A New Look at First Strike Stability

Harrison Schramm
Center for Strategic and Budgetary Assessments
1667 K Street Suite 900
Washington, DC 20006
UNITED STATES

schramm@csbaonline.org

ABSTRACT

During the "Cold War", analysts studied the strategic stability of their nuclear weapons postures, where preventing war was a coequal objective to winning. The First Strike Stability Index was developed to analytically compare the utility of different forces. Since the end of the Cold War, thinking about nuclear conflict in this particular construct fallen out of favor. In this note, we review the concept of First Strike Indices, propose a minor refinement, and provide an historical example. Our goal is to summarize concepts and explore some areas for future development, and provide some minimal worked examples.

1.0 INTRODUCTION

Maintaining a credible nuclear deterrent has been at the core of nuclear powers' military strategy since early in the nuclear era. *Deterrence* is defined as a coercive strategy that relies on holding things the adversary holds dear 'at risk' as a consequence for particular actions. The Cold War development of deterrence strategy has many seminal works, particularly those of Thomas Schelling (Schelling, 1980), who received the Nobel Prize, as well as Herman Kahn (Kahn, 1960).

Several events, specifically the approaching end of the New START Treaty (2021) and a shifting international landscape, have encouraged us to reopen this line of reasoning (Economist, 2018 and US Dept. of State 2018).

Credibly holding adversary positions and forces at risk is at the heart of most deterrence strategies and is in the background of strategies such as Massive Retaliation, Mutually Assured Destruction, Brinkmanship, and Eisenhower's "New Look".

Comparing the utility of strategic forces requires a means of comparison. Measuring the quality of a deterrent can be difficult; invariably it relies on an assumption that the values that an adversary (or ally) attaches to portions of their national power - such as military concentration areas, industrial and natural resources, and population centers - are knowable, quantifiable, and locally stationary. These are difficult assumptions under the most favourable conditions, and we will focus momentarily on two. To say that a situation is quantifiable implies that well-ordered value relationships can be established; specifically, that we can create a scaling by which we say that 'two of x is worth 1 of y'. This is particularly troublesome when the 'x' and 'y' are measured in terms of disparate elements of National Power.

The second assumption – local stationarity – implies that over 'reasonable' time periods, these well ordering relationships do not change. We are unprepared to accept this statement *a priori*. Over the arc of nations' budget cycles, which typically span on the order of 5 to 10 years, this may be true, but in an actual crisis, the relative worth of assets being scored may shift radically at the speed of conflict itself, which will be measured on the order of 30 minute increments. Any serious student of these models needs to be constantly aware that the timescales used in planning do not conform with the time scales used in operation.

The difficulty in measuring does not mean that an attempt should not be made. Trying to measure and gain insights in this topic is the focus of the remainder of this article. In this note, we follow the ideas first laid out by Kent and Thayler, who created the "First Strike Stability Index" (henceforth, FSSI). The core idea of the FSSI is to create an equilibrium point in a two player game, where neither side may improve their prospects unilaterally. We call this type of regime *stable*.

A stable deterrence regime may be best described by considering the following example: Suppose that two players, Alice and Bob, are the holders of nuclear weapons stockpiles, and each time period (day, say), they decide whether or not to launch their nuclear weapons without knowledge of the others' decision. Each assumes that initiating a war - a first strike - will result in their adversary immediately attempting to retaliate. If Alice believes that Bob has a (comparatively) large number of nuclear weapons, or that Bob has strategic 'holdouts' that will survive her onslaught regardless of her actions, she reasons that Bob is, in a sense, indifferent to whether he is attacked first or second. Following a similar line of reasoning, Bob also reasons that Alice is indifferent to being attacked first or retaliating, and deterrence - at least theoretically - is stable. The converse situation - where Alice reasons that she has an overabundance of weapons (comparatively) and that Bob's weapons are exposed - is deterrence unstable, and may lead either party to conclude that Alice's optimal strategy is a so-called "Splendid First Strike" against Bob. It is noteworthy that Bob's conclusion about Alice is as dangerous for Alice as it is for Bob.

