Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2015) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Fragility and destabilizing protest: Combining event and structural data for improved forecasts

Miguel Garces[a], Ian S. Lustick[a,b], Brian P. Levey[c,*]

[a]Lustick Consulting, Inc., Ardmore, PA 19003, United States
[b]University of Pennsylvania, Philadelphia, PA 19104, United States
[c]Strategic Analysis Enterprises, Inc., Williamsburg, VA 23185, United States

Abstract

Social science-based efforts to achieve success forecasting events of interest, including domestic political crises in foreign countries, have advanced considerably in recent years, though controversy exists over claims of success, the validity of methods, and the credibility of codings. One issue that challenges every big data effort is the verifiable operationalization of concepts. That entails devising and deploying understandable and clearly codable proxies, or indicators, that do not distort the effective meaning of the variable being operationalized. An important example of these efforts is the W-ICEWS project funded originally by DARPA and then OSD/ONR. Organized as a team effort by Lockheed-Martin, ATL, the project has sought to improve its performance on a variable known as "Domestic Political Crisis." Difficulties defining that variable have suggested the possibility of narrowing its focus to "Destabilizing Protest"–the occurrence of mainly non-violent unrest that threatens reigning institutions of authority seriously enough to warrant high-level attention from US policy makers. In this paper we report on efforts to develop this variable by combining publicly available assessments of regime fragility with event data measuring mass protest from the W-ICEWS project. We then test the viability of this new variable by making in-sample and out-of-sample forecasts using a multivariate model of social indicators.

Peer-review under responsibility of AHFE Conference.

Keywords: Political instability; Protest; Statistical modeling; Concept formation; Fragility

* Corresponding author. Tel.: 1-484-412-8557.
E-mail address: miguelag@lustickconsulting.com

2351-9789
Peer-review under responsibility of AHFE Conference.

1. Introduction

According to former DARPA Program Manager Sean O'Brien, the purpose of the Integrated Crisis Early Warning System (ICEWS) program is to "develop a comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, sub-national, and international crises in a way that supports decisions on how to allocate resources to mitigate them." [1] Given this definition, questions remain regarding whether a "crisis" includes any level of disturbance in a country or whether a specific sensitivity to that disturbance is required. In this paper specifically, we focus on Domestic Political Crisis (DPC), one of the five key crises identified by ICEWS and its successor, OSD/ONR's Worldwide Integrated Crisis Early Warning System (W-ICEWS).

[1] The W-ICEWS team has struggled to define DPC in ways that can be standardized and clearly codable while honoring the diversity of forms that domestic political crises may take in different types of political systems. One proposal has been to narrow the "event of interest" from any sort of "crisis" to "Destabilizing Protest." Destabilizing Protest is defined as *the occurrence of mainly non-violent unrest that threatens reigning institutions of authority seriously enough to warrant high-level attention from US policy makers*. We will report on efforts to develop this variable by combining publicly available assessments of regime fragility with event data measuring mass protest from the W-ICEWS project. We then test the viability of this new variable by making in-sample and out-of-sample forecasts using a multivariate model of social indicators.

1.1. Defining the problem

The original DARPA definition of a Domestic Political Crisis (DPC) was as follows: "Significant opposition to the government, but not to the level of rebellion or insurgency (for example, power struggle between two political factions involving disruptive strikes or violent clashes between supporters)." [1] One element of this definition that makes it problematic is that it is defined relative to other EOI's, insurgency and rebellion, and therefore cannot be identified independent of prior assessments that other events did not occur. Moreover, the ambiguity of the original DPC definition is bound to make both theoretical and correlation-based forecasts difficult since boundary conditions for applicable theories or rates of association with other events cannot be clearly established.[2] Inevitably, without clear guidance from a standardized and easily interpreted operationalization, coders will (and did) develop relatively ad hoc and case-by-case classification rules. Based on feedback from users of W-ICEWS it appears that despite the apparent intuitively obvious meaning of "domestic political crisis," users have found DPC to be more confusing and less useful than the other EOIs. If one is not sure what a "DPC" is, what is one to make of a forecast that there is a high likelihood that it is likely to occur in country X next month?

