CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Dissertation Thesis

Essays in Empirical Financial Economics

Author: **Diana Žigraiová**

Supervisor: **Doc. PhDr. Ing. Petr Jakubík, Ph.D.**

Academic Year: 2018/2019

Acknowledgements

I would like to thank my advisor, Petr Jakubík, for his guidance and helpful advice in the course of my PhD studies. I wish to thank the Economic Research Department at the Czech National Bank where most of the essays included in this dissertation came to fruition for providing me with the opportunity to conduct research full time.

I equally wish to express my gratitude to co-authors of all my papers. Namely, I am grateful to Tomáš Havránek, Jan Babecký, Kamil Galuščák and Kateřina Šmídková for their inspiring ideas and the opportunity to collaborate with them.

I acknowledge that this thesis received support from the European Union's Horizon 2020 Research and Innovation Staff Exchange programme under the Marie Sklodowska-Curie grant agreement No. 681228 and from the Czech Science Foundation (grant #15-02411S).

Last but not least, I thank my family for their unconditional support.

Declaration

I hereby declare that I compiled this thesis using only the listed resources and literature and that this thesis has not been used to obtain a different or the same degree.

Prague, July 2018

Diana Žigraiová

Bibliographic record

Žigraiová, Diana. *Essays in Empirical Financial Economics.* Dissertation thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, 2018, 171 pages. Advisor: Doc. PhDr. Ing. Petr Jakubík, Ph.D.

Abstract

This dissertation is composed of four essays that empirically investigate three topics in financial economics; financial stress and its leading indicators, the relationship between bank competition and financial stability, and the link between management board composition and bank risk.

In the first essay we examine which variables have predictive power for financial stress in 25 OECD countries, using a recently constructed financial stress index. We find that panel models can hardly explain FSI dynamics. Although better results are achieved in country models, our findings suggest that financial stress is hard to predict out-of-sample despite the reasonably good in-sample performance of the models.

The second essay develops an early warning framework for assessing systemic risks and predicting systemic events over two horizons of different length on a panel of 14 countries. We build a financial stress index to identify the starting dates of systemic financial crises and select crisis-leading indicators in a two-step approach; we find relevant prediction horizons for each indicator and employ Bayesian model averaging to identify the most useful predictors. We find superior performance of the long-horizon model for the Czech Republic.

The theoretical literature gives conflicting predictions on how bank competition should affect financial stability, and dozens of researchers have attempted to evaluate the relationship empirically. In the third essay we collect 598 estimates of the competition-stability nexus reported in 31 studies and analyze the literature using meta-analysis methods. Our findings suggest that the definition of financial stability and bank competition used by researchers influences their results in a systematic way. We find evidence for moderate publication bias. Taken together, the estimates reported in the literature suggest little interplay between competition and stability, even when corrected for publication bias and potential misspecifications.

The fourth essay investigates how composition of Czech bank management boards affects bank risk. We build a unique data set comprising selected biographical information on the management board members of Czech banks over the 2001-2012 period and combine it with individual bank financial data. Next, we apply a machine learning technique – the random forest – to identify the best predictors of bank risk and further interpret the model output. We find non-linear relationships between average directors' age, average director tenure, the proportion of directors holding an MBA and the proportion of non-national directors and the three observed bank risk proxies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The thesis comprises four separate essays that empirically address three distinct issues in financial economics; namely financial stress and its leading indicators, the relationship between bank competition and financial stability, and the link between management board composition and bank risk. Two of the essays investigate the issues from the global perspective while the other two essays focus on the Czech Republic.

A financial stress index (FSI) measures the current state of stress in the financial system by combining several indicators of stress into a single statistic. In the wake of the global financial crisis monitoring financial stability with financial stress indices has become more prominently used by policy institutions (for example, FSI of Hollo et al. (2012) or the stress index of Johansson and Bonthron (2013) for Sweden) and has also attracted numerous research efforts (for example, Misina and Tkacz, 2009, Illing and Liu, 2006). The first two essays make use of a financial stress index in order to identify leading indicators of financial stress. The first essay builds a cross-country FSI and aims to identify those variables that have predictive power for financial stress in a sample of 25 OECD countries. This essay was published in *Journal of Financial Stability*. The second essay constructs an early warning system (EWS) of financial crises that consists of an FSI and a set of leading indicators and evaluates its performance on Czech data. The second essay was published in *Economic Systems*.

The theory does not provide clear guidance on the expected sign of the relationship between bank competition and financial stability. There exist two opposing views; the competition-fragility hypothesis (for example, Keeley, 1990) that argues that competition hampers stability, and the competition-stability hypothesis (for instance, Boyd and De Nicolo, 2005), which advocates that competition makes the financial system more resilient. The third essay applies a meta-regression analysis to resolve the question of how bank competition affects financial stability. This essay was published in *Journal of Economic Surveys*.

The global financial crisis has also highlighted failures and weaknesses in corporate governance arrangements which did not manage to safeguard against excessive risk-taking in a number of financial services companies (Kirkpatrick, 2009). The fourth essay investigates the relationship between the management board composition of banking institutions in the Czech Republic and their risk. The previous version of this essay appeared in Institute of Economic Studies Working Paper series and Czech National Bank Working Paper series. However, the version included in this thesis has been re-estimated using a machine learning technique, the random forest, to address the relatively small sample size and potential nonlinearities. This version is also currently under review in the *Central European Journal of Public Policy*.

The first three essays are co-authored and I estimate my contribution to be 50%, 90% and 75%, respectively. The last essay is authored solely by me.

The Global Financial Crisis has demonstrated the need to monitor and understand systemic risk, that is, the risk of a widespread disruption in providing financial services caused by impairment of parts of or the financial system as a whole which might adversely impact the real economy (IMF, BIS & FSB, 2009). Lo Duca et al. (2017) in their crises database for European countries identify the majority of crises as complex events which incorporate materialisation of several different risks; for example problems in the banking sector, the materialisation of sovereign risk, sudden adjustments of external positions, or significant asset price corrections (for instance, real estate markets). As for the real costs of financial risks, the average output loss associated with systemic crises is estimated to be 8% of GDP on a sample of European countries over the 1970-2016 period (Lo Duca et al., 2017).

Macroprudential policy is generally defined as the use of prudential actions to limit systemic risk (IMF, 2013). Policymakers have focused on mitigating risks to financial stability in three broad forms: collecting the necessary data and developing early warning approaches in order to identify and monitor systemic risk in the financial system; implementing appropriate prudential regulations to enhance institutions' resilience to shocks; and adopting macroprudential and other policies to contain system-wide risks (Orsmond and Price, 2016). The systemic risk monitoring framework contains a range of indicators and methods that need to be combined with qualitative judgment to help provide decision on when to act. In this respect, macroprudential policy takes into consideration three sets of tools (IMF, 2013). First, countercyclical capital buffers and provisions can be activated to increase resilience of the system to shocks. Second, sectoral tools, such as increases in risk weights for lending to particular segments of the credit market, can be implemented to limit build-up of risks in individual sectors. In this respect, Loan-to-Value and Debt-to-Income ratios can be employed to contain vulnerabilities in the residential housing market. Third, to limit funding risks associated with reliance on vulnerable non-core funding liquidity tools, such as Liquidity Coverage Ratio and Net Stable Funding Ratio, can be engaged. In addition, tools to limit risks from interconnectedness and contagion within the financial system, for instance capital surcharges for global and domestic systemically important institutions or limits on excessively large exposures within the system, are also at policymakers' disposal.

In the first essay (Chapter 2) we address the issue of measuring systemic financial stress and of identifying financial stress leading indicators. We aspire to explain dynamics of a FSI as a continuous measure of financial stress in the panel of 25 OECD countries. Since FSIs are widely used in policy institutions for monitoring financial stability and even for activation of macro-prudential tools, it would be useful to identify leading indicators of financial stress so that policymakers may try to avoid increases in financial stress rather than responding to high levels of stress reactively, that is, responding to signals from EWS.

For the purpose of identifying leading indicators of financial stress we use the stress index proposed by Vermeulen et al. (2015). There are three motives for choosing this index. First, the FSI can be consistently calculated for a large sample of countries. Second, it is available for a relatively long time span. Third, it covers a broad range of financial markets in a country and captures indicators frequently included in multi-country stress indexes. Following the literature on early warning models, we collect 20 potential early warning indicators of financial stress. While there is no theoretical literature on determinants of financial stress that could guide our selection, it has been documented that financial stress is related to occurrence of financial crises (Vermeulen et al., 2015). Next, we employ a relatively novel approach to systematically select the most useful indicators to explain our FSI; Bayesian model averaging. Bayesian model averaging (BMA) resolves model uncertainty with respect to explaining our FSI by running many regressions with different subsets of all possible combinations of financial stress leading indicators. We lag all potential leading indicators by 4 quarters (alternatively by 8 and 12 quarters) in order to balance the need to be potentially informative (the information a variable provides is likely to decline with a longer prediction horizon) and the need to allow for timely policy action. Despite the fact that different leading indicators might be informative for explaining financial stress over different horizons and consequently improve the predictive performance of our models, we keep our setting simple in order to have a more straightforward interpretation.

To evaluate the relationship between the seven BMA-preselected variables and our FSI in the cross-country setting we estimate a panel model with country fixed effects. We find that only the lag of the FSI, the money market rate, the world private credit gap, and the unemployment rate are statistically significant. Our results show that in panel context financial stress is very hard to predict, which suggests that financial stress forerunners likely differ across countries. For this reason, we build country models using a country-specific set of leading indicators which is based on running BMA separately on individual country-level data. It turns out that the fit of the country-level models is substantially better than that of the panel model. However, for none of the G7 countries does the model adequately capture the increase in financial stress during 2008-2009. Although better results are achieved in models estimated at the country level, our findings suggest that financial stress is difficult to predict, despite the acceptable in-sample performance of our models.

In the second essay (Chapter 3) we build an early warning system and evaluate its performance for the Czech Republic. EWS can be broadly defined as functional, data-driven approaches that draw attention to variables associated with past crises with the main objective of alerting policymakers of the potential for future crises (Gramlich et al., 2010). Following Lo Duca and Peltonen (2013) we construct a cross-country FSI for the panel of 14 advanced and developing countries and identify the top 30% of FSI values as financial crises. This essay contributes to the early warning literature in two ways. First, we explore a combination of advanced estimation techniques in developing the multivariate EWS framework over the two horizons of different length. We determine useful prediction horizons for potential crisis-leading indicators by means of univariate logit models, and subsequently we systematically select the most useful indicators using BMA. Second, we add to the literature by extending our aggregate EWS framework to a single post-transition country, the Czech Republic.

As a benchmark for crisis dating we use the crisis database by Laeven and Valencia (2008, 2012). Due to differences in drivers behind individual crises episodes we anchor the selection of potential crisis-leading indicators in the existing EWS literature (Lo Duca and Peltonen, 2013; Babecky et al., 2013, 2014; Jakubik and Slacik, 2013). Therefore, the dataset of potential leading indicators for each country in the panel captures growth in domestic and global asset prices, valuations of domestic and global assets, private credit to GDP ratios, and growth in credit for both domestic and global economy. Moreover, we account for interactions between asset price growth and asset valuations on domestic, domestic and global, and global level. In the same spirit, we interact credit growth and leverage level.

In the absence of a clear theoretical framework linking the potential set of leading indicators with crisis occurrences, we again deal with model uncertainty surrounding crisis-leading indicators by applying BMA. We evaluate the usefulness of our EWS framework against the crisis database by calculating a utility function proposed by Alessi and Detken (2011), which combines the model's false signals and missing events with the policymaker's preference between the two. We find that EWS over longer horizon for the Czech Republic outperforms the model over short horizon.

Despite the fact that EWS based on a more homogenous panel of countries achieves higher usefulness, as measured by the utility function, due to unobserved heterogeneity among countries, building a regional EWS for post-transition economies requires some significant simplifications. First, non-convergence of the BMA sampling algorithm due to a large number of potential leading indicators and too few data points introduces doubts about actual usefulness of the selected indicators and their stability. Second, for the same reasons evaluation of the regional EWS performance is reduced to only in-sample analysis. Based on the results, we can conclude that a suitable EWS framework derived from a panel approach for the Czech Republic should incorporate in addition to transition countries also advanced economies that provide longer time series to avoid computational challenges.

The theory does not provide clear guidance on the expected sign of the relationship between bank competition and financial stability. On the one hand, the competition-fragility hypothesis (represented, for example, by Keeley, 1990) argues that competition hampers stability. Strong competition in the banking sector compels banks to take on excessive risks in the search for yield, which leads to overall fragility of the financial system. On the other hand, under the competition-stability hypothesis (for instance, Boyd and De Nicolo, 2005), increased competition makes the financial system more resilient. A competitive banking sector results in lower lending rates, which support firms' profitability, leading to lower credit risk for banks. Moreover, in uncompetitive environments banks are more likely to rely on their too-big-to-fail position and engage in moral hazard (Mishkin, 1999). Since the early 2000s, dozens of researchers have reported estimates of the competition-stability nexus, but their results vary. In the third essay (Chapter 4) we collect all available estimates of the relation between bank competition and financial stability, and examine them using up-to-date meta-analysis methods. Meta-analysis, a quantitative review of the literature, has been used in economics to synthesize estimates of a relationship in question from primary studies. In contrast to narrative surveys that are to a large extent only useful in discussing the reasons for the heterogeneity observed in the results, meta-analysis can provide policymakers and other researchers with clear guidelines concerning the relationship in question.

Our results suggest that the mean reported estimate of the competition-stability nexus is close to zero, even after correcting for publication bias and potential mistakes in measurement. We find support for publication selection against positive results; that is, some authors of primary studies tend to discard estimates inconsistent with the competition-fragility hypothesis.

We find evidence that data characteristics matter for the reported coefficients corresponding to the relationship between competition and stability. Researchers who use heterogeneous samples of countries (including both developed and developing economies) tend to obtain smaller estimates. The effect of competition on stability is larger in developed countries, but even there the positive effects do not seem to be strong. Next, accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus. We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates.

Furthermore, we show that the definition of the proxy for financial stability is important for the results of primary studies. For example, if dummy variables (usually indicating financial crises) are used as a proxy for stability, the authors tend to report much larger estimates than when a continuous measure of financial stability is used. Similarly, the results of primary studies are systematically affected by the choice of the proxy for bank competition. Studies using the H-statistic (a measure that uses the elasticities of banks' revenue with respect to the banks' input prices) tend to report larger estimates of the competition-stability nexus, while studies that employ the Boone indicator (based on the relation between bank performance, measured by profits, and efficiency, measured by marginal costs) usually show smaller estimates. Next, if the researchers ignore the endogeneity problem in regressing financial stability on bank competition (i.e., that stability may also influence competition), they tend to underestimate the effect.

We also find that controlling for supervisory and regulatory conditions in regressions usually decreases the reported estimates, which supports the notion that banking systems with more activity restrictions and greater barriers to entry are more likely to suffer from systemic financial distress (Beck et al., 2006 a,b). In the last step of our analysis we construct a weighted average of all the estimates and give more weight to the ones whose authors avoid potential mistakes in measurement (such as ignoring endogeneity) and that have better publication characteristics (for example, more citations). Nevertheless, the resulting estimate still points to a very weak or non-existent link between bank competition and financial stability.

The global financial crisis has highlighted the need to study, understand, and improve the corporate governance of financial entities. For example, The BCBS (2006) especially advocates studies of a governance structure composed of a board of directors and senior management. In particular, the European Commission (2010) seeks to improve existing corporate governance practices, that is, the functioning, composition, and skills of commercial banks' boards of directors. In the fourth and last essay (Chapter 5) we investigate the effect of the management board composition of commercial banks on bank risk-taking in the Czech Republic over the 2001-2012 period. Specifically, we examine if and how commercial bank management boards affect bank risk-taking in terms of board size, the average age of directors, director tenure, the proportion of female directors, director education level, and the proportion of non-national directors.

To perform the analysis, we prepare a unique data set that comprises selected biographical information on the management board members of Czech banking institutions and combine it with individual bank financial data to serve as control variables. We apply a modern machine learning technique - the random forest - that allows for modelling nonlinear relationships between explanatory variables and the dependent. It can be applied to small data sets with a large number of predictors since it is insensitive to outliers, robust to adding new observations and robust to overfitting.

First, we grow the random forest on the entire set of board and control variables in order to identify the most useful predictors for the three bank risk proxies with respect to model accuracy. Next, we build the random forest on the sets of the identified best predictors for each bank risk measure and interpret the model predictions in terms of individual variable contributions to the random forest outcome, and by means of partial dependence plots.

We find support for the hypothesis of increasing risk aversion with age in the Czech banking sector, although the director age beyond a certain threshold appears to impact bank stability very little.

Next, decreases in average director tenure on board are found to reduce bank stability while increases in tenure enhance stability. This corroborates the view that quality of board advice and expertise increases over time, once new directors gained sufficient knowledge of the firm to perform appropriate decision-making.

As for directors' education, we find evidence that large increases in the proportion of directors with an MBA enhance bank profit volatility. The effect of small positive and negative changes in their proportion on ROA volatility is nonlinear. The finding thus captures both risk-increasing and risk-reducing implications of directors holding an MBA found in the literature.

Finally, we present evidence that when majority of directors on board are foreigners, bank risk, captured by profit volatility and the NPL ratio, increases substantially. This can be linked to overcoming differences arising from different cultural backgrounds and languages.

References

Alessi, L. and Detken, C., 2011. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/bust Cycles: A Role for Global Liquidity. European Journal of Political Economy, vol. 27(3), pp. 520-533.

Babecky J., Havranek T., Mateju J., Rusnak M., Smidkova K., Vasicek B., 2013. Leading Indicators of Crisis Incidence: Evidence from Developed Countries, Journal of International Money and Finance, vol. 35(C), pp. 1-19.

Babecky J., Havranek T., Mateju J., Rusnak M., Smidkova K., Vasicek B., 2014. Banking, Debt, and Currency Crises in Developed Countries: Stylized Facts and Early Warning Indicators, Journal of Financial Stability, vol. 15(C), pp. 1-17.

Basel Committee on Banking Supervision, 2006. Enhancing Corporate Governance for Banking Organizations." Staff reports, Bank for International Settlements.

Beck, T., Demirguc-Kunt, A. & Levine, R., 2006a. Bank concentration, competition, and crises: First results. Journal of Banking and Finance 30: pp. 1581–1603.

Beck, T., Demirguc-Kunt, A. & Levine, R., 2006b. Bank concentration and fragility: Impact and mechanics. In: Stulz, R. and M. Carey (eds), The Risks of Financial Institutions. National Bureau of Economic Research.

Boyd, J. & De Nicolo, G., 2005. The theory of bank risk-taking and competition revisited. Journal of Finance 60: pp. 1329–1342.

European Commission, 2010. Green Paper: Corporate governance in financial institutions and remuneration policies. Staff report.

Gramlich D., Miller G.L., Oet M.V., Ong S.J., 2010. Early warning systems for systemic banking risk: critical review and modeling implications. Banks and Bank Systems, vol. 5(2), 2010.

Hollo, D., Kremer, M., Lo Duca, M., 2012. CISS - A composite Indicator of Systemic Stress in the Financial System, ECB Working Paper Series 1426.

Illing, M., Liu, Y., 2006. Measuring financial stress in a developed country: An application to Canada. Journal of Financial Stability 2 (3), 243–65.

International Monetary Fund & Bank for International Settlements & Financial Stability Board, 2009. Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations. October.

International Monetary Fund, 2013. Key Aspects of Macroprudential Policy. June.

Jakubik P., Slacik T., 2013. Measuring financial (in)stability in Emerging Europe: A new index-based approach. Oesterreichische Nationalbank (Austrian Central Bank), issue 25, June, pp. 102-117.

Johansson, T., Bonthron, F., 2013. Further development of the index for financial stress for Sweden. Sveriges Riksbank Economic Review 2013:1, 1-20.

Keeley, M., 1990. Deposit insurance, risk and market power in banking. American Economic Review 80: pp. 1183–2200.

Kirkpatrick, G. (2009): The Corporate Governance Lessons from the Financial Crisis. Staff report, OECD Steering Group on Corporate Governance.

Laeven, L. and Valencia, F., 2008. Systemic Banking Crises: A New Database. IMF Working Paper No. 08/224.

Laeven, L. and Valencia, F., 2012. Systemic Banking Crises Database: An Update. IMF Working Paper No. 12/163, June 01, 2012.

Lo Duca, M., Peltonen, T., 2013. Assessing Systemic Risks and Predicting Systemic Events. Journal of Banking and Finance, Vol. 37 (2013), No. 7, pp. 2183-2195.

Lo Duca, Marco & Koban, Anne & Basten, Marisa & Bengtsson, Elias & Klaus, Benjamin & Kusmierczyk, Piotr & Lang, Jan Hannes & Detken, Carsten & Peltonen, Tuomas, 2017. A new database for financial crises in European countries. Occasional Paper Series 194, European Central Bank.

Mishkin, F. S., 1999. Financial consolidation: Dangers and opportunities. Journal of Banking and Finance 23: pp. 675–691.

Misina, M., Tkacz, G., 2009. Credit, asset prices, and financial stress. International Journal of Central Banking 5 (4), 95-122.

Orsmond, David & Price, Fiona, 2016. Macroprudential Policy Frameworks and Tools. Reserve Bank of Australia. Bulletin - December Quarter 2016.

Vermeulen, R., Hoeberichts, M., Vašíček, B., Žigraiová, D., Šmídková, K., de Haan, J., 2015. Financial stress indices and financial crises. Open Economies Review, 26 (3), 383-406.

Chapter 2

Leading Indicators of Financial Stress: New Evidence

Abstract

This paper examines which variables have predictive power for financial stress for 25 OECD countries, using a recently constructed Financial Stress Index (FSI). First, we employ Bayesian model averaging to identify leading indicators of stress. Next, we use those indicators as explanatory variables in a panel model for all countries and in models at the individual country level. It turns out that panel models can hardly explain FSI dynamics. Although better results are achieved in country models, our findings suggest that (increases in) financial stress is (are) hard to predict out-of-sample despite the reasonably good in-sample performance of the models.

This paper was co-authored with Bořek Vašíček, Marco Hoeberichts, Robert Vermeulen and Jakob de Haan, and published in the Journal of Financial Stability (2017, 28, pp. 240-257). This paper is dedicated to the memory of Kateřina Šmídková. The authors would like to thank two anonymous referees for their very helpful comments on a previous version of this paper. The views expressed are those of the authors and do not reflect the views of the European Commission, the Czech National Bank or De Nederlandsche Bank.

2.1 Introduction

Financial stress indices (FSIs) are widely used by policymakers as an instrument for monitoring financial stability. A financial stress index measures the current state of stress in the financial system by combining several indicators of stress into a single statistic. According to Holló et al. (2012: 4-5), a FSI "not only permits the real time monitoring and assessment of the stress level in the whole financial system, but it may also .... be used to gauge the impact of policy measures aimed at alleviating financial instability." From a policy perspective, reliably predicting increases in financial stress is crucial, as this would provide policymakers some time to take measures to alleviate stress. As shown by Vermeulen et al. (2015), spikes in financial stress may appear very abruptly. Since FSIs are now widely used in policy institutions for monitoring financial stability and even for activation of macro-prudential tools[1], it would be very useful to identify leading indicators of financial stress so that policymakers may try to avoid increases in financial stress rather than responding to high levels of stress reactively.

So far, leading indicators of financial stress have received limited attention in the literature. However, there is an extensive line of research predicting financial (especially banking) crises in which several methodologies have been employed (summaries are provided by Demirguc-Kunt and Detragiache, 2005; Demyanyk and Hasan, 2010 and Klomp, 2010). Although most of these „early warning" studies assume that crises are caused by identical factors across countries and that therefore standard panel models can be used, some studies depart from this assumption. For example, Klomp (2010) using a random coefficient logit model for about 130 banking crises between 1970 and 2007, concludes that there exists significant heterogeneity in the causes of banking crises. Although high credit growth, negative GDP growth and high real interest rates are, on average, the most important causes of a banking crisis, none of these variables has a significant impact in more than 60 percent of the banking crises. Similarly, several studies applied binary regression trees (e.g. Davis and Karim, 2008), which allows explicitly for the fact that not all crises are alike and accommodates non-linearities by including conditional thresholds. However, it is a nonparametric approach that cannot estimate the marginal contributions of each explanatory variable or confidence intervals for the estimated thresholds.

[1] For instance, the FSI of Holló et al. (2012) is the first item of the Risk Dashboard of the European Systemic Risk Board. In Sweden, the stress index plays a role in discussions of signals that can be used to activate and deactivate countercyclical capital buffers (Johansson and Bonthron 2013).

Only three earlier papers have examined leading indicators of financial stress. Their results are very mixed. Misina and Tkacz (2009) try to identify leading indicators of the financial stress index of Illing and Liu (2006) for Canada. They conclude that business credit and real estate prices emerge as important predictors of financial stress. Slingenberg and de Haan (2011) use a Financial Stress Index for 13 OECD countries to examine which variables help predicting financial stress. Their findings suggest that financial stress is hard to predict. Only credit growth turns out to have some predictive power for most countries. Several other variables have predictive power for some countries, but not for others. Finally, Christensen and Li (2014) employ the signal-extraction approach to monitor the evolution of a number of economic indicators that tend to exhibit unusual behavior in the period preceding a financial stress event. They combine these variables in three different indicators: the summed composite indicator, the extreme composite indicator and the weighted composite indicator. These composite indicators are used to predict the likelihood of the occurrence of financial stress events within a given period of time. Using the IMF Financial Stress Index (Cardarelli et al., 2011) and 12 indicators for 13 OECD countries, the authors conclude that the composite indicator performs best in terms of out of sample predictions.

One important limitation of previous studies is that they look at restricted set of countries and indicators and do not examine to what extent combinations of several leading indicators affect their results. The purpose of this paper is to examine systematically which variables have predictive power for financial stress in a sample of 25 OECD countries and to examine whether these leading indicators have the same predictive power for different countries[2]. For this purpose we use the stress index recently proposed by Vermeulen et al. (2015)[3]. The main reasons for choosing this index are that i) the FSI can be consistently calculated for a large sample of countries, ii) it is available for a relatively long time span and iii) it covers a broad range of financial markets in a country. Furthermore, this index is fairly representative for other cross-country FSIs as

[2] One may wonder why we do not examine leading indicators of financial crises directly. There are two reasons. First, policy makers rely on FSIs in monitoring financial stability. Second, financial crises occur at low frequency in industrial countries, which makes it hard to examine regularities. Therefore, a FSI can be used as left-hand side variable in an early warning model (instead of a crisis dummy). Duprey et al. (2015) combine the two approaches by converting a continuous measure of financial stress into a binary systemic stress dummy for 27 EU countries.

[3] So the purpose of the paper is not to come up with yet another financial stress index. As will be explained in more detail section 2.2, several stress indexes have been suggested. The stress index used in our analysis captures indicators frequently included in multi-country stress indexes (see the online Appendix for a comparison of several widely used FSIs).

explained in detail in Vermeulen et al. (2015).

As a first step, we gather data for more than 20 potential early warning indicators of financial stress. While there is no theoretical literature on determinants of financial stress that could guide our selection, it has been documented that that financial stress is related to occurrence of financial crises (Vermeulen et al., 2015) we use indicators that have all been suggested in the literature on early warning models of financial, namely banking, crises (e.g. Frankel and Rose 1996; Kaminsky et al. 1998; Klomp, 2010) that have been the most common form of financial turmoil in our sample of OECD countries (Babecký et al., 2014). Next, we employ Bayesian model averaging (BMA) to identify which of those variables are related to our FSI. The systematic approach to selection from large set of potential financial stress forerunners is a major improvement compared to previous studies that aimed at more limited sample of countries and examined a narrower set of potential leading indicators (Misina and Tkacz, 2009; Slingenberg and de Haan, 2011 and Christensen and Li, 2014). BMA is a procedure that allows a subset of the most useful leading indicators of financial stress to be selected from the set of all possible combinations of potential leading indicators (Fernandez et al. 2001; Sala-i-Martin et al. 2004). This also differs from common practice in early warning studies, where usually a limited number of (potential) leading indicators are selected on the basis of the authors' judgment, theory or previous empirical studies[4]. The BMA approach allows us to identify the most important leading indicators of financial stress. Next, we use those variables as explanatory variables in a panel model for all our countries and in models at the individual country level (for the G7 countries only). Since policymakers are primarily interested in variables that may predict high levels of or increases in financial stress, we also estimate our models using variables that measure only high levels of FSI or increases in the FSI. It turns out that panel models can hardly explain FSI dynamics suggesting that financial stress forerunner might differ across countries. Although better results are achieved for models estimated at the country level, our findings suggest that (increases in) financial stress is (are) hard to predict. Whereas the in-sample fit of the country level models is very decent (i.e. the models are able to track most of the FSI dynamics), the out-of-sample predictions are far less impressive.

The paper is structured as follows. Section 2.2 discusses the literature on financial stress and presents the Financial Stress Index used in our analysis. Section 2.3 describes

[4] Misina and Tkacz (2009) and Slingenberg and de Haan (2011) follow the procedure common in the early warning literature. They only consider a limited set of potential leading indicators. Christensen and Li (2014) use a different approach that does not allow identifying the predictive power of individual indicators.

our empirical framework. Section 2.4 presents the outcomes of panel and country-level models using leading indicators selected on the basis of a BMA as explanatory variables of (increases in) financial stress. The last section concludes.

2.2 Financial stress and economic outcomes

Several papers have come up with a FSI for one country (e.g. Illing and Liu 2006) or for several countries (e.g. Cardarelli et al. 2011). In general, stress indexes for a single country combine more stress indicators into one statistic than multi-country stress indexes (for an extensive comparison of FSIs we refer to Kliesen et al. 2012)[5]. This is not surprising in view of data availability. For this reason, the index used in our analysis does not include some sectors, notably the real estate sector and securitization markets, even though there are good reasons for including these segments of the financial system in constructing a FSI (cf. Oet et al. 2012).

We employ the FSI developed by Vermeulen et al. (2015), which consists of 5 widely used variables to capture stress in several segments of the financial system (see Table 2.1 for details). This index is fairly representative of indexes used in cross-country analyses[6]. All variables included in the index are standardized, i.e. we subtract the mean and divide by the standard deviation. The index used is the non-weighted sum of the standardized variables included[7]. The interpretation of the FSI is very straightforward. If the index rises above 0, it indicates an increase in stress; if it is below 0, the financial system is stable. The FSI is calculated for 25 OECD countries using EViews 8.1 (see Figure 2.1).

[5] As pointed out by Vermeulen et al. (2015) FSIs have several limitations. First, they generally do not capture interconnectedness. The same holds for certain other characteristics of the financial system, like the systemic importance of certain financial institutions. Finally, Borio and Drehmann (2009) argue that that the lead with which market prices—on which most FSIs rely—point to distress is uncomfortably short from a policy perspective.

[6] Online Appendix O1 compares the index of Vermeulen et al. (2015) and several other FSIs that recently have been proposed.

[7] Vermeulen et al. (2015) show that using the weighting method proposed by Holló et al. (2012) does not lead to very different results. We therefore prefer giving all the variables the same weight as that makes the index easy to interpret.

Table 2.1: Indicators considered and FSI

FSI1	Stock price volatility derived from a one year rolling GARCH(1,1) specification
FSI2	Volatility of monthly changes in the nominal effective exchange rate as calculated by a one year rolling GARCH(1,1) specification
FSI3	Beta of the banking sector, calculated as cov(return banking sector, total market)/variance(total market)
FSI4	Long-term interest rate - US long-term interest rate (measure of sovereign risk). This variable is zero for the US
FSI5	Inverse yield curve - (long-term interest rate – short-term interest rate), i.e. short-term interest rate – long-term interest rate
FSI	Financial stress index is the non-weighted sum of each financial stress indicator (FSI = FSI1 + FSI2 + FSI3 + FSI4 + FSI5).

Source: Vermeulen et al. (2015).

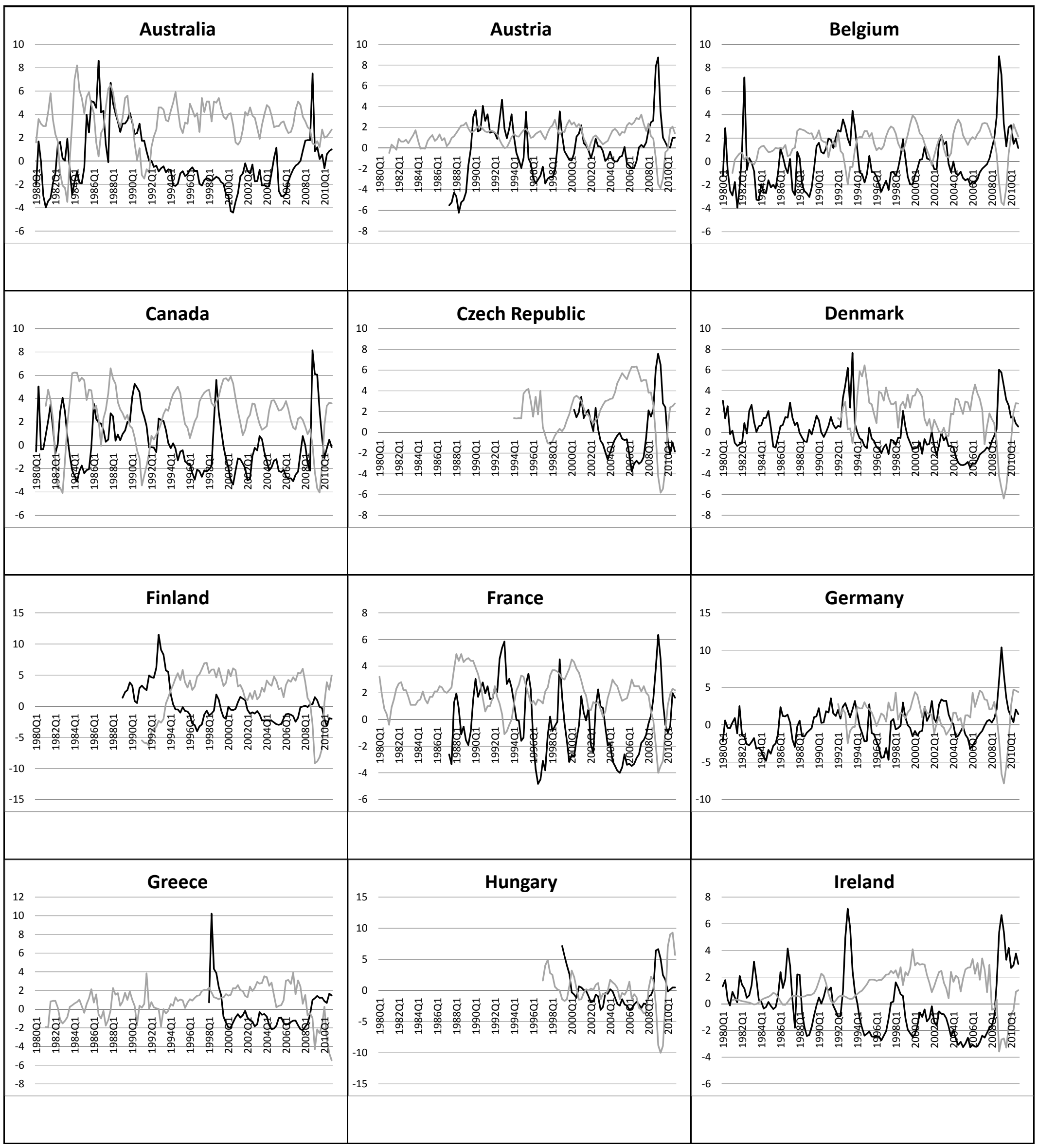
Financial stress indexes have been used for several purposes (see Vermeulen et al. 2015 for a more extensive discussion).[8] For instance, Cardarelli et al. (2011) use their stress index for 17 advanced economies to examine the relationship between financial stress and economic slowdowns. Their findings suggest that episodes of financial turmoil characterized by banking distress are more likely to be associated with deeper and longer downturns than episodes of stress mainly in securities or foreign exchange markets.

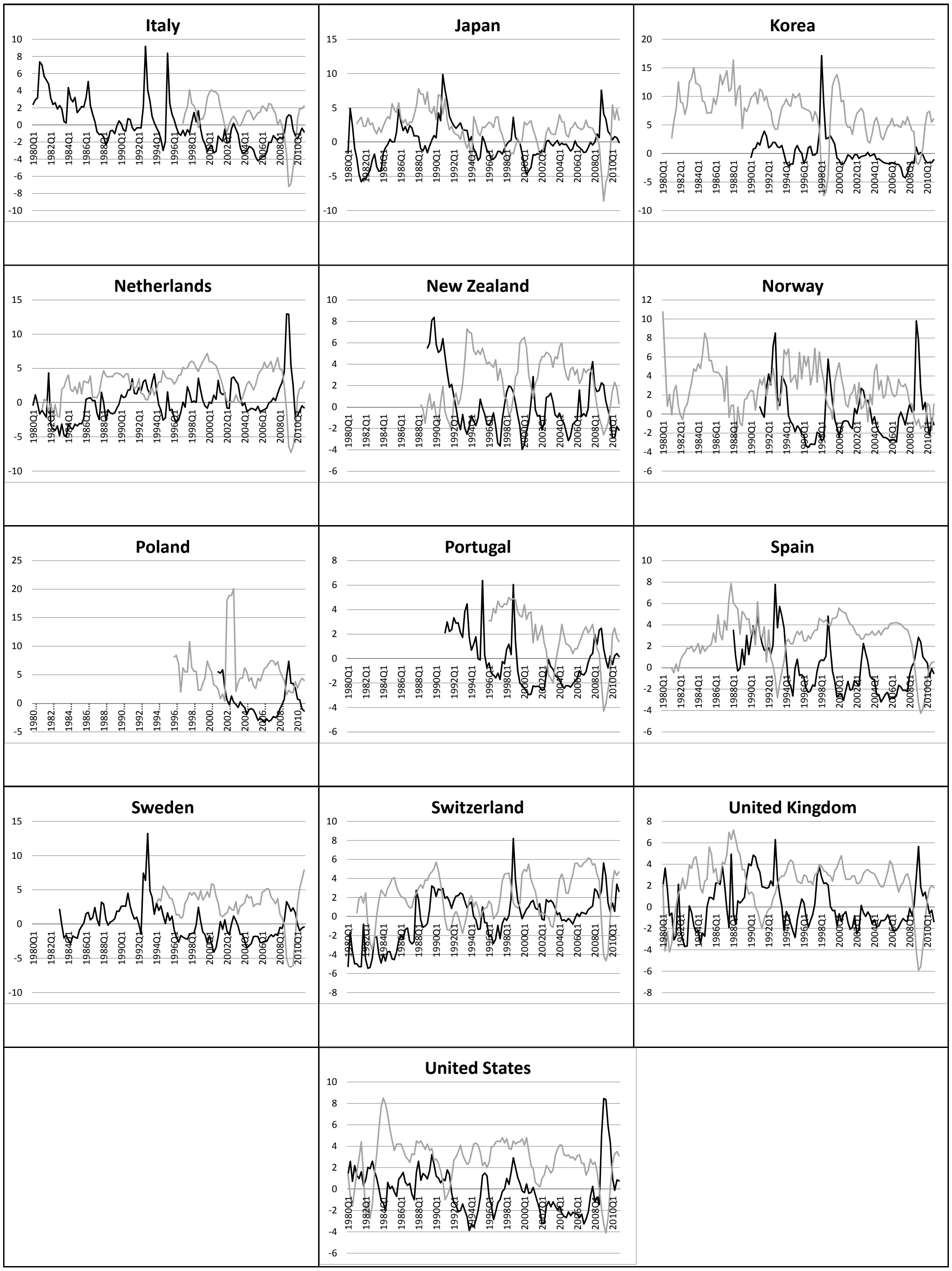
Figure 2.1 shows the FSI used in this paper and year-on-year changes in real GDP (both at quarterly frequency) in 25 OECD countries. Availability of the FSI differs across countries in the time dimension. There is almost an inverse pattern between these two variables in most countries. This pattern is not driven solely by the recent global financial crisis. Periods of above-average financial stress are commonly accompanied by below-average economic growth and vice versa. This inverse pattern is also apparent from Table 2.2 showing the correlation coefficient between the two series at the country

[8] Several recent papers are worth mentioning. Cevik et al. (2013) construct a FSI for Bulgaria, the Czech Republic, Hungary, Poland, and Russia and examine the relationship between financial stress and economic activity. Martin and Milas (2013) estimate Taylor rules in which they include a FSI to model UK monetary policy. Likewise, Baxa et al. (2013) examine whether and how central banks of the USA, the UK, Australia, Canada, and Sweden responded to episodes of financial stress over the last three decades. Mallick and Sousa (2013) analyze the real effects of financial stress in the Euro-zone using VARs. Blot et al. (2015) use the Federal Reserve Bank of St. Louis FSI and the index of Holló et al. (2012) to analyse the relationship between price stability and financial stability. Apostolakis and Papadopoulos (2014) examine financial stress co-movements and spillovers among the G7 economies for the 1981–2009 period, while Apostolakis and Papadopoulos (2015) use the FSI of Balakrishnan et al. (2009) to analyze financial stress spillovers among the banking, securities and foreign exchange markets.

level. While the average contemporaneous correlation between FSIs and GDP growth across countries amounts to -0.37, it is as high as -0.7 for some countries. Moreover, the temporal lead of FSI (vis-à-vis GDP growth) is confirmed by the dynamic correlations. Indeed, it seems that FSI is even more correlated with GDP growth one quarter ahead. Yet, this finding is slightly weaker when we disregard the observations from recent financial crises. However, the correlation between FSI and GDP growth four quarters ahead is much lower and on average only -0.12. So, it seems unlikely that the current level of FSI will help policymakers to mitigate economic losses one year ahead.

Figure 2.1: FSI vs. GDP growth, 1980Q1 - 2010Q4





Note: This figure shows the FSI (black line) and real GDP growth (light grey line) for the 25 OECD countries in our sample.

Table 2.2: Correlation between GDP growth and financial stress

	Full sample			Sub-sample until 4Q 2006		
country	t	t+1	t+4	t	t+1	t+4
AUS	-0.15	-0.21	-0.09	-0.13	-0.18	-0.05
AUT	-0.38	-0.54	-0.40	-0.16	-0.29	-0.32
BEL	-0.40	-0.46	-0.23	-0.25	-0.28	-0.20
CAN	-0.45	-0.48	-0.16	-0.39	-0.40	-0.22
CZE	-0.73	-0.84	-0.57	-0.60	-0.63	-0.74
DNK	-0.59	-0.54	0.00	-0.37	-0.33	0.12
FIN	-0.54	-0.53	-0.34	-0.68	-0.65	-0.44
FRA	-0.39	-0.47	-0.26	-0.24	-0.35	-0.33
GER	-0.56	-0.64	-0.14	-0.35	-0.43	-0.27
GRC	-0.32	-0.30	-0.14	-0.19	-0.28	-0.29
HUN	-0.29	-0.29	0.41	-0.20	0.02	0.39
IRL	-0.67	-0.65	-0.43	-0.55	-0.58	-0.55
ITA	-0.45	-0.51	-0.12	-0.20	-0.33	-0.11
JAP	-0.09	-0.10	0.04	0.13	0.12	0.09
KOR	-0.30	-0.33	0.24	-0.38	-0.43	0.16
NLD	-0.20	-0.36	-0.32	0.08	0.03	-0.10
NZL	-0.55	-0.59	-0.45	-0.65	-0.66	-0.44
NOR	-0.34	-0.33	-0.08	-0.25	-0.25	0.10
POL	-0.25	-0.19	0.41	-0.10	-0.02	0.82
PRT	-0.06	-0.11	-0.01	0.28	0.31	0.40
SPA	-0.37	-0.41	-0.27	-0.39	-0.44	-0.37
SWE	-0.48	-0.40	0.05	-0.14	-0.05	0.05
SWI	0.02	-0.03	-0.04	0.07	0.05	-0.01
UK	-0.29	-0.30	-0.22	-0.20	-0.22	-0.32
US	-0.42	-0.38	0.03	-0.19	-0.22	-0.01
mean	-0.37	-0.40	-0.12	-0.24	-0.26	-0.11

Note: This table shows the correlation between GDP growth and: contemporaneous FSI (columns 1 and 4), FSI one period lagged (columns 2 and 5) and FSI two periods lagged (columns 3 and 6). In columns (1)-(3) the full sample period is used, while in columns (4)-(6) the sample runs until the financial crisis.

2.3 Empirical framework

Given the lack of studies that aim to predict financial stress, we select our list of potential leading indicators from studies on early warning indicators of financial crises following Babecký et al. (2013; 2014) who aim at similar sample of OECD countries. After dropping some variables because of data availability, we are left with a set of 24 potential macroeconomic and financial variables (see Table A2.1 in the Appendix). These variables have all been argued to be linked to financial crises, although studies frequently report different findings for their significance, and sometimes even for their sign (see Table 1 in Klomp, 2010). For instance, Demirgüç-Kunt and Detragiache (1997) argue that high short-term interest rates affect bank balance sheets adversely if banks cannot increase their lending rates quickly enough. Furthermore, Calvo et al. (1993) conclude that capital flows are sensitive to changes in the level of the world interest rate. Large capital inflows and capital flight, particularly in the case of emerging countries, may affect the stability of the financial sector. Our list of variables includes the credit gap and house prices. The motivation for these variables comes from recent research on financial cycles suggesting that they are driven by growth in credit and house prices (see, for instance, Drehmann et al., 2012). The turn of financial cycles comes along with financial instability. Most of our original variables are available at quarterly frequency; for those that are not we use linear interpolation.

Due to the absence of a theoretical framework that links our potential leading indicators to FSI dynamics, the choice of leading indicators to be included in the model needs to be addressed. In principle, we would like to run a regression with our continuous FSI as the dependent variable and all leading indicators as explanatory variables. However, including all potential indicators into one regression is infeasible and would likely lead to many redundant regressors in the specification. We therefore employ Bayesian model averaging (BMA) that deals with the issue of model uncertainty by running many regressions with different subsets of 2^{25} possible combinations of potential variables (Fernandez et al. 2001; Sala-i-Martin et al. 2004)[9]. Thus, under the BMA, many different models γ are estimated based on the following structure:

$$FSI_{i,t} = \alpha_i^{\gamma} + X_{i,t-4}^{\gamma}\beta^{\gamma} + \varepsilon_{i,t}^{\gamma} \quad \varepsilon_{i,t}^{\gamma} \sim (0, \sigma^2 I) \tag{2.1}$$

where $FSI_{i,t}$ is our continuous FSI, α_i^{γ} a constant, β_t^{γ} a vector of coefficients, $\varepsilon_{i,t}^{\gamma}$ an error term and $X_{i,t}^{\gamma}$ a subset of all potential leading indicators. So, each model γ contains a different subset of explanatory variables in $X_{i,t-4}^{\gamma}$. Specifically, all potential leading indicators are lagged by 4 quarters (alternatively by 8 and 12 quarters), which is the

[9] Similarly, Crespo-Cuaresma and Slacik (2009) and Babecky et al. (2014) apply BMA in the context of discrete models of financial crisis occurrence. Furthermore, BMA has also been applied to solve model uncertainty in the field of meta-analysis (e.g. Babecky and Havranek, 2014; Havranek and Rusnak, 2013). Raftery et al. (1997) and Eicher et al. (2011) provide further details on BMA.

common forecasting horizon employed in early warning studies. The aim is to balance the need to be potentially informative (the information a variable provides is likely to decline with a longer prediction horizon) and the need to allow for timely policy action. Therefore, we want to identify the overall macroeconomic conditions that precede financial stress one (alternatively two and three) year(s) ahead. Whereas more complicated lag structures might potentially improve the predictive performance of our models, we prefer to keep our setting simple in order to have a more straightforward interpretation.

BMA gives each model γ a weight, which captures the model's fit (similar to an adjusted R2) and reports weighted averages of the models' regression parameters and standard deviations, using posterior model probabilities from Bayes' theorem:

$$p(M^\gamma | FSI_{i,t}, X^\gamma_{i,t-4}) \propto p(FSI_{i,t} | M^\gamma, X^\gamma_{i,t-4}) p(M^\gamma) \tag{2.2}$$

where $p(M^\gamma | FSI_{i,t}, X^\gamma_{i,t-4})$ is the posterior model probability, $\propto$ a sign of proportionality, $p(y | M^\gamma, X^\gamma_{i,t-4})$ the marginal likelihood of the model and $p(M^\gamma)$ the prior probability of the model. The posterior model distribution of any statistic θ is then obtained from model weighting as follows:

$$p(\theta | M^\gamma, FSI_{i,t}, X^\gamma_{i,t-4}) = \sum_{\gamma=1}^{2^K} p(\theta | M^\gamma, FSI_{i,t}, X^\gamma_{i,t-4}) \frac{p(M^\gamma | FSI_{i,t}, X^\gamma_{i,t-4}) p(M^\gamma)}{\sum_{j=1}^{2^K} p(y | M_j, X^\gamma_{i,t-4}) p(M_j)} \tag{2.3}$$

To express the lack of prior knowledge about the parameters and models, uniform priors are used. For the vector of coefficients β^γ Zellner's g prior is used as Eicher et al. (2011) have shown that the application of the uniform model prior and the unit information prior to the parameters in the model performs well for forecasting. Moreover, a posterior inclusion probability (PIP) is reported for each variable to show the probability with which the variable is included in the true model:

$$PIP = p(\beta^\gamma \neq 0 | y) = \sum_{\beta^\gamma \neq 0} p(M^\gamma | y) \tag{2.4}$$

The large number of potential variables entering into our BMA renders enumeration of all potential combinations of variables not only time consuming but even infeasible (Feldkircher and Zeugner, 2009). Therefore, we use the Markov Chain Monte Carlo (MCMC) sampler developed by Madigan and York (1995) to obtain results for the most important part of the posterior model distribution. The sampler uses a standard birth-death MCMC search algorithm used in most BMA routines to explore the model space

(Feldkircher and Zeugner, 2015). In each iteration step a candidate regressor is drawn $k_C \sim U(1, K)$. A birth step is adding the candidate regressor to the current model M_j if that model did not already include k_c. On the other hand, the candidate regressor is dropped if it is already contained in M_j (death step). The new model is always drawn from a neighborhood of the current one differing only by a single regressor. To compare the sampled candidate model M_i to the current one, the posterior odds ratio is calculated implying the following acceptance probability:

$$\widetilde{p_{\iota J}} = min\left[1, \frac{p(M_i)p\left(Y\middle|M_i\right)}{p(M_j)p\left(Y\middle|M_j\right)}\right]. \tag{2.5}$$

The quality of the MCMC approximation of the actual posterior distribution is linked to the number of draws the sampler is set to go through during the estimation process (iterations). However, the MCMC sampler might start sampling from models that do not yield the best results and only after some time converges to models with high posterior model probabilities. Hence, we discard initial iterations of the sampler (burn-ins).