This and similar lines of reasoning is why fully-developed Nuclear Powers either have or are seeking assured second strike weapons, and furthermore that the United Kingdom has moved to rely solely on the Ballistic Missile Submarine (SSBN).

Much of deterrence ideas from the beginning of the nuclear era to today fundamentally rest on the notion of 'strategic indifference' in one's adversary. Actions or periods of time which violate this assumption represent strategic risk, for both players.

Before proceeding, we would like to point out that one of the key reasons to build this sort of model is to f make

2.0 PREVIOUS WORK

The following summarizes the work of Kent and Thayler (Kent and Thayler 1990, Best and Bracken, 1995). These papers present the FSSI:

$$FSSI_{a,b} = \frac{C_{a,1}}{C_{a,2}} \cdot \frac{C_{b,1}}{C_{b,2}},$$

which says that the stability index between two nations a and b is the ratio of the costs of 'going first' (i.e. attacking) to those of 'going second' (i.e. defending). In the context of a Stackelburg Game By definition, the costs of going second are assumed to be higher than those of going first.

Kent and Thaler (referenced by Best and Bracken) Define cost as:

$$Cost_{self} = Damage_{self} - \lambda Damage_{enemy} + \lambda$$

Where "Damage is defined between 0 and 1. Thus cost is defined between 0 and $1 + \lambda$. Intuitively, cost to self is damage to self minus a bonus for damaging the enemy, plus the factor λ to make the cost of the most favorable situation 0" (Best and Bracken, 1995).

The 'kernel' of the model is:

$$\frac{c_{a,1}}{c_{a,2}} \cdot \frac{c_{b,1}}{c_{b,2}},$$

or, restated, the product of the ratios of the cost of going first divided by the cost of going second for each nation. It is assumed that the cost of going second is at least as great as the cost of going first.

Both players can increase the FSSI of the system by increasing the costs that their adversary takes by going first, or by decreasing their costs for going second. This first case is covered by having a credible second strike capability, and the second case is covered by dispersing or hardening their force.

3.0 SSI – AN ADJUSTMENT TO THE FSSI

Fundamentally, we seek to map a nations' multi-variate profile of their nuclear arsenal and value centers to a single 'score'. Many functions are possible, and finding optimal forms is an open area of research. For purposes of discussion we introduce the **Strategic Security Index (SSI)** as follows:

$$SSI = \frac{W - [W - F](1 - e^{-\eta\frac{A}{T}})}{W} = 1 - (1 - \frac{F}{W})(1 - e^{-\eta\frac{A}{T}})$$

Where:

W is a nations' total number of weapons, which in this case we count as individual missiles or bombs, not warheads.

F is the number of weapons that the nation considers invulnerable (or *F*ensed off). Missiles carried on an idealized SSBN may be a candidate for this class.

A is the number of credible *A*ttacks the adversary can mount against the nation's stockpile. Here we take this to be equivalent to the number of weapons; other factors are possible, such as decrementing for a reserve.

T is the number of independent locations or *T*argets that the Nation presents to its' adversary.

η is a scaling parameter, which corresponds to the number of targets that each attack can successfully eliminate. Note that the *natural* scaling, $\eta = 1$ corresponds to an ~80% effectiveness per weapon. η is an end-to-end performance number, and considers orders, launch, vehicle failures, missed targeting. It is not strictly required that $0 \leq \eta \leq 1$; for MIRV'd warheads or exceptionally large warheads, it is conceivable that one weapon may attack more than one target.

Keen readers will notice that the ratio of Attacks to Targets has an exponential form. Increasing eta tends to magnify the impact of the ratio A/T in the unadjusted index, see below.

