

No paradox here?

Improving theory and testing of the nuclear stability-instability paradox with synthetic counterfactuals¹

Francesco Bailo, School of Communication, University of Technology Sydney

and

Benjamin E. Goldsmith, School of Politics & International Relations, Australian

National University

(ben.goldsmith@anu.edu.au)

Abstract This article contributes to both the theoretical elaboration and empirical testing of the ‘stability-instability paradox,’ the proposition that while nuclear weapons deter nuclear war, they also increase conventional conflict among nuclear-armed states. Some recent research has found support for the paradox, but quantitative studies tend to pool all international dyads while qualitative and theoretical studies focus almost exclusively on the US–USSR and India–Pakistan dyads. This article argues that existing empirical tests lack clearly relevant counterfactual cases, and are vulnerable to a number of inferential problems, including selection on the dependent variable, unintentionally biased inference, and extrapolation from irrelevant cases. The limited evidentiary base coincides with a lack of consideration of the theoretical conditions under which the paradox might apply. To address these issues this article theorizes some scope conditions for the paradox. It then applies synthetic control, a quantitative method for valid comparison when appropriate counterfactual cases are lacking, to model international conflict between India–Pakistan, China–India, and North Korea–US, before and after nuclearization. The article finds only limited support for the paradox when considered as a general theory, or within the theorized scope conditions based on the balance of resolve and power within each dyad.

Keywords: Nuclear weapons, Interstate conflict, Synthetic control method, Nuclear deterrence

¹ Author accepted manuscript for *Journal of Peace Research*, accepted on 28 April 2021.

Biographical statements

BENJAMIN E GOLDSMITH, PhD in Political Science (University of Michigan, 2001); Professor, Australian National University (2017-).

FRANCESCO BAILO, PhD in Government and International Relations (University of Sydney, 2017); Lecturer, University of Technology Sydney (2019-); Recent book: *Online communities and crowds in the rise of the Five Star Movement* (Palgrave, 2020).

Replication data

Data and code for the analysis, along with an Online appendix, available at: <https://www.prio.org/jpr/datasets/> and <https://doi.org/10.7910/DVN/8HLD6E> .

Acknowledgements

Previous versions were presented at Tohoku University (2016), Midwest Political Science Association (2016), Pacific International Politics Conference (2017), Australian National University (2018), and Asian Polmeth (2019). We thank Jiro Akita, Patrick Brandt, Wilfred Chow, Songying Fang, Pedro Franco de Camps Pinto, Kosuke Imai, Koji Kagotani, Haille Na-Kyung Lee, Hatsuru Morita, Barry Oneal, Michael Tomz, and other participants for valuable feedback. We thank Aishu Balaji and Salma Refas for excellent assistance.

Motivation

This article advances theoretical development and testing of the ‘stability-instability paradox,’ a long-standing aspect of nuclear deterrence. The theory expects that while nuclear weapons deter nuclear war, they also lead nuclear-armed states to increase conventional military conflict with each other. ‘[T]he greater the stability of the “strategic” balance of terror, the lower the stability of the overall balance at its lower levels of violence’ (Snyder, 1969: 123; see also Jervis, 1984: 31–34; Zagare, 1992). A recent review finds a cumulation of evidence of ‘limited war’ among nuclear states (Geller, 2017: 12). But our findings suggest the evidence is far from conclusive, even considering scope conditions that sharpen the theory.

Important formal-theoretical and quantitative research appears to support the paradox (Powell, 2015; Rauchhaus, 2009). Other recent work on nuclear security tends to accept it as an established fact, incorporating the paradox as a foundation for related empirical or theoretical investigation (Kydd, 2019; Watterson, 2017). However, at least one quantitative study (Bell & Miller, 2015) does not find support, and some qualitative studies of India and Pakistan deliver equivocal assessments (Ganguly & Hagerty, 2006; Kapur, 2005, 2007; Narang, 2010).

We argue that studies of the stability-instability paradox (hereafter SIP) using traditional quantitative and qualitative methods face significant challenges for valid inference. Gartzke and Kroenig (2017: 1853) call for ‘[m]ore precise theory’ and ‘improved methods of inference’ in the study of nuclear security issues. In this article we clarify theoretical scope conditions for SIP, reflected in two hypotheses based on Powell’s (2015) formal model, and use a newer method across three appropriate cases. Only one of our cases fits the general expectation. There are several challenging aspects of applying synthetic control to interstate conflict data, but it has distinct advantages and adds essential new evidence to the debate.

We first discuss the existing literature on SIP, focusing on barriers to valid inference, and address theoretical foundations. We then describe research design, present empirical analysis for three cases for which we have sufficient data, and conclude.

Barriers to inference

Empirical tests of SIP are vulnerable to several inferential problems. A concern for qualitative studies is selection on the dependent variable (Geddes, 1990). Two dyadic cases dominate the literature: Soviet Union–United States and India–Pakistan. It is hard to imagine these cases were *not* chosen because they appear to confirm the theory. After the USSR acquired nuclear weapons in 1949, the Cold War escalated. After Pakistan became a nuclear power in 1990, tensions over Kashmir increased; the 1999 Kargil War erupted after nuclear tests in 1998. These cases have contributed to the theory’s development and survival (Cohen, 2013; Montgomery & Edelman, 2015), but selection on the dependent variable can lead to misplaced inference and inattention to alternative explanations.

Qualitative studies may also suffer from unintentionally biased interpretation. Motivations behind specific military-security decisions – notoriously difficult to confirm – are key pieces of evidence. If interpretation is consistent with expectations, this can lead to false-positive findings, a type-I inferential error.

Analysts of India–Pakistan relations debate whether the nuclear balance causes Indian restraint and Pakistani belligerence. Evidence may be weighted because it appears to improve the explanatory power of a preferred variable, despite other plausible explanations. A case in point is two prominent authors’ different inferences about the role of international opinion in India’s restraint in the Kargil War.

Kapur (2008: 77) writes: ‘V.P. Malik, Indian Army chief of staff during the Kargil operation, explains that the Indians avoided crossing the Line of Control mainly out of concern

for world opinion: “The political leaders felt that India needed to make its case and get international support” for its position in the conflict. The Indian government believed that it could best do so by exercising restraint even in the face of clear Pakistani provocations.’

But Ganguly (2008: 59) argues: ‘One seemingly plausible explanation for India’s restraint might have been the perceived need to court international public opinion in the aftermath of the nuclear tests. This argument does not withstand scrutiny, however. Given the resentful public mood within the country, an upcoming general election, and the existence of a regime that had few qualms about the use of force to resolve disputes, the inhibitions of global public opinion could not have served as a powerful barrier to the expansion of the conflict. [...] Consequently, even in the absence of incontrovertible public statements, through a process of inference and attribution, one can make a cogent argument that the principal source of Indian restraint was Pakistan’s overt possession of a nuclear arsenal. Indian policymakers, cognizant of this new reality, were compelled to exercise suitable restraint for fear of escalation to the nuclear level.’

Adjudication between such competing assertions does not seem possible given the available evidence. As Leng (2006: 116) writes, ‘[t]here is no solid evidence’ to support particular claims about Indian restraint, only ‘attributional distortion.’

Similarly, qualitative inference may be attempted from evidence that does not logically warrant it. Kapur (2008: 74–75) appears to treat Pakistani Prime Minister Benazir Bhutto’s claims in this way, writing: ‘Pakistani leaders have openly acknowledged nuclear weapons’ emboldening effects. Benazir Bhutto, who served her first term as Pakistani prime minister from 1988 to 1990, stated [...] “Islamabad saw its capability as a deterrence to any future war with India,’ [...] because ‘a conventional war could turn nuclear.’” Logically, acknowledging a nuclear deterrent does not justify inference of ‘emboldening effects.’ Since she was not in power during major instances of emboldened Pakistani behavior such as the 1987 initiation of

the Kashmir insurgency or the Kargil War, causal inference from her statements is further open to question.

Although the evidence seems indeterminate, qualitative empirical support for the paradox hinges on just such judgments, inferring the sources of Indian restraint and Pakistani belligerence. While such over-confident inference is not inherent to qualitative analysis, this review of influential articles suggests it does affect the debate.

