to the outcome of a vote by the many trees (bagging).

Boosting has been proposed by Schapire (1990) and Freund and Schapire (1996) to augment the strength of a "weak learner" (an algorithm that predicts poorly). Specifically, for a given distribution $\mathcal{D}$ of

[4] Note that there is a tradeoff here: setting lower values for the minimum acceptable margin of improvement or the minimum number of observations in a child node will lead to a more accurate prediction (at least within the sample the model uses to learn). However, improving the accuracy of the algorithm within the sample may lead to over-fitting in the sense that the model will perform more poorly out-of-sample.

importance values assigned to each observation in L, and for a given desired error, $\tilde{R}$, and failure probability, ϕ, a *strong learner* is an algorithm that has a sufficiently high probability (at least $1 - \phi$) of achieving an error rate no higher than $\tilde{R}$. A weak learner has a lower probability (less than $1 - \phi$) of achieving the desired error rate. Boosting algorithms for classification create a set of M classifiers, $F = (f_1, \ldots, f_M)$that progressively re-weight the importance of each observation based on whether the previous classifier predicted it correctly or incorrectly. Modifications of the boosting algorithm for classification have been developed for regression trees by Freund and Schapire (1997) and Friedman(2001).

Starting with a $\mathcal{D}_1 = (1/N, \ldots, 1/N)$, suppose that our initial classifier, $f_1 = T$ (single-tree CART, for example)¸ is a "weak learner" in that the misclassification rate, $\hat{R}(d)$ is greater than the desired maximum desired misclassification rate, $\tilde{R}$. Next, for all observations in the learning sample, recalculate the distribution weights for the observations as:

$$\mathcal{D}_2 = \frac{\mathcal{D}_1(i)}{Z_2} \times \begin{cases} \frac{\hat{R}_1(d)}{1-\hat{R}_1(d)} & if\ d_1(\mathbf{x}_i) = y_i \\ 1 & otherwise \end{cases},$$

where Z_m is a scaling constant that forces the weights to sum to one.

The final decision rule for the boosting algorithm is to categorize the outcomes according to $d(\mathbf{x}) = \arg\max_{y \in Y} \sum_{m:d_m(\mathbf{x})=y} \log\left(\frac{1-\hat{R}_m(d)}{\hat{R}_m(d)}\right)$. Using this decision rule and its corresponding predictions, we calculate the estimate of the misclassification rate in the same way as in step (4) of the single tree algorithm.

2.3.2 Bootstrap Aggregating (Bagging)

The bagging method proposed by Breiman (1996) takes random resamples, $\{L^{(M)}\}$, from the learning sample *with replacement* to create M samples using only the observations from the learning sample. Each of these samples will contain N observations – the same as the number of observations in the full training sample. However, in any one bootstrapped sample, some observations may appear twice (or more), others not at all.[5] The bagging method then adopts the rules for splitting and declaring nodes to be terminal described in the previous section to build M classification trees.

To complete steps (3) and (4), bagging needs a way of aggregating the information of the predictions from each of the trees. The way that bagging (and, as we will soon see, a random forest) does this for class variables is through *voting*. For *classification trees* (categorical output variables), the voting processes each observation[6] through all of the M trees that was constructed from each of the bootstrapped samples to obtain that observation's predicted class for each tree. The predicted class for the entire model, then, is equal to the mode prediction of all of the trees. For *regression trees* (continuous output variables), the voting process calculates the mean of the predicted values for all of the bootstrapped trees. Finally, the bagging calculates the redistribution estimate in the same way as it did for the single classification tree, using the predicted class based on the voting outcome.

2.3.3 Random Forests

Like bagging, a random forest is a tree-based algorithm that uses a voting rule to determine the predicted class of each observation. However, whereas the bagging randomizes the selection of the observations for each tree, a random forest may randomize over multiple dimensions of the classifier (Breiman, 2001). The most common dimensions for randomizing the trees are the selection of the input variables for node of each tree, as well as the observations included for constructing each of the trees. We briefly describe the construction of the trees for the random forest ensemble below.

---

[5] Note that the probability that a single observation is selected in each draw from the learning set is $1/N$ Hence, sampling with replacement, the probability that it is completely left out of any given bootstrap sample is $(1 - 1/N)^N$. For large samples this tends to $1/e$. The probability that an observation will be completely left out of all M bootstrap samples, then, is $(1 - 1/N)^{NM}$.

[6] Note that the observations under consideration could be from the in-sample learning set or from outside the sample (the test set).

A random forest is a collection of tree decision rules, $\{d(\mathbf{x}, \Theta_m), m = 1, ..., M\}$, where Θ_m is a random vector specifying the observations and inputs that are included at each step of the construction of the decision rule for that tree. To construct a tree, the random forest algorithm takes to following steps:

i. Randomly select $n \leq N$ observations from the learning sample;[7]
ii. At the "root" node of the tree, select $k \in K$ inputs from **x**;
iii. Find the split in each variable selected in (ii) that minimizes the mean square error at that node and select the variable/split that achieves the minimal error;
iv. Repeat the random selection of inputs and optimal splits in (ii) and (iii) until some stopping criteria (minimum improvement, minimum number of observations, or maximum number of levels) is met.

The bagging method described in the previous sub-section is in fact a special case of a random forest where, for each tree, Θ_m consists of a random selection of $n = N$ observations from the learning sample with replacement (and each observation having a probability of being selected in each draw equal to $1/N$) and sets the number of inputs to select at each node, k, equal to the full length of the input vector, K so that all of the variables are considered at each node.

2.4 Validation and Testing of Predictive Accuracy

Once we have built our learning algorithm, the next issue is to evaluate the validity of our error estimates and the predictive strength of our models. Error estimates ($R[d]$) can sometimes be misleading if the model we are evaluating is over-fitted to the learning sample. These error estimates can be tested out-of-sample or cross-validated using the learning sample.

To test the out-of-sample validity, we simply split the full dataset into two random subsets of *countries*: the first, known as the *learning sample* (or training sample) contains the countries and

7 In contrast to bagging, where the number of observations selected for each tree exactly equals the total number of observations in the learning sample, and the draws always sampled with replacement, the number of observations selected for each tree of the forest can be set to be less than the total size of the learning sample, and can therefore be sampled with or without replacement. This also allows for slightly greater flexibility with respect to stratified or clustered sampling from the learning sample.

observations that will build the models; the second, known as the *test sample*, will test the out-of-sample predictive accuracy of the models. The out-of-sample error rates will indicate which models and specifications perform best, and will help reveal if any of the models are over-fitted.

To validate the error rates, machine learning uses either hold-out validation or cross-validation. In our study, we have used hold-out validation, which involves training the models using one portion (in our case 70% selected at random) of the dataset and measuring the accuracy of the model using the remaining 30%.

3. Data

As a first step in analyzing some preliminary data on terrorism, we have predicted the number of terror attacks using each of the seven models described above (OLS regression, Poisson regression, regression tree, random forest, bagging, and boosting). For our pilot specification, we have included 69 input (or explanatory) variables that cover most of the ones discussed in Gassebner and Luechinger's (2011) survey of the empirical literature on conflict.

We measure our output (or "dependent") variable, Terror Attacks, as the total number of terror attacks in a country in the last five years. This variable comes from the Global Terror Database published by the University of Maryland and covers the years 1970-2014.[8] In order to maintain the spirit of "prediction" in our model, we then consider our input ("explanatory") variables as five-year lagged averages of the preceding five years. Moreover, we only consider the variables at non-overlapping five year intervals so that none of the same information is contained in consecutive time intervals in our sample. In this sense, at any given point in time policy makers will be able to use our model to predict whether a country will be likely to experience a greater or lesser number of terror incidents in the next five years.

