Answers to the reviewer:

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.*

The aim of the paper is to present results of data driven segmentation of the domain of banking crises with publicly available World Bank indicators. The work is different from the current literature mainly in respect of the used methodology. The main insight is that banking crises are not a monolithic phenomenon. The consequences are potentially manifold, especially for future research and development of models. In this paper, in order to demonstrate usefulness of our results in a completely different way, we have used the fact that crises in EU countries in year 2008 are detected as instances of different subpopulations. By a comparative analysis we came to a conclusion that a group of countries in which the financial crises have been very severe have problems with good governance. The result is potentially relevant for policy makers. In spite of the invested effort we have been able to identify neither previous work using data driven methodology in the domain of banking crises nor any result in respect of segmentation of this domain. But we have been very happy to realize that our main conclusion connected with segmentation of the crises in EU in year 2008 is in accordance with the results reported by Francis (2003) that have been obtained by building a model of financial fragility.

2) *Section 2 ("Data"), line 2: do you analyze crises in period 1976-2007 or 1976-2011?*
Correct is 1976-2011. By mistake an incorrect range is given in the first paragraph of Section 2.

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 algorithm 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.*

SD algorithm starts from a list of examples. The examples are described by a set of attribute values that can be numerical or nominal. Each example is either positive or negative. The goal is construction of a set of rules that describe relevant subpopulations of the positive class of examples. The rules are built as conjunctions of features and each feature is a logical test of one attribute value. Features have logical values true and false. Attributes that describe examples can have unknown values but features cannot. Main parameters that define the induction process are:

| | |
|---|---|
| $g$ | generalization parameter. Higher values will result by more general rules that are true for more examples. |
| $min\_support$ | minimal expected quality for each rule. Value defines minimal number of positive examples that must be true for the rule and the minimal number of negative examples that must be false for each feature. Lower values result by more complex solutions and this parameter is implemented by "model complexity" parameter for the service available at http://dms1.irb.hr. |
| $beam\_size$ | number of intermediate solutions (equal 50 for http://dms1.irb.hr implementation). |

Internal variables are:

| | |
|---|---|
| $e$ | number of previous covering. Defined for each positive example as the number for how |

|  |  |
|---|---|
|  | many rules the given example has been true. |
| *cw* | cowering weight. Defined as the value *1/(e+1)* for each positive example. |
| *Q* | weighted quality of a rule or a feature. Defined as *A/(B+g)* where *A* is the sum of *cw* values for all positive examples that are true for the rule (feature) and *B* is the total number of negative examples that are true for the rule (feature). *g* is the previously defined generalization parameter. |

1) Construct the set of features. For all numerical attributes features of the form (att.>value) and (att. <value) are constructed while for nominal attributes features of the form (att. =value) are constructed.
2) Eliminate features that are true for less than *min_support* positive examples or that are false for less than *min_support* negative examples.
3) For all positive examples starting value *e* is equal 0 and cw value is equal 1.
4) Do iteratively steps 5-16. Each iteration constructs one rule.
5) Prepare a starting set of intermediate solutions that consists of single features with highest quality *Q*.
6) Do iteratively steps 7-12. Each iteration improves quality of intermediate solutions.
7) Prepare an empty set of new solutions.
8) For each combination of one intermediate solution and one feature from the set of features do steps 9-10.
9) Build a new rule built as a conjunction of the intermediate solution and the feature. If the new rule is true for less than *min_support* of positive cases, forget it.
10) Compute quality *Q* of the new rule and remember the new rule in the set of new solutions if its quality is among beam_size best new solutions.
11) Copy the set of new solutions into the set of intermediate solutions.
12) Stop if the quality *Q* of any rule in the set of intermediate solutions did not increase compared to the previous iteration.
13) Select the rule with best quality *Q* from the set of intermediate solutions as the best rule.
14) For each positive example that is true for the best rule increase counter *e*. Compute *cw* values.
15) Copy the best rule to set of output rules.
16) Stop when preselected number of rules has been constructed.


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?*

Our interpretation for subgroups 1-3 is that the main driving force is actually the increased credit activity that is generally recognized as a risk fact for banking crises. And that increased credit activity is especially dangerous in countries that are not able to absorb these credits for some reason. In the absence of potentially more useful economic indicators, it turned out that low percentage of active population is related with low credit absorption power of a country. For subgroups 4-5 the interpretation is that banking crises may be expected in countries with some socioeconomic problems. The origins and causes may be various, including political transition, social conflicts and worsening economic conditions. And again, in the absence of more useful indicators it turned out that indicators like stagnating or decreasing life expectancy and stagnating or increasing children mortality may be signs that socioeconomic problems are present in a country. Section 5 is devoted to the analysis of the Subgroups presented in Section 4.


5) *(a) Subgroup 1: why ageing contributes to instability of financial sector? What is the mechanism for this?*

A large proportion of elderly population in developed economies means that the percentage of active population is low. And that reduces a possibility to use the credit potentials in a proper way.

