Predicting Proliferation: High Reliability Forecasting Models of Nuclear Proliferation as a Policy and Analytical Aid[*]

Final Technical Report for PASCC Award #N00244-15-1-0003

Jeffrey M. Kaplow[†] and Erik Gartzke[‡]

Executive Summary

Researchers have now produced a number of quantitative studies of the determinants of nuclear proliferation, using data on all known nuclear weapons programs. But while scholars have laid important groundwork in understanding the causes of nuclear pursuit, these studies are primarily focused on *explaining* rather than *predicting* proliferation. Drawing from existing quantitative work, this project uses statistical learning methods to construct a predictive model of proliferation, focusing on the ability of different nuclear proliferation theories to make accurate out-of-sample predictions. This study makes two contributions to the literature on nuclear proliferation and the larger policy debate. First, it identifies for the first time an empirically grounded set of nuclear "triggers"—conditions under which countries are most likely to shift from latent nuclear capacity to a full-fledged nuclear weapons effort. Understanding these triggers has become increasingly important, as more states have begun to pursue a nuclear hedging strategy in which they seek dual-use nuclear capabilities without committing to a weapons program. Second, this study helps to reconcile conflicting empirical findings in the literature. Predictive analytics provide a new and useful way of understanding the substantive significance of existing empirical findings, and of comparing the relative importance of different theoretical approaches.

---

[*] This research was supported by the Project on Advanced Systems and Concepts for Countering Weapons of Mass Destruction (PASCC) at the Naval Postgraduate School, Award #N00244-15-1-0003. An earlier draft of this project was presented at the Annual Meeting of the American Political Science Association, September 2015, San Francisco, CA.

[†] College of William & Mary; Department of Government. Email: jkaplow@wm.edu

[‡] University of California, San Diego; Department of Political Science. Email: egartzke@ucsd.edu

Assessing the risk of nuclear proliferation is a difficult job, made more challenging by the large amount of sometimes conflicting data available to the intelligence and policy communities. **The analysis presented here attempts to contribute to these assessments by focusing on the predictive power of various indicators of proliferation risk.** Where other studies examine specific cases of proliferation or restraint to gain insight about particular circumstances or causal processes, this project asks which indicators best predict a state's decision to seek nuclear weapons in general.

Analysts have examined a number of theories, causal models, or predictive indicators for nuclear proliferation, including a state's underlying productive capacity, recent conflict behavior, foreign assistance, nuclear rivalry, alliance ties, and international institutions.

- **Among these, the presence of a nuclear rival appears to be the strongest predictor of nuclear proliferation.** This finding suggests that analysts should continue to use the nuclear behavior of rivals as a valuable indicator of proliferation risk, and also points to the continued relevance of concerns about proliferation cascades in the Middle East and East Asia.

- **Two other factors—conflict behavior and the presence of a nuclear ally— significantly contribute to the accuracy of nuclear proliferation predictions.**

- Considering a state's membership in international institutions, its receipt of nuclear assistance, and the presence of latent capability did not significantly improve prediction in this analysis.

This analysis is merely a first step in bringing the new tools of "big data" to bear on the issue of proliferation. In the future, predictive analytics may serve as an important adjunct to expert knowledge of proliferation pathways, helping the intelligence and policy communities focus on the most relevant indicators of proliferation risk.

Introduction

In recent years, scholars have turned increasingly to quantitative tools to study the drivers of the proliferation of nuclear weapons technology. These efforts complement a longstanding body of qualitative work, and add to an admirable diversity of theoretical approaches—from psychological (Hymans 2006), to bureaucratic (Sagan 1996), to normative (Rublee 2009). The study of nuclear proliferation does not lack for theories or for empirical findings.

Few scholars, however, have attempted to synthesize or adjudicate among disparate findings.[1] This is not surprising—the incentives for proposing a novel theory or illuminating a new empirical association frequently outweigh those for efforts to understand collective progress of the field. But this gap makes it difficult to understand the extent of our collective knowledge of nuclear proliferation. The number of theories and findings also creates a substantial barrier when it comes to communicating with policy practitioners. With so many well-studied causes of nuclear proliferation—but no real sense of which is most important in any given case—it is difficult to translate academic findings into recommendations for counterproliferation policy.