[8] When we combine all of the variables, our sample covers 1975-2014, since some entire data sources, like the Database of Political Institutions, don't become available until 1975.

From the Cross-National Time Series (CNTS; Banks, 2015) we take the numbers of assassinations, demonstrations, government crises, guerrilla warfare incidents, purges, riots, and strikes as measures of underlying low-level social instability. We also take the number of cabinet changes and executive changes as measured of political instability, and the effectiveness of the legislature as a measure of political legitimacy.

From the Database of Political Institutions (DPI; Cruz et al., 2016) we take the number of checks on power; executive and legislative indices of electoral competition; legislative, government, and opposition fractionalization indices; government Herfindahl index; and government polarization index as measures of the concentration (or not) of power and accountability (or not) within the government. We then include the changes in veto players, existence of electoral fraud, executive tenure, presence of a military executive, as measures of political stability and executive power. Finally, we include plurality voting and proportional representation as indicators of structural differences in electoral rules.

Next, we take several indices of government quality from the International Country Risk Guide (ICRG; PRS Group, 2015). It is important to remember that, for each of the ICRG indices, a higher value always coincides with "better" outcomes on this dimension of institutional quality. For example, in the case of the "internal conflict" (or "external conflict") index, a higher value for the index somewhat counterintuitively corresponds to less conflict. The same can be said for "ethnic tensions," "religious tensions," and "military in politics" – in each of these cases higher values relate to less of the (bad) thing that the variable name implies. That being said, we include the following indices from the ICRG: the bureaucratic quality and corruption indices as measures of the transparency of government; ethnic tensions, external conflict, internal conflict, law and order, and religious tensions as measures of the levels of latent (or open) social hostility, and the government's ability to ease those hostilities; government stability and investment profile indices as measures of the government's credibility in carrying out stated policies and refraining from expropriation; and democratic accountability and military in politics indices as a measures of the legitimacy and responsiveness of the regime to the public's preferences. We also add the Polity2 index and regime durability from the Polity IV Project as additional measures of legitimacy and responsiveness.

As measures of economic and cultural divisions within society, we include measures of income inequality, as well as ethnic and religious fractionalization. The former comes from the Standardized World Income Inequality Database (SWIID; Solt, 2014). The latter come from Reynal-Querol (2002), which in turn come from the *Atlas Naroda Mira* (Atlas of the People of the World; USSR, 1964).

Finally, we include numerous measures of economic human development from the World Development Indicators from the World Bank. They are: aid and development assistance; arms exports and imports; public education and health spending; female labor force participation; foreign direct investment (FDI); fuel exports; gross domestic product (GDP) per capita; government consumption; the stock of foreign born immigrants; infant mortality; the inflation rate in consumer prices; life expectancy; literacy; military expenditures; military personnel; population and its rate of growth; portfolio investment; primary, secondary, and tertiary school enrollment rates; social contributions; telephones per 100,000 people; the unemployment rate; urban population; and the youth dependency ratio.

Rather than exhaustively describing the distributional characteristics and justifying the inclusion of each variable, we kindly refer the reader to visit Gassebner and Luechinger's (2011) survey and the references therein to the various studies that have already provided such a description and justification. For readers with an interest in some of the characteristics of the observed data in our sample, we have included the descriptive statistics for all 69 variables in Table 1.

We can see from the table that each of our explanatory variables has a problem of omitted values to varying degrees. The tree-based methods (single trees, boosting, bagging, and random forest) can exploit the full information available automatically by using surrogate information, or using the median or mode at that branch of a tree as a best guess for the value of a missing data point. Standard parametric methods (in our case Poisson regression and neural networks) do not do this automatically, and regression methods that do (like full-information maximum likelihood), might do so in ways that give different imputations of missing data.

To resolve this, we pre-process our data using random forest imputation. The basic idea is that we consider a covariate that does not have missing data (in our case conflict), and perform a random forest model to predict that variable (instead of the true variable of interest since that would be "cheating" for running the full model). Next, whenever the algorithm encounters a missing value at any node of a tree, the imputation substitutes the median or mode for that variable and continues with the subsequent splits. The imputed values in each tree therefore exploit the full complement of conditional distribution for that variable based on that tree. Averaging over all of the trees, we obtain imputed values for missing data points that uses as much relevant data about the conditional distribution of the variable as possible. It also has the advantage of creating imputed values that are naturally bounded by the domains of the observed data. Parametric methods like multiple imputation estimate parameters based on an assumed distribution for the missing variables, and depending on the sensitivity of the parameters and the distributions of the covariates, may lead to extreme values outside of the logical bounds for a given variable (e.g. negative income).

3. Results

3.1 Predictive Quality

Table 2 reports the predictive quality of each of the models using the 70 variables. The best models we see to predict the overall number of terror attacks are the single regression tree, random forest and Bagging predictors, which reduce the overall MSE in the learning sample by about 64%, and 63%, and 59%, respectively, compared to the unconditional sample mean. An average of all of the predicted values of all of the models (which sometimes provides a better prediction, especially in cases of classification) improves the MSE by about 49%, while OLS regression improves the MSE by about 26%. However, as we might expect, the trees that use random bootstrapping (bagging and random forest) predict considerably better out of sample, with a test sample MSE reduction of 71% and 70% of the total MSE, respectively. Of particular interest, is the fact that these models achieve a significant reduction in the MSE despite the exclusion of the lagged number of terror attacks in our model,[9] since the pre-existing level of violence has been shown to be one of the strongest predictors of current and future violence in studies of conflict (Bang et al., 2016).

It is worth noting that the Poisson regression model, which tends to yield more valid estimates of causal effects, actually *increases* the MSE of the predictor in comparison to a prediction based on the simple sample mean. This is not quite the case for the neural network model, but we can see that the neural network and boosting models predict fairly poorly both in and out of sample.

The reader will note that our models included the "usual suspect" correlates of terrorism (and then some!) identified in the literature. A good theory should explain a model predictively. An earth centric model of the solar system fails to predict the movement of celestial bodies while a heliocentric model does. That argues in favor of the helio-centric model. Similarly, the most predictive model of terrorism probably comes closest to identifying the variables that explain terrorism the most. Moreover, the reductions in out of sample mean square error in the different models reported here are a quantitative measure of the extent of our knowledge of terrorism. This may be the best we can do to provide a predictive terrorism model in the absence of any way to do randomized controlled trials or actual behavioral experiments. The reader will note that the primacy of tree based models in predicting terrorism suggests there may not even be a universal theory of terrorism – different countries may have different pathways into terrorism. Nevertheless, we do provide the policy-maker with a sense of the best they can

[9] We exclude lagged terror attacks because we are partly looking to predict (a reason to include), but also looking to select a model to build theories and test causal effects (subsequent analyses). Lagged attacks would improve the prediction, but would explain *so much* of the variation that we're not left with much to select a model on.

do to predict and therefore preemptively deter terrorism. Last, the most important predictors in the "best" model identified here should be good candidates for a causal theory of terrorism. It is to this aspect we turn to next.

3.2. Variable Importance

Table 3 reports the variable importance levels (measured as the percentage of the total reduction in MSE that is attributed to that variable) based on the single regression tree, boosting, bagging, and random forest models, which predicts conflict the best. Here, we see that the first five variables in the list account for close to one-third (about 31 percent) of the overall improvement in the random forest model's MSE. We also see that the single strongest predictor of current levels of terrorism is a history of assassinations in that country, which accounts for about 12% of the total reduction in the MSE in the random forest model, and 25% of the reduction for the bagging model and 63% of the reduction for the boosting model. The second strongest predictor, guerrilla war, accounts for about 10% of the MSE reduction for the bagging and forest models, and over 30% of the reduction for the boosting model.