1.2. Developing a new variable

The development of a new variable, also known as concept formation in the social sciences, is a process marked more by creativity, insight, and judgment than by rule-guided calculation. Perhaps the only thing that most authors writing about this topic agree on is that we must define terms in order to have a productive scholarly conversation. According to Barbara Geddes, definitions should be "concrete, unambiguous, and public, so that other scholars can understand the basis for the analyst's judgments." [2] It seems that most recent authors on the subject at least implicitly agree with John Gerring's claim that concept formation is a "highly variable process involving trade-offs." [3] Gerring has eight criteria that must be judged (familiarity, resonance, parsimony, coherence, differentiation, depth, theoretical utility, and field utility) and others use a similar multi-faceted approach to defining their concepts, usually noting its use in the literature, face validity, usefulness, etc. Brian J. Phillips in a recent article defining terrorism points out: "There is no 'true' definition." This is perhaps obvious upon reflection, but

[1] For more information about W-ICEWS, see icews.com. Much of the data used in this exercise is publicly available at the Harvard Dataverse: dataverse.harvard.edu.
[2] DPC forecasts have consistently been one of the least accurate in the W-ICEWS program.

making this point explicitly helps justify our effort to justify a category based on its effectiveness, rather than its correspondence with the fundamental ontology of the universe. [4] In the end concepts are not true; they are either more or less useful for doing work.

For the purposes of this paper, a definition is an unquestioned assumption that sets boundary conditions, identifies concepts, or describes relationships between concepts. The process of developing a definition requires or implies lower level theories about the empirical world that are left unquestioned within the context of the model. Definitions cannot be tested directly (i.e. "Is that really what it is?"), but their effectiveness can be. The effectiveness of a definition can be tested by evaluating the quality of work done on a problem by employing it compared to work done on the same problem or type of problem by deploying a substantially different definition. In this paper we will be defining a new variable that goes beyond the Domestic Political Crisis (hereafter "legacy DPC") as an abstract concept. We will then explicate its operationalization in order to verify that coding rules sort data into categories that really do indicate the presence or absence of the variable, as we define it. We will use three criteria to judge the effectiveness of our new definition and its operationalization. We will evaluate whether it is *measurable* by using automatically coded data, we will gauge if it is *predictable* with a statistical model, and we will discuss whether it might be *meaningful* to users.

1.3. Re-defining a domestic crisis

The ambiguity of the legacy DPC was due in part to the conflation of two key levels for potential political crisis: the government and the regime. David Easton among others proposed a useful theoretical framework for understanding political systems using three levels: the political community, the regime, and the government. [5] The political community refers to the shape, extent, and composition of the population that considers itself to be operating within the same political framework. Also referred to as the "state" level, questions about political community pertain to the extent to which members of a society "are sufficiently oriented toward each other" to form a macro-political unit. [5] The regime level encompasses the "rules of the game," meaning the constitutional order or other overarching principles that regulate governance across the political community. Finally, the government level denotes incumbents in the roles for the exercise of political power established by the regime. [6][7] This three-tiered framework of analysis suggests that the depth of a crisis can be measured in part by whether it affects only incumbents (the government); both incumbents and the "rules of the game" (the regime); or the very integrity of the political community, in addition to raising questions about the fate of the regime and incumbent of the current government. We claim that for potential users of our new variable, regime level instability is the most relevant since it reliably implies a "depth" to the crisis that warrants high-level attention from US policy makers.

Developing a redefined version of the legacy DPC also requires distinguishing between "level of disturbance" and "sensitivity to disturbance." If we bring children to a mattress store they can run, play, jump, and jostle one another without fear of catastrophe and without parents needing to do very much rule-making or rule-enforcing. But if we bring children to a gallery featuring elaborate and delicately displayed objects of the glass-blowing art, we know that if the children behave as they would in a mattress store, catastrophe is almost certain to ensue. In both settings we can expect turbulent behavior; but only in the setting that we know is characterized by "fragility" (the glass-blowing gallery) do we believe that turbulence is likely, or even plausibly, linked to "crisis."

