What Drives Nuclear Proliferation?

An International Studies Quarterly Online symposium

Daniel Nexon
Philipp C. Bleek
Rupal Mehta
Matthew Fuhrmann
Todd S. Sechser
Etel Solingen
Joshua Malnight
Mark S. Bell

DeRaismes Combes, Managing Editor

Published Online, 26 July 2017 v1.0

INTRODUCTION

Daniel Nexon
Georgetown University

What drives governments to develop nuclear weapons? For those concerned in slowing, halting, or reversing the spread of nuclear weapons, the answer to this questions matter immensely. It helps identify what policies are most likely to convince states to abandon the quest for nuclear weapons, to not pursue them in the first place, or even to give up their existing stockpiles. But international-studies analysis provides a host of different, and sometimes contradictory answers. The same is true for the growing body of statistical literature on nuclear weapons. As Mark Bell argues in his *International Studies Quarterly* article, it even "offers many more distinctive explanations for proliferation than there are cases of proliferation in the historical record." Bell conducts a series tests using "extreme bounds analysis, cross-validation, and random forests to evaluate 31 variables identified as significant determinants of proliferation." He finds that, "While some variables—particularly, the pursuit and possession of other weapons of mass destruction, receipt of sensitive nuclear assistance, and some measures of threat—perform better than others, the overall results should give us pause. The majority of variables identified as significant determinants of proliferation fail to provide robust explanations for existing patterns of proliferation. They also offer little predictive ability beyond what we can achieve with an extremely simple model. The quantitative literature on proliferation has, for now, produced more tentative findings than scholars typically understand."

This is, to put it mildly, an important set of conclusions. Thus, we asked a number of scholars of nuclear-weapons policy and proliferation for their reactions to Bell's research note. They do so in this *International Studies Quarterly Online* symposium. Philipp C. Bleek finds Bell's analysis mostly persuasive, and writes that "the fact that so few variables perform well is both surprising and depressing." Rupal N. Mehta generally agrees, and stresses that Bell's analysis "suggests that academic training should more heavily incorporate a more intuitive and complete understanding of the value - and limitations- of statistical analysis." Matthew Fuhrmann and Todd S. Sechser separately weigh in. Fuhrmann engages in a replication analysis of Bell's note. He "extended his extreme bounds analysis in two ways, both of which suggest that some variables may be better at explaining proliferation than [Bell's] analysis implies." Sechser discusses some limitations of Bell's analysis, and concludes that "His research note is not an indictment of quantitative methods; it is an endorsement of them." Etel Solingen and Joshua Malnight evaluate the substantive, methodological, and academic-policy gap issues raised by Bell's research note, and Bell responds to his interlocutors.

Timely, illuminating, a little depressing, and leaves me wanting more

Philipp C. Bleek
Middlebury Institute of International Studies at Monterey

Mark Bell's (2016) meta-analysis of the quantitative literature on the causes of proliferation is both timely and illuminating. My primary criticism of the piece relates to various lacunae. To be fair, this is partly because the merits of Bell's work left me wanting more.

The large (and growing) literature applying sophisticated quantitative techniques to the proliferation puzzle is ripe for the incisive synthesis and broad-brush critiques Bell offers. And much of Bell's analysis sheds significant light, with the caveat that some of Bell's work stretches the bounds of my own methodological training.

Extant literature, including early garbage can (or as Erik Gartzke once described them to me, "throw the spaghetti against the wall and see what sticks") studies (i.e., Singh & Way 2004; Jo & Gartzke 2007), and later studies focused on particular variables[1], has various virtues. Quantitative approaches are a good fit (pardon the pun) for what appears to be the probabilistic and multi-causal nature of proliferation, as Bell notes, though complex causality, contingency, and equifinality interact in potentially more problematic ways with oft-unarticulated modeling assumptions. Bell is also right to observe that when subsequent studies challenge prior ones, for example when my co-author Eric Lorber and I (2014) highlighted and sought to correct flaws in earlier studies assessing the proliferation-tamping role of alliances, that sort of revisionism is a sign of the strength, not weakness, of the literature.

The vast majority of quantitative proliferation studies present bivariate statistics, one or a set of fully specified models, and a bevy of robustness checks. Chris Achen's (2005) trenchant argument for more in-depth engagement with a smaller number of variables has, to date, been name-checked more than taken seriously as a methodological prescription. Bell's work, especially his extreme bounds analysis "assessing the robustness of variables in the presence of various combinations of other variables" is both an important step in that direction and can inform further such steps.

Bell employs both dependent and independent variables from two prominent studies, Singh and Way (2004) and Jo and Gartzke (2007), upon which many others have built. But he does not acknowledge the coding flaws of many of those variables, documented, among other places, in my own work (Bleek and Lorber 2014; Bleek 2010). Bell does, however, attempt to remedy the significant flaws in the extant versions of the nuclear guarantor variable by employing the variable from the alliance study mentioned above. He also highlights challenges related to the complex spatial and temporal dependencies within country-year data. Still, his article never mentions the hazard analysis approach where the dependent variable is time until proliferation (See: Box-Steffensmeier and Jones 1997). The approach, employed by a number of the studies Bell cites, is intended to correct for some

[1] Many of these were included in two special issues of the *Journal of Conflict Resolution*, April 2009 (53:2) and April 2014 (58:3).

of the shortcomings of more conventional regression analyses, though it also entails its own potential issues. Bell's appendix explains that "the software for extreme bounds analysis does not permit the use of survival models," but - as a robustness check - does report a model with splines to capture time dependence. Its results are apparently essentially identical to those reported in the article.

Bell's article left me mulling what else might be included in a broad-brush critique of the literature. Relevant, though definitely stretching the bounds of what Bell could reasonably be expected to cover in a single article, would be discussion of medium-n techniques like Ragin's qualitative comparative analysis (Ragin 2000). And though many of the studies Bell cites marry quantitative and qualitative analysis, few in the proliferation domain combine them in ways that capitalize on their synergistic potential (Editorial 2007).

