Assessing Proliferation Risk: Using Data Analytics to Evaluate Emerging Nuclear Threats

Jeffrey M. Kaplow
Assistant Professor, Department of Government
William & Mary
jkaplow@wm.edu

Abstract

The nuclear proliferation landscape has changed. Twenty years ago, the countries most likely to seek nuclear weapons were easily identified, and the factors that could push them into nuclear pursuit were well understood. Today, widespread latent capability has made the list of countries with the capacity for weapons pursuit much longer, and the potential triggers for nuclear proliferation have grown more complex. To deal effectively with today's nuclear proliferation threats, intelligence and policy analysts need versatile tools to evaluate proliferation risk and focus limited resources and analytic capacity. This article employs the tools of data science to develop a new measure of the likelihood that a country will seek nuclear weapons under particular circumstances: the proliferation risk score. Proliferation risk scores are both scalable and flexible. They can be updated with new data and expanded to cover new factors that might lead to nuclear pursuit, and they can be used in a variety of what-if scenarios to identify factors that represent the greatest risk for a specific country in a specific global situation. They can be helpful to analysts as an adjunct to traditional analysis and as a starting point for structured analytic approaches.

Keywords: nuclear proliferation, risk assessment, risk modeling

Introduction

Assessing the risk of nuclear proliferation has posed an enduring challenge for policymakers and analysts. The risk assessment process builds a deeper and more systematic understanding of the likelihood of proliferation across many potential nuclear aspirants and under a variety of conditions. Assessing proliferation risk is an essential first step in avoiding surprise and designing effective policies to counter the spread of nuclear weapons. The risk assessment process contributes to proliferation research and policymaking in four ways:

1. Proliferation risk assessment allows for triaging analytic and research effort, directing resources to focus on the most urgent among many recognized and "over the horizon" threats.
2. Proliferation risk assessment contributes to early warning of proliferation activity, providing time for a policy response.
3. Proliferation risk assessment allows analysts to evaluate the effect of future policy shifts or changes in the international security environment on the likelihood that particular states will seek weapons.
4. The process of proliferation risk assessment may surface new indicators of proliferation intent that can be applied more broadly to analysis of states of concern.

Proliferation risk assessment is a complex analytic task. Done well, it requires integrating technical and political expertise, working across traditional functional and regional boundaries, and merging qualitative knowledge and data-driven empirical findings. Researchers in and out of government frequently attempt to evaluate the risk of proliferation in a particular country. These studies are valuable, but usually examine only the specifics of the case at hand. They often do not attempt to apply data and findings from multiple historical cases or from other contemporary cases, and they are usually limited to extrapolating from the existing international security environment, domestic conditions, and state of international institutions. That is, it is difficult for these studies to consider how changes in the international context would affect the risk of proliferation.

This article leverages new data and analytic techniques to answer several key questions about the risk of nuclear proliferation. Which states are most likely to seek nuclear weapons? Under what conditions is proliferation most likely? What policies will be most effective in stopping proliferation under particular conditions? What indicators can help analysts and researchers distinguish between nuclear pursuit and the mere development of nuclear energy programs?

Predictive models of nuclear proliferation can help to answer these questions. Building on recent advances in studies of proliferation and nuclear-related data collection efforts, this project applies statistical learning techniques to create flexible quantitative models of proliferation. The output of these models—the proliferation risk score—tells us not only which states seem most likely to seek nuclear weapons under today's international security conditions and counter-proliferation policies, but also allow us to see the effect that varying these circumstances can have on proliferation risk for states of concern. Risk scores thus facilitate a more data-driven form of scenario-based research and “what-if” style analysis.[1] Risk scores also shed light on the factors that have best distinguished between proliferating states and states that have shown nuclear restraint. These factors are likely to be effective indictors of future proliferation intent.

Existing Literature

A rich academic and policy literature has identified a number of important drivers of the decisions of states to seek nuclear weapons. These drivers of proliferation and nuclear restraint can be divided into three broad categories: nuclear motives (demand-side factors), nuclear capabilities (supply-side factors), and international institutions.[2] Analyses of nuclear motives generally focus on a state’s concern about its own security.[3] Studies have found links between nuclear weapons pursuit and a state’s conflict behavior,[4] the proliferation decisions of neighbors or rivals,[5] the presence of an alliance or security guarantee from a nuclear-armed patron,[6] and the type of regime or leader.[7] Nuclear capability, for its part, has been associated with weapons programs both through domestic capacity and through assistance from other states.[8] Finally, a number of analysts see the Treaty on the Non-Proliferation of Nuclear Weapons as effective in

---

[1] Huss 1988; Golfarelli, Rizzi, and Proli 2006.

[2] See Sagan 2011 for a thorough review.

[3] Sagan 1996.

[4] Jo and Gartzke 2007; Fuhrmann 2009.

[5] Fuhrmann 2009; Miller 2014a.

[6] Bleek and Lorber 2014; Gerzhoy 2015; Reiter 2014.

[7] Fuhrmann and Horowitz 2015; Hymans 2006; Way and Weeks 2014.

[8] On domestic capacity, see Singh and Way 2004; Jo and Gartzke 2007. On foreign assistance, see Brown and Kaplow 2014; Fuhrmann 2009; Kroenig 2009.

constraining state behavior; it has been hailed as one of the most successful security treaties in history.[9]

The last decade has seen a renaissance in empirical studies of nuclear proliferation, particularly those making use of quantitative methodologies. This research has introduced a number of useful datasets to test hypotheses about nuclear proliferation, including data on nuclear pursuit,[10] domestic capacity and nuclear latency,[11] nuclear assistance,[12] nuclear security guarantees and nuclear deployments,[13] the presence of other WMD programs,[14] membership in the nuclear nonproliferation regime,[15] nuclear sanctions,[16] and military strikes against nuclear targets.[17]

Taken as a whole, this body of literature is well positioned to help researchers and analysts better understand the risk of proliferation in particular cases. But each of these studies focuses on a relatively narrow hypothesis about the drivers of proliferation; little existing work attempts to synthesize these data and findings. Proliferation risk scores help to fill this gap, leveraging existing data to better understand proliferation risk in cases of interest.