The relationship between these two components can be further illustrated with a geological metaphor found in Brownlee (2007) and Pierson (2003). [8][9] The earth's tectonic plates shift gradually over long periods of time. As the plates move, they occasionally generate seismic tremors of various sizes. Understanding the motion of the plates is crucial to understanding, and predicting, the occurrences of earthquakes. The theoretical research we conduct on long-term change is geared toward understanding how jaggedness associated with the slow motion of the plates is related to the sudden "crises" known as earthquakes. By learning how to track gradual change and detect transition patterns from slow processes of adjustment to non-linear shifts, we would be better prepared to detect fragile situations. In other words, levels of geological disturbance happen everywhere, but only in those regions that are extremely sensitive to that disturbance (such as fault lines) should we be concerned that a real crisis (earthquake) may ensue.

There are four possible outcomes when we consider these two variables (turbulence only, fragility only, calm, and crisis). By using turbulence and fragility in conjunction with the idea of a regime-level crisis, we can identify those situations of "domestic political crisis" likely to be most interesting for policy-makers and analysts, i.e. when substantial turbulence is present in the context of regime fragility.

1.4. Operationalizing the concept

We must now convert our abstract definition above into something more concrete, which we will call "Destabilizing Protest." It is defined as the occurrence of mainly non-violent unrest that threatens reigning institutions of authority seriously enough to warrant high-level attention from US policy makers. In order to operationalize the concept, we need to find a reasonable way to independently measure our two key variables: turbulence and regime fragility. We measure "mainly non-violent unrest" (turbulence) as W-ICEWS mass protest event data and "threatens reigning institutions of authority" (regime fragility) using the State Fragility Index.

Our data representing turbulence comes from the W-ICEWS event dataset, which contains daily data of "who did what to whom?" in 167 countries since January 2001. That data is aggregated up to the monthly level and all protests with any actor towards any other actor is included. In addition, all protests that could not be coded as toward any actor are included as well. Although there are some other event data options, none appear as reliable and broad as the W-ICEWS event data.[3]

There are a number of indexes that attempt to measure the "fragility" of the state, although they often imply different theories and paradigmatic starting points.[4] We've chosen the State Fragility Index because it is relatively well known, covers the time period for the W-ICEWS program, and is only missing six of the 167 W-ICEWS countries. According to the State Fragility Index: "A country's fragility is closely associated with its state capacity to manage conflict; make and implement public policy; and deliver essential services and its systemic resilience in maintaining system coherence, cohesion, and quality of life; responding effectively to challenges and crises, and sustaining progressive development."[10] We believe that this concept of fragility captures the key components necessary to operationalize the sensitivity to disturbance of the reigning institutions of authority.

Next we go through essentially a four-step normalization and conversion process to convert the raw data into a binary event for each country-month in the dataset:

- *Normalizing protest:* First we divide protest by logged population of the country, and then by the total number of events captured by the W-ICEWS event coder. This allows us to normalize our data by population and media coverage. Next we divide the new value for each observation by the mean of the entire dataset, thereby setting the new normalized protest's mean to one.
- *Normalizing State Fragility Index*: Next we take the annual State Fragility Index Value and add one (since the least fragile country can have a value of zero) and we divide the new value for each observation by the mean of the entire dataset, thereby setting the new State Fragility Index's mean to one.
- *"Destabilizing Protest" count:* After that, we multiply the normalized protest by the normalized State Fragility Index to get what we refer to as the "Destabilizing Protest" count variable. If a country has an average level of turbulence and fragility then it will have a value of one for this new variable.
- *Conversion to binary "Destabilizing Protest":* Lastly, we artificially set a cut point so that 10% of our observations will become Destabilizing Protests (DPs) and 90% will not.

The output of our process creates a new binary variable measuring the occurrence of Destabilizing Protest that runs monthly from January 2001 to December 2013 for 161 countries. This variable has a .21 Pearson's correlation

---

[3] There are 340,478 protests in our dataset over every country and about 65% of which are coded without a target. The number of events varies widely in part due to population and media coverage. India has 33,643 protests while many countries have merely dozens since 2001. Our data ends in December of 2013.