After regime-directed violence, two of the next three strongest predictors involve the extent to which the military engages with everyday life and politics. Military personnel and the military in politics index, account for almost 15% of the reduction in MSE combined, on average (slightly more in the single tree, slightly less in the bagging and forest models, and not at all in the boosting algorithm). In between these measures of military engagement, we see religious fractionalization to account for about 7% of the variation on average. Rounding out the top ten predictors are health spending (3.9% of the MSE), the time a time trend (3.8%), population (3.6%), executive tenure (3.2%), and fuel exports (2.6%).

What can we learn from all this? It may be useful to turn to the variables that ended up being poor predictors of terrorist attacks. For example, democratic accountability, investment profile, or ethnic fractionalization are not very good predictors of terrorism. This suggests that theoretical models that purport to explain terrorism through those lenses are not very powerful even though many of the extant models do look at terrorism through these lenses. It is here that our approach affects public choice of anti-terrorism policy the most. Since things like democracy do not appear to predict terrorism, a “war on terror” that emphasizes the spread of Jeffersonian democracy may be less effective than an emphasis on stable civilian led governments that are able to provide public goods like health care effectively.

A good theoretical model of terrorism must be able to explain e.g. the primacy of the specific kind of political uncertainty that assassinations bring to the table; or, how the military influences terrorism when entangled in a country’s politics. Such models do not exist at this time. They should. Further, in the absence of randomized experiments to tie down a causal theory of terrorism our approach may be the best way to identify a set of predictors. Moreover, we suggest that if there is to be a consensus around what sorts of variables cause terrorism they should center on those variables that predict terrorism better than others do. This, we suggest, is a simple scientifically validated heuristic. Further, if there is suspicion that something else may be a better explanation for terrorism the question simplifies to – does adding that variable increase predictive power. Thus, our methodological approach can help further our understanding of terrorism.

Last, researchers often use p-values to say that a hypothesis is correct. Recent guidance from the American Statistical Association (when responding to the fact that a spate of studies that could not be replicated), suggested that this interpretation is a misuse of the *p*-value. A significant *p*-value can come about through sheer chance. We suggest that machine-learning technologies can strengthen the social sciences by adding the element of prediction. A variable with a significant *p*-value but which does not predict the outcome well may well be significant by pure chance. The variability in what matters for explaining terrorism that is rampant in the current literature may be an artifact of chance. On the other hand, a highly significant causal variable (as opposed to one that has, by chance, a low *p*-value) should be able to predict out-of-sample. Thus, prediction adds a layer of discipline to investigative results.

3.3 Partial Dependence

The next step is to analyze *how* each of the variables affects aggregate terror risk. To do this we use a *partial dependence plot* mapping the possible values of the input variable of interest onto the observed incidence of terror attacks. Partial dependence plots display the marginal effect of variable x_k conditional on the observed values of all of the other variables, ($x_{1,-k}$, $x_{2,-k}$, … $x_{n,-k}$) Specifically, it plots the graph of the function:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} f(x_k, x_{i,-k}),$$

where the summand, $f(x_k, x_{i,-k})$, is simply the observed outcome of the number of terror attacks.

We have included the partial dependence plots for the top 10 variables in the variable importance list as figures at the end of our paper. These variables demonstrate some important aspects of both conflict and terrorism.[10]

First, Figure 2 illustrates that assassinations increase terror. There is compelling evidence that assassinations increase the likelihood of political violence (Iqbal & Zorn, 2008). Successful assassinations impede the efficiency with which the regime can take decisions on combating terrorism. This is particularly clear in the case of an autocratic regime with relatively few key decision-makers. This result also holds even for a relatively democratic regime since, at a minimum, replacing an assassinated politician may take time and may change the policy preferences of the regime so as to make it difficult to reach a collective decision in a short span of time. Assassinations thus reduce the price of terror relative to peaceful negotiations and increase the level of terrorism.

[10] We note that factors that play heavily into the prediction of conflict may differ in important ways from those factors that we have shown to predict terror attacks

Figure 3 shows that guerrilla warfare increases terrorism. While there may be some overlap between guerrilla warfare and terrorism, agencies that make strategy over national security policy tend to define them as distinct phenomena. Hence, in some cases we might think of terrorism and guerrilla warfare as different tactics employed by rebel groups towards similar ends. In this context, guerilla warfare reduces the credibility of the regime to protect the population, while simultaneously increasing the price of peace.

Next, in Figure 4, more military personnel also translates into more terror attacks. Here, we note that the efficiency of a military depends on enforcing a cohesive hierarchy. As with any organization, the military is subject to collective action and agency problems. Since both types of problems tend to increase with the size of a group (Olson, 1965), it is not hard to visualize that an increase in the number of military personnel may actually impede its efficiency in combating terrorism. At the same time, a more potent military force may tip the choice of an insurgency away from conventional battle tactics in favor of random terror attacks. In terms of our model, both effects translate into a reduced price of terror and more terrorism.

Interestingly, and as we show in Figure 5, religious fractionalization tends to *reduce* the risk of terror through most of the distribution of values. Religious fractionalization may increase the diversity of preferences within a terrorist organization and hence, reduce its internal cohesion Collier & Hoeffler (2004). Greater religious diversity also poses a challenge to terror mobilization, which requires the leaders to make credible commitments to recruits about the ex post distribution of rents. Shared ethnic or religious ties between the leaders of the coalition and the recruit pool serve to enhance the credibility of such commitments (Weinstein, 2007). It therefore follows that greater religious fractionalization will undermine the signaling mechanism and impede mobilization towards terrorism leading to a reduction in its quantity.

In contrast to the overall size of the military, we see in Figure 6 that increased military involvement in politics *reduces* aggregate terror risk. This provides an interesting counterpoint to the argument that military regimes are more vulnerable to terrorism (Wilson & Piazza, 2013). Geddes (1999:126) observes that:

> '…most professional soldiers place a higher value on the survival and efficacy of the military itself than on anything else… This corporate interest implies a concern with the maintenance of hierarchy, discipline, and cohesiveness within the military; autonomy from civilian intervention; and budgets sufficient to attract high-quality recruits and buy state-of-the-art weapons.'

Hence, a military regime does not have to address the diversity of political preferences on terrorism that may impede efficient decision making in a democratically elected regime. *Too much* of a divide between political leaders and the military could indicate an absence of clear control over security services. As such, greater military presence in politics increases the price of terror and may in fact reduce its total amount.

The remaining PDPs in Figures 7 through 11 depict the impacts of health expenditures, nonparametric time trend, population, executive years in office, and fuel exports, respectively. Consistent with other findings, we find that health expenditure (a quasi-public good) reduces terror risk, while population, executive tenure, and fuel exports all increase the risk.

The PDP's reported here establish just how the most important predictors of terrorism affect terrorist attacks. Notably these predictors do not have linear effects. Good theoretical explanations of terrorism must therefore explain these non-linearities. We have lessons for the policy maker too. It is important for the policy-maker to realize that small increases in health care expenditure reduce the number of terror attacks dramatically – thus initial increases in health care expenditures (and other public goods to the extent health care can be viewed as a proxy for public goods) have the biggest bang for the buck in terms of reducing terrorism.

The PDPs also demonstrate the possibility of *threshold* effects in the impacts of the variables. One or two assassinations in the preceding years may dramatically increase future terror risk, whereas more than that will not have much impact; variation in militarization at moderate levels may not have an impact on terror risk, whereas a very pervasive military state becomes less stable (and skews the payoffs of different forms of violence towards terror); increasing health expenditures from *very* low rates may significantly reduce terror, while marginal increases above moderate levels – while potentially beneficial for other reasons – do not necessarily buy more stability. These equilibrium shifts may suggest that game theoretic strategic models may be better at explaining terrorism over straightforward neo-classical models.