In this study, we weigh a variety of factors that have been found to encourage or prevent nuclear proliferation, according to a simple metric: the ability of these factors to predict nuclear weapons proliferation out-of-sample. By examining the predictive validity of key explanatory variables, we cut through the literature's accumulated findings of

[1] For exceptions, see Gartzke and Kroenig (2009, 2014) and Sagan (2011). For criticisms that the results of statistical tests of nuclear proliferation causes or effects may not be meaningful empirically, see Bell (2015) and Douglass and Lanoszka (2015).

statistical significance and ask, simply, "how much does this factor help us predict proliferation?"

This study has three parts. In the first part, we describe the state of empirical work on the causes of nuclear proliferation and highlight six contenders for the title of best indicator of a nuclear weapons program: capability, conflict behavior, nuclear assistance, rivalry, alliances, and institutions. In the second part, we describe our approach to measuring the predictive validity of these factors, and our efforts to address potential threats to the validity of this analysis. In the third part, we present our findings and suggest several avenues for subsequent research.

Theories of nuclear proliferation

The literature on nuclear proliferation proposes a number of potential drivers of state decisions to seek nuclear weapons. Here, we highlight six theories that scholars have previously examined using quantitative methods. In focusing on these six theories, we are clearly leaving out several influential explanations for state behavior, including theories based on international norms (Rublee 2009; Sagan 1996), domestic politics and regime type (Sagan 1996; Way and Weeks 2014), and economic liberalization (Solingen 1994). We also fail to address directly the relative importance of specific counterproliferation policies, such as international sanctions (Haggard and Nolan 2012; Miller 2014) or military action (Kreps and Fuhrmann 2011). We leave for future research the task of evaluating the relative predictive power of these additional causal factors.

Capability

Nuclear weapons development requires significant indigenous capability. Large-scale weapons programs draw on a country's financial resources, of course, but also call for trained scientists and engineers, access to uranium reserves, the ability to work with specialty metals, and the sustained attention of both political authorities and a large cadre of competent administrators. The size of the undertaking might be expected to deter all but the most technically capable states from launching nuclear weapons programs. Why expend the resources and risk international opposition when the successful development of a nuclear weapon seems so unlikely?

Certainly, a number of states have opted to forego weapons because of the substantial cost of proliferating, but there also are several examples of states that began nuclear weapons programs while lacking almost all of the required resources. Pakistani Prime Minister Zulfikar Ali Bhutto famously said in 1965 that if India develops nuclear weapons, "we will eat grass or leaves, even go hungry, but we will get one of our own." And the spread of civilian nuclear technology has almost certainly significantly reduced the cost of a weapons program since the early days of the nuclear age. States that already operate a nuclear power plant, for example, can apply the same technology to the production of fissile material for a nuclear weapon.

Several scholars have identified a strong link between supply-side factors—including underlying state capability—and a country's likelihood of seeking nuclear weapons. Singh and Way (2004), for example, find that increasing GDP per capita increases the risk that a low-GDP state will pursue nuclear weapons, and Jo and Gartzke

(2007) show that latent nuclear capacity is significantly associated with weapons programs. Gartzke and Kroenig (2009) thus conclude that supply-side factors—to include state capability—"are among the most important determinants of nuclear proliferation."

Conflict behavior

A state's concern about its own security is likely to be a fundamental driver of nuclear weapons pursuit (Sagan 1996). States justify the expense and risk associated with seeking weapons largely in terms of the security benefits they believe will follow from nuclear acquisition. Countries that have been more frequent participants in international conflict—or that expect to be involved in conflict in the future—will find the cost of nuclear weapons development easier to justify. It is no surprise, then, that multiple studies have shown a strong association between a state's conflict behavior and the pursuit of nuclear weapons (Fuhrmann 2009; Jo and Gartzke 2007).[2]

Nuclear assistance

Advocates of supply-side theories of nuclear proliferation have focused on the provision of nuclear assistance to would-be proliferants. This assistance can take several forms. Supplier states may provide sensitive assistance that contributes directly to weapons design or fissile material production efforts.[3] States may sign nuclear cooperation agreements that facilitate the transfer of civilian nuclear technology to be used for power

[2] This is not a universal finding, however. See, for example, Brown and Kaplow (2014).

