Referee report for "Punctuated Equilibrium in Comparative Perspective" (Ms.#33410)
Reviewer #4

Overall I am very positively disposed towards this paper. There are very few studies in political science that consider hypotheses regarding the higher moments of distributions beyond the variance (and typically, just means). There is an emerging literature that tests hypotheses regarding the full distribution of policy process outcomes, mostly done by Bryan Jones and his co-authors. But much of the Jones et al work is in the US context, and this paper usefully extends an example of these methods to the comparative literature. The distributional results replicate some of the findings in the Jones et al papers regarding what they term punctuated equilibria, and extend the existing work to questions of interest to comparative scholars regarding the relative "friction" of institutions. Cross nationally, the authors find that the stage of the policy process matters more to friction than the design of specific institutions.

That said, I do believe there are a few points that must be addressed before this paper makes a compelling case for publication. First, and most importantly, this paper relies too extensively on other papers to generate the hypotheses and hypothesis tests, and does too little to justify the expectations that are set out for the data (and distributions). While the argument regarding the kurtosis of the process outcome distributions, as tests of the punctuated equilibrium and progressive friction hypotheses, do have a strong intuitive appeal, the processes described in this paper are complex and the simple intuition regarding the kurtosis estimates is not sufficient to generate compelling hypotheses. The complexities arise from the fact that the data are serial and potentially heterogeneous, but the tests ignore these likely properties of the data and the tests rely instead on a null hypothesis assumes that the data are IID.

As a way to see how robust the intuition regarding kurtosis is to dependence and heterogeneity, I programmed a stylized version of the percent-change processes into Stata to simulate outcome distributions. I found that while it is true that extreme valued distributions generate leptokurtic percent-change aggregate distributions (in Stata, I had available the t-distribution with a small degree of freedom parameter – including the Cauchy distribution – and the gamma distribution with a small alpha), I found that there are reasonable conditions where the normal distribution also generates leptokurtic percent-change aggregated distributions, even under the assumption that the draws from period to period are independent.

In the simulation I drew 50 independent draws for the proportions in 19 categories, and then aggregated the data to a single variable. I found that if the mean of the period-to-period distribution increased over time, but the variance remained constant, I observed simulated kurtosis values for the aggregated data in the range of 50 to 100. Note that even though the aggregated kurtosis estimates here are very large, the conditions generating the data are consistent with the null of the punctuated equilibrium hypothesis, since the draws for each period are independent normal but are simply mean heterogeneous.

In this simulation, the kurtosis for the aggregated distribution explodes even when the underlying process was normal since the underlying mean increased over the periods, and with a constant underlying process variance the variance of the percent change distribution decreased dramatically over time. This is because the denominator for the percents increased in each period and so the variance of the percents across categories in each period decreased, a result that is analogous to the standard error of an estimate decreases when the sample size increases. Thus the process generated a very few large percent changes when the percents are differenced in the early stages of the series and the aggregated data had a large

kurtosis even though the underlying process was normal. Conversely, I found that if the variance increases proportionately with the mean, then the kurtosis for the aggregated data remains at about 3.

My point here is not to suggest that the results are suspect, but that the hypotheses were not generated in this specific paper, and so it's reasonable for a reader to wonder if there is a counter argument to explain the results in a way that is inconsistent with the main hypotheses of the paper. The findings from my little simulation are relevant to this paper since it's likely the underlying means of the process distributions are increasing over time. This is especially true for budgets over the time period analyzed, which in addition then potentially undercuts the progressive friction hypothesis. Because of this possibility, the authors need to use either simulation or analytical methods that are specific to the percent-change process they are modeling to set out counter arguments such as the one I discovered in my simulation. These should set out conditions regarding dependence and heterogeneity that can generate large kurtosis values even for a normal process, and then rule these counter arguments out with the data using appropriate tests.

One other point regarding the serial nature of the data. By ignoring the serial nature of the data, the authors in effect are throwing out useful information about the year-to-year process. The paper argues that attention to a policy topic shifts once that topic crosses some threshold in a given time period, and then there is an over-reaction to that topic. The paper suggests that the overall distributions tend to skew right, which means that relatively small amounts of attention are taken from a large number of old categories, and large amounts are devoted to a small number of new categories. This finding is potentially interesting since it might speak to the cognitive process that underlies punctuated equilibria, but the result is only ecological, since the data are disaggregated. It is possible the overall distribution is composed of many period-specific distributions, some skewed left and some right, but perhaps with a few of the latter dominating. The appropriate test of whether this over-reaction works differently in the two tails of the distribution would be to examine skew across time periods, and to test whether the period-specific distributions are skewed right.

Finally, I would note that the results, even taken as currently presented in the paper, are over determined in that there are many differences between the institutions of various countries. For example, while it's possible that US institutions have more friction in their design compared to parliamentary systems, it's also possible that the US government is more professionalized, especially in the legislative branch, which could offset inefficiencies in the institutions themselves. Thus, the similarity between countries could still possibly be driven entirely by institutions, and their comparative properties, rather than simply due to people's common cognitive architecture. The authors note this, but I think it needs to be better emphasized in the paper. I should emphasize here that what is perhaps most compelling about this paper is it may open many possibilities for comparative research. While it may be that questions remain regarding friction, punctuation, and cognitive restrictions on policy processes, the basic approach of this paper – to test hypotheses regarding the stochastic process that generate data – will likely give others ideas for how to refine the hypotheses and the tests. Indeed, the novelty of the methods, and the rich possibilities for testing hypotheses regarding processes, may open up avenues of research that we cannot even currently imagine.

Summary Evaluation Very Good - the manuscript is very good in several respects with few minor and fixable problems

Overall Recommendation Revise and resubmit -