As a final cautionary note, we observe that the time trend shows that the long-run downward trend in terror risk that began in the 1970s seems to have abruptly ended around 2008-2009. This highlights the renewed importance of understanding terrorism, both in terms of predicting it better, and in terms of measuring its root causes. This work has sought to shed some light on which of the various factors might be most salient in gaining that understanding.

4. Conclusion.

Our methodological approach helps identify the variables that best predict terrorism. We suggest that any causal explanation of terrorism should start here. A good causal explanation of terrorism should center on variables that predict well. This approach also helps differentiate between competing explanations for why terrorism happens in a way that is not sensitive to model selection or assumptions about underlying distributions of variables – both problems in the current econometric literature on terrorism. At the same time, we suggest that predictive error is augments p-values to get a sense of how much a researcher or a policymaker should rely on a set of results for precisely the problems in the empirical literature noted above. Moreover, we notice that predictors of terrorism have non-linear effect. Econometric point estimates do not capture these non-linearities. Moreover, these non-linearities hint at complex effects that the current crop of theoretical models of terrorism do not capture well. Thus, our methodological approach suggests future terrorism research ought to center on explaining complex non-linear effects of certain kinds of variables (assassinations for example) over others (democracy and economic institutions). Last, but not the least, we suggest that policy makers can use our methodological approach to understand (a) how to predict terror attacks, and (b) what approaches are most likely to deter terror attacks. Such knowledge can help better allocate scarce resources.

Works Cited

Banks, A. S., & Wilson, K. A. (2015). *Cross-National Time-Series Data Archive.* Jerusalem, Israel: Databanks International. Retrieved from http://www.databanksinternational.com

Basuchoudhary, A., & Shughart II, W. F. (2010). On Ethnic Conflict and the Origins of Transnational Terrorism. *Defence and Peace Economics*, *21*(1), 65-87.

Crain, N. V., & Crain, W. M. (2006). Terrorized Economies. *Public Choice*, *128*(1), 317-349.

Cruz, C., Keefer, P., & Scartascini, C. (2016). *Database of Political Institutions Codebook, 2015 Update (DPI2015).* Washington, DC: Inter-American Development Bank.

Enders, W., & Sandler, T. (2006). Distribution of Transnational Terrorism among Countries by Income Class and Geography after 9/11. *International Studies Quarterly*, *50*(2), 367-393.

Faria, J. R., & Arce, D. G. (2005). Terror Support and Recruitment. *Defence and Peace Economics, 16*(4), 263-273.

Fortna, V. P. (2015). Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes. *International Organization*, *69*(3), 519-556.

Frey, B. S., & Luechinger, S. (2003). How to Fight Terrorism: Alternatives to Deterrence. *Defence and Peace Economics*, *14*(4), 237-249.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining.* Cambridge: MIT Press.

Iqbal, Z., & Zorn , C. (2008, March 6). The political consequences of assasination. *Journal of Conflict Resolution*, *52*(3), 385-400.

Laqueur, W. (1999). *The New Terrorism: Fanaticism and the Arms of Mass Destruction.* Oxford: Oxford University Press.

Marshall, M. G., Gurr, T. R., & Jaggers, K. (2015). *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2015.* Vienna, VA: Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/inscrdata.html

Murphy, K. P. (2012). *Machine Learning: A Probabalistic Perspective.* Cambridge, MA: MIT Press.

Pape, R. (2005). *Dying to Win: The Strategic Logic of Suicide Terrorism.* New York: Random House.

PRS Group. (2015). *International Country Risk Guide.* East Syracuse, NY: PRS Group. Retrieved from http://epub.prsgroup.com/products/international-country-risk-guide-icrg

Ripley, B., & Venables, W. (2016). *R Package 'nnet' Version 7.3-12.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://cran.r-project.org/web/packages/nnet/nnet.pdf

Sandler, T. (2014). The Analytical Study of Terrorism: Taking Stock. *Journal of Peace Research, 51*(2), 257-271.

Sepan, A., & Linz, J. J. (2011). Comparative Perspectives on Inequality and the Quality of Democracy in the United States. *Comparative Perspectives on Inequality and the Quality of Democracy, 9*(4), 841-856.

Shughart II, W. F. (2002). September 11, 2001. *Public Choice, 112*(1-2), 1-8.

Shughart II, W. F. (2006). An Analytical History of Terrorism, 1945-2000. *Public Choice, 128*(1-2), 7-39.

Solt, F. (2014). The Standardized World Income Inequality Database. *Version 5.0.* Working Paper.

Ward, M., Greenhill, B. D., & Bakke, K. M. (2010). The Perils of Policy by p-value: Predicting Civil Conflict. *Journal of Peace Research, 47*(4), 363-375.

Tables

Table 1. Variables and descriptive statistics

Variable	Source	Obs	Mean	Std. Dev.	Min.	Max.
Terror Attacks	GTD	6,411	86.88	375.81	0	10,701
Assassinations	CNTS	5,318	0.21	0.84	0.00	18.50
Cabinet Changes	CNTS	5,310	0.44	0.37	0.00	3.50
Demonstrations	CNTS	5,318	0.52	1.15	0.00	14.00
Effectiveness of Leg.	CNTS	5,297	1.74	0.94	0.00	3.00
Executive Changes	CNTS	5,310	0.19	0.28	0.00	3.00
Government Crises	CNTS	5,318	0.13	0.27	0.00	2.67
Guerrilla Warfare	CNTS	5,318	0.12	0.32	0.00	2.60
Purges	CNTS	5,318	0.03	0.13	0.00	2.50
Riots	CNTS	5,318	0.31	1.05	0.00	18.20
Strikes	CNTS	5,318	0.12	0.34	0.00	3.40
Changes in Veto Players	DPI	4,838	0.12	0.15	0.00	1.00
Checks on Power	DPI	4,831	2.52	1.60	1.00	17.00
Exec. Electoral Comp.	DPI	4,850	5.15	2.08	1.00	7.00
Executive Years in Office	DPI	4,859	7.93	7.68	1.00	45.00
Electoral Fraud	DPI	4,214	0.14	0.32	0.00	1.00
Government Frac	DPI	4,428	0.19	0.25	0.00	1.00
Government Herfindahl	DPI	4,428	0.82	0.25	0.02	1.00
Government Polarization	DPI	4,673	0.36	0.69	0.00	2.00
Legislative Frac.	DPI	4,419	0.46	0.30	0.00	1.00
Leg. Electoral Comp.	DPI	4,855	5.41	2.00	1.00	7.00
Military Executive	DPI	4,856	0.21	0.39	0.00	1.00
Opposition Frac	DPI	3,362	0.45	0.27	0.00	1.00
Plurality Voting	DPI	3,877	0.68	0.46	0.00	1.00
Proportional Rep.	DPI	3,474	0.58	0.49	0.00	1.00
Bureaucratic Quality	ICRG	3,376	2.11	1.19	0.00	4.00
Corruption	ICRG	3,376	3.08	1.35	0.00	6.00
Democratic Accountability	ICRG	3,376	3.64	1.62	0.00	6.00
Ethnic Tensions	ICRG	3,376	3.91	1.44	0.00	6.00
External Conflict	ICRG	3,376	9.48	2.22	0.00	12.00
Government Stability	ICRG	3,376	7.45	2.10	1.00	11.50
Internal Conflict	ICRG	3,376	8.61	2.62	0.03	12.00
Investment Profile	ICRG	3,376	6.94	2.34	0.08	12.00
Law and Order	ICRG	3,376	3.60	1.48	0.25	6.00
Military in Politics	ICRG	3,376	3.66	1.80	0.00	6.00
Religious Tensions	ICRG	3,376	4.54	1.35	0.00	6.00
Polity2	Polity IV	4,520	1.16	7.26	-10.00	10.00