In our calculations, we set the number of iterations to 5 million after the initial 1 million iterations are discarded as burn-ins. The correlation obtained between iteration counts and analytical posterior model probabilities exceeds 0.95, which we consider as sufficient convergence. This measure indicates the quality of approximation by showing to what extent the MCMC sampler converged to a good approximation of posterior model probabilities. The use of the uniform model prior means that the expected prior model parameter size equals half the number of potential indicators entered into the Bayesian model averaging. However, after having updated the model prior with data it yields a smaller expected posterior model parameter size as the uniform model prior puts more importance on parsimonious models. We prefer parsimonious models, as policy makers can more easily monitor models with fewer variables. We perform the BMA exercise in R using the bms package developed by Feldkircher and Zeugner (2009).

We select all leading indicators that have a posterior inclusion probability larger than 50% and use those variables as explanatory variables in a panel model for all our countries. Next, we run the BMA at the individual country level (for the G7 countries only). Again, we select for each country the variables with a posterior inclusion probability larger than 50% and estimate an OLS model based on the variables that the BMA selects for the respective countries.

2.4 Empirical Results

2.4.1 Panel analysis with FSI

Figure 2.2 presents the results of the BMA exercise for the panel of 25 OECD countries using a lag of four quarters for the leading indicators. So we test whether an indicator is related to our FSI one year ahead. The indicators used are explained in Table A2.1 of the Appendix, while Table A2.2 shows the availability of the data used. The figure depicts the ranking of the variables according to their post inclusion probability (PIP), i.e. the probability that the variable belongs to the "true" model (right-hand side axis). The colours indicate the sign of the coefficient (blue – positive, red – negative, blank – the variable is missing from the model). This model detects seven variables with a PIP higher than 0.5, which is our rule of thumb to select a variable to further analysis[10]. The coefficients of these seven variables are consistent across the different models, although the signs are not always in line with theoretical priors.

As a robustness check we have estimated the model using lags of 8 and 12 quarters for the leading indicators. The results show that different variables are selected by the BMA-procedure for crises 8 or 12 quarters ahead. The BMA-procedure selects 11 variables with a PIP higher than 0.5 for 8 quarters ahead and 8 variables for 12 quarters ahead. Only the money market rate and unemployment rate are selected for all three forecast horizons.

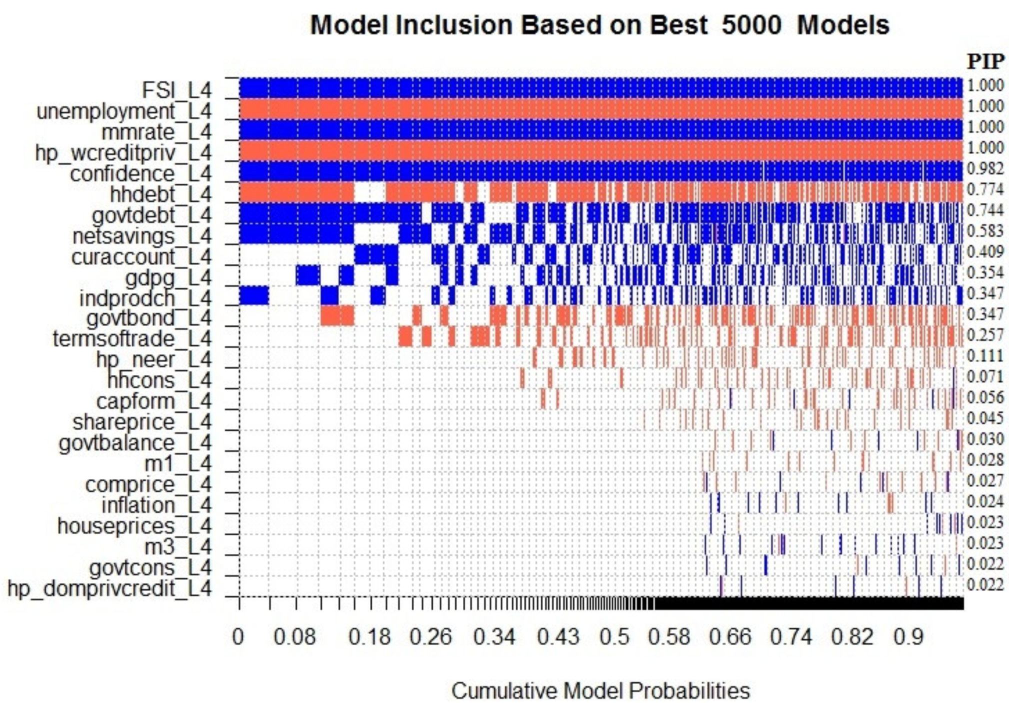
To evaluate the relationship between the seven BMA-preselected variables and our FSI in more detail, we next estimate a panel model with country fixed effects. The first column in Table 2.3 reports the results. It turns out that only four variables are statistically significant, namely the lag of the FSI, the money market rate, the world private credit gap, and the unemployment rate. Most notably, the overall fit of the model is relatively low. Only the money market rate keeps its significance at 8 and 12 quarters ahead (see columns (2) and (3) of Table 2.3). Note that different variables become significant at different forecast horizons, e.g. M3 growth in the eight and twelve quarters ahead forecast. The overall fit of the model further deteriorates. We therefore keep the horizon of four quarters as our benchmark.

[10] Note that the PIP of a variable is a relative probability conditional on the other variables in the model. We deem 0.5 as conservative threshold to disregard irrelevant variables whereas there is no guarantee that the variable with PIP higher than that will be statistically significant at conventional confidence levels in normal regressions.

Finally, we use a pre-crisis sample period that ends in 2006 in order to discard the global financial crisis. As column (4) of Table 2.3 shows, the fit of this model is similar to the model reported in column (1). Apart from money market rate the only other significant variables are commodity prices, exchange rate and government balance.

This panel exercise suggests that it is very difficult to find a set of robust predictors of financial stress across different countries. We have therefore performed a number of other panel exercises, such as allowing for nonlinear effects by using squares and cubes and estimating models for each subcomponent of the FSI, but fail to detect a specification with a substantially higher fit than the benchmark case (these results are available upon request).

Figure 2.2: Bayesian model averaging: leading indicators of FSI, 4Q ahead



Note: Rows = potential FSI predictors. Columns = best models according to marginal likelihood, ordered from left. Full cell = variable included in model, blue = positive sign, red = negative sign. Variables are described in Table A2.1 in the Appendix. L4 means that four lags have been used.

Apparently, within a panel context financial stress is very hard to predict, which suggest that financial stress forerunners might differ across countries. Consequently, forecasting models at the national level may do a better job as not all leading indicators considered may be equally important for all countries (Slingenberg and de Haan, 2011).

Table 2.3: Comparison of results of BMA preselected early warning indicators of FSI (PIP ≥ 0.5) 4, 8 and 12Q ahead for a panel of 25 OECD countries

	(1)	(2)	(3)	(4)
Lags:	4	8	12	4 (pre-crisis)
Constant	-4.75**	-8.35***	-2.12***	-2.4
	(-2.15)	(-3.88)	(-3.6)	(-1.04)
Lag FSI	0.14***			
	(3.38)			
Money market rate	0.37***	0.48***	0.13***	0.41***
	(9.36)	(6.4)	(3.17)	(14.42)
World credit gap	-0.17***			
	(-8.63)			
Unemployment	-0.15***	-0.06	0.02	0.01
	(-3.00)	(-0.65)	(0.17)	(0.15)
Private debt	-0.03			
	(-1.67)			
Production				
Confidence	0.04	0.07***		-0.00
	(1.57)	(3.46)		(-0.03)
Govt. debt	0.01	0.01		
	(1.48)	(1.06)		
Govt. bond yield		-0.28**		
		(-2.49)		
Commodity prices		0.04***		-0.02*
		(3.55)		(-1.8)
Exchange rate		-0.06***		-0.07***
		(-4.86)		(-4.09)
Current account		0.01	0.03	-0.03
		(0.16)	(0.46)	(-0.62)
Capital formation		0.03	0.06***	
		(1.7)	(3.47)	
Stock market		0.01	0.02***	
		(1.66)	(7.3)	
M3 growth		0.05***	0.11***	-0.01
		(3.03)	(5.11)	(-0.87)
Terms of trade			-0.06**	
			(-2.2)	
Govt. balance			0.01	0.08**
			(0.23)	(2.75)
Net savings	0.02			
	(0.91)			
R^2	0.22	0.07	0.09	0.19
Obs.	1586	1486	1386	1186
Count.	25	25	25	25

Note: This table shows results from a panel regression with country fixed effects. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained in Table A2.1 in the Appendix.

Next, we therefore turn our attention to individual countries. In this exercise, we limit ourselves to the G7 countries, which also partially reflects data availability.

2.4.2 Country level analysis with FSI

There are different ways to tackle potential heterogeneity of leading indicators of FSI across countries. The simplest option is to assume that the set of indicators is homogeneous across countries, i.e. to keep the indicators preselected by the panel BMA (as in Figure 2.2), but allow for different marginal effects. The estimation results for these country-specific models (available on request) only give a marginally better fit than the results for the panel model as reported in Table 2.3.

We therefore estimate country models using a country-specific set of leading indicators (based on the BMA results reported in Figure A2.1 in the Appendix). With the exception of Germany the BMA identifies 7 to 10 variables with a PIP above 0.5 for each G7 country. The most striking result is that the fit of the country-level models is substantially better than that of the panel model (see Table 2.4). It is also apparent that there is a lot of cross-country heterogeneity. Interestingly, apart from Japan the lag of the FSI is not significant anymore. Indeed, Figure 2.1 suggests that the persistence of the FSI is relatively low as the index can abruptly change from one quarter to the next.

There are several variables that are significant across various countries although the sign of the coefficients is not always the same. Specifically, we find that falling house prices, decreasing unemployment, decreasing household debt, increasing government bond yields, and increasing government consumption are statistically significant leading indicators of financial stress in at least three out of the seven countries. As our results are derived from a purely statistical approach, we refrain from interpreting them from a theoretical perspective. Still, some findings deserve some attention. First, in contrast to Borio and Lowe (2002), we do not find that credit is a good leading indicator of financial stress. Similarly, Rose and Spiegel (2009, 2010) do not find strong evidence that credit growth is a leading indicator for the recent financial crisis in their cross-country study. Second, our finding that residential real estate prices frequently have good leading-indicator properties is in line with the results of some previous studies, including Adalid and Detken (2007) and Goodhart and Hoffman (2008).

Even allowing for country-heterogeneity, the present approach might be seen as restrictive, as it allows only for a linear relation between each leading indicator and our FSI. Tables A2.3 and A2.4 in the Appendix compare the results of a linear and a non-linear model for two countries where we find some evidence in favour of non-linearities, namely the US and the UK. Specifically, in both countries there seems to be a parabolic relationship between our FSI and house prices. For the US, we also find a non-linearity for M3 growth and for the UK for the world private credit gap. While including these terms further improves the fit of the model, the improvement is only very marginal. We therefore conclude that the linear model seems to be a reasonable approximation of the relationship between the selected leading indicators and financial stress.

Finally, we have re-estimated our model for the US using two alternative financial stress indexes, namely those of the Federal Reserve Bank of Cleveland and of the Federal Reserve Bank of Kansas. Both indices are much broader than our FSI as they do not face the restriction that data should be available for all countries for which we have constructed our FSI. Table A2.5 in the Appendix shows that the estimates are fairly similar to those reported in Table 2.4, so that we conclude that our results are not driven by the use of the FSI proposed by Vermeulen et al. (2015).

Table 2.4: Comparison of results of BMA preselected early warning indicators of FSI (PIP ≥ 0.5) 4Q ahead for individual G7 countries

	USA	UK	JAP	GER	FRA	ITA	CAN
Constant	3.31***	-11.79***	0.05	10.73***	217.45***	-39.82**	-3.47***
	(1.08)	(1.45)	(0.18)	(2.19)	(48.29)	(18.06)	(0.46)
M3 growth	0.34***						0.41***
	(0.08)						(0.08)
House prices	-0.16***	0.19***	0.25***	-1.04***	-0.33***		
	(0.04)	(0.04)	(4.37)	(0.12)	(0.05)		
Domestic credit gap	-0.05		-0.15***		0.23***	0.14***	
	(0.05)		(-7.08)		(0.07)	(0.04)	
Unemployment	-1.29***			-1.54***	-2.38***		
	(0.20)			(0.24)	(0.40)		
Govt. balance	-0.42***	0.09				0.44***	
	(0.10)	(0.06)				(0.13)	
Production	0.14	-0.26***					
	(0.09)	(0.06)					
Private debt	-0.18**	-0.30***			0.18*	-0.39***	-0.14***
	(0.08)	(0.08)			(0.11)	(0.09)	(0.08)
GDP growth	-0.13						
	0.15						
Govt. bond yield	0.57***	1.00***				0.24**	
	(0.10)	(0.10)				(0.12)	

World credit gap		-0.29***				-0.22***	
		(0.07)				(0.07)	
Net savings		0.83***					
		(0.17)					
Capital formation		-0.13***	-0.17***				
		(0.03)	(-5.16)				
Current account		-0.98***					-0.40***
		(0.23)					(0.09)
Exchange rate		0.10**	-0.07***				0.09**
		(0.04)	(-4.05)				(0.04)
Lag FSI			-0.27***				
			(-3.42)				
Money market rate			0.32***				
			(3.3)				
Inflation			0.69***	1.68***	-0.72*		
			(3.95)	(0.31)	(0.39)		
Commodity prices				-0.01	-0.08***	-0.09***	
				(0.02)	(0.02)	(0.02)	
Production					0.66***	0.22***	
					(0.09)	(0.05)	
Confidence					-1.95***	0.42***	
					(0.46)	(0.18)	
Stock market					0.04***		
					(0.01)		
Govt. consumption						0.46***	0.27***
						(0.13)	(0.10)
Terms of trade							-0.16***
							(0.06)
Household cons.							0.36***
							(0.13)
R2	0.57	0.73	0.70	0.56	0.64	0.61	0.49
Obs.	120	89	96	68	91	72	116

Note: This table shows results from OLS regressions. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained in Table A1 in the Appendix.

2.4.3 In-sample and out-of-sample fit

Even though the previous section shows that the in-sample fit of the country level models is relatively decent, the real test of these models is of course how well they predict financial stress out of sample. We therefore re-run the BMA and the regressions for each country using a subsample that ends in 2006. For all countries but the US we find a similar or even slightly better in-sample fit of the models using data up to 2006 compared to the models using data up to 2010 (see Table 2.5). Moreover, Figure 2.3 shows that the in-sample fit of the models using data up to 2006 is quite good.

Table 2.5: Comparison of model fit for whole sample and pre-crisis subsample

	USA	UK	JAP	GER	FRA	ITA	CAN
R^2 (full sample)	0.57	0.73	0.70	0.56	0.64	0.61	0.49
R^2 (subsample)	0.38	0.78	0.78	0.64	0.77	0.70	0.52

Figure 2.4 compares the predicted FSI (based on the parameters of the pre-crisis subsample) and an autoregressive model based on the 4^{th} lag of the FSI[11]. The figure compares out-of-sample rolling forecasts using the coefficients of the model estimated on the subsample ending in 2006 and the respective values of the leading indicators from the 2006Q1 - 2009Q4 period, i.e. corresponding to the prediction horizon from 2007Q1 until 2010Q4 (left-hand side panels in Figure 2.4). The right-hand side panels depict similar out-of-sample forecasts for the autoregressive model.

The results shown in Figure 2.4 are not very encouraging. In fact, for none of the countries does the model adequately capture the increase in financial stress during 2008-2009. This result is quite disappointing in view of the decent in-sample fit of the country models. However, it shall be also noted that financial stress related to global financial crisis might represent rather atypical period as the FSI values recorded during this period for most OECD countries as well as the cross-country financial stress synchronization are historically unique.

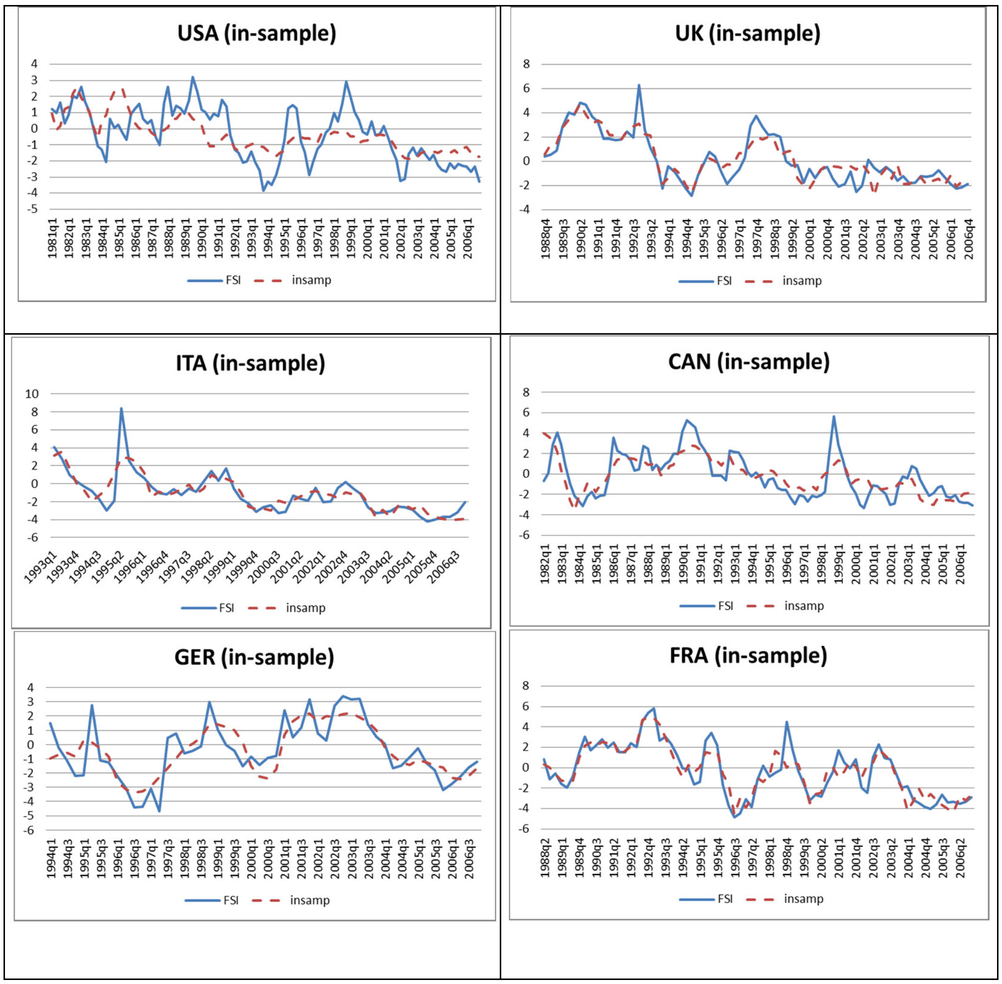
Table 2.6 compares the RMSE of the two models. The out-of-sample performance of the BMA-based leading indicators model is not better than that of the autoregressive model showing the limits of using the selected variables for out of sample forecasts. So even though the explanatory power of the variables selected by the BMA is quite good pre-crisis, trying to predict financial stress during the crisis years using these variables is doing more harm than good.

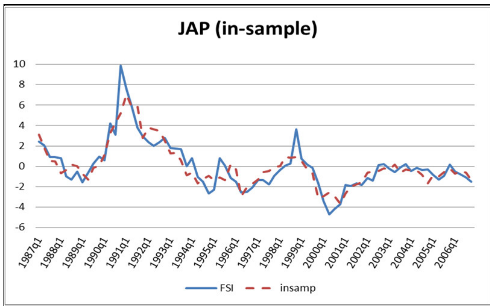
[11] Alternatively, one can choose a random walk as an alternative benchmark model. However, this would be in contrast with the stationary, i.e. mean-reverting, behaviour of the FSI. We opt for the 4^{th} lag because this is a fair comparison with the one year in advance predictive criterion we use in the BMA variable selection procedure.

Table 2.6: Comparison of model fit: Country level models vs. autoregressive models

	USA	UK	JAP	GER	FRA	ITA	CAN
RMSE (BMA)	5.20	7.49	2.47	5.65	9.36	4.53	3.88
RMSE (AR 4$^{\text{th}}$ lag)	4.15	1.94	2.50	3.82	2.65	1.09	3.44

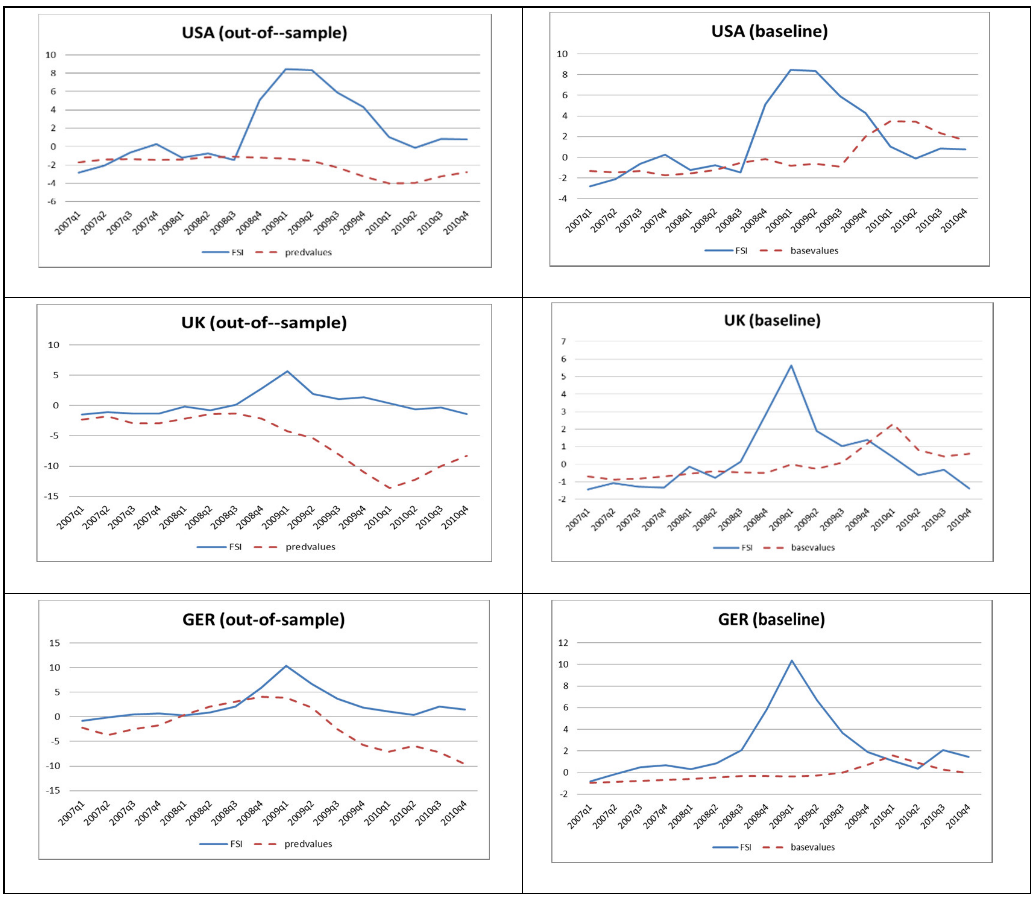
Figure 2.3: In-sample fit of country level models

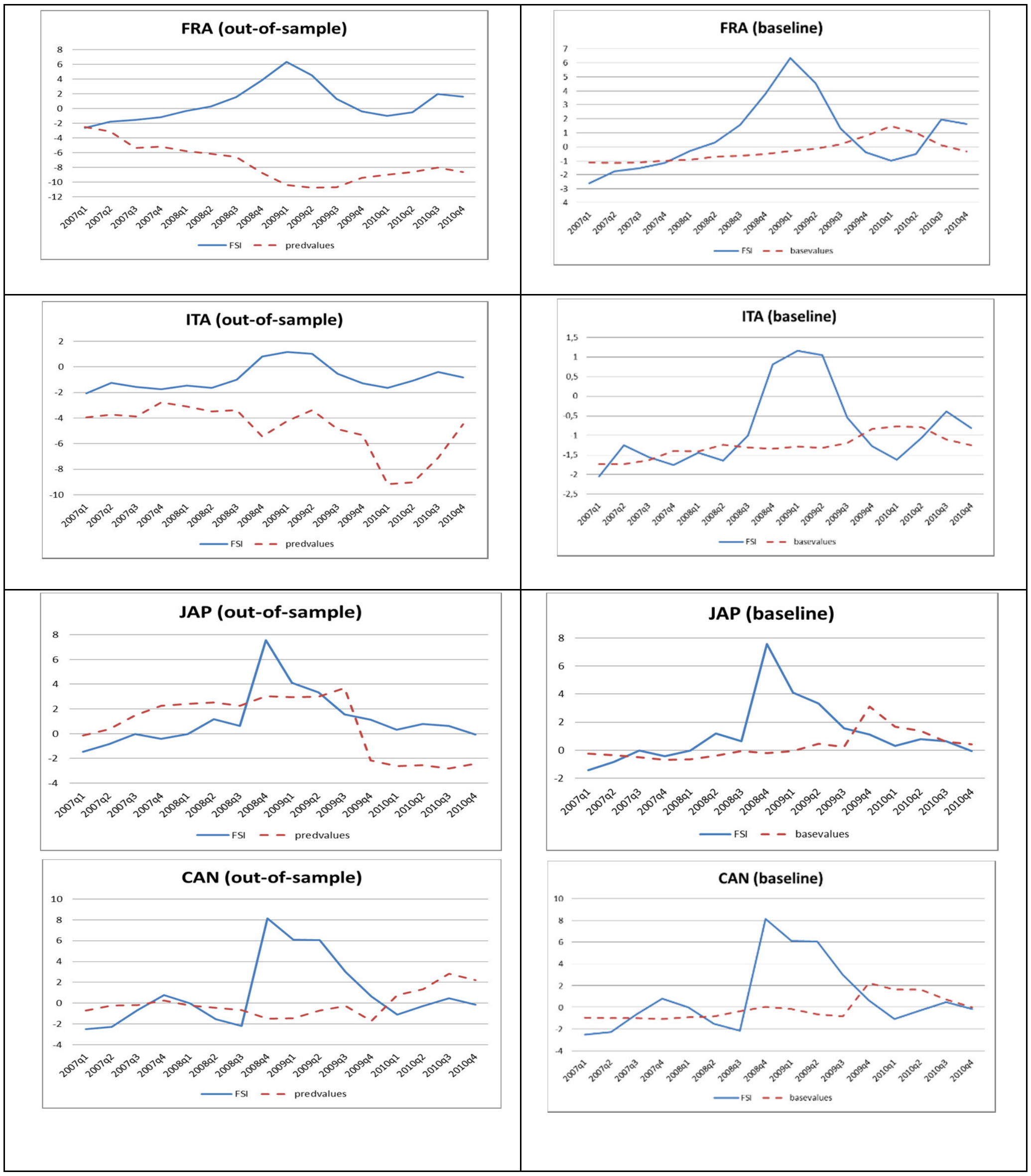




Note: The figures compare the actual level of the FSI with the predicted value (in-sample) according to the models based on BMA-selected variables

Figure 2.4: Out-of-sample fit: country level models vs. autoregressive models





Note: The figures compared the predicted FSI from models based on BMA-selected variables (left panel) and autoregressive models based on the FSI 4th lag only (right-hand side panels).

2.4.4 Thresholds of FSI and increases of FSI

So far, our analysis has been based on predicting the level of our FSI. Since policymakers are primarily interested in variables that may predict high levels or increases in financial stress, we also estimate our models using on the left-hand side a variable that measures whether the FSI is above a particular threshold (in line with Lo Duca and Peltonen, 2013) or the increase of the FSI. First, we transform the FSI into a binary indicator taking value 1 whenever the FSI value is higher than the 80% quantile and 0 otherwise. We estimate logit models with the same country-specific leading indicators as in Table 2.3. Table A2.3 and Figure A2.2 in the Appendix report the results. The findings are largely in line with those in Tables 2.3 and 2.4, i.e. even when we aim at peaks of FSI only, most of the variables are still significant and the model has a decent in-sample fit. However, the out-of-sample performance is poor and also when threshold effects are considered it is difficult to outperform a simple autoregressive model.

An alternative approach is to focus on increases in FSI. We compute year-on-year changes in the FSI and use a Tobit-model to analyze whether the BMA-preselected variables are able to predict increases in the FSI 4 quarters ahead. Table A2.4 in the Appendix presents the results of the Tobit-regressions. Since we transform the data from levels to changes, it is not surprising that the in-sample fit, as measured by Pseudo-R^2, is slightly worse than in previous regressions. The signs and significance of the variables are largely in line with the logit regressions and confirm our earlier results. Again, the out-of-sample performance is relatively poor (not shown).

Conclusions

Rose and Spiegel (2010:15) conclude that "Despite a broad search, we have been unable to find consistent strong linkages between pre-existing variables that are plausible causes of the Great Recession and the actual intensity of the recession." Similarly, our results suggest that it is hard to identify leading indicators of financial stress. We have examined which variables have predictive power for financial stress in a sample of 25 OECD countries, using the Financial Stress Index (FSI) of Vermeulen et al. (2015) which is fairly representative of stress indices used in cross-country analyses. First, we have used Bayesian model averaging to identify leading indicators of our FSI. Next, we have used those indicators as explanatory variables in a panel model for all our countries and in models at the individual country level. It turns out that panel models can hardly explain FSI dynamics. Our models generally do not predict an increase in financial stress before the recent financial crisis. Although the unprecedented nature of the financial crisis may play a role here, our results are in line with previous studies suggesting that financial turmoil in general is hard to predict. Although better results are achieved in models estimated at the country level, our findings suggest that (increases in) financial stress is (are) hard to predict, even though the in-sample performance of our models is quite acceptable.

The results in this paper show that policymakers will face difficulties when trying to proactively avoid potential stress in financial markets. It is a challenging task for models to predict the abrupt changes in financial stress. Furthermore, the potential drivers of financial stress differ across countries and may differ as well across stress episodes. The lack of predictability implies that policymakers need to be equipped with flexible tools to respond quickly to emerging financial stress, since long policy implementation lags may aggravate the financial stress episode and the negative effects on the real economy.

References

Adalid, R., Detken, C., 2007. Liquidity shocks and asset price boom/bust cycles. ECB Working Paper 732.

Apostolakis, G., Papadopoulos, A.P., 2014. Financial stress spillovers in advanced economies. Journal of International Financial Markets, Institutions and Money 32, 128–149.

Apostolakis, G., Papadopoulos, A.P. 2015. Financial stress spillovers across the banking, securities and foreign exchange markets. Journal of Financial Stability 19, 1–21.

Babecký J., Havránek T., 2014. Structural reforms and growth in transition: A meta-analysis. Economics of Transition 22 (1), 13–42.

Babecký J., Havránek T., Matějů J., Rusnák M., Šmídková K., Vašíček B., 2013. Leading indicators of crisis incidence: Evidence from developed countries. Journal of International Money and Finance 35, 1-19.

Babecký J., Havránek T., Matějů J., Rusnák M., Šmídková K., Vašíček B., 2014. Banking, debt and currency crisis: early warning indicators for developed countries. Journal of Financial Stability 15, 1-17.

Balakrishnan, R., Danninger, S., Tytell, I., Elekdag, S.A., 2009. The Transmission of Financial Stress from Advanced to Emerging Economies. IMF Working Paper 09/133.

Baxa, J., Horváth, R., Vašíček, B., 2013. Time-varying monetary-policy rules and financial stress: Does financial instability matter for monetary policy? Journal of Financial Stability 9, 117–138.

Blot, C., Creel, J., Hubert, P., Labondance, F., Saraceno, F., 2015. Assessing the link between price and financial stability. Journal of Financial Stability 16, 71–88.

Borio C., Drehmann, M. 2009. Towards an operational framework for financial stability: 'Fuzzy' measurement and its consequences. BIS Working Paper 284.

Borio, C., Lowe, P., 2002. Asset prices, financial and monetary stability: Exploring the nexus. BIS Working Paper 114.

Calvo, G., Leiderman, L., Reinhart, C., 1993. Capital inflows and real exchange rate appreciation in Latin America: The role of external factors. IMF Staff Papers 40(1), 108-151.

Cardarelli, R., Elekdag, S., Lall, S., 2011. Financial stress and economic contractions, Journal of Financial Stability 7, 78-97.

Cevik, E.I., Dibooglu, S., Kutan, A.M., 2013. Measuring financial stress in transition economies. Journal of Financial Stability 9, 597–611.

Christensen, I., Li, F., 2014. Predicting financial stress events: A signal extraction approach. Journal of Financial Stability 14, 54–65.

Crespo-Cuaresma, J., Slacik, T., 2009. On the determinants of currency crisis: the role of model uncertainty. Journal of Macroeconomics 31 (4), 621–632.

Davis, E. P., Karim, D., 2008. Comparing Early Warning Systems for Banking Crises. Journal of Financial Stability 4(2), 89–120.

Demyanyk, Y., Hasan, I., 2010. Financial crises and bank failures: A review of prediction methods. Omega 38, 315–324.

Demirgüc¸-Kunt, A., Detragiache, E., 1997. The determinants of banking crises in developing and developed countries. IMF Staff Papers 45(1), 81-109.

Demirgüc-Kunt A., Detragiache E. 2005. Cross-country empirical studies of systemic bank distress: a survey. National Institute Economic Review 192, 68-83.

Drehmann, M., Borio, C., Tsatsaronis, K., 2012. Characterizing the financial cycle: don't lose sight of the medium term! BIS Working Paper 380.

Duprey, T, Klaus, B., Peltonen, T., 2015. Dating systemic financial stress episodes in the EU countries. ECB Working Paper 1873.

Eicher, T.S., Papageorgiou, C., Raftery, A.E., 2011. Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. Journal of Applied Econometrics 26 (1), 30–55.

Feldkircher, M., Zeugner, S., 2009. Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging. IMF Working Paper 09/202.

Feldkircher, M., Zeugner, S., 2015. Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. Journal of Statistical Software, vol. 68 (4), pp. 1-37.

Fernandez, C., Ley, E., Steel, M., 2001. Benchmark priors for Bayesian model averaging. Journal of Econometrics 100 (2), 381–427.

Frankel, J.A., Rose, A.K., 1996. Currency crashes in emerging markets: an empirical treatment. Journal of International Economics 41 (3–4), 351–366.

Goodhart, C., Hofmann, B. 2008. House prices, money, credit and the macroeconomy. ECB Working Paper 888.

Havránek, T., Rusnák, M., 2013. Transmission lags of monetary policy: A meta-analysis. International Journal of Central Banking 9(4), 39–76.

Holló, D., Kremer, M., Lo Duca, M., 2012. CISS – A "Composite Indicator of Systemic Stress" in the financial system. ECB Working Paper 1426.

Illing, M., Liu, Y., 2006. Measuring financial stress in a developed country: An application to Canada. Journal of Financial Stability 2 (3), 243–65.

Johansson, T., Bonthron, F., 2013. Further development of the index for financial stress for Sweden. Sveriges Riksbank Economic Review 2013:1, 1-20.

Kaminsky, G.L., Lizondo, S., Reinhart, C.M., 1998. The leading indicators of currency crises. IMF Staff Papers 45 (1), 1–48.

Kliesen K.L., Owyang M.T., Vermann E.K., 2012. Disentangling diverse measures: A survey of financial stress indexes. Federal Reserve Bank of St. Louis Economic Review September/October 2012: 369-397.

Klomp, J., 2010. Causes of banking crises revisited. North American Journal of Economics and Finance 21, 72–87.

Lo Duca M., Peltonen T., 2011. Macro-financial vulnerabilities and future financial stress – assessing systemic risks and predicting systemic events. ECB Working Paper 1311.

Madigan, D., York, J., 1995. Bayesian graphical models for discrete data. International Statistical Review 63 (2), 215–232.

Mallick, S.K., Sousa, R.M., 2013. The real effects of financial stress in the Eurozone. International Review of Financial Analysis 30, 1–17.

Martin, C., Milas, C., 2013. Financial crises and monetary policy: Evidence from the UK. Journal of Financial Stability 9, 654–661.

Misina, M., Tkacz, G., 2009. Credit, asset prices, and financial stress. International Journal of Central Banking 5 (4), 95-122.

Oet, M.V., Bianco, T., Gramlich, D., Ong, S. 2012. Financial Stress Index: A Lens for Supervising the Financial System Federal Reserve Bank of Cleveland Working Paper 12-37.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociological Methodology 25, 111–163.

Rose, A.K., Spiegel, M., 2009. Cross-country causes and consequences of the 2008 crisis: International linkages and American exposure. NBER Working Paper 15358.

Rose, A.K., Spiegel, M., 2010. Cross-country causes and consequences of the crisis: An update. CEPR Discussion Paper 7901.

Sala-i-Martin, X., Doppelhofer, G., Miller, R., 2004. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. American Economic Review 94 (4), 813–835.

Slingenberg, J.W., de Haan, J. 2011. Forecasting financial stress. De Nederlandsche Bank Working Paper 292.

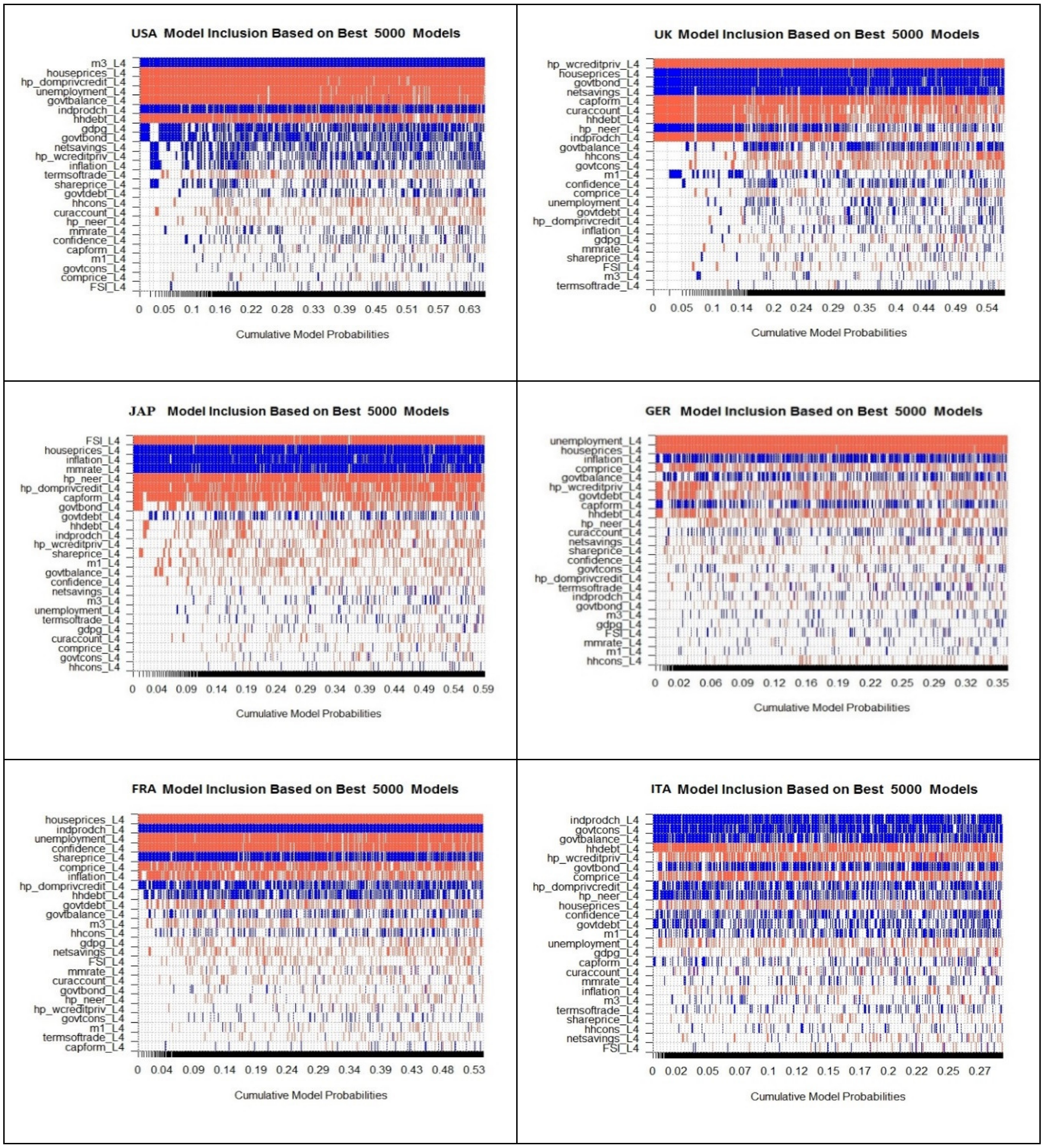
Vermeulen, R., Hoeberichts, M., Vašíček, B., Žigraiová, D., Šmídková, K., de Haan, J., 2015. Financial stress indices and financial crises. Open Economies Review, 26 (3), 383-406.

Appendix

Table A2.1: Variables, transformations and data sources

Variable	*Description*	*Transformation*	*Main source*
Capital formation	Gross total capital formation (constant prices)	% yoy	Statistical offices, OECD
Commodity prices	Commodity prices	% yoy	Commodity Research Bureau
Confidence	Consumer confidence indicator	none	OECD
Current account	Current account (% of GDP)	none	OECD, WDI
GDP growth	Real GDP growth	% yoy	Statistical offices
Govt. balance	Government balance (% of GDP)	none	Statistical offices
Govt. bond yield	10Y government bond yield	none	National central banks
Govt. consumption	Government consumption (constant prices)	% yoy	OECD, statistical offices
Govt. debt	Government debt (% of GDP)	none	WDI, ECB
Household cons.	Private final consumption expenditure (constant prices)	% yoy	Statistical offices
Household debt	Gross liabilities of personal sector growth	% yoy	National central banks, Oxford Economics
House prices	House price inflation	% yoy	BIS, Eurostat, Global Property Guide
Domestic credit gap	Domestic credit to private sector to GDP gap	HP gap	BIS, WDI
World credit gap	Domestic credit to private sector to GDP gap	HP gap	BIS, WDI
Exchange rate	Nominal effective exchange rate gap	HP gap	IFS
Production	Industrial production growth	% yoy	Statistical offices
Inflation	Consumer price inflation	% yoy	Statistical offices, national central banks
M1 growth	M1 growth	% yoy	National central banks
M3 growth	M3 growth	% yoy	National central banks
Net savings	Net national savings (% of GNI)	none	WDI
Stock market	Stock market index growth	% yoy	Reuters, stock exchanges
Money market rate	Money market interest rate	none	IFS
Terms of trade	Terms of trade change	% yoy	Statistical offices
Unemployment	Unemployment rate	none	Statistical offices

Figure A2.1: Bayesian model averaging: early warning indicators of FSI for G7 countries, 4Q ahead



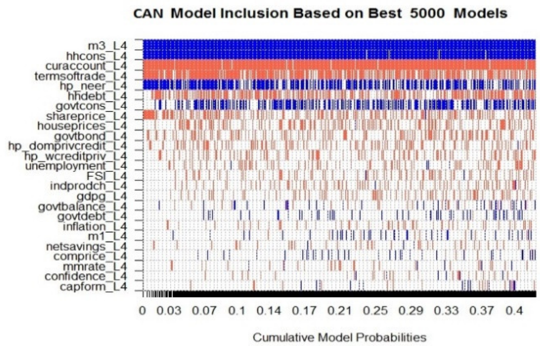


Table A2.2: Variable availability

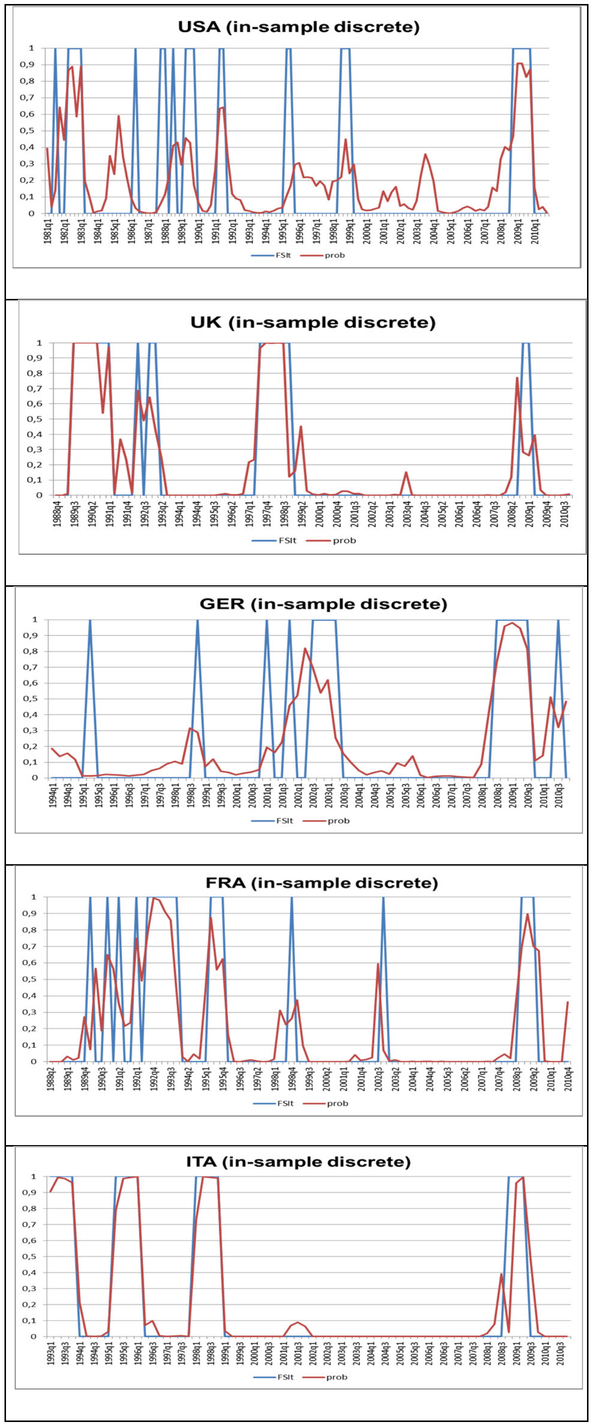
Country	*FSI availability since*	*Availability all predicting variables since*
Australia	1980Q1	1986Q1
Austria	1987Q2	1997Q1
Belgium	1980Q1	1999Q1
Canada	1980Q1	1986Q1
Czech Rep.	2000Q2	2000Q2
Denmark	1980Q1	1996Q1
Finland	1989Q1	1996Q4
France	1987Q2	1987Q2
Germany	1980Q1	1993Q1
Greece	1998Q1	2001Q1
Hungary	1999Q1	2002Q4
Ireland	1980Q1	1998Q1
Italy	1980Q1	1992Q1
Japan	1980Q2	1986Q1
Korea	1990Q1	1998Q4
Netherlands	1980Q1	1994Q1
New Zealand	1989Q1	1990Q1
Norway	1991Q1	1992Q3
Poland	2001Q1	2001Q2
Portugal	1991Q1	1999Q1
Spain	1988Q1	1994Q1
Sweden	1983Q1	1999Q1
Switzerland	1980Q1	1990Q1
UK	1980Q1	1987Q4
US	1980Q1	1986Q1

Table A2.3: Comparison of results of BMA preselected early warning indicators of FSI (PIP ≥ 0.5) 4 Q ahead using linear and nonlinear model – the US

	USA linear	*USA non-linear*
Constant	3.31***	-11.91***
	(1.08)	(3.97)
M3 growth	0.34***	0.75***
	(0.08)	(0.22)
M3 growth^2		-0.04**
		(0.02)
House prices	-0.16***	-0.21***
	(0.04)	(0.05)
House prices^2		0.01***
		(0.00)
Domestic credit gap	-0.05	-0.07
	(0.05)	(0.05)
Domestic credit gap^2		0.01
		(0.01)
Unemployment	-1.29***	-0.91
	(0.20)	(0.85)
Unemployment^2		0.02
		(0.06)
Govt. balance	-0.42***	-0.14
	(0.10)	(0.13)
Govt. balance^2		0.12
		(0.03)
Production	0.14	0.24***
	(0.09)	(0.09)
Production^2		0.00
		(0.01)
Household debt	-0.18**	0.55
	(0.08)	(0.36)
Household debt ^2		-0.03*
		(0.02)
GDP growth	-0.13	0.10
	0.15	(0.27)
GDP growth ^2		-0.09
		(0.04)
Govt. bond yield	0.57***	2.26***
	(0.10)	(0.44)
Govt. bond yield^2		-0.09***
		(0.03)
R^2	0.57	0.71
Adjusted R^2		
Obs.	120	120
Count.	1	1

Note: This table shows results from OLS-regression. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained in Table A2.1 in the Appendix.

Figure A2.2: In-sample fit of country level models - logit model



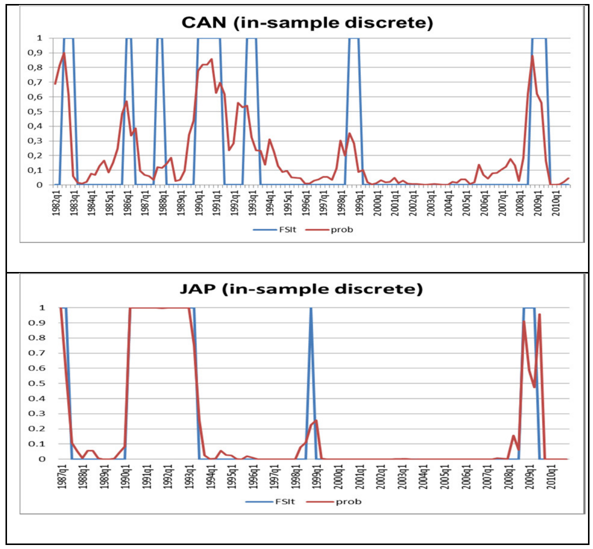


Table A2.4: Comparison of results of BMA preselected early warning indicators of FSI (PIP ≥ 0.5) 4 Q ahead using linear and nonlinear model – the UK

	UK linear	*UK non-linear*
Constant	-11.79***	-3.31
	(1.45)	(5.14)
House prices	0.19***	0.22***
	(0.04)	(0.07)
House prices^2		-0.01***
		(0.00)
Govt. balance	0.09	-0.15
	(0.06)	(0.19)
Govt. balance^2		-0.03
		(0.03)
Production	-0.26***	-0.24***
	(0.06)	(0.07)
Production^2		0.01
		(0.01)
Household debt	-0.30***	-0.24
	(0.08)	(0.31)
Household debt^2		0.01
		(0.01)
Govt. bond yield	1.00***	0.73
	(0.10)	(0.82)
Govt. bond yield^2		-0.00
		(0.05)
World credit gap	-0.29***	-0.23*
	(0.07)	(0.12)
World credit gap^2		0.04**
		(0.01)
Net savings	0.83***	0.22
	(0.17)	(0.36)
Net savings^2		-0.02
		(0.09)

Capital formation	-0.13***	-0.08
	(0.03)	(0.06)
Capital formation^2		-0.00
		(0.00)
Current account	-0.98***	1.19
	(0.23)	(0.94)
Current account^2		0.34**
		(0.13)
Exchange rate	0.10**	0.13***
	(0.04)	(0.04)
Exchange rate^2		0.01
		(0.01)
R^2	0.73	0.79
Adjusted R^2	0.69	0.73
Obs.	89	89
Count.	1	1

Note: This table shows results from OLS regressions. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained Table A2.1 in the Appendix.