Common quantitative methods also have shortcomings for testing SIP. They can address selection problems by pooling dyads, and large-n data are typically more consistently coded. However, in spite of its general framing, the paradox seems to apply to a limited number of cases (for example, dyads with the potential for both conflict and nuclear-weapons acquisition). Without delineation of scope, extrapolation from irrelevant cases is likely, artificially reducing standard errors and distorting findings. Interstate conflict and nuclear weapons acquisition are relatively rare events. The vast majority of dyads will be non-nuclear and non-conflictual; inclusion in analysis might not be theoretically appropriate (Table I)². The likely distortion would be a misleadingly pacific non-nuclear counterfactual, again leading to type-I errors.

² *MID level* is our conflict indicator developed from the Militarized Interstate Dispute hostility and severity indicators, detailed in the Data section

Table I. Conflict rates among nuclear, mixed, and non-nuclear dyads, 1954-2010

	State 1	State 2	Number of nuclear states in dyad	MID level median	MID level mean	MID level max	Dyad-years
5 nuclear dyads with most conflict	IND	PAK	2	3.33	3.54	5.92	21
	USA	CHN	2	0.00	1.09	4.50	47
	USA	USR RUS	2	1.00	1.00	4.50	57
	USA	PAK	2	0.00	0.93	2.67	21
	USR RUS	CHN	2	0.00	0.80	4.50	47
5 nuclear dyads with no conflict	UKG	FRA	2	0.00	0.00	0.00	51
	USA	FRA	2	0.00	0.00	0.00	51
	FRN	CHN	2	0.00	0.00	0.00	47
	UKG	ISR	2	0.00	0.00	0.00	44
	USR RUS	IND	2	0.00	0.00	0.00	23
All non-nuclear dyads (n=13,813)			0	0.00	0.01	7.00	513,622
All mixed dyads (n=1627)			1	0.00	0.04	7.00	51,484
All nuclear dyads (n=44)			2	0.00	0.25	5.92	1115

Notes: we have chosen 5 of the longest duration and most geographically representative dyads with no conflict, from among the 25 nuclear dyads with no conflict in the dataset.

But general arguments about SIP lead to hypotheses like the following: ‘Symmetric nuclear dyads are more likely to experience conventional war and low-level conflict than are nonnuclear dyads’ (Bell & Miller, 2015: 76) and ‘The probability of crisis initiation and limited uses of force between two states will increase when both states possess nuclear weapons’ (Rauchhaus, 2009: 263). Each compares nuclear dyads to all others, which is inappropriate if most dyads are outside the theory’s scope.

Average conflict levels among relevant categories clarifies both the importance of scope conditions and pitfalls of extrapolation (Table I). Non-nuclear dyads and those with only one nuclear state have much lower average conflict than jointly nuclear dyads. But it is unlikely that most such dyads, like Tanzania–Turkey or Russia–Rwanda, provide relevant counterfactuals. Many jointly nuclear dyads also may fall outside the theory’s scope, such as allies like France–United Kingdom or those with no history of conflict like China–Israel.

Rauchhaus's study is path-breaking in not selecting on the dependent variable and clearly measuring 'low' and 'high' conflict levels. He finds support for the paradox: '[e]vidence suggests that while nuclear weapons promote strategic stability, they simultaneously allow for more risk-taking in lower intensity disputes' (Rauchhaus, 2009: 258). However, the analysis uses all dyads in the international system, likely extrapolating from irrelevant cases, and does not distinguish between nuclear states that are rivals, allies, or states with no conflict history.

Rauchhaus's findings therefore are based equally on France–United Kingdom and US–USSR, and on dyads unlikely to acquire nuclear weapons or experience conflict. Should we make a prediction for nuclear Botswana–Costa Rica? In pooled dyadic analysis, this case has equal weight to that of India–Pakistan, contributing to counterfactual inference as a non-nuclear dyad. Bell and Miller address some shortcomings in Rauchhaus's study, including controlling for previous conflict, and find that jointly nuclear dyads are not more conflict-prone. However, this does not address the issues of scope conditions and extrapolation.

Cognizant of the inferential challenges, we develop scope criteria based in part on our interpretation of the formal theory of Powell (2015). Our use of synthetic control (Abadie, Diamond & Hainmueller, 2010, 2015; Abadie & Gardeazabal, 2003), discussed below, allows us to focus on three relevant cases with sufficient pre- and post-treatment time-series data for analysis, while increasing the rigor of comparative inference. Using the language of experiments, synthetic control predicts the post-treatment difference between a variable of a given unit and a hypothetical untreated version of the same variable, estimated by 'synthesizing' multiple partially relevant units. It is appropriate for testing a counterfactual proposition when there are few or no valid comparator cases available – precisely the difficulty empirical studies of SIP have faced.

Hypotheses

It is a truism that we should be glad to lack empirical evidence of the causes of nuclear war, and to have little evidence of conventional fighting between nuclear powers. Nevertheless, as for study of the consequences of nuclear weapons in general (Gartzke & Kroenig, 2016), study of SIP has suffered from this paucity of cases and evidence.

SIP was developed with one case primarily in mind, the US–USSR Cold War dyad. The theory originated with Snyder (1969). Jervis (1984)³ and Glaser (1990) made central contributions.⁴ It posits a trade-off between the degree of strategic stability at the level of all-out nuclear war, and the likelihood of conventional conflict. Strategic (in)stability is the risk of conflict *escalation* to all-out nuclear war. Snyder (1969: 123) hypothesized, ‘stability in the strategic nuclear balance tends to destabilize the conventional balance.’ The idea was core to Cold-War debates about US conventional and nuclear postures, the extent to which ‘MAD [mutual assured destruction] makes the world safe for conventional war’ (Glaser, 1990: 224), and therefore the need for enhanced conventional deterrent and/or greater strategic instability through, for example, more emphasis on tactical nuclear weapons (Snyder, 1969: 120).

The other frequently studied case is India–Pakistan. As Kapur (2005: 127–128) writes: ‘Most scholars attribute ongoing violence in the region to a phenomenon known as the ‘stability/instability paradox.’ According to the paradox, strategic stability, meaning a low likelihood that conventional war will escalate to the nuclear level, reduces the danger of

³ Jervis (1984: 154–55) ultimately argues conventional instability is not exacerbated under strategic stability because the challenger could not be sure ‘the situation would remain under control.’

⁴ Snyder’s essay originally appeared in 1965. Henry Kissinger (1957) also wrote on the topic (see Geller, 2017: 7).

launching a conventional war. But in lowering the potential costs of conventional conflict, strategic stability also makes the outbreak of such violence more likely.’

We take the hypotheses noted above from recent studies (Bell & Miller, 2015; Rauchhaus, 2009) to represent the standard SIP proposition, and restate it as our first hypothesis. We define ‘nuclear dyads’ as those in which each state is believed to possess nuclear weapons that pose a plausible deterrent to the other.

Hypothesis 1. Nuclear dyads will experience conventional conflict at a level substantially above the non-nuclear counterfactual.

Considering the existing literature, it appears that each theoretical or qualitative study tends to focus on a single dyad, while quantitative analyses include all. Each approach ignores fundamental issues of the theory’s scope, invoking some version of Hypothesis 1. But Powell’s (2015) game-theoretic analysis addresses SIP in ways that imply limiting conditions. We use it as a guide.

Balance of resolve, balance of power, and SIP conditions

Powell addresses SIP in a model that formalizes the role of conventional military balance in nuclear brinkmanship. Our propositions represent our interpretation, not formal hypotheses he developed. We discuss scenarios that inform the two additional hypotheses we are able to test (the Online appendix elaborates further hypotheses).⁵

⁵ Future investigation could explore temporal conditions. Will nuclear dyads eventually reach a new equilibrium (e.g., US–USSR détente)?

In Powell's model, a nuclear-armed challenger state decides how much conventional force to use in conflict initiation with a nuclear-armed defender. The more conventional force the challenger uses, the higher likelihood of achieving its objectives, provided conflict does not escalate to nuclear war. However, by using more conventional force, the challenger simultaneously increases the chance of escalation to nuclear war. The defender must choose whether to respond in a way that further risks escalation. The confrontation is both a contest of nuclear resolve, and a conventional fight. The balance of forces on the ground, and thus the expected outcome of a conventional conflict, matter for nuclear-risk calculations.

Rather than assuming, as nuclear deterrence theory usually does, that conventional military balance between two nuclear adversaries would not affect nuclear confrontation and the logic of MAD, Powell explicitly models conventional military balance as an element affecting nuclear risk and deterrence.