Table 1. Variables and descriptive statistics (continued)

Variable	Source	Obs	Mean	Std. Dev.	Min	Max
Regime Durability	Polity IV	4,569	23.99	28.73	0.00	198.00
Ethnic Fractionalization	Reynal-Querol	4,749	0.45	0.28	0.01	0.96
Religious Fractionalization	Reynal-Querol	4,749	0.28	0.23	0.00	0.78
Income Inequality (Gini)	SWIID	3,350	38.52	9.87	16.49	69.35
Area	WDI	6,110	682,865	1,717,163	2	16,400,000
Off. Aid & Dev. Assistance	WDI	4,045	0.08	0.11	-0.01	0.76
Arms Exports	WDI	1,703	0.01	0.08	0.00	1.50
Arms Imports	WDI	3,976	0.04	0.12	0.00	3.32
Education Spending	WDI	3,436	4.45	2.32	0.59	44.30
Foreign Direct Investment	WDI	4,602	2.80	4.72	-32.30	72.50
Female Labor Force Part.	WDI	3,293	50.12	17.55	9.20	90.80
Fuel Exports	WDI	3,875	16.82	28.33	0.00	100.00
GDP per Capita	WDI	4,807	9,560.35	16,016.19	65.64	141,000.00
Government Consumption	WDI	4,538	16.47	6.87	3.37	84.50
Health Spending	WDI	2,647	3.48	2.21	0.01	18.36
Immigrant Stock	WDI	4,975	8.07	13.75	0.03	86.80
Infant Mortality	WDI	5,103	48.11	40.72	2.18	174.00
Inflation	WDI	4,168	32.94	254.00	-17.60	6522.40
Life Expectancy	WDI	5,074	64.79	10.58	24.30	82.50
Literacy Rate	WDI	1,549	73.42	23.01	10.90	100.00
Military Expenditures	WDI	2,995	2.74	3.03	0.09	48.60
Military Personnel	WDI	3,092	1.88	2.23	0.06	35.80
Population	WDI	5,190	30.94	116.87	8.82	1,316.00
Population Growth	WDI	5,190	1.80	1.44	-4.84	15.50
Portfolio Investment	WDI	4,000	0.01	0.16	-0.02	4.88
Primary Enrollment	WDI	4,763	97.05	22.35	15.80	208.00
Secondary Enrollment	WDI	4,407	60.84	33.35	2.13	155.60
Social Contributions	WDI	1,203	17.11	15.02	0.00	59.97
Telephones	WDI	5,127	14.70	18.58	0.01	103.42
Tertiary Enrollment	WDI	4,135	18.62	19.36	0.00	99.20
Unemployment	WDI	3,007	9.03	6.78	0.20	59.50
Urban Population	WDI	5,190	50.33	24.51	4.18	100.00
Youth Dependency	WDI	5,000	62.07	23.94	19.44	114.40

Table 2. MSEs for the various learning models

	Learning Sample		Test Sample	
	MSE	% Decrease	MSE	% Decrease
OLS Regression	107,708.05	25.71%	98,119.17	26.12%
Poisson Regression	151,539.85	-4.52%	139,385.78	-4.96%
Neural Network	144,695.12	0.20%	132,389.28	0.31%
Regression Tree	52,038.41	64.11%	80,182.62	39.62%
Boosting Predictor	141,677.19	2.28%	129,790.58	2.27%
Bagging Predictor	59,866.71	58.71%	40,202.12	69.73%
Random Forest	54,271.19	62.57%	38,504.85	71.01%
Average of All Predictors	74,564.39	48.57%	76,391.30	42.48%
Total MSE	144,987.24		132,802.82	

Table 3. Variable importance rankings

Variable	Tree	Bagging	Boosting	Forest	Average
Assassinations	7.618	24.930	62.966	12.388	14.979
Guerrilla War	2.677	10.735	30.698	9.436	7.616
Military Personnel	15.482	4.166	0.000	2.555	7.401
Religious Frac	12.386	4.761	0.000	3.218	6.788
Military Politics	12.682	1.913	1.082	3.529	6.042
Health Spending	3.765	4.390	2.499	3.548	3.901
Year	1.882	5.704	0.000	3.888	3.825
Population	0.947	3.568	0.394	5.562	3.359
Exec Yrs in Office	6.441	1.940	0.000	1.315	3.232
Fuel Exports	6.193	1.455	0.000	1.227	2.958
Dem Accountability	5.222	1.243	0.000	1.411	2.625
Effectiveness of Leg	0.000	3.104	0.000	3.041	2.048
Aid & Assistance	2.528	0.973	0.000	1.826	1.775
Gini	0.981	2.106	0.000	2.083	1.723
Tertiary Enrollment	2.053	0.752	0.000	2.023	1.609
Female LFPR	0.000	1.226	2.361	3.213	1.480
Portfolio Investment	0.000	2.695	0.000	1.600	1.432
Area	1.858	1.194	0.000	0.837	1.297
Arms Imports	1.425	1.402	0.000	1.009	1.279
Strikes	0.662	1.369	0.000	1.711	1.247
Ethnic Tension	0.733	0.467	0.000	2.409	1.203
Checks	1.702	1.248	0.000	0.615	1.188
Internal Conflict	0.969	0.125	0.000	2.257	1.117
Telephones	1.882	0.435	0.000	0.697	1.005
Law Order	1.322	0.560	0.000	0.966	0.950
GDP pc	0.235	0.842	0.000	1.736	0.938
Urban Population	1.710	0.404	0.000	0.645	0.920
Ethnic Frac	0.469	1.336	0.000	0.885	0.897
Polity 2	0.000	1.355	0.000	1.244	0.866
Investment Prof	0.321	1.019	0.000	1.169	0.836
Legislative Frac	1.425	0.532	0.000	0.549	0.835
Riots	0.307	0.729	0.000	1.382	0.806
Primary Enrollment	1.425	0.170	0.000	0.687	0.761
Arms Exports	0.000	1.233	0.000	0.950	0.728
Demonstrations	0.179	0.464	0.000	1.511	0.718
Unemployment	0.000	0.813	0.000	1.310	0.708
Religious Tension	0.000	0.300	0.000	1.601	0.634
Infant Mortality	0.000	0.773	0.000	1.046	0.606

Table 3. Variable importance rankings

Variable	Tree	Bagging	Boosting	Forest	Average
Secondary Enrollment	0.000	0.848	0.000	0.765	0.537
Immigrant Stock	0.016	0.672	0.000	0.816	0.501
Reg Durability	0.248	0.547	0.000	0.609	0.468
Gov Consumption	0.000	0.369	0.000	0.824	0.398
Gov Stability	0.618	0.089	0.000	0.441	0.383
Corruption	0.075	0.472	0.000	0.584	0.377
Life Expectancy	0.000	0.242	0.000	0.889	0.377
Youth Dependency	0.000	0.286	0.000	0.842	0.376
FDI	0.000	0.372	0.000	0.719	0.363
Fraud	0.346	0.106	0.000	0.421	0.291
Opposition Frac	0.000	0.304	0.000	0.560	0.288
Inflation	0.207	0.326	0.000	0.300	0.278
External Conflict	0.259	0.140	0.000	0.428	0.276
Bureaucratic Qual	0.000	0.277	0.000	0.501	0.259
Leg. Elec. Comp.	0.248	0.192	0.000	0.295	0.245
Exec. Elec. Comp.	0.000	0.262	0.000	0.465	0.242
Literacy Rate	0.000	0.143	0.000	0.574	0.239
Population Growth	0.000	0.394	0.000	0.302	0.232
Proportional Rep	0.000	0.368	0.000	0.326	0.231
Social Contributions	0.000	0.173	0.000	0.515	0.229
Military Expend	0.167	0.212	0.000	0.277	0.219
Purges	0.331	0.051	0.000	0.179	0.187
Education Spending	0.000	0.243	0.000	0.304	0.182
Military Exec	0.000	0.085	0.000	0.316	0.134
Government Herfindahl	0.000	0.082	0.000	0.235	0.106
PluralityVoting	0.000	0.093	0.000	0.196	0.096
Gov Polarization	0.000	0.063	0.000	0.198	0.087
Government Frac	0.000	0.093	0.000	0.087	0.060
Cabinet Changes	0.000	0.028	0.000	0.085	0.038
Changes in Vetoes	0.000	0.065	0.000	0.018	0.028
Executive Changes	0.000	0.020	0.000	0.044	0.021
Government Crises	0.000	-0.050	0.000	-0.191	-0.080

