# Opponent's Report on Dissertation Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague
Opletalova 26, 110 00 Praha 1, Czech Republic
Phone: +420 222 112 330, Fax: +420 222 112 304

| | |
|---|---|
| Author: | **Mgr. Diana Žigraiová** |
| Advisor: | Doc. PhDr. Ing. Ing. Petr Jakubík Ph.D. Ph.D. |
| Title of the Thesis: | Essays in Empirical Financial Economics |
| Type of Defense: | **DEFENSE** |
| Date of Pre-Defense: | February 21, 2018 |
| Opponent: | Prof. RNDr. Jiří Witzany, Ph.D. |

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?
b) Is the thesis based on relevant references?
c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?
d) Do the results of the thesis allow their publication in a respected economic journal?
e) Are there any additional major comments on what should be improved?
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

*(Note:* The report should be at least 2 pages long.)

**Content of the Report:**

The Dissertation Thesis is based on four empirical research papers, the first two focusing on predictions of Financial Stress Indicator and systemic events, the third on the relationship between bank competition and financial stability, and the fourth on the dependence of the risk of Czech banks on management board composition.

The answer to questions a), b) and d) above is positive: The first three papers have been already published in respected journals and the fourth as a CNB Working Paper. The thesis contains an overview of the current state of knowledge, it is based on relevant references, and there are original contributions of the author. Regarding question e), I had a few remarks, questions, and recommendations during the pre-defense that have been mostly resolved or answered in the submitted final version of the thesis.

**The first paper** uses a Financial Stress Index (FSI) that is not constructed exactly the same way as the one in the second paper. In this case, the main goal is to find leading financial and macroeconomic indicators in or order to predict the continuous FSI variable. The BMA-MCMC methodology is used in order to estimate the Bayesian Post Inclusion Probability (PIP) and select the most important lagged regression variables. The same remark as in the second paper (first in the pre-defense version), regarding a standard versus non-standard approach to the selection of variables, applies. While in-sample performance appears very good, the out-sample performance indicated only visually in Fig. 2.4 looks quite poor. I recommend also reporting out-of-sample R2 and RMSE performance measures to make a more objective comparison. The poor out-of-sample performance again calls for reconsideration of the modeling approach. Finally, section 3.4.4 briefly reports an alternative binary (FSI) outcome regression with same issues as outlined for the second paper.

*Final defense notes:* The general remarks have been resolved in the final version providing a better description of the BMA technique and interchanging the first and the second paper. I accept the explanation that the notion of "BMA" in literature corresponds to the models used by the author. The author points out that the out-of-sample RMSE have been reported in Table 3.6 (in the original numbering, 2.6 in the final version). But I am still missing out-of-sample R2 that would be normally reported since it is much easier to assess performance using R2 than with RMSE that depends on the scale. Can the author explain why R2 is not reported? It should be also noted that the BMA model RMSE turns out to be significantly larger for all countries compared to the benchmark AR(4) model. Could the author comment the poor performance of the BMA model compared to the benchmark model?

**The second paper** tries to develop an Early Warning System (EWS) for financial crises. My first question is why the paper uses a rather complicated methodological approach instead a straightforward one. Specifically, why is the original "crisis – no crisis" binary outcome indicator transformed to a continuous indicator FSI that is again transformed to binary FSI used finally for the development? The construction of FSI makes sense for a VAR or similar regressions where a continuous target variable is needed but in this case, the original binary crisis indicator is all we need. Note that the Type I and Type II errors of the binary FSI indicator are quite large according to Fig 3.1B introducing unnecessary additional noise into the model. Secondly, while the BMA selection of variables approach is an interesting alternative, in my view a more standard approach to build a logit model should be used, at least as a benchmark model. Specifically, there should be a preselection of variables based e.g. on univariate Gini (or ROC) from the long list of candidate explanatory variables. The preselected variables might be transformed to make their relationship to crisis log-odds ratio approximately linear (hopefully improving performance of the final model), and the final set of variables could be selected using the standard forward, backward, or stepwise selection procedure. The resulting model could be more robust providing better out-of-sample results compared to the reported relatively weak out-of-sample results based on the machine learning BMA approach. In addition, the final model performance should be tested against the "true" crisis indicator (not binary FSI) where even worse out-of-sample performance could be unfortunately expected. I also have not found any standard descriptive statistics indication the number of observations, and in particular of ones and zeros, highlighting the problems in the model development.