[3] The number of potential suppliers of sensitive assistance has been increasing in recent years. See Braun and Chyba (2004).

production or other non-weapons purposes. And states may receive multilateral nuclear assistance from international organizations such as the International Atomic Energy Agency.

Each form of nuclear assistance has the potential to make a nuclear weapons program less costly. Even civilian nuclear assistance can have this effect; the dual-use nature of nuclear power programs means that the same technology can be applied to research and development of nuclear weapons if the recipient state makes the decision to seek a bomb. By lowering the barriers to a successful nuclear weapons effort, nuclear assistance may make states more likely to take the initial step and launch a nuclear weapons program. Studies that have examined the role of nuclear assistance in encouraging proliferation have found that sensitive assistance, bilateral civilian assistance, and multilateral civilian assistance are strongly associated with the decisions of states to seek nuclear weapons (Brown and Kaplow 2014; Fuhrmann 2009; Kroenig 2009).

Nuclear rivalry

Most nuclear weapons states have joined the nuclear club at least partially in response to the nuclear pursuit of another state. The Manhattan Project was driven partly by fears of German nuclear efforts. The Soviet Union sought weapons, in part, to keep step with the United States. China's nuclear program spurred India's program, which in turn pushed Pakistani nuclear development. If nuclear rivalry is a primary driver of nuclear weapons programs, then we should worry about proliferation cascades or dominoes, in which the pursuit of nuclear weapons by one state prompts weapons programs in others, which lead to programs in other states, and so on.

Quantitative findings on the role of nuclear rivalry, however, have surprisingly been less than emphatic. In most of his quantitative models, Fuhrmann (2009) finds that having a nuclear rival makes states more likely to seek their own nuclear weapons. Other studies, however, find this relationship is not robust to multiple model specifications (Bleek and Lorber 2014; Brown and Kaplow 2014; Kroenig 2009). Jo and Gartzke (2007) report the counterintuitive finding that states are actually less likely to seek nuclear weapons when they have nuclear rivals.[4] The mixed nature of existing evidence is grounds for additional analysis, something we address here.

Nuclear alliances

Faced with external threats, states may seek nuclear weapons for protection, or they may turn to allies that already possess nuclear weapons. This latter option, if available, may be significantly cheaper than embarking on one's own nuclear weapons program. To the extent that these two strategies are substitutes, we might expect that countries that are able to rely on the "nuclear umbrella" of an ally should be less likely to seek nuclear weapons themselves. At times, the US nuclear umbrella has been explicitly exercised as a tool of nonproliferation—a means intended to convince allied states that they do not need to develop their own nuclear deterrent.

Here, too, quantitative studies of nuclear proliferation have come to different conclusions. Fuhrmann (2009) and Brown and Kaplow (2014) find no relationship

[4] This result may, they note, be attributable to concerns about preventive conflict.

between the presence of a nuclear ally and the likelihood that a state seeks nuclear weapons. Other work identifies some nonproliferation benefit to a nuclear umbrella, but not in all model specifications (Jo and Gartzke 2007; Kroenig 2009; Singh and Way 2004). In a reanalysis of these studies, using modified data, Bleek and Lorber (2014) find strong support for nuclear alliances in discouraging proliferation.

Institutions

Since the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) came into force in 1970, only four states have acquired nuclear weapons, and no state has acquired nuclear weapons while also a member of the NPT.[5] For some, this is strong evidence of the effectiveness of the NPT in constraining state behavior. The NPT has even been hailed as one of the most successful security treaties in world history (Cirincione 2008), chiefly because it is seen as a key factor in keeping the number of nuclear weapons states far below the dire predictions of experts in the 1960's.