Table A2.5: Robustness: different Financial Stress Indexes for the United States

	FSI	*CFSI*	*KCFSI*
Constant	1.585**	0.333	1.234
	(0.50)	(1.23)	(0.99)
M3 growth	0.158***	0.203**	0.153***
	(0.04)	(0.07)	(0.04)
House prices	-0.073***	-0.064*	-0.102**
	(0.02)	(0.02)	(0.03)
Domestic credit gap	-0.022	0.039	0.042
	(0.02)	(0.03)	(0.02)
Unemployment	-0.600***	-0.288	-0.432***
	(0.09)	(0.15)	(0.12)
Govt. balance	-0.195***	-0.181*	-0.145**
	(0.05)	(0.08)	(0.05)
Production	0.065	0.03	0.026
	(0.04)	(0.07)	(0.04)
Household debt	-0.082*	-0.022	-0.018
	(0.04)	(0.08)	(0.06)
GDP growth	-0.063	0.156	0.139*
	(0.07)	(0.10)	(0.06)
Govt. bond yield	0.267***	0.006	0.06
	(0.05)	(0.10)	(0.06)
R2	0.567	0.548	0.65
N	120	78	84

Note: FSI is the financial stress index used in this paper, while CFSI and KCFSI are the financial stress indexes constructed by the Federal Reserve Bank of Cleveland and the Federal Reserve Bank of Kansas City, respectively[12]. Both the CFSI and KCFSI are constructed at the quarterly frequency by taking their mean value during a specific quarter. In order to ease the comparison of coefficients all FSI series are standardize with mean zero and standard deviation 1. *** indicates significance at 1%, ** at 5% and * at 10% level.

Table A2.6: Results of BMA preselected early warning indicators for extreme values of FSI (PIP ≥ 0.5) 4Q ahead for individual G7 countries – logit model

	USA	*UK*	*JAP*	*GER*	*FRA*	*ITA*	*CAN*
M3 growth	0.52***						0.36**
	(0.18)						(0.17)
House prices	-0.28	1.41***	0.93**	-1.04**	-0.58**		
	(0.74)	(0.50)	(2.22)	(0.41)	(0.28)		
Domestic credit gap	-0.21***		-0.41*		0.77***	0.05	
	(0.08)		(-1.77)		(0.27)	(0.17)	
Unemployment	-1.77***			-1.74***	-4.47***		
	(0.59)			(0.53)	(1.26)		
Govt. balance	-0.42*	1.08***				2.91***	
	(0.23)	(0.40)				(0.50)	
Production	0.57***	-0.26***			1.72***	0.98**	
	(0.19)	(0.06)			(0.48)	(0.44)	
Household debt	-0.30	-1.19**			0.06	-1.46***	-0.05
	(0.29)	(0.54)			(0.33)	(0.39)	(0.17)
GDP growth	-0.98***						
	(0.33)						
Govt. bond yield	0.53**	4.40***				2.04***	
	(0.25)	(1.28)				(0.56)	
World credit gap		-1.30***				-0.67	
		(0.43)				(0.48)	
Net savings		2.77***					
		(0.06)					
Capital formation		-0.46*	-0.88**				
		(0.26)	(-2.4)				
Current account		-5.10***					-0.51**
		(2.23)					(0.24)
Exchange rate		0.39**	-0.24***				0.34***

[12] The data are available at: https://research.stlouisfed.org/fred2/series/CFSI ; https://research.stlouisfed.org/fred2/series/KCFSI.

		(0.19)	(-2.78)				(0.12)
Lag FSI			-1.66*				
			(-1.84)				
Money market rate			2.71**				
			(2.36)				
Inflation			3.13	1.74**	-2.76**		
			(1.36)	(0.72)	(1.23)		
Commodity prices				-0.01	0.00	-0.03	
				(0.03)	(0.71)	(0.09)	
Confidence					-4.70***	-1.75*	
					(1.20)	(1.04)	
Stock market					0.07**		
					(0.03)		
Govt. consumption						1.92***	-0.11
						(0.70)	(0.16)
Terms of trade							-0.08
							(0.15)
Household cons.							0.22
							(0.27)
Pseudo R^2	0.33	0.72	0.82	0.40	0.55	0.82	0.36
Obs.	120	89	96	68	91	72	116

Note: This table shows results from Logit-regressions. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained in Table A2.1 in the Appendix.

Table A2.7: Results of BMA preselected early warning indicators for increases of FSI (PIP ≥ 0.5) 4Q ahead for G7 countries – tobit model

	USA	**UK**	**JAP**	**GER**	**FRA**	**ITA**	**CAN**
M3 growth	0.471***						0.528***
	(0.12)						(0.13)
Lag FSI			-1.56***				
			(-6.64)				
House prices	-0.214***	0.242***	0.41***	-0.803**	-0.254**		
	(0.05)	(0.07)	(4.38)	(0.24)	(0.08)		
Domestic credit gap	-0.353***		-0.16***		-0.037	0.218***	
	(0.08)		(-3.75)		(0.13)	(0.06)	
Unemployment	-1.387***			-0.557	-2.380***		
	(0.33)			(0.37)	(0.55)		
Govt. balance	-0.180	0.329*				0.079	
	(0.19)	(0.16)				(0.16)	
Production	0.479***	0.001			0.569***	0.486***	
	(0.13)	(0.15)			(0.14)	(0.11)	
Household debt	0.282	-0.325*			0.256	-0.255	-0.344**
	(0.16)	(0.13)			(0.16)	(0.13)	(0.13)
GDP growth	-0.519*						
	(0.24)						
Govt. bond yield	-0.092	0.702***				-0.175	
	(0.17)	(0.16)				(0.20)	

World credit gap		-0.596***				-0.362***	
		(0.11)				(0.10)	
Net savings		-0.063					
		(0.34)					
Capital formation		-0.028	-0.22***				
		(0.06)	(-4.45)				
Current account		-0.565					0.179
		(0.38)					(0.13)
Exchange rate		0.100	-0.05**				0.273***
		(0.07)	(-2.00)				(0.07)
Inflation			0.91***	1.453*	-1.425**		
			(4.33)	(0.56)	(0.49)		
Commodity prices				0.056	0.009	-0.120***	
				(0.03)	(0.03)	(0.03)	
Confidence					-2.318***	0.572*	
					(0.59)	(0.26)	
Stock market					0.075***		
					(0.01)		
Govt. consumption						0.644**	0.197
						(0.19)	(0.16)
Household cons.							1.079***
							(0.24)
Terms of trade							-0.366***
							(0.10)
Money market rate			0.22				
			(1.26)				
pseudo R^2	0.21	0.17	0.38	0.09	0.15	0.22	0.15
N	120	89	92	68	91	72	116

Note: This table shows results from Tobit-regressions for annual changes in FSI, where changes in FSI<0 are censored. *** indicates significance at 1%, ** at 5% and * at 10% level. Variables in column 1 are explained in Table A2.1 in the Appendix.

Chapter 3

Systemic Event Prediction by an Aggregate Early Warning System: An Application to the Czech Republic

Abstract

This work develops an early warning framework for assessing systemic risks and for predicting systemic events over the short horizon of six quarters and the long horizon of twelve quarters on the panel of 14 countries, both advanced and developing. First, we build Financial Stress Index to identify starting dates of systemic financial crises for each country in the panel. Second, early warning indicators for assessment and prediction of systemic risk are selected in a two-step approach; we find relevant prediction horizons for each indicator by a univariate logit model followed by the application of Bayesian model averaging to identify the most useful indicators. Finally, we observe performance of the constructed EWS over both horizons on the Czech data and find that the model over the long horizon outperforms the EWS over the short horizon. For both horizons, out-of-sample probability estimates do not deviate substantially from their in-sample estimates indicating a good out-of-sample performance for the Czech Republic.

This paper was co-authored with Petr Jakubík and published in Economic Systems (2015, 39, pp. 553-576). The views expressed in this paper are those of the authors and do not necessarily reflect those of the institutions the authors are affiliated with. Financial support from Grant Agency of the Czech Republic (GACR 14-02108S and GACR 403/10/1235) is gratefully acknowledged.

3.1 Introduction

In the wake of the recent global crisis, research in the area of financial stability including Early Warning Systems (EWS) has attracted renewed attention. EWS can be characterized as functional, data-driven approaches that draw attention to variables associated with past crises with the main objective of alerting policy-makers of the potential for future crises (Gramlich et al., 2010). Essentially EWSs are based on the existence of causality between crises and crisis-driving factors and the possibility of crisis-driving factors identification ex ante. In the financial context, EWSs can be used for risk prediction of both a single financial institution risk from microeconomic point of view as well as the risk of an entire financial system, i.e. macroeconomic risk. Of the aforementioned risks we focus on the latter in this study.

In general, systemic risk can be defined as “the possibility that an event will trigger a negative feedback loop that significantly affects financial markets’ ability to allocate capital and serve intermediary functions, which, in turn, will create spillover effects on the real economy that have no clear self-healing mechanism” (Hendricks, Kambhu, and Mosser, 2007). As such the functioning of the financial system is impaired to the extent that economic growth and welfare suffer materially (Lo Duca and Peltonen, 2013).

This paper contributes to the early warning literature in two ways. First, we explore a combination of advanced estimation techniques in developing our multivariate EWS framework over the two horizons of differing length on the panel of 14 countries. Our EWS allows for different relevant prediction horizons for potential leading indicators, determined by univariate logit models, and it employs a relatively novel systematic approach to selecting the most useful crisis leading indicators; Bayesian model averaging. Second, the paper extends the scope of the early warning literature by investigating the performance of our aggregate EWS framework for a single country, the Czech Republic.

The paper is organized as follows. Section 3.2 explains methodology, develops Financial Stress Index and identifies systemic event episodes from calculated FSI. Section 3.3 deals with the identification of leading indicators for systemic events detection. Section 3.4 evaluates performance of our systemic events probability framework on the panel of countries over the short and the long horizon. Section 3.5 applies and evaluates performance of the developed EWS to the Czech Republic. Section

3.6 checks validity and investigates the performance of an alternative regional panel EWS for the Czech Republic. The last section concludes.

3.2 Financial Stress Indicator

Despite the fact that root causes of financial crises throughout history are often diverse along with their propagating channels and market segments, which are consequently affected, it is still interesting to compare these events in terms of systemic stress levels reached. For this reason, a general objective of constructing a financial stress index (FSI) is to measure, in an analytical way, the level of instability (frictions, stresses) within a financial system and to present the findings in a single statistic.

Formerly, the literature on financial crises has substantially depended on historical narratives of crisis episodes, that is mostly for banking crises connected with bank capital erosion and disruption of lending; cases which typically demanded public intervention (Caprio and Klingebiel, 2006). Other such documented episodes further banking crisis episodes with those of currency crises that exhibit reserves depletion and/or major changes in exchange rate mechanism (e.g. Kaminsky and Reinhart, 1999). Despite the fact that these historical crises narratives provide a wide database of crisis episodes, there has been an outbreak of a more analytically based research that aspires to quantify financial stress within the economy by means of a single comprehensive statistic, FSI. The underlying reason for this branch of research is the existence of several drawbacks linked to the above-mentioned historical approaches to crises identification. First, these crisis episodes are known ex post to have large output effects and often require large public intervention while high stress episodes of little macroeconomic impact are often disregarded. Second, episodes identified by historical approaches usually spread over considerable time periods and thus incorporate stresses of varying magnitudes, making it challenging to identify stress peak dates. Third, as databases tend to focus on banking and currency crises, security market stress or liquidity squeezes are easily overlooked, e.g. Long-Term Capital Management collapse of 1998. To avoid these problems, extreme values of a composite indicator, FSI, are used for financial crises identification.

Furthermore, extreme values of FSI allow researchers to exactly identify start and end dates of high systemic stress episodes since crises databases, such as the widely used database by Laeven and Valencia (2008 and 2012), typically provide only a year (or years) when financial crises took place without specifying an exact month/quarter of their onset and end. Another reason for using elevated values of FSI for crises dating is the need of policymakers to be able to monitor systemic stress in the economy with a certain regularity, that is, some policy institutions, such as ECB and their Composite Indicator of Systemic Stress (CISS), construct FSI on a monthly basis. Elevated values of an FSI would thus signal accumulation of stress and widespread imbalances in the economy with potential real consequences - financial crises. Discretization of a continuous FSI measure thus allows for prompt policy actions without having to wait for a true binary crisis indicator to be available via crises databases.

To ensure robustness of our Early Warning System, we develop a measure of financial stress within the economy, FSI, for the panel of 14 countries. The panel consists of EU and OECD member countries. In addition, we include Argentina, Russia and Thailand, countries that either underwent several episodes of financial turmoil, have linkages to other countries in the CEE region or exhibited elevated financial stress during the Asian crisis of 1997, respectively. To verify the suitability of a global panel in our analysis, we also select a subset of three economies from the same region to create a more homogenous panel and derive regional EWS as well as evaluate its performance in Section 3.5.

Following Lo Duca and Peltonen (2013) we construct the FSI on a quarterly basis using the suggested components in their paper. Their FSI incorporates indicators from main segments of domestic financial market since the impact of a negative shock on the economy is typically observable in several of its segments. This accounts for fundamental characteristics of systemic financial stress widely documented in the literature (e.g. Hakkio and Keeton, 2009; Fostel and Geneakoplos, 2008). FSI by Lo Duca and Peltonen (2013) also uses a minimum set of indicators to avoid data availability issues across time and countries. Despite this simplification, including only indicators for vital parts of the economy in the FSI still captures financial stress levels well. Moreover, adding more components does not significantly change the shape of the resulting FSI (Hollo, Kremer and Lo Duca, 2012). Conversely, inclusion of too many indicators "could potentially contaminate the FSI with noisy indicators" (Cardarelli, Elekdag and Lall, 2011).

For the above mentioned reasons, the FSI is calculated by aggregating the following 5 components:

1. *Negative quarterly returns of the main equity index*, calculated from equity returns which were multiplied by -1 so that negative returns increase financial stress while positive returns are set to 0.
2. *Realised volatility of the main equity index*, calculated by determining standard deviation of the main equity index values over the last 12 months leading to each observation date.

3. *Realised volatility of the nominal effective exchange rate*, resulting from computing standard deviation of nominal effective exchange rate values over the last 12 months leading to each observation date.
4. *TED spread*, measured as the difference between 3-month interbank rate and 3-month Treasury bill rate. This component represents the credit risk associated with interbank lending. The higher the TED spread the more the default risk on interbank loans is perceived.
5. *Realised volatility of the yield on 3-month Treasury bills*, calculated as standard deviation of 3-month Treasury bill yields over the last 12 months preceding each observation date.

For two countries in our panel (Thailand, Turkey) an alternative set of indicators is developed due to data unavailability. The alternative FSI is also computed for every other country in the sample as a robustness check to verify that FSI captures high stress periods appropriately. These indicators are aggregated into an alternative FSI (see Figure A3.1 in Appendix) which differs from the originally constructed FSI in 2 components. Namely the last 2 components (4 and 5) are substituted by the following indicators:

i. *Inverted interest rate spread*, calculated as the difference between interest rate paid by banks on demand, savings or time deposits minus interest rate charged by banks on loans. In general, the measure is used as a proxy for profitability in a banking sector.
ii. *Realised volatility of the yield on long-term government bonds*, calculated as a standard deviation of long-term government bond yields over the last 12 months preceding each observation date.

Such composition of the FSI accounts for the four fundamental characteristics of the financial stress widely documented in the literature (e.g. Hakkio and Keeton, 2009; Fostel and Geneakoplos 2008):

- *Significant shifts in asset prices* (captured here through main equity index returns)
- *A sudden increase in risk or uncertainty* (captured here through realised volatility of the main equity index, T-bill rate realised volatility, alternatively through realised volatility of yield on government bonds and realised volatility of nominal effective exchange rate)
- *Abrupt changes in liquidity* (expressed here by TED spread)
- *State of the banking system* (its health is approximated here by interest rate spread as a proxy for profitability)

Each of the 5 components of the FSI is transformed before aggregation in order to facilitate measuring and cross-country comparison of financial stress levels by each FSI. Therefore, every observation at every point in time (year's quarter) for each indicator is assigned the value equal to the percentile it represents of the country-specific distribution function for this indicator. The values of thus transformed observations for each component range from 0 to 1 included. The individual stress components are designed in such a way that their higher values representing higher percentiles of their distributions signal increased financial stress levels.

The transformed variables are then aggregated into the FSI according to the following formula:

$$FSI_{i,t} = \sum_{j=1}^{5} w_j \cdot Ind_{j,i,t} \tag{3.1}$$

where j represents each indicator of FSI, i indicates a country within the panel and t stands for the quarter an observation falls into. The FSI is thus a continuous measure at quarterly frequency that is calculated as a weighted average of the 5 transformed indicators for each country i at each quarter t. Market-equal weighting is chosen for indicator aggregation, i.e. placing a weight of 25% on each market represented within FSI. Moreover, as the method of combining individual stress measures into FSI is not clearly defined and is still subject to research efforts (e.g. Illing and Liu, 2006; Oet et al., 2011; Louzis and Vouldis, 2011), we opt for the more neutral market-equal weighting approach to combining individual indicators into FSI. Hence, the market-equal weighting scheme accounts for the cross-country nature of our aggregate model. The distribution of weights among individual indicators is thus as follows:

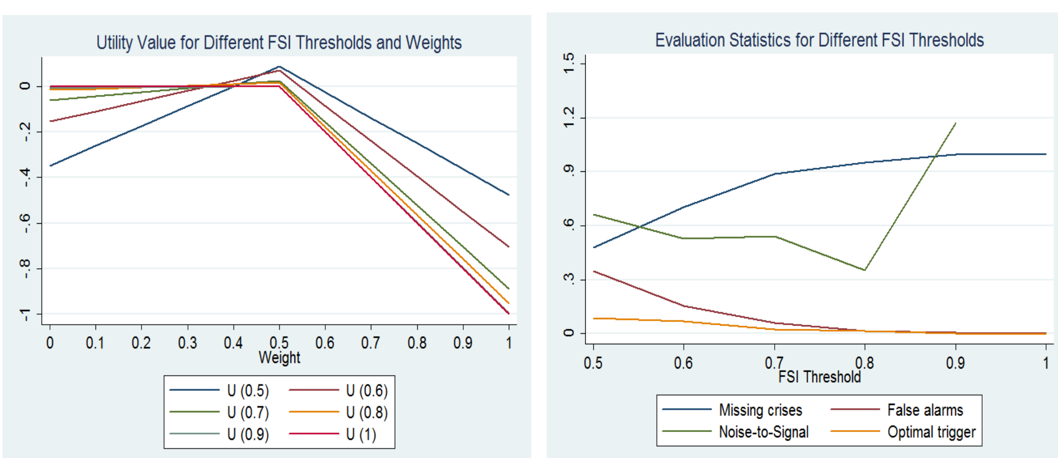
- 12.5% for negative returns of the main equity index
- 12.5% for the realised volatility of the main equity index
- 25% for the realised volatility of the nominal effective exchange rate
- 25% for TED spread or inverted interest rate spread
- 25% for the realised volatility of 3-month Treasury bill rate or long-term government bond yield volatility

Systemic events identification from financial stress measure, FSI, is crucial to EWS framework as it indicates crisis occurrence/absence that is used as a dependent within the EWS. Due to the fact that FSI was calculated as the simple average of financial stress captured in different markets by selected indicators and expressed as the percentile value of these indicators' country distributions, it represents average attained levels of stress in the economy as a whole in each time period. Hence, it is possible to set a certain value of FSI as a threshold, which once exceeded, would signal the occurrence of a systemic event. In this spirit, the threshold of 0.7 was chosen for systemic event occurrence, which in turn identifies 30% of highest stress periods for each country as crises. This calibration allows for correct identification of most systemic events in this study as compared with the crises database compiled by Laeven and Valencia (2008) (see Figure 3.2A in Appendix). To test sensitivity, we set the threshold for crises detection to different values of FSI, from 0.5 to 1, and evaluate the rate of success in detecting systemic events for individual thresholds. Ratio of missed crises, false alarms and Noise-to-Signal ratio are used in this robustness check.

Panel A in Figure 3.1 shows that preference weight of 0.5 in probability utility function (U)[1] maximizes the utility level over all FSI thresholds for crises identification. Panel B indicates that type I error of missing crises increases with higher FSI thresholds while type II error of false alarms decreases. Optimal trigger, i.e. maximized utility for each FSI threshold, falls with higher thresholds. Noise-to-Signal ratio is minimized for the threshold of 0.8.

Figure 3.1: Evaluation statistics of alternative FSI thresholds

A) B)



[1] $U = Min[\mu, 1-\mu] - \left(\mu * \left(\frac{C}{A+C}\right) + (1-\mu) * \left(\frac{B}{B+D}\right)\right)$, where $\left(\frac{C}{A+C}\right)$ is the proportion of missing signals, $\left(\frac{B}{B+D}\right)$ the proportion of false signals, μ expresses policy-maker's preference towards either error type.

Overall, there is a trade-off between errors of type I and II for increasing thresholds while maximized utility of FSI falls. Noise-to-Signal is the lowest for the threshold of 0.8. However, setting FSI threshold to 0.8 fails to detect any crisis quarter for Argentina, Czech Republic, Hungary, Russia, Sweden, Thailand and Turkey and thus reduces our pool of observable crises periods. For this reason, we maintain the selected FSI threshold of 0.7 for our analysis.

In order to ensure that FSI behaves as an appropriate early warning indicator by signalling upcoming systemic events, it needs to equal 1 in periods leading to the outbreak of these events. The horizon for signalling upcoming systemic events of two different lengths, short and long, is of interest in our work. For this purpose, the two models (short and long) are built to account for the appropriate signalling of upcoming crises over each of these horizons.

Therefore, in the short model FSI is set to 1 in 6 quarters leading to an event as this time length should be sufficient for policy makers to prepare adequate policy response (Lo Duca and Peltonen, 2013). Furthermore, in line with Lo Duca, Peltonen (2013) and Bussiere, Fratzscher (2006) the so-called periods of economic recovery, i.e. transitions from systemic events to tranquil periods, are excluded from the sample since during these periods "economic variables go through an adjustment process before reaching again the path they have during tranquil periods", that could consequently lead to a „post crisis bias" (Bussiere and Fratzscher, 2006). In practice this means that FSI is set to 0 after a crisis outbreak, i.e. in periods during which it originally remained above the set threshold. Additionally, FSI is assigned 0 in all tranquil periods shorter than 6 quarters as any subsequent high stress periods could still be continuations of previous systemic events (Lo Duca and Peltonen, 2013). As for the long model with the horizon of 12 quarters, the binary FSI transformation applies the same reasoning that was implemented for the transformation over the short horizon. In this case, FSI is set to 1 in 12 quarters preceding a systemic event outbreak and to 0 in all other periods[2].

3.3 Leading Indicators for Systemic Events Detection

In regard to constructing a framework for assessment and probabilistic prediction of systemic events, it is essential to include among potential leading indicators variables

[2] Tables A3.1 – A3.4 in the Appendix present information on the number of event and non-event (i.e. 1 and 0) instances contained in the FSI for both the short and the long model and for models estimated on the regional subpanel of countries. In addition, Tables A3.2 and A3.4 present FSI value statistics contained in the effective sample, i.e. after missing observations were excluded from the dataset.

with the capacity to capture presence of imbalances within both the domestic and the global economy that may lead to an outbreak of a systemic event. The different theoretical settings behind the explanation of individual crises episodes generate alternative sets of potential explanatory variables for the probability of a crisis occurring (Crespo-Cuaresma and Slacik, 2009). For this reason, the initial set of variables in this study is based on indicators that tend to appear in existing early warning system mechanisms, such as Lo Duca and Peltonen (2013), Babecky et al. (2013, 2014), Jakubik and Slacik (2013). Our set contains not only domestic and global variables but also interactions between selected domestic variables, between global variables and between domestic and global variables.

In this spirit, for each country in the panel, the growth in domestic asset prices is approximated by real annual growth of the local MSCI index while asset price valuations are expressed by the ratio of equity market capitalization to GDP. As for leverage, it is measured by the ratio of private credit to GDP while growth in a country's bank credit is approximated by private credit annual growth. Moreover, interaction between domestic asset price growth and asset price valuations as well as interaction between domestic credit growth and leverage level is expressed as the product of the two variables that should capture the dynamics. The same set of variables and their interactions as for domestic economy is prepared also for the global economy. In an attempt to capture additional fragilities that emerge when the overheating of the domestic economy coincides with the vulnerabilities in the global conditions, interactions between domestic and global variables were included as products of relevant variables. Despite the fact that this approach to quantifying variable interactions might be too general, it helps us avoid setting the interactions arbitrarily for individual countries. Global variables were approximated by GDP-weighted averages of four large economies within the sample, i.e. Euro area, Japan, United Kingdom and United States (Lo Duca and Peltonen, 2013).

Apart from these variables the set of potential leading indicators includes proxies for macroeconomic conditions on the domestic level as well as on the global level. The short and long trends are derived from Hodrick-Prescott filter with values of the smoothing parameter of 1600 and 400 000, respectively. All indicators are in quarterly frequency. However, the variables from Table 3.1 that are indicated as obtained from the World Bank (WB) were initially in annual frequency, thus to ensure their quarterly frequency for the purposes of this analysis a decomposition by cubic-match method is applied (Babecky et al., 2013). Real variables within the dataset are calculated by deflating a nominal variable by the consumer price index (CPI). Ultimately, the set of amassed variables covers the period between 1990Q1 and 2013Q1 for 14 countries altogether.

Table 3.1: Set of potential leading indicators

Indicator	*Description*	*Source*	*Indicator*	*Description*	*Source*
Real GDP	year-on-year change	OECD, NCB	Real private credit annual growth	interaction between global and domestic variables	BIS
Real M2	year-on-year change	IMF, NCB	Private credit/GDP	interaction between global and domestic variables	BIS
Real money	year-on-year change	IMF, NCB	Real MSCI annual growth x Global market capitalization/GDP	interaction between global and domestic variables	WB, www.msci.com
M2	share of GDP	IMF, NCB	Private credit growth x Global Private credit/GDP	interaction between global and domestic variables	BIS
Money	share of GDP	IMF, NCB	CPI	year-on-year change	IMF, OECD, NCB
Real domestic credit	year-on-year change	IMF	Real effective exchange rate	period-on-period change	BIS
Government deficit	share of GDP	IMF, NCB, Reuters	Global real private credit	year-on-year change	BIS
Government debt	share of GDP	OECD, NCB, Reuters	Global market capitalization	share of global GDP	WB
Private credit	share of GDP	BIS	Global private credit	share of global GDP	BIS
Real MSCI index	deviation from HP trend (short)	www.msci.com	Global private credit growth x Global private credit/GDP	interaction between global variables	BIS
Reserves	period-on - period change	IMF, OECD	Global real GDP	year-on-year change	OECD, NCB
Trade balance	period-on - period change	IMF, OECD	Global CPI	year-on-year change	IMF, OECD, NCB
Current account/GDP	share of GDP	OECD, NCB	Real private credit	year-on-year change	BIS
Unemployment rate	share of labour force	IMF, NCB	Market capitalization	share of GDP	WB
Gross fixed capital formation	period-on-period change	IMF	Private credit growth x Private credit/GDP	interaction between domestic variables	author based on BIS

Industrial production	period-on-period change	IMF, OECD, NCB	**Market capitalization/GDP**	interaction between global and domestic variable	author based on WB
Non-performing loans	share of total loans	WB			

Notes: NCB stands for national central banks and HP stands for Hodrick-Prescott filter.

The set of potential leading indicators is prepared in a way to ensure their stationarity, i.e. expressing indicators by mostly growth rates. For nonstationary variables, their stationarity is ensured by first differencing.

Furthermore, as leading indicators are explanatory variables and FSI a dependent in latter analysis, avoidance of potential correlations between indicators of systemic events and the variables from the FSI composition needs to be kept in mind when building the set of indicators. Therefore, correlations were checked for critical indicators. Those variables for which null of no correlation is rejected, are ultimately excluded from the analysis. This step is crucial for regressions used in identification of optimal lags of indicators later in this section. The resulting list of potential leading indicators can be found in Table 3.1.

3.3.1 Selecting Optimal Lags

Optimal lags selection for the indicators to be included in EWS poses a challenging question as different indicators might be able to discern the probability of a systemic event occurrence with a varying lead time length. In this view, various indicators are capable of issuing either a late warning for a 1-3Q horizon ahead or an early warning for 4-8Q ahead of a systemic event materialization as specified in Babecky et al. (2013). Generally, the indicator lags selection is conditional upon researchers' expert opinion (Kaminsky and Reinhart, 1999) or to allow for publication lags of selected indicators (Lo Duca and Peltonen, 2013). As the literature has not come to a consensus on identifying optimal indicator lags, we deal with this issue by introducing a quantitative approach, inspired by Babecky et al. (2013, 2014).

Babecky et al. (2013, 2014) choose a panel vector autoregression model to account for differing dynamics of indicators in regards to systemic event occurrences. In contrast to that paper, we obtain important lags for each indicator from a univariate logit model with FSI as a dependent (transformed into binary form) and an indicator with lags from 0 to 8 (in quarters) as independent variables. Previous exclusion of indicators correlated with FSI allows for investigation of optimal lags by means of univariate regressions. This setting investigates the dynamics of each indicator and FSI separately with the aim to extract lags that are relevant for explanation of systemic event occurrences as defined by binary FSI. Moreover, logit model is chosen for this purpose to maintain consistency throughout the entire analysis. Lags of each indicator that emerged significant from these univariate logit models are included in building our EWS.

The method for relevant lag selection is performed twice with the same set of initial indicators from Table 3.1, once for FSI in the short form, i.e. flashing 1 in six quarters preceding the identified outbreak of a systemic event, and once for FSI in the long form, i.e. flashing 1 in twelve quarters preceding the identified outbreak of a systemic event. Finally, after the inclusion of the relevant lags the set of potential indicators expands from 33 as presented in Table 3.1 to 78 for the short model, i.e. with FSI in the short form, and to 74 for the long model with FSI in the long form.

3.3.2 Addressing Model Uncertainty

As there is no unique theoretical framework linking the potential set of variables with crisis occurences, the issue of model uncertainty surrounding the choice of early warning indicators needs to be dealt with (Crespo-Cuaresma and Slacik, 2009). An ideal solution would be running regressions with different subsets of selected lagged indicators to ensure robustness of results. However, addressing model uncertainty this way manually would be very time consuming. For this reason, we employ Bayesian model averaging (BMA)[3] that resolves the issue by running many regressions with different subsets of 2^{78} possible combinations of indicators for the short model and 2^{74} combinations for the long model. In economics, several applications of BMA to solve model uncertainty are performed in the area of meta-analysis (e.g. Babecky and Havranek, 2014; Havranek and Rusnak, 2013). In the area of financial stability, Crespo-Cuaresma, Slacik (2009) and Babecky et al. (2013, 2014) make use of BMA to identify determinants of crisis episodes.

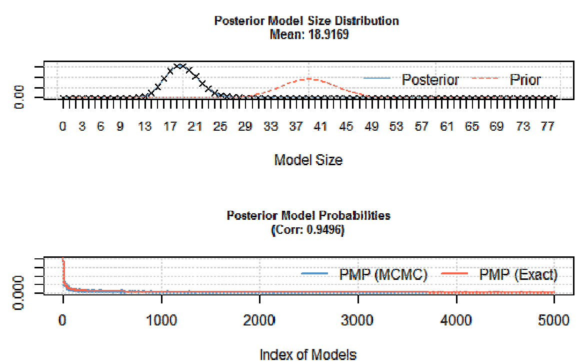
Compared to the studies by Babecky et al. (2013, 2014) the set of potential EWIs we input into BMA is more focused on credit indicators and their interactions. In line with Lo Duca and Peltonen (2013) we also incorporate asset price and credit developments and valuation levels. Furthermore, we also calculate indicator interactions on global, and global and domestic level.

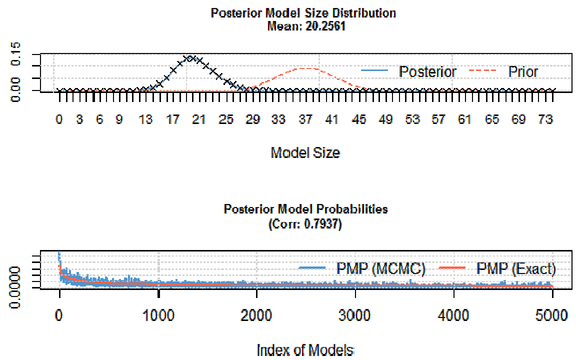
[3] The computation is performed in R using „BMS“ package by Feldkircher and Zeugner (2009).

Due to a large number of potential variables (and their lags) we input into BMA, enumeration of all potential combinations of variables becomes not only time consuming but with increasing number of variables even infeasible (Feldkircher and Zeugner, 2009). Therefore, the standard Markov Chain Monte Carlo (MCMC) birth-death sampler developed by Madigan and York (1995), which is used in majority of BMA routines (Feldkircher and Zeugner, 2015), is applied to obtain results on the most important part of the posterior model distribution and thus deliver as precise estimates as possible. The quality of the MCMC approximation to the actual posterior distribution, i.e. the correlation of MCMC approximation results and the analytical ones, is linked to the number of draws the sampler is set to go through during the estimation process (iterations). However, as the MCMC sampler might start sampling from models that might not yield the best results and only after some time converge to models with high posterior model probabilities, it is advisable to discard these initial iterations (burn-ins).

For sets of potential indicators in both models, the number of iterations is set to 45 000 000 after the initial 2 000 000 are discarded as burn-ins. The correlations obtained between the MCMC and analytical results for the short and the long model are 0.9496 and 0.7937, respectively, which could be considered a sufficient convergence. Figure 3.2 details these results as well as it shows prior and posterior model size distributions for both models.

Figure 3.2: Convergence and model size distributions for the short and the long model





As is discernible from Figure 3.2, uniform model prior was employed in the computations so that expected prior model parameter size equals half the number of potential indicators entered into the Bayesian model averaging. However, after having updated the model prior with data it yields a smaller expected posterior model parameter size since parsimonious models are preferred.

BMA technique ultimately identifies the following 12 indicators as useful in crisis signalling over the short horizon of 6 quarters (short model): real money growth and its 4th lag, the 5th and 8th lags of real MSCI deviation from short Hodrick-Prescott trend, the 2nd lag of unemployment rate, the 7th lag of global private credit annual growth, the 7th lag of interaction between global private credit annual growth and global private credit/GDP, the 4th and 5th lag of interaction between private credit annual growth and global private credit/GDP, the 1st lag of global real GDP annual growth, the 1st lag of real GDP annual growth and global CPI annual growth. As for other selected variables, domestic GDP annual growth, unemployment rate, real money growth and global CPI growth are also considered informative and will be therefore included in the final short model.

For the EWS over the long horizon of 12 quarters the following 13 variables are selected as the most informative: the 1st lag of real domestic credit annual growth, the 4th lag interaction between domestic private credit annual growth and global private credit/GDP, the 5th lag of unemployment rate, the 1st and 8th lag of domestic CPI annual growth, the 4th lag of interaction between domestic and global real private credit annual growth, real domestic GDP annual growth, global market capitalization/GDP and its 3rd lag, the 8th lag of MSCI deviation from short Hodrick-Prescott trend, the 1st lag of global private credit annual growth, the 1st lag of interaction between global private credit annual growth and global private credit/GDP and M2/GDP ratio.

In line with common findings in the literature (e.g. Alessi and Detken, 2011; Lo Duca and Peltonen, 2013), credit and private credit indicators both domestic and global as well as their interactions are found useful for the models over both horizon lengths. Overall 5 credit indicators are to be included into EWS over the long horizon and 4 over the short horizon. Moreover, the ratio of global market capitalization/GDP is selected even twice over the long horizon, which coincides with the finding by Lo Duca, Peltonen (2013) that this is the most useful global indicator, i.e. the most useful indicator overall, in their study. As for asset prices, they are an important indicator in both models, though only their 8th lag appears in the long model.

Same as for the short model indicators, apart from credit indicators domestic GDP, CPI growth and unemployment rate are selected into the long model. As for money aggregates, the ratio of M2/GDP is selected for the long model as opposed to real money growth that appears in the short model.

3.4 Systemic Events Probability Framework

Having selected appropriate indicator lags and indicators themselves, we focus now on estimating the joint impact of useful indicators on the probability of a systemic event. As the dependent variable is FSI in a binary form, a logistic regression is applied to the data to ascertain the relation between useful indicators of vulnerabilities and crisis probability (advocated e.g. by Demirguc-Kunt and Detragiache, 2005). To estimate logit model, the maximum likelihood estimation technique is used which yields coefficient estimates that are consistent and asymptotically efficient as well as asymptotic standard errors of the coefficient estimates (Cramer, 2003)[4].

3.4.1 Short Model Estimation and Performance

The short logit model contains binary FSI, with values of 1 in 6 quarters preceding the pre-defined outbreak of a systemic event and 0 in all other periods, on the left-hand side, i.e. the dependent variable. On the right-hand side there are 12 useful indicators, the outcome of BMA technique. However, this model initially displays high collinearity between 2 of the indicators, the seventh lag of global annual private credit growth and the seventh lag of the interaction between global annual private credit growth and global

[4] All calculations are performed in R using package "verification" that follows the process outlined in Mason and Graham (2002).

private credit over GDP. Therefore, to avoid collinearity among independent variables, the seventh lag of the interaction between global annual private credit growth and global private credit over GDP is omitted from the model. All in all, the final short model comprises 11 indicators. We then fit the model to all available data, then to data until 2011 and to data only until 2006. For each model in-sample predictions are computed. The out-of-sample predictions for each model are calculated over the pre-crisis period of the Global crisis, i.e. 2006Q1-2008Q1, and over the period of 2011Q1-2013Q1, which is the time this study was first conducted.

Due to the nature of logit model, the coefficient estimates for independent variables are log-odds ratios. However, in order to estimate more precisely the extent of the change in crisis likelihood given a change in an independent variable, an exponential of the log-odds ratio indicates actual odds of materialization of an event. For a negative relationship between the explanatory and the dependent variable odds lie between 0 and 1, in case of a positive relationship they exceed 1.

Table 3.2: Short model estimation on all available data

	Coefficient	Std. Error	z	p-value	
Const	-2.26005	0.452736	-4.9920	5.98e-07	***
realmoneygl4	-12.5276	2.26024	-5.5426	2.98e-08	***
MSCIhpshortl5	5.34506	1.51967	3.5172	0.00044	***
Uratel2	22.1122	6.88392	3.2122	0.00132	***
Globpcredgl7	40.2897	10.0369	4.0142	5.97e-05	***
MSCIhpshortl8	1.85216	0.928	1.9959	0.04595	**
realmoneyg	-7.32671	2.27894	-3.2150	0.00130	***
Int_pcgxglopcGDPl4	11.6145	9.21088	1.2610	0.20732	
Int_pcgxglopcGDPl5	23.2149	9.92291	2.3395	0.01931	**
GlobGDPgl1	14.747	9.168	1.6085	0.10772	
realGDPgl1	20.7753	6.63864	3.1294	0.00175	***
GlobCPIg	31.2841	19.104	1.6376	0.10151	

Mean dependent var	0.192453	S.D. dependent var	0.394599
McFadden R-squared	0.452249	Adjusted R-squared	0.406020
Log-likelihood	-142.1812	Akaike criterion	308.3624
Schwarz criterion	359.6369	Hannan-Quinn	328.4320

Likelihood ratio test: Chi-square(11) = 234.783 [0.0000]

Note: The number following each indicator states an indicator‘s lag (in quarters), * represents significance on 10%, ** on 5% and *** on 1% significance level.

Only one variable, real money growth (realmoneyg) and its fourth lag (realmoneygl4), have a negative relationship with the likelihood of an event occurrence. A unit change in all other independent variables increases the odds of a crisis by more than 1.

Table 3.3: In-sample performance of short logit models

In-sample performance of short logit models							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Short truncated till 2006	0.436	0.812	87.56	100	0.133	0.959	1.75E-08
Short truncated till 2011	0.372	0.756	88.70	84.44	0.122	0.937	2.66E-38
Short on full data	0.335	0.805	89.43	73.53	0.092	0.908	6.38E-38

As shown in Table 3.3 the short model performance is measured by several indicators: maximum utility measure (U), threshold for which the model's utility is maximized, percentage correctly predicted (PCP)[5], percentage of crises predicted[6], Noise-to-Signal ratio (NtS ratio)[7], ROC area[8] and p-value[9].

The logit model performance can be assessed in a framework that takes into account missing systemic events, false signal emissions and policy-maker's preferences. This analysis follows the approach by Alessi and Detken (2011), which allows finding the optimal early warning thresholds for indicators and thus rank them with respect to their

[5] The utility-maximizing threshold is used as a cut-off

[6] It is calculated as the number of periods when signal was correctly issued over the number of periods in which the signal should have been issued (sum of "correct signal" periods and "missing signal" periods).

[7] A useful model is supposed to have a NtS of less than 1, a value of 1 would result if a model provides purely random signals (Kaminsky et al., 1998).

[8] A value of 1 indicates a perfect model while a random forecast would have the ROC area equal to 0.5. The ROC area calculation follows Mason and Graham (2002).

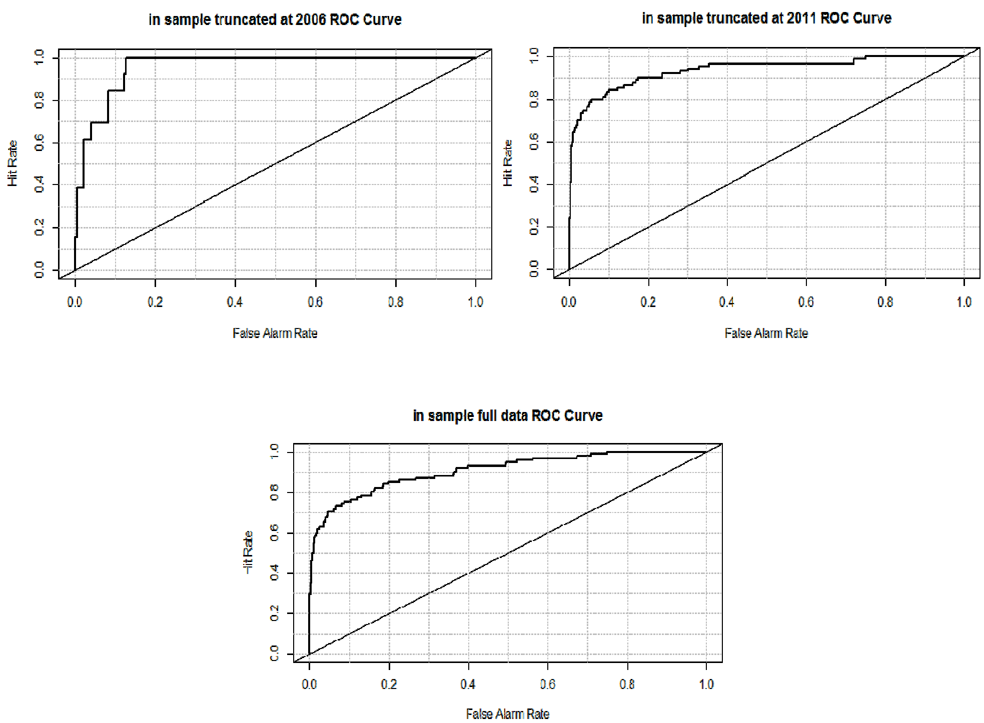
[9] p-value helps estimate the adequacy of a model forecast via ROC area and is related to Mann-Whitney U statistics. The statistics tests the null of the area under the ROC curve equal to 0.5 or the forecast has no skill.

crisis detecting usefulness. The objective of this analysis is to find a threshold for prediction resulting from each model that maximizes the utility function U. In order to find the optimal threshold, all the predictions of FSI are transformed into percentile values of their distribution function. Every such percentile value is then set as a threshold for which the value of utility function is computed. The threshold which maximizes the utility function, apart from minimum and maximum value of the distribution, is consequently chosen as optimal. In our analysis, we apply policymaker's preference neutrality towards missing signals and false alarms in the utility function. Checking for different preferences between the false signals and missing events in our analysis reveals that preference neutrality maximizes the utility value over all different preference specifications in aggregate EWS regressions as well as for regressions on the Czech data. In addition, setting policymakers' preferences to 0.5 is considered a more stringent test of model performance than other lower thresholds (Davis et al., 2011). The results can be provided by authors upon request. The crisis dating needed for these calculations is provided from the crises database by Laeven and Valencia (2008, 2012) while for the Euro area and the global economy the crisis dating includes only one systemic event, i.e. the global current crisis, within the observed period of 1990Q1-2013Q1.

According to these calculations the best-performing in-sample short model is the one estimated on truncated data until 2006. It has the highest U measure, the percentage of crises predicted and area under ROC curve which is also highly significant with p-value of 1.75E-08. On the other hand, the model's NtS ratio is the largest out of the compared in-sample short models while the percentage correctly predicted is the lowest. On the whole, the in-sample performance of the short model appears to be more than satisfactory verifiable by low p-values and signifying strong rejection of the null of no forecast skill for all three fittings of the short model.

Areas under ROC curves presented in Table 3.4 are obtained from ROC curve plots in Figure 3.3. The further the ROC curve for a model is from the diagonal, the larger the discrimination (analogy with Gini coefficient as a measure of inequality), i.e. the higher the forecast's skill to anticipate correctly the occurrence or non-occurrence of pre-defined events.

Figure 3.3: ROC curve plots for in-sample performance of short logit model estimated on data until 2006, until 2011 and on all available data



Once in-sample performance of the short logit model is validated, we verify its out-of-sample performance. This check is performed to estimate a model's forecasting ability. The results of forecasts for the model on truncated data till 2011 over the period of 2011Q1-2013Q1 and those of the model on truncated data until 2006 over pre-crisis period of 2006Q1-2008Q1 are summarized in Table 3.4 below.

Table 3.4: Out-of-sample performance of short logit models

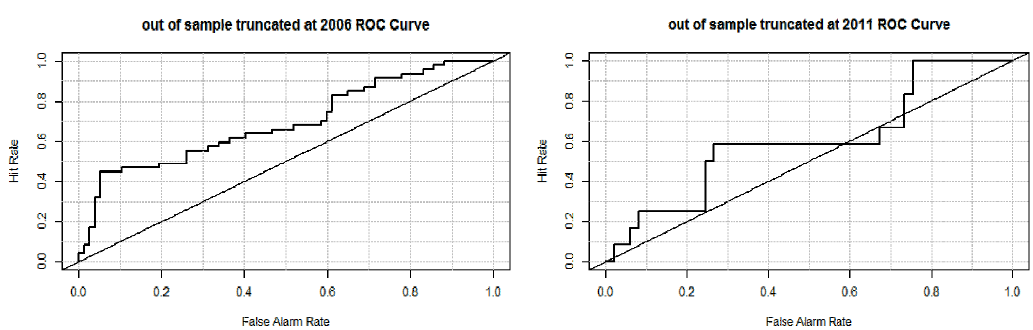
Out-of-sample performance of short logit models							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Short truncated till 2006	0.197	0.796	75	44.68	0.145	0.691	0.00019
Short truncated till 2011	0.159	0.666	68.85	58.33	0.490	0.599	0.150506

As expected the out-of-sample performance of the model is lower compared to its in-sample results. The maximum utility is about half of that for in-sample performance and all other performance measures decrease (apart from NtS ratio which increases) indicating weaker performance in general. The better out-of-sample performance is for the model on truncated data till 2006 due to having higher utility measure, PCP, lower NtS ratio and a larger area under ROC curve, which is significant on 0.02% significance level. However, its percentage of crises predicted is lower than that for the model on data truncated till 2011. The worse out-of-sample performing short model, on data truncated until 2011, does not differ dramatically in terms of performance measures from the better one apart from NtS ratio that is almost 0.5 and the area under ROC curve of 0.599 which is significant only on 16% significance level even if the model itself is not a random forecast (area of 0.599 is still larger than 0.5).

Overall, the short model on truncated data until 2006 is ranked as the best performing by its U measure both in-sample and out-of-sample. However, out-of-sample the model experiences almost 55% fall in its utility, 14.5% decline in its PCP, the fall of 55.3% in its percentage of crises predicted, 9% rise in its NtS ratio while the area under ROC curve shrinks by 28%. In comparison, the out-of-sample performance of the worse model, estimated on data up till 2011, declined from its in-sample performance by 57% for U, 22.4% for PCP, 31% for percentage of crises predicted and by 36% for ROC area while its NtS ratio shot up by 302% to the level of almost 0.5.

In addition, the best ranked model, estimated on truncated data till 2006, appears to be more stable when estimated out-of-sample than the second best ranked short model (see Figure 3.4).

Figure 3.4: ROC curve plots for out-of-sample performance of short logit model estimated on data up till 2006 and till 2011



3.4.2 Long Model Estimation and Performance

In the long model, the dependent is the binary FSI, with values of 1 in 12 quarters preceding the pre-defined occurrence of a systemic event and 0 in all other periods, on the left-hand side while there are 13 indicators, selected from BMA technique, on the right-hand side. Similarly to the short model, the long model also displays high collinearity between 2 of the indicators, i.e. the first lag of global annual private credit growth and the first lag of the interaction between global annual private credit growth and global private credit over GDP. To avoid this issue, the first lag of the interaction between global annual private credit growth and global private credit over GDP is omitted in the spirit of the short model analysis.

After this adjustment the final long model containing 12 indicators is fitted, as in the case of the short model, to all available data, to data truncated till 2011 and truncated till 2005. For each model in-sample predictions are calculated. Similarly to the short model, out-of-sample predictions are computed over the pre-crisis period of the Global crisis, i.e. 2005Q1 - 2008Q2, and over the period at which this study was originally conducted, i.e. 2011Q1 - 2013Q1.

Next, the analysis resumes the structure of that for the short model. As such, indicator coefficients are estimated for each of the sub periods followed by in-sample and out-of-sample performances of long logit models.

Table 3.5: Long model estimation on all available data

	Coefficient	Std. Error	z	p-value	
const	-10.9152	1.23256	-8.8557	8.31e-019	***
rdomcredl1	-1.07904	2.10306	-0.5131	0.60790	
Int_pcgxglopcGDPl4	50.8465	13.926	3.6512	0.00026	***
Uratel5	8.77642	5.82841	1.5058	0.13212	
CPIgl8	4.54716	4.36299	1.0422	0.29731	
Int_realpcredgl4	-309.252	109.757	-2.8176	0.00484	***
CPIgl1	-54.7731	8.75562	-6.2558	3.96e-010	***
realGDPg	-14.7407	5.81822	-2.5335	0.01129	**
GlobmcapGDP	5.26878	1.47088	3.5821	0.00034	***
MSCIhpshortl8	1.59052	0.949851	1.6745	0.09403	*
Globpcredgl1	2.48314	10.0189	0.2478	0.80425	
GlobmcapGDPl3	7.26515	1.38609	5.2415	1.59e-07	***
M2GDP	2.06582	0.844418	2.4464	0.01443	**

Mean dependent var	0.275362	S.D. dependent var	0.447102
McFadden R-squared	0.459095	Adjusted R-squared	0.419078
Log-likelihood	-175.7201	Akaike criterion	377.4402
Schwarz criterion	433.5163	Hannan-Quinn	399.3503

Likelihood ratio test: Chi-square(12) = 298.286 [0.0000]

Note: the number following each indicator states an indicator‘s lag (in quarters), * represents significance on 10%, ** on 5% and *** on 1% significance level.