---

[9] Fuhrmann and Lupu 2016; Coe and Vaynman 2015; Cirincione 2008.

[10] Bleek and Lorber 2014; Jo and Gartzke 2007; Singh and Way 2004.

[11] Jo and Gartzke 2007; Fuhrmann and Tkach 2015.

[12] Brown and Kaplow 2014; Fuhrmann 2009; Kroenig 2009.

[13] Bleek and Lorber 2014; Fuhrmann and Sechser 2014.

[14] Horowitz and Narang 2014.

[15] Carcelli et al. 2014.

[16] Miller 2014b; Reynolds and Wan 2012.

[17] Fuhrmann and Kreps 2010.

Using Proliferation Risk Scores

Proliferation risk scores and other forms of quantitative analysis cannot substitute for expert analysis. The issues associated with nuclear pursuit—both political and technical—are too complex to be accurately represented in the data fed into a quantitative analysis. The approach taken here, however, can still be useful as an aid to expert analysis in at least three ways: interrogating existing cases of concern, identifying new over-the-horizon threats, and facilitating a more data-driven form of scenario analysis and other structured analytic techniques.

First, proliferation risk scores allow analysts to interrogate a particular case that has already been identified as a proliferation risk. This is the approach shown in the scenarios above. Analysts working on a particular country scenario can use the model to examine some of the underlying dynamics of the case and pose structured hypotheticals related to the country of concern. For example, if analysts wish to investigate the proliferation dynamics for a country like Estonia, for example, they can begin with a set of inputs suggested by the most recent data. Estonia's projected economic output in 2021, combined with its recent involvement in international disputes, the status of NATO and the nonproliferation regime, and nuclear development in the country all factor into a baseline assessment of proliferation risk. From there, analysts can examine how changes in these variables affect Estonia's proliferation risk scores. Proliferation risk scores can reveal which factors might act as a proliferation trigger, pushing Estonia to seek nuclear weapons, and which policy levers are most likely to encourage nuclear restraint.

A particular challenge in evaluating future threats is that there are often many potential targets that seem equally unlikely, at least at first glance. Because resources for conducting long-term analysis are limited—analytic capacity is rightly focused on the most likely near-term threats—intelligence and policy organizations must make difficult choices about where to focus their attention. Proliferation risk scores can help analysts identify the over-the-horizon proliferation risks that merit additional expert analysis. Analysts might begin with a particular future scenario, such as the weakening of alliance ties between the United States and its global partners, or a downturn in international trade. From there, analysts can generate a list of states of particular concern given this change in the global strategic environment. This approach is likely to surface new countries of concern that may not be current targets of analytic attention. Once identified, these new candidate countries could be subjected to additional analysis by adjusting the dynamics of the future scenario, as described above, or by using more traditional qualitative analytic techniques.

Finally, proliferation risk scores facilitate a more data-driven approach to qualitative structured analytic techniques. Analysts engaged in scenario planning efforts or what-if analysis frequently must imagine possible alternative futures.[18] These scenarios are designed to be plausible, in the sense that they are consistent with existing trends, even if they are not necessarily likely outcomes. Proliferation risk scores can help reduce some of the speculation inherent in these efforts by linking data-based outcomes

[18] See, for example, Heuer, Jr. and Pherson 2020.

to assumptions about future trends. Traditional scenario analysis or alternative futures analysis invites analysts to imagine several potential drivers of future outcomes. These drivers can be used to calculate new proliferation risk scores for particular scenarios. For example, if a driver of one alternative future is the global economy, analysts can examine risk scores for both high-GDP and low-GDP scenarios, so that discussion about particular countries of concern can be guided by the available data.

Materials & Methods

Countries vary in their risk of nuclear proliferation. Some countries strongly desire weapons but lack the resources to develop them, while others have the means but not the motive. Considering even this simple framework—nuclear capability and nuclear willingness—is enough to give a pretty good sense of which countries should worry analysts when it comes to proliferation. In general, countries of proliferation concern tend to be relatively rich nations that fear for their own security, but there are some prominent exceptions. Pakistan, Libya, and North Korea, weapons-pursuers all, seemed by most conventional measures of national wealth to lack the resources for a nuclear weapons program. Romania, Yugoslavia, and South Africa sought nuclear weapons despite seeming fairly secure even without them (although their respective leaders did not see it that way). A simple understanding of the drivers of proliferation ends up missing some important cases of nuclear pursuit.

To build a measure of proliferation risk, this project uses the output of a quantitative model of nuclear proliferation that incorporates multiple variables, drawing from the extensive literature on the drivers of nuclear weapons programs. Academic studies have identified significant associations between the likelihood of proliferation and a state's conflict behavior,[19] its economic strength,[20] nuclear capability and the diffusion of nuclear technology,[21] the proliferation decisions of neighbors or rivals,[22] the presence of an alliance or security guarantee from a nuclear-armed patron,[23] the type of regime or leader,[24] receipt of bilateral or multilateral nuclear assistance,[25] and membership in the Treaty on the Non-Proliferation of Nuclear Weapons (NPT).[26]

Building on the findings from these studies, this project constructs a quantitative model of proliferation, in which the dependent variable is whether or not a country has a nuclear weapons program in a given year.[27] Explanatory variables are derived from the studies listed above, with a few additions, and are summarized in Table 1. Like most existing quantitative work in nuclear proliferation, the data are structured as a pooled time series with the country-year as the unit of analysis. Each observation in the data

---

[19] Brown and Kaplow 2014; Fuhrmann 2009; Jo and Gartzke 2007; Singh and Way 2004.

[20] Fuhrmann 2009; Singh and Way 2004.

[21] Bleek and Lorber 2014; Jo and Gartzke 2007.

[22] Fuhrmann 2009; Miller 2014a.

[23] Bleek and Lorber 2014; Gerzhoy 2015; Reiter 2014.

[24] Fuhrmann and Horowitz 2015; Hymans 2006; Singh and Way 2004; Way and Weeks 2014.

[25] Brown and Kaplow 2014; Fuhrmann 2009; Kroenig 2009.

[26] Bleek and Lorber 2014; Fuhrmann 2009; Jo and Gartzke 2007.