Of course, others are less impressed. Detailed case studies of nuclear restraint—the decisions of high-risk states to forgo weapons—have found little role for the NPT (Reiss 1995; Rublee 2009). Many argue that the NPT is in crisis, and that whatever constraining power it once had is now eroding (Perkovich 2006; Sauer 2006; Williams and Wolfstahl 2005). At least one analyst has suggested that the NPT is likely to do more harm than good in the future (Wesley 2005). Summarizing the state of research on the impact of the NPT, Potter (2010) laments that "NPT advocates and critics alike typically

[5] One state that acquired nuclear weapons, South Africa, later gave them up. North Korea only reached nuclear weapons status after withdrawing from the NPT. India, Israel, and Pakistan are nuclear capable countries that never joined the NPT.

assert their preferred views about the merits of the Treaty and its (in)dispensable contribution in retarding the spread of nuclear weapons."

A number of quantitative studies have examined the role of the NPT in at least a cursory way. These analyses typically include as a control variable a measure of whether a state is a member of the NPT in a given year. This variable, when included in a model of nuclear pursuit, is often significantly associated with a decreased likelihood of proliferation (Bleek and Lorber 2014; Brown and Kaplow 2014; Fuhrmann 2009; Jo and Gartzke 2007). As these authors are quick to acknowledge, however, the significant negative correlation between NPT membership and nuclear weapons programs does not tell us much about the independent role of the NPT, because there is likely to be a significant selection effect with regard to treaty membership. If states are more likely to join the NPT when they have no intention of proliferating, as we might expect, then this correlation may have little to do with the constraining power of the treaty.[6]

The predictive power of proliferation theory

Most quantitative studies of nuclear proliferation aim to identify a causal link between a particular factor and a state's propensity to seek nuclear weapons. In service of this goal, scholars 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, moreover, make some attempt to account for alternative explanations for their results, often by controlling for confounding variables

[6] For an analysis that seeks to avoid this selection problem, see Kaplow (2013).

in a regression model. This mode of analysis is useful in understanding whether an individual factor affects the outcome of interest.[7]

We take a somewhat different approach. We make no attempt to test hypotheses or identify, except in an indirect way, individual factors that matter for nuclear proliferation. Instead we inquire, with the above theories as our guide, how well one can predict proliferation. How much do each of the above theories contribute to our ability to make accurate predictions?

A focus on out-of-sample predictive validity has several advantages. First, it moves us away from the questionable emphasis in the quantitative literature on statistical significance as a metric for a successful result. Second, testing the performance of models within our data sample—as we do, for example, in traditional regression models—is a kind of teaching to the test. This practice risks overfitting our models and mistaking idiosyncrasies in our data for real-world trends. Finally, framing our results in terms of predictive validity helps to bridge the gap between academics and policy practitioners, positioning quantitative models as a kind of analogy to the work that policy and intelligence analysts do routinely. Rather than asking policy professionals to interpret a set of regression results, we can express our findings in terms of how well we were able to make predictions beyond the data used to construct our models.

[7] We are sympathetic to, but do not address here, the larger critique of this analytic approach and its reliance on statistical significance as an indicator of a worthy result.

Below, we describe our analysis of the out-of-sample predictive validity of six major theories of nuclear proliferation, including the variables we use and our modeling and cross-validation strategy.

Data and variables

Our data are structured as a pooled time series with the country-year as the unit of analysis. Each observation in the data represents an individual state in a given year. Our data run from 1945—the dawn of the nuclear age—to 2010. The dependent variable in this analysis is a dichotomous measure of whether a state begins a nuclear weapons program in a given year, using data updated from Jo and Gartzke (2007).[8] Following some in the quantitative nuclear literature, we drop all observations from the data once a state has begun a nuclear weapons program (or acquired a weapon). The state reenters the dataset after its program ends (or its nuclear weapons have been given up), and it is once again at risk of pursuing nuclear weapons. This data structure makes for a much harder test of the predictive power of theories of nuclear proliferation, because the best predictor of whether or not a state will have a nuclear weapons program next year is whether a state has a nuclear weapons program this year. By eliminating from the data any hints provided by the presence of an existing program, we make prediction significantly more difficult.