Bell is right to highlight that variables that explain proliferation are also likely to affect the strength of efforts to prevent proliferation. This is Nick Miller's point when he argues that proliferation by rivals is in fact a very significant spur - a finding that runs counter to my own dissertation results - but that there are few cases of "reactive proliferation" because the hegemonic United States takes that possibility seriously and acts to prevent it (Miller 2014). Whether Miller is right or not, this problem bedevils proliferation analysis more generally.

I feel compelled to push back against Bell's repeating of King, Keohane, and Verba's (1994) canard that qualitative studies selecting on the dependent variable risk generating misleading inferences. As many qualitatively and quantitatively-oriented scholars have since argued, KKV's application of a quantitative logic of inference to the qualitative domain is inapt, and the literature has moved on (i.e. Mahoney 2010).

These comments are mostly down-in-the-weeds methodological ones. More broadly, the fact that so few variables perform well is both surprising and depressing. That said, I suspect —and my colleague Matt Fuhrmann's partial replication seems to confirm—that this is partly a function of how statistically demanding many of Bell's techniques are relative to the modest case universe, as well as his decision to include certain variables not widely employed in the literature that likely have confounding interactions with others. I was also surprised by which variables did better and which did worse. For example, in my own work I have found that a formal alliance with a nuclear-armed protector correlates very strongly with the absence of proliferation, whereas Bell finds little support for that variable. Conversely, my own modeling suggests that whether receiving sensitive nuclear assistance correlates with proliferation is rather susceptible to model specification, whereas that is one of the few variables for which Bell finds more robust support. I hope others will weigh in to try to help explain how variables that are both theoretically compelling and robustly supported by extant studies find so little support here.

The Search Continues: The Progress of Quantitative Nuclear Studies and the Bridges Yet to Cross

Rupal Mehta
University of Nebraska-Lincoln

Mark Bell (2016) surveys the landscape of the determinants of nuclear proliferation. Bell offers a strong and analytically-driven critique to the current approach of many large-n quantitative studies and highlights room for improvement and empirical advances. As a contributor to this scholarship, I am sympathetic to and motivated by Bell's concerns and appeals for more carefully constructed research designs and modeling choices.[2] I am also similarly convinced that our field can continue to improve upon this existing work by employing nuanced data and novel design.[3]

Yet, my primary criticism of this piece stems from two potential implications of Bell's analysis. The first is that the onus of analytical responsibility rests on the quantitative scholars (rather than on both theoretical and statistical work alike). Second, I examine Bell's suggestion of the potentially more limited role that this research should play in policy-making.

The introduction of quantitative scholarship on nuclear proliferation came at a time when our theoretical understanding of the two biggest issues related to nuclear weapons (its causes and consequences on the international system) was muddled. In part, this research agenda was attempting to make sense of the growing field of theoretical work that had yet to be tested on the universe of proliferation cases (Singh and Way 2004; Jo and Gartzke 2007; Bleek and Lorber 2014).[4] Some of this early wave of scholarship was trying to wade through competing logics that each offered contradictory implications, and provide a first cut test of which, if any, had empirical merit. Quantitative nuclear scholarship, to its credit, has done a remarkable job in demonstrating that much of the early theoretical work on the causes and consequences of nuclear weapons is similarly tentative and has produced mixed findings (Thayer 1995; Sagan 1996/1997; Hymans 2006; Solingen 2007; Tannenwald 2007; Paul 2000). Without this initial phase of large-n testing (with the accepted concerns on explanatory and predictive power), it is important to realize that our understanding of why states pursue nuclear weapons would be even more uncertain.

Based on this piece, I think Bell would agree that we as scholars and analysts must return to first principles. As he states in his study, "this literature, however, offers many more

---

[2] See Erik Gartzke, Jeffrey M. Kaplow, and Rupal N. Mehta. 2014. "The Determinants of Nuclear Force Structure." *Journal of Conflict Resolution*. 58(3): 481-508. Also, see Gene Gerzhoy, Rupal N. Mehta, and Rachel Whitlark. 2015. "Assessing the Benefits and Burdens of Nuclear Latency." *Working Paper*.

[3] See Matthew Fuhrmann and Benjamin Tkach. 2015. "Almost Nuclear: Introducing the Nuclear Latency Dataset." *Conflict Management and Peace Science*. 32(4):443-461 and Michael C. Horowitz and Neil Narang. 2014. "Poor Man's Atomic Bomb? Exploring the Relationship Between "Weapons of Mass Destruction." *Journal of Conflict Resolution*, 58(3):509-525.

[4] See also Bell's discussion on selecting the dependent variable. See also, Gary King, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Quantitative Research*. Princeton: Princeton University Press.

distinctive explanations for proliferation than there are cases of proliferation in the historical record." This similarly suggests the need for more thorough and rigorous theoretical thinking about the explanatory and predictive power of various theoretical mechanisms. As every good singular modeling choice must be theoretically-driven, so too should our broader approach to multivariate testing. If some variables are shown to have weak (or potentially harmful) predictive power, it is incumbent upon researchers to determine whether we must move beyond these explanations in the broader scholarship. It may also then become necessary to invest more time in clearly examining the logic and mechanisms behind these causes of proliferation, and inductively and deductively develop new logic as our understanding continues to progress.

Bell also suggests that scholars should be more careful in providing policy implications based on the findings from quantitative scholarship. As a fervent believer in the weaknesses associated with probabilistic research, I am often hesitant to provide a clearly binary recommendation for an inherently more complex, multi-causal reality. As such, I support Bell's appeal for being more cautious in offering suggestions for how policy-makers can better predict instances of proliferation.

Yet, one of the unfortunate side effects of social science research on on-going human phenomena is that we're sometimes going to get it wrong. As our understanding of intentions and motivations change (in nuclear decision-making and other arenas), so too do actors' preferences, strategies, and behavior. Our research (both theory and analysis) is trying to keep up with an ever-evolving world, and to limit the dissemination of our findings until we're absolutely certain of its predictive power (and until there is no chance that we'll get it wrong) may be imprudent. Policymakers are also operating in an asymmetric-information environment and if our research is able to contribute even a little to ongoing debates and discussions that may prevent harmful and more disastrous outcomes, perhaps it is worth sometimes 'getting it wrong.'