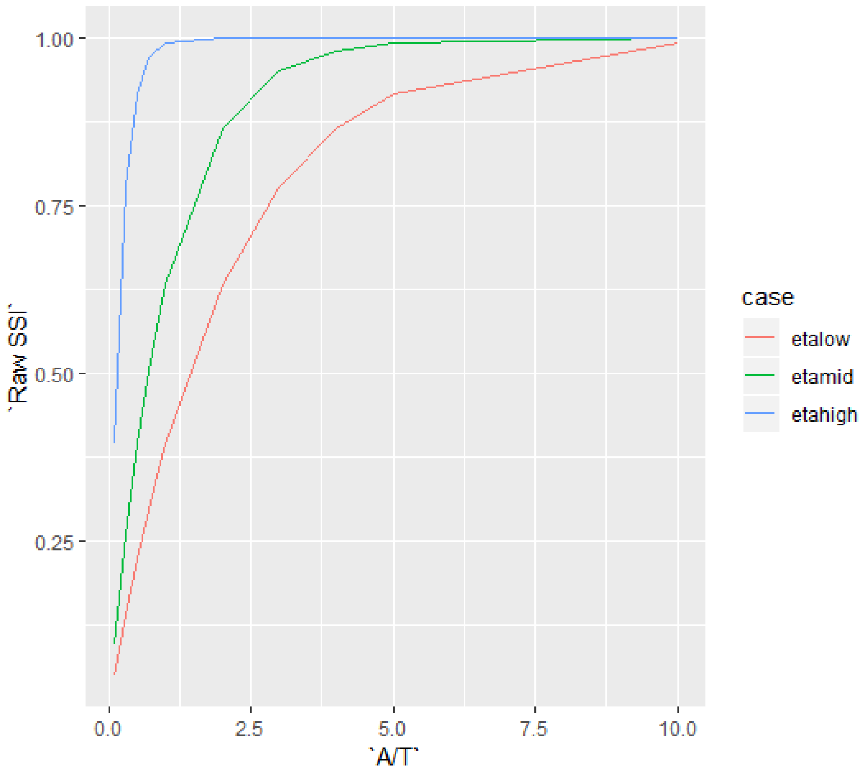

Figure 3.1 The impact of varying the parameter eta on the unadjusted SSI. Increasing Eta tends to increase the sensitivity to the ration of attacks to targets, in the direction of decreasing stability.

As with the FSSI, Nations with high SSI will tend to feel secure in their stockpiles, where nations with lower SSI will tend to feel less secure.

It should be noted that he SSI is a dimensionless quantity, and is only meaningful for making comparisons between different instances. Statements such as "An SSI of .5 means a 50% chance of nuclear war" are *not valid.*

We close this section by exercising our model against the Global Nuclear Weapons database, comparing the US and Soviet Union over the history of the nuclear age.