We use three variables to measure a state's latent nuclear weapons *capability*. First, we use the country's real GDP per capita to capture its overall level of national resources (Gleditsch 2002). Second, we use a seven-point composite measure of latent

[8] Using alternative pursuit dates from Singh and Way (2004) yields similar results.

nuclear capacity from Jo and Gartzke (2007). This measure runs only through 2001, however, so we assume for the purposes of this analysis that states maintain the same level of underlying nuclear capacity from 2001 through 2010. This assumption is probably not far from the truth—the individual components of the Jo and Gartzke measure, such as uranium deposits, change infrequently. Third, we use a dichotomous measure of nuclear production capability—a dummy variable equal to one when a state produces any electricity from nuclear power in a given year (World Bank 2015). States with civilian nuclear infrastructure may find it far easier to pursue development of nuclear weapons.

To capture *conflict behavior*, we use a simple five-year moving average of the number of militarized interstate disputes (MIDs) in which a state was involved (Ghosn, Palmer, and Bremer 2004). Alternative operationalizations, including the count of MIDs over the last five years, three years, and one year, and dummy variables representing MID involvement over these timeframes, produce roughly the same results.

Two variables represent theories of *nuclear assistance* in our analysis. First, we use Fuhrmann's (2009) measure of civilian nuclear cooperation—the cumulative number of nuclear cooperation agreements signed by a state. Second, we also include in our models a measure of multilateral nuclear assistance. This is equal to the number of fuel cycle-related technical cooperation projects—administered by the International Atomic Energy Agency—in which a state participated in a given year (Brown and Kaplow 2014).

Nuclear rivalry may also help to predict proliferation. We include in our models a dichotomous variable set to one if a state had a rival with a nuclear weapon in a given year, using rivalry data from Thompson and Dreyer (2011). Alternative variable specifications substitute the presence of a rival with a nuclear weapons program—rather than one that has already obtained at least one nuclear weapon—but the results are the same.

To capture the role of *alliances* in predicting proliferation, we adopt a dichotomous variable set to one if a state has a defense pact with a nuclear state in a given year, using alliance data from the Correlates of War Project (Gibler and Sarkees 2004).

We follow common practice in this literature and measure the impact of *institutions* with a dummy variable set to one if a state is a member of the NPT in a given year. NPT membership data is drawn from Carcelli et al. (2014).

To facilitate comparisons with existing models of nuclear proliferation, we also include in our models a cubic polynomial of the number of years since a state last pursued nuclear weapons or, for those states that have never had a weapons program, the number of years since 1945. This variable—equivalent to the peace years variable often used in studies of international conflict—addresses temporal autocorrelation in the logit and probit models frequently used in this literature (Carter and Signorino 2010). For our purposes, this variable also has a useful substantive interpretation as a stand-in for the history of a state's nuclear adventures. The behavior of states that recently engaged in nuclear weapons development may rightly fall under more scrutiny than those that have been good nuclear citizens since the start of the nuclear age.

Statistical learning model

We test the predictive accuracy of these variables using a statistical learning model—the support vector machine (SVM). An SVM represents data as points in multidimensional space, 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). Different statistical learning methods enjoy different levels of predictive accuracy depending on the underlying data structure. We make no representation about the relative effectiveness of SVMs versus other common statistical learning methods such as neural networks or random forests; we limit ourselves to SVMs for simplicity. Future work could apply alternative statistical learning models or supplement our SVMs with a model-averaging strategy such as committee methods or bootstrap aggregation, in which predictions are averaged across bootstrap samples to reduce variance (Hastie, Tibshirani, and Friedman 2001).

Statistical learning approaches are less frequently employed in political science, but are commonly used in computer science and statistics. 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. The pursuit of nuclear weapons is a rare event. Thus even the strongest drivers of proliferation probably exert relatively little influence on proliferation decisions in the vast majority of cases. But in states that are at high risk of proliferating—that is, in the cases we care most about—these factors may matter a great deal.[9] The linear regression

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

models and their close relatives (such as logit and probit) that are used most often in the quantitative nuclear proliferation literature are not flexible enough to capture complex non-linear relationships that are likely to be present in these data.

Cross-validation and predictive metrics

We examine predictive accuracy using a 3-fold cross-validation procedure of the kind commonly used in the computer science and machine learning literatures (Arlot and Celisse 2010). We divide our data randomly into three parts.[10] Two parts of the dataset are used as training data to construct the model, while the third part is used to test the predictive power of the model. The divided data is then shuffled and the process repeated, so that each piece of the original dataset serves once as the test data. To avoid any bias introduced by the initial division of the data, we repeat this entire process 10 times using different random subsamples and average the results.