The ambition of good social research is to answer genuine and significant puzzles in our field, use strong evidence to corroborate theoretically-driven hypotheses, and potentially to amend and extend the paradigm of known phenomena.[5] While, as Bell points out, there is still much work to be done to accomplish this aims, I take comfort in the progress we've made to date.

While some may read this piece and question the utility of quantitative analysis (despite Bell's intentions to the contrary), I believe that this study suggests that academic training should more heavily incorporate a more intuitive and complete understanding of the value—and limitations—of statistical analysis. I take this piece (and related literature) as a call to arms to further delineate and strengthen the manner in which quantitative scholarship can improve our understanding of nuclear decision-making, including specifically the causes and consequences of nuclear proliferation and reversal, as well as other complex political phenomena.[6]

[5] I thank David Lake and Mathew D. McCubbins for this introduction to the philosophy of science.

[6] See Rupal N. Mehta. "Buying Off Friends and Foes: The Determinants of Nuclear Reversal." *Working Paper 2015*.

ASSESSING THE UTILITY OF STATISTICAL MODELS FOR EXPLAINING AND PREDICTING NUCLEAR PROLIFERATION

Matthew Fuhrmann
Texas A&M University

Research on the causes of nuclear proliferation has expanded considerably over the last decade.[7] Recent studies suggest that several factors influence how and why nuclear weapons spread: security threats, alliance politics, superpower coercion, domestic political institutions, international norms and treaties, and leader beliefs, among others (e.g., Hymans 2006, Jo and Gartzke 2007, Way and Weeks 2014, Miller 2014, Monteiro and Debs 2014). We do not fully understand, however, how important these factors are in shaping the proliferation process. Bell's article (2016) is therefore a welcome addition to the literature.

Bell has two objectives: (1) to assess how well existing variables from quantitative studies *explain* patterns of proliferation and (2) to evaluate the utility of these variables for *prediction*. He examines the explanatory and predictive power of 31 variables using extreme bounds analysis, cross-validation, and random forests.

Bell's analysis shows that the majority of variables identified as correlates of proliferation in the existing literature "fail to provide robust explanations for existing patterns of proliferation." In addition, he finds that these variables offer "little predictive power." On the basis of this evidence, Bell concludes that the quantitative literature has "produced more tentative findings than scholars typically understand" and that "the overall results should give us pause."

These findings are important and deserve to be taken seriously, especially since the type of analysis Bell carries out can play a critical role in advancing knowledge. I fully support his view that we should carefully assess the robustness of our findings and evaluate the predictive power of our models.

To build on Bell's initial effort, I extended his extreme bounds analysis in two ways, both of which suggest that some variables may be better at explaining proliferation than his analysis implies.

First, I excluded certain independent variables from the analysis based on theoretical considerations. Bell examines every possible five-variable statistical model from the 31 possible independent variables. This approach is appropriate given his research objective: to identify factors that are robustly *correlated* with nuclear proliferation. If one instead seeks to isolate the *causal* effect of one variable on nuclear proliferation (to the extent possible when using observational data), however, the inclusion of certain variables may introduce post-treatment bias.

Consider, for example, the argument that former rebels are more likely to seek nuclear weapons in office (Fuhrmann and Horowitz 2015). Bell's list of variables includes many factors that are caused by rebel experience, such as the initiation of military disputes

[7] The author thanks Mark Bell for sharing his replication files.

(Horowitz and Stam 2014). Including measures of military conflict in a model designed to estimate the effect of rebel experience on proliferation, therefore, introduces post-treatment bias. However, Bell includes rebel experience with other covariates that are caused by having a former rebel in office because his analysis is not specifically designed to evaluate the relationship between rebel experience and proliferation.

I replicated Bell's extreme bounds analysis while removing six variables that seemed most likely to be caused by rebel experience: domestic unrest, biological weapons pursuit, chemical weapons pursuit, chemical weapons possession, enduring rivalry, and military disputes. This leaves 25 variables from which five-variable models are estimated.

The results reveal that rebel experience is robustly associated with nuclear weapons pursuit in the positive direction (see Figure 1).[8] The findings are similar when I exclude biological and chemical weapons only, allowing domestic unrest, enduring rivalry, and military disputes to remain in the models. The insignificance of rebel experience in Bell's analysis, then, appears to be driven by the inclusion of biological and chemical weapons programs.

Thus, preventing variables that are post-treatment to rebel experience (especially biological and chemical weapons) from appearing in the models seems to change the inference we draw. It is worth noting that land borders – the "pre-treatment" control for a state's security environment – emerges as statistically significant in this analysis, too. Other variables that are insignificant in Bell's analysis – for example, personalistic regimes (see Way and Weeks 2014) – may similarly be important for explaining proliferation once we reduce the risk of post-treatment bias.

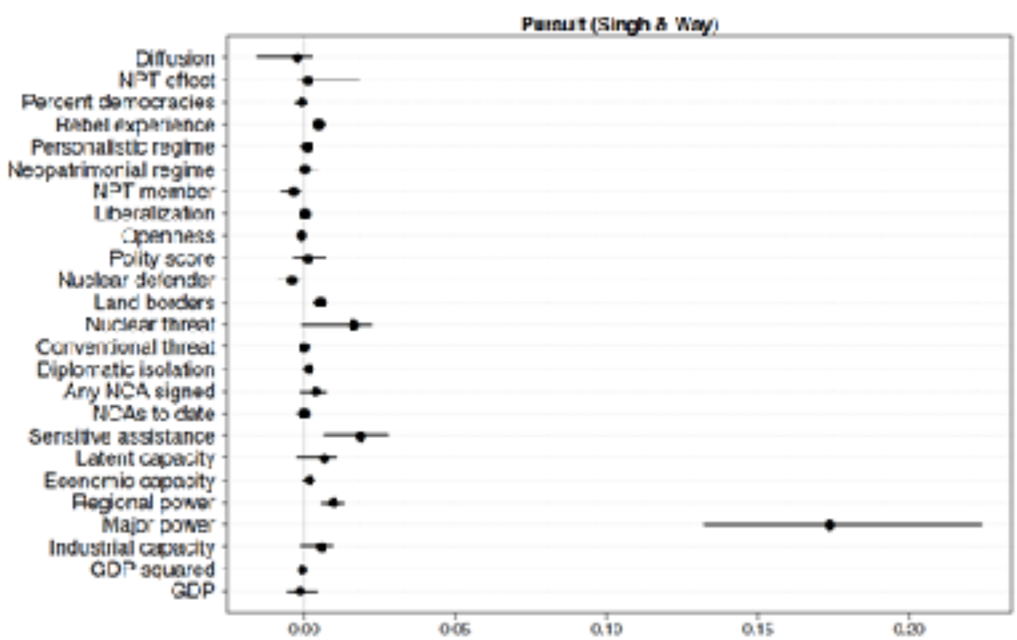

Figure 1. Results from extreme bounds analysis with post-treatment variables excluded. These findings are based on 10,000 models randomly drawn from all possible models.