Figures

Figure 1. ANN Diagram Example

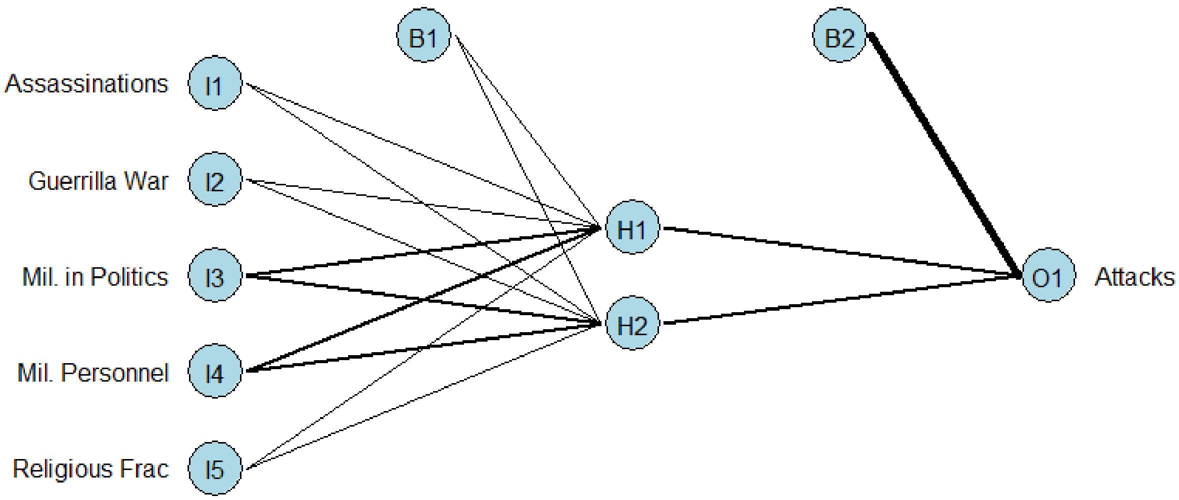


Figure 2. Partial dependence plot: assassinations

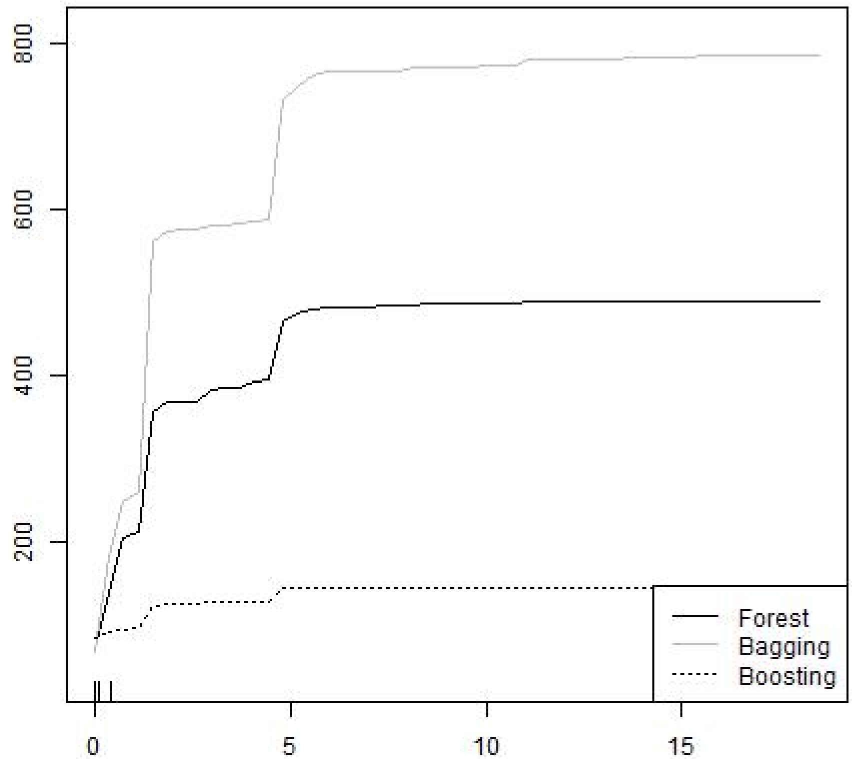


Figure 3. Partial dependence plot: guerrilla warfare

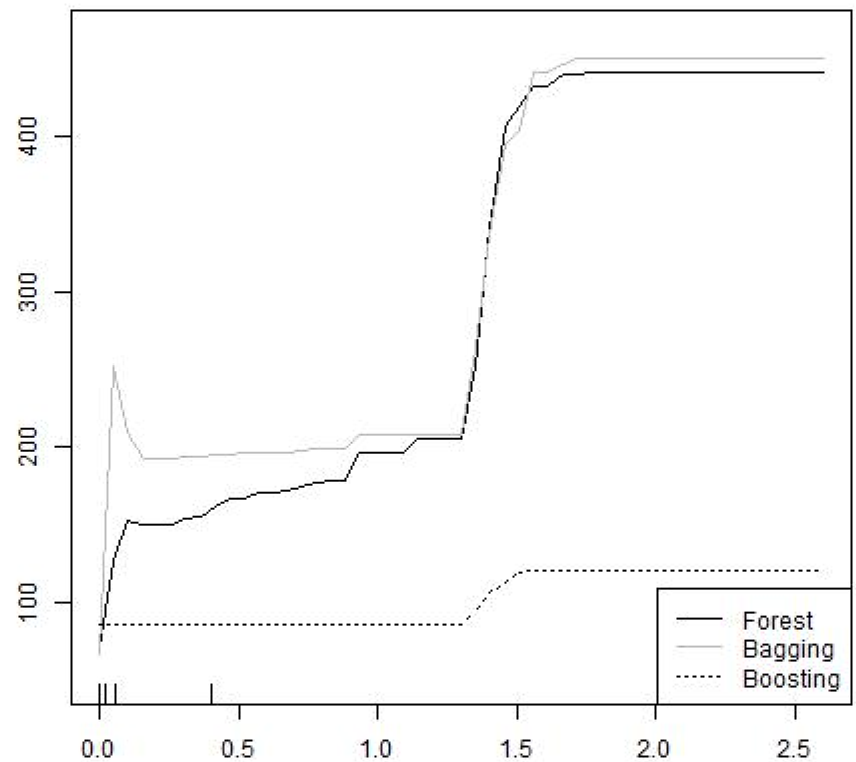