The fact that nuclear proliferation is an exceedingly rare event—only about 0.3 percent of observations in our data correspond with initiation of a nuclear weapons development program—complicates this process in two important ways. First, 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 we select, 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 (Sun, Wong, and Kamel 2009).

[10] 2-fold cross validation yields similar results.

While there is no single solution to this problem, we can mitigate the issue by oversampling the rare class in the data or undersampling the prevalent class. Here, we do both, adopting an algorithm known as SMOTE (synthetic minority oversampling technique) (Chawla et al. 2002). This algorithm works by adjusting the training data for our models. 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 testing data that is set aside and then used to evaluate predictive accuracy for the model.

Rare events also complicate the task of measuring predictive success. Overall accuracy—how many of the test data points the model successfully predicted—is a poor measure of success when rare events are involved, because the model can achieve very high levels of overall accuracy without providing any leverage against the problem of interest. A model that always predicts that states will not proliferate would accurately predict more than 99 percent of cases in our dataset, but such a model is not particularly useful as a guide to countering proliferation.

As an alternative, we adopt two metrics that are sensitive to changes in the model's ability to predict proliferation—rather than just non-proliferation. First, we examine the area under the ROC curve (AUC) (Swets 1988). 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. A perfect model, one that correctly predicts all cases, will have an AUC of 1.

As a second metric for predictive success, we use the F_1 score. This metric balances two elements of the model's predictions. The first is the model's positive predictive value, the share of "yes" predictions that turn out to be correct. The second is the model's 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 the feature—nuclear proliferation—that we care most about.

Testing predictive power

Our overall results show fairly strong predictive power for the full model of proliferation. The SVM with all variables included has an average AUC of 0.76. This compares quite favorably to a naïve model made up just of the cubic polynomial of time since the last weapons program, which has an AUC of 0.15. But the AUC metric includes model performance in predicting negative cases—non-nuclear pursuit—as well as positive cases. A less rosy picture is presented by the F_1 score, with an average value of 0.05 across the full model.

These metrics are more meaningful when the contributions of individual theories of nuclear proliferation are considered. Table 1 shows the results for each metric when adding a given set of variables into an otherwise fully specified SVM. That is, each row

Table 1: Predictive power of theories of nuclear proliferation

	AUC	*Percent of AUC*	*F_1 score*	*Percent of F_1 score*
Full model	0.764	100%	0.053	100%
Capability	0.022	2.96%	0.004	8.16%
Conflict behavior	0.035	4.80%	0.006	12.77%
Nuclear assistance	0.009	1.19%	0.003	6.00%
Rivalry	0.031	4.23%	0.011	26.19%
Alliances	0.017	2.28%	0.005	10.42%
Institutions	-0.022	-2.80%	0.001	1.92%

Average results from 10 runs of a 3-fold cross-validation procedure using a support vector machine. Values indicate the change in the specified metric when adding the given variables to an otherwise fully specified model.

can be interpreted as the contribution that an individual theory makes to the out-of-sample prediction of nuclear weapons programs. These values are also expressed as a percentage of the total metric, and can be seen as the increase (or decrease) in predictive performance we gain from adopting that theory of nuclear proliferation in our analysis.

The stand-out predictor among these theories is the presence of a nuclear rival. Adding rivalry to the SVM increases the AUC by more than four percent and the F_1 score by more than 25 percent, on average. While nuclear rivalry has not fared well in omnibus quantitative tests of the correlates of proliferation, here it excels. This is not unusual—the factors that matter most for proliferation in an absolute sense may not be those that make the greatest contribution to efforts to make predictions out of sample. This finding suggests that the nuclear behavior of rivals should continue to be adopted as a valuable indicator of proliferation risk, and also points to the continued relevance of concerns about proliferation cascades in the Middle East and elsewhere.

The predictive power of alliance-based theories of nuclear proliferation also is somewhat surprising given the mixed findings of the wider quantitative proliferation literature. Adding a measure of nuclear umbrellas to the SVM increases the F_1 score by more than ten percent.