[27] There is, of course, some disagreement among analysts over when countries started and ended nuclear efforts. This analysis uses data updated from Jo and Gartzke 2007.

represents an individual state in a given year. The data run from 1939—when countries first began exploring nuclear weapons as a serious possibility—to 2010.

Most quantitative studies of nuclear proliferation aim to identify a statistical association between a particular factor and a state's propensity to seek nuclear weapons. In service of this goal, scholars frequently conduct some form of regression analysis, and report whether variables of interest—representing their key causal factors—achieve statistical and perhaps substantive significance. These analyses usually make some attempt to account for alternative explanations for their results, often by controlling for confounding variables in a regression model. This mode of analysis is useful in understanding whether an individual factor affects the outcome of interest.

Table 1: Explanatory variables in a model of proliferation risk

Factor	**Measure**	**Data Source**
Economic capacity	Real GDP per capita	K. S. Gleditsch (2002)
Nuclear capacity	Lab-scale enrichment or reprocessing facility	Fuhrmann and Tkach (2015)
	Operational nuclear reactor	Kaplow (2020)
Nuclear diffusion	Log of years since 1938	
Nuclear ally	Defense pact with nuclear state	Gibler (2019)
Nuclear rival	Enduring rival with a nuclear weapons program	Goertz, Diehl, and Balas (2016)
Conflict behavior	Interstate armed conflicts over the last 5 years (moving average)	N. P. Gleditsch et al. (2002)
Regime type	Polity IV score	Marshall, Jaggers, and Gurr (2010)
Regime membership	Member of the NPT	Carcelli et al. (2014)
Nuclear assistance	Bilateral civilian nuclear cooperation agreement	Fuhrmann (2009)
	IAEA fuel cycle-related technical cooperation	Brown and Kaplow (2014)
Foreign policy affinity	Ideal point derived from UN General Assembly voting	Bailey, Strezhnev, and Voeten (2017)

The goal of statistical modeling in this project, however, is somewhat different—less concerned with the explanatory power of a given variable than with leveraging the predictive power of the model as a whole. Rather than focusing on the statistical significance of the variables themselves, the outcome of interest is a prediction, an assessment of proliferation risk. Given what is known about all of these factors, how likely is it that this individual case, a given country in a given year, is engaged in a nuclear weapons program?

To better capture proliferation risk, the predictive results of the model are evaluated out of sample. That is, the model is constructed using one set of data, and then the model's performance is evaluated using another set of data that has been reserved for this purpose. A focus on out-of-sample predictive validity has several advantages. First, it moves away from the questionable emphasis in quantitative analysis on statistical significance as a metric for a successful result. Second, testing the performance of models within the data sample—as one does, for example, in traditional regression models—is a kind of teaching to the test. This practice risks overfitting models and mistaking idiosyncrasies in the data for real-world trends. Finally, expressing results in terms of out-of-sample predictive validity captures a kind of substantive significance in the model, providing a better sense of how the model performs in the real world.

To make predictions out of sample, proliferation risk scores employ a leave-one-out cross-validation procedure of the kind commonly used in computer science and machine learning.[28] First, a particular country is excluded from the data. Then, the remaining dataset is used to construct the model.[29] Finally, that model is used to generate predictions of proliferation risk for the country that had been left out of the analysis. This procedure is repeated for each country, so that the predictions made are

[28] Arlot and Celisse 2010.

[29] The dataset excludes country-years that occur once a country has actually acquired nuclear weapons. These countries can no longer properly be considered to be "pursuing" nuclear weapons.

always out of sample—the country for whom nuclear pursuit is predicted is never considered when constructing the model in the first place.

As with all data analysis, there are complications. Not all variables are available for all country-years in the dataset. For example, Fuhrmann's data on nuclear cooperation agreements is not available after 2003, and states involved in an active conflict are often missing a Polity score—a measure of democracy—in that year.[30] One option for dealing with missing data would be to drop these observations or these variables from the analysis. Instead, the cross-validation process is repeated separately for each missingness pattern in the data. That is, the leave-one-out cross-validation procedure is first conducted for each country that has no missing data. A new dataset is then created that omits the variable on nuclear cooperation agreements and the leave-one-out cross-validation procedure is conducted for each country that now has no missing data. A new dataset is then created with a different pattern of missingness and the process repeated until predictions are generated for each country. The prediction actually used in the analysis is the one generated by the dataset with the fewest missing variables for that country. This process allows for best leveraging the data available, while still generating a prediction for each country-year in the dataset.[31]

Another complication stems from the fact that nuclear proliferation is a rare event—only about 7 percent of observations in the data correspond with a nuclear

[30] Fuhrmann 2009; Marshall, Jaggers, and Gurr 2010.

[31] For a review of related approaches to addressing missingness in the context of out-of-sample prediction, see Conroy et al. 2016; Fang et al. 2019; and García-Laencina, Sancho-Gómez, and Figueiras-Vidal 2010.

weapons program. Statistical models are constructed to best fit all of the data in the training sample. When that data is overwhelmingly an example of non-nuclear pursuit, the best-fitting model is likely to err in the direction of explaining those more prevalent cases. That is, the model selected, almost by definition, is designed to explain the more frequent case in the dataset. This issue of class imbalance is a familiar problem in the computer science literature.[32]

While there is no single solution to this problem, the issue can be mitigated by oversampling the rare class in the data or undersampling the prevalent class. Here, both approaches are used, adopting an algorithm known as SMOTE (synthetic minority oversampling technique).[33] This algorithm works by adjusting the training data for the models (the data used to construct the model). It adds to the number of cases of nuclear pursuit in the data, generating new, synthetic cases using a nearest neighbor method. It also reduces the number of non-pursuit cases in the data through systematic undersampling. The result is a more balanced sample of cases in the training data, facilitating better prediction of the rare event. Note that this procedure has *not* been used to adjust the out-of-sample data that is set aside and then used to evaluate predictive accuracy for the model.

Predictions are generated using a statistical learning model—the support vector machine (SVM).[34] An SVM represents data as points in multidimensional space,

[32] Sun, Wong, and Kamel 2009.