From long model estimation on the full data sample, it is observable that 4 independent variables, the 1st lag of real domestic credit growth, the 4th lag of interaction between domestic and global real private credit growth, the 1st lag of CPI annual growth and annual growth of real GDP, have a negative relationship with the dependent, a likelihood of a systemic event occurrence. A unit change in all other explanatory variables increases the odds of a crisis occurrence by more than 1. The model validation is performed the same way as for the short model by comparing model on full data, model estimated on truncated data till 2011 and model on truncated data till 2005.

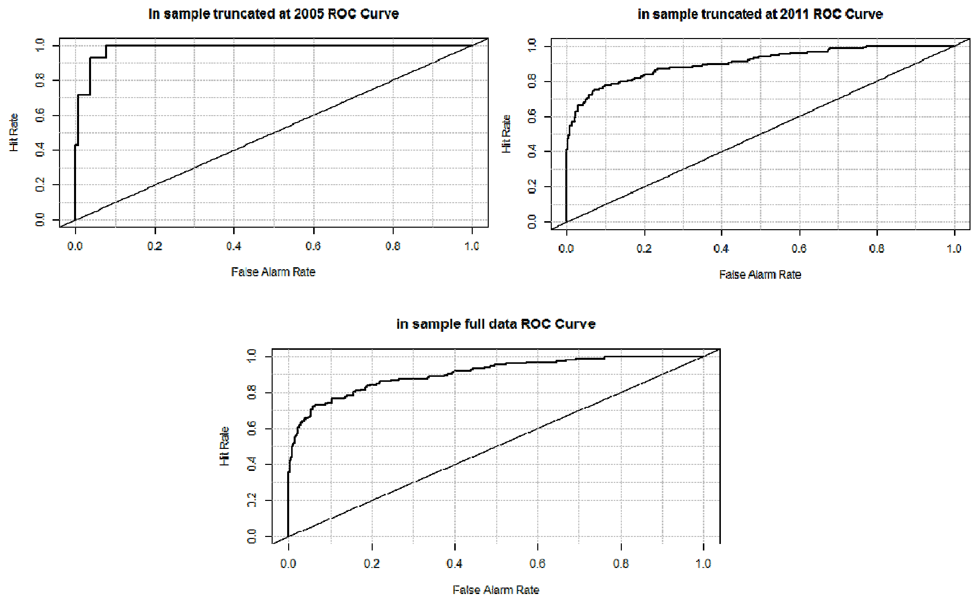
Table 3.6: In-sample performance of long logit models

In-sample performance of long logit models							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Long truncated till 2005	0.461	0.845	92.31	100	0.084	0.984	1.04E-09
Long truncated till 2011	0.339	0.726	87.37	75	0.099	0.905	1.18E-46
Long on full data	0.334	0.753	87.86	73.03	0.089	0.905	2.44E-49

In the spirit of the short model analysis, the best in-sample performing long model is the one estimated on data up till 2005. This model boasts the highest PCP, percentage of crises predicted as well as area under ROC curve while it has the lowest NtS ratio of only 0.084.

All in all, the differences in performance measures between the best performing long model in-sample and the second best are not very substantial with the largest difference of 36% for U measure.

Figure 3.5: ROC curve plots for in-sample performance of long logit model estimated on data up till 2005, till 2011 and on all available data



Now, it is of interest to analyse the model's performance out-of-sample and to detect the differences.

Table 3.7: Out-of-sample performance of long logit models

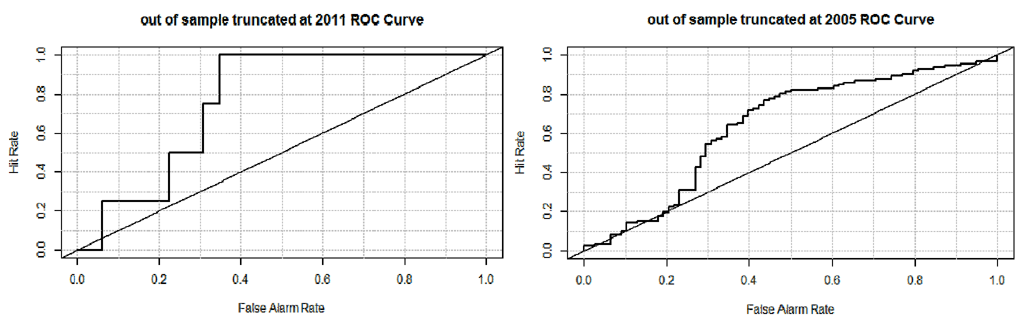
Out-of-sample performance of long logit models							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Long truncated till 2011	0.327	0.596	66.04	100	0.367	0.765	0.041465
Long truncated till 2005	0.166	0.365	67.89	76.79	0.584	0.639	0.000579

The first look reveals that the better performing long model is not the one estimated on data up till 2005 as is the case for short models but the one estimated on data till 2011 and projected over the last couple of years. The U measure of the better performing model is double of that for worse performing one. The percentage of crises predicted for this model is 23% higher than that of its counterpart while NtS ratio is 37% lower and area under ROC curve is almost 20% larger. However, despite the larger ROC area the better out-of-sample model is significant only on 5% level while the worse model's ROC area is significant on 0.06%.

In comparison to the in-sample performance of the model estimated on data up till 2011, its out-of-sample performance measures decline by 3.7% for U, 24.4% for PCP and 15.5% for area under ROC curve. Other measures increase out-of-sample, namely percentage of crises predicted by 25% and NtS ratio by 270%.

As for the worse out-of-sample performing model but the best one in-sample, estimated on data up till 2005, its U measure falls by 64%, PCP by 26.5%, percentage of crises predicted by 24.2%, its area under ROC curve by 35% while its NtS ratio rockets by 595% to almost 0.6, all out-of-sample.

Figure 3.6: ROC curve plots for out-of-sample performance of long logit model estimated on data up until 2011 and till 2005



To conclude, comparatively it appears that out-of-sample performance of the model estimated on data until 2011 deteriorates less than that of the model estimated on truncated data till 2005 making the better model more stable when estimated both in-sample and out[10].

3.5 Model Application to the Czech Republic

There are several advantages of deriving EWS for a single country from a panel EWS. First, a panel approach ensures robustness of the constructed early warning framework and limits influence of potential data problems on the resulting model.

[10] Tables A3.5 and A3.6 in the Appendix present out-of-sample prediction performance of both the short and the long model for which crisis-leading indicators were selected using two alternative approaches; first, the Recursive Feature Elimination (RFE) from the full set of potential indicators (Table A3.5) and second, the backward variable selection on the pre-selected subset of useful indicators (Table A3.6). The tables also offer out-of-sample performance comparison of these alternative models with models whose indicators were selected using BMA.

Second, due to more available data a panel approach allows for the use of more novel data-intensive techniques for crisis determinants identification, such as BMA. Third, especially for developing and transition economies a panel approach to building a country-specific EWS mitigates estimation difficulties related to data restrictions. For these reasons, we observe and evaluate in this section how well our panel EWS framework performs over short and long horizon for the Czech Republic, respectively.

3.5.1 Performance of the Short Model for the Czech Republic

We apply the short EWS with 11 independent variables and an intercept to the full Czech data from 1990Q1 till 2013Q1 as well as only to truncated data until 2011 with the objective of evaluating the model's predictions of systemic stress, i.e. the quality of its in-sample forecasting performance. Since data availability issues are coupled with the fact that the binary FSI in the short form for the Czech Republic contains only zeros until 2006Q3, we are unable to evaluate the short model's in-sample performance on truncated data until 2006. To remain consistent with the panel EWS performance evaluation in Section 3.4, we do not report out-of-sample performance evaluation for the short model truncated till 2006, either.

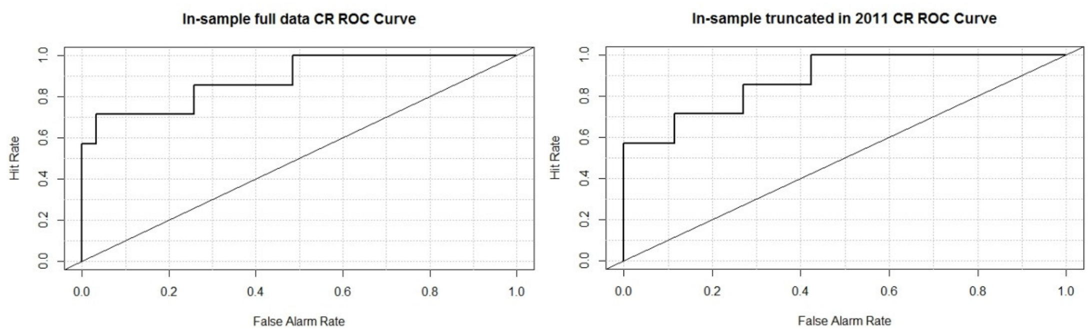
The same measures that were employed to assess the model's performance on panel data are also applied here.

Table 3.8: In-sample performance of the short model for the Czech Republic

In-sample performance of the short model for the Czech Republic							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Short full data	0.34	0.86	92.1	71.43	0.05	0.889	0.000329
Short truncated till 2011	0.30	0.78	84.8	71.43	0.16	0.885	0.000533

As evidenced from Table 3.8 the short model estimated on all available data for the Czech Republic performs better in all performance statistics than the model on truncated data. The model successfully predicts 92.1% of observations as well as 71.43% of systemic events. The quality of in-sample forecast is captured by the area under ROC curve and no discrimination line (the diagonal), that attains for both a value of at least 0.88 while its p-value is quite low. This indicates a good forecasting skill of the short model on both data samples.

Figure 3.7: ROC curves for in-sample performance of the short logit model on full data and truncated data for the Czech Republic



3.5.2 Performance of the Long Model for the Czech Republic

In similar fashion, we apply the long model with 12 most useful indicators from BMA technique to all available Czech data and to truncated data ending by 2011. The same as for the short model, we are unable to evaluate the long model's in-sample performance on truncated data until 2005 due to absence of 1s in binary FSI in the long form until 2005Q1. For this reason, we do not report out-of-sample performance evaluation for the long model truncated till 2005, either.

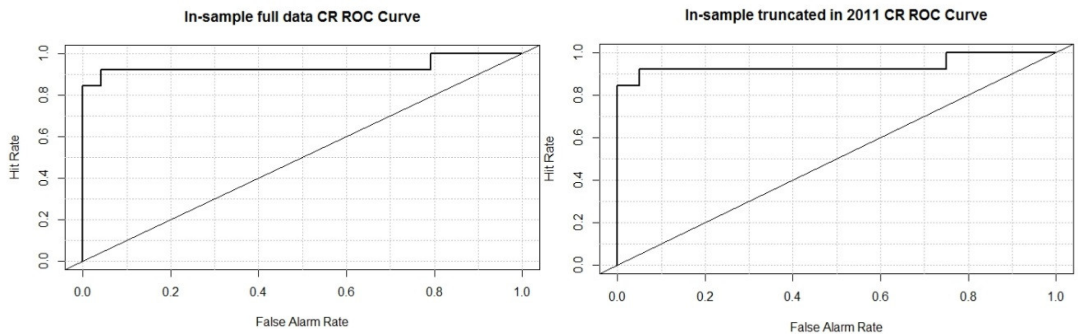
The objective is the same as for the short model estimation on the Czech data; to assess the model's in-sample predictive ability. For this purpose, a set of performance measures is applied to evaluate the model's prediction of the binary dependent.

Table 3.9: In-sample performance of the long model for the Czech Republic

In-sample performance of the long model for the Czech Republic							
Model	U	Threshold	PCP	% crises predicted	NtS ratio	ROC area	p-value
Long full data	0.441	0.67	94.6	92.3	0.045	0.936	7.408E-07
Long truncated till 2011	0.437	0.63	93.9	92.3	0.054	0.938	1.584E-06

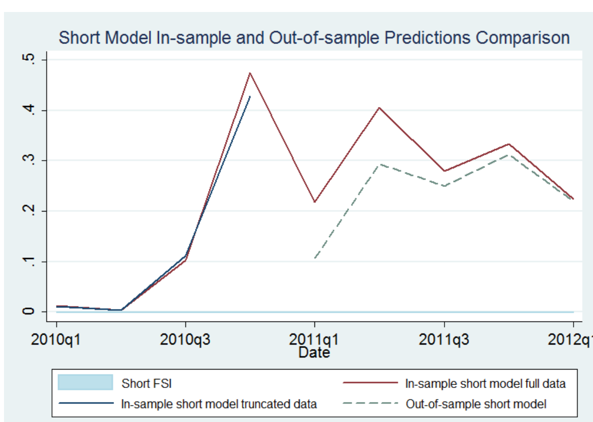
According to performance measures, the long model on both samples performs very well for the Czech Republic. Within the sample both models exceed the utility value of 0.4, predict 92.3% of systemic events and attain the area under ROC curve of more than 0.93 for which p-value is quite low.

Figure 3.8: ROC curves for in-sample performance of the long logit model on full data and truncated data for the Czech Republic



Overall, the best performing models for the Czech Republic are the long models followed by models over the short horizon. The highest ranking model is the one estimated on all available data over the long horizon. Ultimately, highest ranking model, designed to anticipate crises within long horizon of 12 quarters, outperforms the short model on all data in terms of utility measure by 0.1, percentage of crises predicted by almost 21% and ROC area by 0.05. On the other hand, both models perform similarly well in terms of noise-to-signal ratio and percentage of observations correctly predicted.

Figure 3.9: In-sample, out-of-sample predictions comparison of the long and the short model for the Czech Republic



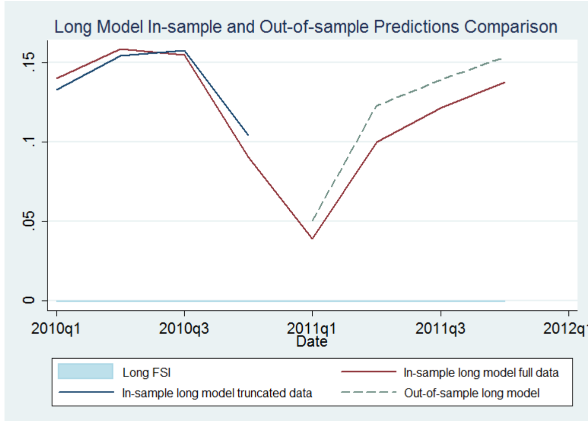


Figure 3.9 depicts systemic event probabilities estimated by both models not only in-sample on full Czech data and on truncated data until 2011 but also out-of-sample from 2011 onwards. The out-of-sample estimates do not diverge substantially from in-sample probabilities for either model. Furthermore, out-of-sample forecast of the long model exhibits closer proximity to the model's in-sample prediction, confirming superior performance of the long models over short models for the Czech Republic.

3.6 Robustness Checks

In this section we investigate the suitability of our large cross-country panel EWS to predict systemic events and the model's performance for the Czech Republic over both horizons. For this purpose, we construct an alternative EWS in line with Candelon et al. (2008), consisting of a subpanel of the three economies that belong to the same geographical region, i.e. Hungary, Czech Republic and Euro area. Such a more homogenous cross-country EWS should allow for countries pooling in a panel model without losing information and affecting model estimation. Candelon et al. (2008) suggest that countries should be pooled into a panel model without losing information when they form an optimal cluster based on Hausman poolability test technique derived in Kapetanios (2003). However, in our case finding the optimal country cluster is hindered by an additional uncertainty in our analysis, i.e. model uncertainty with respect to selecting useful early warning indicators.

3.6.1 Alternative Short Model Estimation and Performance

We proceed to construct our alternative EWS built on a regional subpanel of countries the same way as in the main analysis. First, we address uncertainty in regards to selecting useful early warning indicators by means of BMA. Again, we run BMA on the same set of 78 potential indicators for the short model. The selected useful indicators after having excluded collinear variables for the short model are presented in Table 3.10, which also presents short model estimated on all available data for the subpanel of countries.

Table 3.10: Short model estimation on full data sample

Coefficient	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*	
(Intercept)	-19.060	8.021	-2.376	0.017497	*
CPIg_7	85.648	33.067	2.590	0.009593	**
realGDPg_1	65.851	21.858	3.013	0.002590	**
Globpcredg_8	102.767	46.644	2.203	0.027580	*
realM2g_4	-48.498	17.395	-2.788	0.005303	**
U_rate_2	109.181	60.067	1.818	0.069119	.
mcapGDP_7	12.779	3.647	3.504	0.000458	***
CPIg_1	72.936	33.358	2.186	0.028783	*
Globpcredg_7	123.031	42.371	2.904	0.003688	**
GlobpcgxglobpcGDP_4	33.351	11.020	3.027	0.002474	**
curaccGDP_8	58.743	32.699	1.796	0.072419	.

Note: the number following each indicator states an indicator‘s lag (in quarters), . represents significance on 10% , * on 5%, ** on 1% and *** on 0.1% significance level.

The short model identifies global private credit indicators, market capitalisation over GDP ratio, unemployment rate and growth of monetary aggregate M2 among others as useful indicators for crises identification. These results should, however, be interpreted with caution as the Monte Carlo Markov Chain (MCMC) algorithm used in BMA does not satisfactorily converge to the underlying distribution.

In our case, the failure to converge, despite specifying the same number of iterations and burn-ins of the algorithm as in the construction of the baseline EWS, might stem from the fact that there are too many potential indicators relative to the number of observations in the subpanel which in turn might cause the selected indicators vary when repeatedly running BMA. In addition, the more complicated the distribution of marginal likelihoods, the more difficult it will be for the MCMC sampler to attain a satisfactory correlation with the underlying distribution (Feldkircher and Zeugner, 2009).

Next, we evaluate the alternative short EWS model's performance. We present the evaluation only for models estimated on full data sample as the subpanel of countries with fewer observations prevents us from estimating truncated models and performing out-of-sample assessment consistently over both horizons.

In addition to the problem of collinearity, which we address in building our EWS, discrete data regressions can also become unstable from separation. Separation or perfect prediction arises when some linear combination of the predictors is perfectly predictive of the outcome (Albert and Anderson, 1984 and Lesaffre and Albert, 1989). In order to solve separation, independent variables are gradually removed until the final model is identifiable. However, according to Zorn (2005) this approach may result in removing the strongest predictors from the model. Therefore the technique to employ in case of perfect prediction is Bayesian inference. The Bayesian estimation of logistic regression is used for both, the short and the long model, on the regional panel data as the traditional maximum likelihood estimation suffers from perfect prediction. This demonstrates by producing abnormally large coefficient as well as standard error estimates while p-value equals 1 for all coefficient estimates[11].

Gelman, Jakulin, Pittau and Su (2008) adapt the classical maximum likelihood algorithm within logit model in a way to obtain approximate posterior inference for the coefficients β. The standard logistic regression algorithm, upon which this technique expands, proceeds by approximately linearizing the derivative of the log-likelihood, solving by means of weighted least squares, and then iterating this process, each step evaluating the derivatives at the latest estimate $\hat{\beta}$ (McCullagh and Nelder, 1989). We do not introduce any additional information about prior distribution in the logit model estimation for the regional panel of countries.

Table 3.11 shows in-sample performance statistics of the short EWS constructed from the regional subpanel compared to the large panel EWS from the main analysis.

[11] To yield stable coefficient estimates for logit models via Bayesian inference the "arm" package in R, built to accompany the paper by Gelman, Jakulin, Pittau and Su (2008), was used for the calculations.

Table 3.11: In-sample short models' performance on full data

In-sample short models' performance on full data							
	U	Threshold	PCP	% crises predicted	NtS	ROC area	p-value
Panel	0.34	0.81	89.43	73.53	0.09	0.91	0.00
Subpanel	0.41	0.61	86.51	100.00	0.00	0.97	0.00

We can observe that regional subpanel EWS outperforms the large cross-country EWS in terms of utility, noise-to-signal ratio, area under ROC curve and percentage of predicted crises. These results support the findings in the literature that crisis determinants may to a large extent differ across regions (Davis et al., 2011; Candelon et al., 2008).

3.6.2 Alternative Long Model Estimation and Performance

The robustness check for the long model estimated on the subpanel of countries is conducted in the same way as for the model over the short horizon. 74 potential indicators are input into BMA and the resulting long model estimated on full data in the subpanel is presented in Table 3.12.

Table 3.12: Long model estimation on full data sample

Coefficient	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*	
(Intercept)	-62.234	21.104	-2.949	0.00319	**
U_rate_5	287.706	104.934	2.742	0.00611	**
Globpcredg_1	-92.197	65.067	-1.417	0.15649	
GlobmcapGDP_3	31.723	11.611	2.732	0.00629	**
real.money.g	-46.898	18.255	-2.569	0.01020	*
realM2g	108.134	39.802	2.717	0.00659	**
NPL_8	-87.820	58.143	-1.510	0.13094	
mcapGDP	-8.165	6.395	-1.277	0.20166	
GlobmcapGDP	14.826	7.325	2.024	0.04296	*

Note: the number following each indicator states an indicator‘s lag (in quarters), . represents significance on 10% , * on 5%, ** on 1% and *** on 0.1% significance level.

The same as the short model, the long model identifies global private credit indicators, market capitalisation over GDP ratio, unemployment rate and growth of monetary aggregate M2 as useful for crises identification. Monte Carlo Markov Chain (MCMC) algorithm used in BMA does not satisfactorily converge to the underlying distribution for the same reasons as stated above, thus the results should be interpreted with caution.

Table 3.13 shows comparison of in-sample performance statistics of the long EWS constructed from the regional subpanel and the large panel EWS from the main analysis.

Table 3.13: In-sample long models' performance on full data

In-sample long models' performance on full data							
	U	Threshold	PCP	% crises predicted	NtS	ROC area	p-value
Panel	0.33	0.75	87.86	73.03	0.09	0.91	0.00
Subpanel	0.47	0.62	96.40	100.00	0.00	0.99	0.00

Consistently with the short model, regional subpanel EWS outperforms the large cross-country EWS in all performance statistics. Overall, we detect a comparatively more substantial improvement in the long model performance.

3.6.3 Application to the Czech Republic

Now we turn to evaluating the performance of our regional panel EWS for the Czech Republic compared to the baseline EWS from the main analysis. In the same vein, Tables 3.14 and 3.15 present the performance evaluation of the short and the long model for the Czech Republic, respectively.

Table 3.14: In-sample short models' performance on full Czech data

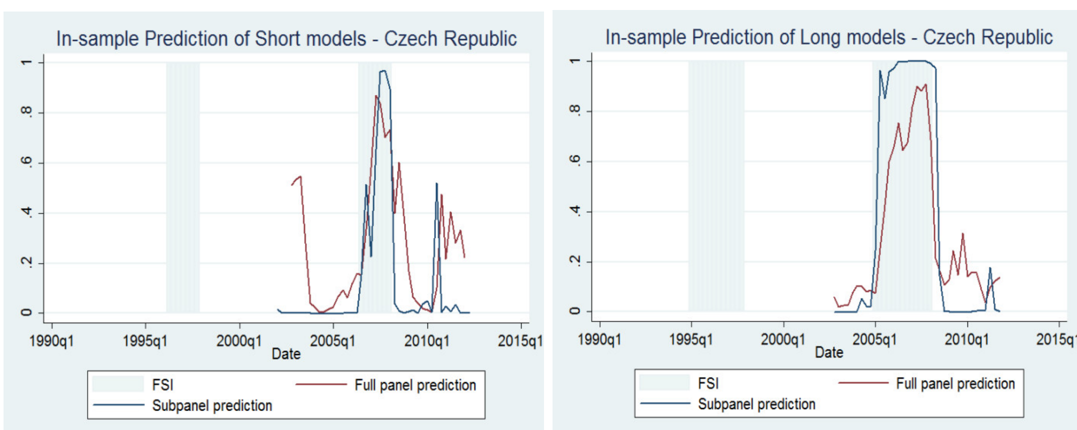
In-sample short models' performance on full data for the Czech Republic							
	U	Threshold	PCP	% crises predicted	NtS	ROC area	p-value
Panel	0.34	0.86	92.11	71.43	0.05	0.89	0.00
Subpanel	0.49	0.83	97.62	100.00	0.00	0.99	0.00

Table 3.15: In-sample long models' performance on full Czech data

In-sample long models' performance on full data for the Czech Republic							
	U	Threshold	PCP	% crises predicted	NtS	ROC area	p-value
Panel	0.44	0.67	94.59	92.31	0.05	0.94	0.00
Subpanel	0.48	0.64	97.30	100.00	0.00	0.98	0.00

Regional EWS outperforms baseline EWS over both horizons for the Czech Republic. Furthermore, it almost attains maximum utility and predicts 100% of crisis events, supporting the notion that regional or more homogenous panel models perform better than large cross-country panel EWS (e.g. Davis et al., 2011). Figure 3.10 plots in-sample predictions of the baseline EWS and regional EWS models estimated over both horizons.

Figure 3.10: In-sample predictions of baseline and regional EWS over both horizons for the Czech Republic



Despite its superior performance, building a more homogenous regional EWS in our analysis requires some significant simplifications. First, non-convergence of the sampling MCMC algorithm in BMA due to a large number of potential indicators and too few data points introduces doubts about actual usefulness of the selected indicators and their stability after repeatedly running BMA on the same dataset. Second, the regional EWS performance evaluation is reduced to only in-sample analysis. Too few data points for the countries in the regional EWS, i.e. Hungary, Czech Republic and Euro area, do not allow for estimating models truncated in 2006 and executing out-of-sample performance analysis over the period of the Global crisis.

Due to these obstacles in building regional EWS, we can conclude that our baseline cross-country panel model is the preferred EWS framework with which to perform an early warning exercise for the Czech Republic. Since the Czech Republic is a post-transition country, Czech macroeconomic time series are quite short, spanning over less than two decades, as is the case for other countries in the region. Therefore, a suitable EWS framework derived from a panel approach for the Czech Republic should incorporate also advanced economies that provide longer time series to avoid computational challenges.

Conclusions

The aim of this study is to develop EWS framework for predicting systemic events for the Czech Republic. We develop two models on the panel of 14 countries, i.e. the short model to allow prediction of events over the horizon of 1.5 years and the long model over the horizon of 3 years. We validate the models' performance on the panel of 14 countries and subsequently observe their skill when applied to the Czech data.

First, the Financial stress index (FSI), measuring the level of financial stress within the financial system, is constructed for each country within the panel following Lo Duca and Peltonen (2013). To aggregate individual subindices from equity, foreign exchange, money and securities markets into FSI, a market-equal weighting is employed due to the cross-country nature of the analysis. FSI thus reports average level of systemic stress in the economy at each point in time (quarter). Moreover, FSI transformed into binary form is used for identification of starting dates of country-specific systemic events.

Second, uncertainty in regards to the inclusion of potential leading indicators into EWS that would best explain crises occurrence is resolved by Bayesian model averaging (BMA) technique. We relax the assumption of a common fixed horizon at which all potential indicators issue early warning signals and detect indicators' relevant lags for signal emission by univariate logit models. Overall, the selected indicators for both horizons are in accordance with the literature which identifies credit indicators as the most useful (Alessi and Detken 2011; Borio and Lowe 2004) as well as their domestic and their global and domestic interactions (Lo Duca and Peltonen, 2013).

Third, the binary logit model containing the BMA selected indicators is estimated for both horizons on the panel. Over the short horizon the best performing model, both in-sample as well as out-of-sample, is the one estimated on data till 2006 with its out-of-sample performance tested over the pre-crisis period of the Global crisis (2006Q1-2008Q1).

The best performing long model in-sample is the one estimated on data till 2005 while out-of-sample it is the one estimated on data until 2011 and projected over the next two years. However, out-of-sample performance of the model estimated on data until 2011 comparatively deteriorates less than that of the model estimated on data till 2005, which makes the model with better out-of-sample performance more stable.

Fourth, observing the performance for the Czech Republic, the highest ranking model in-sample is the model over the long horizon estimated on full data. The model manages to correctly predict more than 92% of systemic events, attains utility of more than 0.44 and its area under ROC curve exceeds 0.93 indicating a very good in-sample predictive skill of the model.

Next, we observe also out-of-sample performance of the constructed EWS over both horizons on the Czech data from 2011 onwards. The out-of-sample estimates of systemic event probabilities do not deviate substantially from their in-sample estimates indicating good out-of-sample performance of the built EWS for the Czech Republic.

Last but not least, comparing the performance of our EWS with a more homogenous regional EWS framework reveals a weaker skill of our baseline EWS. When constructing the regional EWS we faced computational issues in the form of instability of selected crises indicators (insufficient convergence of the sampling algorithm) and impossibility to perform out-of-sample early warning exercise (short macroeconomic time series for the transition economies in the region). Consequently, our baseline EWS remains the preferred early warning framework.

References

Albert, A., Anderson, J. A., 1984. On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. Biometrika 71, pp. 1-10.

Alessi, L. and Detken, C., 2011. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/bust Cycles: A Role for Global Liquidity. European Journal of Political Economy, vol. 27(3), pp. 520-533.

Babecky, J., Havranek,T., 2014. Structural Reforms and Growth in Transition. The Economics of Transition. The European Bank for Reconstruction and Development, vol. 22(1), pp. 13-42, 01.

Babecky J., Havranek T., Mateju J., Rusnak M., Smidkova K., Vasicek B., 2013. Leading Indicators of Crisis Incidence: Evidence from Developed Countries, Journal of International Money and Finance, vol. 35(C), pp. 1-19.

Babecky J., Havranek T., Mateju J., Rusnak M., Smidkova K., Vasicek B., 2014. Banking, Debt, and Currency Crises in Developed Countries: Stylized Facts and Early Warning Indicators, Journal of Financial Stability, vol. 15(C), pp. 1-17.

Borio, C., Lowe, P., 2004. Securing Sustainable Price Stability: Should Credit Come Back from the Wilderness?. Bank for International Settlements Working Papers, No. 157.

Bussiere, M. and Fratzscher, M., 2006. Towards a New Early Warning System of Financial Crises. Journal of International Money and Finance, 25, pp. 953-973.

Candelon, B., Van den Berg, J., Urbain, J.-P., 2008. A Cautious Note on the Use of Panel Models to Predict Financial Crises. Economics Letters, vol. 101 (1), pp. 80-83.

Caprio, G., Klingebiel, D., 2006. Episodes of Systematic and Borderline Financial Crises. World Bank,Washington.

Cardarelli, R., S. A. Elekdag and S. Lall, 2011. Financial Stress and Economic Contractions. Journal of Financial Stability, Vol. 7, pp. 78-97.

Cramer, J.S., 2003. Logit Models from Economics and Other Fields. Cambridge University Press.

Crespo Cuaresma, J., Slacik, T., 2009. On the Determinants of Currency Crises: The Role of Model Uncertainty. Journal of Macroeconomics, vol. 31(4), pp. 621-632.

Davis, E. P., Karim, D., Liadze, I., 2011. Should Multivariate Early Warning Systems for Banking Crises Pool Across Regions? Rev World Econ, vol. 147 (4), pp. 693-716.

Demirguc-Kunt, A., Detragiache, E., 2005. Cross-Country Empirical Studies of Systemic Bank Distress: A Survey. IMF Working Paper, No. 96/05, Washington 2005.

Feldkircher, M. and Zeugner, S., 2009. Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging. IMF Working Paper No. 09/202.

Feldkircher, M., Zeugner, S., 2015. Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. Journal of Statistical Software, vol. 68 (4), pp. 1-37.

Fostel, A. and Geneakoplos, J., 2008. Leverage Cycles and the Anxious Economy. American Economic Review, Vol. 98, No. 4, pp. 1211-1244.

Gelman, A., Jakulin, A., Pittau, M.G., Yu-Sung Su, 2008. A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. The Annals of Applied Statistics 2008, Vol. 2(4), pp. 1360–1383.

Gramlich D., Miller G.L., Oet M.V., Ong S.J., 2010. Early warning systems for systemic banking risk: critical review and modeling implications. Banks and Bank Systems, vol. 5(2), 2010.

Hakkio, S. C. and W. R. Keeton, 2009. Financial Stress: What Is It, How Can It be Measured, and Why Does It Matter?. Federal Reserve Bank of Kansas City, Economic Review, Second Quarter, pp. 5-50.

Havranek, T., Rusnak, M., 2013. Transmission Lags of Monetary Policy: A Meta-Analysis. International Journal of Central Banking, vol. 9(4), pp. 39-76, December.

Hendricks, D., Kambhu, J., Mosser, P., 2007. Systemic Risk and the Financial System. Federal Reserve Bank of New York (Ed.); Economic Policy Review, vol. 13 (2), pp. 65-80.

Hollo, D., Kremer, M., Lo Duca, M., 2012. CISS - A composite Indicator of Systemic Stress in the Financial System, ECB Working Paper Series 1426.

Illing, M., Liu, Y., 2006. Measuring Financial Stress in a Developed Country: An Application to Canada. Journal of Financial Stability 2 (October (3)), pp. 243–265.

Jakubik P., Slacik T., 2013. Measuring financial (in)stability in Emerging Europe: A new index-based approach. Oesterreichische Nationalbank (Austrian Central Bank), issue 25, June, forthcoming.

Kaminsky G., Lizondo S., Reinhart C., 1998. Leading indicators of currency crises. International Monetary Fund Staff Papers 45, pp. 1-48.

Kaminsky G., Reinhart C., 1999. The twin crises: The causes of banking and balance of payments problems. American Economic Review 89(3): June, pp. 473-500.

Kapetanios, G., 2003. Determining Poolability of Individual Series in Panel Datasets. University of London Queen Mary Economics Working Paper No. 499.

Laeven, L. and Valencia, F., 2008. Systemic Banking Crises: A New Database. IMF Working Paper No. 08/224.

Laeven, L. and Valencia, F., 2012. Systemic Banking Crises Database: An Update. IMF Working Paper No. 12/163, June 01, 2012.

Lesaffre, E., Albert, A., 1989. Partial Separation in Logistic Discrimination. J. Roy. Statist. Soc. Ser. B 51, pp. 109-116.

Lo Duca, M., Peltonen, T., 2013. Assessing Systemic Risks and Predicting Systemic Events. Journal of Banking and Finance, Vol. 37 (2013), No. 7, pp. 2183-2195.

Louzis, D. P., Vouldis, A. T., 2011. A Financial Systemic Stress Index for Greece. Paper presented at the First Conference of the Macro-prudential Research (MaRs) network of the European System of Central Banks in Frankfurt am Main, October.

Madigan, D. and York, J., 1995. Bayesian Graphical Models for Discrete Data. International Statistical Review, vol. 63(2), pp. 215–232.

Mason, S.J., Graham, N.E., 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of Royal Meteorological Society (2002), vol. 128, pp. 2145–2166.

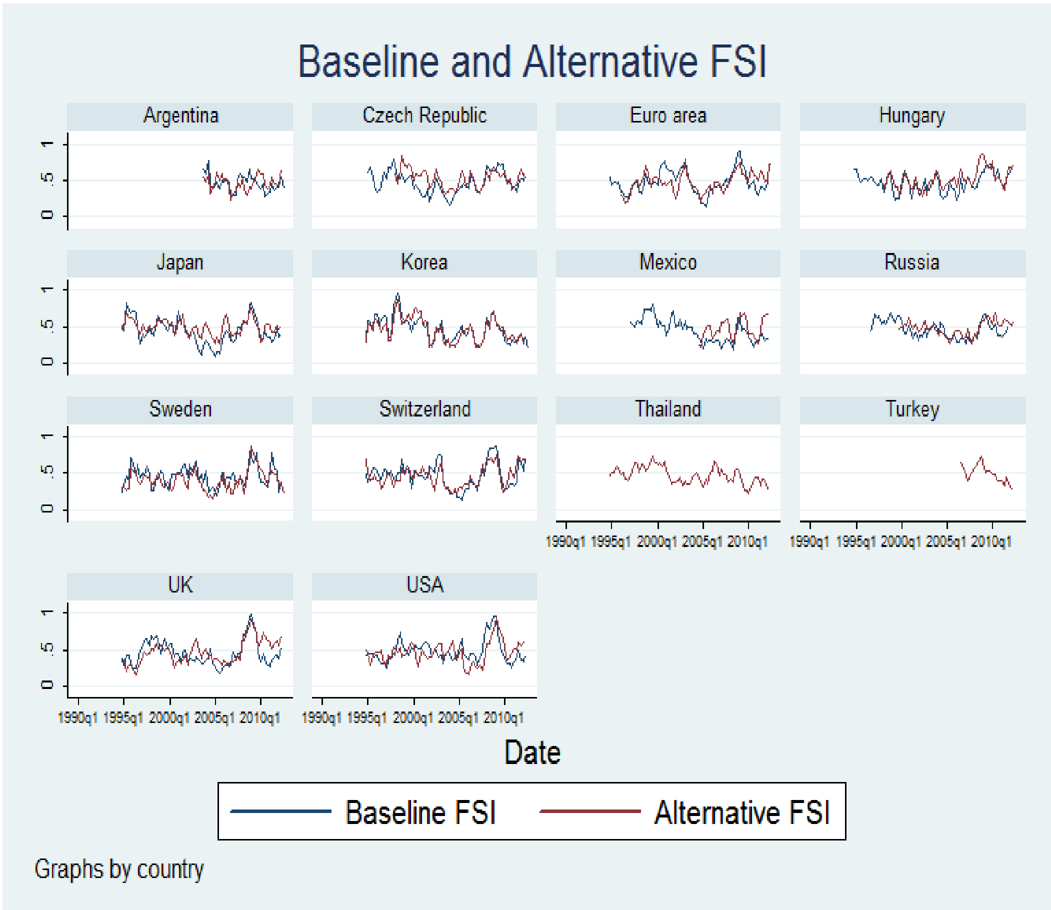
McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd ed. Chapman and Hall, London.

Oet, M. V., R. Eiben, T. Bianco, D. Gramlich and S. J. Ong, 2011. The Financial Stress Index: Identification of Systemic Risk Conditions. Federal Reserve Bank of Cleveland, Working Paper 11-30, November.

Zorn, C., 2005. A solution to separation in binary response models. Political Analysis vol. 13, pp. 157-170.

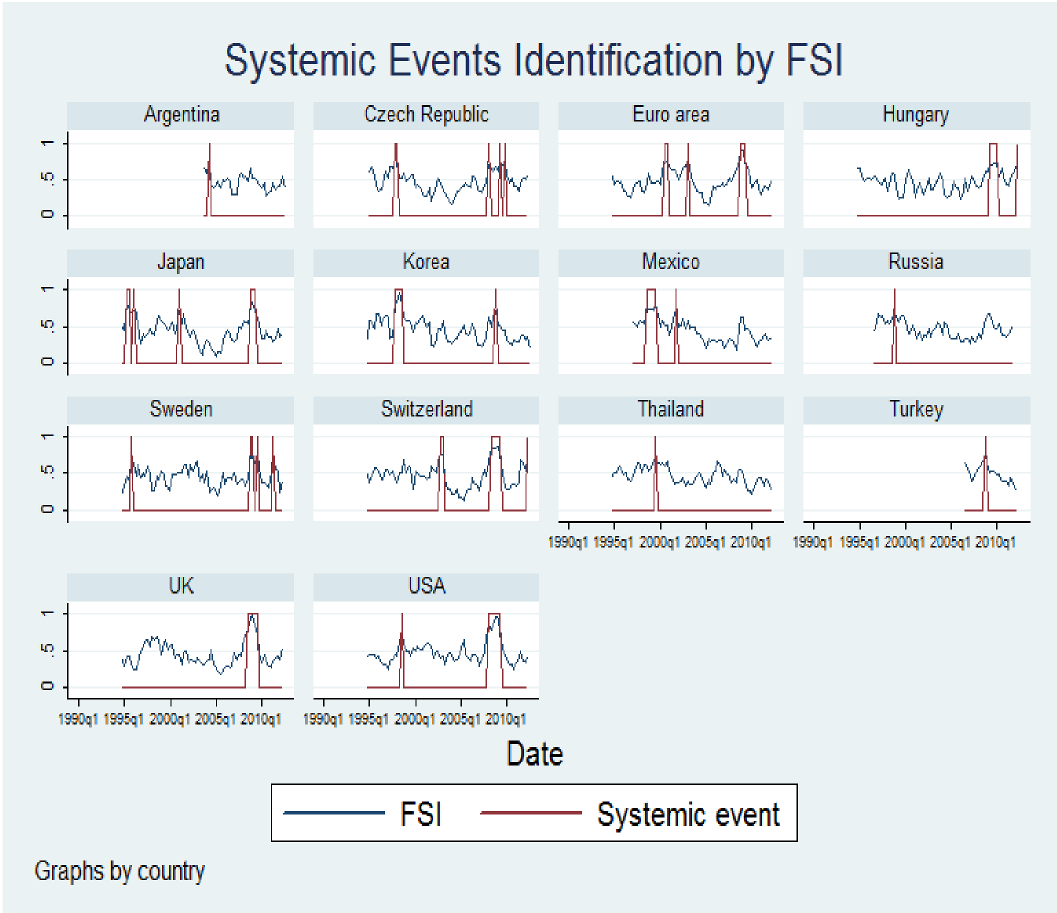
Appendix

Figure A3.1: Baseline FSI and Alternative FSI comparison



Notes:
Argentina: FSI registers the highest stress in 2004 in the wake of debt, systemic banking and currency crises with starting dates of 2001, 2001 and 2002, respectively. Czech Republic: The FSI peaks in 1997 following systemic banking crisis starting in 1996. Euro area: FSI attains its highest values during crisis period in 2008/2009. Hungary: FSI identifies higher stress levels during the recent global crisis than during the Hungarian systemic banking crisis of 1991. Japan: FSI registers increased stress during the recent global crisis and before the outbreak of the Japanese systemic banking crisis in 1997. South Korea: FSI peaks during the Korean systemic banking crisis of 1997. Mexico: Due to data restrictions FSI misses Mexican banking crisis of 1994 but reflects major insolvency of Mexican banks by the year 2000. Russia: The FSI captures the highest stress levels during the Russian crisis of 1998. Sweden: FSI peaks during the recent global crisis and also reflects increased stress in 1995, in the aftermath of the Swedish systemic banking crisis. Switzerland: FSI indicates high stress during the recent global crisis. Thailand: only alternative FSI is constructed due to data restrictions, it reflects well the high stress during the Asian crisis of 1997/1998. Turkey: only alternative FSI is constructed covering the shortest time period of all indices due to data restrictions. United Kingdom and United States: Both FSIs peak during the recent global crisis.

Figure A3.2: Systemic Events Identification by Means of FSI Threshold Value of 0.7



Notes:
Argentina: Systemic event taking place in 2004Q2, near the time of the Argentinian systemic banking crisis. Czech Republic: The threshold exceeded in 1997Q4 and 1998Q1 during the country-specific systemic banking crisis and during the recent global crisis. Euro area: The threshold exceeded in the second half of 2000, the beginning of 2003 and during the recent crisis. Hungary: Event episodes recognized during the global crisis. Japan: The threshold exceeded in periods around the Japanese systemic banking crisis and during the global crisis. South Korea: The threshold identifies the presence of systemic events in the Korean crisis period and during the global crisis. Mexico: Events are identified in the wake of the Mexican systemic banking and currency crises. Russia: The recognized systemic event falls into the Russian crisis period. Sweden: The threshold exceeded in the country-specific systemic banking and the global crisis periods. Switzerland: Systemic events mainly recognized during the global crisis. Thailand: The threshold exceeded after the Asian crisis outburst. Turkey: Due to data constraints the systemic event recognized during the global crisis only. United Kingdom and United States: Both indices exceed the threshold during the global crisis.

Table A3.1: Descriptive Statistics of FSI – Full Dataset

SHORT MODEL				
		Full panel	*Panel truncated in 2011*	*Panel truncated in 2006*
FSI VALUE	1	179	165	88
	0	715	645	444
NO. OF OBS		894	810	532
LONG MODEL				
		Full panel	*Panel truncated in 2011*	*Panel truncated in 2005*
FSI VALUE	1	271	265	131
	0	623	545	349
NO. OF OBS		894	810	480

Table A3.2: Descriptive Statistics of FSI – Effective Dataset

SHORT MODEL				
		Full panel	*Panel truncated in 2011*	*Panel truncated in 2006*
FSI VALUE	1	102	90	13
	0	428	379	180
NO. OF OBS		530	469	193
LONG MODEL				
		Full panel	*Panel truncated in 2011*	*Panel truncated in 2005*
FSI VALUE	1	152	148	14
	0	400	351	155
NO. OF OBS		552	499	169

Table A3.3: Descriptive Statistics of FSI – Full Subpanel

SHORT MODEL				
		Full subpanel	*Subpanel truncated in 2011*	*Subpanel truncated in 2006*
FSI VALUE	1	49	43	21
	0	163	151	113
NO. OF OBS		212	194	134
LONG MODEL				
		Full subpanel	*Subpanel truncated in 2011*	*Subpanel truncated in 2005*
FSI VALUE	1	64	64	25
	0	148	130	97
NO. OF OBS		212	194	122

Table A3.4: Descriptive Statistics of FSI – Effective Subpanel

SHORT MODEL				
		Full subpanel	*Subpanel truncated in 2011*	*Subpanel truncated in 2006*
FSI VALUE	1	33	27	5
	0	93	81	43
NO. OF OBS		126	108	48
LONG MODEL				
		Full subpanel	*Subpanel truncated in 2011*	*Subpanel truncated in 2005*
FSI VALUE	1	39	39	0
	0	72	60	27
NO. OF OBS		111	99	27

Table A3.5: Comparison of Model Out-of-sample Performance – All Variables

SHORT MODEL		
	BMA selection	*RFE selection*
	AUC	*AUC*
TRUNCATED IN 2006	0.691	0.589
TRUNCATED IN 2011	0.599	0.638
LONG MODEL		
	BMA selection	*RFE selection*
	AUC	*AUC*
TRUNCATED IN 2005	0.639	0.555
TRUNCATED IN 2011	0.765	0.937

Notes: The table presents comparison of out-of-sample performance of the short and long logit models for which the selection of leading indicators was performed using BMA on the full set of potential indicators versus applying Recursive Feature Elimination (RFE) on the same full set of potential indicators.

Table A3.6: Comparison of Model Out-of-sample Performance – Pre-selected Variables

SHORT MODEL		
	BMA selection	*Backward selection*
	AUC	*AUC*
TRUNCATED IN 2006	0.779	0.733
TRUNCATED IN 2011	0.673	0.734
LONG MODEL		
	BMA selection	*Backward selection*
	AUC	*AUC*
TRUNCATED IN 2005	0.572	0.734
TRUNCATED IN 2011	0.919	0.772

Notes: The table presents comparison of out-of-sample performance of the short and long logit models for which the selection of leading indicators was performed using BMA on the pre-selected subset of useful indicators (i.e. those with AUC greater than 0.5) versus applying backward variable selection on the same subset of pre-selected useful indicators.

Chapter 4

Bank Competition and Financial Stability: Much Ado about Nothing?

Abstract

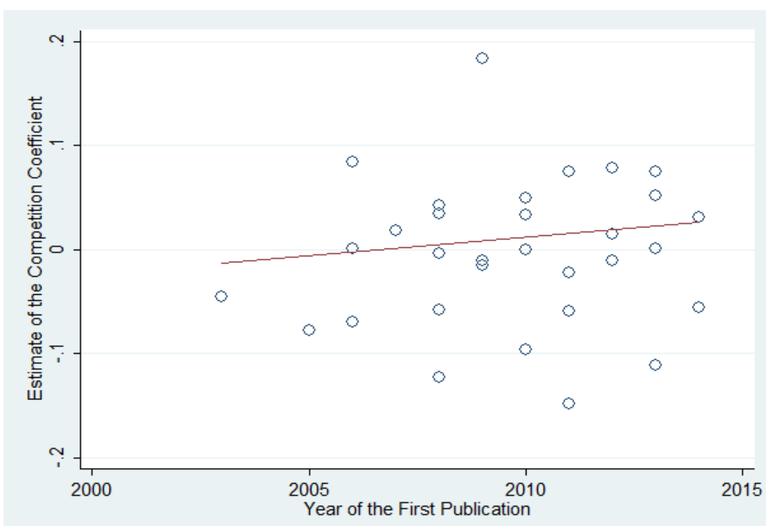
The theoretical literature gives conflicting predictions on how bank competition should affect financial stability, and dozens of researchers have attempted to evaluate the relationship empirically. We collect 598 estimates of the competition-stability nexus reported in 31 studies and analyze the literature using meta-analysis methods. We control for 35 aspects of study design and employ Bayesian model averaging to tackle the resulting model uncertainty. Our findings suggest that the definition of financial stability and bank competition used by researchers influences their results in a systematic way. The choice of data, estimation methodology, and control variables also affects the reported coefficient. We find evidence for moderate publication bias. Taken together, the estimates reported in the literature suggest little interplay between competition and stability, even when corrected for publication bias and potential misspecifications.

This paper was co-authored with Tomáš Havránek and published in Journal of Economic Surveys (2016, 30(5), pp. 944-981). The authors are grateful to Tom Stanley, Martin Paldam, Pedro Bom, Jost Heckemeyer, Josef Brechler, Bořek Vašíček, Jan Babecký, seminar participants at the Czech National Bank, participants of the 2014 MAER-Net Colloquium in Athens, and anonymous reviewers of Journal of Economic Surveys for their helpful comments on previous versions of the manuscript. Diana Zigraiova acknowledges support from the Czech Science Foundation (grant #15-02411S); Tomas Havranek acknowledges support from the Czech Science Foundation (grant #P402/12/G097). An online appendix with data and code is available at http://meta-analysis.cz/competition.

4.1 Introduction

The theory does not provide clear guidance on the expected sign of the relationship between bank competition and financial stability. On the one hand, the competition-fragility hypothesis (represented, for example, by Keeley, 1990) argues that competition hampers stability. Strong competition in the banking sector forces banks to take on excessive risks in the search for yield, which leads to overall fragility of the financial system. On the other hand, under the competition-stability hypothesis (for instance, Boyd and De Nicolo, 2005), increased competition makes the financial system more resilient. A competitive banking sector results in lower lending rates, which support firms' profitability, leading to lower credit risk for banks. Moreover, in uncompetitive environments banks are more likely to rely on their too-big-to-fail position and engage in moral hazard (Mishkin, 1999). Since the early 2000s, dozens of researchers have reported estimates of the competition-stability nexus, but their results vary. As Figure 4.1 shows, the reported results do not converge to a consensus number, complicating our inference from the literature.

Figure 4.1: The reported estimates of the competition-stability nexus do not converge



Notes: The figure depicts the median partial correlation coefficients corresponding to the effects of banking competition on financial stability reported in individual studies. The horizontal axis measures the year when the first drafts of the studies appeared in Google Scholar. The line shows the linear fit (the slight upward trend is not statistically significant).