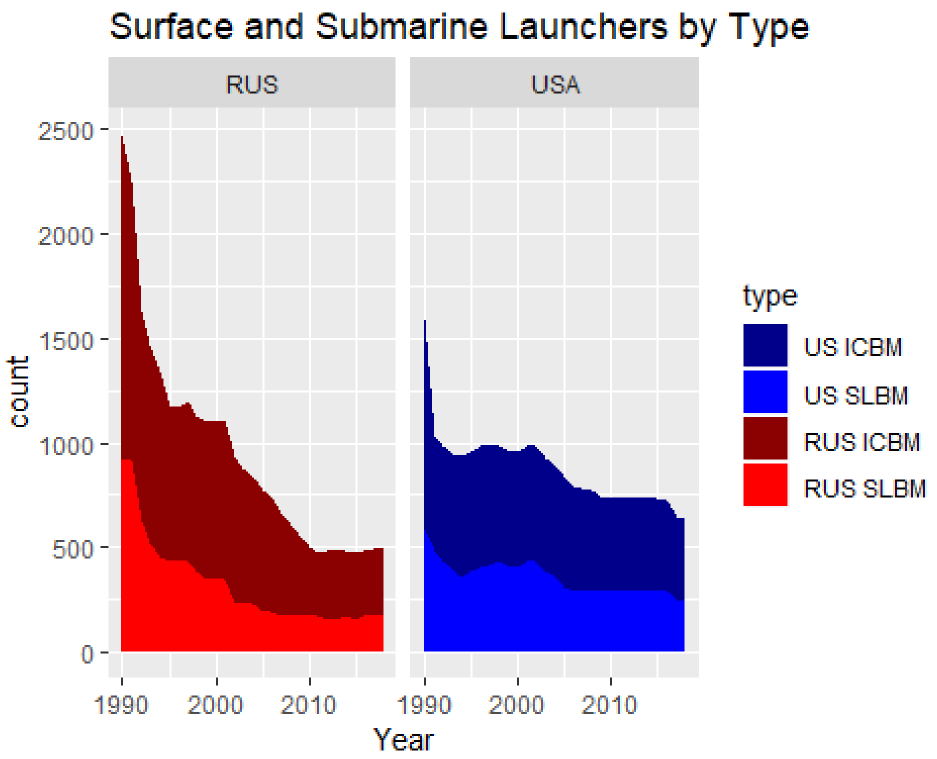

Figure 4-1: US and Soviet / Russian surface and submarine launchers by type. Since 1990, there has been a marked decline in each nation's total launchers for both types.

Using this data and Equation (1) above, we can compute the SSI for the United States and Russia over the past 20 years as follows:

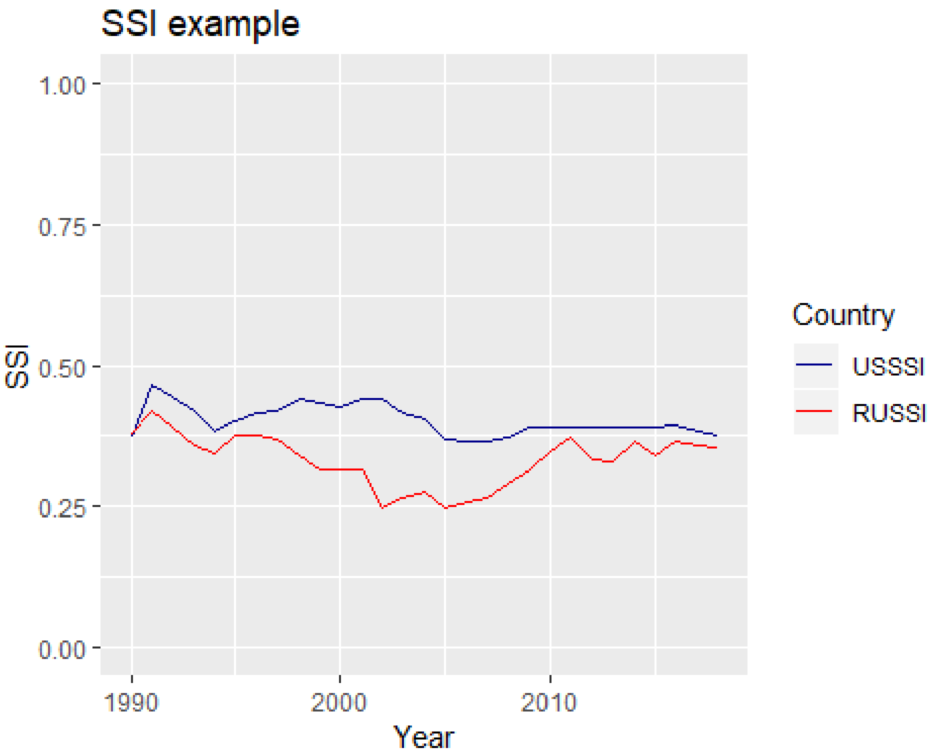

Figure 4-2: A simple exercise of the SSI model, showing the strategic 'index' for each nation in time. The United States index over this time period has been relatively stable, while the Russian index took a significant dip in the mid 2000's.

4.1 The Effect of Dispersion on Stability

Even toy models are useful. Looking at the form of Equation (1), we can see that diluting the number of weapons per target decreases the effectiveness of any independent attack and therefore increases stability.