[8] In this analysis, the dependent variable is nuclear weapons pursuit based on Singh and Way (2004).

Some findings in the extant quantitative literature on nuclear proliferation may be tenuous, as Bell's extreme bounds analysis suggests. Yet, other results are probably more robust than his article implies. To fully appreciate the *causal* effects of independent variables on proliferation, scholars should estimate theoretically appropriate models when assessing the robustness of their findings.

To his credit, Bell recognizes this: he prevents highly post-treatment variables from appearing in the same model and reports relevant findings in the Appendix to his article. His analysis may not go far enough in addressing this issue, however, as suggested by the results shown in Figure 1.

Second, I replicated Bell's extreme bounds analysis with an alternate dependent variable. In the conclusion to his article, Bell suggests that newly available data may shed light on the causes of proliferation. Indeed, more fine-grained measures may add to our understanding of proliferation dynamics. Drawing on the Nuclear Latency dataset (Fuhrmann and Tkach 2015), I created a variable measuring whether a state operates advanced nuclear technology – in particular, uranium enrichment or plutonium reprocessing plants. These dual-use technologies provide the building blocks for a bomb program.

Figure 2 shows that twelve variables are robustly correlated with the possession of advanced nuclear technology. By contrast, Bell finds that six or fewer variables are robustly associated with nuclear weapons exploration, pursuit, or acquisition. At the very least, then, a greater number of factors appear to matter when we examine an earlier stage in the proliferation process. This may be partially because we have more variation to exploit: whereas only 10 countries have ever built nuclear weapons, 31 states have possessed the foundation for a bomb program since World War II.

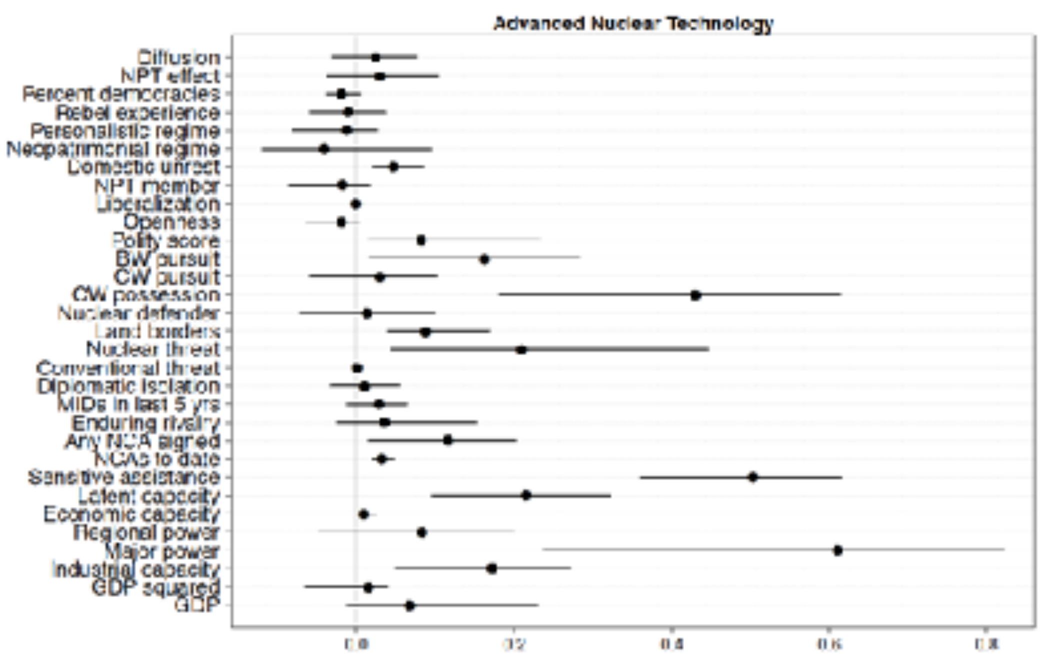

Figure 2. Results from extreme bounds analysis with alternate dependent variable. These findings are based on 10,000 models randomly drawn from all possible models.

Bell's prediction-related findings raise many important questions, too. On some level, these results are not surprising. Nuclear proliferation occurs relatively infrequently. Predicting rare events – especially out of sample – is notoriously difficult (see, e.g., Ward et. al 2010). Cross-validation requires the analyst to carve up a dataset into a "training" and "test" set. A rare event, therefore, becomes even rarer when one carries out this kind of analysis.

Bell codes the dependent variables based on the initiation (rather than the continuation) of a given proliferation stage. This is appropriate given that several previous studies take this approach (e.g., Singh and Way 2004). At the same time, however, this means that there are a relatively small number of "1's" in the dependent variable to begin with, making 'out of sample' predictions exceedingly difficult.

One could alternatively code proliferation "1" for all years in which a given outcome occurs. Iran, for example, could be coded as pursuing nuclear weapons from 1985 to 2003, rather than just in 1985. Our statistical models may be better at predicting whether a country has a nuclear weapons program in a given year, even if they struggle to explain the initiation of a bomb program.

Moving forward, scholars should consider how we might improve the predictive power of our models. Bell's analysis gets us going in the right direction, but there is still more work to be done.

A PIVOTAL MOMENT IN PROLIFERATION RESEARCH

Todd Sechser
University of Virginia

In the last decade, the academic literature on the causes of nuclear proliferation has undergone a quantitative renaissance. A decade ago, the study of proliferation belonged almost exclusively to qualitative scholars, who prioritized explaining individual cases of nuclear acquisition and restraint. Beginning with articles by Singh and Way (2004) and Jo and Gartzke (2007), however, the field began to turn to quantitative methods to explore why states build (or pursue, or explore) nuclear weapons. A new crop of scholars argued that quantitative methods offered insights that other techniques could not.

