The paper uses a novel machine-learning algorithm to identify groups of countries with different "pre-conditions" for a banking crisis. The authors find that some countries experience the crises for financial reasons (e.g. rapid credit growth), whereas in other countries the common predecessor is the deteriorating socio-economic situation.

In my opinion, the major contribution of the paper is that it applies a novel statistical procedure to an important economic issue - identifying the common precursors of banking crises.

While I do like the novel approach that the paper takes, I have concerns with the way it is written and the results that the authors obtain. Please see more detailed comments below:

1) Introduction: What is the research question that the authors try to answer in the paper? What is the relevant literature? How is this paper different from the current literature? How are the results different? Are there any new insights about the banking crises? All these questions should be answered in the Introduction, but they are not. Instead, details about the data or statistical methods can be left out, although the method could be described in short.

2) Section 2 ("Data"), line 2: do you analyze crises in period 1976-2007 or 1976-2011? In general, I would move this section after "Methodology", as it is confusing to read about control group, positive and negative cases before one reads about the methodology itself.

3) Section 3 ("Methodology"): since the SD method is quite novel for the economists, I think the paper should present more detailed and more formal description of this method (despite providing the citations). How do you form subgroups? How do you search over rules? What is the threshold for Q? It is probably a good idea to make list of actions that the algorith does: from taking in the pre-formatted data to spitting out the subgroups. Otherwise it is really hard to subject the method and the results to a peer review.

4) Section 4 ("Induced subgroups of crises"): my major concerns is with this section. The results do not seem to be intuitive. I think it is not enough just to describe them - the authors need to explain the logic behind, if they do believe that there is one. The case for intuition gets stronger given the fact that the SD method is basically a "black box": so why should we believe what it tells us?

(a) Subgroup 1: why ageing contributes to instability of financial sector? What is the mechanism for this?
(b) Subgroup 2: why high social security expenditures contribute to banking crises? Why did you even identify subgroup 2 by "high social security"? Is it only due to "expenditures for public health" rule?
(c) Subgroups 1 and 2 are basically identical comprising of developed economies. Aren't the authors simply capture this fact? All developed economies are characterized by high life expectancy and aged population.
(d) Subgroups 4 and 5 are again very hard to distinguish from each other. And again, what is the mechanism going from socio-economic problems to banking crisis?
(e) In Section 5 authors write that crises occur in societies with expansive credit activity, but which are not able to absorb these credits in a proper way. In Subgroup 1, authors claim, it is due to ageing population, in Subgroup 3 it is due to large percentage of young population. Why is it so? I have not heard such arguments before? Can authors provide citations to substantiate their claim?

5) Section 6: I'd drop it and use the free space to expand Section 4. I do not see how governance indicators help predict crises. Take Subgroup 1. There you have two groups of countries, which differ from each other by the dynamics of their WGIs. Yet, all these countries experienced banking crisis. So where is the prediction? That is leaving aside the fact that WGIs themselves are mostly perception-based indicators, so deteriorating socio-economic situation in a country may drive their negative dynamics despite the real quality of governance.

6) Instead of Section 6 (if there is space, or as a next research projects), authors could take their grouping of countries, and check whether it helps predict banking crises in 2012-13. Which subgroups (if any) would Cyprus and Slovenia fall to? Can any other "potential candidates" be identified?