[4] For example, the Fragile States Index (ffp.statesindex.org), Global Political Risk Index from the Eurasia Group (www.eurasiagroup.net), and the Index of State Weakness from the Brooking Institute (www.brookings.edu/research/reports/2008/02/weak-states-index). Some preliminary testing has shown that the Fragile States Index and State Fragility Index rankings by country are highly correlated for 2013 (~.9 Pearson's Correlation).

with the legacy DPC, and 24.8% of the 161 countries never experience an event (compared to 47.8% for legacy DPC.) Between five and twenty percent of countries experience a destabilizing protest during any given month and the overall percentage of cases is artificially set to 10%. The destabilizing protest dataset provides significantly more variation per country, for example in places like Bangladesh and Haiti where the legacy DPC was a constant *true* or where there were no legacy DPCs in countries like Israel or El Salvador. In addition, there are a number of countries that do have relatively high levels of mass protest, but not experience Destabilizing Protest events due to their low fragility, including Australia, Canada, Germany, New Zealand, Spain, United Kingdom.

To show how the measure works, see Figure 1 below. There are four lines in each image, representing the normalized values for mass protest (dashed line), State Fragility Index (dotted line), the cut point required for an event (flat line at ~2.7), and the Destabilizing Protest count variable (solid line). The grey solid bars show when a country is experiencing a Destabilizing Protest.

In France, we can see that moderate levels of protest do not trigger a Destabilizing Protest, the reason being that France is quite resilient (low fragility) according to the State Fragility Index. However, if mass protest were to reach a point about three times the current maximum for France, it would trigger an event. In Thailand, we can see that the State Fragility Index has remained moderate and steady for the past decade, so a few mass protest spikes have caused Destabilizing Protest events. In Figure 1, we can see two countries in the Middle East, Saudi Arabia and Libya. Due to Libya's more fragile state, Destabilizing Protests have occurred regularly in the dataset, even before the Arab Spring. Besides an increase in recent protests post-2011, Libya has also see a marked rise in its fragility, leading to higher sensitivity to mass protests and therefore a greater frequency of Destabilizing Protests. In Saudi Arabia, on the other hand, there have been a number of mass protests since 2011, but many spikes did not reach the severity of Destabilizing Protest due to the relative resilience of the country.