But the quantitative approach has recently come under attack. Sagan (2011)critiqued the quantitative approach in his 2011 review article, arguing that nuclear proliferation is simply too rare to justify the use of large-*N* statistical methods. Moreover, he argued, quantitative studies of proliferation usually employ causal variables that are under-theorized and crudely measured, yielding a hodgepodge of fragile and inconsistent findings (see also Montgomery and Sagan 2009).

Mark Bell's (2016) outstanding research note mounts a different but equally serious challenge to this literature. More than a decade after Singh and Way's pathbreaking article, Bell takes stock of the field's progress in isolating the underlying causes of proliferation. His assessment is not encouraging.

Collectively, the literature on proliferation has identified dozens of independent variables that have turned up statistically significant in one study or another. But academia has a well-known publication bias: statistically significant findings are more likely to be published than null results. Published studies therefore may paint a skewed picture. Relationships that appear significant in one study may be fragile, or they may be substantively trivial when weighed against other factors.

Indeed, this is exactly what Bell finds in his assessment of the proliferation literature. Using a variety of computationally intensive statistical techniques, Bell evaluates the strength of more than 30 different causal variables and finds that few, if any, perform as advertised. Across literally millions of regression models, the average explanatory power of almost all of these causal variables – including external threats, outside technical assistance, leader background and regime type, and economic capacity – is zero. Worse, he finds that none of these variables consistently improves our ability to predict out-of-sample cases of proliferation.

One limitation of Bell's analysis is that it cannot say anything about causal factors that scholars have not measured quantitatively. For example, studies have pointed to the importance of norms (Rublee 2009), psychology (Hymans 2006), and bureaucratic parochialism (Sagan 1996) as drivers of proliferation, none of which quantitative scholars have been able to convincingly measure. By focusing only on the quantitative literature, Bell's research note cannot – and does not claim to – assess the robustness of these factors. So we therefore should be careful not to interpret his piece as an indictment of the proliferation literature as a whole.

Nonetheless, Bell poses a serious challenge to the quantitative agenda in proliferation research. His findings seem to reinforce Sagan's (2011: 233)contention that the quantitative approach to proliferation has yet to uncover any truly robust findings. It is unsettling that even the most basic and straightforward findings in the quantitative proliferation literature collapse in the face of Bell's tests. (One wonders how studies of, say, militarized disputes would fare in similar tests.) How, then, does the field move forward from here?

I think there are at least three key lessons. First, as in any scientific enterprise, improving measurement must be a central goal of the proliferation literature. Sagan (2011: 228-31) and others have criticized measures of the dependent variable (nuclear capability) in quantitative studies, but Bell's research note seems to suggest that independent variables are actually the key culprit here. For example, scholars have long emphasized national security as a primary driver of proliferation (e.g., Thayer 1995), yet Bell finds that security metrics perform poorly in quantitative models. Is this because security factors don't matter in proliferation decisions, or because we are simply measuring them poorly? One suspects the latter. A key goal therefore must be to improve quantitative metrics so that our theories can be adequately tested. While there are strong professional temptations to locate new and increasingly novel explanations for proliferation, Bell's results suggest that scholars first need to improve tests of the theoretical models we already have.

A second lesson carries more far-reaching implications for how we approach proliferation research. Quantitative studies of proliferation tend to be organized around individual independent variables. These studies often take the following form: (1) assert that nuclear proliferation is important; (2) hypothesize that a particular factor contributes to it; (3) demonstrate that this factor is statistically correlated with proliferation. (This formula, of course, is not unique to the study of proliferation.) But as Bell demonstrates, the aggregate result of this approach has been underwhelming. Bell's findings imply that we should pay less attention to the bare statistical significance of individual variables and more attention to whether those variables improve our ability to explain and predict proliferation. Proliferation scholars should regularly report metrics such as out-of-sample prediction rates and marginal reduction in error as alternative ways of assessing the importance of their findings.

Third, Bell's analysis reaffirms – rather than undermines – the case for quantitative methods in the study of proliferation. One key advantage of quantitative methods is that they allow us to compare the relative importance of independent variables that all appear to "matter" (Fuhrmann, Kroenig, and Sechser 2014). Deciding which variables can be safely ignored is a critical step in predicting complex political phenomena. Quantitative methods help scholars discriminate between essential and nonessential causal factors. While most of the independent variables scrutinized by Bell seem to fall into the "nonessential" category, it is worth noting that his tests would not have been possible without quantitative tools. His research note is not an indictment of quantitative methods; it is an endorsement of them.

In short, Bell has raised the bar for future quantitative studies of proliferation, which will now need to demonstrate much more than statistical significance in a few selected models. Bell is to be commended for making a timely and indispensable contribution to the nuclear proliferation literature, and scholars will be grappling with his findings for a long time.

More Noise than Signal in Proliferation Studies?

Etel Solingen and
University of Califronia, Irvine

Mark Bell's article (2016) is a welcome contribution to the unavoidable task of evaluating a research program that exhibits the problems he identifies: the failure of most quantitative studies to offer strong explanations for proliferation patterns, and their inability to predict out-of-sample cases. His findings resonate with those of other proliferation experts. The existing quantitative literature, argues Bell, produces more tentative findings than scholars typically understand. We concur fully with the first part of the last sentence but believe that the broader community of experts typically *does* understand the serious limitations of most quantitative studies (and qualitative ones) on this topic.

