Predicting Rogue State Missile Launches

Thomas J. Brailey

January 2020

1 Introduction

Why do rogue states publicize their testing of missiles? Oftentimes, missile tests in states such as North Korea, Iran, Iraq, and Pakistan, highlight issues with their current capabilities or end in failure , . As such, these states cannot credibly commit to challenging the hegemon in a conflict, so publicizing missile tests with such frequency seems counterintuitive. Nevertheless, rogue states are often faced with international condemnation, sanctions, and reciprocal weapons testing, suggesting that these missile tests are, to some degree, credible. Part of the threat is, of course, that rogue states' technology will improve and, at some future point in time, will pose a credible threat to their opponents. Following that argument, given the frequency of testing and the apparent high failure rate, it would seem that rogue states may be better off conducting fewer tests with a higher success rate if they wish to show that their technology is indeed improving, yet as the data show, missile testing has increased in frequency and by number of launches at once despite low success rates. I argue that missile testing can be used as a signaling tool outside of interstate conflict and outside of a context of war; that they are also used as a signaling strategy amidst political events such as United Nations Security Council resolutions and visits by heads of state of other nations. We can predict that missile tests are more likely to occur following a sanction or around the time of a head of state visit. Missile testing is thus a way of signaling displeasure at resolutions made toward the rogue state and a way of presenting power during ceremonious and diplomatic

1

events on a bilateral level. As such, rogue states are still showing their military capabilities, but are doing so within the context of state diplomacy. I build a novel dataset focusing on four states generally considered to be "rogue" and conduct rigorous statistical analyses, the results of which show support my hypotheses.

2 Overview

There exists a great deal of literature focusing on the strategic logic and potential outcomes of signaling, especially with regards to states of the international community vis-à-vis one another and vis-à-vis the rogue states themselves (Gartzke et al., 2017; Handberg, 2016; Crawford, 1982; McManus, 2016; Fearon, 1997; Fearon, 1995; de Guevara and Kühn, 2011). Signaling techniques in this context revolve around the balance of power between states, where signals of strength present states with more favorable bargaining conditions. Before advancing, it is worth defining some terms. Central to the paper, missile testing is the firing or controlled explosion of rocket-propelled munitions. There are several types of missile which will be discussed in the paper and in the dataset, the most common of which are SRBMs (short-range ballistic missiles), MRBMs (medium-range ballistic missiles), ICBMs (intercontinental ballistic missiles), and SLVs (satellite-launched vehicles). All these weapons can be nuclear-fitted, so when discussing sanctions, nuclear sanctions and missile test sanctions will be used interchangeably. Moreover, based on existing literature, I characterize states such as North Korea, Iran, Iraq, and Pakistan as so-called "rogue states", states that "show contempt for international norms by repressing their own populations, promoting international terrorism, seeking weapons of mass destruction and standing outside the global community" (Henriksen, 2001).

For states such as North Korea, Iran, Iraq, and Pakistan, missile testing and launches in a context of war fall under the bargaining model as "signaling to demonstrate resolve" (Fearon, 1997; Gartzke et al., 2017). Though these states are perhaps less developed economically and

militarily than their opponents, and though they are not credibly committing to, or signaling, war, missile testing forces other states to acknowledge the rogue state's capabilities. Various country-specific reports suggest missile testing is a means to an end. That, in the case of Pakistan, the "change in the variety and tempo of launches seems to suggest that Pakistan wants to have in place a credible missile-based deterrence as soon as possible," and in the case of North Korea, missile tests indicate that areas in the United States could become vulnerable (Chandrashekar et at., 2006; Newhouse, 2001; Merrell and Abrahams, 2019). While it cannot be disputed that the overarching goal of any missile program is to serve as a deterrence and to ensure security, the missile tests themselves are studied only through the lens of crisis bargaining. They do not provide information on potentially important political events occurring in close temporal proximity which could serve to explain the timing and number of tests themselves. Further, very little, if any, academic literature exists on in-depth analysis of missile testing and missiles programs in rogue or developing states; most of the information is limited to speculative news reports . While this information is helpful and provides case study evidence which can be used to support my argument, it does not allow for any form of quantitative analysis.