Figure 4. Partial dependence plot: military personnel

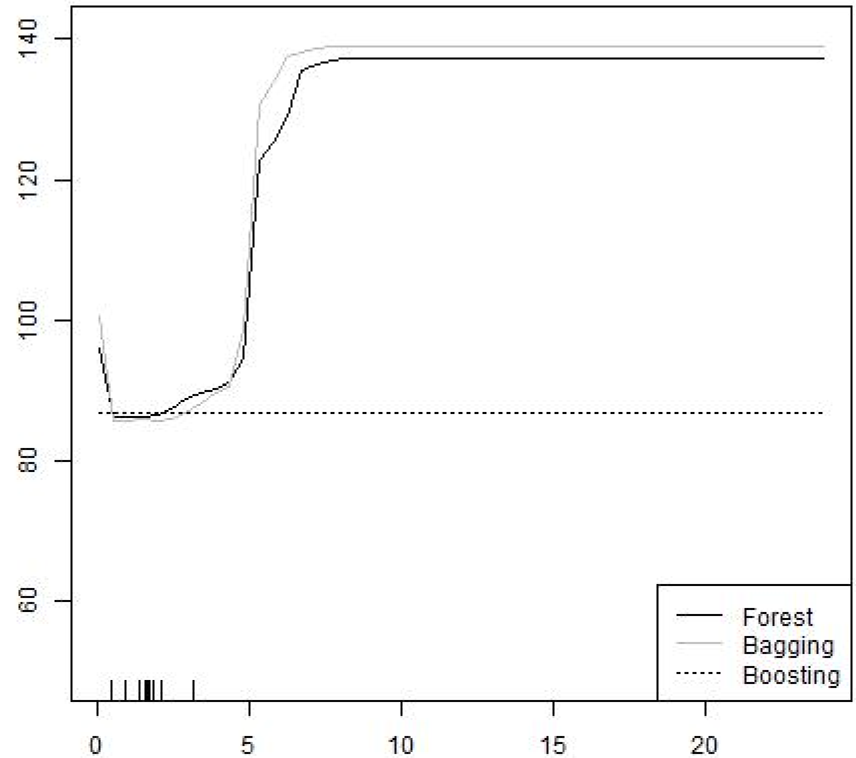


Figure 5. Partial dependence plot: religious fractionalization

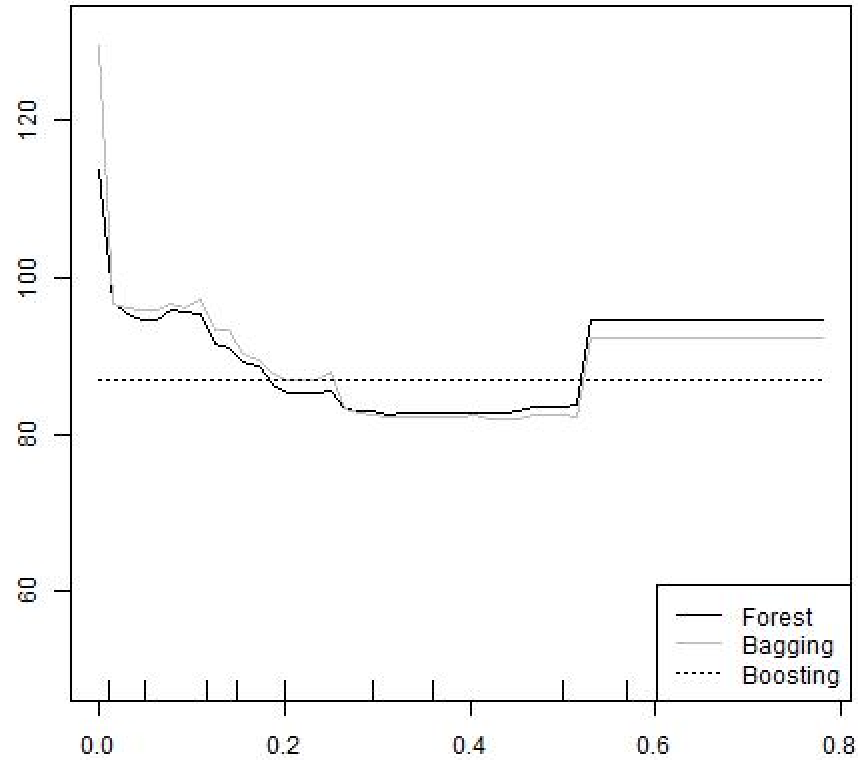
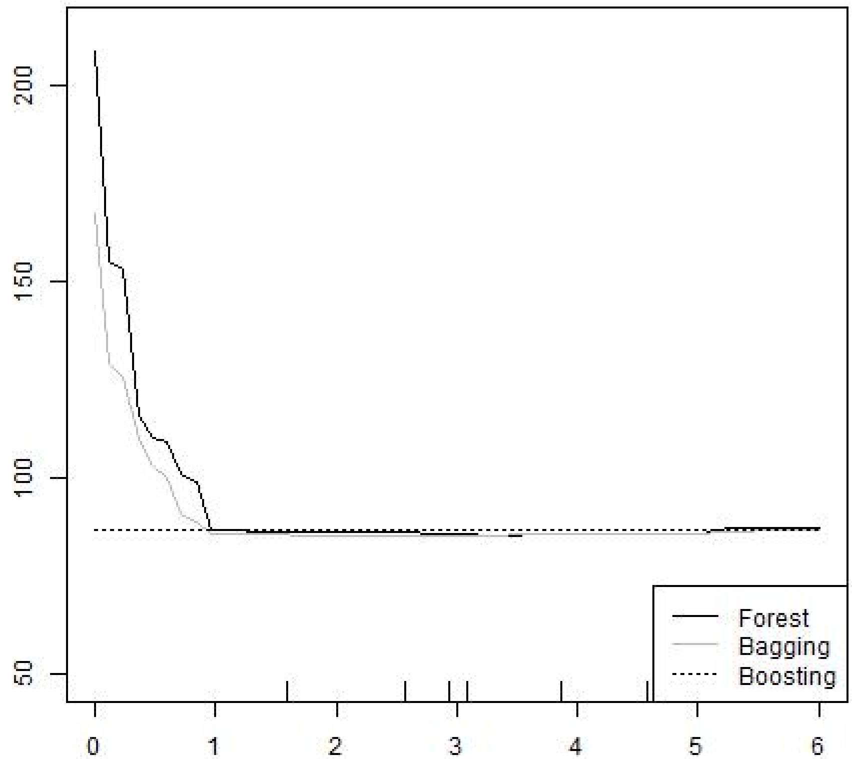