What are those limitations of quantitative studies according to Bell? First, there are many more distinctive explanations than cases. Second, most variables identified as significant determinants of proliferation "fail to provide robust explanations for existing patterns." Third, studies question the robustness of each other's quantitative findings. Fourth, they provide little sense of the hierarchy of importance of different explanations. Fifth, they offer "little predictive ability beyond what we can achieve with an extremely simple model" (which, incidentally but unstated in Bell's piece, can be of a qualitative kind). Sixth, they are not transparent about those limitations. Seventh, they typically model the effects of variables as constant across time and space. Here Bell reiterates a point others have made: studies must control for the world-time under which nuclear weapons are developed or eschewed, such as pre- or post-NPT era (Solingen 2007). Because of all of the above and more, Bell concludes that weak correlations between proliferation and many variables in extant quantitative studies offer no proof whatsoever that those variables do not in fact cause or prevent proliferation. In other words, the absence of evidence is not evidence of absence, as is sometimes argued in court. Most of these shortcomings are well known, and some can afflict qualitative studies as well (for an extensive review, see Wan and Solingen 2015 and Solingen 2007).

Bell's evidence for these deficiencies stems from his application of modern statistical and machine learning techniques. "Extreme bounds" analysis examines the robustness of variables across many possible model specifications, partially addressing what some label "model uncertainty" (e.g. Droguett and Mosleh 2008). "Cross-validation" examines how well a sample of cases predicts out-of-sample cases. Bell finds that out-of-sample prediction is quite poor overall, though certain variables do better in in-sample prediction. "Random forests" seek to maximize explained variation through strategic divisions of the data and can be useful in principle for finding complex relationships within the data. But Bell's results from applying those techniques are even more damning: no variables consistently explain or predict proliferation.

What accounts for the apparent poor performance of certain variables in quantitative studies, according to Bell? First, the models often neglect indirect causal pathways, which are far more difficult to capture. Hence they have little to say about their actual causal strength. Second, the deficient operationalization of variables--inadequate measures for underlying concepts or theories--is another Achilles heel. Two examples illustrate the consequences of invalid indicators in our view. First, Bell suggests that many measures of threat may perform poorly because threats "must be filtered through elite perceptions before they affect proliferation decisions."

However, other studies have already stipulated and tested in a significant number of *relevant* cases that "concerns with existential security are never perfunctory reflections of structural considerations …but rather the product of domestic filters that convert such considerations into different policies" and that "domestic survival models may be seen as filters through which security is defined" providing "a better handle on the operational implications of security predicaments" (Solingen 2007: 4,6, 53, 72, 259, 285, and Solingen 1998). Second, a frequently used variable, trade openness, does not capture whether dominant coalitions are "internationalizing" or "inward looking" (a *political* variable with attendant consequences for nuclear choices according to the same theory). President Park Chung-hee adopted an internationalizing model in 1964 under very low levels of trade openness (TO), as did others. Rising ratios can expand the beneficiaries of TO but can also buttress inward-looking counter-movements. The relative strength of internationalizers may or may not dovetail with TO levels; the former cannot be inferred from the latter and must be gauged independently. The relationship between TO and coalitional models is not linear but the product of domestic political contestation and institutional variation. Furthermore, particular global world-times and context can mobilize forces behind inward-looking nationalist or internationalizing banners. Both examples thus point to potential failures to operationalize and measure underlying theories and concepts.

We concur with Bell that – when designed and operationalized appropriately--quantitative analysis might still be useful for specifying the relative weight of variables. This is not a unique virtue of quantitative studies, however. Rigorous qualitative work can: (a) advance falsifiable arguments; (b) assess them against competing claims; (c) be no less "evidence-based," pace Bell; (d) be more effective at discovering, dissecting and assessing causal pathways; (e) not select invariably on the dependent variable, as Bell asserts they do; and, crucially (f) be more invested in developing the kind of strong theoretical justifications that Bell calls for. Bell regards the inclusion of "the universe of cases" as a strength of quantitative studies, presumably avoiding selection bias. But there is wide discrepancy about what the appropriate "universe of cases" should be. Furthermore, serious concerns arise when the chosen "universe" exacerbates heterogeneity and decreases validity. Bell acknowledges as much when he suggests analyzing subsets of the data. He also argues that quantitative analysis can "explicitly [model] the probabilistic and multi-causal processes that likely cause proliferation." While that may be true in theory, his own results suggest it is rarely so in practice. Most quantitative models are generally linear and rely on additive linearity to account for multi-causal processes. At the very least we would expect significant interaction terms in regression models (to Bell's credit, random forests does attempt to solve this problem). Failure to include these terms renders the values of average effects relatively useless, particularly for temporal changes. Indeed, we may lack the data to appropriately model proliferation with statistical certainty despite attempts to multiply observations. And, in any event, the latter are not truly independent temporally or spatially.

Bell finds quantitative studies seriously limited in providing useful policy insights. He is right (and that may apply to some qualitative studies as well). Policy-makers and experts—often dismissive of quantitative findings--are progressively more likely to associate a state's probability of "going nuclear" with, say, the political strength of ruling coalitions seeking greater openness to the global economy (a variable omitted from the 31 scrutinized in quantitative studies). That may suggest that there is growing attention to evidence that decisions to abandon nuclear weapons since the 1970s have been strongly associated—causally and temporally--with decisions to embrace the global economy. This significant regularity emerges from extensive and systematic comparative analysis across regions (Solingen 2007). The P5+1/Iran 2015 nuclear agreement may well be designed to encourage a nascent shift in an internationalizing direction. The final fate of Iran's nuclear program hangs-- to a significant degree --in the balance between those who seek to deepen the course of economic openness and those who oppose it (Esfandiari 2015). Having said that, Bell's conclusion that no single variable is likely to "deterministically cause proliferation" seems uncontroversial. Understanding the scope conditions under which variables operate is where the real action should be (Sil and Katzenstein 2010).

Some concluding suggestions. First, quantitative methods that embrace uncertainty such as Bayesian models may not solve all modeling dilemmas. They can, however, provide more accurate estimation of our knowledge and incorporate it through specification of priors. Second, we wholeheartedly concur with Bell's plea for more explicit theorizing and modelling of the data-generating processes through which one expects proliferation to occur. Theorizing can range from game-theoretic to various other tools. Bell shows that adding variables into linear models and then estimating their marginal effects has been generally fruitless. Third, quantitative studies could complement rigorous comparative work based on deep knowledge of all or most cases involved. Hypotheses (and new observable implications) can be tested with hoop, smoking gun, straw in the wind, most and least likely criteria and other tests (Van Evera 1999; Ragin 2000, 2008; Solingen 2007, 2008; Mahoney 2012). Fourth, efforts across theoretical and methodological lines should be far more attentive to a (strangely enough) neglected causal mechanism: politics.