In our above analysis, we assumed that each weapon was *independently* vulnerable to attack; i.e. that the amount of dispersion was equal to the number of weapons. In practice, this is almost never the case; weapons tend to be grouped in installations. For the next example, we consider for both the US and Russians in year 2010 the effect of varying the number of concentration areas on the Stability Index.

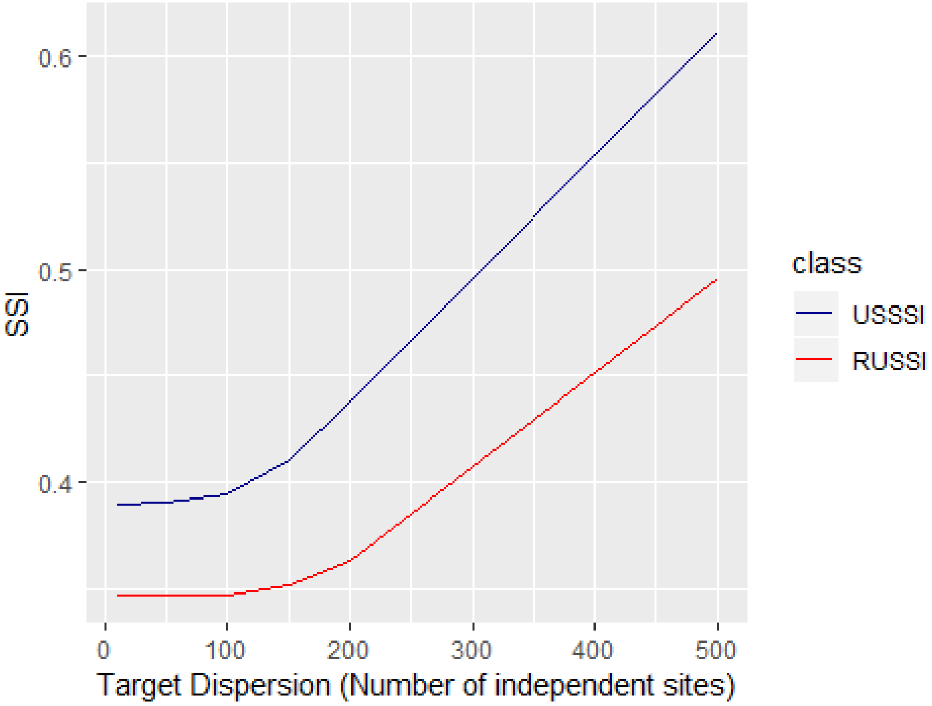

Figure 4-3: The impact of dispersion on stability. Increasing dispersion in this model increases stability. It may be possible for a nation to increase stability with dispersion vice increased weapons.

5.0 EXTENSIONS

We conclude two extensions to the model, dealing with two elements of real-world uncertainty; Shifting Alliances and the potential "Shock" of shifting technology.

5.1 Multipolarity and Alliances

Real competitions in a multipolar world with alliances are rarely monolithic, but instead form *cliques*. These cliques are centered around a 'patron' state, with the satellite clique members being willing - to a degree - to contribute to the defense of their patron, and conversely, to expect help in their defense. NATO and the Warsaw Pact could both be considered historical examples of this type of clique. One simple way to approach this situation is to `pool' both the resources (Weapons) and vulnerabilities (Targets) and treat the clique as a monolithic entity. In an hypothetical 'perfect alliance', this approach would be valid.

A more nuanced approach is to consider the contributions of each clique member, and consider the index from each clique member's point of view, assuming that each clique member remains an independent actor.

In this case, our model becomes:

$$SSI = \frac{W - [W - F](1 - e^{-\frac{\sum_j \eta_j \mu_j A_j}{\sum_i \xi_i \tau_i}})}{W},$$

Where

i is the index for each nation in the clique (including the *patron*),

j is the index for each nation in the Adversary clique,

W and F represent the Patron's Weapons inventory and Fenced component (as before),

μ is the relative liability from adversaries assumed by joining the clique, $0 \leq \mu \leq 1$.

η_j is the effectiveness of each Nation's arsenal in the adversary clique.