[33] Chawla et al. 2002.

[34] For a general discussion, see Steinwart and Christmann 2008.

developing a set of statistical rules that maximize the gap between points of one type (states that seek nuclear weapons) and points of another type (states that forgo weapons programs). Statistical learning approaches like SVMs are commonly used in computer science and statistics and have been increasingly employed in the social sciences. Statistical learning is particularly well suited to problems in which the relationships between variables are highly conditional, as they are likely to be in the case of nuclear proliferation. Because the pursuit of nuclear weapons is a rare event, even the strongest drivers of proliferation probably exert relatively little influence on proliferation decisions in the large majority of cases. But in states that are at high risk of proliferating—that is, in the cases analysts care most about—these factors may matter a great deal.[35] The linear regression models and their close relatives (such as logit and probit) that are used most often in quantitative analysis of nuclear proliferation are not flexible enough to capture the complex non-linear relationships that are likely to be present in these data.

What comes of all this is a set of predictions about the likelihood of proliferation in each case. The output of the model is a predicted probability—the proliferation risk score. This is a percentage chance assigned by the model that this country will pursue a nuclear weapon in this year. Predictions of over 50 percent are commonly considered to be a "yes" prediction; these are cases where the model guesses the country is pursuing a nuclear weapon.

[35] See Beck, King, and Zeng 2000 for an application of this argument to international conflict.

Evaluating Measures of Proliferation Risk

What should analysts look for in a measure of proliferation risk? Assessments of risk are a form of prediction. A good measure should correctly predict proliferation, in the sense that countries flagged by the measure as risky should be more likely to seek nuclear weapons, while countries the measure sees as low risk should not have a nuclear weapons program. A good measure of proliferation risk will provide reliable predictions, measured quantitatively, but will also have good face validity. That is, the measure's predictions will generally seem reasonable to those with substantive knowledge of nuclear proliferation.

A measure that predicts proliferation accurately is preferred, but overall accuracy is not a useful measure of success when rare events like proliferation are involved, because a model can achieve very high levels of overall accuracy without providing any leverage against the problem of interest. If a model always predicts that states will not proliferate, it will accurately predict nearly all of the observations in the data, but such an approach is not particularly useful as a means of understanding proliferation risk.

The data science literature on "class imbalance"—situations where one outcome (proliferation, in this case) is much rarer in the data than the other outcome (non-proliferation)—provides some guidance on predictive metrics.[36] While there is no single accepted solution to this problem, a common approach is to adopt metrics for accuracy that are more sensitive to the ability to assess proliferation itself, rather than just non-

[36] Kotsiantis, Kanellopoulos, and Pintelas 2006.

proliferation.[37] Two such metrics are used here. First, model accuracy is evaluated using the area under the Receiver Operating Characteristics (ROC) curve—abbreviated AUC for Area Under [the ROC] Curve.[38] On one axis of the ROC curve is the rate of false positives—the number of cases in which the model incorrectly predicted that a state would seek weapons, divided by the total number of cases of nuclear non-pursuit. On the other axis is the true-positive rate—the number of cases in which the model correctly predicted proliferation divided by the total number of cases of nuclear weapons programs. These rates are plotted against each other across a range of thresholds for positive and negative predictions. A perfect assessment of proliferation risk, one that correctly predicts all cases of both proliferation and nuclear restraint, would have an AUC of 1. Random guessing would yield an AUC of 0.5.

As a second metric for predictive success, models are evaluated using the F_1 score. This metric balances two elements of the assessment of proliferation risk. The first is positive predictive value, the share of "yes" predictions that turn out to be correct. The second is sensitivity, the share of real proliferation episodes that the model correctly identifies. The F_1 score is the harmonic mean of these two factors. F_1 scores closer to 1 indicate a greater level of predictive success, while scores closer to zero indicate more incorrect predictions of nuclear proliferation.

[37] For a review of common accuracy metrics in the context of class imbalance, see Luque et al. 2019.

[38] Swets 1988.

Results

By these quantitative metrics, the model performs extremely well. It successfully discriminates between cases of proliferation and non-proliferation, and it is particularly effective when considered by metrics that privilege positive predictions—cases of actual nuclear pursuit. Table 2 provides evaluation metrics for the model of proliferation risk. For comparison, the table also includes a qualitative measure of proliferation risk adapted from Coe and Vaynman, which uses declassified documents and other sources to enumerate all countries assessed as nuclear-capable by the United States and Soviet Union in the early years of the nuclear nonproliferation regime.[39] The model performs well in comparison to this qualitative approach.

False positives and false negatives provide a better sense of the face validity of the proliferation risk model. The former are cases that the model judged to have a high probability of nuclear pursuit, but which the data indicate did not have a nuclear weapons program; the latter are cases in which there was a nuclear weapons program, but the model predicted there would be nuclear restraint. False positives and false negatives for the proliferation risk model are shown in Table 3.

Iran and Egypt are well represented among the false positives. Both cases are plausible. Many experts see Egypt as having had nuclear weapons ambitions, even if it

[39] Coe and Vaynman 2015. Evaluation of this qualitative measure is limited to the years 1957–1974, to better match the declassified source material. The performance of the measure decreases if a broader timeframe is used.

Table 2: Predictive metrics for measures of proliferation risk

	AUC (area under the ROC curve)	**F_1 Score**	**Incorrect predictions**
Proliferation risk model	0.89	0.52	2.8 percent
Qualitative risk (1957–1974)	0.81	0.26	15.3 percent

had not taken concrete action to launch a nuclear weapons program.[40] Bleek, for example, codes Egypt as "exploring" nuclear weapons from 1955–1980.[41] Whether Iran truly halted its nuclear weapons program in 2003, as asserted by the US intelligence community, has remained in dispute.[42] A number of non-governmental analysts judge Iran as continuing its nuclear weapons program through the 2000s.[43] Both Iran and Egypt had ample resources to bring to bear on nuclear development, worried about rivals with nuclear weapons or nuclear weapons programs, and lacked an alliance with a nuclear-armed patron.