*Final defense notes:* Regarding, the crisis – no crisis indicator discussion, I understand the author's arguments for the FSI indicator, although I would still prefer to see a comparison with a development based on the original indicator. I cannot fully agree with the discussion of BMA versus standard stepwise or backward elimination presented in the final thesis response to the referees. It should be pointed out that stepwise selection of variables is a banking industry workhorse in the area of credit risk modeling and so the general arguments in favor of BMA and against the stepwise selection approach should be viewed with certain suspicion. I appreciate that the author developed benchmark models using the backward (not stepwise that usually performs better) selection and compared their performance to the BMA result in tables reported in Table A3.6. However, the results look surprising indicating instability of the models since the AUC between the two approaches, BMA and backward selections, differs by as much as 50% in both directions. If both models were correctly developed the differences should be so large. Could the author provide some explanation of the large differences in AUC?

**The third paper** focuses on a different issue of bank competition and financial stability. It is a meta-analysis paper based on 31 selected studies reporting almost 600 estimates of the regression coefficient between variously defined competitiveness measures and financial stability measures. The main conclusion is that the literature suggests almost no inter-play between competition and stability with a slight publication bias towards the negative relationship (higher competitiveness causing lower stability). The paper is very detailed, in my view maybe too long, but overall it is well written. One small remark is that equation (4.8) can be obtained dividing (4.7) by $SE^2$ and not just SE as stated on p. 102.

**The fourth paper** presents an application of another machine learning method called Random Forests (RF) to the relationship between management board composition and stability (riskiness) of Czech banks. The author collects data from annual reports of 21 Czech banks (2001-2012) such as age of directors, education level, proportion of foreign directors etc., and defines the explained variable as the Z-score, NPL ratio, or profit volatility. The RF technique is then applied providing certain expected (as larger board member tenure implying larger stability) and unexpected results (e.g. higher proportion of MBA among board members implying lower stability) discussed in the conclusion. In my view, as with other data-mining methods, there is a significant danger of data overfitting that is not considered much by the author. In case of single regression trees, in order to achieve robustness, it is important to keep the number of observations in the terminal nodes above certain limit. This can be achieved, for example, by pruning the tree. It is surprising that the author claims that isolated leaves with a few (outlier) observations present rather an advantage. It is true that the overfitting issue can be partially solved by the RF approach but also in this case there are a number of parameters specifying how the individual trees are sampled that need to be fine-tuned in terms of the in-sample versus out-sample performance. My recommendation is to report more details on the RF "growing" procedure and cross-validation test results in order to support plausibility of the conclusions.

*Final defense notes:* The author has added a few additional tables and figures with descriptive statistics, random forest hyper-parameters and reporting their cross-validation performance. Why R2 is not reported rather than MSE? What is the definitions of error rate reported in Figures A5.2-4 (when the dependent variable is continuous)? I am still surprised that there is no limit on the size of the terminal tree nodes or on the depth of the trees. Specifically, according to Table A5.1 there is less than 200 observations and the tree shown in Figure A5.5 has terminal nodes with less than 2% of total observations, i.e. at most 3 which is quite unusual and leads to overfitting. Can the author explain why there is no limit on the random forest tree size?

Overall, answering the questions c) and f), in my opinion, **the thesis would be defendable at my home institution and I recommend the thesis for defense without substantial changes.**

| | |
|---|---|
| Date: | 2. 10. 2018 |
| Opponent's Signature: | |

| | |
|---|---|
| | |
| Opponent's Affiliation: | Prof. RNDr. Jiří Witzany, Ph.D.<br>VŠE |