Referee Reviews for "An Empirical Evaluation of Explanations for State Repression"

1st Round (10/25/13)

Reviewer 1

I very much like this paper. I think it has the potential to significantly advance the existing literature by providing some insights into what is//is not important for further investigation. This said, I believe that some revisions are necessary to improve the work.

1) I think that theory should not be ignored.

For example, I think that the article should be framed around some theoretical approach: e.g., rational choice, principal-agent arguments or prospect theory - using these to help develop theoretical expectations which would in turn help sort results. At present, it is not clear what the author thinks about why repression is applied and how the various independent variables influence this expectation. Theory would help the author juxtapose domestic vs. international factors as well as diverse aspects of political institutions. THese differences are found in the empirical analysis but in the current version of the manuscript it is not quite clear what one is to do with the diversity. At present, these divisions are ignored and the author acts as if all variables are theoretically equivalent (i.e., that they are all equally likely to be supported). Even the studies conducted are not of equal value. For example, I would argue that all of the internationally-oriented articles mis-specify the domestic factors highlighted by Bueno De Mesquita et al. as well as Davenport and Armstrong.

Theory would help the author think about state repression. For instance, at present there is no way to think about the various repressive indicators used. The CIRI subcomponents are perhaps reflective of important dimensions: overt vs. covert, constraining vs. eliminating citizens/challengers and selective vs. indiscriminate behavior. PTS combines scope, lethality and the degree of political targeting. It is possible that the different CIRI measures provide insight into what is driving PTS. I do not like the fact that the different measures are

simply tossed in without thinking about what they are intended to measure and why different variables influence/don't influence them

2) Related to this, I do not find keeping lagged repression out of the model compelling. The theoretical explanation being drawn upon here includes bureaucratic inertia. Whether or not diverse variables can add explanatory power when the past is considered seems important. Historically, this is one of the variables that has mattered the most.

3) I think that the measurement issues need to be addressed. I immediately thought that the use of Polity was problematic because I knew that the sub-dimensions include elements of repression. Why wait for putting this in the conclusion. Related to this though is the problem of civil war. I realize that previous work has included this in their models but given that repressive behavior is part of the concept, this seems conceptually unacceptable. Run models without this variable in addition to noting this conceptual problem.

4) I'm wondering if the different models employed can be tied to specific theoretical explanations. Do the different approaches provides insights into the use of state repression through their assumptions, ability to consider interactions or some other criteria?

5) The authors state that they cannot think of a reason why democracy would be related to political imprisonment but this seems generally consistent with democratic attempts to hide political repression discussed by Foucault as well as Rejali.

Reviewer 2

This manuscript presents two ways to examine the significance of a whole host of the (seemingly) most important independent variables predicting state repression. The repression literature has become a stable and respected area of scholarship in political science, and it is built on a number of statistical analyses. These studies have introduced a large number of variables to this field that most new scholarship treats as canonical, so that most studies include GDP, civil war, population, etc. without theorizing or even thinking twice about their inclusion. In my experience, reviewers also ask for these canonical controls, even when the scholar's theory does not call for them as potential confounders. This paper performs two types of analysis?cross-validation and random forests?that examine the predictive power of these variables: how much do each of these variables improve our ability to predict state repression? The authors find that very few of these canonical variables actually go far in predicting state repression, with the major exceptions of civil war, natural resources, and democratic institutions, especially courts.

I love this paper. The empirical analyses do a critical service to the domestic conflict literature (covering both human rights and civil wars) in highlighting which of the variables we see so commonly are likely to be important in baseline models predicting repression. More importantly, they demonstrate the value of doing this kind of substantive analysis of our models, to ensure that variables in an estimated model have more meaning than meeting an arbitrary cutpoint. I'm fully on board with the importance of such studies, and I think this one is valuable to the domestic conflict field.

The paper is also very well-written and the analyses well-executed. The authors know exactly what they're doing and they present their results effectively. In other words, I believe this to be a quality manuscript that represents an important contribution to the field of domestic conflict and its respective subfields. I recommend that the authors be invited to revise and resubmit the work to APSR.