ξ_i is the advantage in dispersion - due to basing rights or cooperative agreements within the clique (number of targets each clique member presents).

This development is myopic from the each member's point of view, in that the other members of the clique contribute to her offensive capability but do not contribute to her vulnerability. An area for future study is the relative commitment that satellites make to the mutual defence compared with the risks they take on by joining the clique.

The central tension is the commitment that each clique member displays, as well as the other clique member's beliefs about this commitment. The difficulty with cliques is that the only way to be certain of any members' commitment is to actually test the alliance in war. The parameter η_j captures the clique's confidence in the reliability of each member, which is measured both by the reliability of their weapons, as well as their likelihood of contributing to the mutual defence.

A model like this may be useful in determining how a potential ally might increase or decrease the stability of an entire deterrence regime. Note that this is a more nuanced question than 'How will this potential ally increase my own security', because it takes into account the impact on the entire system.

As an example case, consider a (completely fictional) clique led by Rivendell, with other members Rohan and Bree. They are allied strategically against the combined armies of Isengard and Mordor.

Name	Dispersion	Number of Targets
Rivendell	1	10
Bree	.3	15
Rohan	.2	12
The Shire	*.5*	*10*

Name	Effectiveness	Number of Attacks	Commitment
Mordor	5	20	.1
Isengard	12	10	.4

Using equation (2) above, we compute the stability index of .41 without including The Shire in the clique, and .35 with The Shire. This indicates that including this member in the clique will have a slightly adverse effect on stability.

5.2 Shocks and "Strategic Surprises"

Another extension to the model involves the notion of *Strategic Surprise*, specifically, what happens if a series of weapons in the 'Fenced' or invulnerable category become - suddenly - vulnerable.

Here we are interested in measuring the 'Delta SSI' or change in SSI for a given time change, and perhaps find the epoch of greatest risk for technological development; recalling - from the introductory paragraphs - that instability as measured by the SSI may be as damaging to the innovator as to the laggard.

Symbolically, we would be interested in:

$$max_{t^*}(SSI(t^*_-) - SSI(t^*_+))$$

Where t^* denotes the time of maximum instability given the development programs of $_{a,b}$, and the subscripts $_{+,-}$ indicate the infinitesimal epochs preceding and succeeding the potential changes in technology. These 'maximization epochs' represent areas of heightened risk that may be offset by technological or policy actions.