South Korea in 1979 and France in 1948 are also considered by the model as very likely to have a nuclear weapons programs, and again these are plausible cases. Some analysts have argued that South Korea—despite significant pressure from the United States—did not fully shutter its nuclear weapons program until as late as 1981 [44]. Most

[40] Rublee 2006; Rublee 2009a.

[41] Bleek 2017.

[42] Arnold et al. 2021; National Intelligence Council 2007.

[43] Bleek 2017; Jo and Gartzke 2007; Singh and Way 2004.

[44] Bleek 2017.

see

Table 3: False positives and negatives from the proliferation risk model

False positives		False negatives	
Country	**Years**	**Country**	**Years**
Iran	2006–2010	Libya	1973 1975
Egypt	1997–1998 2003	Taiwan	1967
Iran	1983	Libya	1979
Egypt	1996 2001–2002	Taiwan	1968–1969
Iran	2004	Libya	1976–1977
South Korea	1979	Taiwan	1970–1971
Egypt	1994–1995 1999–2000	Libya	1974 1978
France	1948	Yugoslavia	1950–1953 1982–1985 1986–1987
Egypt	1988	Libya	1970–1972
Czechoslovakia	1979 1981–1983	Syria	2002–2003

France as launching a full nuclear weapons program in the mid-1950s, but some code France as “exploring” nuclear weapons immediately after World War II.[45]

The most questionable case on the false positives list is Czechoslovakia in the late 1970s and early 1980s. Czechoslovakia is not known to have had nuclear weapons

[45] Ibid.; Singh and Way 2004.

ambitions or to have made any attempt to realize them if it did. But the model might be forgiven for seeing similarities between this case and the case of Romania, which engaged in fledgling nuclear weapons efforts around this time.[46] While Czechoslovakia hewed closer to the Warsaw Pact line than did Romania in this period, it also had an extensive civilian nuclear infrastructure fueled by Soviet nuclear assistance.[47] Its nuclear latency in those years—Czechoslovakia operated a lab-scale reprocessing facility beginning in 1977—suggests the model's assessment is not unreasonable.[48]

Turning to false negatives, the model's biggest misses come from Libya in the 1970s and Taiwan in the late 1960s and early 1970s. Yugoslavia, in the early 1950s and the 1980s, and Syria in 2002–2003 also make the list. Not coincidentally, all of these cases would also be missed by simply identifying states operating enrichment and reprocessing (ENR) facilities. None of these countries had ENR capabilities in these time periods, although Taiwan operated research reactors throughout this period and Yugoslavia ran research reactors and a power reactor through the 1980s.[49] Further, most of these cases are the subject of disagreement among scholars who have undertaken detailed coding of nuclear weapons programs. Bleek codes Taiwan's activities as mere

[46] For a detailed discussion of nonproliferation dynamics in the Warsaw Pact, see Lanoszka 2018.

[47] A 1979 US intelligence assessment highlights Czechoslovakia as possessing "the most highly developed nuclear program in Eastern Europe, including a broad-based research program in reactor and fuel-cycle technology, a large nuclear power program, and a well-developed nuclear industry." See National Foreign Assessment Center 1979.

[48] Fuhrmann and Tkach 2015.

[49] Kaplow 2020.

"exploration," rather than "pursuit."[50] Jo and Gartzke do not see Libya as engaging in a nuclear weapons program at all.[51] Bleek codes Yugoslavia's first attempt at pursuing nuclear weapons as beginning only in 1953—after the earliest period flagged by the model as a false negative—and Singh and Way do not consider any of Yugoslavia's nuclear work as meeting the threshold for nuclear pursuit.[52] Even the early years of the Syria program are debatable, as construction of the al-Kibar nuclear reactor—later bombed by Israel—probably did not commence until 2001 or 2002.[53]

This list of false positives and false negatives suggests some face validity for the model's predictions. Given that not every case will be predicted correctly, it is encouraging to see arguable cases among those that the model missed. Coupled with strong predictive metrics, there is reason to believe the proliferation risk model is a useful way of distinguishing those cases that carry some proliferation risk.

High-Risk Cases

Proliferation risk scores can help identify the set of states most likely to seek nuclear weapons under particular circumstances. It seems reasonable to begin with today's strategic environment. Which states appear most likely to have a nuclear weapons program today? Table 4 shows the states with the highest proliferation risk scores, using the most recent information in the dataset assembled for this project.

---

[50] Bleek 2017.

[51] Jo and Gartzke 2007.

[52] Bleek 2017; Singh and Way 2004.

[53] Albright and Brannan 2008; Office of the Director of National Intelligence 2008.

Table 4: Countries with the highest current proliferation risk scores

Proliferation Risk Score Rank	**Country**
1	Syria
2	Iran
3	Saudi Arabia
4	Vietnam
5	Afghanistan
6	Libya
7	Taiwan
8	Jordan
9	Lebanon
10	Cuba

The countries on this list are largely those with some latent nuclear capability, economic resources, and/or significant external threats. The inclusion of some of these states comes as no surprise. Iran figures prominently on this list and is undoubtedly the country of most proliferation concern as of this writing. Taiwan—facing an existential threat from a nuclear-armed neighbor—is perennially featured on expert lists of most likely proliferants, as is Saudi Arabia. Others, like Syria and Libya, would be of increased concern absent internal conflict, a dynamic which may not be adequately recognized by the model.

But some of these states are harder to reconcile with expert opinion. Vietnam, for example, is a staunch supporter of nuclear nonproliferation globally and an early adherent to the nuclear weapons ban treaty. While it is engaged in a long-term dispute with other claimants over territory in the South China Sea, it is not a country that is generally thought of as facing substantial security threats. It likely features on this list by virtue of its nuclear latency. Vietnam operates a nuclear research reactor, has received significant nuclear technical assistance from the International Atomic Energy Agency, and in 2014 concluded a nuclear cooperation agreement with the United States.

The high proliferation risk scores of countries such as Vietnam and Jordan illustrate both the benefits and pitfalls of a model-based approach to assessing proliferation risk. On the one hand, if this method produced a list of most-likely proliferants that was identical to the conventional wisdom, it would be of little use as an aide to analysis. On the other hand, a high risk score for a country like Vietnam, with its strong nonproliferation credentials, might lead some to question the predictive efficacy of the underlying model.