Finally, Bell's concerns seem specific to the proliferation literature, not a blanket criticism of quantitative studies. So are the points we raise here. The choice of appropriate method remains subordinated to the question one seeks to address and the availability of sufficient *relevant* cases (positive or negative), as we argue in our own work-in-progress. All methods applied to understanding nuclear proliferation (a topic rampant with secrecy walls) share difficulties with adequate and reliable data, but some do so more than others. Hence collaboration across methods may give us a better handle on the problem. Alas, such efforts remain few and far between, a casualty of entrenchment in methodological silos (no pun intended).

Uncertainty, Inferences, and the Study of Nuclear Proliferation

Mark Bell
University of Minnesota

I am grateful to *ISQ* for facilitating this symposium, and to Philipp Bleek, Matthew Fuhrmann, Rupal Mehta, Todd Sechser, and Etel Solingen and Joshua Malnight for participating. It is a pleasure to share a (virtual) platform with such accomplished scholars and to have them engage so constructively with my work.

In my article, "Examining Explanations for Nuclear Proliferation," I sought to evaluate the quantitative literature on the causes of nuclear proliferation. That literature collectively identifies a large number of variables as statistically significant determinants of proliferation. But it does not provide us with a good understanding of the *relative strength* of these different variables in explaining proliferation, nor whether any of them might allow us to predict proliferation. Using three techniques—extreme bounds analysis, cross-validation, and random forests—I found that few variables provide strong explanations for proliferation or offer much in the way of predictive capacity. I thus concluded that the quantitative literature on the causes of proliferation has produced more tentative findings than scholars typically understand. We should be more careful about claiming that this literature has identified factors that drive, and predict, nuclear proliferation.

All of the contributors to this symposium raise important points and questions that deserve further consideration. I lack the space to address all of these points here, so I focus on a few that I consider to be the most pressing.

Mehta worries about a possible implication of my article: that academics should not seek to communicate with policymakers until we are certain about our conclusions. That was not my intended conclusion, and I suspect Mehta and I are actually in agreement. That is, I fully endorse academics engaging with policymakers, and do not think we should wait for absolute certainty before doing so (we will be waiting a long time if we do!). I do think, however, that we need to be careful that we communicate not only our findings, but their relative uncertainty.

I consider the accurate communication of uncertainty to be a core part of the scientific enterprise. Scholars—whatever methods they use—should not fear doing so. This is especially true when we study rare and complex phenomenon such as nuclear proliferation: what Robert Jervis calls "the strangeness of the nuclear world" may not prove at all easy to boil down into straightforward recommendations. In short, uncertainty in our findings is a sign of good rather than bad research. We should acknowledge and embrace that uncertainty. While policymakers may have an intuitive sense of these limitations and uncertainties, as Solingen and Malnight suggest, scholars also have a professional obligation to be clear about them.

Fuhrmann argues that the extreme bounds analysis I use may underestimate the effect of some variables because of post-treatment bias. For example, he contends that, in order to estimate the effect of rebel leadership on proliferation, we should not control for a series of measures of a state's security environment—such as past military disputes or domestic unrest. Because past disputes or domestic unrest may be caused by having a rebel in office, including them in the analysis may dilute the effect of rebel leadership. Fuhrmann is right that extreme bounds analysis evaluates the importance of (to take Fuhrmann's example) rebel leadership across a wide range of models—some of which include variables that are plausibly post-treatment to rebel leadership. If we are certain that that these variables are caused by rebel leadership (and not by other factors that cause proliferation), then we should, in fact, expect that extreme bounds analysis will underestimate the effect of rebel leadership.

However, I don't think we enjoy that kind of certainty. For example, past disputes or the presence of domestic unrest are not just post-treatment to rebel leadership. They may also cause rebel leaders to assume control of governments, and may also cause proliferation directly. If so, excluding these variables may *overestimate* the effect of rebel leadership as much as including them may underestimate it. In essence, we can find arguments for both including and omitting these variables from our analysis. How, then, should we estimate the effect of rebel leadership in the face of uncertainty about whether these variables should be included in our models? In my view, it is precisely when we have uncertainty of this sort that extreme bounds analysis is *most* valuable. It allows us to assess—in a systematic way—the robustness and sturdiness of existing findings to the inclusion or exclusion of a variety of plausible explanatory variables.

Bleek questions why certain variables perform well in the analysis and others perform badly. I agree with him that further investigation into this would be useful. However, the suggestion I make in the article—that variables that are highly causally proximate to proliferation tend to perform well, while further removed causes tend to perform less well —fits the two factors that Bleek highlights. The receipt of sensitive nuclear assistance performs relatively well. This may be partly explained by the fact that receiving such assistance, as Scott Sagan and Alex Montgomery have argued, is very conceptually close to nuclear exploration itself. The second variable Bleek mentions—having a nuclear-armed ally —affects proliferation incentives less directly, which may explain why it performs somewhat less well.

Solingen and Malnight, as well as Seschser, raise important points about the role of variables and mechanisms that quantitative scholars have struggled to measure. These include norms, leader psychology, and the attitudes of domestic coalitions towards the global economy. I agree with them: quantitative methods require variables that can be measured "at scale" across thousands of observations. This often means that quantitative scholars are forced—through no fault of their own—to use proxies for the underlying, theoretically relevant, but hard to measure variables that we care about. When dealing with rare events—where mismeasurement in one or two cases may significantly alter our results—this can prove problematic. Similarly, quantitative methods can shed light on causal mechanisms (e.g., Imai et al 2011). But the assumptions required to do so are generally extensive. An advantage of qualitative methods is that by focusing on fewer cases, they allow us to pay more attention both to measuring the variables in the cases under examination and examining the mechanisms and causal processes at work.