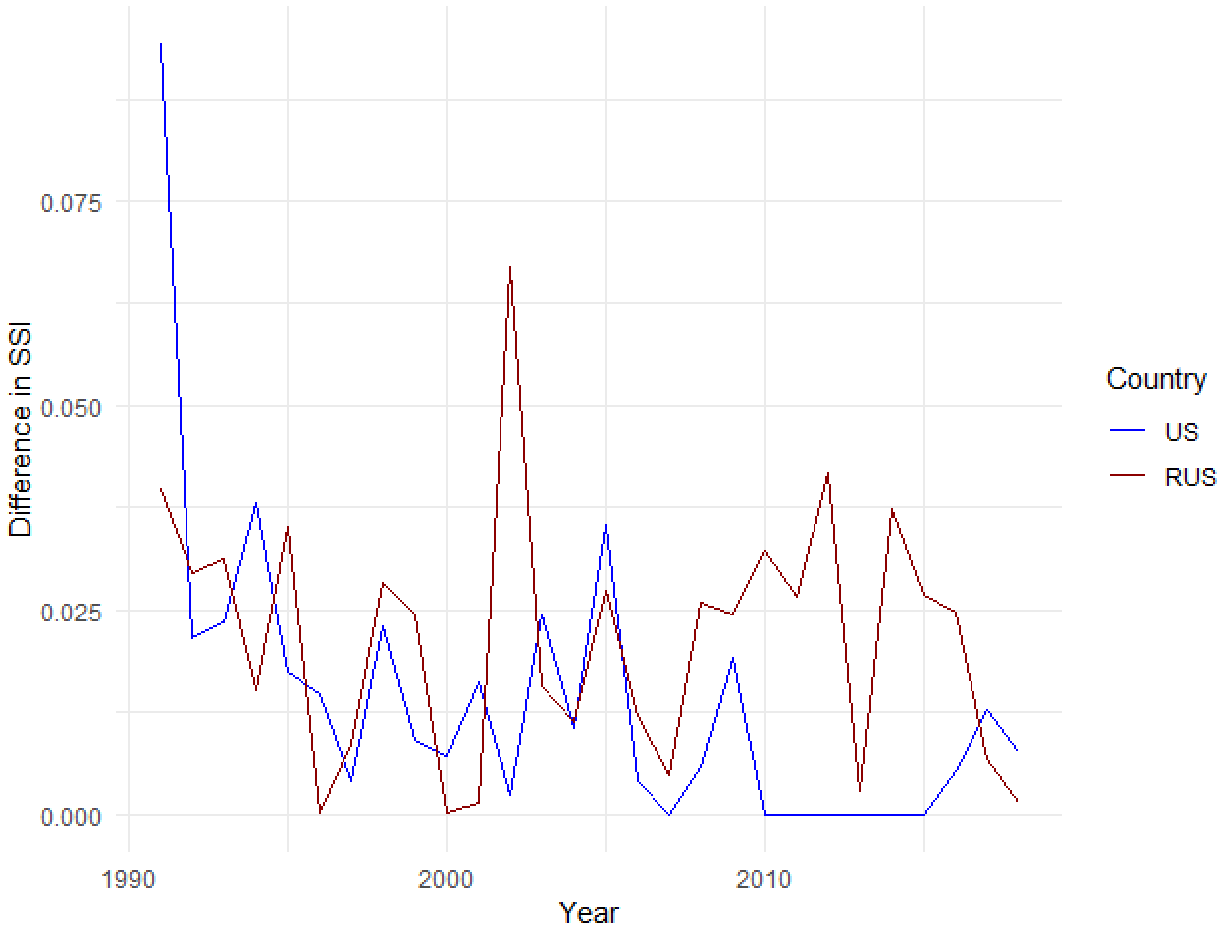

Figure 5-1: Differences in Strike Stability in time. In this example, the greatest change – and risk of technological surprise – occurred for Russia in 2002.

6.0 CONCLUDING REMARKS AND FUTURE WORK

In this short note we have presented a (minor) refinement to the First Strike Stability Index - the Strategic Stability index. This is by no means a complete story. The best that models like this can do is to aid in strategic thought by creating concrete measures for comparing different strategies. They also help decision-makers sketch out different policy alternatives in an environment which is internally consistent.

6.1 Future Work

We are interested in using public statement sentiment data, as derived from public policy statements and media reports during the Cold War, to determine analytically the values of η. Specifically, we are interested in knowing how sentiment 'leads' or 'lags' development, and cases where deploying fewer weapon may lead to a more stable deterrence regime.

REFERENCES

[1] Best, M., Bracken, J. (1995). "First Strike Stability in a Multipolar World." Management Science 41 (2) 298-321.

[2] Economist Magazine (2018). "A Farewell to Arms Control," May 5th issue. London, England.

[3] Kent, G., Ochmanek, D., Spirtas, M, and Prinie, B. (2008). "Thinking About America's Defense - an Analytical Memoir," RAND Corporation, Santa Monica, CA

[4] Kent, G. and Thaler, D. (1990). "First Strike Stability and Strategic Defenses," RAND Project Air Force Technical Report No. R-3918-AF. RAND Corporation, Santa Monica.

[5] Kahn, H. (1960). "On Thermonuclear War," Princeton University Press.

[6] Powell, R. (1989) "Crisis Stability in the Nuclear Age," The American Political Science Review. Vol 83, No 1, 61-76.

[7] Schelling, T. (1980) The Strategy of Conflict. Harvard University Press, Cambridge Mass.

[8] United States Department of State (2018). "New Start" Web Reference https://www.state.gov/t/avc/newstart/ retrieved 23 August 2018.