If the top ten list of proliferation risk were entirely populated with states like Vietnam, there would be some cause to worry. A well-functioning predictive model should be able to identify countries like Iran as a high-risk case; regardless of Tehran's weapons intent, it is openly building up a capability that could be used to produce nuclear weapons. But, at the same time, the purpose of the model is to surface out-of-the-box ideas about proliferation risk. That Vietnam has some characteristics in common with past cases of nuclear proliferation seems worth identifying.

Identifying High-Risk Global Conditions

Proliferation risk scores can help identify the conditions under which nuclear proliferation becomes more likely, both for specific states in particular circumstances, and for nuclear proliferation generally. By adjusting the data to reflect hypothetical changes in the global environment, analysts can assess the relative impact of these changes on the risk of nuclear proliferation. Table 5 shows the effect of some possible changes in international security and their effect on proliferation risk scores.

The breakdown of the nuclear nonproliferation regime, here represented by a substantial increase in the number of states seen as cheating on their NPT commitments, has a dramatic effect on proliferation risk scores, with the average risk of proliferation nearly tripling in this scenario. Less extreme but still substantial changes in global proliferation risk result from weakening of US alliance ties and increasing international disputes for those states that are already prone to conflict. Granting additional nuclear latency to countries that are in the process of developing such capabilities increases the average proliferation risk score by about 13 percent, and a shift toward authoritarian regimes for states currently lacking strong democracies or autocracies leads to an average increase in proliferation risk scores of about 12 percent.

This approach facilitates a form of what-if analysis in which the consequences for proliferation of particular policy approaches can be examined in a more systematic way. Often such analyses involve merely a general sense of the policy trade-offs built into international security decision-making. It is particularly tempting for policy analysts to overvalue the short-term impact of bilateral relations over more amorphous multilateral

Table 5: Effect of changing international conditions on average proliferation risk score

Change in Global Condition	**Increase in Average Proliferation Risk Score**
Breakdown of nonproliferation regime	192 percent
Weakening of US alliance ties	29 percent
Increase in international disputes	27 percent
Increase in nuclear latency	13 percent
Increase in authoritarianism	12 percent

policy goals, like strengthening the nuclear nonproliferation regime. Proliferation risk scores can provide a reminder of the scale of the impact broad policy changes can have on the global nuclear proliferation environment.

Proliferation in East Asia

A similar approach can be applied to a particular hypothetic scenario, yielding insights about the proliferation dynamics in specific countries or regions. Japan and South Korea, for example, are nearly always among the nations seen as a potential proliferation risk.[54] North Korea's continuing nuclear development in the region, coupled with long-term security competition with China and with each other, provide ample security incentives to seek at least a latent nuclear capability, if not a full-fledged nuclear weapons arsenal. Both Japan and South Korea have an extensive domestic nuclear infrastructure that could be quickly turned to weapons purposes if it desired. Both

[54] On South Korea's nuclear ambitions, see Hersman and Peters 2006; and Fitzpatrick 2016. On Japan, see Furukawa 2012; Rublee 2009b; and Solingen 2010.

countries have considered nuclear weapons in the past. Japan examined its nuclear weapons policy in the late 1960s and early 1970s amidst deciding whether to join the NPT, and it reconsidered that policy again in the mid-1990s and early 2000s when faced with a rising nuclear threat from North Korea.[55] South Korea conducted a small-scale nuclear weapons development effort in the mid-to-late 1970s, ultimately shuttering its work under US pressure.[56]

Proliferation risk scores can help examine what factors are most essential for achieving nuclear restraint by these two key nations in East Asia. This what-if analysis begins with a baseline level of proliferation risk for both states, based on the latest information available in the dataset assembled for this project. Using a baseline model of proliferation risk, Japan has a proliferation risk score of 0.24 and South Korea has a proliferation risk score of 1.34. These scores translate to the model's assessment of the probability of a nuclear weapons program in each state, so Japan is seen as having a 0.24 percent chance of seeking nuclear weapons, while South Korea has a 1.34 percent chance of pursuing weapons.

From this baseline, it is straightforward to adjust factors that may matter for nuclear proliferation according to hypothetical scenarios for the East Asia security environment. Imagine, for example, that the United States walks back security guarantees for these key allies, or otherwise leads Japan and South Korea to question the

[55] Kishi 2018; Green and Furukawa 2008; Rublee 2010.

[56] Fitzpatrick 2016; Bleek 2017.

US commitment to the region.[57] Analysts can then recalculate the proliferation risk score under these new, hypothetical conditions, to better understand the predicted effect that particular changes in policy or in the general security environment have on the risk of proliferation. Table 6 shows several possible adjustments and the corresponding change in proliferation risk scores for each country.

Weakening the US alliance has a significant effect on both states. For Japan, the loss of the US alliance increases the proliferation risk score by a factor of 8, to about 1.9. The effect is more muted in South Korea, which begins with a higher baseline level of risk, but still nearly quadruples the baseline level of risk. The effects of increased international disputes and a breakdown in the nuclear nonproliferation regime are smaller, but these effects compound when combined. In a scenario in which there is both an increase in international disputes and a breakdown in the nuclear nonproliferation regime, Japan's risk of proliferation increases by 9 times, and South Korea's by 3 times. A wider breakdown in regional security, including a weakening US alliance, has a particularly dramatic effect on the chances of proliferation in Japan, which sees a more than 40 times increase in proliferation risk. The model assesses that there would be about a 10 percent chance that Japan would seek nuclear weapons in that scenario.

[57] Akita 2020.

Table 6: Effect of changing conditions on proliferation risk in Japan and South Korea

Condition	**Increase in Proliferation Risk Score for Japan**	**Increase in Proliferation Risk Score for South Korea**
Weakening US alliance	8.0x	3.7x
Increase in international disputes	2.9x	1.2x
Breakdown in nonproliferation regime	4.1x	2.6x
Increase in international disputes and breakdown in nonproliferation regime	9.2x	3.3x
Weakening US alliance, increase in international disputes, and breakdown in nonproliferation regime	41.7x	9.3x

Conclusions

This project constructed a predictive model of nuclear proliferation and introduced proliferation risk scores—a model-based assessment of the risk of proliferation given a particular set of circumstances. This article illustrates some of the ways that proliferation risk scores can be used as an adjunct to expert analysis: surfacing cases of proliferation risk for additional scrutiny, assisting in a more systematic approach to what-if and scenario analysis, and helping to quantify the effect of changes in nonproliferation policy and the international security environment on the risk of nuclear proliferation.