3   Hypotheses

Missile testing can be used as a signaling tool for state diplomacy, outside the context of a large-scale, interstate war. While often characterized by the media as a threat of war against the hegemon, missile testing can be a reactionary signal of displeasure or a display of power coinciding with political events, such as sanctions or diplomatic events. To put it more plainly, my concept is the act of testing a missile and my conceptual definition posits that rogue states are more likely to conduct missile testing around important political and diplomatic events. As such, the unit of analysis is year/month/country/missile test. The missile test variable will take a value of 1 if a missile test occurred in that month, 0

if otherwise. As mentioned, I operationalize the concept of a diplomatic event as either a United Nations Security Council resolution against the country in question and head of state visits. Heads of state visits are either the head of the rogue state visiting another country, or the head of state of a foreign country visiting the rogue state. The operationalization of this concept will be discussed further later in the paper. For robustness, I include several other variables including lagged variables for the UNSC resolution and a dummy variable for whether the rogue state was in the midst of a crisis or war. Lastly, when hand-coding the political event columns, I made sure not to include UNSC resolutions that were in direct response to a missile test, thus ruling out the possibility of reverse causality.

One issue faced with time-varying variables is the arbitrary parameterization of time-intervals. Put differently, how does one quantify the difference between a missile test occurring on the last day of one month versus the first day of the next month. If a missile test falls within the first month then it may turn out to be statistically significant, though a missile test that occurred a day later may be put into a different month and return no statistical significance. To combat this issue, I generated lagged variables to account for these arbitrary categories of time.

It is worth noting that, as is the case with most quantitative studies in international relations, my research design is an observational study, drawing from a range of academic texts, government reports, and the like. While I am unable to exercise control over the assignment of my "Y" variable—as is the case in a controlled experiment—I am not faced with the same ethical and monetary costs associated with running an experiment or conducting a survey. That said, neither form of experiment would lend itself particularly well to the study in hand.

Given the existing literature on signaling tactics exhibited by rogue states, a number of related hypotheses may be formulated. H1 posits that the timing of missile tests is predictable based on the occurrence on a political event.

$\mathbf{H_1}$: If a political event that directly addresses a rogue state occurs, that state is more

likely to exhibit some form of weapons test in close temporal proximity.

In reality, one cannot prove this hypothesis as there are myriad event types which are beyond the scope of this paper. As such, the initial hypothesis has been separated into two sub-hypotheses (H1a and H1b). Initially, the event column was aggregated to include all events deemed relevant to the study and that could be collected within a reasonable time. Having this aggregated variable was an ecological fallacy and would essentially mean that X would almost always take the value of 1, as there is almost always a political event occurring each month. By disaggregating the event type, one can analyze whether specific event types exert more of an effect on a state's likelihood to conduct a missile test. Given the time constraints of my data collection, I identify two types of event—United Nation Security Council Resolutions and heads of state visits—that I believe may provide explanation for rogue state missile testing, H1a and H1b.

H_{1a}: If a United Nations Security Council Resolution passes that directly addresses a rogue state, that state is more likely to exhibit some form of weapons test in close temporal proximity.

H_{1b}: If a either a head of state visits a rogue state, or the rogue state's head of state visits another country, the rogue state is more likely to exhibit some form of weapons test in close temporal proximity.

Given that binary dependent variables are limited in the models that can be run on them—at least within the scope of this paper—I added a count variable which looks at the number of missiles fired during each test. This leads us to our alternative hypothesis:

H_2: Larger displays of capability, characterized as multiple missiles fired during one test, are more likely to occur in close temporal proximity to a political event.

The null hypothesis (H0), we can assume that there is no predictive ability of significant political events on a rogue state's proclivity to conduct missile tests.

H_0: There exists no discernable relationship between a political event and missile tests conducted by rogue states.

Here we expect no statistically significant results from our models (p ¿ 0.1).

4   Data

Initial data analysis began on a database of North Korean missile launches developed by the James Martin Center for Nonproliferation Studies (CNS) . A database on Iran, again from the CNS, was joined . Data on missile launches in Pakistan and Iraq were hand-coded, drawing on tables and qualitive reports. From there, the dependent variable was expanded out using the tidyr package in R, changing the unit of analysis from launch/day to country/year/month/missile test. The event and crisis columns were all hand-coded and were informed by United Nations reports, academic papers, and historical documents. As the data stand, there are 1728 observations, four countries, and data from 1984 through to 2019. This dataset has the potential to include more information on political events, missile specifications (as well as other forms of capability signaling), additional countries, and other relevant information. Thus, although this database is in its infancy, it has the potential to become a valuable information source for the field of comparative international relations. The layout of the data as it stands will be discussed further in the limitations section. That said, by combining multiple countries' missile launch data, it is currently the only dataset of its kind.