When the literature lacks patterns visible at first sight, narrative surveys are useful in discussing the reasons for the heterogeneity observed in the results, but they cannot provide policy makers and other researchers with clear guidelines concerning the relationship in question. Our aim in this paper is to collect all available estimates of the relation between bank competition and financial stability, and examine them using up-to-date meta-analysis methods. Meta-analysis is most commonly applied in medical research to synthesize the results of clinical trials, and the use of this method dates back at least to Pearson (1904). Meta-analysis later spread to the social sciences, including economics and finance, and examples of early applications are summarized by Stanley (2001). Recent applications of meta-analysis include Chetty et al. (2011), who explore the intertemporal elasticity of substitution in labor supply, Doucouliagos et al. (2012), who investigate the link between chief executives' pay and corporate performance, and Babecky and Havranek (2014), who evaluate the impact of structural reforms on economic growth.

We collect 598 estimates of the competition-stability nexus from 31 studies published between 2003 and 2014, and present, to our knowledge, the first meta-analysis on the topic. We do not find evidence for any robust relationship between bank competition and financial stability: either the positive and negative effects of competition offset each other, or current data and methods do not allow researchers to identify the relationship. This conclusion holds even when we account for publication selection bias and potential misspecifications in the literature.

The studies estimating the effect of bank competition on financial stability differ greatly in terms of the data and methodology used. We account for 35 aspects of studies and estimates, including the length of the sample, regional coverage, the definitions of key variables, the inclusion of controls, the estimation methodology, and publication characteristics (such as the number of citations of the study and the impact factor of the journal). We explore how these aspects affect the reported estimates, and use Bayesian model averaging (BMA; Raftery et al., 1997) to address model uncertainty. BMA is especially useful in meta-analysis, because for many study aspects there is no theory telling us how they should influence the results. Our findings indicate that researchers' choices concerning the data used, the definitions of key variables, and the estimation methodology affect the reported estimates systematically. We also find that highly cited studies published in good journals tend to report larger estimates of the competition-stability nexus. Finally, using all the estimates we construct a synthetic study, for which we select the methodology and publication aspects that we prefer (such as control for endogeneity and the maximum number of citations). The resulting estimate of the competition-stability nexus is very small.

The paper is organized as follows. Section 4.2 briefly discusses the related literature on the topic and explains how the effect of bank competition on financial stability is estimated. Section 4.3 explains how we collect the estimates and re-compute them to a common metric (partial correlation coefficients). Section 4.4 tests for the presence of publication bias. Section 4.5 describes the sources of heterogeneity in the literature and provides estimates of the competition-stability nexus conditional on our definition of best practice. In Section 4.6 we perform robustness checks using, among other things, alternative priors for BMA and alternative weights. The last section concludes. Appendix A presents diagnostics of the BMA exercise; the online appendix at http://meta-analysis.cz/competition includes an extensive robustness check using a more homogeneous subsample of estimates, additional results, and also lists the studies included in the meta-analysis.

4.2 Estimating the Effect of Bank Competition on Financial Stability

The impact of bank competition on financial stability remains a controversial issue in the theoretical literature. Two opposing theories – the competition-stability hypothesis and the competition-fragility hypothesis – can be used to justify the conflicting results often found in empirical studies.

The competition-fragility hypothesis asserts that more competition among banks leads to instability of the financial system. Marcus (1984) and Keeley (1990) model theoretically the "charter value" proposition, where banks choose the risk level of their asset portfolios. In the setting of limited liability, bank owners, who are often given incentives to shift risks to depositors, tend to engage only in the upside part of the risk-taking process. In more competitive systems, this behavior places substantial emphasis on profits: banks have higher incentives to take on excessive risks, which leads to higher instability of the system in general. In addition, in competitive systems the incentives of banks to properly screen borrowers are reduced, which again contributes to system fragility (Allen and Gale, 2000; Allen and Gale, 2004; Boot and Thakor, 1993). Conversely, when entry barriers are in place and competition in the sector is limited, banks have better profit opportunities and larger capital cushions and, therefore, are not prone to taking aggressive risks. In this framework highly concentrated banking systems contribute to overall financial stability (Boot and Greenbaum, 1993; Hellman, Murdoch, and Stiglitz, 2000; Matutes and Vives, 2000).

The competition-stability hypothesis, on the other hand, proposes that more competitive banking systems imply less fragility of the financial system. Specifically, Boyd and De Nicolo (2005) show that lower client rates facilitate lending as they reduce entrepreneurs' cost of borrowing. Lower costs of borrowing raise the chance of investment success, which, in turn, lowers banks' credit portfolio risk and leads to increased stability within the sector. Some theoretical studies reveal that banks in uncompetitive systems are more likely to originate risky loans, which pave the way to systemic vulnerabilities (Caminal and Matutes, 2002). Similarly, Mishkin (1999) stresses that, in concentrated systems, regulators are prone to implement too-big-to-fail policies that encourage risk-taking behavior by banks.

Overall, it appears that empirical studies conducted for individual countries do not find conclusive evidence for either the stability-enhancing or the stability-deteriorating view of competition (Fungacova and Weill, 2009; Fernandez and Garza-Garciab, 2012; Liu and Wilson, 2013). Some of the cross-country literature shows that more competitive banking systems are less likely to experience a systemic banking crisis (Beck et al., 2006a; Schaeck et al., 2009). In contrast, other studies (Yeyati and Micco, 2007; Uhde and Heimeshoff, 2009; Boyd et al., 2006) reveal that in more competitive systems bank failures tend to be more frequent. Further research also provides evidence that in more concentrated systems banks have higher capital ratios, which offsets the possibly stronger risk-taking behavior on their part (Berger et al., 2009; Schaeck and Cihak, 2012).

In this meta-analysis we focus on variants of the following model used in the literature to examine the effect of bank competition on stability:

$$Stability_{it} = \alpha + \beta \cdot Competition\ Measure_{it} + \sum_{k=1}^{N} \gamma_{kit} X_{kit} + e_{it}, \qquad (4.1)$$

where i is a bank index and t a time index and X is a set of control variables, both bank-specific and country-specific. Measures of stability and competition tend to vary across individual studies, as we will discuss later in this section (the various estimation methods used by researchers will be discussed in Section 4.5). We are interested in the coefficient β; positive estimates of the coefficient imply a positive effect of bank competition on financial stability, and vice versa.

Bank stability is often measured in an indirect way: that is, by considering individual or systemic banking distress, effectively the negative of stability. In this spirit, the non-performing loan (NPL) ratio is often used as a fragility indicator. Nevertheless, the NPL ratio only covers credit risk and cannot be directly linked to the likelihood of bank failure (Beck, 2008). Another measure of individual bank distress extensively used in the literature is the Z-score (e.g. Boyd and Runkle, 1993; Lepetit et al., 2008; Laeven and Levine, 2009; Cihak and Hesse, 2010). This measure indicates how many standard deviations in return on assets a bank is away from insolvency and, by extension, from the likelihood of failure. The Z-score is calculated as follows:

$$Z_{it} = \frac{ROA_{it} + {}^{E_{it}}/_{TA_{it}}}{\sigma_{ROA_{it}}}, \tag{4.2}$$

where ROA is the rate of return on assets, E/TA is the ratio of equity to total assets, and σ_{ROA} is the standard deviation of the return on assets. Bank profitability, measured by ROA and ROE (return on equity), profit volatility, approximated by ROA and ROE volatility, and bank capitalization, expressed by the capital adequacy ratio (CAR) or the ratio of equity to total bank assets, are additional measures of individual bank distress frequently used in the literature. Moreover, some studies (e.g. Beck, Demirguc-Kunt, and Levine, 2006a,b) model fragility in the banking sector by means of systemic banking crisis dummies. Other studies (such as Fungacova and Weill, 2009) apply individual bank failure dummies or measures of a bank's distance-to-default to proxy financial stability.

Concerning the proxies for competition, the Lerner index is one of the indicators frequently employed in the literature. This index quantifies the price power capacity of a bank by expressing the difference between price and marginal cost as a percentage of the price:

$$Lerner_{it} = \frac{\left(P_{TA_{it}} - MC_{TA_{it}}\right)}{P_{TA_{it}}}, \tag{4.3}$$

where $P_{TA_{it}}$ is the price of total assets, expressed in practice by total revenues to total bank assets, and $MC_{TA_{it}}$ is the marginal cost of total assets for bank i. The index thus takes values between 0 and 1, with the values of 0 and 1 reached only in the case of perfect competition and under pure monopoly, respectively. Alternatively, the degree of competition in the banking sector can be measured by the so-called H-statistic, introduced by Panzar and Rosse (1987). The H-statistic measures competition by summing the elasticities of a bank's revenue with respect to its input prices. Another competition measure, the Boone (2008) indicator, applied by Schaeck and Cihak (2012), for example, expresses the effect of competition on the performance of efficient banks and offers an organization-based explanation for how competition can improve stability.

In addition, concentration ratios were originally used as bank competition proxies: for instance, the Herfindahl-Hirschman index and the C3 concentration ratio, which indicates the share of the three largest banks' assets in the total assets of the country's banking system. Nevertheless, some studies (e.g. Claessens and Laeven, 2004) have shown that bank concentration is not an adequate indicator of the competitive nature of the system, as concentration and competition highlight different banking sector characteristics. In the spirit of better erring on the side of inclusion in meta-analysis (Stanley, 2001), we also collect estimates that measure competition by the inverse of concentration, and conduct a robustness check where we exclude these estimates.

4.3 The Data Set of Competition-Stability Estimates

The first step in any meta-analysis is to collect estimates from primary studies. We search for studies relevant to our meta-analysis using the Google Scholar and RePEc search engines and the following combinations of keywords: "competition" and "stability," "competition" and "fragility," "concentration" and "stability," and "concentration" and "fragility." We collect both published and unpublished studies, and try to include as many papers as possible. Since we need standard errors of the estimates to be able to use up-to-date meta-analysis methods, we have to omit studies that do not report statistics from which standard errors can be computed. In the end, we are left with 31 studies, which report 598 estimates; the oldest study in our sample was published in 2006. We also collect 35 variables reflecting the context in which researchers obtain their estimates. Our data collection strategy, as well as all other aspects of this meta-analysis, conform to the Meta-Analysis of Economics Research Reporting Guidelines (Stanley et al., 2013).

Given the broad scope of the measures used in the literature to proxy for both bank competition and financial stability, it is imperative that we recompute the individual estimates to a common metric. Because some stability proxies measure financial fragility and some competition proxies investigate how uncompetitive the market is (for example, larger values of the Lerner index imply a less competitive nature of the system), we adjust the signs of the collected estimates so that they directly reflect the relationship between competition and stability. After this adjustment the collected estimates imply either that higher competition increases bank stability or that higher competition decreases bank stability, and they could be compared with each other if all studies used the same units of measurement.

Due to the inconsistency in the use of measurement units of regression variables in the literature, we transform the reported estimates into partial correlation coefficients (PCCs). The PCC is a unitless measure of the strength and direction of the association between two variables, competition and stability in our case, while holding other variables constant (Stanley and Doucouliagos, 2012). The PCCs enable us to directly compare estimates reported in different studies. This technique is widely used in meta-analysis research nowadays; a related application can be found, for example, in Valickova et al. (2014).

The partial correlation coefficient is calculated according to the following formula:

$$PCC = \frac{t}{\sqrt{t^2 + df}}, \tag{4.4}$$

where t is the t-statistic of the reported coefficient and df denotes the number of degrees of freedom used for the estimation. The corresponding standard errors of the PCC are calculated as follows:

$$SEPCC = \sqrt{\frac{(1 - PCC^2)}{df}}. \tag{4.5}$$

Moreover, if the primary study assumes a quadratic relationship between competition and stability and thus reports two coefficients associated with the measure of competition, the overall impact on stability needs to be linearized using the following formula:

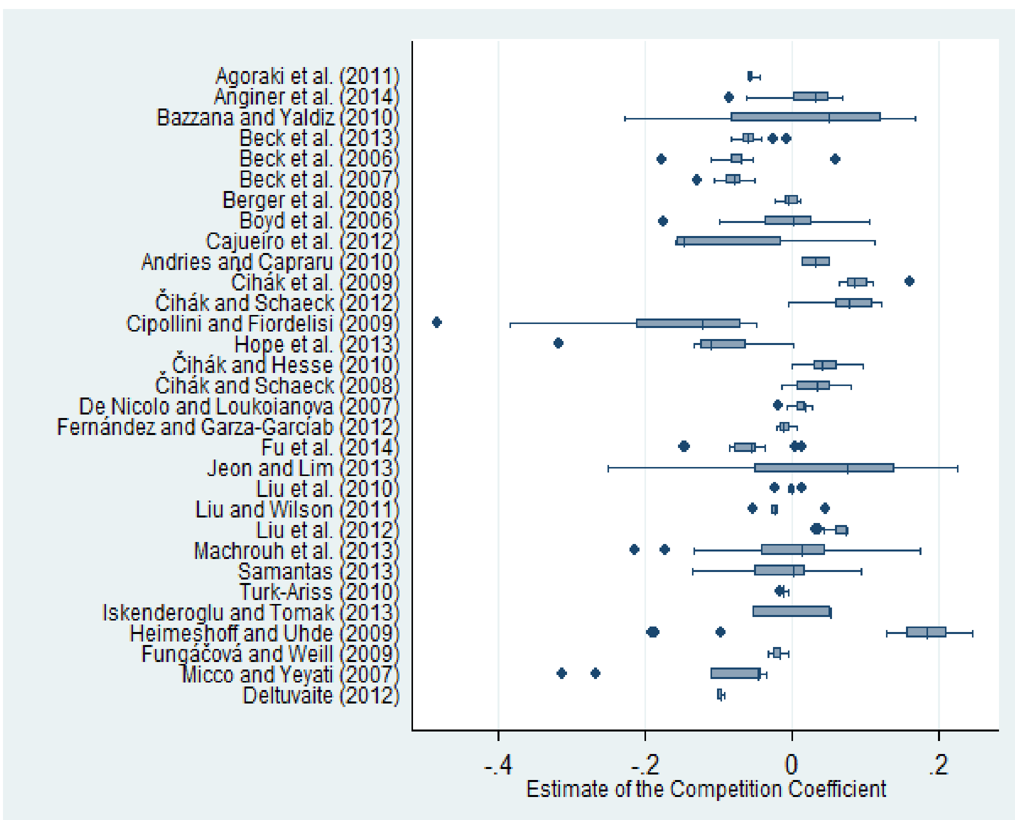
$$\beta = \widehat{\beta_1} + 2\widehat{\beta_2}\bar{x} \qquad SE(\beta) = \sqrt{SE\left(\widehat{\beta_1}\right)^2 + 4SE\left(\widehat{\beta_2}\right)^2 \bar{x}^2}, \tag{4.6}$$

where $\widehat{\beta_1}$ is the estimate of the competition coefficient for the linear term, $\widehat{\beta_2}$ is the estimate of the competition coefficient for the quadratic term, $\bar{x}$ is the sample mean of the competition measure in the study, $SE\left(\widehat{\beta_1}\right)$ is the standard error of the reported coefficient for the linear term, and $SE\left(\widehat{\beta_2}\right)$ is the standard error of the reported coefficient for the quadratic term. The covariance term is omitted from the $SE(\beta)$ formula due to the unavailability of the original data. The resulting coefficient of bank competition after linearization is subsequently transformed into the PCC in line with equations (4) and (5).

Figure 4.2 depicts the within- and between-study dispersion in the partial correlation coefficients of the competition-stability estimates reported in the 31 studies that we examine in this meta-analysis. It is apparent that the literature is highly heterogeneous, both between and within studies. Meta-analysis will help us to formally trace the sources of this heterogeneity.

Table 4.1 shows summary statistics for all the estimates and for two subsamples of the estimates that evaluate the effect for developed and developing countries. The left-hand part of the table shows unweighted means, while the right-hand part shows means weighted by the inverse of the number of estimates reported per study. All the means are close to zero, indicating little interplay between competition and stability. The estimates for developed countries are slightly larger than those for developing and transition countries. (The overall mean is slightly negative, while the means for both developing and developed countries are positive, which suggests that studies that mix these two groups tend to find smaller estimates of the effect.) Nevertheless, all these values are negligible and would be classified as implying no effect according to the guidelines for the interpretation of partial correlation coefficients in economics (Doucouliagos, 2011).

Figure 4.2: Variability in the estimated competition coefficients across individual studies



Notes: The figure shows a box plot of the PCCs of the competition coefficient estimates (the PCCs of the β estimates from equation (4.1)) reported in individual studies. Full references for the studies included in the meta-analysis are available in the online appendix.

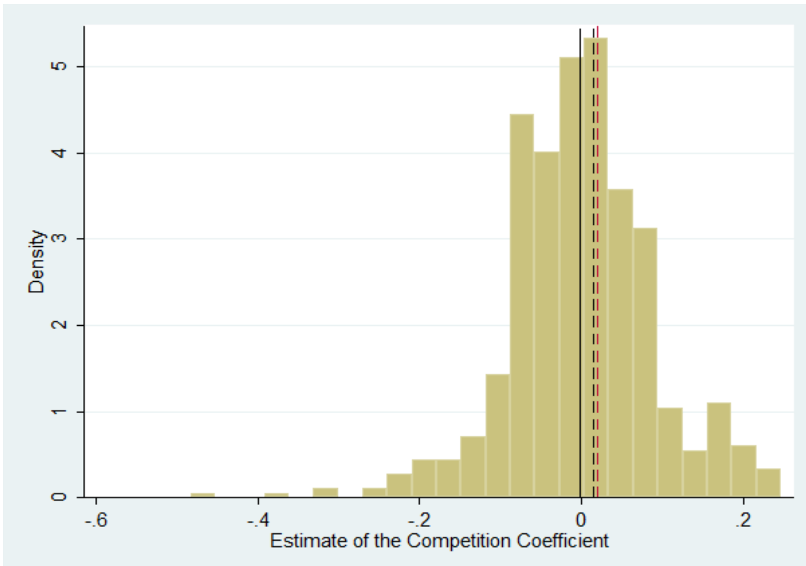
Table 4.1: Estimates of the competition effect for different country groups

	Unweighted			*Weighted*			No. of estimates
	Mean	95% Conf. Int.		Mean	95% Conf. Int.		
All	-0.001	-0.025	0.023	-0.012	-0.035	0.011	598
Developed	0.020	-0.032	0.073	0.011	-0.030	0.052	201
Developing and transition	0.001	-0.022	0.023	-0.019	-0.051	0.012	194

Notes: The table presents the mean PCCs of the competition coefficient estimates (the PCCs of the β estimates from equation (4.1)) over all countries and for selected country groups. The confidence intervals around the mean are constructed using standard errors clustered at the study level. In the right-hand part of the table the estimates are weighted by the inverse of the number of estimates reported per study.

Figure 4.3 depicts the distribution of the partial correlation coefficients of all the competition coefficient estimates. It appears that the PCCs are symmetrically distributed around zero with a mean of -0.0009, while the mean of the study-level medians is also close to zero and equals 0.0099. We also report the mean of the PCCs of the estimates that are reported in studies published in peer-reviewed journals, as opposed to those reported in unpublished manuscripts. In total, 21 of the 31 studies in our sample were published in peer-reviewed journals, yielding 376 estimates of the competition coefficient. The mean for published studies is 0.0116: it appears that journals tend to report slightly larger estimates of the competition coefficient compared to the grey literature.

Figure 4.3: Studies published in journals report slightly larger estimates



Notes: The figure shows the histogram of the PCCs of the competition coefficient estimates (the PCCs of the β estimates from equation (4.1)) reported in individual studies. The solid vertical line denotes the mean of all the PCCs. The dashed lines denote the mean of the median PCCs of the estimates from the studies and the mean of the PCCs of those estimates that are reported in studies published in peer-reviewed journals, respectively.

4.4 Testing for Publication Bias

Publication selection bias arises when an estimate's probability of being reported depends on its sign or statistical significance. Rosenthal (1979) refers to this phenomenon as the "file drawer problem," implying that researchers may hide estimates that are either insignificant or have a counterintuitive sign in their file drawers, and seek instead to obtain new estimates that would be easier to publish. A number of studies, e.g., by DeLong and Lang (1992), Card and Krueger (1995), and Ashenfelter et al. (1999), identify publication selection bias in empirical economics. In addition, Doucouliagos and Stanley (2013) conduct a survey of meta-analyses and find that most fields of empirical economics suffer from publication bias. The bias tends to inflate the mean estimates reported by empirical studies. For example, Doucouliagos and Stanley (2009) estimate that the adverse employment effect of minimum wage increases is seriously overstated in the published empirical literature. In our case there are opposing theories concerning the effect of competition on stability, so both positive and negative estimates are publishable, which might alleviate publication bias. In this section, we test for potential publication bias in the literature evaluating the competition-stability nexus before we proceed with the analysis of heterogeneity in the next section.

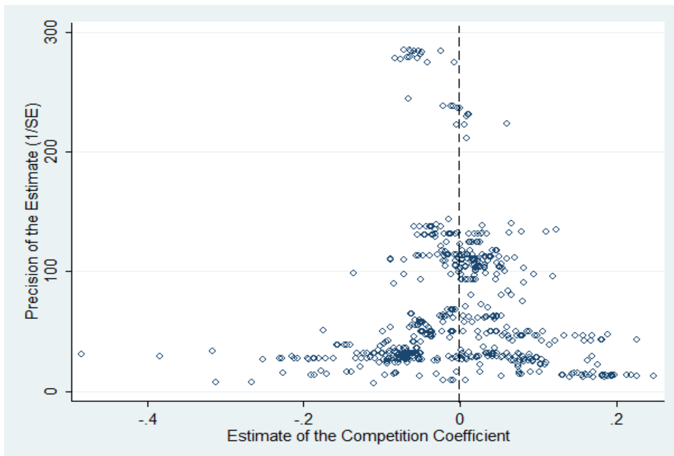
We start with visual tests for the presence of publication bias. The most commonly applied graphical test uses the so-called funnel plot (Egger et al., 1997), which depicts the magnitude of the estimated effect on the horizontal axis and precision (the inverse of the estimated standard error) on the vertical axis. The most precise estimates (located at the top of the funnel) should be close to the true underlying effect. With decreasing precision, the estimates get more dispersed; overall, they should form a symmetrical inverted funnel. If there is publication bias in the literature, the funnel is either asymmetrical due to the exclusion of estimates of a certain sign or size, or hollow due to the omission of insignificant estimates, or displays both these properties.

Figure 4.4A shows the funnel plot for the PCCs of all the competition coefficient estimates reported in the studies, while Figure 4.4B depicts the funnel plot for the median values of the PCCs of the estimates reported in individual studies. We observe that both funnels are relatively symmetrical, and the most precise estimates are close to the mean reported PCC of the estimates. Moreover, the funnels are not hollow, and even estimates with very little precision (and large p-values) at the bottom of both plots are reported. Therefore, we can infer that these funnel plots do not point to the presence of publication bias in the competition-stability literature, as opposed to the findings in most other fields in economics and finance (for example, Havranek and Irsova, 2011; Havranek and Irsova, 2012; Havranek et al., 2012).

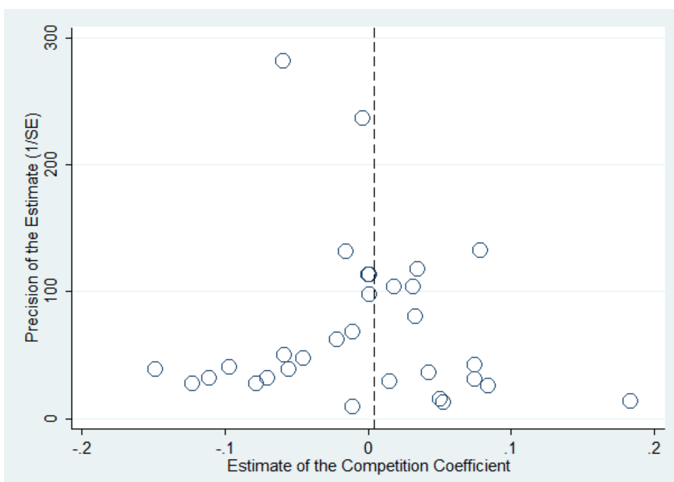
A more rigorous approach to testing for publication bias consists in funnel asymmetry tests. These tests explore the relationship between the collected coefficient estimates and their standard errors following the methodology suggested by Card and Krueger (1995). In the presence of publication selection, the reported estimates are correlated with their standard errors. For example, if negative estimates are omitted, a positive relationship appears between the reported coefficient estimates and their standard errors because of heteroskedasticity in the equation (Stanley, 2008).

Figure 4.4: Funnel plots do not suggest strong publication bias

A) All estimates



B) Median estimates from studies



Notes: In the absence of publication bias the funnel should be symmetrical around the most precise estimates of the competition coefficient (the PCC of the β estimate from equation (4.1)). The dashed vertical lines denote the mean of the PCCs of all the estimates in Figure 4.4A and the mean of the study-level medians reported in Figure 4.4B.

Similarly, researchers who prefer statistical significance need large estimates to offset large standard errors. Thus, we estimate the following equation:

$$PCC_i = \beta_0 + \beta_1 SE(PCC_i) + \varepsilon_i, \tag{4.7}$$

where PCC_i is the partial correlation coefficient of the competition coefficient estimate, $SE(PCC_i)$ is the standard error of the partial correlation coefficient, β_0 is the mean PCC corrected for the potential publication bias, β_1 measures the extent of publication bias, and ε_i is a disturbance term. Equation (4.7) is commonly called the funnel asymmetry test, as it follows from rotating the axes of the funnel plot and inverting the values on the new horizontal axis so that it now shows standard errors instead of precision.[1]

The results of the funnel asymmetry tests are presented in Table 4.2. The coefficient estimates in the upper part of the table result from fixed effects[2] estimation with standard errors clustered at the level of individual studies and from instrumental variable estimation (where the number of observations is used as an instrument for the standard error). Fixed effects control for method or other quality characteristics specific to individual studies. We also report results for the subsample of estimates reported in published studies to see whether they show different levels of publication selection bias. The bottom half of the table presents results from regressions weighted by the inverse of the number of estimates reported per study in order to diminish the effect of studies reporting many estimates. In all specifications in Table 4.2, both coefficient estimates are significant at least at the 5% level. A moderate negative publication bias is present, and the estimated size of the competition-stability effect beyond publication bias appears to be close to zero, especially for weighted results. For unweighted results we obtain small effect sizes according to the guidelines by Doucouliagos (2011) for partial correlations reported in the field of industrial organization.

The magnitude of the publication bias is slightly larger in published studies than in unpublished manuscripts, but the difference is not statistically significant. We consider it remarkable that the fixed effects and instrumental variable specifications yield very similar results. In meta-analysis it is important to check for endogeneity of the standard error, because very often it can happen that the meta-analyst cannot collect all relevant information on the methodology used in the primary studies. If the meta-analyst omits

[1] It is worth noting at this point that authors of primary studies do not directly report partial correlation coefficients; we compute the PCCs from the statistics the authors provide. Because the PCCs are nonlinear transformations of the original estimates and standard errors, a linear relation between estimates and standard errors does not translate into a linear relation between PCCs and SE(PCC). In consequence, our publication bias estimates might be biased downwards.

[2] We use the term „fixed effects" in the panel-data sense common in economics. In meta-analysis, though, the term „fixed effects" is frequently used to point at the assumption that the population effect is fixed (and does not vary randomly across studies, as opposed to random effects estimation).

an aspect of methodology that influences both the reported coefficients and their standard errors in the same direction, he or she will obtain biased estimates of the magnitude of the publication bias. Our results suggest that in the case of the competition-stability nexus endogeneity is not an important issue.

Table 4.2: Funnel asymmetry tests show moderate publication bias

Unweighted regressions	Fixed Effects	Fixed Effects_Published	Instrument	Instrument_Published
SE (publication bias)	-1.671**	-1.898**	-1.614***	-2.291***
Constant (effect beyond bias)	0.044**	0.073**	0.043***	0.086***
No. of estimates	598	376	598	376
No. of studies	31	21	31	21
Weighted regressions	Fixed Effects		Fixed Effects_Published	
SE (publication bias)	-1.568***		-1.636***	
Constant (effect beyond bias)	0.034***		0.044***	
No. of estimates	598		376	
No. of studies	31		21	

Notes: The table presents the results of the regression specified in equation (4.7). The standard errors of the regression parameters are clustered at the study level. Published = we only include published studies. Fixed Effects = we use study dummies. Instrument = we use the logarithm of the number of observations in equation (4.1) as an instrument for the standard error and employ study fixed effects. The regressions in the bottom half of the table are estimated by weighted least squares, where the inverse of the number of estimates reported per study is taken as the weight. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

Equation (4.7), however, suffers from heteroskedasticity, because the explanatory variable directly captures the variance of the response variable. To achieve efficiency, many meta-analysis applications divide equation (4.7) by the corresponding standard error, i.e., they multiply the equation by the precision of the estimates. This specification places more emphasis on precise results. Dividing equation (4.7) by the corresponding SE of the PCC, we obtain the following equation:

$$t_i = \beta_1 + \beta_0(1/SE(PCC_i)) + \mu_i, \tag{4.8}$$

where β_0 is the mean PCC of the coefficient estimate corrected for the potential publication bias, β_1 measures the extent of publication bias, and t_i is the corresponding t-statistic. Table 4.3 below presents results from the heteroskedasticity-corrected equation (4.8).

Table 4.3: Heteroskedasticity-corrected funnel asymmetry tests confirm the presence of publication bias

Weighted by precision	Fixed Effects	Fixed Effects_Published	Instrument	Instrument_Published
1/SE (effect beyond bias)	0.005	0.065	0.019**	0.053***
Constant (publication bias)	-0.757	-4.000*	-1.706**	-3.344***
No. of estimates	598	376	598	376
No. of studies	31	21	31	21
Weighted by precision and no. of observations	**Fixed Effects**		**Fixed Effects_Published**	
1/SE (effect beyond bias)	0.013		0.056**	
Constant (publication bias)	-1.539**		-4.339**	
No. of estimates	598		376	
No. of studies	31		21	

Notes: The table presents the results of the regression specified in equation (4.8). The standard errors of the regression parameters are clustered at the study level. Published = we only include published studies. Fixed Effects = we use study dummies. Instrument = we use the logarithm of the number of observations in equation (4.1) as an instrument for the standard error and employ study fixed effects. The regressions in the bottom half of the table are additionally weighted by the inverse of the number of estimates reported per study. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

We can observe from Table 4.3 that publication bias is not equally strong across all specifications, in contrast to Table 4.2. Moreover, the true underlying effect beyond publication bias is only significant when equation (4.8) is estimated by means of instrumental variables or by fixed effects for the subsample of published studies. Table 3 confirms that the competition-stability effect beyond publication bias is indeed close to zero, as no estimate surpasses the threshold defined by Doucouliagos (2011) to denote at least a weak effect. The story changes for publication bias, which now seems to be much stronger in published studies than in unpublished manuscripts, which would suggest that journal editors or referees prefer papers that show results consistent with the competition-fragility hypothesis.

For evaluation of the extent of publication bias, Doucouliagos and Stanley (2013) provide guidelines for the value of the constant in the funnel asymmetry test specified by equation (4.8). They identify that the literature suffers from substantial selectivity if $\hat{\beta}_1$ from equation (4.8) is statistically significant and, at the same time, $1 \leq |\hat{\beta}_1| \leq 2$.

Both conditions hold for the value of the constant estimated by fixed effects and weighted by the inverse of the number of observations, as well as for the constant in regressions estimated by the instrumental variable method. The values of the coefficient estimated in Table 4.3 for published studies are even larger than 2, which would suggest severe publication bias according to the guidelines by Doucouliagos and Stanley (2013). Nevertheless, we believe the overall evidence points to only moderate publication bias, because the corrected estimates of the competition-stability nexus are close to the simple mean of all the estimates uncorrected for publication bias.

4.5 Why the Reported Coefficients Vary

4.5.1 Variable Description and Methodology

In this section we add the characteristics of the studies and estimates into equation (4.7) to explore what drives the heterogeneity in the literature. We do not weight the resulting equation by precision as is the case in equation (4.8): weighting by the estimates' precision introduces artificial variation into variables that are defined at the study level (for example, the impact factor of the study) or that tend to vary little within studies (for example, sample size). In contrast, we weight the regressions by the inverse of the number of estimates reported per study to give the same importance to each study in our data set. In the next section we also perform a robustness check for regressions not weighted by the number of estimates per study.

Table 4.4 describes all the variables that we collect from the primary studies. For each variable the table also shows the mean, the standard deviation, and the mean weighted by the inverse of the number of estimates reported per study. For ease of exposition we divide the collected variables into eight groups.

Table 4.4: Overview and summary statistics of regression variables

Variable	*Description*	*Mean*	*SD*	*WM*
Data characteristics				
Competition coefficient	The coefficient capturing the effect of bank competition on financial stability (recomputed to the partial correlation coeff.)	-0.001	0.090	-0.012
SEPCC	The estimated standard error of the competition coefficient	0.027	0.022	0.029
Samplesize	The logarithm of the number of cross-sectional units used in the competition-stability regression	7.835	1.615	7.760

T	The logarithm of the number of time periods (years)	2.224	0.743	2.264
sampleyear	The mean year of the sample period on which the competition-stability regression is estimated (base: 1992,5)	8.889	4.328	9.340
Countries examined				
developed	equals 1 if the researcher only examines OECD countries	0.336	0.473	0.366
developing and transition	equals 1 if the researcher only examines non-OECD countries	0.324	0.469	0.376
Reference case: mixed	equals 1 if the researcher examines both OECD and non-OECD countries (omitted category)	0.339	0.474	0.258
Design of the analysis				
quadratic	equals 1 if the square of the competition coefficient is included in the regression	0.119	0.324	0.217
endogeneity	equals 1 if the estimation method accounts for endogeneity	0.635	0.482	0.713
macro	equals 1 if the competition-stability regression is estimated using country-level data	0.256	0.437	0.133
averaged	equals 1 if the competition-stability regression uses variables in the form of country-level averages over banks	0.120	0.326	0.085
Treatment of stability				
dummies	equals 1 if stability is measured by a crisis dummy or a bank failure dummy	0.142	0.349	0.129
NPL	equals 1 if stability is measured by non-performing loans as a share of total loans	0.050	0.218	0.095
Zscore	equals 1 if stability is measured by the Z-score statistic	0.452	0.498	0.537
profit_volat	equals 1 if stability is measured by ROA volatility or ROE volatility	0.075	0.264	0.039
profitability	equals 1 if stability is measured by ROA or ROE	0.043	0.204	0.045
capitalization	equals 1 if stability is measured by the capital adequacy ratio (CAR) or the equity-total assets ratio	0.069	0.253	0.040
DtoD	equals 1 if stability is measured by Logistic R2 Merton's distance-to-default or probability of bankruptcy	0.065	0.247	0.047
Reference: other_stability	equals 1 if stability is measured by a less frequently used method (omitted category)	0.104	0.305	0.069
Treatment of competition				
Hstatistic	equals 1 if competition is measured by the H-statistic	0.090	0.287	0.098
Boone	equals 1 if competition is measured by the Boone indicator	0.075	0.264	0.108
Concentration	equals 1 if competition is measured by concentration measures C3 or C5	0.157	0.364	0.147

Lerner	equals 1 if competition is measured by the Lerner index	0.360	0.480	0.414
HHI	equals 1 if competition is measured by the Herfindahl-Hirschman index	0.266	0.442	0.197
Reference: other_competition	equals 1 if competition is measured by a less frequently used method (omitted category)	0.052	0.222	0.037
Estimation methods				
Logit	equals 1 if the logit or probit model is used in the estimation of the competition-stability regression	0.172	0.378	0.161
OLS	equals 1 if OLS is used in the estimation	0.137	0.344	0.115
FE	equals 1 if fixed effects are used in the estimation	0.229	0.421	0.136
RE	equals 1 if random effects are used in the estimation	0.067	0.250	0.043
GMM	equals 1 if GMM is used in the estimation	0.182	0.386	0.309
TSLS	equals 1 if two-stage least squares are used in the estimation	0.149	0.356	0.110
Reference: other_method	equals 1 if a less frequently used method is employed (omitted category)	0.064	0.244	0.126
Control variables				
regulation	equals 1 if regulatory/supervisory variables are included in the competition-stability regression	0.239	0.427	0.282
ownership	equals 1 if bank ownership is controlled for in the competition-stability regression	0.166	0.372	0.271
global	equals 1 if macroeconomic variables are included in the competition-stability regression	0.794	0.405	0.764
Publication characteristics				
citations	The logarithm of the number of Google Scholar citations normalized by the difference between 2015 and the year the study first appeared in Google Scholar (collected in July 2014)	2.045	1.222	1.790
firstpub	The year when the study first appeared in Google Scholar (base: 2003)	6.453	2.979	6.677
IFrecursive	The recursive impact factor of the outlet from RePEc (collected in July 2014)	0.243	0.210	0.205
reviewed_journal	equals 1 if the study is published in a peer-reviewed journal	0.629	0.484	0.677

Notes: SD = standard deviation. WM = mean weighted by the inverse of the number of estimates reported per study. All variables except for citations and the impact factor are collected from studies estimating the competition coefficient from equation (4.1). The search for studies was terminated on July 1, 2014, and the list of studies included is available in the online appendix. Citations are collected from Google Scholar and the impact factor from RePEc.

Group 1 – Data characteristics: We control for the number of cross-sectional units and time periods used to estimate the competition coefficient in equation (4.1). Ceteris paribus, we intend to place more weight on studies that use larger samples to minimize the potential small-sample bias, and it is therefore important to check whether such studies yield systematically different results. Although being correlated with the standard error, the number of cross-sectional units and time periods bring additional information to our model, and the results can suggest whether the bias identified in the previous section is due to publication selection or small samples. Moreover, we control for the age of the data used in the primary studies by including the variable sampleyear, which represents the midpoint of the data period used by researchers. Although Figure 4.1 suggests no significant time trend in the estimates of the competition-stability nexus, perhaps the literature can be shown to converge to a particular result when data and method heterogeneity in primary studies is controlled for.

Group 2 – Countries examined: We account for potential cross-country heterogeneity by including dummies for developed (OECD member) countries and developing and transition (non-OECD) countries. The characteristics of the banking sector (measured, for example, by the credit-to-GDP ratio) differ greatly between developed and developing countries, which can affect the results of primary studies. In our sample, 34% of all the collected estimates are obtained using a sample of developed countries, while 32% of estimates are extracted from studies focusing on developing and transition countries. The reference case for this group of dummy variables is estimation that mixes these two groups.

Group 3 – Design of the analysis: We control for the general design of the studies in our sample, captured by the variables quadratic, endogeneity, macro, and averaged. First, the dummy variable quadratic controls for the inclusion of the square of the competition measure in the regressions. In total, 12% of the estimates in our sample have to be linearized because researchers test for possible nonlinear relationships between bank competition and stability (in the next section we will discuss how our results change when we conduct separate meta-analyses of the linear and quadratic term). The dummy variable endogeneity reflects whether individual studies account for potential endogeneity in their analysis, either by employing estimation methods with instruments or by using lagged values of bank competition in equation (4.1). Later we also include dummy variables for estimation methods, some of which control for endogeneity. Nevertheless, the correlations between these variables and endogeneity do not exceed 0.42. Next, the dummy variable macro assigns the value 1 to an estimate if the estimate is calculated using data constructed at the aggregate level, as opposed to studies using bank-level data. The motivation behind this control emerges from the narrative literature survey by Beck (2008), who notes that bank-level studies tend to obtain smaller estimates of the competition effect, perhaps because they fail to capture spillovers to other sectors of the economy. Finally, the dummy variable averaged assigns the value 1 to an estimate if the regressors in equation (4.1) in the original study are constructed as country-level averages over banks, even though the data are technically bank-level. This simplification decreases the variance available for the estimation, and might lead to aggregation bias. 12% of the collected competition effect estimates are extracted from studies that use explanatory variables in the form of averages over the observed period in their regressions (e.g. Berger et al., 2009; Levy Yeyati and Micco, 2007).

Group 4 – Treatment of stability: Due to the large diversity of the approaches to measuring financial stability in the literature, it is possible that a portion of the variation in the competition coefficient estimates is due to a different definition of stability. We distinguish between the seven most common approaches. Some researchers use dummy variables representing either the outbreak of a systemic banking crisis or a bank failure (e.g. Beck et al., 2006 a,b; Fungacova and Weill, 2009). Popular methods for measuring individual bank stability include the ratio of non-performing loans to total bank loans, the Z-score, an aggregate measure of bank stability, fluctuations in the return on assets (ROA) or the return on equity (ROE) as indicators of bank profit volatility, ROA or ROE as measures of bank profitability, measures of capitalization, the capital adequacy ratio or equity to assets ratio, and measures of distance to default. The reference case for this group of dummy variables accounts for additional approaches to quantifying financial stability that are used less frequently, such as the ratio of loan loss reserves to total assets, the ratio of deposits to total bank liabilities, or the shareholder value ratio expressed as economic value added over the capital invested by shareholders.

Group 5 – Treatment of competition: Similarly to the indicators of stability, there is large diversity in the approaches to quantifying competition within the banking sector. We control for the five most commonly used measures. We include Panzar and Rosse's (1987) H-statistic and Boone's (2008) index. Quite frequently, measures of market structure are applied to assess the intensity of competition in the sector; concentration ratios are one type of such measures. For 36% of the estimates in our sample, competition is measured via the Lerner index. Herfindahl-Hirschman indices (HHI) are another example of market structure measures extensively used in the literature. Overall, market structure measures are used to compute 42% of the estimates in the sample (e.g., by Beck et al., 2006 a,b; Berger et al., 2009; Boyd et al., 2006; Cipollini and Fiordelisi, 2009). We decide to include the estimates arising from the use of these market structure measures in our analysis despite the recent assertions in the literature that concentration is not a suitable proxy for a lack of competition (e.g. Claessens and Laeven, 2004; Bikker, 2004). As a robustness check in the online appendix, we estimate the impact of competition on stability after excluding these potentially misspecified estimates from our sample. The reference case for this group of dummy variables covers alternative and infrequently used proxies of market competition, e.g. the extent of entry barriers into banking and percentage of applications to enter banking denied (Anginer et al., 2014), market pressure dummy (Jeon and Lim, 2013), and market power calculated as the difference between total revenues and total costs over total bank revenues (Bazzana and Yaldiz, 2010).

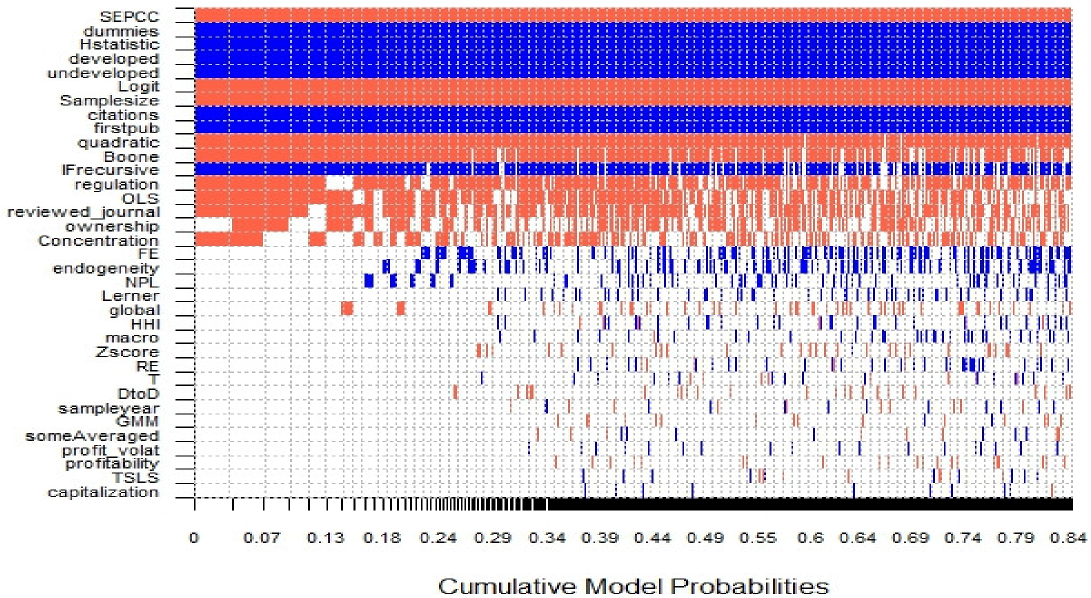
Group 6 – Estimation methods: We control for six different estimation methods in our analysis: logit, OLS, FE, RE, GMM, and TSLS. Based on the findings of many previous meta-analyses, we assume that different methods might systematically affect the resulting estimates of the competition coefficient. As to the frequency of use, 17% of estimates originate from logit estimation, 14% from OLS, 23% from fixed effects, 7% from random effects, 18% from GMM, and 15% from TSLS. In our data set the variable reflecting the use of logit is not identical to the variable that captures the use of dummy variables on the left-hand side, because some of the studies that employ dummy variables use linear estimation techniques. Moreover, other studies, e.g. Cipollini and Fiordelisi (2009), incorporate either random effects or GMM estimators into logit and probit models, which we in turn classify into the RE or GMM categories. The reference case for this group represents sporadically used estimation methods in the literature, for example Tobit regressions (Fu et al., 2014; Turk Ariss, 2010), generalized least squares (Liu et al., 2012), and weighted least squares (Levy Yeyati and Micco, 2007).

Group 7 – Control variables: The most commonly used controls in the estimation of the competition-stability relationship in equation (4.1) are regulatory and supervisory variables such as capital stringency, supervisory power, the investor protection index, economic and banking freedom, the share of market entry restrictions or governance (e.g. Cihak et al., 2009; Beck et al., 2006 a,b; Beck et al., 2013; Anginer et al., 2014; Agoraki et al., 2011), ownership controls, i.e., foreign and state bank ownership (e.g. Bazzana and Yaldiz, 2010; Berger et al., 2009; De Nicolò and Loukoianova, 2007) and macroeconomic variables defined at the country level, such as GDP growth and real interest rate (Agoraki et al., 2011), trade as share of GDP, private credit as GDP share (Anginer et al., 2014), terms of trade, inflation, M2 share of reserves (Beck et al., 2006a,b) or exchange rate (Boyd et al., 2006). Including macroeconomic variables in regressions in original studies aims to proxy the economic climate (e.g. Beck et al., 2006a,b). Specifically, short-term real interest rates reflect the banks' cost of funds that may impact bank profitability via default rates. Similarly, foreign exchange risk, measured by exchange rate depreciation and the ratio of M2 to foreign exchange reserves, captures a bank's vulnerability to abrupt capital outflows. Moreover, credit growth controls for potential large credit expansion that can lead to asset price bubbles and upon their burst to a subsequent crisis in the sector. Regulatory and supervisory controls are used in 24% of regressions, ownership variables in 17% of estimations, and macroeconomic variables are used as controls in 79% of regressions.

Group 8 – Publication characteristics: We control for study quality by including the number of citations. This control reflects additional aspects of study quality not captured by other variables described above. Although the number of citations is an imperfect control for quality (and may be also influenced by the results of the study), we find it appealing to place more weight on highly-cited studies, other things (especially data and methodology) being equal. To control for the potential time trend in the literature, we add the year when each study first appeared in Google Scholar. Another control we use to account for study quality is the recursive RePEc impact factor of the outlet. Finally, in order to evaluate whether studies published in peer-reviewed journals report systematically different estimates in comparison to unpublished studies after we control for data and methodology, we include a corresponding dummy variable.

We would like to run a regression with the PCC of the estimates of the competition coefficient as the dependent variable and all the variables introduced above as explanatory variables. Nevertheless, including all of the variables at the same time is infeasible as we would probably obtain many redundant regressors in the specification. With such a large number of explanatory variables, we initially do not know which ones should be excluded from the model. An ideal approach would be to run regressions with different subsets of independent variables to ensure that our results are robust: to this end, we employ Bayesian model averaging (BMA) to resolve the model uncertainty problem, an issue that is inevitable in meta-regression analysis. BMA runs many regressions with different subsets of all the 2^{35} possible combinations of explanatory variables (we have 35 regressors at our disposal). To make the estimation feasible, we employ the Monte Carlo Markov Chain algorithm to go through the most promising of the potential models (we use the bms package for R developed by Feldkircher and Zeugner, 2009). BMA gives each model a weight, which can be thought of as an analogy of the adjusted R-squared, to capture the model's fit. Finally, BMA reports weighted averages from the models for posterior mean values of regression parameters and posterior standard deviations, which capture the distribution of regression parameters across individual models. Moreover, a posterior inclusion probability (PIP) is reported for each variable to show the probability with which the variable is included in the true model. Raftery et al. (1997) and Eicher et al. (2011) provide further details on BMA in general. Detailed diagnostics of our BMA exercise can be found in Appendix A.

Figure 4.5: Bayesian model averaging – model inclusion



Notes: The response variable is the PCC of the estimate of the competition coefficient (the PCC of the β estimate from equation (4.1)). All regressions are weighted by the inverse of the number of estimates reported per study. Columns denote individual models; the variables are sorted by posterior inclusion probability in descending order. Blue color (darker in grayscale) = the variable is included and the estimated sign is positive. Red color (lighter in grayscale) = the variable is included and the estimated sign is negative. No color = the variable is not included in the model. The horizontal axis measures the cumulative posterior model probabilities. Numerical results of the BMA estimation are reported in Table 4.5. A detailed description of all the variables is available in Table 4.4.

4.5.2 Results

Figure 4.5 shows the results of the BMA exercise. The columns in the figure denote the individual regression models, while their width indicates the models' posterior probabilities. The variables are sorted by their PIP in descending order. If the sign of a variable's regression coefficient is positive, it is denoted by blue color (darker in grayscale). Conversely, if the sign of a variable's coefficient is negative, it is colored in red. Where a variable is excluded from a model, the corresponding cell is left blank. The horizontal axis measures the cumulative model probabilities: the models that are the most successful in explaining the heterogeneity in the estimates of the competition effect are on the left, and we can see that they include less than a half of all the variables.

The numerical results of the BMA exercise are reported in Table 4.5. On the right-hand side of the table we also report the results of OLS estimation with standard errors clustered at the level of individual studies. From this "frequentist check" we exclude the variables that prove to be irrelevant for the explanation of the variability in the literature (that is, have PIP lower than 0.5). The OLS regression thus includes 15 variables identified by BMA to help explain the variation in the reported competition effects. Overall, OLS with clustered standard errors yields results consistent with BMA for variables with high inclusion probabilities. The signs of the variables' regression parameters are the same and the size of their parameter estimates is similar as well. Therefore, we can conclude that our results are robust to error-clustering, as BMA by definition does not cluster standard errors in the estimation. Eicher et al. (2011) provide a framework for the identification of the strength of the variables' effect in BMA. The effect of a variable is considered weak if the corresponding PIP is between 0.5 and 0.75, substantial if it is between 0.75 and 0.95, strong if it is between 0.95 and 0.99, and decisive if it exceeds 0.99.