6) *(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?*

Necessary conditions of Subgroup 2, as also for other detected subgroups, can be interpreted in different ways. We agree that more appropriate names may be found. Meaning of selected supporting factors may be also useful in that process. Concretely for Subgroup 2 we think that high life expectancy and low children mortality are the results of high expenditures for public health but also that other factors like high road traffic safety, low crime rates, environmental protection, and working conditions are relevant as well. That is the reason that we prefer the term "high social security" although a better solution may exist.

7) *(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.*

Subgroups 1 and 2 overlap significantly and both are characteristic for developed economies. It would be more appropriate if subgroup discovery algorithm had constructed one bigger subgroup instead of these two. Reasons for such an outcome may be in the characteristics of available data as well as in the choice of parameters that we have used, especially for the generalization parameter $g$. In Section 4 we have presented the results in the form we have actually got them by the application of the subgroup discovery algorithm while in Sections 5 and 6 we analyse them as one group.

It is true that necessary conditions for Subgroup 2 actually describe all developed economies. This is a consequence of the fact that in the year 2008 we had banking crises in many developed economies and in the used data set there is small total number of examples of developed economies without crises. But if we notice that Subgroups 1 and 2 significantly overlap and that a necessary condition for Subgroup 1 is high slope of domestic credits to private sector while one of supporting condition for Subgroup 2 is high absolute value of these credits then it is clear that appropriate interpretation is that high credit activity in developed economies is dangerous and not the fact that an economy is developed.

8) *(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?*

The application of the algorithm resulted in two similar subgroups that partially overlap. The reasons are the same as for Subgroups 1 and 2 and in the Section 5 we analysed them as one subgroup. In no way we are suggesting the existence of the causality relation between socio-economic problems and banking crises. A more appropriate interpretation is that some other problems or processes are in the background and that they have been the cause for both of them. For example, it is known that all countries in transition had significant socio-economic problems and banking crises. Also, in our interpretation in Section 5 we mention Congo and Kenya that suffered from political conflicts resulting by both socio-economic problems and banking crises.

9) *(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?*

Out interpretation is that the common characteristic for Subgroup 1 and Subgroup 3 is a relative small percentage of active population. We do not know for scientific analysis of connections between credit absorption capacities of a country and banking crisis outburst, but we think it must exist. It is known that high availability of credits supports national economy but also that expansive credit activity, which is over some limit, may result by a banking crisis. And we think that at least one part of the credit absorption limit

of a country depends on the number of active population because profit necessary for the repayment of credits can be formed only by the work of this population.

> 10) 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.

Section 6 is central part of the paper for two reasons. The first is that there are statistically relevant differences in respect of good governance data for two groups of EU countries that all experienced banking crises in the year 2008. And the countries like Greece and Italy for which it is known that the crises have been very severe and long-lasting are in the group with bad indicator values, especially in respect of control of corruption. The second reason is that the grouping of the countries is based on the results of subgroup discovery. The significance of the difference detected in respect of good governance data (that have not been used in the induction process) is an indirect proof that subgroups have sense. Our result does not demonstrate predictive quality of good governance data in respect of banking crises but confirms that financial fragility, at least for developed EU countries, is related with good governance. Table 2 is not included for predictive reasons but to indicate that problems in Greece and Italy are far from their solution and that there are also other EU countries that have similar problems.
It is true that World Bank good governance indicators are perception based. But Francis (2003) has demonstrated correlation between financial fragility and quality of governance on a significantly larger set of indicators and a part of them has been more objectively measured.

> 11) 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?

The main goal of the work is not construction of predictive models. The subgroup discovery methodology is appropriate for descriptive analysis tasks because it generates rules that can be interpreted by humans and in this way enables insight into properties of induced subpopulations. But, in order that there exists a real connection between properties and subpopulations, the predictive quality of constructed rules should be as good as possible. Precision is one of the measures used for the evaluation of the quality of predictive models. It is defined as a ratio between the number of correct predictions and the total number of predictions. For rules consisting of necessary conditions for Subgroups 1-5 the precision measured on the used training set is 0.94, 0.84, 0.76, 0.71, and 0.81, respectively. Typically, when used on independent test sets the measured quality of models is significantly lower and hard to estimate. By using World Bank data for the period 2010-2012 it can be noticed that Cyprus satisfies necessary conditions to be in Subgroup 1 while Slovenia satisfies conditions for Subgroups 2 and 5. In Subgroup 1 are besides Cyprus also Singapore and Hong Kong. Conditions of Subgroup 5 are satisfied by in total 27 countries and 13 of them satisfy also the supporting condition of high annual money and quasi money growth. Here we list five of them with largest money growth in the year 2011: Estonia, Chile, Sri Lanka, Malaysia, and Vietnam.