The methods introduced here are flexible and scalable. As new information becomes available, the statistical learning model underlying proliferation risk scores can be updated to improve predictive accuracy. For analysts focused on specific aspects of nonproliferation policy, such as the effectiveness of nonproliferation sanctions, small additions to the underlying data and relatively minor changes to the modeling approach used here could yield insights as to the conditions under which these policy levers are most likely to succeed. The same approach can also be applied as an aid to traditional analysis in other areas of international security, such as in assessing the risk of conflict or terrorism.[58]

[58] Ward, Greenhill, and Bakke 2010.

Funding
This research was funded by the Defense Threat Reduction Agency (DTRA) through the Project on Advanced Systems and Concepts for Countering Weapons of Mass Destruction (PASCC) at the US Air Force Academy, Award #FA7000-19-2-0022.

Data Availability Statement
The data presented in this study are openly available from www.jkaplow.net/research.

Acknowledgments
The author gratefully acknowledges the many contributions to this project of the students of NukeLab, the undergraduate social science research lab at William & Mary's Global Research Institute, and in particular Madeline Kelleher, Keaton Lee, Erica Bucchieri, Elizabeth Brown, Jameson Crouse, and Katie Spofford.

Conflicts of Interest
The author declares no conflict of interest.

References

Akita, Hiroyuki. 2020. Trump demands Japan and South Korea pay for nuclear umbrella. *Nikkei Asia*.

Albright, David, and Paul Brannan. 2008. *The Al Kibar Reactor: Extraordinary Camouflage, Troubling Implications*. ISIS Report. Institute for Science and International Security.

Arlot, Sylvain, and Alain Celisse. 2010. A survey of cross-validation procedures for model selection. *Statistics Surveys* 4: 40–79.

Arnold, Aaron, Matthew Bunn, Caitlin Chase, Steven E. Miller, Rolf Mowatt-Larssen, and William H. Tobey. 2021. The Iran nuclear archive: impressions and implications. *Intelligence and National Security* 36 (2): 230–242.

Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. Estimating Dynamic State Preferences from United Nations Voting Data. *Journal of Conflict Resolution* 61 (2): 430–456.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. Improving Quantitative Studies of International Conflict: A Conjecture. *American Political Science Review* 94 (1): 21–35.

Bleek, Philipp C. 2017. *When Did (and Didn't) States Proliferate? Chronicling the Spread of Nuclear Weapons*. Discussion Paper. Cambridge, MA: Project on Managing the Atom, Belfer Center for Science and International Affairs.

Bleek, Philipp C., and Eric B. Lorber. 2014. Security Guarantees and Allied Nuclear Proliferation. *Journal of Conflict Resolution* 58 (3): 429–454.

Brown, Robert L., and Jeffrey M. Kaplow. 2014. Talking Peace, Making Weapons IAEA Technical Cooperation and Nuclear Proliferation. *Journal of Conflict Resolution* 58 (3): 402–428.

Carcelli, Shannon, Erik Gartzke, Rebecca Davis Gibbons, and Jeffrey M. Kaplow. 2014. The Nuclear Regime Complex: A New Dataset. *Paper prepared for the 2014 International Studies Association Annual Convention, Toronto, Canada*.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*: 321–357.

Cirincione, Joseph. 2008. *Bomb scare: the history and future of nuclear weapons*. New York, NY: Columbia University Press.

Coe, Andrew J., and Jane Vaynman. 2015. Collusion and the Nuclear Nonproliferation Regime. *The Journal of Politics* 77 (4): 983–997.

Conroy, Bryan, Larry Eshelman, Cristhian Potes, and Minnan Xu-Wilson. 2016. A dynamic ensemble approach to robust classification in the presence of missing data. *Machine Learning* 102 (3): 443–463.

Fang, Fang, Wei Lan, Jingjing Tong, and Jun Shao. 2019. Model Averaging for Prediction With Fragmentary Data. *Journal of Business & Economic Statistics* 37 (3): 517–527.

Fitzpatrick, Mark. 2016. *Asia's latent nuclear powers: Japan, South Korea and Taiwan*. Adelphi 455. London, UK: International Institute for Strategic Studies.

Fuhrmann, Matthew. 2009. Spreading Temptation: Proliferation and Peaceful Nuclear Cooperation Agreements. *International Security* 34 (1): 7–41.

Fuhrmann, Matthew, and Michael C. Horowitz. 2015. When Leaders Matter: Rebel Experience and Nuclear Proliferation. *The Journal of Politics* 77 (1): 72–87.

Fuhrmann, Matthew, and Sarah E. Kreps. 2010. Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941-2000. *Journal of Conflict Resolution* 54 (6): 831–859.

Fuhrmann, Matthew, and Yonatan Lupu. 2016. Do Arms Control Treaties Work? Assessing the Effectiveness of the Nuclear Nonproliferation Treaty. *International Studies Quarterly* 60 (3): 530–539.

Fuhrmann, Matthew, and Todd S. Sechser. 2014. Nuclear Strategy, Nonproliferation, and the Causes of Foreign Nuclear Deployments. *Journal of Conflict Resolution* 58 (3): 455–480.

Fuhrmann, Matthew, and Benjamin Tkach. 2015. Almost nuclear: Introducing the Nuclear Latency dataset. *Conflict Management and Peace Science* 32 (4): 443–461.

Furukawa, Katsuhisa. 2012. Japan's Nuclear Option. In *Over the Horizon Proliferation Threats*, edited by James J. Wirtz and Peter R. Lavoy, 13–32. Stanford, CA: Stanford University Press.

García-Laencina, Pedro J., José-Luis Sancho-Gómez, and Aníbal R. Figueiras-Vidal. 2010. Pattern classification with missing data: a review. *Neural Computing and Applications* 19 (2): 263–282.