5   Methodology

5.1   Core Model

Given that the dependent variable is binary our core model is a multivariate logistic regression. This model shows us the rogue states' latent propensity to conduct a missile test and looks at the marginal effect on the log-odds of a one-unit change in our "x" values. The equation for our logistic regression can be written as such:

$$logit(missiletest) = B_0 + B_1(Year) + B_2(Country) + B_3(UNSCResolution)+$$

$$B_4(HeadofStateTravel) + B_5(HeadofStateVisit) + B_6(Crisis) + E \tag{1}$$

"p" is defined as the probability that Y is equal to 1, and "E" is our error term. Our beta-one variable will show us if there is a relationship between the passage of time and the number of missile tests over time, which could be a confounding variable. Beta-two addresses another confounder; whether there is a statistically significant difference between the number of missile tests and the rogue state itself. Betas-three, -four, and -five, are my core independent variables which I seek to establish association with "p". Beta-six represents whether being in the midst of war increases the log-likelihood of engaging in missile testing.

5.2   Supplementary Models

As mentioned, having a binary dependent variable is somewhat constraining with regards to the type of analyses that can be run. Further, the difference between the outputs of logit and probit models is negligible given the nature of the data. I felt it necessary to conduct additional analyses to confirm my hypotheses as well as to better understand my count dependent variable. This variable shows not only when the missile tests occurred, but also how many missiles were fired, as is the basis of H2. I ran a base ordinary least squares regression, followed by three other variants which focused on the interaction effects of my crisis and year variables. From there I ran a heteroscedasticity check and reran my initial linear model with robust standard errors. I then ran a Poisson model to account for the existing heteroscedasticity in my data. Lastly, as is the case with quantitative international relations, it is difficult to prove that our observations do not suffer from contagion. It is safer to assume that, to some degree with some of our cases, certain events influence others. As such, I ran a negative binomial model to account for this possibility. The findings of these supplementary models, including visualizations and regression tables, can be found in the

appendix section.

6   Findings

Our models point to some interesting results. Firstly, we see our year variable exhibiting a statistically significant association with the occurrence of missile testing. This confirms the various reports in our literature review that indeed the frequency of missile tests have been increasing over the years. With regards to the countries in our dataset, we see that being Iraq or Pakistan reduces decreases the log-odds of conducting a missile test by 3.509 and 1.337 respectively. This seems logical given that these two countries have the most missing values—they would likely level out to 0 if these variables had similar completeness to the Iran and North Korea variables. It seems that based on our logistic regression, we can confidently reject at least part of hypothesis H1b in that head of state travels do not seem inherently related to missile testing. Though head of state visits to a rogue state increases the log-odds of conducting a missile test by 0.584, this is only statistically significant and the one-star (p ¡ 0.1) level. Whether a rogue state is in the middle of a conflict seems to exert a very weak negative effect on missile testing, falling outside the range of statistical significance. The issue with this interpretation is that, given the hostile behavior of these states, they are almost always in the midst of some form of war or conflict. This is clear in the data as the crisis variable almost always takes the value of 1. To overcome this, and conduct relevant comparisons, one would need to include states which are not as frequently in conflict. For example, by incorporating non-rogue polities, we could better establish the effect conflict has on missile testing, though this is outside the scope of this paper. However, for the purpose of this paper, a non-statistically significant result helps to prove my initial theory, as I argue that signaling methods can exist outside of a context of war. Because being in a crisis does not exert an effect on missile testing, signaling strategies can occur outside of the traditional capability-displaying theory.

It seems that there exists a statistically significant (at the three-star level; p ¡ 0.01) relationship between UNSC resolutions and missile tests. Further, as our lagged variables do not exhibit statistical significance, it seems we can infer that these missile tests were conducted as an immediate response to the resolutions. A UNSC resolution increases the log-odds of a missile test by 1.411. Moreover, though not reported in the regression output , the p-value for this independent variable is 1.81e-08 which allows us to conclude that this result is indeed statistically significant. Thus, we can, with confidence, reject our null hypothesis that there is no statistically significant relationship between political events and missile tests conducted by rogue states.

In this model, the log-likelihood and Akaike Information Criterion (AIC) values do not give much insight into the effectiveness of the model. Having run several iterations of the logistic regression including interaction effects and the omission of variables, the log-likelihood and AIC exhibited very little variation. The model presented in this paper gave the best log-likelihood value, indicating it was the "best" model. It is also worth noting, as is well visualized by the above plots, the latent propensities of rogue states to conduct missile tests is very small. As such, given the data we have, it is difficult to determine whether these differences in latent propensity matter, but this will be discussed further in the limitations section of the paper.