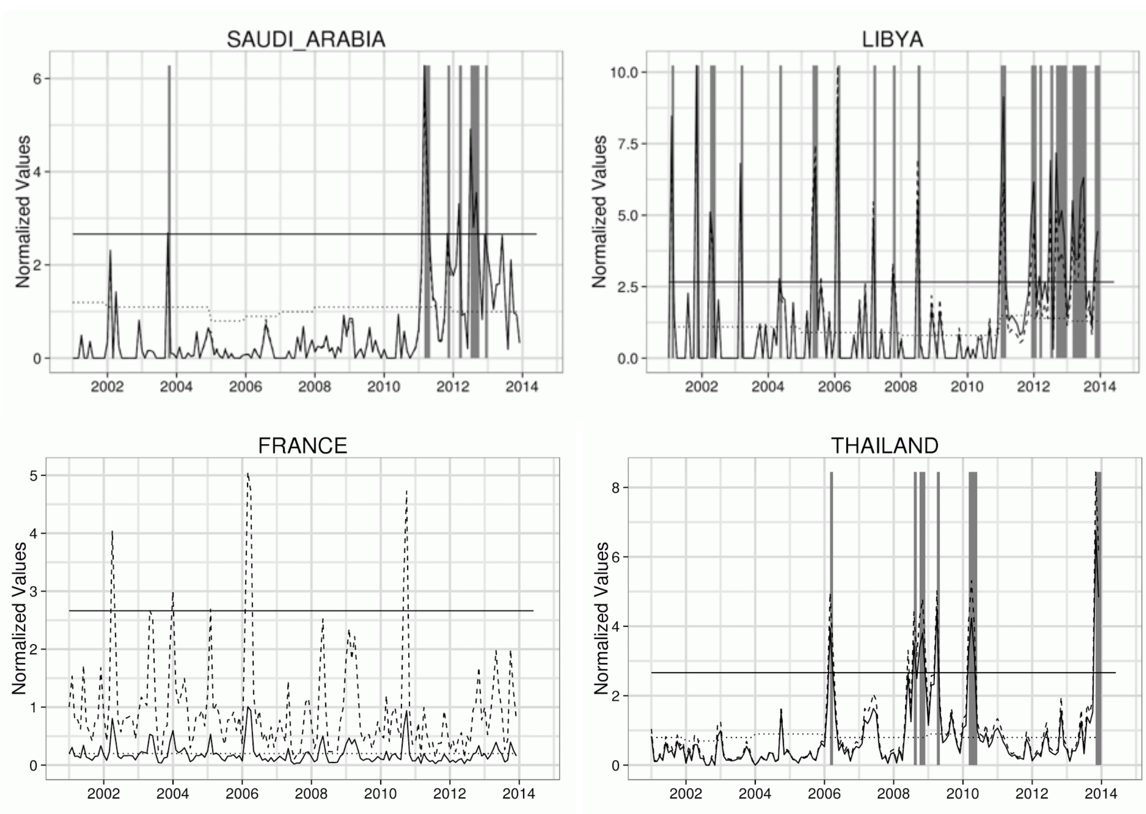

Fig. 1. These images show when a country is experiencing a Destabilizing Protest event (solid bars). The line represent normalized values for mass protest (dashed line), State Fragility Index (dotted line), the cut point required for an event (flat line at ~2.7), and the Destabilizing Protest count variable (solid line).

1.5. Forecasting the new variable

In order to measure the predictability of our new variable, we need to if it is actually something that could be anticipated given current modeling strategies. We developed three models of Destabilizing Protest based on interactions between various sectors in society. Model 1 captures repressive actions taken by the government toward society. Model 2 focuses on social actors' protest, verbal conflict, and cooperative actions with the government.

Model 3 includes dissident protest and hostile actions (violent behavior) directed at the government. These models were then combined to make an ensemble forecast for Destabilizing Protest. The individual component models are relatively 'weak learners', but the resulting ensemble model produced satisfactory results. Recall, or the true-positive rate for the ensemble was 75% in-sample and 78% out-of-sample and overall accuracy both in-sample and out-of-sample was approximately 80%.

Figure 5 shows the in-sample (blue bar at the bottom), out-of-sample (green bar), and forecast (red bar) predictions for three base models (gray) and the ensemble model (red) for a select group of countries. Vertical shaded regions denote the presence of a Destabilizing Protest. Australia and Greece do not experience a Destabilizing Protest and their respective ensemble models predictions are unsurprisingly low. By contrast, Bangladesh and Sri Lanka experience sporadic Destabilizing Protests and the ensemble tends to trend up around the time of these periods. Both Egypt and Yemen experience prolonged periods of Destabilizing Protests and this is where we see the best performance from the ensemble model.

2. Effectiveness of our definition

Earlier in our paper we claimed we would compare our new variable to the ICEWS legacy DPC in terms of its measurability by using automatically coded data, its predictability using a statistical model, and its meaningfulness for users. Each of these considerations is important for judging its effectiveness.

2.1. Measurability

Although it is undeniable that our variable is technically measurable since all of the required data is strictly quantitative, whether the measures are measuring what we think they are is a different question. The W-ICEWS event data has been estimated to be the most accurate event dataset currently available, but we know that certain event types and countries will be less accurate than others. Measuring incidents of protest using this method is a best guess and will introduce some noise into the result. Additionally, the State Fragility Index is coded using a number of specific country-level inputs, some of which will be less reliable than others due to measurement error. It could also be true that the index is "wrong," in that a country could be more fragile than it appears given the best social