The concept 'balance of resolve' characterizes each side's relative commitment to the issues at stake. Powell (2015: 590) argues: 'states in the midst of a nuclear crisis frequently appear to face a fundamental trade-off between bringing more military power to bear and raising the risk of escalation to nuclear war. When deciding whether or not to escalate, a state can often take steps that more fully exploit its military capabilities and potential. This increases the chances of prevailing if any subsequent fighting remains limited and the conflict does not escalate to a catastrophic nuclear exchange. But these steps also make it more likely that the crisis will ultimately end in this way.'

The conventional balance can change each side's perceived risk, allowing anticipation of a conventional conflict's likely outcome. This can indicate whether one state will be backed into a corner, caught between a devastating or existentially threatening conventional outcome, and the option of escalating to nuclear conflict. This chance of 'situations that are so dire that the states are sure to escalate' makes knowledge of the conventional balance important,

transforming nuclear brinkmanship from a pure contest of resolve, to one in which the balance of resolve interacts with the balance of conventional capabilities. ‘The model’s defining feature is that the amount of power the challenger brings to bear affects the stability of the conflict’ (Powell, 2015: 598–599).

Powell (2015: 590) identifies two conventional-conflict levels and suggests that the risk of nuclear war will increase conflict at ‘low’ but not ‘high’ levels. He does not specify criteria, although his examples of low-level conflict include the 1999 Kargil War, which reaches the standard war threshold of 1000 battle deaths.⁶ That is, a low-level conflict might be restrained, rather than all-out, conventional war. Powell (2015: 610) explains: ‘A high degree of instability deters India from bringing its conventional superiority to bear against Pakistan. This in turn enables Pakistan to pursue low-level conflict which otherwise would be deterred by the threat of a large-scale Indian conventional retaliation.’

This highlights the importance of the counterfactual: whether nuclear rivals will use more conventional force against each other *than they would have if they had not acquired nuclear weapons*. When ‘there is no existential threat’ and conventional capabilities and resolve determine outcomes, challengers are potentially willing to use their full military capability in conflict if they are more resolved than defenders; but challengers are deterred if defenders are more resolved and the balance of power does not favor the challenger (Powell, 2015: 607). Even if the weaker challenger is more resolved, ‘[i]n the absence of any risk of [nuclear] escalation’ a stronger defender can use its full capabilities to defeat the challenge without worrying about transforming the conflict into one of nuclear resolve (Powell, 2015:

⁶ Uppsala Conflict Data Program states: ‘fighting claimed over 1000 lives’ (accessed 16 November 2017 at <http://ucdp.uu.se/#/statebased/422>). Kapur (2008: 74), also states that it crossed this threshold.

612). This ‘highlight[s] in a very simple way some of the incentives a weak state has to “go nuclear” and thereby be able to transform a contest of strength into one of resolve’ (Powell, 2015: 612).

Situations in which the challenger will use little or no force include when both challenger and defender have low resolve over an issue. A challenge is unlikely, and any degree of nuclear instability would further reduce the risk of conventional conflict. With low resolve, the greater the risk of escalation, the lower the challenger’s force level, even when the balance of power favors the challenger: ‘a strong but less-resolute challenger brings less and less power to bear as the situation becomes less stable [...]. Indeed, if the level of instability is high enough, the chances of prevailing [...] are too low to make the use of force worthwhile and there is no challenge’ (Powell, 2015: 610).

But conditions of imbalanced resolve invite greater conventional conflict. Specifically, imbalanced resolve coupled with power imbalance induces dangerous dynamics: the weaker state either fears exploitation, in the case of low resolve, or, if highly resolved, perceives the opportunity to use strategic instability to achieve gains it could not achieve through conventional means alone. A clear prediction of Powell’s model is that, relative to the non-nuclear counterfactual, weaker but more resolved challengers will be emboldened to use force. According to Powell (2015: 610, 612, 614) this scenario fits the India–Pakistan confrontation over Kashmir, in which Pakistan is more resolved on the issue, but India enjoys an advantage in conventional capabilities, as well as the US ‘in a confrontation with a nuclear-armed “rogue” state’ such as North Korea.

To organize the implications of this logic parsimoniously, we focus on the dyadic distributions of resolve and power. We identify three possible conditions for the balance of

resolve within a dyad: symmetric low resolve, symmetric high resolve, and asymmetric.⁷ The balance of conventional military power can be characterized as either symmetric (balanced) or asymmetric (imbalanced). Interaction gives six possible dyad-types (Table II). Beyond general Hypothesis 1 above, we present hypotheses for two scenarios we can test given available data (shaded): asymmetric power and resolve, and symmetric low resolve.⁸ The ‘paradox’ is only expected to emerge in the former scenario, while the latter falls outside the scope of SIP.

Table II. Conditioning scenarios for the stability-instability paradox

		<u>Balance of resolve</u>		
		Symmetric (low)	Symmetric (high)	Asymmetric
<u>Balance of power</u>	Symmetric	Less conventional conflict.	Equal or less conventional conflict.	Equal or more conventional conflict.
	Asymmetric	Less conventional conflict.	Equal or more conventional conflict.	More conventional conflict.

Hypothesis 2. Nuclear dyads with asymmetric resolve and an imbalance of conventional power will experience conventional conflict at a level substantially above the non-nuclear counterfactual.

Hypothesis 3. Nuclear dyads with symmetric low resolve will experience conventional conflict at a level below the non-nuclear counterfactual.

⁷ Powell distinguishes theoretical challengers and defenders, roles that are not fixed for real-world dyads; we assume either state can potentially take either role.

⁸ For additional hypotheses see Online appendix.

We test Hypothesis 2 on India–Pakistan and North Korea–United States, and Hypothesis 3 on China–India. While we are not able to test other relevant dyads, which lack pre-treatment data, we present time-series conflict-level graphs for three in the Online appendix: China–US, China–USSR, and US–USSR. It is questionable whether *any* conforms to the pattern expected by the general theory. Specifically, conflict increases for US–USSR and China–US immediately after US nuclear acquisition in 1945, but in neither case is there a clear additional jump after the dyad becomes jointly nuclear, and both eventually decline substantially. Nuclear asymmetry between China and the USSR, from Soviet acquisition in 1949 to PRC acquisition in 1964, marks a low-conflict period, increasing substantially when the dyad becomes jointly nuclear, potentially supporting SIP expectations. But Sino-Soviet tensions subside after two decades, and are not markedly different from the dyad’s pre-1949 conflict levels. Thus only a selective and shorter time-series provides descriptive support for SIP.

Research design

Method

Synthetic control (Abadie, Diamond & Hainmueller, 2010) is a comparative approach to the analysis of intervention (treatment) effects, allowing inference about a binary intervention in observational time-series data when there is a single treated unit and no valid counterfactual (control) unit. An estimate of the counterfactual unit is instead ‘synthesized’ from a weighted combination of non-treated units. Political-science applications include the impact of political shocks on economic outcomes (Bove & Nisticò, 2014; Dorsett, 2013; Horiuchi & Mayerson, 2015; Sanso-Navarro, 2011) and democratization (Kennedy, 2014); ours is the first to interstate conflict we know of.

Among the strengths of the method are its lack of reliance on assumptions about the data generating process (Samartsidis et al., 2019) and ability to account for bias from omitted time-variant covariates (Abadie, Diamond & Hainmueller, 2010: 495, 504; Billmeier & Nannicini, 2013: 984, 987; Costalli, Moretti & Pischedda, 2017: 82–83). Criticisms focus on violation of assumptions, such as random assignment of the treatment (Samartsidis et al., 2020), or questions about predictor selection (Botosaru & Ferman, 2019; Ferman, Pinto & Possebom, 2020). However when a reasonably long time-series is available, and there is only a single treated unit, synthetic control (with sensitivity analysis) is appropriate and there is no other generally superior method (Samartsidis et al., 2019).

Acquisition of (deterrent) nuclear weapons by the second state in a dyad is the treatment inducing SIP. We must model the post-treatment counterfactual, but appropriate comparators are hard or impossible to identify. Specifically, we want to model the conflict level in the post-treatment period, absent the treatment (e.g., a non-nuclear India–Pakistan from 1990). No existing dyad is likely to be similar enough, nor is the pre-treatment dyad because it is not possible to rule out other post-treatment changes that might affect conflict. Neither would all dyads with only one nuclear state be a valid set of counterfactual cases. Most will include a second state that is either not likely to acquire nuclear capabilities and/or not a rival of the nuclear state (Table I). Nuclear dyads can also emerge quickly, as did India–Pakistan, while the relevant history of rivalry can stretch back decades. Synthetic control allows us to create an artificial comparator case from a broader dataset, such that a set of variables predicts very similar conflict patterns pre-treatment for both the real and synthetic dyads.