Table 4.5: Explaining heterogeneity in the estimates of the competition coefficient

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Competition effect	Post. Mean	Post. SD	PIP	Coef.	Robust Std. Err.	P-value
Data characteristics						
SEPCC	-1.7883	0.2046	1.0000	-1.1940	0.6511	0.067
Samplesize	-0.0367	0.0035	1.0000	-0.0240	0.0089	0.007
T	0.0005	0.0039	0.0517			
sampleyear	0.0000	0.0005	0.0456			
Countries examined						
developed	0.2015	0.0219	1.0000	0.1761	0.0295	0.000
developing and transition	0.1072	0.0169	1.0000	0.0985	0.0262	0.000
Design of the analysis						
quadratic	-0.0533	0.0124	0.9971	-0.0441	0.0128	0.001
endogeneity	0.0100	0.0212	0.2371			
macro	0.0025	0.0124	0.0699			
someAveraged	-0.0004	0.0047	0.0397			
Treatment of stability						
dummies	0.2115	0.0282	1.0000	0.1841	0.0194	0.000
NPL	0.0020	0.0060	0.1323			
Zscore	-0.0005	0.0027	0.0630			
profit_volat	0.0006	0.0051	0.0371			
profitability	-0.0003	0.0030	0.0354			
capitalization	0.0001	0.0029	0.0271			
DtoD	-0.0013	0.0078	0.0504			
Treatment of competition						
Hstatistic	0.1083	0.0217	1.0000	0.1140	0.0181	0.000

Boone	-0.0709	0.0313	0.8974	-0.0583	0.0225	0.010
Concentration	-0.0185	0.0226	0.4742			
Lerner	0.0036	0.0130	0.1217			
HHI	0.0023	0.0108	0.0847			
Estimation methods						
Logit	-0.1874	0.0230	1.0000	-0.1599	0.0190	0.000
OLS	-0.0352	0.0244	0.7558	-0.0382	0.0184	0.038
FE	0.0113	0.0211	0.2774			
RE	0.0018	0.0115	0.0581			
GMM	-0.0003	0.0029	0.0402			
TSLS	-0.0001	0.0030	0.0323			
Control variables						
regulation	-0.0321	0.0197	0.7982	-0.0356	0.0138	0.010
ownership	-0.0147	0.0175	0.4811			
global	-0.0017	0.0058	0.1156			
Publication characteristics						
citations	0.0497	0.0092	1.0000	0.0461	0.0095	0.000
firstpub	0.0219	0.0044	1.0000	0.0233	0.0033	0.000
IFrecursive	0.1060	0.0528	0.8749	0.0964	0.0477	0.043
reviewed_journal	-0.0249	0.0186	0.7254	-0.0151	0.0142	0.289
Constant	-0.0004	NA	1.0000	-0.1184	0.0860	0.169
Studies		31			31	
Observations		598			598	

Notes: The response variable of the PCC of the β estimate from equation (4.1). PIP = posterior inclusion probability. Post. SD = posterior standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.5. The standard errors in the frequentist check are clustered at the study level. More details on the BMA estimation are available in Table A4.1 and Figure A4.1. A detailed description of all the variables can be found in Table 4.4.

The results of our BMA exercise support the notion of the presence of publication bias (the regression coefficient on the standard error is similar to the one presented in Section 4.4); it seems that positive and insignificant estimates are underreported in the literature, because researchers tend to prefer results that are consistent with the competition-fragility hypothesis. Next, the larger the size of the data sample, the smaller the reported coefficient appears to be. As for country coverage, it seems that the estimates for developed countries tend to be slightly larger than those for non-OECD countries. The use of a quadratic relationship between competition and financial stability is associated with estimates that are on average 0.05 smaller, and the corresponding variable has a decisive posterior inclusion probability. Interestingly, the choice between micro and macro data in specifying the empirical exercise in primary studies does not influence the results significantly.

When financial stability is proxied by dummy variables for financial distress, the resulting competition coefficient estimates tend to be inflated by 0.21. In contrast, the use of macro data does not affect the reported results in a systematic way. This finding is at odds with the literature survey by Beck (2008, p. 6), who notes that "while bank-level studies do not provide unambiguous findings on the relationship between competition and stability, cross-country studies point mostly to a positive relationship." Similarly, our results contrast the finding by Schaeck and Cihak (2012), who argue that banks have higher capital ratios in more competitive environments, and thus that capitalization is one of the channels through which competition enhances stability. On the contrary, after controlling for many other method choices, we find that the use of capitalization as a proxy for stability does not affect the reported estimates of the effect of competition on stability.

As for the measures of competition, the reported estimates tend to be larger by 0.11 when Panzar and Rosse's (1987) H-statistic is used to measure bank competition. This systematic measurement issue could be due to the fact that the H-statistic imposes restrictive assumptions on a bank's cost function that are only valid when the market in question is in equilibrium (Beck, 2008). When competition is measured by the Boone index, the estimations yield smaller effects on stability (by 0.05) and the explanatory power of this variable measured by the PIP is substantial. Concerning the suitability of market structure measures of competition, i.e., concentration ratios and HHI, neither of these measures was selected in our BMA exercise as useful in explaining the variation in the literature. To further check the robustness of this result, we repeat the BMA analysis in the online appendix after excluding coefficient estimates obtained from regressions where competition was proxied by measures of concentration and HHI.

Regarding estimation methods, our results suggest that estimating equation (4.1) by a logit or a probit model tends to decrease the competition coefficient estimates by 0.19, while estimation by ordinary least squares (therefore, ignoring potential endogeneity) causes a moderate downward bias of about 0.04. Controlling for regulatory and supervisory measures decreases the estimated coefficient by approximately 0.03, which is in line with the arguments raised by Barth, Caprio, and Levine (2004) and Beck et al. (2006 a,b).

All publication characteristics that we control for have relatively high posterior inclusion probabilities. A higher recursive impact factor and more study citations are associated with larger reported estimates. Conversely, peer-reviewed journals seem to publish estimates 0.02 smaller than those reported in unpublished manuscripts, though the inclusion probability for this variable suggests only a weak effect. Moreover, our results indicate that the reported estimates of the competition coefficient increase over time.

As a final step of our analysis, we attempt to calculate the mean estimate of the competition-stability nexus after correcting for potential misspecifications and placing greater weight on estimates published in quality outlets. This part of our analysis is the most subjective as it requires a definition of "best practice" in estimating the competition coefficient. For each variable deemed useful by the BMA exercise, i.e., with PIP larger than 0.5, we plug in a preferred value, a sample minimum or a sample maximum, or, in the case of no preference, a sample mean. Then we compute a linear combination of regression parameters and obtain the value of the partial correlation coefficient conditional on our definition of best practice. We plug in the sample maxima for the size of the data set, the recursive impact factor, and the number of citations. We also prefer if the study is published in a peer-reviewed journal, if the estimation controls for regulation measures, as a higher degree of restrictions on banks' activities and barriers to bank entry is linked to systemic banking distress (Barth, Caprio, and Levine, 2004; Beck et al., 2006 a,b), and if the researcher uses the Boone index, a relatively novel approach to measuring competition arising from the industrial organization literature.

Because our focus rests primarily on the most precise competition coefficient estimates, we plug in the value 0 for the standard error of the PCC of the estimate (similarly as in Section 4.4, this approach corrects for publication bias). We also prefer if OLS is not used for the estimation of the competition-stability nexus, because it does not account for potential endogeneity. We prefer if a continuous variable is used as a proxy for stability, and if simple logit is not used for the estimation (again, because it does not allow for addressing endogeneity). We plug in zero for the dummy variable that corresponds to the assumed quadratic relation between competition and stability; in this case we have to linearize the estimates, which might induce a bias. We prefer if the H-statistic is not used in the estimation, because, as we have mentioned, it imposes restrictive assumptions on a bank's cost function that are only valid when the market in question is in equilibrium (Beck, 2008). We plug in sample means for all the other variables.

Table 4.6 summarizes the results of our best-practice estimation. Apart from the baseline results reported in the left-hand part of the table, we also report results for unweighted regressions (discussed in more detail in the next section) in the right-hand part. The column denoted "Diff." shows the difference between the best-practice coefficient estimates and the simple means of the reported coefficients presented in Table 4.1 for all countries, developed countries, and developing and transition countries.

Table 4.6: Best-practice estimates of the competition coefficient

Best practice	Weighted				Unweighted			
	Estimate	95% Conf. Int.		Diff.	Estimate	95% Conf. Int.		Diff.
All countries	0.022	-0.022	0.066	0.034	0.038	0.000	0.076	0.039
Developed	0.096	0.049	0.144	0.085	0.091	0.045	0.137	0.071
Developing and transition	0.019	-0.035	0.072	0.038	0.055	0.011	0.099	0.054

Notes: The table presents estimates of the competition coefficient for selected country groups implied by Bayesian model averaging and our definition of best practice. We take the regression coefficients estimated by BMA with PIP $>$ 0.5 and construct fitted values of the competition coefficient conditional on control for publication characteristics and other aspects of methodology (see the text for details). Diff. = the difference between these estimates and the means reported in Table 4.1. The confidence intervals are constructed using study-level clustered standard errors estimated by OLS. The right-hand part of the table presents results based on the robustness check using unweighted regressions (Table 4.8 in the next section).

In general, all the best-practice coefficient estimates are larger than the means reported in Table 4.1, which captures both the correction for publication bias and alleged misspecifications. Concerning the baseline results, however, only the estimate for developed countries is positive and statistically significant at the 5% level. Nevertheless, based on the guidelines for the interpretation of the size of partial correlation coefficients in economics (Doucouliagos, 2011), even the largest estimate reported in Table 4.6 represents merely a small effect. According to the classic Cohen (1988) guidelines, the estimate is below the threshold set for small effects. Overall, even the best-practice exercise does not reveal any important effect of bank competition on financial stability.

4.6 Robustness Checks

In this section we present the results of four robustness checks, which we obtain by estimating the model presented in the previous section with some modifications. First, we report the results of BMA when employing alternative priors (g-prior and model size). Second, we present the results for unweighted regressions with the same priors for BMA as in the baseline estimation in Section 4.5. Third, we only use frequentist methods (OLS and fixed effects). Fourth, we use inverse-variance weights, which are more common in meta-analysis.

The baseline estimation presented in the previous section employs the unit information prior for Zellner's g-prior. In this setting, the prior contains the same amount of information as one observation in the data set, and the prior is commonly used in the literature. Moreover, the uniform model prior used in the baseline specification gives the same prior probability to each model; Eicher et al. (2011) show that their choice of priors often delivers the best predictive performance. Nevertheless, the uniform model prior favors models with the mean number of regressors, i.e., $35/2 = 17.5$, because they are the most numerous among all the possible model combinations. Therefore, our first alternative specification uses a beta-binomial prior that places the same probability on each model size, in contrast to each model (Ley and Steel, 2009). We accompany the beta-binomial model prior with the BRIC g-prior as in Fernandez et al. (2001).

Table 4.7 presents the results of our BMA exercise with alternative priors. The results are qualitatively as well as quantitatively very similar to those of the baseline specification. We observe no significant differences in the magnitude of the posterior means of individual variables, and the same statement holds for their posterior inclusion probabilities. The subset of regressors identified as useful (with PIP above 0.5) fully coincides with that of the baseline specification.

Table 4.7: Results with alternative BMA priors

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Competition effect	Post. Mean	Post. SD	PIP	Coef.	Robust Std. Err.	P-value
Data characteristics						
SEPCC	-1.7527	0.2120	1.0000	-1.1940	0.6511	0.067
Samplesize	-0.0362	0.0036	1.0000	-0.0240	0.0089	0.007
T	0.0003	0.0034	0.0373			
sampleyear	0.0000	0.0005	0.0335			
Countries examined						
developed	0.1976	0.0248	1.0000	0.1761	0.0295	0.000
developing and transition	0.1030	0.0188	1.0000	0.0985	0.0262	0.000
Design of the analysis						
quadratic	-0.0517	0.0141	0.9884	-0.0441	0.0128	0.001
endogeneity	0.0159	0.0269	0.3037			
macro	0.0028	0.0132	0.0672			
Averaged	-0.0004	0.0043	0.0310			
Treatment of stability						
dummies	0.2179	0.0315	1.0000	0.1841	0.0194	0.000
NPL	0.0012	0.0047	0.0818			
Zscore	-0.0004	0.0023	0.0427			
profit_volat	0.0004	0.0043	0.0255			
profitability	-0.0002	0.0024	0.0236			
capitalization	0.0001	0.0024	0.0186			
DtoD	-0.0007	0.0060	0.0313			

Treatment of competition						
Hstatistic	0.1074	0.0228	1.0000	0.1140	0.0181	0.000
Boone	-0.0637	0.0375	0.8020	-0.0583	0.0225	0.010
Concentration	-0.0182	0.0244	0.4183			
Lerner	0.0032	0.0128	0.0946			
HHI	0.0021	0.0107	0.0659			
Estimation methods						
Logit	-0.1883	0.0237	1.0000	-0.1599	0.0190	0.000
OLS	-0.0296	0.0265	0.6208	-0.0382	0.0184	0.038
FE	0.0160	0.0258	0.3261			
RE	0.0020	0.0119	0.0521			
GMM	-0.0002	0.0023	0.0272			
TSLS	-0.0002	0.0031	0.0258			
Control variables						
regulation	-0.0313	0.0205	0.7625	-0.0356	0.0138	0.010
ownership	-0.0129	0.0176	0.4014			
global	-0.0013	0.0051	0.0837			
Publication characteristics						
citations	0.0476	0.0101	1.0000	0.0461	0.0095	0.000
firstpub	0.0207	0.0050	1.0000	0.0233	0.0033	0.000
IFrecursive	0.0958	0.0622	0.7699	0.0964	0.0477	0.043
reviewed_journal	-0.0211	0.0198	0.6028	-0.0151	0.0142	0.289
Constant	-0.0004	NA	1.0000	-0.1184	0.0860	0.169
Studies		31			31	
Observations		598			598	

Notes: The response variable is the competition effect. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.5. The standard errors in the frequentist check are clustered at the study level. In this specification we use the beta-binomial prior advocated by Ley & Steel (2009) (the prior model probabilities are the same for all model sizes) and the BRIC g-prior following Fernandez et al. (2001). More details on the BMA estimation are available in Table A4.2 in Appendix A. A detailed description of all the variables is available in Table 4.4.

Second, we run the BMA exercise with the same priors as in our baseline specification but for regressions not weighted by the inverse of the number of observations reported in studies (Table 4.8). In this case studies with fewer reported competition coefficient estimates become less influential in the meta-analysis, and the results are dominated by papers that produce many estimates. In this robustness check the BMA only selects 14 variables with inclusion probability higher than 0.5 as opposed to 15 variables in the baseline specification. In addition, the results of the robustness check suggest that measuring stability by means of bank profitability tends to lower the coefficient estimate by 0.03. In contrast, estimating equation (4.1) by fixed effects or instrumental variables increases the estimated competition coefficient by 0.05, with a decisive PIP in both cases. These findings are consistent with our results from the previous section, where we report that using OLS (which disregards endogeneity) is associated with smaller reported estimates.

Furthermore, including controls for bank ownership decreases the reported estimate by 0.06 with a decisive PIP. This finding supports the results by Barth, Caprio, and Levine (2004), who argue that bank ownership matters for bank stability. In particular, they find that foreign bank entry tends to be positively related to banking system stability, while government ownership impacts competitiveness as well as stability in a negative way. In contrast to the baseline specification, here we do not find the following aspects important: controlling for a nonlinear relationship between competition and stability, measuring competition via the Boone index, estimating equation (4.1) by means of OLS, controlling for regulation and supervision in the banking sector, and publication of the study in a peer-reviewed journal. As for the signs and magnitudes of the estimated coefficients for individual regressors, they broadly coincide with the baseline specification. Nevertheless, the robustness check shows less evidence for publication bias in the literature. Also, the estimated coefficients for dummy variables reflecting developed and developing countries are much smaller, shrinking the difference between the implied competition coefficients for the different country groups.

Table 4.8: Results for unweighted regressions

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Competition effect	Post. Mean	Post. SD	PIP	Coef.	Robust Std. Err.	P-value
Data characteristics						
SEPCC	-0.7259	0.5667	0.7003	-0.5768	0.7862	0.4630
Samplesize	-0.0258	0.0082	1.0000	-0.0248	0.0092	0.0070
T	0.0008	0.0034	0.0735			
sampleyear	0.0006	0.0015	0.1946			
Countries examined						
developed	0.1529	0.0172	1.0000	0.1519	0.0175	0.0000
developing and transition	0.1127	0.0172	1.0000	0.1156	0.0170	0.0000
Design of the analysis						
quadratic	0.0012	0.0050	0.0755			
endogeneity	0.0056	0.0110	0.2461			
macro	-0.0103	0.0161	0.3408			
Averaged	0.0000	0.0024	0.0219			
Treatment of stability						
dummies	0.1861	0.0281	1.0000	0.1660	0.0176	0.0000
NPL	0.0138	0.0249	0.2739			
Zscore	0.0091	0.0166	0.2660			
profit_volat	0.0176	0.0238	0.4350			

profitability	-0.0281	0.0233	0.6587	-0.0451	0.0246	0.0660
capitalization	0.0101	0.0196	0.2437			
DtoD	-0.0015	0.0080	0.0674			
Treatment of competition						
Hstatistic	0.1294	0.0223	1.0000	0.1123	0.0173	0.0000
Boone	-0.0021	0.0088	0.0873			
Concentration	0.0159	0.0244	0.3626			
Lerner	0.0136	0.0211	0.3566			
HHI	0.0103	0.0199	0.2488			
Estimation methods						
Logit	-0.1304	0.0303	0.9999	-0.1275	0.0121	0.0000
OLS	0.0000	0.0019	0.0214			
FE	0.0621	0.0134	1.0000	0.0503	0.0113	0.0000
RE	0.0128	0.0204	0.3355			
GMM	0.0000	0.0018	0.0221			
TSLS	0.0532	0.0132	0.9999	0.0515	0.0147	0.0000
Control variables						
regulation	0.0002	0.0020	0.0281			
ownership	-0.0595	0.0096	1.0000	-0.0588	0.0289	0.0420
global	0.0016	0.0054	0.1033			
Publication characteristics						
citations	0.0377	0.0063	0.9996	0.0407	0.0087	0.0000
firstpub	0.0179	0.0033	0.9997	0.0205	0.0029	0.0000
IFrecursive	0.0470	0.0419	0.6405	0.0490	0.0379	0.1960
reviewed_journal	0.0019	0.0080	0.0807			
Constant	-0.1269	NA	1.0000	-0.1263	0.0870	0.1460
Studies		31			31	
Observations		598			598	

Notes: The response variable is the competition effect. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.5. The standard errors in the frequentist check are clustered at the study level. In this specification we do not weight the regressions by the inverse of the number of estimates reported per study. More details on the BMA estimation are available in Table A4.3 in Appendix A. A detailed description of all the variables is available in Table 4.4.

Third, we only use frequentist methods in estimations; the results are reported in Table 4.9. The left-hand panel of the table shows OLS, while the right-hand panel shows fixed effects estimation. The advantage of the fixed effects estimation is that it removes all idiosyncratic effects of individual studies (such as study quality) on results, but it also automatically removes all variables constant within studies and greatly decreases the variance of some of the method variables. The results are, in general, consistent with our baseline estimation: in both specifications we find evidence for downward publication bias, and the best-practice estimate of the effect of bank competition on financial stability is small and statistically insignificant (see Table C11 in the online appendix for best-practice estimation related to this robustness check).

There are, of course, some differences in the estimated magnitude and statistical significance for some of the moderator variables. Using OLS, we find that estimates obtained with the Lerner index as a proxy for competition tend to be somewhat larger than those computed with other competition proxies. Moreover, the use of fixed effects in primary studies tends to be associated with larger reported coefficients. Apart from that, it seems that Bayesian and frequentist methods bring very similar results. We observe more differences when we use study fixed effects, which wipe out a large portion of variance in our data. With fixed effects, the variable controlling for the number of cross-sectional units used in the primary study becomes insignificant. In contrast, the dummy variable that equals one for macro-level studies becomes statistically significant and large, indicating that macro studies to report much larger estimates of the competition-stability nexus compared to bank-level studies. The choice of the measure of competition seems to be more important now than in our baseline estimation. Interestingly, however, the two dummy variables that reflect the use of competition proxies based on concentration (variables HHI and Concentration) are not jointly statistically significant – which we corroborate in the online appendix by conducting a separate meta-analysis of the estimates that are obtained using concentration measures. Estimation methods influence the results significantly: the use of OLS, fixed effects, and GMM is associated with larger reported coefficients than the use of other estimation methods.

Fourth, we estimate our baseline model with weights based on the inverse variance of the reported estimates instead of the inverse of the number of estimates reported in a study. The left-hand panel of Table 4.10 shows the results of BMA; the right-hand panel shows OLS. Once again, we find evidence for downward publication bias: the corresponding variable has high posterior inclusion probability, although in the OLS regression its statistical significance decreases (nevertheless, that is due to the inclusion of many potentially redundant variables in the frequentist setting). The best-practice estimate, reported in Table C12 in the online appendix, once again shows no effect of bank competition on financial stability, which is consistent with our previous results. The individual regression coefficients are broadly similar to the baseline case, with the exception of the choice of a proxy for stability: now it seems to be more important, and each choice can be expected to bring different results. Similarly to the previous case, however, the largest difference is caused by the use of a binary variable on the left-hand side in primary studies; this method choice is associated with competition-stability estimates 0.2 larger compared with the reference case.

Table 4.9: Results for frequentist methods

Response variable:	OLS			Fixed effects		
Competition effect	Coef.	Robust Std. Err.	P-value	Coef.	Robust Std. Err.	P-value
Data Characteristics						
SEPCC	-1,5708	0,8567	0.067	-1,6234	0,6912	0.026
Samplesize	-0,0363	0,0110	0.001	0,0148	0,0212	0.491
T	0,0141	0,0107	0.188	-0,0511	0,0268	0.067
sampleyear	0,0040	0,0033	0.222	0,0057	0,0032	0.082
Countries examined						
developed	0,1689	0,0211	0.000		(omitted)	
undeveloped	0,1008	0,0166	0.000	0,1020	0,0760	0.189
Design of the analysis						
quadratic	-0,0080	0,0204	0.694	-0,0071	0,0135	0.604
endogeneity	0,0240	0,0292	0.410	-0,0292	0,0163	0.084
macro	-0,0040	0,0364	0.914	0,1882	0,0138	0.000
someAveraged	-0,0023	0,0285	0.935	0,0226	0,0151	0.146
Treatment of stability						
dummies	0,2232	0,0373	0.000		(omitted)	
NPL	0,0299	0,0259	0.250	0,0239	0,0232	0.310
Zscore	0,0116	0,0249	0.641	0,0172	0,0228	0.456
profit_volat	0,0284	0,0206	0.168	0,0192	0,0214	0.378
profitability	-0,0142	0,0270	0.600	-0,0048	0,0270	0.860
capitalization	0,0184	0,0240	0.443	0,0052	0,0254	0.838
DtoD	-0,0157	0,0337	0.641	0,0217	0,0284	0.452
Treatment of competition						
Hstatistic	0,1629	0,0308	0.000	0,0577	0,0201	0.007
Boone	0,0010	0,0271	0.970	0,0744	0,0112	0.000
Concentration	0,0351	0,0356	0.324	0,0709	0,0346	0.050
Lerner	0,0485	0,0188	0.010	0,0721	0,0189	0.001
HHI	0,0444	0,0257	0.084	0,0654	0,0252	0.014
Estimation methods						
Logit	-0,1481	0,0405	0.000		(omitted)	
OLS	-0,0022	0,0218	0.919	0,0225	0,0108	0.045
FE	0,0624	0,0247	0.011	0,0392	0,0180	0.038
RE	0,0317	0,0382	0.406	-0,0042	0,0182	0.819
GMM	0,0014	0,0159	0.932	0,0437	0,0206	0.043
TSLS	0,0393	0,0230	0.087	0,0223	0,0186	0.239
Control variables						
regulation	-0,0184	0,0138	0.181	0,0062	0,0104	0.558
ownership	-0,0341	0,0227	0.133	-0,0193	0,0311	0.539
global	0,0112	0,0176	0.524	0,0239	0,0152	0.125

Publication characteristics						
citations	0,0408	0,0146	0.005		(omitted)	
firstpub	0,0159	0,0067	0.017		(omitted)	
IFrecursive	0,0890	0,0363	0.014		(omitted)	
reviewed_journal	-0,0042	0,0271	0.876		(omitted)	
Constant	-0,1350	0,1124	0.230	-0,1783	0,1656	0.290
Studies		31			31	
Observations		598			598	

Notes: The response variable of the PCC of the β estimate from equation (4.1). PIP = posterior inclusion probability. Post. SD = posterior standard deviation. In the frequentist check we include all explanatory variables. The standard errors in the frequentist check are clustered at the study level. The regressions are estimated by weighted least squares, where the inverse of the number of estimates reported per study is taken as the weight. The left-hand side of the table presents the results of OLS regression while the right-hand part presents the results of fixed effects regression. A detailed description of all the variables can be found in Table 4.4.

Table 4.10: Results for specifications weighted by inverse variance

Response variable:	Bayesian model averaging			OLS		
Competition effect	Post. Mean	Post. SD	PIP	Coef.	Robust Std. Err.	P-value
Data Characteristics						
SEPCC	-1.4216	NA	1.0000	-1.0152	0.9359	0.278
Samplesize	-0.0279	0.0039	1.0000	-0.0276	0.0067	0.000
T	0.0031	0.0054	0.3013	0.0177	0.0097	0.069
sampleyear	0.0001	0.0007	0.0828	0.0043	0.0034	0.201
Countries examined						
developed	0.1266	0.0074	1.0000	0.1257	0.0158	0.000
undeveloped	0.0730	0.0066	1.0000	0.0729	0.0216	0.001
Design of the analysis						
quadratic	0.0007	0.0029	0.0817	0.0059	0.0120	0.620
endogeneity	0.0004	0.0034	0.0935	0.0060	0.0187	0.747
macro	-0.0035	0.0071	0.2561	-0.0123	0.0190	0.519
someAveraged	0.0003	0.0023	0.0644	0.0010	0.0125	0.934
Treatment of stability						
dummies	0.2132	0.0269	1.0000	0.2346	0.0240	0.000
NPL	0.0441	0.0096	0.9996	0.0427	0.0231	0.065
Zscore	0.0384	0.0065	1.0000	0.0377	0.0192	0.049
profit_volat	0.0580	0.0075	1.0000	0.0576	0.0187	0.002
profitability	0.0134	0.0122	0.6193	0.0193	0.0153	0.207
capitalization	0.0407	0.0076	1.0000	0.0402	0.0228	0.078
DtoD	0.0703	0.0106	1.0000	0.0776	0.0303	0.010
Treatment of competition						
Hstatistic	0.0795	0.0134	1.0000	0.0866	0.0309	0.005
Boone	0.0000	0.0020	0.0414	0.0153	0.0155	0.323
Concentration	-0.0002	0.0030	0.0427	0.0084	0.0220	0.701
Lerner	0.0000	0.0011	0.0419	0.0058	0.0055	0.297

HHI	0.0001	0.0014	0.0446	0.0108	0.0091	0.234
Estimation methods						
Logit	-0.1250	0.0275	0.9995	-0.1366	0.0352	0.000
OLS	-0.0002	0.0020	0.0481	0.0111	0.0227	0.626
FE	0.0698	0.0077	1.0000	0.0757	0.0207	0.000
RE	-0.0008	0.0074	0.0483	-0.0235	0.0619	0.704
GMM	0.0002	0.0017	0.0469	-0.0005	0.0197	0.980
TSLS	0.0504	0.0063	1.0000	0.0559	0.0218	0.010
Control variables						
regulation	0.0003	0.0014	0.0723	0.0049	0.0059	0.409
ownership	-0.0028	0.0069	0.1872	-0.0255	0.0210	0.226
global	0.0005	0.0020	0.0854	0.0083	0.0143	0.561
Publication characteristics						
citations	0.0226	0.0074	0.9398	0.0282	0.0131	0.032
firstpub	0.0094	0.0031	0.9350	0.0050	0.0058	0.389
IFrecursive	-0.0007	0.0049	0.0575	-0.0131	0.0446	0.768
reviewed_journal	0.0028	0.0084	0.1445	0.0108	0.0213	0.612
Constant	-0.0004	0.0114	0.0457	-0.0845	0.0856	0.324
Studies		31			31	
Observations		598			598	

Notes: The response variable of the PCC of the β estimate from equation (4.1). PIP = posterior inclusion probability. Post. SD = posterior standard deviation. In the frequentist check we include all explanatory variables. The standard errors in the frequentist check are clustered at the study level. The regressions are estimated by weighted least squares where the inverse of the estimates' variance is taken as the weight. A detailed description of all the variables can be found in Table 4.4.

We provide more robustness checks and additional results in the online appendix. Pages 1-6 of the "additional results" file in the online appendix describe how excluding estimates produced using a concentration-based proxy for bank competition does not alter our main conclusions. Next, as an anonymous referee suggests, the strength of publication bias in the literature may be associated with the affiliations of the authors of primary studies. We try to estimate the funnel asymmetry test for the sub-sample of studies written by researchers not affiliated with policy institutions (such as central banks, ministries, and supra-national institutions). The results, reported in Table C1, show that the coefficient for publication bias loses statistical significance. Nevertheless, this finding is mostly due to the decreased number of the degrees of freedom available for estimation. Table C2 includes an interaction term of the standard error and a dummy variable for studies co-authored by researchers affiliated with policy institutions; the interaction is insignificant, which implies that the extent of publication bias is similar among these two groups of studies. Moreover, the underlying effect of competition on stability corrected for any potential bias is small and statistically insignificant in all these specifications.

Our data sample includes several studies that were published in the same journal. Patterns of publication selection might vary across journals, so in Table C3 we exclude all studies published in journals from which we have more than one study. Similarly to the case of excluding studies co-authored by researchers affiliated with policy institutions, the statistical significance of the publication bias coefficient decreases, because we have much less observations available in the regression. Nevertheless, the two groups of studies do not differ substantially in the magnitude of publication bias, as illustrated by Table C4, where we add an interaction of the standard error and a dummy variable that equals one for journals that provide more than one study for our data set.

Next, we investigate potential bi-directional publication bias in the literature. The disadvantage of the classical funnel asymmetry test is that it only identifies publication bias in one direction. If insignificant estimates, both positive and negative, are discarded, the coefficient on the standard error will be a biased estimate of the extent of publication selection in the literature (nevertheless, the estimate of the underlying effect corrected for the bias will be unbiased, because the classical funnel asymmetry test effectively filters out the net publication bias, either downward or upward). The authors that favor the competition-stability hypothesis might treat negative estimates with suspicion, while the authors preferring the competition-fragility hypothesis might tend to discard positive estimates. The results of the test of bi-directional bias are reported in Tables C5 and C6. We follow Bom and Ligthart (2014) and replace the standard error in the funnel asymmetry test by interactions of the standard error and dummy variables that equal one if the estimate of the competition-stability nexus is positive and negative, respectively. In most specifications the estimated coefficients for these interaction terms are quantitatively similar, but we reject the hypothesis that they are equal. Thus our results are consistent with the presence of some bi-directional publication selection, and we conclude that our estimates of the extent of publication selection presented in Section 4 are probably downward biased. The corrected mean effect of competition on stability is still close to zero.

A third of the studies in our sample investigate potential non-linearity in the effect of competition on stability by including both a linear and quadratic form of the competition measure on the right-hand-side of the regression. Our default approach in this meta-analysis is to approximate the first-order effect using the sample mean of the competition proxy and the delta method to calculate the corresponding standard error. In Tables C7 and C8 we present separate funnel asymmetry tests for the estimates of linear and quadratic terms reported in studies that include both terms into the regression. We find some evidence for downward publication bias among the estimates of the linear term, which is consistent with our baseline results (but now with much fewer degrees of freedom our results are less precise, decreasing statistical significance). More importantly, estimates of the mean effect for both the linear and quadratic term are virtually zero, which corroborates our conclusion that, on average, the available empirical literature does not point to any relationship between bank competition and financial stability. Further, in Tables C9 and C10 we show that excluding non-linear estimates does not alter our conclusions concerning publication bias in the literature: the non-linear estimates show a similar pattern of publication selection. Our results are consistent with several studies that fail to find non-linearity in the effect of competition on stability, such as Agoraki et al. (2011), Turk Ariss (2010), and Fungacova and Weill (2009).

Several general remarks on our methodology are in order. We prefer to use weights based on the inverse of the number of estimates presented in each paper in contrast to weights based on the inverse variance of each estimate, which are typically employed in meta-analysis. We have five reasons for this choice. First, although multiple Monte Carlo simulations (for example, Stanley and Doucouliagos, 2015) show that inverse-variance weights bring optimal results in meta-analysis, these simulations do not consider the case when each study reports several estimates of the effect in question, and moreover if the number of estimates per study varies. When weights are not constant across panels, the interpretation of the weighted results with panel data is unclear, which is why some statistical packages (for example, Stata) do not allow the use of such weights with panel estimators. Second, in applications of meta-analysis researchers typically include variables defined at the level of individual studies, such as the number of citations or publication year. With multiple estimates reported per study the introduction of inverse-variance weighting brings artificial variation to the study-level variables, because they suddenly vary within-studies (and are heavily correlated with other weighted variables). Again, it is not clear how to interpret such results, and there have been no Monte Carlo simulations that would help us with inference.

Third, in meta-analysis the reported standard errors are likely to be endogenous with respect to the reported point estimates. Certain method choices (for example, simple OLS versus instrumental variables) influence both the standard errors and the point estimates. If the influence of the method on the two statistics goes in the same direction, a large coefficient in the funnel asymmetry test may simply reflect this endogeneity instead of any publication or small-sample bias (moreover, as meta-analysis becomes better known in economics, standard errors themselves might become the target of publication selection in order for researchers to increase the weight of their results in meta-analyses). One solution is to use the inverse of the square root of the number of observations as an instrument for the standard error, because this instrument is proportional to the standard error, but not likely to be correlated with method choices. It is unclear how to interpret results of a specification where the employed weights are potentially endogenous to both the response and explanatory variable.

Fourth, inverse-variance weights are highly sensitive to outliers in precision. In most meta-analyses there are a couple of studies that report very small standard errors for no obvious reasons other than idiosyncratic methodology, and very often they also report small point estimates (this issue is connected to the endogeneity problem). The meta-analyst can either omit these studies, which is difficult to justify, winsorize these observations (as in Havranek et al., 2015b), or include them as they are. The differences between these three approaches increase dramatically when inverse-variance weights are used. Fifth, the weights based on the inverse of the number of observations reported in a study give each study the same importance, which in our opinion is more intuitive than to give each study a weight based on the number of estimates it reports (which is what happens when we do not use our preferred weights). Certainly more research is needed to determine the optimal weighting scheme in meta-analysis with panel data. An important step in this direction is presented by Reed et al. (2015), but they unfortunately do not consider the case when different primary studies report a different number of empirical estimates.

A second non-standard feature of our analysis is the reliance on Bayesian model averaging instead of frequentist methods used in most economics meta-analyses (especially OLS or its inverse-variance-weighted variations). While we show that using frequentist instead of Bayesian methods would not change our main results much, we prefer to use BMA. A common objection to BMA is the claim that the method is atheoretical, throwing in many potential explanatory variables and using statistical techniques to find the most important ones. The problem is that in meta-analysis we always have a large number of explanatory variables that might (or might not) potentially influence the reported point estimates. For some of them our economic intuition is stronger, for some of them weaker; nevertheless, we want to control for all the major aspects of data, methodology, and publication characteristics (as recommended by Stanley et al., 2013). The economic theory rarely helps us decide which of the variables we should omit, and the choice between BMA and OLS with sequential t-tests (the standard approach in meta-analysis) is not connected to this issue. Sequential t-tests are not statistically valid, because each subsequent test does not take into account that the result is conditional on the previous one. BMA, in contrast, can be thought of as an extension of the typical frequentist practice in which different specifications with various control variables are estimated to evaluate the robustness of results.

We admit, however, that in our experience BMA and sequential t-tests often yield similar results, although there is no reason why this finding should hold in general. Then a meta-analyst faces a trade-off between a method that is statistically valid and one that is easier to compute. We opt for the first one, and would recommend other meta-analysts to do so when the number of potential explanatory variables in meta-analysis is large (10 may be an acceptable rule of thumb, although the threshold is obviously arbitrary). With less than 10 variables we believe the meta-analyst does not have to resort to sequential t-tests, but simply evaluate the OLS regression with all variables, and additionally several robustness checks. Most economics meta-analysis, however, have more than 10 explanatory variables, which makes BMA an attractive method for this field, because it helps tackle model and parameter uncertainty. BMA techniques similar to those employed in this paper have already been used in economics meta-analyses by Moeltner and Woodward (2009), Irsova and Havranek (2013), and Havranek and Irsova (2015). Havranek et al. (2015a) propose a modification for the case when a group of explanatory variables are strongly predicted to be important by economic theory: these variables are fixed in BMA, which means that they are included in all estimated models, while the subsets of control variables vary.

Conclusions

We conduct a meta-regression analysis of 598 estimates of the relationship between bank competition and financial stability reported in 31 studies. We complement the previous narrative reviews of the literature (Beck, 2008; Carletti and Hartmann, 2002) with a formal treatment of publication bias and heterogeneity in estimations of the competition-stability nexus. Our results suggest that the mean reported estimate of the relationship is close to zero, even after correcting for publication bias and potential misspecification problems. We find evidence for publication selection against positive results; that is, some authors of primary studies tend to discard estimates inconsistent with the competition-fragility hypothesis. To uncover the dependence of the reported estimates on the aspects of study design, we employ Bayesian model averaging, which helps us address model uncertainty.

Our results indicate that data characteristics matter for the reported coefficients corresponding to the competition-stability nexus. Researchers who use heterogeneous samples of countries (including both developed and developing economies) tend to obtain smaller estimates. The effect of competition on stability is larger in developed countries, but even there the positive effects do not seem to be strong. Next, accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus. We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates. In contrast, it does not seem to matter for the results whether the authors of primary studies use micro or macro data.

Furthermore, we show that the definition of the proxy for financial stability is important for the results of primary studies. For example, if dummy variables (usually indicating financial crises) are used as a proxy for stability, the authors tend to report much larger estimates than when a continuous measure of financial stability is used. In a similar vein, the results of primary studies are systematically affected by the choice of the proxy for bank competition. Studies using the H-statistic tend to report larger estimates of the competition-stability nexus, while studies that employ the Boone index usually show smaller estimates; nevertheless, we find no evidence of systematic differences between the results of the studies that use competition measures and the studies that use concentration as a proxy for competition. Next, if the researchers ignore the endogeneity problem in regressing financial stability on bank competition, they tend to underestimate the effect.

We also find that controlling for supervisory and regulatory conditions in regressions usually decreases the reported estimates, which supports the notion that banking systems with more activity restrictions and greater barriers to entry are more likely to suffer from systemic financial distress (Beck et al., 2006 a,b). Finally, studies that receive more citations and are published in journals with a high impact factor tend to report larger estimates of the competition-stability nexus. In the last step of our analysis we construct a weighted average of all the estimates and give more weight to the ones whose authors avoid potential misspecifications (such as ignoring endogeneity) and that have better publication characteristics (for example, more citations). Because several potential misspecifications influence the results in opposite ways, the resulting estimate still points to a very weak or non-existent link between bank competition and financial stability.

The principal limitation of meta-analysis is that it can only correct for problems in the literature that have already been addressed by some researchers. If, in contrast, all studies in the field share a common misspecification that causes a systematic bias, meta-analysis gives biased results as well. It is possible that the underlying effect of banking competition on financial stability is nonzero, but that the data and methods that are currently used in the literature do not allow researchers to identify such an effect. Nevertheless, we show that the bulk of the existing empirical literature provides little support for either the competition-fragility or competition-stability hypothesis.

References

Agoraki, M. K., Delis, M. D. & Pasiouras, F., 2011. "Regulations, competition and bank risk-taking in transition countries." Journal of Financial Stability 7(1): pp. 38–48.

Allen, F. & Gale, D., 2000. "Comparing financial systems." MIT Press, Cambridge, MA.

Allen, F. & D. Gale, 2004. "Competition and financial stability." Journal of Money, Credit, and Banking 36: pp. 453–480.

Anginer, D., Demirguc-Kunt, A. & Zhu, M., 2014. "How does competition affect bank systemic risk?" Journal of Financial Intermediation 23(1): pp. 1–26.

Ashenfelter, O., Harmon, C. & Oosterbeek, H., 1999. "A review of estimates of the schooling/earnings relationship, with tests for publication bias." Labour Economics 6(4): pp. 453–470.

Babecky, J. & Havranek, T., 2014. "Structural reforms and growth in transition: A meta-analysis." Economics of Transition 22(1): pp. 13–42.

Barth, J. R., Caprio, Jr., G. & Levine, R., 2004. "Bank supervision and regulation: What works best?" Journal of Financial Intermediation 13(2): pp. 205–248.

Bazzana, F. & Yaldiz, E., 2010. "The effect of market power on bank risk taking in Turkey." Financial Theory and Practice 34(3): pp. 297–314.

Beck, T., 2008. "Bank competition and financial stability: Friends or foes?" Policy Research Working Paper Series 4656, The World Bank.

Beck, T., De Jonghe, O. & Schepens, G., 2013. "Bank competition and stability: Cross-country heterogeneity." Journal of Financial Intermediation 22(2): pp. 218–244.

Beck, T., Demirguc-Kunt, A. & Levine, R., 2006a. "Bank concentration, competition, and crises: First results." Journal of Banking and Finance 30: pp. 1581–1603.

Beck, T., Demirguc-Kunt, A. & Levine, R., 2006b. "Bank concentration and fragility: Impact and mechanics." In: Stulz, R. and M. Carey (eds), The Risks of Financial Institutions. National Bureau of Economic Research.

Berger, A. N., Klapper, L. & Turk-Ariss, R., 2009. "Banking competition and financial stability." Journal of Financial Services Research 35: pp. 99–118.

Bikker, J. A., 2004. "Competition and efficiency in a unified European banking market." Cheltenham, UK: Edward Elgar.

Bom, P. R. D. & Ligthart, J. E., 2014. "What Have We Learned From the Three Decades of Research on the Productivity of Public Capital?" Journal of Economic Surveys 28(5): 889-916.

Boone, J., 2008. "A new way to measure competition." Economic Journal 118: pp. 1245–1261.

Boot, A. & Greenbaum, S., 1993. "Bank regulation, reputation and rents: Theory and policy implications." In: Mayer C. and X. Vives (eds), Capital Markets and Financial Intermediation, pp. 262–285. Cambridge: University Press.

Boot, A. W. A. & Thakor, A. V., 1993. "Self-interested bank regulation." American Economic Review 83(2): pp. 206–212.

Boyd, J. & De Nicolo, G., 2005. "The theory of bank risk-taking and competition revisited." Journal of Finance 60: pp. 1329–1342.

Boyd, J. H. & Runkle, D. E., 1993. "Size and performance of banking firms: Testing the predictions of theory." Journal of Monetary Economics 31(1): pp. 47–67.

Boyd, J. H., De Nicolo, G. & Jalal, A. M., 2006. "Bank risk-taking and competition revisited: New theory and new evidence." IMF Working Paper No. 06/297.

Caminal, R. & Matutes, C., 2002. "Market power and bank failures." International Journal of Industrial Organisation 20: pp. 1341–1361.

Card, D. & Krueger, A. B., 1995. "Time-series minimum wage studies: A meta-analysis." American Economic Review 85(2): pp. 238–243.

Carletti, E. & Hartmann, P., 2002. "Competition and stability: What's special about banking?" Working Paper Series 0146, European Central Bank.

Chetty, R., Guren, A., Manoli, D. & Weber, A., 2011. "Are micro and macro labour supply elasticities consistent? A review of evidence on the intensive and extensive margins." American Economic Review 101(3): pp. 471–475.

Cihak, M. & Hesse, H., 2010. "Islamic Banks and Financial Stability: An Empirical Analysis." Journal of Financial Services Research 38(2): pp. 95-113, December.

Cipollini, A. & Fiordelisi, F., 2009. "The impact of bank concentration on financial distress: The case of the European banking system." Centro Studi di Banca e Finanza (CEFIN) (Center for Studies in Banking and Finance) 09021, Universita di Modena e Reggio Emilia, Facoltà di Economia "Marco Biagi".

Claessens, S. & Laeven, L., 2004. "What drives bank competition? Some international evidence." Journal of Money, Credit and Banking 36: pp. 563–83.

Cohen, J., 1988. Statistical Power Analysis in the Behavioral Sciences. 2nd ed. Hillsdale: Erlbaum.

De Nicolò, G. & Loukoianova, E., 2007. "Bank ownership, market structure and risk." IMF Working Papers 07/215, International Monetary Fund.

DeLong, J. B. & Lang, K., 1992. "Are all economic hypotheses false?" Journal of Political Economy 100(6): pp. 1257–1272.

Doucouliagos, H., 2011. "How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics." Economics Series 2011/5, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, H. & Stanley, T. D., 2013. "Are all economic facts greatly exaggerated? Theory competition and selectivity." Journal of Economic Surveys 27(2): pp. 316–339.

Doucouliagos, H. & Stanley, T. D., 2009. "Publication selection bias in minimum wage research? A meta-regression analysis." British Journal of Industrial Relations 47(2): pp. 406–428.

Doucouliagos, H., Hama, J. & Stanley, T. D., 2012. "Pay for performance and corporate governance reform." Industrial Relations: A Journal of Economy and Society 51(3): pp. 670–703.

Egger, M., Smith, G. D., Scheider, M. & Minder, C., 1997. "Bias in meta-analysis detected by a simple, graphical test." British Medical Journal 316: pp. 629–634.

Eicher, T. S., Papageorgiou, C. & Raftery, A. E., 2011. "Default priors and predictive performance in Bayesian model averaging, with application to growth determinants." Journal of Applied Econometrics 26(1): pp. 30–55.

Feldkircher, M. & Zeugner, S., 2009. "Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in Bayesian model averaging." IMF Working Papers 09/202, International Monetary Fund.

Fernandez, C., Ley, E. & Steel, M. F. J., 2001. "Benchmark priors for Bayesian model averaging." Journal of Econometrics 100(2): pp. 381–427.

Fernandéz, R. O. & Garza-García, J. G., 2012. "The relationship between bank competition and financial stability: A case study of the Mexican banking industry." Centre for Global Finance Working Paper No. 03/12. University of the West of England.

Fu, X., Lin, Y. & Molyneux, P., 2014. "Bank competition and financial stability in Asia Pacific." Journal of Banking and Finance 38(C): pp. 64–77.

Fungacova, Z. & Weill, L., 2009. "How market power influences bank failures: Evidence from Russia." BOFIT Discussion Papers 12/2009, Bank of Finland, Institute for Economies in Transition.

Havranek, T., Horvath, R., Irsova, Z., & Rusnak, M., 2015a. "Cross-Country Heterogeneity in Intertemporal Substitution." Journal of International Economics, forthcoming.

Havranek, T. & Irsova, Z., 2011. "Estimating vertical spillovers from FDI: Why results vary and what the true effect is." Journal of International Economics 85(2): pp. 234-244.

Havranek, T. & Irsova, Z., 2012. "Survey Article: Publication Bias in the Literature on Foreign Direct Investment Spillovers." Journal of Development Studies 48(10): pp. 1375-1396.

Havranek, T. & Irsova, Z., 2015. "Do Borders Really Slash Trade? A Meta-Analysis." William Davidson Institute WP 1088.

Havranek, T., Irsova, Z. & Janda, K., 2012. "Demand for gasoline is more price-inelastic than commonly thought." Energy Economics 34(1): pp. 201-207.

Havranek, T., Rusnak, M. & Sokolova, A., 2015b. "Habit Formation in Consumption: A Meta-Analysis." Czech National Bank WP, available at meta-analysis.cz/habits.

Hellman, T., Murdoch, K. & Stiglitz, J. E., 2000. "Liberalization, moral hazard in banking and prudential regulation: Are capital requirements enough?" American Economic Review 90: pp. 147–165.

Irsova, Z. & Havranek, T., 2013. "Determinants of Horizontal Spillovers from FDI: Evidence from a Large Meta-Analysis." World Development 42(C): pp. 1-15.

Jeon, J. Q. & Lim, K. K., 2013. Bank competition and financial stability: A comparison of commercial banks and mutual savings banks in Korea. Pacific-Basin Finance Journal, Elsevier, vol. 25(C), pp. 253-272.

Keeley, M., 1990. "Deposit insurance, risk and market power in banking." American Economic Review 80: pp. 1183–2200.

Laeven, L. & Levine, R., 2009. "Bank governance, regulation and risk taking." Journal of Financial Economics 93: pp. 259–275.

Lepetit, L., Nys, E., Rous, P. & Tarazi, A., 2008. "Bank income structure and risk: An empirical analysis of European banks." Journal of Banking and Finance 32: pp. 1452–1467.

Levy Yeyati, E. & Micco, A., 2007. "Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk." Journal of Banking and Finance 31: pp. 1633–1647.

Ley, E. & Steel, M. F. J., 2009. "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions." Journal of Applied Econometrics 24(4): pp. 651–674.

Liu, H. & Wilson, J. O. S., 2013. "Competition and risk in Japanese banking." European Journal of Finance 19(1): pp. 1–18.

Liu, H., Molyneux, P. & Nguyen, L. H., 2012. "Competition and risk in South East Asian commercial banking." Applied Economics 44(28): pp. 3627–3644.

Marcus, A. J., 1984. "Deregulation and bank policy." Journal of Banking and Finance 8: pp. 557–565.

Matutes, C. & Vives, X., 2000. "Imperfect competition, risk taking and regulation in banking." European Economic Review 44: pp. 1–34.

Mishkin, F. S., 1999. "Financial consolidation: Dangers and opportunities." Journal of Banking and Finance 23: pp. 675–691.

Moeltner, K. & Woodward, R. (2009). "Meta-Functional Benefit Transfer for Wetland Valuation: Making the Most of Small Samples." Environmental & Resource Economics, 42(1): pp. 89–108.

Panzar, J. C. & Rosse, J. N., 1987. "Testing for monopoly equilibrium." Journal of Industrial Economics 35: pp. 443–456.

Pearson, K., 1904. "Report on certain enteric fever inoculation statistics." British Medical Journal 2(2288): pp. 1243–1246.

Raftery, A. E., Madigan, D. & Hoeting, J. A., 1997. “Bayesian model averaging for linear regression models.” Journal of the American Statistical Association 92: pp. 179–191.

Reed, W. R., Florax, R. J. G. M., & Poot, J., 2015. “A Monte Carlo Analysis of Alternative Meta-Analysis Estimators in the Presence of Publication Bias.” Economics e-journal discussion paper 2015-9.

Rosenthal, R., 1979. “The ‘file drawer problem’ and tolerance for null results.” Psychological Bulletin 86: pp. 638–641.

Schaeck, K. & Cihak, M., 2014. “Competition, efficiency, and stability in banking.” Financial Management 43: pp. 215–241.

Schaeck, K. & Cihak, M., 2012. “Banking competition and capital ratios.” European Financial Management 18(5): pp. 836–866.

Schaeck, K., Cihak, M. & Wolfe, S., 2009. “Are competitive banking systems more stable?” Journal of Money, Credit and Banking 41(4): pp. 711–734.

Stanley, T. D, 2001. “Wheat from chaff: Meta-analysis as quantitative literature review.” Journal of Economic Perspectives 15(3): pp. 131–150.

Stanley, T. D., 2008. “Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection.” Oxford Bulletin of Economics and Statistics 70(1): pp. 103–127.

Stanley, T. D. & Doucouliagos, H., 2012. “Meta-regression analysis in economics and business.” Routledge Advances in Research Methods.