Conflict behavior is also a significant predictor of state decisions to seek nuclear weapons. Considering a state's recent history of dispute involvement increases AUC by almost five percent, on average, and the F_1 score by more than 12 percent. This finding is more consistent with the existing quantitative literature on nuclear proliferation, which sees a strong correlation between conflict history and propensity to seek nuclear weapons.

The predictive power of supply-side theories of nuclear proliferation is weaker in our tests than other theoretical approaches. These factors seem to be less useful in distinguishing between high-risk cases of proliferation. Measure of latent capability—GDP, nuclear capacity, and nuclear power production—add only three percent to AUC and 8 percent to the F_1 score. And measures of nuclear assistance—civilian nuclear cooperation agreements and multilateral nuclear aid—do even worse by these metrics. These findings are interesting given the collective focus in the policy community on issues of nuclear latency and the risks associated with indigenous nuclear infrastructure in countries of concern. Of course, supply-side factors probably matter a great deal in terms of a state's decision to seek weapons, but in this analysis they fall short as effective indicators of nuclear interest.

Finally, institutional theories of nuclear non-proliferation do especially poorly in our predictive tests. Adding a variable representing NPT membership to the SVM actually diminishes its ability to predict nuclear pursuit according to the AUC metric, and has little effect on the F_1 score. This speaks directly to the NPT's inability to effectively signal state intentions one way or the other. In some ways, the treaty is a victim of its own success. Because nearly every state is now a member of the institution—only India, Pakistan, Israel, North Korea, and newly independent South Sudan remain outside the treaty—it does not tell us much about the nuclear intentions of its members.

Conclusion

Unlike other quantitative studies of nuclear proliferation—which primarily seek to identify causal links between some explanatory variable and state decisions to seek nuclear weapons—this project focuses on the value of different factors as *indicators* of proliferation risk. In doing so, it provides a testing ground for comparing extant theories of proliferation and asking which is most useful in predicting proliferation out of sample.

This approach can be useful to policymakers because it mimics the real-world analytic process. We train our analysis on the body of data that we have available, then use that analysis to predict future events. Predictive validity is a kind of substantive significance—a way to ensure that our findings are not chasing the noise in our data, but rather represent some relationship that exists in the real world. Predictive models excel at showing us which factors are best able to distinguish between the borderline cases of proliferation—the cases of greatest policy concern. These factors are not usually the most obvious drivers of nuclear weapons programs; it is the more subtle and conditional relationships that are likely to matter most in accurately predicting proliferation. These factors become strong candidates for new indicators of proliferation that might be used by the intelligence and policy communities.

Our analysis need not end with these simple metrics of predictive success, however. The output of our predictive model is not merely a set of yes-or-no predictions. Instead, of each state in each year, the model assesses the probability that the state will pursue nuclear weapons under a particular set of conditions. This feature of the predictive model allows for a range of useful analysis in future work. For example, we might investigate the dogs that did not bark—states that the model assesses had a strong likelihood of proliferating but which did not actually choose to seek weapons. Further examination might reveal that these states share particular characteristics that help to mitigate the risk of proliferation. We might also look in more detail at the model's false negatives—the states that the model did not think would proliferate, but which ultimately sought nuclear weapons. Subjecting these cases to extra scrutiny might help analysts to better understand assumptions that are not being made explicit in the study of proliferation risk.

References

Arlot, Sylvain, and Alain Celisse. 2010. "A Survey of Cross-Validation Procedures for Model Selection." *Statistics Surveys* 4: 40–79.

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

Bell, Mark S. 2015. "Examining Explanations for Nuclear Proliferation." *International Studies Quarterly*.

Bleek, Philipp C., and Eric B. Lorber. 2014. "Security Guarantees and Allied Nuclear Proliferation." *Journal of Conflict Resolution* 58(3): 429–54.

Braun, Chaim, and Christopher F. Chyba. 2004. "Proliferation Rings: New Challenges to the Nuclear Nonproliferation Regime." *International Security* 29(2): 5–49.

Brown, Robert L., and Jeffrey M. Kaplow. 2014. "Talking Peace, Making Weapons IAEA Technical Cooperation and Nuclear Proliferation." *Journal of Conflict Resolution* 58(3): 402–28.