The synthetic conflict variable is estimated from a weighted combination of characteristics of non-treated units placed in a ‘donor pool.’ We use the Synth R package to do this (Abadie, Diamond & Hainmueller, 2011). The characteristics of the treated unit are combined in a single-row matrix X_0 , which includes the pre-treatment observed values of the

independent variables and the mean of the dependent variable. The characteristics of the non-treated units are similarly combined in matrix X_1 with as many rows as donors. Synth estimates a vector of weights W such that $\|X_1 - X_0 W\| V$ is minimized, computing two vectors of weights: W describes the relative contribution of each control unit to the synthetic unit, while V describes the relative predictive power of the independent variables. Generally, vector V should minimise the synthetic estimator's mean error. We choose V to minimise mean squared prediction error over the entire pre-treatment period (Nash, 1979).

The main advantage of the method is that comparison is not between the treated dyad and a specific (but inappropriate) non-treated dyad, but between the treated dyad and an optimal (in terms of matching pre-treatment conflict behavior) combination of non-treated dyads. Our donor pool for each treated dyad is a selection of untreated dyads with similar conflict frequency in the pre-treatment period (details below and in Online appendix).

We include both non-nuclear dyads and dyads in which one country had nuclear weapons as potential donors. This produces superior pre-treatment prediction, and the pre-treatment dyad includes both conditions for China–India and India–Pakistan.

Cases

Three nuclear dyads a) have a security rivalry, and b) provide sufficient pre-nuclearization time-series data to successfully implement synthetic control: India–Pakistan, China–India, and North Korea–US. For each, we identify when they became jointly nuclear, and categorize both balance of power and the balance of resolve. Based on recent studies (Jo & Gartzke, 2007; Gartzke & Kroenig, 2009; Rauchhaus, 2009), India and Pakistan's rivalry went nuclear in 1990, when Pakistan became a nuclear state, and India and China became a nuclear dyad with India's 1988 acquisition of the capability.

The case of North Korea is more complicated. In 2002 US Secretary of State Colin Powell stated ‘We now believe they have a couple of nuclear weapons and have had them for years,’ but it is not clear whether he believed these were usable or meant only for testing. North Korea declared itself a nuclear state with a ‘nuclear deterrent’ in 2005 and conducted its first, small, test in 2006. We acknowledge it was not until a 2013 test that Pyongyang more convincingly demonstrated capability for nuclear warheads deliverable by missile (Kristensen & Norris, 2018: 45). However, given the uncertainty around its capabilities and its plausible ability to threaten to use nuclear force, for example against US troops in South Korea, we assess that it acquired nuclear weapons capable of posing a modest but non-ignorable deterrent to US action by 2006.⁹

Categorizations of resolve and power, to which we now turn, indicate which hypothesis is relevant. We assess that India–Pakistan and North Korea–US are asymmetric in both resolve and power, predicting greater conventional conflict after nuclear acquisition by Pakistan and North Korea (Hypothesis 2), while China–India has symmetric low resolve over the issues at stake (and asymmetric power), predicting less conventional conflict after 1988 (Hypothesis 3).

Balance of power

Assessing military balance we consider overall capabilities, but acknowledge that an effective balance might exist even if one state has greater overall capabilities. Paul (2006: 13), discussing India and Pakistan, contends that the local balance of power is especially important for nuclear dyads because conventional war is likely to be short. We categorize a dyad as having relatively symmetric power balance when the weaker state has greater than two thirds (66%) of the stronger’s capabilities. All three dyads are imbalanced by this criterion. Only China–India

⁹ Robustness tests assess alternative nuclearization dates.

approaches balance, with India having on average 47% of China's capabilities, as measured by the Correlates of War composite index of national capabilities (Singer, Bremer & Stuckey, 1972).¹⁰ By contrast, Pakistan has on average less than 20% of India's capabilities, and North Korea less than 5% of US capabilities. Choosing a lower threshold, say 40% or 34%, would not affect our conclusions because our prediction for China–India with symmetric balance of power and symmetric low resolve would be unchanged under Hypothesis 3.

Balance of resolve

Powell defines resolve as 'the highest risk of an all-out nuclear war [a state] would be willing to run in order to prevail;' 'How much risk a state could credibly threaten to run would depend on what was at stake [...].' (Powell, 2015: 592, 594).¹¹ It is important to avoid characterizations of resolve derived from accounts of the behavior we seek to explain – conflict patterns when both states possess nuclear weapons. We therefore consult sources on the general history of bilateral relations, including those available in the *Oxford Research Encyclopedia*, accounts of country and area generalists, and accounts written prior to nuclearization. A full discussion is provided in the Online appendix.

We find that Pakistan should be more resolved in confrontations over Kashmir because the issue is considered directly related to an existential threat for Pakistan, but not India. The border dispute associated with China–India military confrontations is of secondary importance to each after the 1960s, indicating mutual low resolve. While defending South Korea remains

¹⁰ Version 4.0 accessed 1 June 2016 at: <http://www.correlatesofwar.org/data-sets/downloadable-files/national-material-capabilities-v4-0>.

¹¹ We prefer Powell's definition since we base our hypotheses on his theory. Kroenig (2013) argues resolve should also consider expected costs of nuclear war.

important for the United States, the issue at stake has for decades been the status and freedom of action of the North. The balance of resolve tilts in favor of North Korea.

While we acknowledge that these judgements might be questioned, the practical impact for our analysis would likely be to further weaken support for the scope conditions, i.e., if India–Pakistan were judged to have symmetric high resolve or China–India were judged to have asymmetric resolve. Only if North Korea–US had symmetric high resolve would support for the scope conditions improve. We believe it is clear that the issues at stake for North Korea involve regime survival, unlike the US. DPRK leaders can credibly claim there is little difference for them between losing a conventional war with the US and starting a nuclear one – the regime is unlikely to survive either. Our interpretation of the North Korea–US dyad is also consistent with that of Powell himself, who characterized both it and India–Pakistan as including a ‘weaker but more resolute’ side (2015: 612).

Data for synthetic control analysis

We must measure interstate conflict and a set of predictor variables for our three dyads of interest and their respective donor-pool dyads. The donor pool cannot include cases also subject to the treatment, and the synthetic unit should fit the outcome pattern in the pre-treatment period closely, to reliably represent the counterfactual pattern post-treatment (Abadie, Diamond & Hainmueller, 2015). The latter is especially of concern because interstate conflict is a rare event measured categorically, while synthetic control is commonly used with interval-scale outcomes such as continuous economic variables. We address this by creating a more nuanced measure of conflict. The problem of predicting rare events is well-known; there is no technical fix (Galar et al., 2012). Thus while we cannot expect pre-treatment fit to be as close as some other applications, this indicates the difficulty of the analytical task. It is not limited to synthetic

control, also affecting other analytical approaches to SIP, given the small number of positive cases.¹²

The donor pool for each synthetic dyad should be a collection of cases similar enough to the actual dyad that collectively they are likely to provide a good set of empirical predictors for its conflict pattern. We choose dyads with a history of militarized conflict similar to that of the dyad of interest, pre-treatment. We exclude dyads for which both states have nuclear weapons, or containing either state in the dyad under study, to ensure the synthetic unit is indeed non-treated, as well as states that have experienced exogenous shocks relevant for subsequent conflict behavior (Abadie, Diamond & Hainmueller, 2015: 497; such as the 2003 occupation of Iraq, see Online appendix).

Data

Our dependent variable is a moving average of dyadic conflict based on military hostility and fatalities. Following Rauchhaus (2009), we combine two indicators from the Militarized Interstate Dispute (MID) dataset (Palmer et al., 2015). We create a variable ranging from 0 through 6, with higher values indicating greater conflict, which we call *MID level*_{5mar}.¹³

Specifically, *Hostility level* is measured on a 5-point scale, ranging from ‘no militarized action,’ to ‘threat,’ ‘display’ and ‘use’ of force, to ‘war.’ *Fatality* is a seven-level ordinal variable measuring battle-related deaths in specific ranges. Our coding rules are identical for

¹² It might be argued that an analytical, non-predictive, approach is less affected by data imbalance because methods like Firth or rare-events logit (King & Zeng, 2001) can be used. This assumes a large enough sample size and set of positive outcomes to run valid, stable models. Given the plausible scope of SIP, we do not believe these conditions can be met.