My article offers a number of avenues for future research. Many of these suggestions are relevant to qualitative, quantitative, and mixed-methods scholars alike. As Mehta correctly points out, responsibility for improving our understanding of the causes of proliferation is shared across scholars working in different methodological traditions and across academic disciplines. I am not the first to make suggestions of this sort. But I am happy that the participants in the symposium endorse many of them. For example, Bleek and Sechser both agree on the need to focus on improving the measurement of important variables. Sechser also affirms the importance of quantitative scholars moving beyond tests of statistical significance in assessing the importance of variables. Solingen and Malnight, as well as Mehta, endorse theorizing additional observable implications of our theories. Such theorizing would allow for additional tests—whether qualitative or quantitative—of our theories. It would also potentially allow scholars to place less reliance on country-year data. Solingen and Malnight agree with the need for further exploration of how the causes of proliferation have evolved over time. Lastly, Fuhrmann's contribution uses his recently-collected data on nuclear latency to hint at the potential utility of new data sources. He shows that more explanatory variables may be robustly correlated with this new outcome variable than with the outcomes that the literature has used up to this point. This is a promising finding that future work should build on.

Overall, this symposium suggests that there are many fruitful paths for future research on the causes of proliferation to pursue. I look forward to watching this literature develop in the coming years.

References

Achen, Christopher H. (2005). "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong" *Conflict Management and Peace Science* Volume 22.

Bell, Mark S. 2016. Examining explanations for nuclear proliferation. *International Studies Quarterly,* 60(3), pp.520-529.

Bleek, Philipp C. 2010. "Why Do States Proliferate? Quantitative Analysis of the Exploration, Pursuit, and Acquisition of Nuclear Weapons." In William C. Potter with Gaukhar Mukhatzhanova, *Forecasting Nuclear Proliferation in the 21st Century, Volume 1: The Role of Theory.* Stanford University Press.

Bleek, Philipp C. and Erik B. Lorber. 2014. "Security Guarantees and Allied Nuclear Proliferation." *Journal of Conflict Resolution*. 59(1):74-92.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. "Time is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4):1414-1461.

Droguett, Enrique López, and Ali Mosleh. "Bayesian Methodology for Model Uncertainty Using Model Performance Data." *Risk Analysis* 28, no. 5 (October 1, 2008): 1457–76.

Editorial (2007). "The New Era of Mixed Methods." *Journal of Mixed Methods Research* 1(1): 3-7.

Esfandiari, Haleh. 2015. "How Jason Rezaian's Conviction Is a Warning to Iran's President Rouhani." *Wilson Center*. October 13. https://www.wilsoncenter.org/article/how-jason-rezaians-conviction-warning-to-irans-president-rouhani.

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

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

Gartzke, Erik, Jeffrey M. Kaplow, and Rupal N. Mehta. 2014. "The Determinants of Nuclear Force Structure." *Journal of Conflict Resolution*. 58(3): 481-508.

Gerzhoy, Gene Rupal N. Mehta, and Rachel Whitlark. 2015. "Assessing the Benefits and Burdens of Nuclear Latency." *Working Paper.*

Horowitz, Michael C. and Neil Narang. 2014. "Poor Man's Atomic Bomb? Exploring the Relationship Between "Weapons of Mass Destruction." *Journal of Conflict Resolution*, 58(3): 509-525.

Horowitz, Michael C., and Allan C. Stam. (2014) How Prior Military Experience Influences the Future Militarized Behavior of Leaders. *International Organization* 68(3):527–559.

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

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." American Political Science Review 105 (4): 765-789.

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

King, Gary Robert O. Keohane, and Sidney Verba. 1994. Designing Social Inquiry: Scientific Inference in Quantitative Research. Princeton: Princeton University Press.

Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62(1):120-47.

______. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41 (4): 570–97.

Mehta, Rupal N. 2015. "Buying Off Friends and Foes: The Determinants of Nuclear Reversal." *Working Paper.*

Miller, Nicholas L. 2014. The Secret Success of Nonproliferation Sanctions. *International Organization* 68(4):913–944.

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

Monteiro, Nuno P., and Alexandre Debs. (2014) The Strategic Logic of Nuclear Proliferation. *International Security* 39(2):7–51.

Montgomery, Alexander H., and Scott D. Sagan. 2009. "The Perils of Predicting Proliferation." Journal of Conflict Resolution 53 (2): 302-328.

Paul, T.V. 2000. Power versus Prudence: Why Nations Forego Nuclear Weapons. Montreal; McGill-Queen's University Press.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. University of Chicago Press.

______. 2008. Redesigning Social Inquiry: Fuzzy Sets and beyond. Chicago: University of Chicago Press.

Rublee, M.R., 2009. *Nonproliferation norms*: Why states choose nuclear restraint. University of Georgia Press.

Sagan, Scott D. 1996/1997. "Why Do States Build Nuclear Weapons? Three Models in Search of a Bomb." *International Security* 21(3):54-86.

_______. 2011. "The Causes of Nuclear Weapons Proliferation." *Annual Review of Political Science* 14 (1): 225–44. doi:10.1146/annurev-polisci-052209-131042.

Sil, Rudra, and Peter J. Katzenstein. 2010. "Analytic Eclecticism in the Study of World Politics: Reconfiguring Problems and Mechanisms across Research Traditions." *Perspectives on Politics* 8 (02): 411–31.

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

Solingen, Etel. 1998. Regional Orders at Century's Dawn: Global and Domestic Influences on Grand Strategy. Princeton University Press.

———. 2007. Nuclear Logics: Contrasting Paths in East Asia and the Middle East. Princeton University Press.

———. 2008. Solingen, Etel. "Theory and Method in the Study of Nuclear Proliferation" Paper presented at the annual meeting of the ISA's 50th Annual Convention.

Tannenwald, Nina. 2007. The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945. Cambridge, UK: Cambridge University Press.

Thayer, Bradley A. 1995. "The Causes of Nuclear Proliferation and the Utility of the Nuclear Non-Proliferation Regime." *Security Studies* 4(3):463-519.

Van Evera, Stephen. 1997. Guide to Methods for Students of Political Science. Cornell University Press.

Wan, Wilfred, and Etel Solingen. 2015. "Why Do States Pursue Nuclear Weapons (or Not)." In *Emerging Trends in the Social and Behavioral Sciences*. John Wiley & Sons, Inc.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. (2010) The Perils of Policy by P-Value: Predicting Civil Conflicts. *Journal of Peace Research* 47(4):363–375.

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