Stanley, T. D. & Doucouliagos, H., 2015. “Neither Fixed nor Random: Weighted Least Squares Meta-Analysis.” Statistics in Medicine, forthcoming.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S. & Rost, K., 2013. “Meta-analysis of economics research reporting guidelines.” Journal of Economic Surveys 27(2): pp. 390–394.

Turk Ariss, R., 2010. “On the implications of market power in banking: Evidence from developing countries.” Journal of Banking and Finance 34(4): pp. 765–775.

Uhde, A. & Heimeshoff, U., 2009. “Consolidation in banking and financial stability in Europe: Empirical evidence.” Journal of Banking and Finance 33: pp. 1299–1311.

Valickova, P., Havranek, T. & Horvath, R., 2014. “Financial development and economic growth: A meta-analysis.” Journal of Economic Surveys, forthcoming.

Yeyati, E. L. & Micco, A., 2007. “Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk.” Journal of Banking and Finance 31: pp. 1633–1647.

Appendix A: BMA Diagnostics

Table A4.1: Summary of BMA estimation, baseline estimation

Mean no. regressors	Draws	Burnins	Time	No. models visited
16.7873	2.00E+06	1.00E+06	8.946665 mins	428100
Modelspace 2^K	% visited	% Topmodels	Corr PMP	No. Obs.
3.4e+10	0.0012	85	0.9991	598
Model Prior		g-Prior	Shrinkage-Stats	
uniform / 17.5		UIP	Av=0.9983	

Notes: In this specification we employ the priors suggested by Eicher et al. (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure A4.1: Model size and convergence, baseline estimation

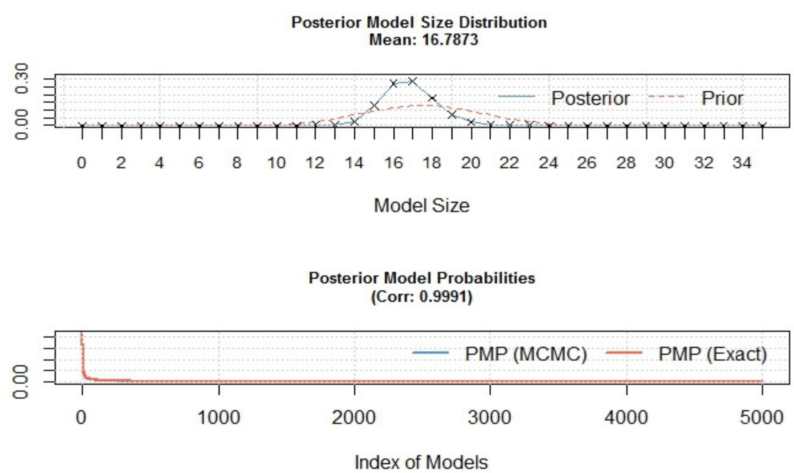


Table A4.2: Summary of BMA estimation, alternative priors

Mean no. regressors	Draws	Burnins	Time	No. models visited
15.9075	2.00E+06	1.00E+06	9.343995 mins	340418
Modelspace 2^K	% visited	% Topmodels	Corr PMP	No. Obs.
3.4e+10	0.00099	92	0.9991	598
Model Prior		g-Prior	Shrinkage-Stats	
random / 17.5		BRIC	Av=0.9992	

Notes: The “random“ model prior refers to the beta-binomial prior used by Ley & Steel (2009): the prior model probabilities are the same for all possible model sizes. In this specification we set Zellner’s g prior in line with Fernandez et al. (2001).

Figure A4.2: Model size and convergence, alternative priors

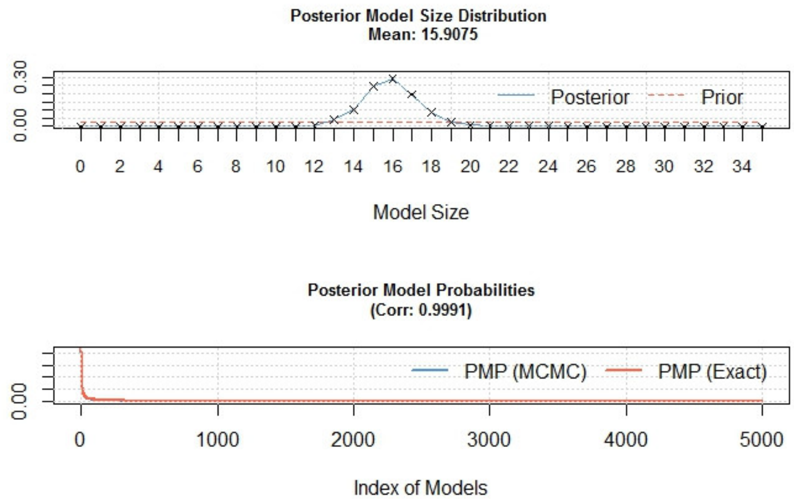
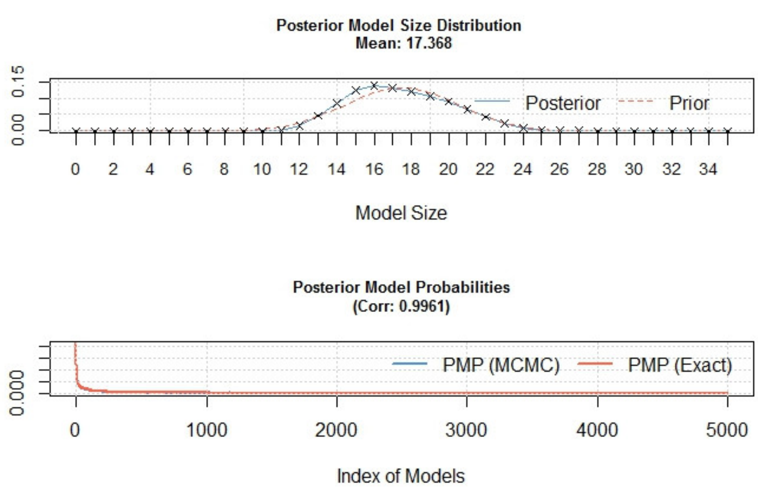


Table A4.3: Summary of BMA estimation, unweighted regressions

Mean no. regressors	Draws	Burnins	Time	No. models visited
17.3680	2.00E+06	1.00E+06	9.077281 mins	543559
Modelspace 2^K	% visited	% Topmodels	Corr PMP	No. Obs.
3.4e+10	0.0016	69	0.9961	598
Model Prior		g-Prior	Shrinkage-Stats	
uniform / 17.5		UIP	Av=0.9983	

Notes: In this specification we employ the priors suggested by Eicher et al. (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure A4.3: Model size and convergence, unweighted regressions



Chapter 5

Management Board Composition of Czech Banking Institutions and Bank Risk: The Random Forest Approach

Abstract

The paper investigates how the management board composition of banking institutions affects their risk in the Czech Republic. For this purpose, we build a unique data set comprising selected biographical information on the management board members of Czech financial institutions holding a banking license over the 2001-2012 period and combine it with individual bank financial data. We apply a machine learning technique – the random forest – to identify the best predictors of bank risk and further interpret the model output. We find non-linear relationships between average directors' age, average director tenure, the proportion of directors holding an MBA and the proportion of non-national directors and the observed bank risk proxies. As for average directors' age, it appears to impact bank stability very little beyond a certain threshold. Decreases in average director tenure on board are found to reduce bank stability while increases in tenure enhance stability. In terms of directors' education, large increases in the proportion of directors with an MBA enhance bank profit volatility. Furthermore, if the majority of directors on board are foreigners, bank risk, captured by profit volatility and the NPL ratio, increases substantially.

I would like to thank Kamil Galuscak, Tomas Havranek, Michal Hlavacek, Petr Jakubik, Evzen Kocenda, Jitka Lesanovska, Ornella Ricci, Borek Vasicek, and Tracy Xu for their helpful comments on the previous versions of this paper. The previous version of this paper was released as Czech National Bank Working Paper 2015/14 and as Charles University Institute of Economic Studies Working Paper 2016/2.

5.1 Introduction

The recent global crisis put financial stability and financial supervision research in the spotlight. In 2009, the OECD Steering Group on Corporate Governance (Kirkpatrick, 2009) highlighted the need to pay special attention to commercial bank corporate governance issues. They concluded that "the financial crisis can be to an important extent attributed to failures and weaknesses in corporate governance arrangements. When they were put to a test, corporate governance routines did not serve their purpose to safeguard against excessive risk-taking in a number of financial services companies." This aspect of financial supervision has been supported by the Basel Committee on Banking Supervision (BCBS), which has drawn attention to the need to study, understand, and improve the corporate governance of financial entities. The BCBS (2006) especially advocates studies of a governance structure composed of a board of directors and senior management.

In the Czech Republic, the Act on Banks 21/1992 governs the organizational structure of financial entities holding a banking license. This legislation requires banks to implement policies that ensure diversity in the members of governing bodies, for example, in their profiles and backgrounds, views, and sets of competencies. Such diversity can lead to a wider pool of resources and expertise, generating more discussion, more monitoring, and more challenges in the boardroom, as stated in the European Commission's 2011 Green Paper. In particular, the European Commission (2010) seeks to improve existing corporate governance practices, i.e., the functioning, composition, and skills of commercial banks' boards of directors.

Following these endeavors, this paper focuses on investigating the effect of the management board composition of commercial banks on bank risk-taking behavior in the Czech Republic over the 2001-2012 period. Specifically, the paper aims to examine if and how commercial bank management boards affect bank risk-taking in terms of board size, the average age of directors, director tenure, the proportion of female directors, director education level, and the proportion of non-national directors.

To the best of the author's knowledge this is the first study of the economic effects of bank management board composition using a machine learning approach – the random forest. The random forest is a non-parametric data mining technique that allows for modelling nonlinear relationships between explanatory variables and the dependent. It can be applied to small data sets with a large number of predictors since it is insensitive to outliers, robust to adding new observations and to overfitting (Breiman, 2001). In addition, relatively recent developments in the random forest analysis – variable contributions and partial dependence analysis – increase transparency of this machine learning technique and enable interpretation of its results that is comparable to regression models.

The composition of commercial bank boards and its risk-taking implications are not sufficiently explored in the corporate governance literature. To the author's knowledge, the only other studies to have addressed this issue are those by Berger et al. (2014) with a focus on Germany, by Pathan (2009), and by Erkens et al. (2012). However, two of these studies – Pathan (2009) and Erkens et al. (2012) - use market-based proxies for bank risk-taking which are not applicable to many transition countries of the CEE region, whose banks are not commonly listed on stock exchanges. Moreover, most studies focus on advanced countries, while relatively little is known about the corporate governance structure and its role in the banking sectors of emerging economies. So far, relatively few studies (Adams and Mehran, 2008; Caprio et al., 2007; Levine, 2004) have focused on corporate governance issues in banks, even though core aspects of corporate governance can be applied to them.

As for the use of machine learning techniques in corporate governance analysis, Liang et al. (2016) and Wu (2010) apply data mining algorithms to predict corporate bankruptcy using selected corporate governance and financial indicators of Taiwanese firms while Pai et al. (2011) use support vector machines to detect management fraud in Taiwan.

Focusing on Czech banking institutions, the analysis is performed for bank management boards in a system of corporate governance with two-tier boards. In two-tier systems, the management board, chaired by the CEO, runs the corporation and reports to the supervisory board. The supervisory board, on the other hand, performs a monitoring role equivalent to that of non-managing directors in the one-tier system found in Anglo-Saxon countries. The supervisory board thus appoints and dismisses members of the management board on behalf of the shareholders. Members of the supervisory board cannot simultaneously hold positions on the board of directors, and vice versa. The two-tier system thus allows for clear separation between inside directors, who run the bank and hold positions on the board of directors, and outside directors, i.e., members of the supervisory board. According to the literature, this board design has risk-taking implications. Adams and Ferreira (2007) found that increasing board independence in a one-tier system makes a CEO less likely to disclose information to non-managing directors, thereby hindering their involvement in management decisions. This, in turn, results in less well informed top management decisions and has direct consequences for risk-taking. However, in two-tier systems the CEO does not face this trade-off in disclosing information and, because shareholders' interests are aligned with those of the supervisory board, the monitoring of managing directors is more intensive and leads to less risk-taking (Berger et al., 2014).

This paper aims to reveal a more efficient management board composition in terms of risk-taking in the Czech banking sector with potential implications for the financial sector stability.

The paper is structured as follows. Section 5.2 describes the development and specific features of the Czech banking sector and formulates our research hypotheses. Section 5.3 builds the data set for investigating the research question and presents overview of board composition variables. Section 5.4 describes the methodology applied, Section 5.5 presents our findings, and the last section concludes.

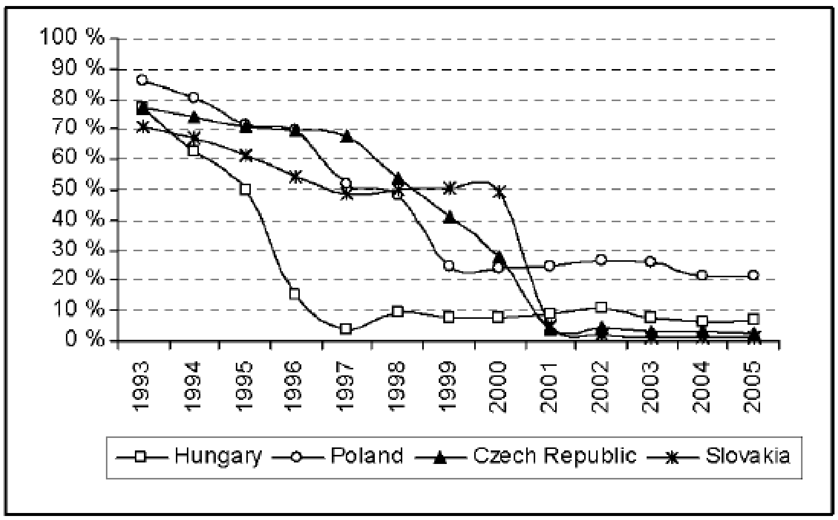
5.2 Czech Banking Sector and Research Hypotheses

The current commercial banking sectors in the Visegrad Four countries, i.e., Hungary, Poland, the Czech Republic, and Slovakia, emerged following the breakup of the state bank (monobank) system combined with the issuing of licenses to new banks. At the start of the transformation process, a two-tier banking system had to be created, with the central bank ensuring macroeconomic stability - and in the Czech case also supervision of commercial banks - and commercial banks contributing to efficient credit allocation. The Czech Republic, along with other post-communist countries, faced problems that made the transformation process difficult: (i) no managerial and supervisory know-how; (ii) no market history of potential lenders; (iii) great uncertainty regarding the outcome of entrepreneurial projects; (iv) inherited bad loans; and (v) no adequate legal framework and regulation (Tuma, 2002).

After the two-tier banking system was formed in 1990, the large Czech banks were transformed into joint-stock companies in 1992 and partially privatized. Nevertheless, the state kept controlling stakes in these banks until the late 1990s. Banking licenses were granted quite freely to newly created banks in the early 1990s and the market was opened to foreign bank branches. This led to a fast increase in the number of banks during this time period.

During the period of economic boom of 1994 - 1996, triggered by inflows of foreign short-term capital and subsequent growth of the money supply, serious problems started to emerge in the sector of small banks due to bad loans and other balance sheet weaknesses. The economic recession in 1997 - 1998 worsened the excessive credit risk that Czech banks had taken on owing to their poor corporate governance (Tuma, 2002). At the end of 1999, non-performing loans constituted more than 40% of the loans granted by large banks, while the same indicator for small Czech-owned banks even exceeded 50%.

Figure 5.1: Proportion of State Control in the Visegrad Four Countries



Notes: Figure 5.1 shows the evolution of the proportion of state control in banks in the Visegrad Four countries as measured by the asset share of banks owned by the state. Source: Kocenda et al. (2007)

During the later stages of the transformation process in the second half of the 1990s, the share of foreign owners in the equity capital of Czech banks grew sharply. The new shareholders of Czech banks are foreign banks based mostly in Belgium, France, and Austria. The state is currently involved in only two banks specializing in government programs in the areas of export promotion and support for small businesses. The overall evolution of bank privatization in the Czech Republic and the other Visegrad Four countries is summarized in Figure 5.1. Figure 5.1 shows the proportion of state control in banks as measured by the asset share of banks owned by the state. The Czech Republic managed to achieve full banking privatization by 2001, as observed by Kocenda et al. (2007).

As a result of the banking sector transformation and consolidation process there are 23 institutions that are holders of a banking license granted by the Czech National Bank in the Czech Republic as of beginning 2016. Moreover, almost 97% of the Czech banking sector's balance sheet assets are controlled by foreigners according to Financial Stability Report 2011/2012 (CNB Financial Stability Department, 2012).

Next, for our analysis we rely on the precondition that the composition of a bank's top management team affects corporate decision-making and, in turn, corporate outcomes, as supported, for example, by Graham et al. (2013) and Adams and Ferreira (2009). Direct evidence that personal traits affect financial outcomes has been presented by Kaplan et al. (2012) who based on data from CEO candidate interviews conclude that their abilities and execution skills are related to subsequent firm performance. This allows for empirical examination of the research question in this paper. In particular, we focus on the following aspects to assess the effect of management board composition on bank risk-taking behavior:

1. *Average Age of Directors*

Empirical evidence suggests a negative relationship between age and risk-taking, as given by Campbell (2006) for investment behavior, Bucciol and Miniaci (2011) for households' risk attitudes, and Sahm (2007) and Grable et al. (2009) based on survey evidence.

2. *Proportion of Female Directors*

There are two contrasting outlooks on how women affect economic outcomes. First, women are more risk averse than men in financial decision-making. This finding is supported by Jianakoplos and Bernasek (1998), Sunden and Surette (1998), and Agnew et al. (2003). Furthermore, women being less overconfident than men makes them less prone to making poor investment decisions, as shown by Barber and Odean (2001), Niederle and Vesterlund (2007), and Goel and Thakor (2008).

Second, in the corporate governance literature, female directors are, however, more likely to take risks than men (Adams and Funk, 2012). A number of studies show that female directors execute excessive monitoring, which reduces shareholder value (Almazan and Suarez, 2003; Adams and Ferreira, 2007), and make poorer investment decisions, as they face greater obstacles than men in gathering information (Bharath et al., 2009). Owing to the dual effect of women on risk-taking in the literature, both effects of female director representation in management boards - increasing as well as reducing risk-taking - should be investigated.

The effect of female representation in boards on economic outcomes is currently of particular interest due to the adoption of legislative measures regulating female board representation in some European countries (e.g. Norway, France, the Netherlands, and Belgium).

3. Education Level

There is a dual effect of directors' educational background on corporate risk behavior. First, the survey by Graham and Harvey (2001) shows that directors holding an MBA employ sophisticated valuation techniques more than directors without such a degree. These sophisticated valuation methods should reduce the risks to the firm.

Second, directors with an MBA are also shown to be more aggressive and employ riskier firm policies (Bertrand and Schoar, 2003). Following Berger et al. (2014), who found a risk-reducing effect of directors with a PhD, we also focus in our analysis on the effect of directors holding a PhD on bank risk. As there are no directors holding both a PhD and an MBA in our sample, this allows us to check if managing directors holding different degrees affect bank riskiness differently. Overall, both the risk-reducing and risk-increasing effect of education on corporate risk-taking should be examined.

4. Proportion of Non-national Directors

The literature typically finds a positive effect of foreign directors on firm performance, as foreign directors might bring new technology and modern managerial techniques into the firm (Oxelheim and Randoy, 2003). On the other hand, Masulis et al. (2012) find that foreign independent directors can provide valuable international expertise and advice to firms but could weaken the board's monitoring and disciplining role. The European Commission's 2010 Green Paper (European Commission, 2010) shares this outlook, as it finds that "some interviewed companies highlighted the importance of foreign board members for international companies while others underlined the difficulties deriving from different cultural backgrounds and languages." Therefore, we hypothesize that foreign directors can either reduce bank riskiness via the modern managerial techniques and better skills they bring into the bank, or increase bank risk due to their unfamiliarity with local market or banking sector specificities and due to the obstacles they face in overcoming cultural and language barriers in the boardroom.

5. Board Size

There is a dual outlook in the corporate governance literature on the number of directors on management boards, i.e., board size. On the one hand, larger boards potentially offer more experience and knowledge and better advice (Dalton et al., 1999) as well as assigning more people to supervision. On the other hand, boards with too many directors face considerable problems with coordination, communication, and decision-making (Lipton and Lorsch, 1992; Jensen, 1993). Greater difficulty in achieving compromises in large decision-making groups results in bigger boards adopting less extreme decisions (e.g. Nakano and Nguyen, 2012). This leads to the hypothesis that larger boards are associated with lower corporate risk-taking.

6. Director Tenure

There is again a dual outlook in the literature on the impact of director tenure on firm performance and, by extension, on firm risk as one of the attributes of firm performance. Huang (2013) finds that board tenure can be positively or negatively related to firm value depending on firm characteristics. In more complex firms with greater advisory needs, board members are more likely to require more time to gain sufficient knowledge to perform appropriate strategic decision-making. Consequently, the quality of board advice and expertise increases over time, with positive implications for firm performance. However, as the effect of board tenure is determined by the trade-off between the marginal benefits of learning and the marginal costs of entrenchment, Huang (2013) also finds that the marginal costs of entrenchment might quickly dominate over the benefits of learning in firms with greater monitoring needs. This implies decreasing firm value with increasing board tenure. However, in the one-tier board system McNulty et al. (2012) find that firm financial risk-taking decreases when tenure of executive directors is significantly greater than that of non-executives.

5.3 Data Set and Descriptive Statistics

To investigate the effect of management board composition on risk-taking, we need to combine two types of data sets. The first data set is prepared by the author from the annual reports of 21 Czech institutions holding a banking license granted by the Czech National Bank[1]. This data set is unique and includes selected information on banks' management board members. In particular, we collect data on the average age of directors, the size of the management board, the average length of time directors hold their positions, the proportion of female directors, the proportion of directors holding a

[1] The remaining two banks, which are also holders of banking licenses, are excluded from the analysis, as, unlike other commercial banks, they primarily serve government schemes in the areas of export support and assistance for small businesses. Moreover, they are state-controlled and, as such, management board decisions in these banks might be motivated by other factors than those in their privately-owned counterparts.

PhD or an MBA, and the proportion of non-national directors[2]. The management board characteristics evolution is presented in Subsection 5.3.3.

The second data set contains financial data on individual banks extracted primarily from the Bankscope database. As described in Section 5.2, the 1990s were a turbulent time for the Czech Republic, characterized by banking privatization and consolidation of the banking sector. Moreover, by 2001 full banking privatization had been achieved (Kocenda et al., 2007) and the Czech banking sector had gained its current defining characteristics, for example, in terms of being almost exclusively owned by foreign investors (Tuma, 2002; CNB Financial Stability Department, 2012). For the reasons given above, and to control for potential bank survivor bias, the combined data set covers the period of 2001 - 2012.

5.3.1 Bank Risk Measures

In order to analyze how management board composition affects bank risk-taking we use three conventional indicators of bank risk, that is, the Z-score, the NPL ratio and profit volatility which can be derived from bank financial statements. Moreover, the use of these indicators will make our results consistent with most studies dealing with board composition issues, as performance indicators extracted from financial reports are used abundantly in the literature.

The Z-score has been frequently used to analyze the determinants of bank risk-taking in the pre-crisis period (e.g. Laeven and Levine, 2009; Foos et al., 2010; Altunbas et al., 2012; Demirguc-Kunt and Huizinga, 2010). Moreover, the measure has been widely used to capture bank stability in studies investigating the relationship between bank competition and financial stability – Agoraki et al. (2011), Anginer et al. (2014), Berger et al. (2009), Nicolo and Loukoianova (2007), and Cihak and Hesse (2010), to mention the most prominent ones. The Z-score indicates how many standard deviations in the return on assets a bank is away from insolvency and, by extension, the likelihood of failure:

$$Z - score_{i,t} = \frac{ROA_{i,t} + E_{i,t}/TA_{i,t}}{sROA_{i,t}}, \tag{5.1}$$

[2] Despite the evidence provided by Minton et al. (2014) on the importance of directors' financial expertise in bank risk-taking, our analysis does not consider this director characteristic due to data limitations.

where i takes values from bank 1 to bank 21 and t indicates a year from 2001 to 2012. $ROA_{i,t}$ captures the return on assets of bank i at time t, $E_{i,t}/TA_{i,t}$ measures the ratio of a bank's equity capital to its total assets, and $sROA_{i,t}$ measures the volatility of a bank's return on assets calculated as a three-year moving average.

Another popular risk proxy is the ratio of non-performing loans to total bank loans (the NPL ratio). This is a measure of credit quality with regard to banks' lending practices. Similarly to the Z-score, the NPL ratio is used abundantly as a fragility indicator in the bank competition-stability literature - see, for example, Cihak and Schaeck (2012), Agoraki et al. (2011), Yeyati and Micco (2007), and Berger et al. (2009). Nevertheless, the NPL ratio only covers credit risk and cannot be directly linked to the likelihood of bank failure (Beck, 2008).

Next, the volatility of the return on assets (sROA), calculated as a three-year moving average, is also used as a proxy for bank risk. This measure of individual bank distress focuses on bank profitability, in particular on the volatility of bank profits, and is frequently used in the literature along with other indicators of bank risk, i.e., the Z-score and the NPL ratio (Beck et al., 2013; Cihak and Schaeck, 2012; Uhde and Heimeshoff, 2009; Liu et al., 2012).

5.3.2 Bank Control Variables

To estimate the effect of management board composition on bank risk, we also need to control for individual bank characteristics in our analysis, by including the following variables:

First, bank size, expressed as the ratio of a bank's total assets to the Czech banking sector's total assets, accounts for the fact that larger banks have a greater capacity to absorb risk and that some banks are too big to fail. Therefore, a positive relation is expected between bank size and risk-taking.

Second, the logarithm of total assets is added to account for asset growth in first differences. In times of fast asset growth, banks are characterized by a different amount of risk-taking.

Third, according to Keeley (1990) incentives to take risks are reduced if a bank has a large charter value. Charter value can be defined as the future economic rents a bank can obtain from its access to markets that are to a large extent protected from competition. Hutchison and Penacchi (1996) show that the ratio of demand deposits to total deposits is a good proxy for a bank's charter value. A negative relation is expected between risk-taking and charter value.

Fourth, the share of Tier I capital in total capital, calculated as the ratio of Tier I capital to Tier I and Tier II capital, is also included, as well as the capital adequacy ratio (CAR), calculated as the ratio of Tier I and Tier II capital to risk-weighted assets. Since capital increases monitoring and reduces moral hazard incentives (Morrison and White, 2004; Allen et al., 2011), a negative relation is expected between Tier I capital share and risk-taking.

Fifth, balance sheet indicators, customer loans to total assets and off-balance sheet items to total assets, are to be considered, too. The effect of loan exposure is risk-increasing whereas off-balance sheet items have ambiguous effect on risk. As shown by Dionne and Triki (2005) hedging by means of off-balance sheet items is risk-reducing. In contrast, off-balance sheet items can be seen as alternative risky investments and as such have risk-increasing impact.

Sixth, a merger dummy that takes a value of one if the bank engaged in a merger and zero otherwise should be included, as mergers often coincide with board composition changes. In addition, we control for different bank categories by business model (general commercial bank/building society), by size (large/small and medium sized) and by capitalization (well capitalized/adequately capitalized) by means of dummy variables.

Seventh, to incorporate the time aspect, a numeric time control is included.

Eighth, we include Czech GDP growth to account for state of the economy in our analysis[3].

Last, the parent bank's risk appetite needs to be accounted for in the analysis, as almost 97% of the Czech banking sector's balance sheet assets are controlled by foreigners (CNB Financial Stability Department, 2012). This control assumes that there is a link between the riskiness of the foreign parent bank and its Czech affiliate. It is measured in the same way as domestic bank risk-taking to keep the analysis consistent.

[3] We have tried including 2 additional macroeconomic control variables - interest rate spread between long-term and short-term Czech government bonds and Czech unemployment rate - which are often considered in the literature on bank risk. We do not report these results since adding the additional macroeconomic controls neither changes the main findings nor do these variables appear among the most useful predictors of bank risk.

The final data set is of annual frequency. Table 5.1 provides an overview of the data and lists the sources for each variable.

Table 5.1: Overview of Variables in the Data Set

Variable	Expected sign	Description	Source
Risk measures			
NPLL		Share of non-performing loans in total loans	Bankscope
logZ		Logarithm of Z-score (profitability and capitalization over volatility of profits, calculated over 3-year period)	Bankscope
sROAA		3-year ROA volatility	Bankscope
Board variables			
BoardsizeL1	+/-	Lag of number of directors on management board	Annual reports
AvrageL1	-	Lag of average age of directors	Annual reports
AvrboardtenL1		Lag of average number of years over which directors hold their positions on board	Annual reports
SharefemL1	+/-	Lag of proportion of female directors on board	Annual reports
SharePhDL1	-	Lag of proportion of directors with PhD on board	Annual reports
ShareMBAL1	+/-	Lag of proportion of directors with MBA on board	Annual reports
ShareforeignL1	+/-	Lag of proportion of foreign directors on board	Annual reports
Control variables			
TAg	+	Growth rate of total bank assets	Bankscope
Banksize	+	Share of bank's total assets in banking sector's total assets	Bankscope
Charterval	-	Bank's demand deposits over total deposits, used as proxy for charter value	Bankscope
Tier1	-	Share of Tier I capital in bank’s capital	ICD
CAR		Capital adequacy ratio, calculated as sum of Tier I and Tier II capital to risk-weighted assets	ICD
CusloansTA	+	Customer loans over bank’s total assets	Bankscope
OffBSTA	+/-	Off-balance sheet items over bank’s total assets	Bankscope
MergerDummy		equals 1 if bank engaged in a merger in given year	Annual reports
Dbank		equals 1 if the institution is general commercial bank	
DSS		equals 1 if the institution is building society	
Dlarge		equals 1 if Banksize exceeds 75th percentile	

Dsmall	equals 1 if Banksize is below 75th percentile	
Dwellcap	equals 1 if Tier1 exceeds 50th percentile	
Dlow	equals 1 if Tier1 is below 50th percentile	
time	Numeric control for time dimension	
GDPg	Annual Czech GDP growth rate	ARAD
Mother bank risk measures		
mNPLL	Parent bank's share of non-performing loans in total loans	Bankscope
logmZ	Parent bank's Z-score (profitability and capitalization over volatility of profits, calculated over 3-year period)	Bankscope
msROAA	Parent bank's 3-year ROA volatility	Bankscope

Notes: The expected signs should be reversed for the Z-score, as this is a proxy for bank stability. Equation 5.1 provides the definition of the Z-score. ICD = the Czech National Bank's internal regulatory information database, ARAD = the Czech National Bank's statistics database. All variables were transformed to ensure stationarity.

5.3.3 Overview of the Czech Banking Sector and Management Board Characteristics

Czech banking institutions consist of general commercial banks and building societies, a specialized type of banks that concentrate on gathering savings for home construction purposes and providing loans for new home construction and renovation and whose product receives state support. In the Czech Republic there were 5 building societies and 16 general commercial banks as of beginning 2016 that we include in the sample. The state is involved in the remaining two banks, which serve specific government schemes and are thus excluded from our analysis.

As for the corporate board variables in our data set, we observe that average director age and tenure increased over 2001 - 2012, while average board size decreased over the same period. However, for Czech building societies board size did not change much on average. Overall, the proportion of women on management boards fell, with the exception of general commercial banks, for which this proportion fluctuates over time. As for the education level of directors, the proportion of directors holding a PhD on the management boards of all banking institutions and general commercial banks decreased, whereas there were no directors with a PhD on the boards of building societies at any time over the sample period. On the other hand, the proportion of directors with an MBA rose over time for the entire sector and for building societies. In general commercial banks, the proportion of directors holding an MBA appears to be similar at the sample end to that in 2001. The proportion of non-national directors decreased over time in general commercial banks while falling more dramatically in building societies over 2001 - 2012.

5.4 Methodology

5.4.1 Endogeneity, Regression Trees and Random Forest

Endogeneity is a frequent problem in corporate governance (Hermalin and Weisbach, 2003). In our case, not only does board composition affect risk-taking, but the reverse implication (risk-taking affecting management board composition) might also be an issue. Therefore, to prevent endogeneity in our analysis we lag all director characteristics by one period, i.e. one year.

To explain bank risk-taking we employ a nonparametric machine learning algorithm – the random forest – which consists of ensembles of decision trees. A classification and regression tree (CART) is a partitioning algorithm that recursively identifies the indicators and the respective thresholds which are able to best split the sample into the relevant classes or numerical values (for regression problems) in predictive data modelling (Breiman et al., 1984). The output of the predictive model is a tree-like structure (see Figure A5.5 in Appendix), grown on a subset of the board and control variables presented in Section 5.3. A tree has one root node, two branches exiting from each parent node and multiple terminal nodes (leaves). The algorithm considers all available explanatory variables and threshold levels and selects such a variable and threshold that results in the two purest subsamples by minimizing the mean square error. MSE, the mean square error, is a standard statistical concept that measures the average of the squares of the errors between the random forest prediction and the actual values of the risk-taking measure, and is defined as follows:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}\left(\widehat{y_{i,t}} - y_{i,t}\right)^2, \tag{5.2}$$

where i takes values from bank 1 to bank 21, t indicates a year from 2001 to 2012 and N is the number of observations of the predicted bank risk measure.

After the first best split of the data is selected, the model further divides the two subsamples by finding the best split at each subsequent node. The algorithm stops when either a stopping rule is implemented (for example, a minimal terminal node is set) or when splitting does not further improve performance of the model, that is there are no further decreases in the MSE. As a result, a decision tree is robust to outliers since these are isolated in leaves.

The downside of implementing a single decision tree approach is its lack of robustness when adding additional variables or observations to the dataset. Furthermore, a single tree if allowed to grown fully might overfit the data. To correct for these disadvantages we use the random forest technique proposed by Breiman (2001) that aggregates many trees. The random forest also benefits from robustness against outliers and does not necessitate any distributional assumptions (such as normal distribution of predictors). In his original paper on random forests Breiman (2001) shows random forests do not overfit data due to the Law of Large Numbers which can be satisfied by building a sufficiently large number of trees. He gives the following principles for growing of individual trees within the random forest:

1. If the number of cases in the training set is N, sample N cases at random - but with replacement, from the original data. This sample is the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

Point 1 is enabled by default in the Sci-kit implementation of the random forest in Python that we use. We address point 2 by optimizing the maximum number of variables to be considered for splitting at each node of every tree. We search for the optimal m over the interval $\langle 2, \max\_var \rangle$, where max_var is the maximum number of independent variables entering the model. We also adhere to point 3 and let trees grow fully.

Breiman (2001) shows that random forest error rate depends on the correlation between any two trees in the forest and on the strength of individual trees. Increasing the correlation between trees increases the forest error rate while increased strength of individual trees, which manifests by a low error rate of a single tree, decreases the forest error rate.

With respect to point 2 reducing m reduces both the correlation and the strength while increasing it increases both. Breiman (2001) suggests that an optimal range of m

exists somewhere in between and needs to be found. Furthermore, he claims that m is the only adjustable parameter to which random forest is somewhat sensitive.

We optimize for the number of trees in the forest and for the maximum number of variables considered for splitting at each node m by means of a six-fold cross-validation. Cross-validation is a machine learning technique that splits the entire sample into six equally-sized subsamples, fits the model on five of these subsamples and assesses the model's performance on the last subsample that is held out. This serves as another safeguard against overfitting and preserves number of observations in small samples since it does not necessitate splitting the sample into separate training and validation sets. Table A5.2 in the Appendix presents results of the random forest optimization for all three bank risk proxies and Figures A5.2 - A5.4 show the random forest error rate as a function of the number of trees.

Another advantage of the random forest is that it allows us to identify the most useful variables for prediction of the risk-taking measure which can be derived from the loss of model accuracy (that is, MSE deterioration) when values of one variable are permuted between instances (Breiman, 2001; Svetnik et al., 2003; Liaw and Wiener, 2002). This predictor space reduction property of the random forest is especially useful for small datasets with a relatively large number of explanatory variables that cannot be meaningfully included within a regression.

To interpret results of a nonparametric approach such as the random forest model we calculate variable contributions of the selected useful variables for prediction in the random forest, as proposed by Kuz'min et al. (2011). Variable contributions are calculated separately for each observation and provide information about the direction (positive/negative) and magnitude of influence of a given variable on the model prediction.

The decision tree structure – each node being assigned a value, a variable being associated with every decision, and contribution of decisions along the path in a tree to the final outcome – allows us to compute variable contributions to the model prediction. Unlike regression, the CART prediction can be decomposed into bias and variable contributions on the predicted instance level. Since the random forest is an ensemble of trees, its prediction equals the average of predictions of individual trees and can be decomposed as follows:

$$prediction_{RF}(y) = \frac{1}{J}\sum_{j=1}^{J} bias_j(y) + \left(\frac{1}{J}\sum_{j=1}^{J} contr_j(1,y) + \cdots + \frac{1}{J}\sum_{j=1}^{J} contr_j(n,y)\right), \tag{5.3}$$

where $prediction_{RF}(y)$ is the random forest predicted value of y, $\frac{1}{J}\sum_{j=1}^{J} bias_j(y)$ is the average of bias term of individual trees j in the forest, and the term in brackets is the sum of averages of each variable contribution from individual trees.

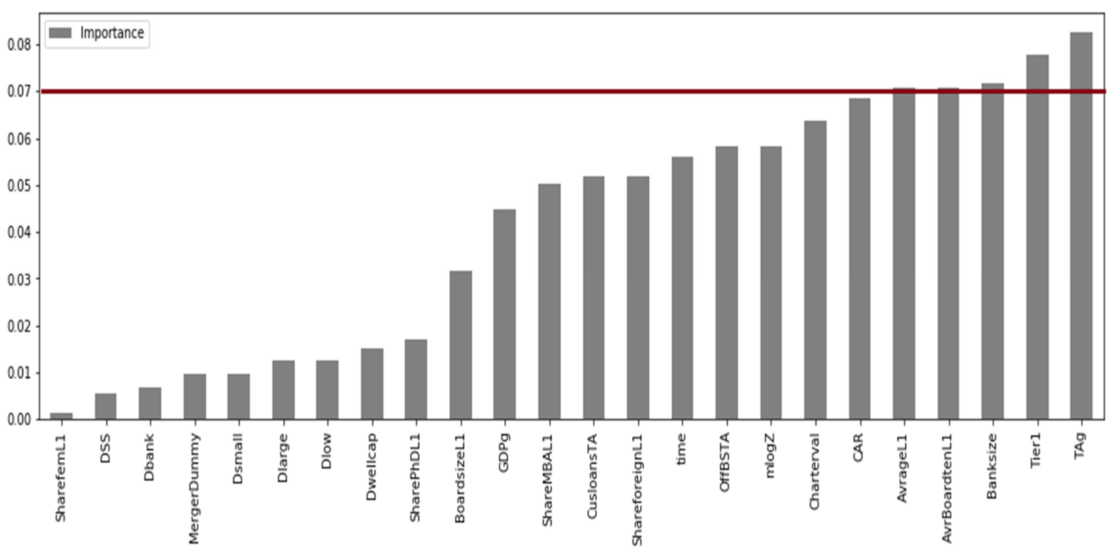
In our analysis, we calculate average contributions of selected variables for our sample overall, that is, by averaging the model predictions over each bank-year observation and obtaining the corresponding mean variable contributions.

5.5 Results

5.5.1 Variable Selection

First, we grow the random forest for each bank risk proxy on the entire set of board and control variables as presented in Table 5.1. This allows us to rank the explanatory variables by their importance with respect to model accuracy and select the most useful predictors on which we subsequently grow the resulting random forest model. Figure 5.2 shows the variable ranking by importance for bank stability measured by Z-score.

Figure 5.2: Random Forest Variable Importances for Z-score Prediction



Notes: The red line represents the variable selection threshold of 0.07.

We aim to select at least four most useful variables to explain bank riskiness and for this reason we set the importance threshold equal to 0.07. The predictors whose importance exceeds this threshold were retained and used for growing the resulting random forest on the parsimonious set of variables. For all three dependents the random forest assigns to 2-3 independent variables much higher importance score than to the rest, that is, the marginal decrease in importance between the 2-3 variables ranked first and the rest is substantial. However, to include more than two predictors into the model we lower the threshold to 0.07 in case of Z-score and NPL ratio, and to 0.065 for profit volatility which ensures the resulting 5, 5 and 4 predictors, respectively.

By means of variable importance, we identify average age of directors, average director tenure, bank size, total bank asset growth and Tier I as useful predictors for bank stability measured by Z-score. In the same spirit, we find that proportion of directors with MBA, proportion of non-national directors, average director tenure and total bank asset growth are useful for explaining bank profit volatility. As for the NPL ratio, proportion of non-national directors, total bank asset growth, Tier I, GDP growth and the NPL ratio of the parent institution emerged as best predictors.

However, there is a potential caveat to identifying the most important variables by means of variable importance, as extracted from the random forest. Since random forests are biased in favor of variables with more levels (that is, with a large number of values) on which to split at individual nodes, the variable importance scores from random forest would not be reliable for datasets with many categorical variables. Our dataset contains 7 dummy variables altogether to control for bank mergers, type of banking institution, level of bank capitalization, and bank size. For this reason, the importance score of these dummy variables might be underestimated and their ranking pushed back in Figure 5.2.

5.5.2 Variable Contributions Analysis

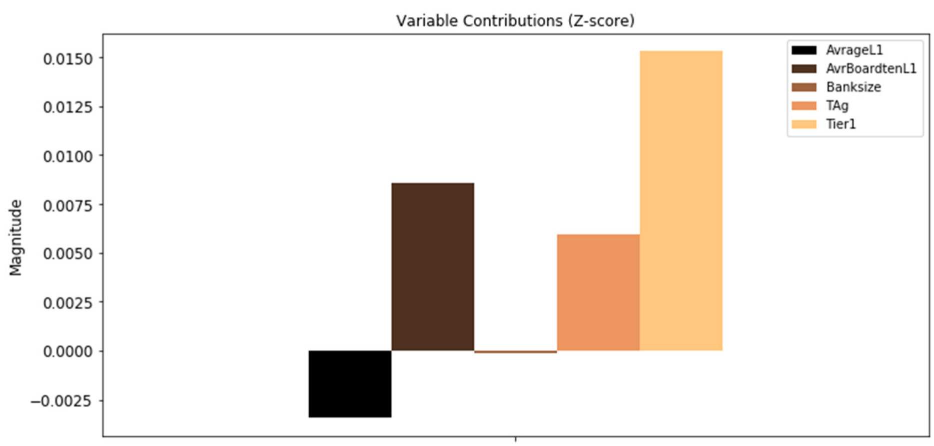
Figure 5.3 below plots variable contributions of the selected useful predictors from Subsection 5.1 to the random forest output for all bank-year observations on average.

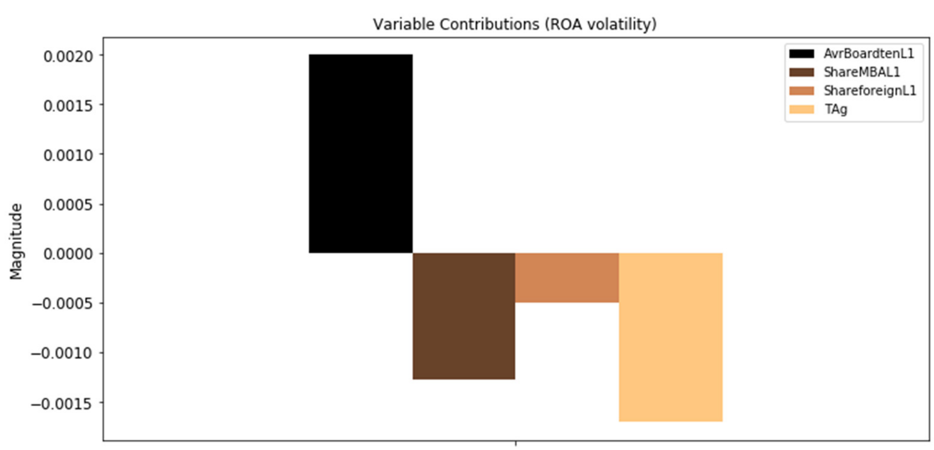
For the Czech banking sector overall the random forest prediction decomposition shows that average directors' age contributes to Z-score prediction negatively on average while average director tenure contributes to banks' stability positively on average. This result contradicts the evidence commonly found in the literature that with increasing director age a firm performance improves. However, further analysis in Subsection 5.5.3 uncovers a non-linear relationship between bank stability and average directors' age.

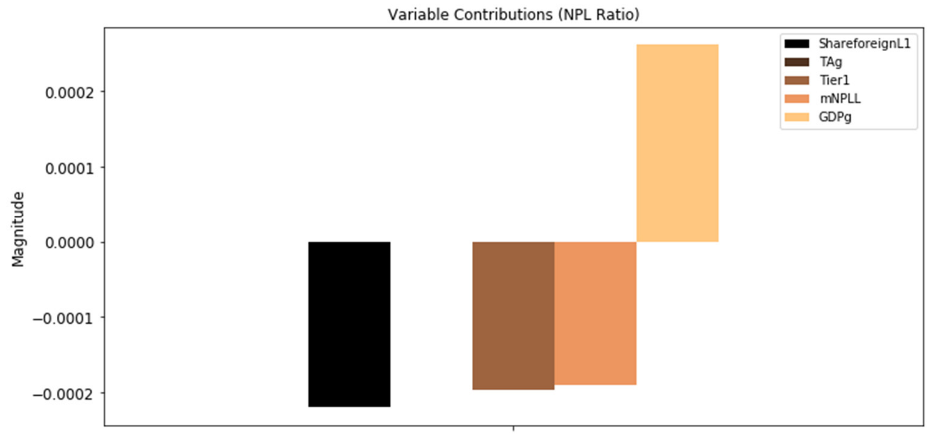
Next, proportion of directors with MBA on board contributes to bank profit volatility in a negative way. This finding corroborates the evidence that directors holding an MBA employ sophisticated valuation methods and thus reduce risks to the firm (e.g. Graham and Harvey, 2001). Moreover, foreign directors on board also appear to contribute to profit volatility negatively on average, confirming the finding in the literature that foreign directors bring new expertise into the firm which improves its performance (Oxelheim and Randoy, 2003). On the other hand, average time for which directors hold their seats on board appears to increase profit volatility on average which is in contrast with positive average contribution of director tenure to bank stability. These findings highlight the dual outlook of the impact of director tenure on firm performance found in the literature (Huang, 2013).

As for NPL ratio, larger proportions of non-national directors increase bank loan portfolio quality (that is, by decreasing NPL ratio). This is in line with decreasing average effect of foreign directors on bank profit volatility.

Figure 5.3: Variable Contributions to the Random Forest Prediction – Banking Sector Overall







Notes: Columns represent direction and magnitude of average contributions of individual identified predictors to bank risk, measured by Z-score, profit volatility and NPL ratio, respectively.

As for the control variables, growth of total bank assets favorably contributes to Z-score prediction on average while it also decreases profit volatility. However, growth of bank assets has no average overall effect on NPL ratio. This confirms the hypothesis that larger banks can absorb risks more easily. Capitalization, captured by Tier I ratio, contributes to bank stability in a favorable manner on average while it also decreases NPL ratio. It needs to be kept in mind, however, that the discussed variable contributions only indicate in which direction and by how much individual variables contribute to prediction of the bank risk measures on average over all bank-year observations. They do not provide us with an insight into modelled pairwise dynamics of independent variables vis-à-vis the dependent.

5.5.3 Partial Dependence Analysis

While in the previous subsection we analyzed average contributions of variables to average random forest prediction over all bank-year instances, now we turn to the analysis of the dynamic relationships between identified useful corporate board variables for bank risk prediction and the bank risk measure itself. For this purpose, we use partial dependence plots, a tool first introduced by Friedman (2001) that tracks how the value of a single predictor influences the random forest bank risk predictions after the influence of all the other predictors in the model has been averaged out. For example, in case of linear regression models, the resulting plots would be straight lines with slopes equal to model parameters.

Figure 5.4 shows non-linear nature of the relationships between board variables and bank risk proxies. Average directors' age below approximately 41 years has a stability-reducing effect in the Czech banking institutions. However, with increasing average directors' age the risk-enhancing effect levels out and affects bank stability very little. Our finding is thus generally in line with the hypothesis of increasing risk aversion with age presented in the literature (Campbell, 2006; Bucciol and Miniaci, 2011), although the relationship is not linear.

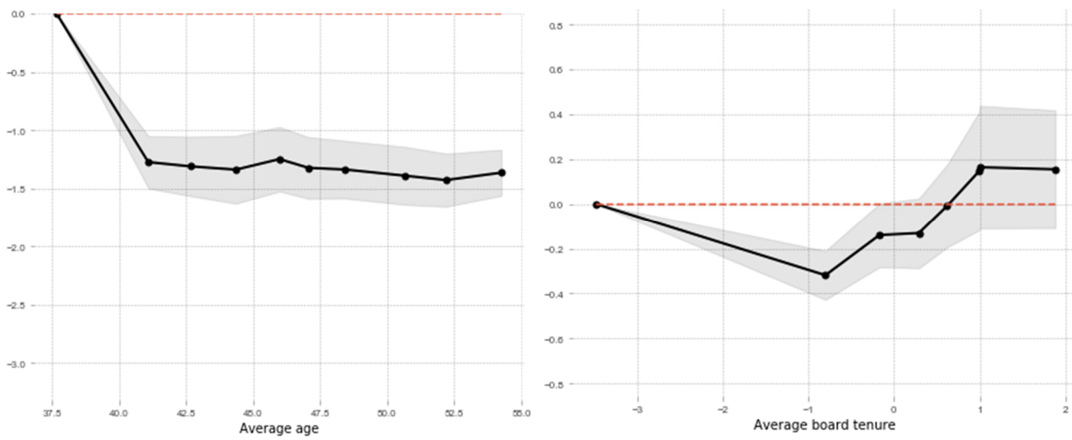
As for average director tenure, decreases in board tenure reduce bank stability whereas increases in tenure affect bank stability favorably. The result is in line with Huang (2013) who finds that the effect of board tenure is determined by the trade-off between the marginal benefits of learning and the marginal costs of entrenchment and depends on the firm complexity. In the same vein, decreases in director tenure on board increase bank risk, as measured by profit volatility, while moderate increases in tenure lower profit volatility.

Next, decreases in the proportion of directors with an MBA do not affect profit volatility while small negative changes and small positive changes in their proportion appear to decrease ROA volatility. On the other hand, larger increases in the number of directors with an MBA enhance profit volatility. This finding seems to encompass both effects of directors with an MBA in the literature – moderate numbers of such directors reduce riskiness due to their knowledge of superior valuation techniques while too many directors holding an MBA on board appear to act more aggressively and employ riskier firm policies.