¹³ Further details on variables and data sources are in the Online appendix.

each state in the dyad. We first collapse *Hostility level* to a 3-point scale, because threat, display, and use of force do not always represent an ordinal progression of conflict intensity. For example, a display of force can be a major troop mobilization, while a use of force can be a minor clash, and a threat can include threatening to use nuclear weapons (Ghosn, Palmer & Bremer, 2004: 145; using the full 5-point scale does not change our results - see Online appendix). We then add to this the fatality level using a simple formula giving more weight to fatalities incurred when the hostility coding is lower:

$$Hostility\ fatality_{it} = Hostility_{it} + (Fatality_{it}/Hostility_{it}).^{14}$$

$Hostility_{it}$ is the variable described above for country i in year t ranging from 1 for no militarized action, 2 for a threat, display, or use of force, and 3 for war. $Fatality_{it}$ is a variable for country i in year t coded 0 when no battle-related deaths are suffered by the country in conflict with the other in that year, 1 when there are 1-25, 2 for 26-100, 3 for 101-250, 4 for 251-500, 5 for 501-999, and 6 for 1000 or more deaths in that year.

To infer a single variable describing the country-to-country relationship, we sum *Hostility fatality* for each state, creating a dyad-level variable ranging from 3 through 8, and subtract 2 to adjust the range to 1 through 6. Dyad-years with no MID are coded 0, giving a variable ranging from 0 through 6, which we call *MID level*. We then take the 5-year right-aligned moving average to create the dependent variable, *MID level*_{5mar}.

We compiled a dataset describing the year-by-year relations among 167 countries, 1954-2010, with 14,566 distinct dyads and 566,221 dyad-year observations. Not surprisingly, the distribution of *MID level* is strongly skewed towards zero (99.5% of dyad-years). For each

¹⁴ To avoid division by 0 we add 1 to Hostility and then subtract 1 from *Hostility fatality*.

specification of the model we sought to identify a donor pool of fewer than 30 dyads with the most similar conflict levels, in the pre-treatment period, to that of the treated dyad. We base this on guidance and examples in Abadie, Diamond, and Hainmueller (2015: 500).

Specifically, while it is important to have a sufficiently large donor pool to allow the synthetic dyad to approximate the pre-treatment conflict pattern, expanding the pool risks overfitting and interpolation bias. In addition to sensitivity checks (Online appendix), limiting donor-pool size helps avoid these problems. For each specification we identified non-nuclear dyads for the donor pool based on average *MID level* similarity to the treated dyad. We selected all dyads within n standard deviations of the treated dyad's 10-year moving average of *MID level*. In practice, n ranges from 11 to 15 standard deviations and the donor pool ranges from 17 to 27 dyads (full details in the Online appendix). We emphasize that only the dyads that maximize pre-treatment fit from the donor pool actually contribute ('donate') to the synthetic dyad.

The predictor variables include twenty-one geographic, military, political and economic indicators, drawn from the interstate conflict literature, with some transformations to fit our predictive exercise. We briefly describe them here, with further details on data and sources in the Online appendix.

The following predictors are measured as 5-year right-aligned moving averages due to their time-variant characteristics: the absolute value of the difference between the military capabilities of each state; the ratio of military capabilities of the weaker to the stronger state; an ordinal measure of the strength of alliance ties within each dyad; the dyadic interaction of military personnel divided by total population for each state; the dyadic interaction for the one-year change in military expenditure for each state; the interaction of each state's trade volume with the other; the interaction of each state's total exports to all trading partners; average MID

Hostility level in the dyad's geographic region; the number of years of dyadic peace since the last MID, and its squared and cubed terms¹⁵; and the number of years the dyad has existed.

Annual values for the following predictors are used: distance between capital cities (natural log); a dummy indicator for contiguity on land or across less than 25 miles of water; a dummy indicator for dyads in which one state has nuclear weapons; ordinal indicators of joint democracy and joint autocracy, the number of great powers in the dyad, and a measure of the political distance between the dyad's regime types.

Results

We briefly summarize our results, then discuss each analysis. The results do not support Hypothesis 1, the unqualified existence of the stability-instability paradox. While there is support for the expectations of SIP for the India–Pakistan dyad, there is no support found with China–India or North Korea–United States.

However, the results for India–Pakistan and China–India do conform to the expectations based on Powell's model, supporting Hypotheses 2 and 3, respectively. We are somewhat cautious about interpreting the results for North Korea–US given the limited post-2006 data, and potential questions of data quality for North Korea, but we see no reason to believe that the dyad was substantially more conflictual than would have been predicted by the

¹⁵ While we acknowledge that there may be other important predictors of conflict, inclusion of this cubic polynomial to capture linear and non-linear impacts of recent conflict frequency helps account for these. In addition, omission of some factors like the underlying ethnic-rivalry characteristics of a dyad, for example India–Pakistan, are constants that should not bias our post-treatment prediction (relative to pre-treatment fit), although it may reduce the overall accuracy of the prediction.

synthetic case for the post-2010 period, so Hypothesis 2 is not supported for this case. Our analysis with the best available evidence, therefore, does not support SIP as a general theory (Hypothesis 1: 1 of 3 cases supported) for nuclear rivals, and provides at best tentative support within scope conditions (Hypothesis 2: 1 of 2 cases supported). We discuss each set of results in turn.

In the model for India–Pakistan, four donor dyads contribute to the synthetic dyad (Table III). Egypt–Israel and Greece–Turkey, long-standing rivals, together account for over 88% of the synthetic dyad. Nicaragua–Costa Rica and Chile–Argentina also have histories of border disputes leading to militarized conflict.

We achieve a reasonable fit for the conflict pattern pre-treatment, with visual correspondence of the actual and synthetic *MID level*_{5mar} lines up to 1990 (Figure 1). In the pre-treatment period the root mean squared prediction error (RMSPE), which measures the lack of fit, is 0.642, in the context of the standard deviation for *MID level* (without the moving average) for the India–Pakistan dyad during that time of 1.753.

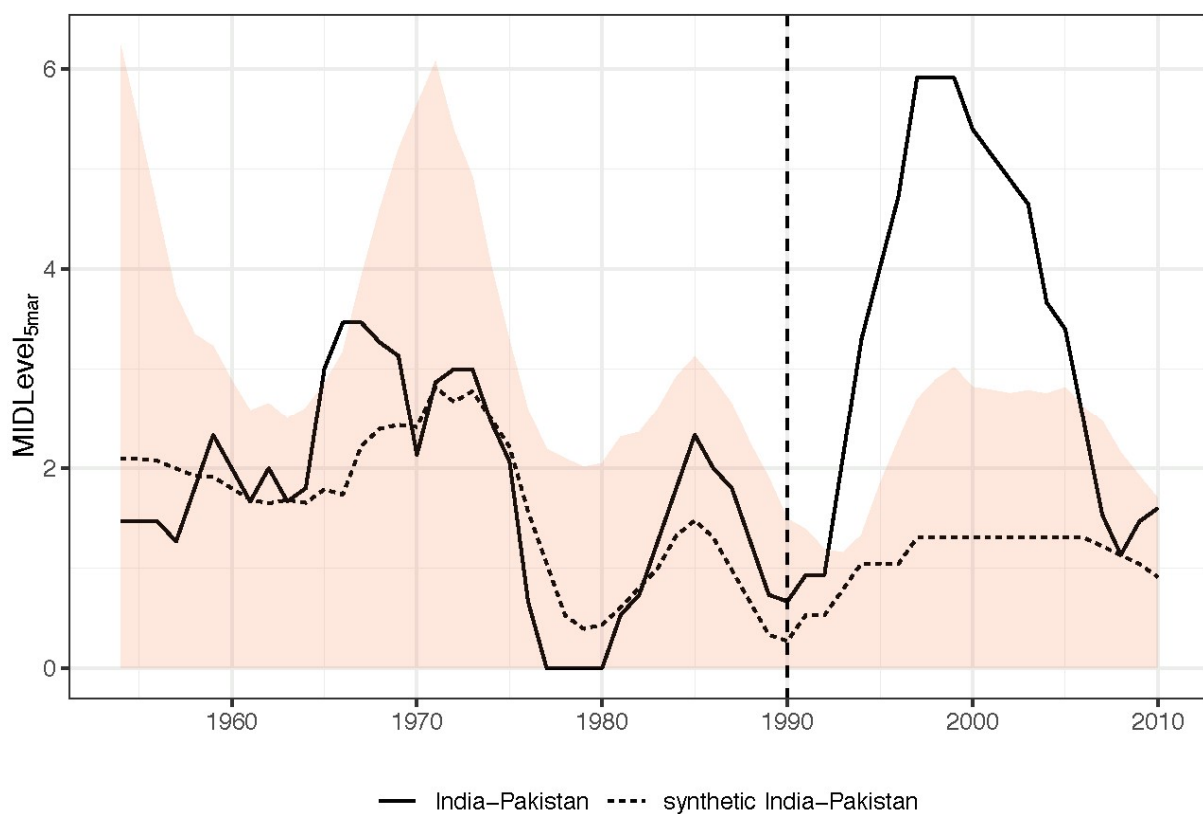
Post-treatment, there is a clear jump in conflict for the treated (actual) dyad after 1990, while the synthetic dyad experiences little change from the previous decade.¹⁶ The shaded area in each figure is a smoothed representation of the average conflict distribution among the donor pool, with the upper bound representing the highest conflict level within two standard