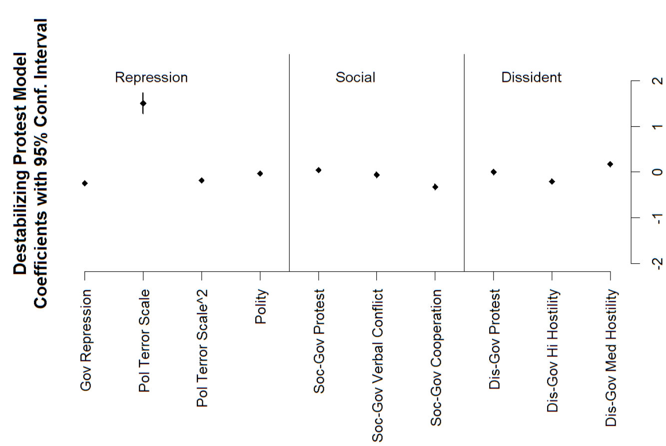

Fig. 2. Model coefficients with 95% confidence intervals are shown for three models of Destabilizing Protest: Government Repression Model, Social-Government Interactions Model, and Dissident-Government Interactions Model.

science theory and data available. Lastly, our artificial ten percent cut point is not ideal since it requires knowledge of the values from the entire dataset. A better method might be to work with users of the variable to determine a training set of key edge cases that could then be used to develop an optimal cut point that clearly codes as many edge cases as possible.

However, even with these downsides the Destabilizing Protest is more precisely and consistently measurable than our legacy DPC, which required dozens of ad hoc assumptions, arbitrary coding rules, and very subjective decision-making by coders. Although the legacy DPC provides some useful distinctions between countries that are or are not experiencing domestic crisis, the fact that it is less easily measured has been a problem.

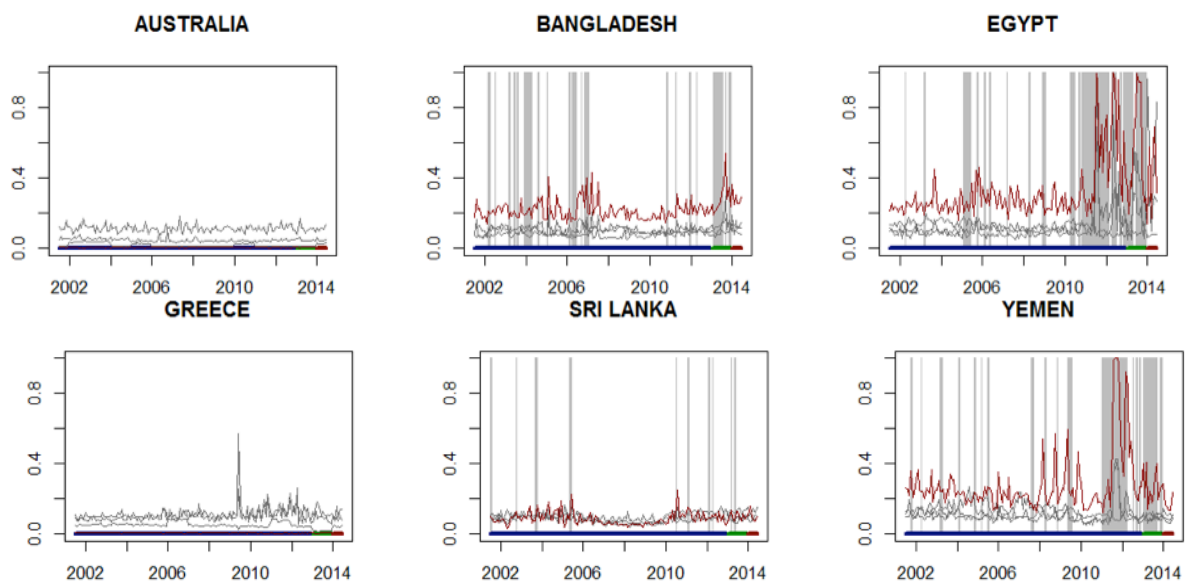

Fig. 3. These charts show the in-sample (blue bar at the bottom), out-of-sample (green bar), and forecast (red bar) as well as predictions for three base models (gray) and the ensemble model (red).