Figure 6. Partial dependence plot: military in politics



the

Figure 7. Partial dependence plot: health expenditure

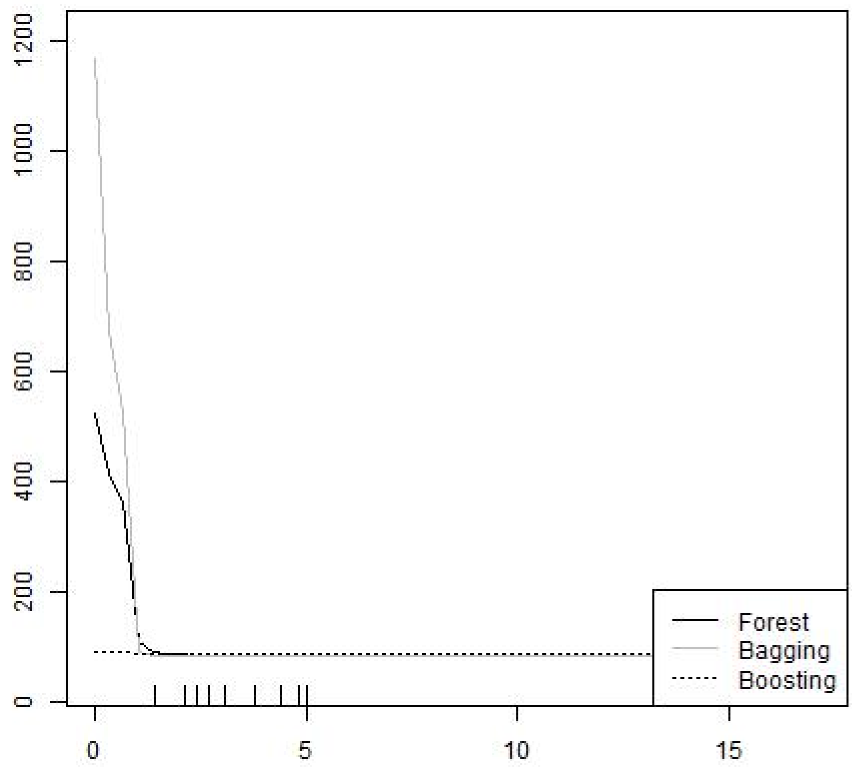


Figure 8. Partial dependence plot: year

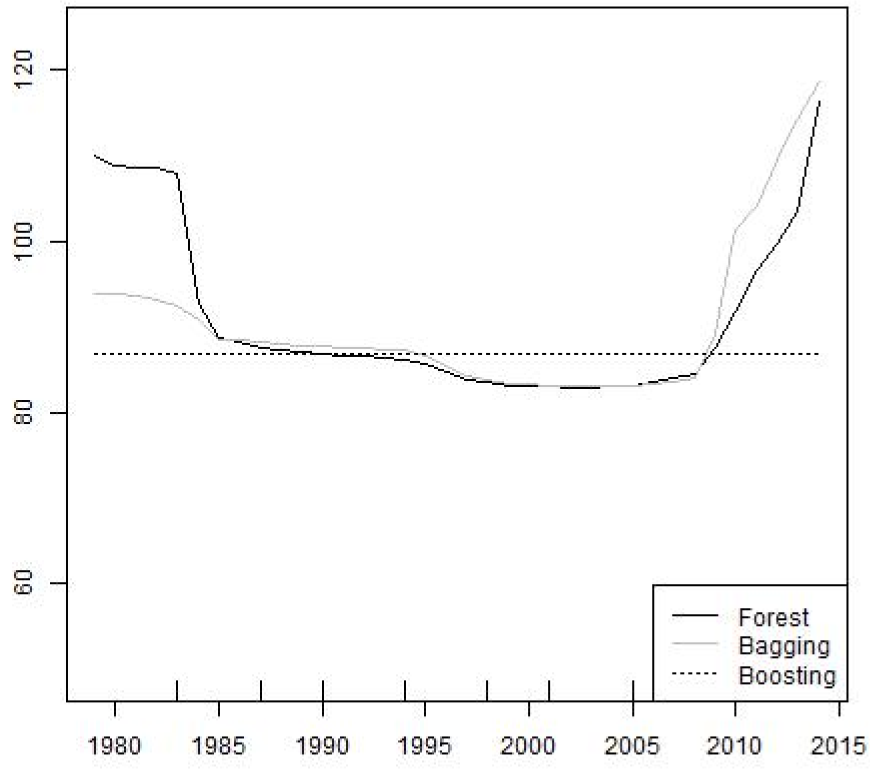


Figure 9. Partial dependence plot: population

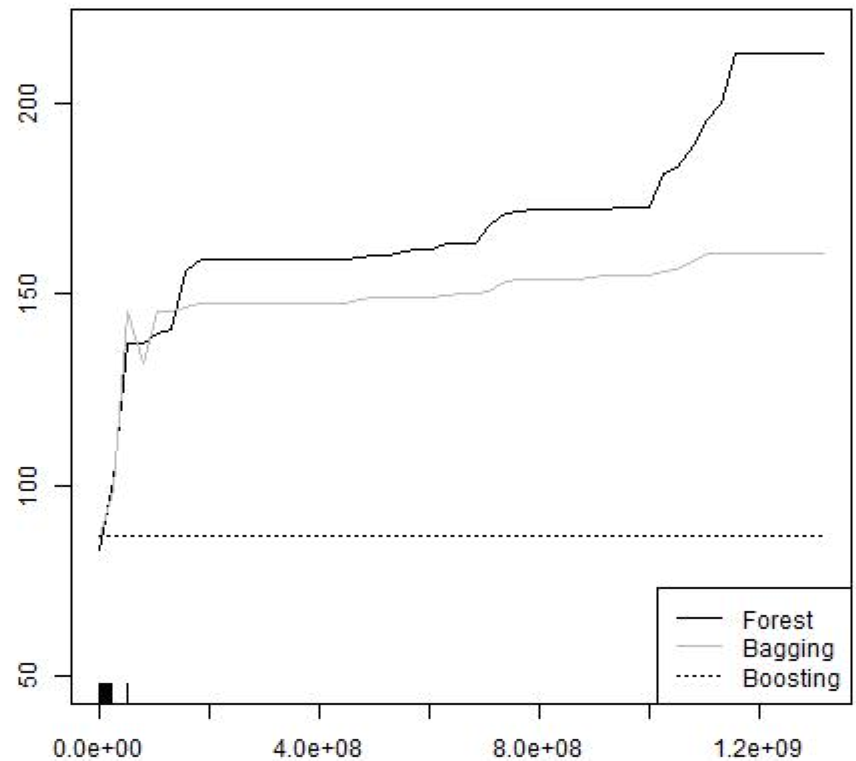


Figure 10. Partial dependence plot: executive years in office

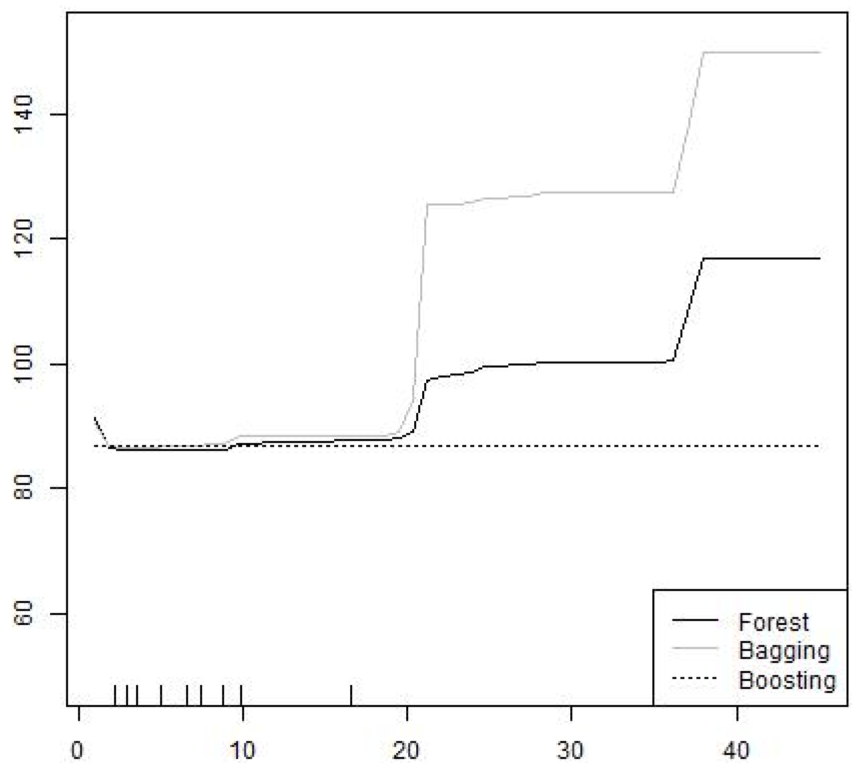


Figure 11. Partial dependence plot: fuel exports