¹⁶ While the 1965 and 1971 India-Pakistan wars had more battle-deaths than the 1999 Kargil war, they have a lower 5-year average *MID level*, due to two factors. First, the *MID Fatality* scale gives any conflict with at least 1000 deaths a score of ‘6,’ so all three wars are equivalent. Second, the 1965 and 1971 wars are spikes in the *MID* data, preceded by low-conflict years, while the 1999 war was preceded by relatively high values on each indicator, making the right-aligned 5-year average higher in 1999 and surrounding years.

deviations of the mean.¹⁷ The conflict level for India–Pakistan substantially exceeds this range only in the post-treatment period. The result is consistent with the general SIP as stated in Hypothesis 1, and much of the qualitative literature. It is also consistent with Hypothesis 2, since the dyad has both an imbalance of conventional capabilities and an imbalance of resolve.

However, as discussed, this case has garnered analysts’ attention precisely because it appears to fit the theory, contributing to the theory’s survival and development, raising concerns about selection on the dependent variable and possible alternative explanations. Given these reservations, if this is the only case consistent with theoretical expectations, it may be inappropriate to claim clear empirical support.

Figure 1. India–Pakistan and synthetic India–Pakistan MID level_{5mar}



¹⁷ A wider range of 3 standard deviations would cover the entire graph, while a narrower range would create too low a threshold for supposedly exceptional conflict levels.

Table III. Synthetic India–Pakistan donor dyads

Donor Dyad	Weight
Egypt-Israel	0.565
Greece-Turkey	0.324
Nicaragua-Costa Rica	0.106
Chile-Argentina	0.005

Turning to China–India, we note that these states fought a brief border war in 1962, and experienced serious tensions in 1967, 1987 (and 2017, 2020). Synthetic China–India provides a reasonably good fit for pre-treatment conflict, shown in Figure 2. The RMSPE is 1.029 with a standard deviation for *MID level* of 1.677. The fit is best from the late 1960s onward. The synthetic dyad under-estimates the actual dyad’s level of violence in the 1960s and around the 1987 dispute. However, even with this apparent low bias in the estimate, *MID level*_{5mar} for the treated dyad is well below what the general theory would predict (Hypothesis 1), and close to the synthetic estimate in the post-treatment period. China–India only exceeds the two-standard-deviation smoothed range of conflict for the donor dyads during the 1960s. The synthetic dyad includes three donor dyads, Nicaragua–Costa Rica, Greece–Turkey, Honduras–El Salvador (Table IV), with histories of intense but intermittent confrontations, in this sense a logical set for China–India. It is important to restate that synthetic control is an appropriate method specifically when valid comparator cases are not available – thus we do not claim that any of these dyads, on its own, is a valid comparator. Rather, the synthetic combination of their weighted predictors provides the best comparison given available data.

The analysis does not support the expectations of SIP. Rather, conflict between China and India post-1988 is close to the predicted non-nuclear counterfactual. Robustness tests (Online appendix) provide no evidence of a treatment effect at 1988. The border dispute that flared in 1984 with troop movements such as new patrols and deployments, was largely resolved with Indian Prime Minister Rajiv Gandhi’s 1988 Beijing visit, although our 5-year moving average carries some of the residual of those tensions into the post-treatment period.

Post-treatment conflict was initially lower than the predicted counterfactual, but both counterfactual and actual conflict hover around zero post-treatment. This is reasonably consistent with Hypothesis 3, supporting our contention that a dyad with symmetric low resolve is outside the scope of SIP (there is no evidence of a treatment effect – but we cannot expect conflict lower than zero).

Figure 2. China–India and synthetic China–India MID level_{5mar}

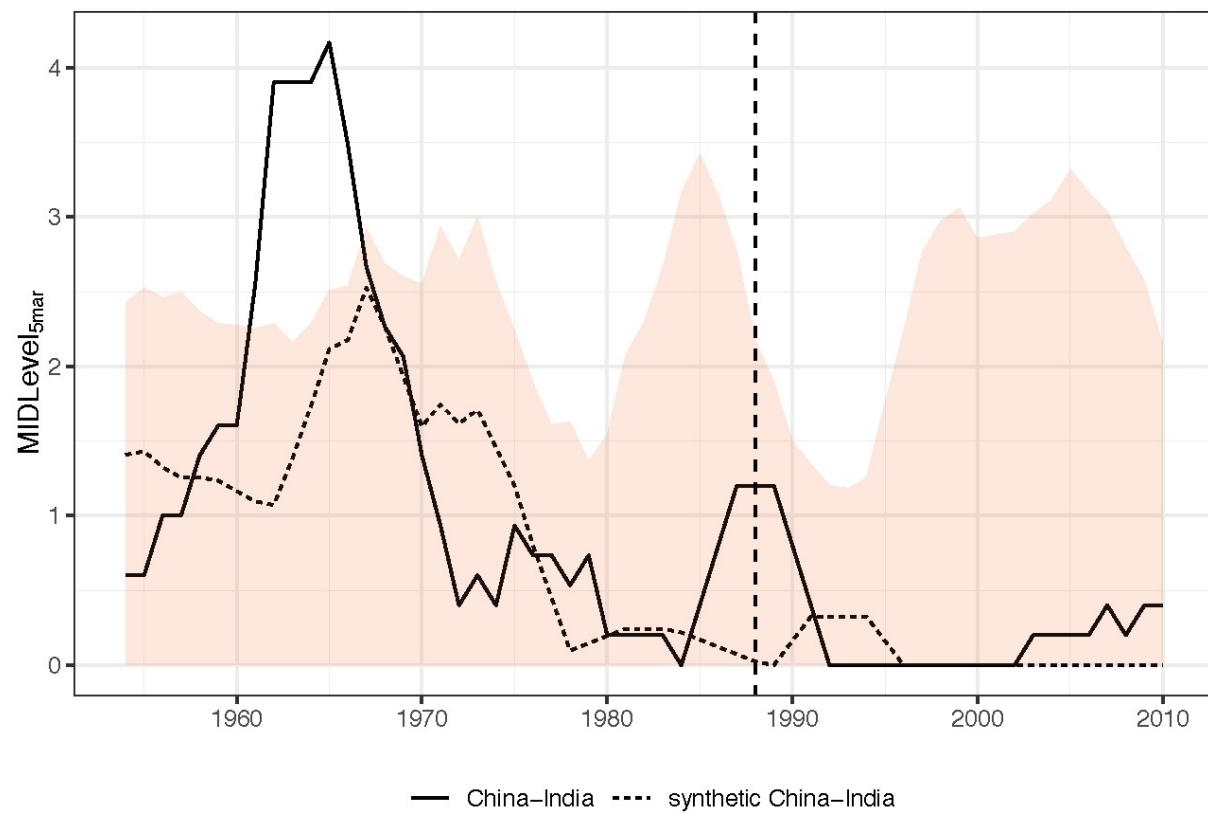


Table IV. Synthetic China–India donor dyads

Donor Dyad	Weight
Nicaragua-Costa Rica	0.531
Greece-Turkey	0.347
Honduras-El Salvador	0.122

North Korea–US has asymmetric power and resolve. Similar to India–Pakistan, the weaker state is more resolved. A limitation of this case is the relatively brief period of post-treatment data (four years, 2007-2010). The pre-treatment analysis achieves a noisier but still reasonable correspondence between real and synthetic dyads, tracking well especially from the mid-1980s. RMSPE is 1.199 given a standard deviation for *MID level* of 1.887, providing confidence in the post-treatment synthetic dyad. The pre-treatment tendency of the synthetic dyad is again to underestimate the conflict level of the real dyad. North Korea–US exceeds the two-standard-deviation smoothed range of donor-pool conflict only in the pre-treatment period (Figure 3). The pre-treatment under-estimation of *MID level*_{5mar} suggests that evidence for SIP should more easily emerge post-treatment, but again it does not. The level of post-treatment conflict is actually lower than the counterfactual level, inconsistent with SIP and with our expectations in Hypothesis 2. There are five donor dyads (Table V), all but one of which (China–Thailand) have substantial histories of conflict, providing a robust donor set. Robustness tests do not provide evidence of a treatment effect at 2006.