My main concern is with the contribution of the piece to political science in general. The manuscript does not explain why the study is important to the field at large: why is this problem with the repression literature representative of a problem common to a great number of literatures? What do we learn from this study *that we did not know* and that can be applied to other types of studies? Most of my points below are subpoints of this main idea and are thus mainly concerns with *framing the contribution* rather than critiques of the research itself.

While these analyses are very new to the repression literature, they are not new in political science or even conflict (e.g., Ward, Greenhill, and Bakke 2010). The authors could go much further in explaining why these particular tests are needed. Why is it important to demonstrate the predictive ability of variables? What do we learn that statistical significance does not tell us? This may seem obvious, but this piece could transform the repression literature if the authors can make a convincing argument that our existing studies are doing a disservice to human rights practices by not addressing this issue. I mean, we all know we should make our models explain as much variance as possible, but most scholars believe they are already doing that. So what is different here, in the concept of predictability, that people are currently missing? The paragraphs at the beginning of section 3 do not go far enough to explain why current methods do not imply generalizability, nor how, in a general sense one can imply generalizability.

The authors do a very nice job of explaining what these tests are doing, and how they assess predictive ability (though I still don't quite understand how decision trees are determined by the approach). I'd like to see a discussion of how scholars can use these in their own work. These will only transform the literature so that they are commonly used if they are accessible to scholars. Teach us how to use them, or how to otherwise identify the predictive ability of our empirical models. Must we rely on these authors to run these analyses every 10 years with our newly introduced variables of interest? It would be better to demonstrate what

scholars should do to demonstrate the importance of their work *in each study*. So how should we apply this to our own work?

Please do not assume we all remember what "parcomp" represents as a subcomponent of Polity scores. Always use words when discussing variables.

The figures are excellent?both descriptive and accessible.

I want to repeat a question I wrote above, because I think that answering it is key to identifying what's unique and critical about this research: why is this problem with the repression literature representative of a problem common to a great number of literatures? In particular, how are the authors solving a common problem that none of us are addressing well? Drawing this out and making it clear will make this piece fit well in our most general of journals.

Reviewer 3

My review focuses on the Random Forests methodology and discussion in the paper. Overall, I see this paper as a timely and appropriate use of the methods for political science/international relations. Despite their popularity in other fields, recursive partitioning methods are extremely new to political science. As such, the paper is an important contribution to the literature. The methods, process, visual representation(graphs) and discussion of results are . The minor revision I suggest is to present the RF method in a clearer and less "jargony" manner. This is a new method for political science, yet - in general - it is relatively straightforward and intuitive. I believe researchers and APSR readers would benefit from a more "gentle" overview and introduction to this method. This should also promote a better overall understanding of your paper and quicker uptake of the method by interested resesearchers.

I recommend looking at some of the language that will be more familiar to social scientists used in the following paper to describe RF - Stroble, Carolin, James Malley and Gerhard Tutz. 2009. "An Introduction to Recursive Partitioning: Rationale, Application and charateristics of Classification and Regression Trees, Bagging and Random Forests." Psychological Methods 14(4): 323-348.

2nd Round (1/30/2014)

Reviewer 1

I like this piece even more than the first time but believe that some additions/clarifications would make it stronger, clearer and broader in its implications:

1) provide some summary statement of the guiding theoretical orientation for most of the work identified (e.g., rationalist, structuralist, etc.), work

through what the basic argument is and how this relates the variables, hypotheses and models selected. This is not addressed at all in the current version and thus we are left without an understanding of why prior researchers did what they did.

2) related to this, what are the implications of the compilation of variables and the theories associated with them. there is a comment buried on page 23 that there are important policy implications involved but this is part of the picture. the dominant theories in political science are involved in the discussion of state repression and the operationalizations/concepts in this community are drawn from the larger literature. the piece still reads as isolated. good references/acknowledgement to weingast et al.

3) i feel that the disaggregated stuff is just tossed in at the moment. the literatures has people that study indices like pts and cirri (e.g., poe/tate and cingranelli and richards), there are people that study single repressive actions (e.g., hathaway), there are people that study distinct combinations (e.g., personal integrity as well as civil liberties restrictions [davenport]) and some work that looks at a bunch of different types (e.g., cingranelli and richards). there is no discussion of the different approaches or the implications of the different approaches. this should be done in order to better contextualize your contribution.

