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).