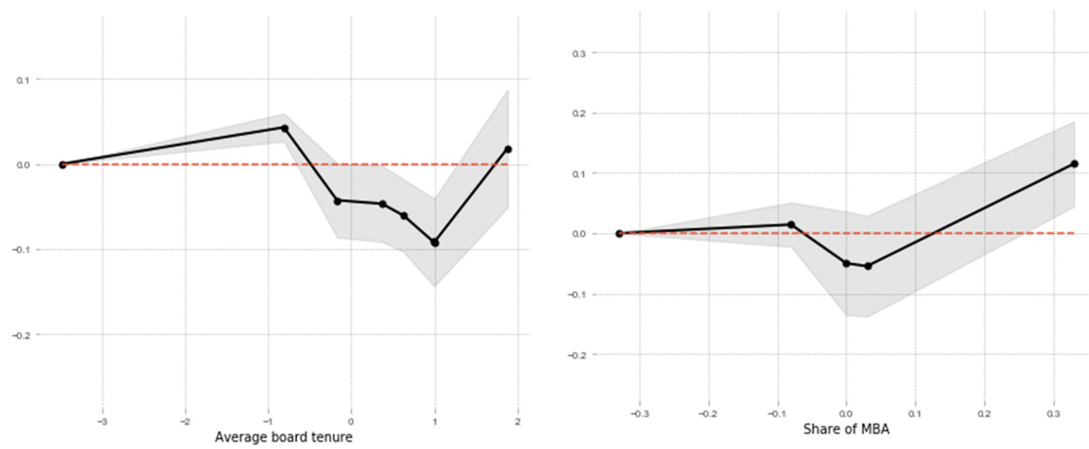
Finally, small proportions of non-national directors on Czech bank boards do not impact riskiness meaningfully. However, it appears that when about half of directors on board are foreigners, bank risk, measured by profit volatility and NPL ratio, starts to increase substantially. The result supports the evidence presented in the European Commission's 2010 Green Paper that some companies experience difficulties derived from foreign directors' different cultural backgrounds and languages.

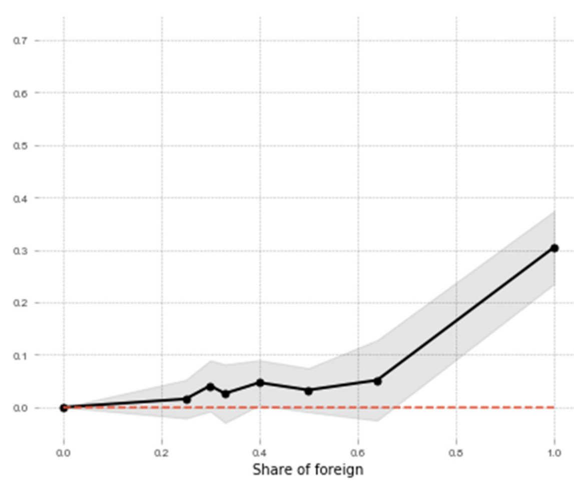
Figure 5.4: Partial Dependent Plots for Board Variables

(A) Dependent: Z-score

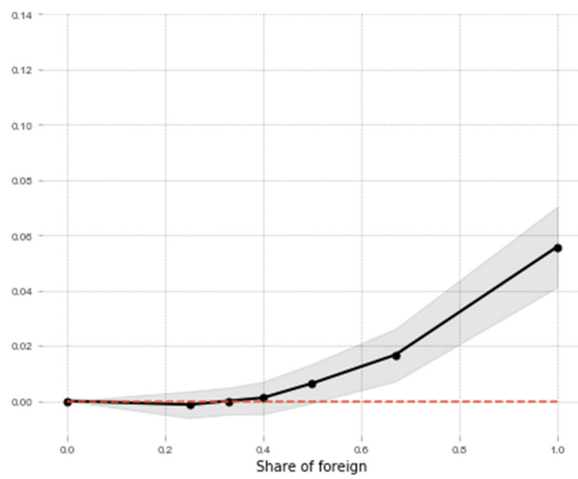


(B) Dependent: ROA volatility





(C) Dependent: NPL ratio



Notes: Figures in Panel A depict how Z-score changes with changes in average director age and average director tenure. Panel B shows the dynamics between profit volatility and average director tenure, proportions of directors holding an MBA, and non-national directors, respectively. Panel C illustrates the dependence of NPL ratio on different proportions of non-national directors on board. Red dashed line marks the zero line and the shaded area around the curve highlights its standard deviation confidence band.

Overall, while the feature contributions analysis in Subsection 5.5.2 provided an insight into the magnitude and direction of influence of each variable on average random forest prediction over all cross-sections and years in the sample, the partial dependence plots are informative of the pairwise dynamics between board variables and bank risk on the model level. As such, partial dependence plots allow for interpretation of the random forest model predictions which is comparable to that of linear regression models.

Conclusions

In this paper, we investigate how the management board composition of banking institutions affects risk-taking behavior in the Czech Republic using the random forest, a machine learning technique. To perform the analysis, we prepare a unique data set that comprises selected biographical information on the management board members of Czech banking institutions and combine it with individual bank financial data to serve as control variables.

First, we grow the random forest on the entire set of board and control variables in order to identify the most useful predictors for the three bank risk proxies with respect to model accuracy. Next, we built the random forest on the sets of the identified best predictors for each bank risk measure and interpret the model predictions in terms of individual variable average contributions to the random forest outcome, and by means of partial dependence plots.

For the Czech banking sector over the 2001-2012 period we find non-linear relationships between average directors' age, average director tenure, the proportion of directors holding an MBA and the proportion of non-national directors and their respective bank risk proxies.

We find support for the hypothesis of increasing risk aversion with age in the Czech banking sector, although the director age beyond a certain threshold appears to impact bank stability very little.

Next, decreases in average director tenure on board are found to reduce bank stability while increases in tenure enhance stability. This corroborates the view that quality of board advice and expertise increases over time, once new directors gained sufficient knowledge of the firm to perform appropriate decision-making.

As for directors' education, large increases in the proportion of directors with an MBA enhance bank profit volatility. The effect of small positive and negative changes in their proportion on ROA volatility is nonlinear. The finding thus captures both risk-increasing and risk-reducing implications of directors holding an MBA in the literature.

Last but not least, we present evidence that when majority of directors on board are foreigners, bank risk, captured by profit volatility and the NPL ratio, increases substantially. This can be linked to overcoming differences arising from different cultural backgrounds and languages.

We do not find that the remaining collected director biographical data, that is, gender, PhD-level education and the number of managing directors on board, are important predictors of bank risk measures in our random forest framework.

Before we conclude it should be mentioned that in addition to directors' characteristics other factors, such as form and magnitude of managers' compensation (in the form of deferred remuneration or benefits), might have non-negligible impact on firm performance. In particular, Cassell et al. (2012) suggest that CEOs with large inside debt holdings prefer investment and financial policies that are less risky. However, due to sensitivity of directors' remuneration statistics and its limited availability in the Czech context, including this control into our analysis would be very problematic at present. Nevertheless, extending the presented analysis with data on directors' remuneration would be an interesting avenue for further research.

References

Adams, R. B. & D. Ferreira (2007): "A Theory of Friendly Boards." Journal of Finance 62(1): pp. 217-250.

Adams, R. B. & D. Ferreira (2009): "Women in the boardroom and their impact on governance and performance." Journal of Financial Economics 94(2): pp. 291-309.

Adams, R. B. & P. Funk (2012): "Beyond the Glass Ceiling: Does Gender Matter?" Management Science 58(2): pp. 219-235.

Adams, R. B. & H. Mehran (2008): "Corporate performance, board structure, and their determinants in the banking industry." Staff Reports 330, Federal Reserve Bank of New York.

Agnew, J., P. Balduzzi, & A. Sunden (2003): "Portfolio Choice and Trading in a Large 401(k) Plan." American Economic Review 93(1): pp. 193-215.

Agoraki, M.-E. K., M. D. Delis, & F. Pasiouras (2011): "Regulations, competition and bank risk-taking in transition countries." Journal of Financial Stability 7(1): pp. 38-48.

Allen, F., E. Carletti, & R. Marquez (2011): "Credit Market Competition and Capital Regulation." Review of Financial Studies 24(4): pp. 983-1018.

Almazan, A. & J. Suarez (2003): "Entrenchment and Severance Pay in Optimal Governance Structures." Journal of Finance 58(2): pp. 519-548.

Altunbas, Y., S. Manganelli, & D. Marques-Ibanez (2012): "Bank Risk during the Financial Crisis: Do business models matter?" Working Papers 12003, Bangor Business School, Prifysgol Bangor University (Cymru/Wales).

Anginer, D., A. Demirguc-Kunt, & M. Zhu (2014): "How does competition affect bank systemic risk?" Journal of Financial Intermediation 23(1): pp. 1-26.

Barber, B. M. & T. Odean (2001): "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." The Quarterly Journal of Economics 116(1): pp. 261-292.

Basel Committee on Banking Supervision (2006): "Enhancing Corporate Governance for Banking Organizations." Staff reports, Bank for International Settlements.

Beck, T. (2008): "Bank competition and financial stability: friends or foes?" Policy Research Working Paper Series 4656, The World Bank.

Beck, T., O. De Jonghe, & G. Schepens (2013): "Bank competition and stability: Cross-country heterogeneity." Journal of Financial Intermediation 22(2): pp. 218-244.

Berger, A., L. Klapper, & R. Turk-Ariss (2009): "Bank Competition and Financial Stability." Journal of Financial Services Research 35(2): pp. 99-118.

Berger, A. N., T. Kick, & K. Schaeck (2014): "Executive board composition and bank risk taking." Journal of Corporate Finance 28(C): pp. 48-65.

Bertrand, M. & A. Schoar (2003): "Managing With Style: The Effect Of Managers On Firm Policies." The Quarterly Journal of Economics 118(4): pp. 1169-1208.

Bharath, S. T., M. Narayanan, & H. Seyhun Nejat (2009): "Are women executives disadvantaged?" University of Michigan.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984): "Classification and Regression Trees." Wadsworth Statistics/Probability Series. Belmont, CA.

Breiman, L. (2001): "Random forests." Machine Learning 45 (1): pp. 5–32.

Bucciol, A. & R. Miniaci (2011): "Household Portfolios and Implicit Risk Preference." The Review of Economics and Statistics 93(4): pp. 1235-1250.

Campbell, J. Y. (2006): "Household Finance." Journal of Finance 61(4): pp. 1553-1604.

Caprio, G., L. Laeven, & R. Levine (2007): "Governance and bank valuation." Journal of Financial Intermediation 16(4): pp. 584-617.

Cassell, C. A., S. X. Huang, J. M. Sanchez, M. D. Stuart (2012): "Seeking safety: The relation between CEO inside debt holdings and the riskiness of firm investment and financial policies." Journal of Financial Economics 103(3): pp. 588-610.

Cihak, M. & H. Hesse (2010): "Islamic Banks and Financial Stability: An Empirical Analysis." Journal of Financial Services Research 38(2): pp. 95-113.

Cihak, M. & K. Schaeck (2012): "Competition, efficiency, and stability in banking." Financial Management 43(1): pp. 215-241.

CNB Financial Stability Department (2012): CNB Financial Stability Report 2011/2012. Czech National Bank. ISBN 978-80-87225-38-7.

Dalton, D. R., C. M. Daily, J. L. Johnson, & A. E. Ellstrand (1999): "Number of directors and financial performance: A meta-analysis." The Academy of Management Journal 42(6): pp. 674-686.

Demirguc-Kunt, A. & H. Huizinga (2010): "Are banks too big to fail or too big to save? International evidence from equity prices and CDS spreads." Policy Research Working Paper Series 5360, The World Bank.

Dionne, G. & T. Triki (2005): "Risk Management and Corporate Governance: the Importance of Independence and Financial Knowledge for the Board and the Audit Committee." Cahiers de recherche 0515, CIRPEE.

Erkens, D. H., M. Hung, & P. Matos (2012): "Corporate governance in the 2007-2008 financial crisis: Evidence from financial institutions worldwide." Journal of Corporate Finance 18(2): pp. 389-411.

European Commission (2010): "Green Paper: Corporate governance in financial institutions and remuneration policies." Staff report.

European Commission (2011): "Green Paper: The EU corporate governance framework." Staff report.

Foos, D., L. Norden, & M. Weber (2010): "Loan growth and riskiness of banks." Journal of Banking & Finance 34(12): pp. 2929-2940.

Friedman, J.H. (2001): "Greedy Function Approximation: A Gradient Boosting Machine." Annals of Statistics 29: pp. 1189-1232.

Goel, A. M. & A. V. Thakor (2008): "Overconfidence, CEO Selection, and Corporate Governance." Journal of Finance 63(6): pp. 2737-2784.

Grable, J. E., S. McGill, & S. Britt (2009): "Risk tolerance estimation bias: the age effect." Journal of business and economics research 7(7): pp. 1-12.

Graham, J. R. & C. R. Harvey (2001): "The theory and practice of corporate finance: evidence from the field." Journal of Financial Economics 60(2-3): pp. 187-243.

Graham, J. R., C. R. Harvey, & M. Puri (2013): "Managerial attitudes and corporate actions." Journal of Financial Economics 109(1): pp. 103-121.

Hermalin, B. E. & M. S. Weisbach (2003): "Boards of directors as an endogenously determined institution: a survey of the economic literature." Economic Policy Review (Apr): pp. 7-26.

Huang, S. Z. (2013): "Zombie board: Board tenure and firm performance." Technical report, Research Collection School of Accountancy Singapore Management University.

Hutchison, D. E. & G. G. Pennacchi (1996): "Measuring Rents and Interest Rate Risk in Imperfect Financial Markets: The Case of Retail Bank Deposits." Journal of Financial and Quantitative Analysis 31(03): pp. 399-417.

Jensen, M. C. (1993): "The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems." Journal of Finance 48(3): pp. 831-80.

Jianakoplos, N. A. & A. Bernasek (1998): "Are Women More Risk Averse?" Economic Inquiry 36(4): pp. 620-30.

Kaplan, S. N., Klebanov, M. M. & M. Sorensen (2012): "Which CEO Characteristics and Abilities Matter?" The Journal of Finance 67(3): pp. 973-1007.

Keeley, M. C. (1990): "Deposit Insurance, Risk, and Market Power in Banking." American Economic Review 80(5): pp. 1183-1200.

Kirkpatrick, G. (2009): "The Corporate Governance Lessons from the Financial Crisis." Staff report, OECD Steering Group on Corporate Governance.

Kocenda, E., J. Hanousek, & P. Ondko (2007): "The Banking Sector in New EU Member Countries: A

Sectoral Financial Flows Analysis." Czech Journal of Economics and Finance (Finance a uver) 57(5-6): pp. 200-224.

Kuz'min, V.E., Polishchuk, P.G., Artemenko, A.G. & Andronati, S.A. (2011): "Interpretation of qsar models based on random forest methods." Molecular Informatics 30(6-7): pp. 593-603.

Laeven, L. & R. Levine (2009): "Bank governance, regulation and risk taking." Journal of Financial Economics 93(2): pp. 259-275.

Levine, R. (2004): "The Corporate Governance of Banks - a concise discussion of concepts and evidence." Policy Research Working Paper Series 3404, The World Bank.

Liang, D., Lu, C.-C., Tsai, C.-F. & Shih, G.-A. (2016): "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study." European Journal of Operational Research 252 (2): pp. 561-572.

Liaw, A., Wiener, M. (2002): "Classification and regression by random Forest." R News 2 (3): pp. 18–22.

Lipton, M. & J. W. Lorsch (1992): "A Modest Proposal for Improved Corporate Governance." Business Lawyer 48: pp. 59-77.

Liu, H., P. Molyneux, & L. H. Nguyen (2012): "Competition and risk in South East Asian commercial banking." Applied Economics 44(28): pp. 3627-3644.

Masulis, R. W., C. Wang, & F. Xie (2012): "Globalizing the boardroom - The effects of foreign directors on corporate governance and firm performance." Journal of Accounting and Economics 53(3): pp. 527-554.

McNulty, T., Florackis, C. & P. Ormrod (2012): "Corporate Governance and Risk: A Study of Board Structure and Process." Research Report 129, The Association of Chartered Certified Accountants.

Minton, B. A., J. Taillard, & R. Williamson (2014): "Financial expertise of the board, risk taking, and performance: Evidence from bank holding companies." Journal of Financial and Quantitative Analysis 49(2): pp. 351-380.

Morrison, A. & L. White (2004): "Crises and Capital Requirements in Banking." CEPR Discussion Papers 4364, C.E.P.R. Discussion Papers.

Nakano, M. & P. Nguyen (2012): "Board size and corporate risk-taking: Further evidence from Japan." Corporate Governance: An International Review 20(4): pp. 369-387.

Nicolo, G. D. & E. Loukoianova (2007): "Bank ownership, market structure, and risk." Proceedings 1058, Federal Reserve Bank of Chicago.

Niederle, M. & L. Vesterlund (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?" The Quarterly Journal of Economics 122(3): pp. 1067-1101.

Oxelheim, L. & T. Randoy (2003): "The impact of foreign board membership on firm value." Journal of Banking & Finance 27(12): pp. 2369-2392.

Pai, P.-F., Hsu, M.-F. & Wang, M.-C. (2011): "A support vector machine-based model for detecting top management fraud." Knowledge-Based Systems 24 (2): pp. 314-321.

Pathan, S. (2009): "Strong boards, CEO power and bank risk-taking." Journal of Banking & Finance 33(7): pp. 1340-1350.

Sahm, C. R. (2007): "Stability of risk preference." Finance and Economics Discussion Series 2007-66, Board of Governors of the Federal Reserve System (U.S.).

Sunden, A. E. & B. J. Surette (1998): "Gender Differences in the Allocation of Assets in Retirement Savings Plans." American Economic Review 88(2): pp. 207-11.

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P. & Feuston, B.P. (2003): "Random forest: a classification and regression tool for compound classification and QSAR modeling." Journal of Chemical Information and Computer Sciences 43: 1947-1958.

Tuma, Z. (2002): "Banking sector development in the Czech Republic." In "East-West Conference 2002, Austrian National Bank, Vienna".

Uhde, A. & U. Heimeshoff (2009): "Consolidation in banking and financial stability in Europe: Empirical evidence." Journal of Banking & Finance 33(7): pp. 1299-1311.

Wu, W.-W. (2010): "Beyond business failure prediction." Expert Systems with Applications 37 (3): pp. 2371-2376.

Yeyati, E. L. & A. Micco (2007): "Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk." Journal of Banking & Finance 31(6): pp. 1633-1647.

Appendix

Table A5.1: Summary Statistics

Variable	No. of obs.	Mean	St. Dev.	Min	Max
Risk measures					
NPLL	133	0.05	0.05	0.00	0.25
logZ	169	3.93	1.06	1.51	7.45
sROAA	169	0.33	0.41	0.00	2.92
Board variables					
BoardsizeL1	167	4.18	1.49	2.00	9.00
AvrageL1	177	45.71	5.36	35.75	62.67
AvrBoardtenL1	163	0.24	1.13	-3.83	1.88
SharefemL1	163	0.00	0.07	-0.33	0.33
SharePhDL1	163	0.00	0.05	-0.33	0.25
ShareMBAL1	163	0.00	0.10	-0.33	0.33
ShareforeignL1	177	0.34	0.27	0.00	1.00
Control variables					
Banksize	168	0.00	0.01	-0.12	0.02
TAg	168	0.15	0.22	-0.12	2.31
Charterval	147	-0.01	0.12	-0.56	0.41
Tier1	158	0.91	0.11	0.60	1.33
MergerDummy	188	0.10	0.30	0.00	1.00
Dbank	252	0.76	0.43	0.00	1.00
DSS	252	0.24	0.43	0.00	1.00
Dlarge	252	0.19	0.39	0.00	1.00
Dsmall	252	0.81	0.39	0.00	1.00
Dwellcap	252	0.29	0.45	0.00	1.00
Dlow	252	0.71	0.45	0.00	1.00
CusloansTA	156	0.02	0.06	-0.15	0.16
OffBSTA	149	0.00	0.07	-0.24	0.57
CAR	172	-0.01	0.09	-0.53	0.49
time	252	6.50	3.46	1.00	12.00
GDPg	231	2.66	3.21	-4.77	6.81
Mother bank risk measures					
mNPLL	139	0.07	0.09	0.00	0.76
mlogZ	194	3.62	1.07	1.03	5.61
msROAA	175	-0.03	0.98	-3.01	3.28

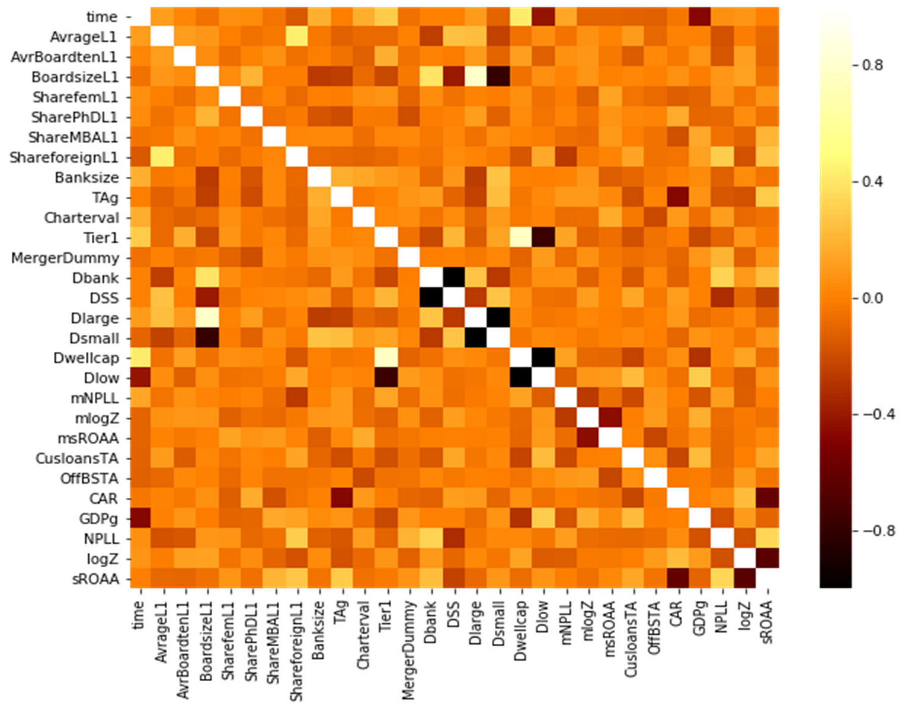
Notes: Descriptive statistics of the variables. Table 5.1 presents definitions of the variables and lists their sources.

Table A5.2: Random Forest Optimization Results

Random forest on all variables			
Dependent	*Z-score*	*NPL*	*sROAA*
6-fold cross-validation MSE	0.95	0.001	0.05
Max. variables for splitting	2	11	5
No. of trees	110	70	80
Random forest on identified useful variables			
Dependent	*Z-score*	*NPL*	*sROAA*
6-fold cross-validation MSE	0.88	0.001	0.05
Max. variables for splitting	2	2	2
No. of trees	60	200	120

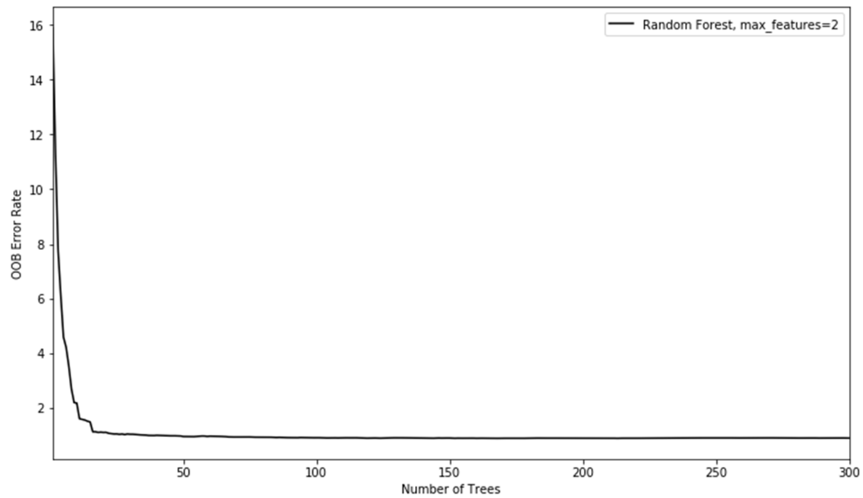
Notes: Optimization statistics from six-fold cross-validation performed on the full set of independent variables and on the subset of identified best predictors. The full list of independent variables can be found in Table 5.1. MSE stands for mean squared error.

Figure A5.1: Variable Correlations



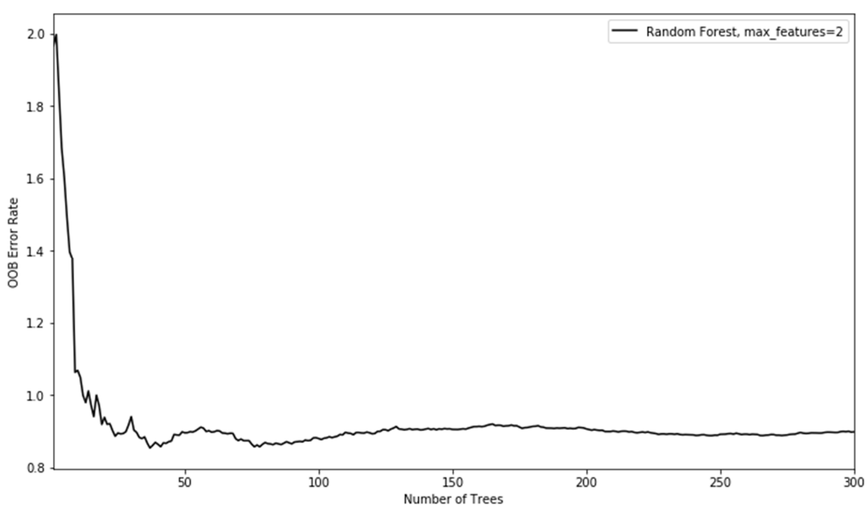
Notes: Correlations between variable pairs as well as between individual predictors and dependent variables. The legend on the right indicates sign and magnitude of pairwise correlations by color. Table 5.1 presents definitions of the variables and lists their sources.

Figure A5.2: Random Forest Error – Z-score



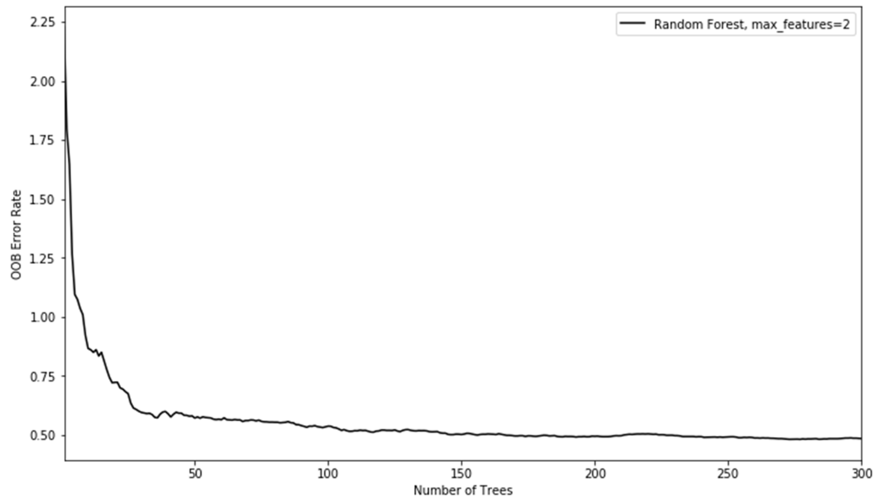
Notes: Error rate of the random forest as a function of number of trees grown on the subset of identified best predictors of Z-score. Maximum number of variables considered for splitting at each node is 2.

Figure A5.3: Random Forest Error – ROA volatility



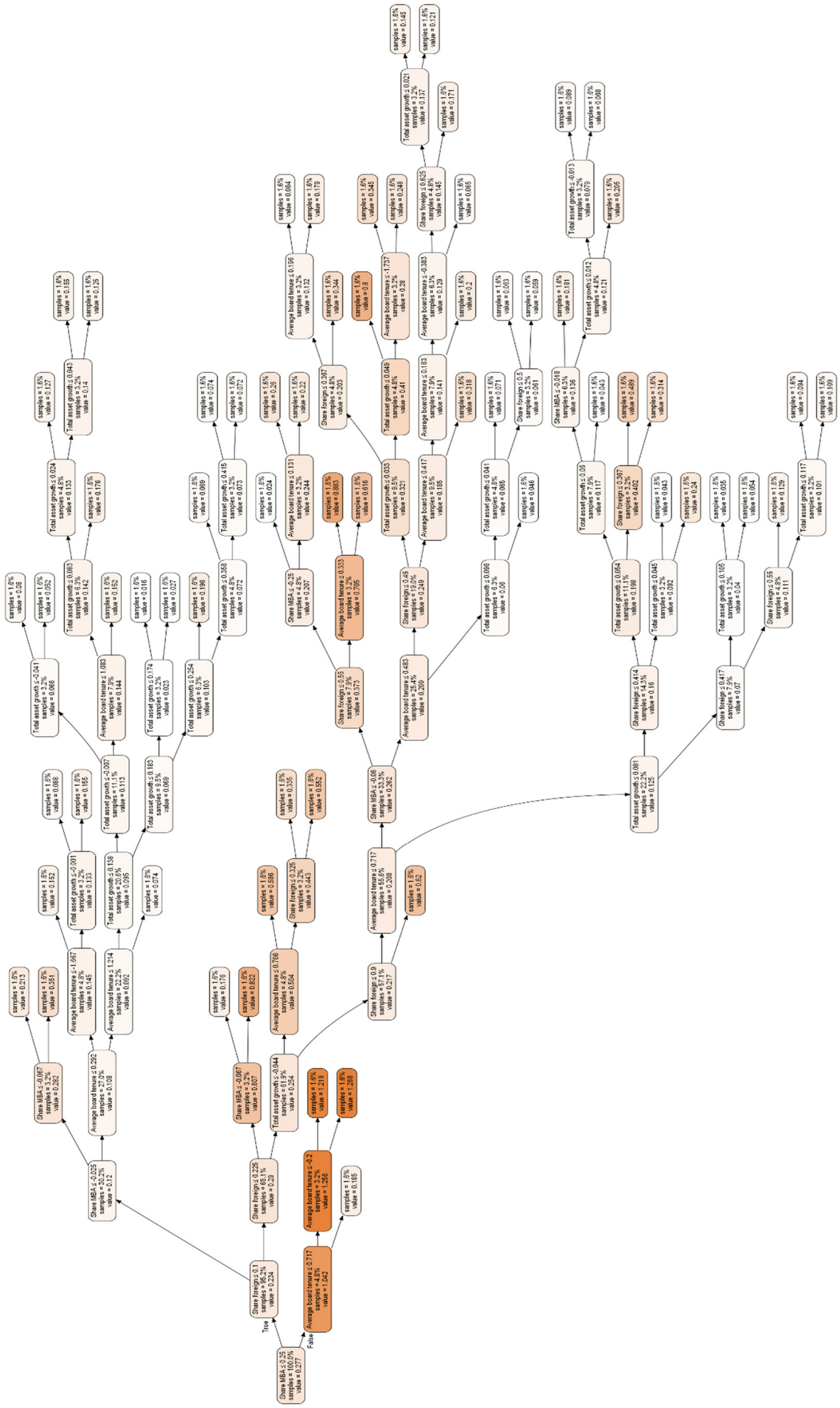
Notes: Error rate of the random forest as a function of number of trees grown on the subset of identified best predictors of ROA volatility. Maximum number of variables considered for splitting at each node is 2.

Figure A5.4: Random Forest Error – NPL ratio



Notes: Error rate of the random forest as a function of number of trees grown on the subset of identified best predictors of NPL ratio. Maximum number of variables considered for splitting at each node is 2.

Figure A5.5: A Decision Tree



Notes: The first decision tree of the random forest grown on identified best predictors of ROA volatility. Each node has an assigned value and partitions data into two subsamples based on a decision rule (True/False). The darker the color of the node the smaller the subset of data the node partitions.

Response to Referees

I would like to thank all the referees for their helpful and insightful comments that improved this dissertation and lead to its more cohesive form. The referees' comments are written in italics while my responses are in roman. The response is organized by referee.

Comments by Prof. RNDr. Jiří Witzany, Ph.D.

1. *I recommend to outline the applied BMA and MCMC method in more detail in the first or second chapter of the thesis. The BMA is described in the second paper and not in the first where it is already used. The details of the MCMC sampler are not disclosed at all, the description can be probably found in the references. Nevertheless, the reader should not be left guessing what the exact design of the MCMC that can be obviously set up in different ways is.*

The MCMC sampler used is the standard birth-death sampler that is applied in the majority of BMA routines (Feldkircher and Zeugner, 2015). A paragraph explaining the details on this sampler type was added in Section 2.3. In addition, the sampler type was also specified in subsection 3.3.2. Furthermore, chapters 2 and 3 were swapped in order to ensure the reader's comfort in regard to the applied BMA methodology.

2. *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 the continuous indicators 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 2.1B introducing unnecessary additional noise into the model.*

There are two main reasons for using extreme values of a continuous financial stress measure, an FSI. First, FSIs allow researchers to exactly identify start and end dates of high systemic stress episodes since crises databases typically provide only a year (or years) when financial crises took place without specifying an exact month/quarter of their onset and end. Second, policymakers need to be able to monitor systemic stress in the economy with a certain regularity which is done by constructing composite financial stress indices. Discretization of such a continuous measure, an FSI, thus allows for prompt policy actions without having to wait for a true binary crisis indicator to be available via a crises database. This explanation was also added to Section 3.2.

3. *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 as at least 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.*

Yes, there are different approaches to variable selection and identification of the underlying model, the simplest of which is stepwise variable selection. Nevertheless, the stepwise selection procedure is not statistically valid and suffers from several drawbacks (Koop, 2003). Its main issue is one-at-a-time or sequential variable elimination (backward selection) since it is possible that a useful variable is excluded each time the test is performed (i.e. the model is re-estimated on the new subset of variables). In addition, the removal of less significant predictors tends to increase the significance of the remaining variables in the model which might lead to overstating their importance. Finally, stepwise variable selection tends to pick models that are smaller than desirable for prediction purposes (Roecker, 1991). On the other hand, the BMA approach does not impose only one model as the "true" one on the underlying data and instead constructs many different models over the variable space whose fit is subsequently weighted. In other words, significance testing by stepwise/sequential selection routines leads to only one potentially useful combination of predictors, disregarding the fact that subsets of different variables might be equally successful at explaining the dependent.

Nevertheless, in line with the suggestions raised I have applied a backward stepwise selection to the initial set of 78 predictors for the short model and 74 predictors for the long model. However, the backward selection as implemented in Stata 14 fails to eliminate redundant variables and to identify only meaningful predictors of the dependent when the number of explanatory variables is very large, as is the case here. For this reason, to be able to implement a sequential elimination of variables comparable to stepwise selection on the entire initial set of potential early warning indicators I used Python Scikit-learn Recursive Feature Elimination (RFE) procedure. RFE recursively removes predictors and builds a model on those predictors that remain. It uses the model accuracy (set to maximizing the area under ROC curve) to identify which variables (and combination of variables) contribute the most to predicting the dependent. In this respect, RFE repeatedly constructs a model, chooses the worst performing variable, sets this variable aside and repeats the process with the remaining variables. Subsequently, variables are ranked based on when they were eliminated. Comparison of out-of-sample performance of models containing RFE-selected indicators with that of the original models can be found in Table A3.5. In addition, I also performed variable selection on the subset of pre-selected indicators based on area under ROC curve (AUC). In particular, indicators with AUC greater than 0.5 were retained out of the original 78 and 74 indicators for the short model and the long model, respectively, to be subsequently used in the traditional backward selection procedure in Stata 14. To enable a fair comparison of this technique with the BMA approach to variable selection, I also run BMA on the same subset of pre-selected variables and compare their out-of-sample performance in Table A3.6. I also refer the reader to these new tables in footnote 10 in Section 3.4.

Overall, based on the reported results it cannot be concluded that the models based on the sequential selection achieve superior performance to BMA in our setting.

4. *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.*

In line with my explanation of using elevated values of FSI for crises dating in point 2 testing the resulting models' performance against the "true" crisis dummy would not be very practical for policymaking because of lack of timeliness with which the true crisis dummy is available over time and because of insufficient information regarding start or end dates of crises in the crises databases. The tables containing descriptive statistics of the dependent with respect to the number of observations and zero and one instances are provided in Tables A3.1 – A3.4 in the Appendix to Chapter 3. Furthermore, I added footnote 2 in Section 3.2 to refer the reader to these statistics. Tables A3.2 and A3.4 provide FSI statistics on the effective sample for each model highlighting the problems related to the model development mentioned.

5. *Second paper: The same remark as above 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. 3.4 looks quite poor. I recommend also to report out-of-sample R2 and RMSE performance measures to make a more objective comparison.*

My response regarding the standard (i.e. stepwise) versus the non-standard (BMA) selection of leading indicators and drawbacks of the stepwise procedure given in point 3 applies here. As for the out-of-sample performance of the G7 country models, their RMSE is reported in Table 3.6 for both models based on BMA-selected variables and for simple AR(4) models (the benchmark).

6. *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.*

Details on the random forest growing and parameter tuning were added to Section 5.4. The resulting random forest parameters that were the outcome of the tuning exercise using cross-validation are provided in Table A5.2 in the Appendix for each dependent. Descriptive statistics of the variables included in the analysis are presented in Table A5.1 while the random forest error rate as a function of number of trees in the forest are presented in Figures A5.2, A5.3 and A5.4 for each dependent. Additionally, a new control variable for macroeconomic conditions, annual Czech GDP growth, was added to the analysis. The results - the identified useful predictors and the pairwise variable dynamics vis-a-vis the dependent - remain unchanged. Moreover, I have tried including 2 additional macroeconomic controls, interest rate spread between long-term and short-term Czech government bonds and Czech unemployment rate, which are often considered in the literature on bank risk. However, adding these two additional controls does not change the main results and these variables do not appear among the most useful predictors of bank risk, either.

Comments by Doc. PhDr. Petr Teplý, Ph.D.

1. *Page 14: the distribution of weights among individual financial stress index (FSI) indicators is chosen for indicator aggregation, i.e. placing a weight of 25% on each market represented within FSI. Did Diana provide robust testing of these weights? For instance, one might argue that higher volatility of equity indices might have a higher impact on financial stability than the TED spread (I also recommend replacing commas by decimal points in the numbers on this page).*

Alternative approaches for indicator aggregation within an FSI have been explored in the literature. However, many studies which focus on construction of a cross-country FSI have used simple equal variance weighting, i.e. arithmetic average of FSI components; for example, the FSI by Caldarelli, Elekdag and Lall (2011) for 17 advanced economies, the FSI by Lo Duca and Peltonen (2013) for 28 economies, and even Yiu, Ho and Jin (2010) for construction of their FSI for Hong Kong.

The FSI by Lo Duca and Peltonen (2013) whose components we used in construction of the FSI used in Chapter 3 uses a slightly different aggregation of subindices – they apply equal variance weighting as opposed to our market-equal weighting. Despite this departure from their approach, Figure A3.2 in the Appendix shows our resulting FSI successfully captures historical episodes of elevated financial stress in individual countries in the panel.

In addition, for cross-country analysis a neutral (market-equal) weighting of subindices within an FSI seems more appropriate due to inherent country heterogeneity present in such panels. Assigning more weight on stress arising from equity market in countries that are predominantly banking sector oriented (e.g. emerging economies) would be misleading for financial stability purposes since banking sector is relatively more important for these countries. Since our panel contains both equity market and banking sector oriented economies we believe market-equal weights are the most suitable. Also, the commas were replaced by decimal points as suggested.

2. *Page 43: The author suggests adding to the EWS also data on advanced countries with longer time series. Is it general recommendation or is there any specific "cut-off" number of countries or periods that would improve the model for the Czech Republic significantly?*

The observation to add also developed countries to the EWS is general in the sense that more homogenous subpanels of emerging (developing) economies suffer from data restrictions and insufficiently long time series. Furthermore, many emerging economies experienced stresses in relation to domestic crises as opposed to advanced economies that registered very high levels of financial stress during the Global Financial Crisis. As a result, crises episodes such as the Russian crisis in 1998 or the Asian crisis in 1997 might be difficult to capture by EWS containing purely emerging economies due to the short time series problem.

In addition, computationally more intensive techniques, such as BMA, need sufficient number of observations in the resulting panel to arrive to a good convergence with the underlying distribution. As demonstrated on our subpanel insufficient convergence resulted in instability of posterior inclusion probabilities of leading indicators and therefore their usefulness for forecasting systemic events is limited.

I cannot give any exact minimum number of countries/periods needed as a cut-off for making EWS estimation less challenging. However, the number of observations in the final panel needs to be sufficiently large to allow for investigation of different combinations of candidate indicators. The more candidate indicators there are the larger the final panel should be to avoid computational issues. Including a large number of emerging countries into the more homogenous subpanel would be helpful but data limitations to overcome are likely to be substantial.

3. *Page 52: The author correctly states that "Vermeulen et al. (2015) show that using the weighting method proposed by Holló et al. (2012) does not lead to very different results". I understand that it is done to make the index easy to interpret (i.e. the same weights are applied). I have a similar question to my Question 1.1. Which indicator in Table 3.1 is the most important for financial stability from a view of policy-makers according to the author?*

My response to point 1 raised by the referee applies. For this reason, importance of individual subindices of the FSI for policymakers would be likely country-dependent, i.e. for small open economies NEER volatility might be more important than, for instance, for USA (Hakkio and Keeton, 2009).

From frequentist point of view, however, the most frequent type of risk across events is banking risk, which materialised 31 times in complex events and twice in isolation (Lo Duca et al., 2017).

In terms of real costs of financial risks, Lo Duca et al. (2017) find that for EU countries sovereign risk stands out. In their global crises dataset Laeven and Valencia (2013) also find that output losses stemming from banking or sovereign crises are significantly higher than those related to currency crises. Therefore, due to their costliness policymakers might pay increased attention to banking and sovereign risk segments of the financial system (FSI3 and FSI4). However, events that are both domestic and external in nature also coincide with higher output losses (Lo Duca et al., 2017) compared with purely domestic events. For this reason, assessing stresses in the remaining segments of the financial system, that is, via subindices FSI1, FSI2 and FSI5, is important and contributes to the complete picture of the financial stress levels in the economy as captured by FSI.

4. *The author on page 70 says that "The lack of predictability implies that policymakers need to be equipped with flexible tools to respond quickly to emerging financial stress, since long policy implementation lags may aggravate the financial stress episode and the negative effects on the real economy." Could be Diana more specific on the suggested flexible tools for policymakers?*

In the light of the results of Chapter 2 we suggest that policymakers cannot rely on having a time period of several quarters ahead of systemic risks build-up during which corresponding macroprudential tools can be applied to counter negative effects of future elevated financial stress. This calls for a greater flexibility of macroprudential tools at policymakers' disposal.

In line with these results, ESRB (2016) also calls for flexibility in the use of macroprudential tools to counter diverse risks arising from heterogeneous economic and financial conditions across EU countries. The proposed instrument flexibility by ESRB (2016) should extend to scope of exposure, institutions the instruments can be applied to (i.e. allowing national authorities to take action) and to timeliness of instruments' implementation. Effectively, ESRB (2016) identifies burdensome activation (lengthy approval) and notification procedure for some of the instruments which can induce inaction bias. For instance, procedures to counteract risks related to real estate exposures were identified as having potential to aggravate systemic risks due to their lengthy implementation.

5. *Page 124: The author summarizes that "We find evidence for publication selection against positive results; that is, some authors of primary studies tend to discard estimates inconsistent with the competition-fragility hypothesis." I am afraid that this fact is common in the recent science (not limited to economics). Does Diana have any recommendation how to avoid such bias toward to "conventional" conclusions?*

One suggestion how to limit publication selection bias would be for journal editors to subject empirical studies presenting unconventional results to careful scrutiny in the form of numerous robustness, methodology and data checks. Afterwards, if the unconventional results remain unchanged and robust to different specifications the study should be accepted for publication.

Another way to verify consistency and credibility of results in empirical studies is to conduct replication studies. However, replications do not seem to be in high demand in economics, nor are they attractive for empirical researchers since the likelihood of publishing a replication study in a major journal is rather low compared to publishing original studies (Hamermesh, 2007).

6. *Page 154: Diana concludes, among others, that decreases in average director tenure on board are found to reduce bank stability while increases in tenure enhance stability. This implies short-termism or myopic behavior of managers focusing on profit seeking. EBA guidelines (EBA/GL/2015/22) tries to solve it through deferred remuneration. I will appreciate to add discussion over this issue to the final version of the thesis.*

Indeed, in wake of the global financial crisis executive compensation practices have become subject of scrutiny by practitioners as well as by academics. Deferred compensation is one of the components of inside debt holdings of a firm – it represents unsecured and unfunded liabilities of a firm which renders these executive holdings sensitive to default risk similar to that faced by other outside creditors (e.g. Edmans and Liu, 2011). Empirical evidence by Cassell et al. (2012) suggests that CEOs with large inside debt holdings prefer investment and financial policies that are less risky. A paragraph highlighting the relevance of managers' compensation for firm risk was added in the conclusions to Chapter 5.

Despite the fact that including remuneration statistics would certainly offer an interesting avenue for further research as well as it would extend the current analysis, data on remuneration could not be included in the analysis due to their sensitivity and their limited availability in the Czech context.

Comments by Dr. Martin Gächter

1. *While this approach is indeed innovative, one may argue that the explanation of the methodology (in section 5.4) might be a bit too short. A more intuitive explanation would be helpful for the reader to understand why this specific method was applied in this case. While the results are very interesting, I think that a discussion of the caveats (e.g. methodological drawbacks, omitted variable*

bias because the observed variables may be correlated with other crucial drivers etc.) in the conclusion would be desirable.

Section 5.4 was extended with the details on the random forest growing procedure and explanation of the random forest parameter tuning. The results of random forest optimization for each of the dependents are now presented in Table A5.2 and in Figures A5.2 – A5.4 in the Appendix. Furthermore, to increase transparency descriptive statistics of the variables included in the analysis are presented in Table A5.1.

As for the caveats, a paragraph on drawbacks of the applied method with respect to identifying useful predictors was added to subsection 5.5.1. In addition, a potential issue with respect to managers' remuneration and its relevance for risk which could not be addressed in the analysis is mentioned in the last paragraph in the conclusion.

To further test stability of the results, I added a new control variable for macroeconomic conditions, annual Czech GDP growth. The results - the identified useful predictors and the pairwise variable dynamics with the dependents - remain unchanged. Moreover, I have tried including 2 additional macroeconomic controls, interest rate spread between long-term and short-term Czech government bonds and Czech unemployment rate, which are often considered in the literature on bank risk. However, adding these two additional controls does not change the main results nor do these variables appear among the most useful predictors of bank risk.

Comments by the Advisor

1. *Given that the paper is focused on the application of the model to the Czech economy while the second paper presented in the next chapter deals with 25 OECD countries, I would propose to swap the second and third chapter. It would make more logical order and better connection of the topics. Additionally, this change would provide some comfort to the reader as the second paper contains a full description of the Bayesian selection method used in both papers, while the first paper provides only the reference.*

The second and third chapter were swapped in the final version of the dissertation to ensure the reader's comfort as suggested.

2. *The equal weighting used to construct the Financial Stress Index might be questionable as there is a difference importance of equity, bond, exchange rate and interest rate markets as well as the banking sectors among the investigated*

economies. However, this simplifications is commonly used in the literature. Finally, I would suggest to author to consider to add some short paragraph into the first introductory chapter to explain how early warning systems fit into the macroprudential toolkit and what are the other tools that could be used.

The reasoning behind the chosen weighting of subindices within the FSI in our cross-country application is provided in my response to point 1 raised by Doc. PhDr. Petr Teplý, Ph.D.

A paragraph on macroprudential policy and the tools available within the macroprudential framework was added to the introductory chapter and precedes the overview of the topics covered in the dissertation thesis.

3. *Chapter 5: Perhaps some further discussion on the limitations of the conducted analysis might be added. I would recommend to report more on descriptive statistics of data sample and results in the Appendix.*

The Table A5.1 presenting descriptive statistics of the variables included in the analysis was added in Appendix. My response to the comment raised by Dr. Gächter gives potential caveats in connection with the methodology and the research question under investigation.

4. *Additionally, the format and styles of all figures and tables should be aligned across all four chapters. During the pre-defense, the author could further elaborate on the contribution of the thesis to the current macroprudential policy dissuasions and its practical implications. It could be also further consider to add some introductory remarks on financial stability and its importance for sustainable economic development into the first chapter before providing the summary of all four studies included into the thesis.*

I have made attempts to format the tables as similarly as possible given the different type of information they present throughout the thesis. As for the figures, harmonization of their appearance is limited since the figures were produced in four different types of software.

A paragraph on systemic risk and the threat it poses for economic growth was included in the introductory chapter.

References

Alex Edmans & Qi Liu, 2011. Inside Debt. Review of Finance. European Finance Association, vol. 15(1), pp. 75-102.

Cardarelli, Roberto & Elekdag, Selim & Lall, Subir, 2011. Financial stress and economic contractions. Journal of Financial Stability, vol. 7(2), pp. 78-97.

Cassell, Cory A. & Huang, Shawn X. & Manuel Sanchez, Juan & Stuart, Michael D., 2012. Seeking safety: The relation between CEO inside debt holdings and the riskiness of firm investment and financial policies. Journal of Financial Economics, vol. 103(3), pp. 588-610.

Craig S. Hakkio & William R. Keeton, 2009. Financial stress: what is it, how can it be measured, and why does it matter? Economic Review, Federal Reserve Bank of Kansas City, issue Q II, pp. 5-50.

European Systemic Risk Board, 2016. ESRB Response to the European Commission's Consultation Document on the 'Review of the EU Macro-prudential Policy Framework', October.

Feldkircher, M., Zeugner, S., 2015. Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R. Journal of Statistical Software, vol. 68 (4), pp. 1-37.

Hamermesh Daniel S., 2007. Viewpoint: Replication in economics. Canadian Journal of Economics/Revue canadienne d'économique, vol. 40 (3), pp. 715-733.

Koop, G., 2003. Bayesian Econometrics. John Wiley & Sons.

Lo Duca, M., Peltonen, T., 2013. Assessing Systemic Risks and Predicting Systemic Events. Journal of Banking and Finance, Vol. 37 (2013), No. 7, pp. 2183-2195.

Lo Duca, Marco & Koban, Anne & Basten, Marisa & Bengtsson, Elias & Klaus, Benjamin & Kusmierczyk, Piotr & Lang, Jan Hannes & Detken, Carsten & Peltonen, Tuomas, 2017. A new database for financial crises in European countries. Occasional Paper Series 194, European Central Bank.

Luc Laeven & Fabian Valencia, 2013. Systemic Banking Crises Database. IMF Economic Review, International Monetary Fund, vol. 61(2), pp. 225-270.

Roecker, Ellen B., 1991. Prediction error and its estimation for subset - selected models. Technometrics, vol. 33, pp. 459–468.

Yiu, Matthew S. & Ho, Wai.-Yip Alex & Jin, Lu, 2010. A Measure of Financial Stress in Hong Kong Financial Market: The Financial Stress Index. Hong Kong Monetary Authority Research Note No. 02/2010.