Figure 3. North Korea–United States and synthetic North Korea–United States $MID_{level_{5mar}}$

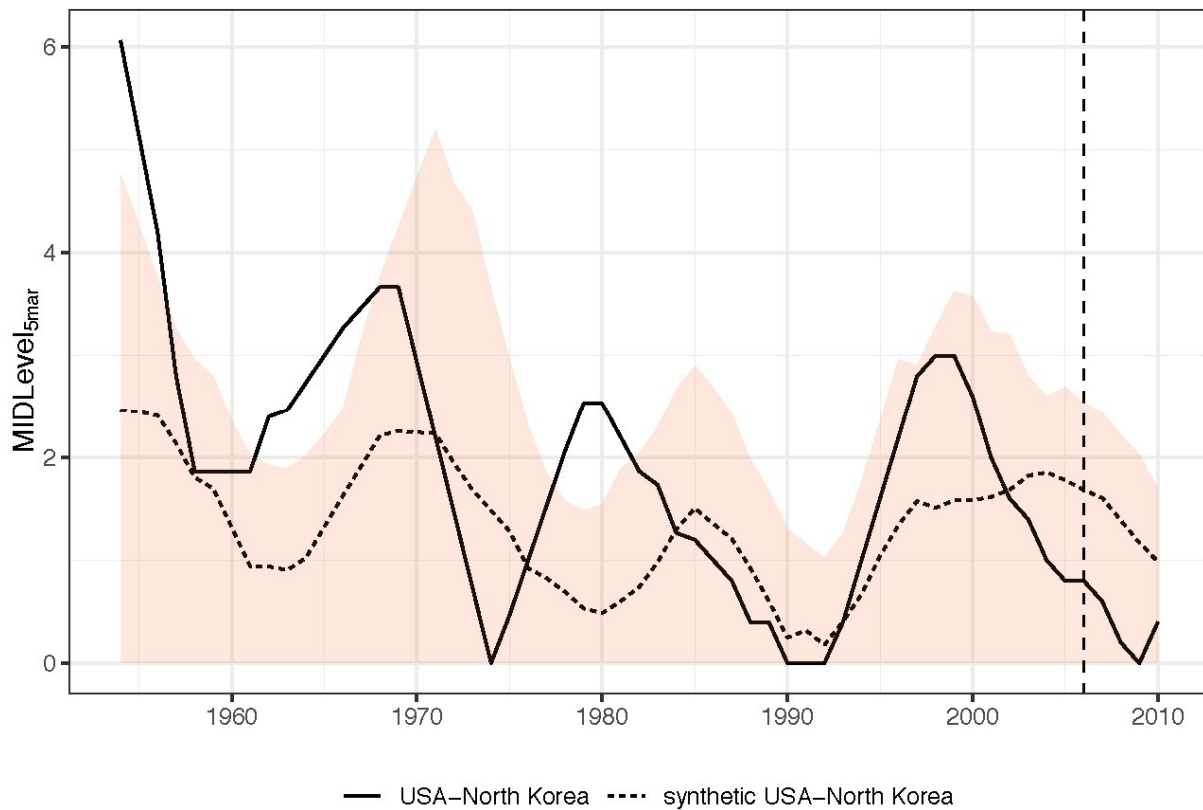


Table V. Synthetic North Korea–United States donor dyads

Donor Dyad	Weight
Israel-Saudi Arabia	0.354
Greece-Turkey	0.285
Yugoslavia-Turkey	0.177
China-Thailand	0.174
China-Taiwan	0.010

Because we know there was a major incident affecting the dyad in 2010, the impact of which could be diluted in the averaged data, we also assessed *MID level* (without the 5-year moving average) as the dependent variable, but the results (Online appendix) do not change our conclusions. Specifically, during joint US–South Korean anti-submarine exercises, the

South Korean Navy corvette *Cheonan* was sunk, almost certainly by a North Korean submarine, killing 46 crew. For the North Korea–US dyad this reached *MID level 2*.¹⁸

What is missing for the averaged and annual analyses for DPRK–US is a clear post-treatment gap as seen for India–Pakistan. Although we do not have MID data after 2010, the post-2010 historical record strongly suggests no such gap would emerge. The dyad experienced only low-level conflict during the Obama administration’s ‘strategic patience’ strategy, and tensions during the initial years of the Trump administration involved missile launches and verbal sparring. In the absence of a counterfactual estimate post-2010, however, we can only contrast that with previous decades, which witnessed the 1968 seizure of a US intelligence ship, downing of a US Air Force plane in 1969 killing 31, killing of two US troops in the demilitarized zone in 1976, exploding a South Korean civilian airliner in 1987 killing 115, a 202 deadly North-South naval clash, multiple assassination attempts on the South’s presidents, and hostile rhetoric from all sides. Even with our limited data, and regardless of when we might mark the dyad’s nuclearization, no support for SIP emerges in the North Korea–US case.

Conclusions

Our results reinforce the contention that it is important to consider the conditions for valid counter-factual inference regarding the stability-instability paradox, both expanding consideration to relevant jointly nuclear cases beyond US–USSR and India–Pakistan, and avoiding invalid large-*n* comparisons. But doing this potentially undermines the general theory and does not reveal clear empirical support for conditional hypotheses. While there has been near-exclusive focus on whether the paradox helps explain conflict patterns for India and

¹⁸ The MID dataset records a North Korean use of force and a US show of force. Only South Korea suffered fatalities, which is not included in our dyadic measure for North Korea–US

Pakistan since 1990, reflected in Kapur's observation that '[m]ost scholars' make this attribution, we suggest that inference from this case alone is biased by selection on the dependent variable. Causation is attributed to nuclear weapons, although the qualitative evidence is often interpreted through the prism of the theory.

A relatively unexamined case, China–India, exhibits no increase in conflict after India acquired nuclear weapons. This is inconsistent with the stability-instability paradox as a general theory (Hypothesis 1). However, if we follow Powell and consider the balance of power and resolve as fundamental conditioning factors, then this case of symmetric low resolve would fall outside the scope of the paradox, and no effect on conflict could be a logical expectation. But our analysis of the North Korea–US dyad undermines the theoretical scope conditions. Although our expectations – and Powell's (2015: 612) – for this dyad are the same as for India–Pakistan, we do not find empirical support for the paradox in conflict levels after 2006, which are lower than the synthetic non-nuclear counterfactual.

We believe that at a minimum, our synthetic control analysis is an important, clarifying complement to existing studies. But we also contend that existing quantitative studies inappropriately pool all dyads, and existing qualitative studies have fundamental inferential challenges given the available evidence, while selecting on the dependent variable. To further advance studies of SIP, and the implications of nuclear weapons in general, our analysis shows the importance of clearly defining scope conditions and carefully considering the bases for counter-factual inference. Synthetic control is a valuable tool for this, but our theoretical elaboration also suggests how qualitative analyses can better select a range of cases based on theoretical priors, and similarly how quantitative studies might limit their samples to theoretically relevant dyads.

Our results cast doubt on the theory, even after we give greater precision for its applicability. There appears to be little systematic empirical support that is not vulnerable to

claims of bias due to selection on the dependent variable, extrapolation from irrelevant cases, or fraught interpretation of indeterminate qualitative evidence. This finding raises a further fundamental question that deserves investigation: if this important aspect of nuclear deterrence theory is in doubt, what are the implications for scholars' and practitioners' broader understanding of nuclear weapons in international relations?

The implications, we believe, would be in direct contradiction to the conclusions of Geller's (2017: 25) recent comprehensive review of the literature, which places great stock in the cumulation of evidence for an increase in 'limited war' among nuclear states, claiming that '[c]rises among nuclear powers have a higher probability of escalating.' Our study suggests that this is a cumulation of similarly problematic analyses, in need of reconsideration. Our analysis strongly suggests that any further study of the stability-instability paradox should be informed by new theorizing of its scope, which must include reconsideration of basic assumptions and causal mechanisms.