Reviewer 2

Reading this paper again confirms for me its importance to repression studies and the quality of the work the authors have done to address a critical problem in repression scholarship. In my opinion, the authors have responded well and in full to the vast majority of the concerns raised by the editors and the reviewers, and the current manuscript is a marked improvement over what was already a very strong paper.

In this brief review, I will not point out each of the things the revision does well, though I want to stress that I think the paper is of high quality. I wish to point out my remaining concern, and I recommend that the editors take it into consideration, though I do recommend publication of the article.

I take issue with the discussion of the conceptualization/measurement problems of civil wars and democracy. The empirical estimates suggest that civil war, political competition, youth bulges, judicial independence, and natural resources have the most predictive value in terms of increasing accurate prediction over basic models. Of these five measures that are shown to have predictive value, the authors take particular issue with the coding of civil war and political competition, arguing that the measurement of these concepts include the very concept of repression and therefore should not be used as control variables in models predicting repression. In other words, they do a post hoc analysis in an attempt to undermine the estimated results that these variables have predictive

value. While I do not disagree that there are serious issues with using measures with such overlap, I believe that the authors are overstepping the bounds of their analysis and the claims they can make for future research.

The purpose of this article is to evaluate the ability of *specific measures* to add predictive value to baseline models predicting repression. While the authors do spend a bit of time discussing the concepts that the measures approximate in the literature review, all these models do is assess whether these particular measures add predictive value to general models. The analyses here do not: (1) assess the value of *concepts* for predicting repression (which no empirical model can do definitively, of course), nor do they (2) evaluate *explanations* or theories of state repression (which the title explicitly claims).

This would seem to be a nitpicky point, but the language used throughout the paper is slippery with these distinctions. If the authors choose to go into why these MEASURES are not appropriate, they should take great care not to suggest that the CONCEPTS are not appropriate. For instance, there are ways to demonstrate that violent dissent leads to repression that is not reliant on using civil wars as the variable (Davenport 2007 book, Conrad and Moore 2010, Danneman and Ritter Forthcoming). And while political competition is problematic as coded by Polity, that's not to say that democracy (or more likely, its component parts), is not useful as a concept, as the authors point out with the DeMeritt & Young example. The authors make very strong claims in the introduction, literature review, and the discussion of the empirics that suggest they have results that overturn these concepts, when *the findings are that they predictive power.*

Furthermore, the focus on CW and polcomp, while not unwarranted, seems a bit ad hoc. What led the authors to focus on these variables that demonstrated high predictive value, but not on youth bulges or natural resources? The purpose of the paper is not to assess all of the 30-odd measures that we use in terms of their coding or conceptual appropriateness, and the authors do not do so systematically. Yet the language used in the lit review ("we argue?" as if it is a critical point of the paper) and elsewhere suggest the authors are putting meaningful importance on these arguments that are neither fully fleshed out nor a systematic study across the measures used. For instance, on page 6, the authors write, "This claim is strongly supported by our analysis, and we discuss the implications of this problem in more detail below"?the phrasing of which implies that the authors have empirical evidence that repression affects political competition, which they do not. They *infer* that the strong results are because of the correlation, and they point out the problems in the coding.

I want to stress that I don't disagree with their points, but I question the appropriateness of the post hoc theorizing and the strength of these claims FOR THIS PAPER, given the actual goals and analyses therein. It seems to me that this analysis would be better shortened - serve as a warning - and reserved for more careful analysis in another paper when the authors can consider the concepts at work and alternative ways to measure them. Additionally, the

authors should take care regarding language slippage between concepts and measures regarding the claims they make on these two particular variables of concern. I also think that the title should be something like "An Empirical Evaluation of Predictors of State Repression" (or similar), since the authors are not evaluating theories/explanations and make it explicit that this is not their goal.

Finally, as an aside, I think the point about international factors having more complex relationships than scholars tend to propose when including them in their models is rock-solid and so very cool. Way to be.