Gerzhoy, Gene. 2015. Alliance Coercion and Nuclear Restraint: How the United States Thwarted West Germany's Nuclear Ambitions. *International Security* 39 (4): 91–129.

Gleditsch, Kristian Skrede. 2002. Expanded Trade and GDP Data. *Journal of Conflict Resolution* 46 (5): 712–724.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39 (5): 615–637.

Goertz, Gary, Paul Francis Diehl, and Alexandru Balas. 2016. *The Puzzle of Peace: The Evolution of Peace in the International System*. Oxford, UK: Oxford University Press.

Golfarelli, Matteo, Stefano Rizzi, and Andrea Proli. 2006. Designing what-if analysis: towards a methodology. In *Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, 51–58.

Green, Michael J., and Katsuhisa Furukawa. 2008. Japan: New Nuclear Realism. In *The Long Shadow: Nuclear Weapons and Security in 21st Century Asia*, edited by Muthiah Alagappa, 347–372. Stanford, CA: Stanford University Press.

Hersman, Rebecca K. C., and Robert Peters. 2006. Nuclear U-Turns. *The Nonproliferation Review* 13 (3): 539–553.

Heuer, Jr., Richards J., and Randolph H. Pherson. 2020. *Structured Analytic Techniques for Intelligence Analysis*. CQ Press.

Horowitz, Michael C., and Neil Narang. 2014. Poor Man's Atomic Bomb? Exploring the Relationship between "Weapons of Mass Destruction". *Journal of Conflict Resolution* 58 (3): 509–535.

Huss, William R. 1988. A move toward scenario analysis. *International Journal of Forecasting* 4 (3): 377–388.

Hymans, Jacques E. C. 2006. *The Psychology of Nuclear Proliferation: Identity, Emotions, and Foreign Policy*. Cambridge, UK: Cambridge University Press.

Jo, Dong-Joon, and Erik Gartzke. 2007. Determinants of Nuclear Weapons Proliferation. *Journal of Conflict Resolution* 51 (1): 167–194.

Kaplow, Jeffrey M. 2020. All the World's Nuclear Reactors. Dataset.

Kishi, Toshimitsu. 2018. *Deliberations on Japanese Nuclear Policy During the Sato Administration: Studies by the Cabinet Research Office*. Tokyo, Japan: National Graduate Institute for Policy Studies.

Kotsiantis, Sotiris, Dimitris Kanellopoulos, and Panayiotis Pintelas. 2006. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering* 30 (1): 25–36.

Kroenig, Matthew. 2009. Importing the Bomb: Sensitive Nuclear Assistance and Nuclear Proliferation. *Journal of Conflict Resolution* 53 (2): 161–180.

Lanoszka, Alexander. 2018. Nuclear Proliferation and Nonproliferation among Soviet Allies. *Journal of Global Security Studies* 3 (2): 217–233.

Luque, Amalia, Alejandro Carrasco, Alejandro Martín, and Ana de las Heras. 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition* 91: 216–231.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2010. *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010*. Vienna, VA: Center for Systemic Peace.

Miller, Nicholas L. 2014a. Nuclear Dominoes: A Self-Defeating Prophecy? *Security Studies* 23 (1): 33–73.

Miller, Nicholas L. 2014b. The Secret Success of Nonproliferation Sanctions. *International Organization* 68 (04): 913–944.

National Foreign Assessment Center. 1979. *Nuclear Programs of Eastern Europe*. Intelligence Assessment.

National Intelligence Council. 2007. *Iran: Nuclear Intentions and Capabilities*. National Intelligence Estimate.

Office of the Director of National Intelligence. 2008. Background Briefing with Senior U.S. Officials on Syria's Covert Nuclear Reactor and North Korea's Involvement.

Reiter, Dan. 2014. Security Commitments and Nuclear Proliferation. *Foreign Policy Analysis* 10 (1): 61–80.

Reynolds, Celia L., and Wilfred T. Wan. 2012. Empirical Trends in Sanctions and Positive Inducements in Nonproliferation. In *Sanctions, Statecraft, and Nuclear Proliferation*, edited by Etel Solingen, 56–122. Cambridge, UK: Cambridge University Press.

Rublee, Maria Rost. 2006. Egypt's Nuclear Weapons Program. *The Nonproliferation Review* 13 (3). Routledge: 555–567.

Rublee, Maria Rost. 2009a. *Nonproliferation norms: why states choose nuclear restraint*. Athens, GA: University of Georgia Press.

Rublee, Maria Rost. 2009b. The Future of Japanese Nuclear Policy. *Strategic Insights* 8 (2): 1–11.

Rublee, Maria Rost. 2010. The Nuclear Threshold States. *The Nonproliferation Review* 17 (1): 49–70.

Sagan, Scott D. 1996. Why Do States Build Nuclear Weapons?: Three Models in Search of a Bomb. *International Security* 21 (3): 54–86.

Sagan, Scott D. 2011. The Causes of Nuclear Weapons Proliferation. *Annual Review of Political Science* 14 (1): 225–244.

Singh, Sonali, and Christopher R. Way. 2004. The Correlates of Nuclear Proliferation: A Quantitative Test. *Journal of Conflict Resolution* 48 (6): 859–885.

Solingen, Etel. 2010. The Perils of Prediction: Japan's Once and Future Nuclear Status. In *Forecasting Nuclear Proliferation in the 21st Century*, edited by William C. Potter and Gaukar Mukhatzhanova, 2:131–157. Vol. 2. Palo Alto, CA: Stanford University Press.

Steinwart, Ingo, and Andreas Christmann. 2008. *Support Vector Machines*. Springer Science & Business Media.

Sun, Yanmin, Andrew K. C. Wong, and Mohamed S. Kamel. 2009. Classification of imbalanced data: a review. *International Journal of Pattern Recognition and Artificial Intelligence* 23 (04): 687–719.

Swets, John A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240 (4857): 1285–1293.

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

Way, Christopher, and Jessica L. P. Weeks. 2014. Making It Personal: Regime Type and Nuclear Proliferation. *American Journal of Political Science* 58 (3): 705–719.