While this will not be discussed in great depth in the main portion of this paper, my count models all show similar results to my initial logistic regression, again providing support for H1a and H2. They also point to the statistical significance of the head of state visit variable. Further, given that I ran more count dependent variable models that I did binary dependent variable models, I have more fit statistics that help understand the effectiveness of each model.

6.1 Limitations

There are several limitations to this study, most of which arise from the sparseness of the data and the temporal scope of the research project. Though these limitations exist, I argue that they are not so severe as to discredit the study itself. I believe that these findings will only be bolstered by further data collection.

Many of the issues associated with the sparseness of the dataset. Given that the data on missile launches are few and far between for all countries, let alone the more secretive rogue states, it is hard to estimate how valid our findings are. If our data is not representative of the real world, we cannot claim our data has construct validity which reduces the credibility of our claims. Had I more time to conduct my data collection, I would have collected data on all countries. This would allow for a more reliable and valid prediction of strategic missile testing by rogue states. Given that my data come from reputable sources, and multiple models report similar outcomes across all my independent variables, I can assume that my data and methodologies are reliable. Though, in a bid to improve reliability, it would be beneficial to harness data from several sources to ensure that my findings are the same across sources. That said, this dataset is, as far as I am aware, the only one of its kind and may prove a useful resource in comparative studies of military equipment.

While a more complete dataset would undoubtedly be beneficial for the credibility of my study, one permanent limitation of my study is the inability to apply causal inference models to draw conclusions about the nature of missile testing strategy. In the context of my study, causal inference would be employed to look at the causal effect of one specific "X" variable which would either be UNSC resolutions or heads of state visits, or all political events. As mentioned, this would either be too specific or far too broad to make any valid causal claims. Moreover, implementing a causal design would benefit from having more data on all countries, as well as information on a whole range of variables relating to political events, military and technological capabilities, and other pertinent aspects of international relations. These variables would be difficult, if not impossible, to objectively quantify, and

the time required to do so is well outside the scope of this paper.

Similarly, the very nature of international relations and state strategic interactions prevents us from making truly causal claims. There exists myriad confounders and variables—outside those previously mentioned—which cannot be quantified nor accounted for in conventional parametric models. As I can only hope to achieve a strong predictive model, my findings have limited policy implications. That said, this is by no means the motivation behind the paper. Rather, I aimed to show that signaling strategies exist outside of a context of war, and that missile testing can be predicted by significant events that impact that state.

Lastly, as a result of the sparse data the results from the various models I ran show very small changes to the dependent variable. While some of these results are statistically significant, the true effect of each independent variable may well be larger but are masked by the high frequency of 0 and NA values in the dataset. As it stands, it is hard to gauge whether the effects exerted by each independent variable matter in the grand scheme of missile testing and signaling writ large. Again, the only way to improve on this aspect of the study is more data collection and a more complete dataset. That said, the results generated from my binary and count models do show strong relationships in the existing data which provides some optimism and impetus for future studies and replication.

7   Conclusion

Existing literature posits that signaling tends to occur between two states amidst some sort of conflict or crisis. These states, the argument goes, utilize techniques that display their military capabilities and technological advancement. The data that I have collected point to a novel use of signaling outside of the traditional interstate conflict literature. Signaling, specifically missile testing, can be predicted with a high level of confidence following resolutions made by the United Nations Security Council, and, to a lesser degree, visiting heads to state. Thus, we see signaling practices used outside of a context of war and instead within

the context of diplomatic relations. These findings are interesting to the extent that we are able to better understand strategies employed by rogue states that exhibit asymmetric disadvantage on the world stage. We see those states employ military-style signaling tactics when engaging in bilateral-relationship building and when being sanctioned by the international community. But what are the implications of these results? While this study and the data upon which it is based is in its infancy, I believe that further data collection and research will give a more in-depth understanding of how rogue states conduct diplomacy with the international community when faced with military, economic, and political asymmetry.

8 Appendix

8.1 Additional core hypothesis visualizations

8.1.1 Predicted versus expected values of logistic regression

8.1.2 Logistic regression coefficient values and confidence intervals

8.1.3 Residuals for logistic regression

8.2 Data visualizations

8.2.1 Summary of data

8.2.2 Summary of variable value frequencies (TestCount)

8.3 Supplementary models

8.3.1 Linear models

[1]

---

[1]Model (1) is the baseline model. (2) removes Crisis and Year columns. (3) is the Crisis:Year interaction effect excluding the interacting variables. (4) is the interaction effect plus the interacting variables.