Carcelli, Shannon, Erik Gartzke, Rebecca 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*.

Carter, David B., and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3): 271–92.

Cirincione, Joseph. 2008. *Bomb Scare: The History and Future of Nuclear Weapons*. New York: Columbia University Press.

Douglass, Rex W., and Alexander Lanoszka. 2015. "Two Warheads Passing in the Night: On Modeling Nuclear Proliferation." Paper presented at the annual meeting of the American Political Science Association, 3-6 September, San Francisco, CA.

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

Gartzke, Erik, and Matthew Kroenig. 2009. "A Strategic Approach to Nuclear Proliferation." *Journal of Conflict Resolution* 53(2): 151–60.

———. 2014. "Nuclear Posture, Nonproliferation Policy, and the Spread of Nuclear Weapons." *Journal of Conflict Resolution* 58(3): 395–401.

Ghosn, Faten, Glenn Palmer, and Stuart A. Bremer. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21(2): 133–54.

Gibler, Douglas M, and Meredith Reid Sarkees. 2004. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816–2000." *Journal of Peace Research* 41(2): 211–22.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46(5): 712–24.

Haggard, Stephan, and Marcus Nolan. 2012. "Engaging North Korea: The Efficacy of Sanctions and Inducements." In *Sanctions, Statecraft, and Nuclear Proliferation*, ed. Etel Solingen. Cambridge, UK: Cambridge University Press, 232–60.

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): 167–94.

Kaplow, Jeffrey M. 2013. "State Compliance and the Track Record of International Institutions." *Paper prepared for the 2013 Annual Meeting of the American Political Science Association*.

Kreps, Sarah E., and Matthew Fuhrmann. 2011. "Attacking the Atom: Does Bombing Nuclear Facilities Affect Proliferation?" *Journal of Strategic Studies* 34(2): 161–87.

Kroenig, Matthew. 2009. "Importing the Bomb: Sensitive Nuclear Assistance and Nuclear Proliferation." *Journal of Conflict Resolution* 53(2): 161–80.

Miller, Nicholas L. 2014. "The Secret Success of Nonproliferation Sanctions." *International Organization* 68(04): 913–44.

Perkovich, George. 2006. "The End of the Nonproliferation Regime?" *Current History* 105(694): 355–62.

Potter, William C. 2010. "The NPT & the Sources of Nuclear Restraint." *Daedalus* 139(1): 68–81.

Reiss, Mitchell. 1995. *Bridled Ambition: Why Countries Constrain Their Nuclear Capabilities*. Washington, D.C.: Woodrow Wilson Center Press.

Rublee, Maria Rost. 2009. *Nonproliferation Norms: Why States Choose Nuclear Restraint*. Athens, GA: Univ of Georgia Press.

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

———. 2011. “The Causes of Nuclear Weapons Proliferation.” *Annual Review of Political Science* 14(1): 225–44.

Sauer, Tom. 2006. “The Nuclear Nonproliferation Regime in Crisis.” *Peace Review* 18(3): 333–40.

Singh, Sonali, and Christopher R. Way. 2004. “The Correlates of Nuclear Proliferation: A Quantitative Test.” *Journal of Conflict Resolution* 48(6): 859–85.

Solingen, Etel. 1994. “The Political Economy of Nuclear Restraint.” *International Security* 19(2): 126–69.

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.

Thompson, William R., and David R. Dreyer. 2011. *Handbook of International Rivalries: 1494-2010*. Washington, DC: Congressional Quarterly Press.

Way, Christopher, and Jessica L. P. Weeks. 2014. “Making It Personal: Regime Type and Nuclear Proliferation.” *American Journal of Political Science* 58(3): 705–19.

Wesley, Michael. 2005. “It’s Time to Scrap the NPT.” *Australian Journal of International Affairs* 59(3): 283–99.

Williams, Joshua, and Jon B. Wolfstahl. 2005. *The NPT at 35: A Crisis of Compliance or a Crisis of Confidence?* New York: Carnegie Endowment for International Peace.

World Bank. 2015. *World Development Indicators*. http://www.worldbank.org/research.