2.2. Predictability

After running our three models and ensemble forecast, we found that both the in-sample and out-of-sample predictions for Destabilizing Protest performed fairly well. Besides the accuracy and recall metrics, however, we believe that the model results could provide more detail for analysts who pay careful attention to changing forecasts. For example, although the forecast is generally low in Sri Lanka, there are spikes during the months with actual events, so an analyst might be alerted not that an event is "probable" but only that its likelihood has increased above the base rate. Similarly, the volatility of a forecast over the past several months might be an indicator to an analyst that some signals for a potential event might be appearing. Overall, these models show that there is an 80% likelihood of correctly identifying whether an event will occur in a given month.

2.3. Meaningfulness

The legacy DPC lacked meaningfulness for users because it was too broad and captured a number of different types of political phenomena. This new measure likely has similar problems, but does so in a more obvious operational way. Mass protest can be triggered by a myriad of phenomena in the political world, but they all lead to the same type of political mobilization. Although we are not measuring the low-level causes, we are measuring the actual observed behavior of political actors. Future versions of this variable could attempt to disaggregate protest into types, or even measure other forms of political mobilization.

We do believe that this variable has face validity in relation to its meaningfulness to users. There are a number of problems that have been resolved from the legacy DPC (these include aforementioned countries with long-running legacy DPCs or no legacy DPCs) and the variable captures much of what we think should be important to policy makers when attempting to determine the domestic political stability of a particular country. It is reasonable that policymakers could use mass protest as a reliable indicator of heightened levels of political mobilization in contexts that may well matter dramatically (that are fragile) and therefore may require more attention and analysis. For example, cases like Libya show that an increase in a country's fragility does indeed increase its likelihood of experiencing a crisis. On the other hand, countries like France should not (and do not according to our operationalization) exhibit Destabilizing Protest due to the country's low-level of fragility. Thus this variable can handle both stability and change in fragility levels, which in turn can increase or decrease the likelihood of an event occurring. In addition, the component variables of Destabilizing Protest can be examined separately to allow users to learn how turbulence and fragility are being measured more precisely.

3. Conclusion

In this paper we have outlined the process of developing a new domestic crisis variable for the W-ICEWS project and evaluated its measurability, predictability, and meaningfulness. Having shown that the principles of fragility and turbulence can be combined systematically and precisely into a variable with some face validity, next steps include more validation and verification to assess the contribution studying this variable can make to analytic, explanatory, and forecasting tasks of interest to both scholars and policymakers.

Acknowledgements

Support for the research reported in this paper was received from the Office of Secretary of Defense/Office of Naval Research (Contract # N00014-14-C-0387).

References

[1] O'Brien, Sean P. "Crisis early warning and decision support: Contemporary approaches and thoughts on future research." International Studies Review 12.1 (2010).
[2] Geddes, Barbara. Paradigms and sand castles: Theory building and research design in comparative politics. University of Michigan Press, 2003, quoted in Phillips, Brian J. "What is a terrorist group? Conceptual issues and empirical implications." Terrorism and Political Violence. (2014).
[3] Gerring, John. "What makes a concept good? A criterial framework for understanding concept formation in the social sciences." Polity (1999): p. 357.
[4] Lakatos, Imre. History of science and its rational reconstructions. Springer Netherlands, 1971.
[5] Easton, David. 1957. "An Approach to the Analysis of Political Systems." World Politics 9 (3).
[6] Easton, David. 1959. "Political Anthropology." Biennial Review of Anthropology 1
[7] Easton, David. 1965. A Framework for Political Analysis. Englewood Cliffs, NJ: Prentice-Hall
[8] Brownlee, Jason. 2007. Authoritarianism in an Age of Democratization. Cambridge: Cambridge University Press.
[9] Pierson, Paul. 2003. "Big, Slow-Moving, and...Invisible: Macrosocial Processes in the Study of Comparative Politics." In Comparative Historical Analysis in the Social Sciences, edited by James Mahoney and Dietrich Rueschemeyer. Cambridge: Cambridge University Press. 177-207.
[10] Marshall, Monty G., and Benjamin R. Cole. "Global Report 2014: Conflict, governance, and state fragility." (2014). http://www.systemicpeace.org/vlibrary/GlobalReport2014.pdf.