References

- Abadie, Alberto, Alexis Diamond & Jens Hainmueller (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105(490): 493–505.
- Abadie, Alberto, Alexis Diamond & Jens Hainmueller (2011) Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software* 42(13).
- Abadie, Alberto, Alexis Diamond & Jens Hainmueller (2015) Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2): 495–510.
- Abadie, Alberto & Javier Gardeazabal (2003) The economic costs of conflict: A case study of the Basque Country. *The American Economic Review* 93(1): 113–132.
- Bell, Mark S & Nicholas L Miller (2015) Questioning the effect of nuclear weapons on conflict. *The Journal of Conflict Resolution* 59(1): 74–92.
- Billmeier, Andreas & Tommaso Nannicini (2013) Assessing economic liberalization episodes: A synthetic control approach. *The Review of Economics and Statistics* 95(3): 983–1001.
- Botosaru, Irene & Bruno Ferman (2019) On the role of covariates in the synthetic control method. *The Econometrics Journal* 22(2): 117–130.
- Bove, Vincenzo & Roberto Nisticò (2014) Coups d'état and defense spending: A counterfactual analysis. *Public Choice* 161(3/4): 321–344.
- Cohen, Michael D (2013) How nuclear South Asia is like Cold War Europe. *The Nonproliferation Review* 20(3): 433–451.
- Costalli, Stefano, Luigi Moretti & Costantino Pischedda (2017) The economic costs of civil war: Synthetic counterfactual evidence and the effects of ethnic fractionalization. *Journal of Peace Research* 54(1): 80–98.
- Dorsett, Richard (2013) The effect of the troubles on GDP in Northern Ireland. *European Journal of Political Economy* 29(March): 119–133.
- Ferman, Bruno, Cristine Pinto & Vitor Possebom (2020) Cherry picking with synthetic controls. *Journal of Policy Analysis and Management* 39(2): 510–532.
- Galar, M, A Fernandez, E Barrenechea, H Bustince & F Herrera (2012) A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(4): 463–484.

- Ganguly, Sumit (2008) Nuclear stability in South Asia. *International Security* 33(2): 45–70.
- Ganguly, Sumit & Devin T Hagerty (2006) *Fearful Symmetry: India-Pakistan Crises in the Shadow of Nuclear Weapons*. Seattle, WA: University of Washington Press.
- Gartzke, Erik & Matthew Kroenig (2009) A strategic approach to nuclear proliferation. *The Journal of Conflict Resolution* 53(2): 151–160.
- Gartzke, Erik & Matthew Kroenig (2016) Nukes with numbers: Empirical research on the consequences of nuclear weapons for international conflict. *Annual Review of Political Science* 19(1): 397–412.
- Gartzke, Erik & Matthew Kroenig (2017) Social Scientific Analysis of Nuclear Weapons: Past Scholarly Successes, Contemporary Challenges, and Future Research Opportunities. *Journal of Conflict Resolution* 61(9): 1853–1874.
- Geddes, Barbara (1990) How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis* 2(ed): 131–150.
- Geller, Daniel S (2017) Nuclear weapons and international conflict: Theories and empirical evidence. In: *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press (<https://dx.doi.org/10.1093/acrefore/9780190228637.013.347>).
- Ghosn, Faten, Glenn Palmer & Stuart A Bremer (2004) The MID3 data set, 1993–2001: Procedures, coding rules, and description. *Conflict Management and Peace Science* 21(2): 133–154.
- Glaser, Charles L (1990) *Analyzing Strategic Nuclear Policy*. Princeton, NJ: Princeton University Press.
- Horiuchi, Yusaku & Asher Mayerson (2015) The opportunity cost of conflict: Statistically comparing Israel and synthetic Israel. *Political Science Research and Methods* 3(3): 609–618.
- Jervis, Robert (1984) *The Illogic of American Nuclear Strategy*. Ithaca, NY: Cornell University Press.
- Jo, Dong-Joon & Erik Gartzke (2007) Determinants of nuclear weapons proliferation. *Journal of Conflict Resolution* 51(1): 167–194.
- Kapur, S Paul (2005) India and Pakistan's unstable peace: Why nuclear South Asia is not like Cold War Europe. *International Security* 30(2): 127–152.
- Kapur, S Paul (2007) *Dangerous Deterrent: Nuclear Weapons Proliferation and Conflict in South Asia*. Stanford, CA: Stanford University Press.
- Kapur, S Paul (2008) Ten years of instability in a nuclear South Asia. *International Security* 33(2): 71–94.

- Kennedy, Ryan (2014) Fading colours? A synthetic comparative case study of the impact of ‘colour revolutions’. *Comparative Politics* 46(3): 273–292.
- King, Gary & Langche Zeng (2001) Logistic regression in rare events data. *Political Analysis* 9(2): 137–163.
- Kristensen, Hans M & Robert S Norris (2018) North Korean nuclear capabilities, 2018. *Bulletin of the Atomic Scientists* 74(1): 41–51.
- Kroenig, Matthew (2013) Nuclear superiority and the balance of resolve: Explaining nuclear crisis outcomes. *International Organization* 67(1): 141–171.
- Kydd, Andrew H (2019) The sturdy child vs. the sword of Damocles: Nuclear weapons and the expected cost of war. *Security Studies* 28(4): 645–676.
- Leng, Russell J (2006) Realpolitik and learning in the India-Pakistan enduring rivalry. In: TV Paul (ed.) *The India-Pakistan Conflict: An Enduring Rivalry*. New Delhi: Cambridge University Press / Foundation Books, 103–127.
- Montgomery, Evan Braden & Eric S Edelman (2015) Rethinking Stability in South Asia: India, Pakistan, and the Competition for Escalation Dominance. *Journal of Strategic Studies* 38(1–2): 159–182.
- Narang, Vipin (2010) Posturing for peace? Pakistan’s nuclear postures and South Asian stability. *International Security* 34(3): 38–78.
- Nash, John C (1979) *Compact Numeral Methods for Computers: Linear Algebra and Function Minimisation*. New York, NY: Adam Hilger.
- Palmer, Glenn, Vito D’Orazio, Michael Kenwick & Matthew Lane (2015) The MID4 dataset, 2002–2010: Procedures, coding rules and description. *Conflict Management and Peace Science* 32(2): 222–242 (<https://correlatesofwar.org/data-sets/MIDs/mids>).
- Paul, TV (2006) Causes of the India-Pakistan enduring rivalry. In: TV Paul (ed.) *The India-Pakistan Conflict: An Enduring Rivalry*. New Delhi: Cambridge University Press / Foundation Books, 3–24.
- Powell, Robert (2015) Nuclear brinkmanship, limited war, and military power. *International Organization* 69(3): 589–626.
- Rauchhaus, Robert (2009) Evaluating the nuclear peace hypothesis: A quantitative approach. *Journal of Conflict Resolution* 53(2): 258–277.
- Samartsidis, Pantelis, Shaun R Seaman, Silvia Montagna, André Charlett, Matthew Hickman & Daniela De Angelis (2020) A Bayesian multivariate factor analysis model for evaluating an intervention by using observational time series data on multiple

- outcomes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183(4): 1437–1459.
- Samartsidis, Pantelis, Shaun R Seaman, Anne M Presanis, Matthew Hickman & Daniela De Angelis (2019) Assessing the causal effect of binary interventions from observational panel data with few treated units. *Statistical Science* 34(3): 486–503.
- Sanso-Navarro, Marcos (2011) The effects on American foreign direct investment in the United Kingdom from not adopting the Euro. *JCMS: Journal of Common Market Studies* 49(2): 463–483.
- Singer, J David, Stuart Bremer & John Stuckey (1972) Capability distribution, uncertainty, and major power war, 1820-1965. In: Bruce Russett (ed.) *Peace, War, and Numbers*. Beverly Hills, CA: SAGE.
- Snyder, Glenn (1969) The balance of power and the balance of terror. In: Dean C Pruitt & Richard C. Snyder (eds) *Theory and Research on the Causes of War*. Englewood Cliffs: Prentice Hall, 114–126.
- Watterson, Christopher J (2017) Competing interpretations of the stability–instability paradox: the case of the Kargil War. *The Nonproliferation Review* 24(1–2): 83–99.
- Zagare, Frank C (1992) NATO, rational escalation and flexible response. *Journal of Peace Research* 29(4): 435–454.