Faculty of Social Sciences
Valtiotieteellinen tiedekunta

FINANCIAL FRAGILITY

–

EMPIRICAL STUDIES ON CRISES AND REFORMS

Eero Tölö

DOCTORAL DISSERTATION
To be presented for public discussion with the permission of the Faculty of Social Sciences of the University of Helsinki, at online seminar organized by the University of Helsinki, on the 19th of April, 2021 at 15 o'clock.

Helsinki 2021

Doctoral Programme in Economics, Helsinki, 2021

Supervisors:
Esa Jokivuolle, Bank of Finland, and
Antti Ripatti, University of Helsinki

Pre-examiners:
Steven Ongena, University of Zurich, and
Siem Jan Koopman, Vrije Universiteit Amsterdam

Opponent:
Paul Wachtel, New York University, Stern School of Business

Publications of the Faculty of Social Sciences No. 183/2021.
Valtiotieteellisen tiedekunnan julkaisuja No. 183/2021.

ISSN 2343-273X (Print) ja
ISSN 2343-2748 (Online)

ISBN 978-951-51-6335-6 (Paperback)
ISBN 978-951-51-6336-3 (PDF)

Unigrafia
Helsinki 2021

Abstract

The thesis “Financial fragility – empirical studies on crises and reforms” consists of an introduction and four empirical studies. The introduction provides a more detailed summary of the articles and the empirical methods than is provided here.

Paper 1 entitled “Indicators used in setting the countercyclical capital buffer” is a comprehensive study of early warning indicators of banking crises. These indicators can be used to help guide the decisions on the level of the countercyclical capital buffer. The study examines a large set of early warning indicators in a robust comparable setup. The study corroborates the view by the literature that credit-based indicators are important. Additionally, we find various price-based indicators to be useful.

Second paper, “Predicting financial crises with recurrent neural networks”, studies the state-of-the-art methods to predict financial crisis events. All these methods make use of subset of variables similar to those covered in paper 1. It finds that deep neural networks based on the Long-Short Term Memory (LSTM) architecture or Gated Recurrent Units (GRU) deliver superior performance compared to the more basic models.

Paper 3 is entitled “Do banks’ overnight borrowing rates lead their CDS price? Evidence from the Eurosystem”. It is based on interbank overnight loans filtered from unique payment system data on interbank transactions that take place in the European TARGET2 large value payment system. The study finds that the overnight loan prices can lead the CDS price especially during periods of financial stress. The interpretation is that the private overnight loan rates contain private information not present in the public CDS quotes. Overall, the results suggest that the bank-specific overnight borrowing rates can be a useful short-term risk indicator for banks.

The last paper “Have Too-Big-To-Fail Expectations Diminished? Evidence from the European Overnight Interbank Market” uses the same data source as paper 3 and investigates whether the post-crisis regulation has affected the pricing of loans in the overnight market. Specifically, the article studies whether the perceived too-big-to-fail subsidies of large banks have decreased following the implementation of bank resolution and recovery directive (BRRD) in EU. The

article finds a gradual decline in the overnight loan rate differential between small and large banks that coincides with the gradual implementation of the new directive. However, the decline in the rate differential does not occur at the exact implementation dates of the BRRD directive. Rather, we observe a decline in the funding cost advantage of large banks when actual bail-in events take place during the sample period.

Acknowledgements

I would like to use this opportunity to thank my co-authors, supervisors, employers, colleagues, and family.

The work presented in this dissertation has benefited from the contributions by my co-authors Esa Jokivuolle, Simo Kalatie, Helinä Laakkonen, and Matti Virèn. I'm grateful to my employers Päivi Heikkinen (at the Oversight Division of Bank of Finland), Paavo Miettinen and Katja Taipalus (at the Financial Stability Division of Bank of Finland), and Jouko Vilmunen (at the Research Department of Bank of Finland). I would also like to thank my many colleagues and others over the years for useful discussions and comments including and not limited to Sampo Alhonsuo, Gene Ambrocio, Tuulia Asplund, Nina Björklund, Zuzana Fungacova, Adam Gulan, Eleanora Granciero, Wouter den Haan, Jyrki Haajanen, Markus Haavio, Iftekhar Hasan, Matti Hellqvist, Seppo Honkapohja, Karlo Kauko, Miska Kuhalampi, Lauri Jantunen, Mikael Juselius, Juha Kilponen, Tommi Korpela, Kasperi Korpinen, Kimmo Koskinen, Markku Lanne, Jani Luoto, Otso Manninen, Peter Palmroos, Hanna Putkuri, Pertti Pylkkönen, Mikko Sariola, Peter Sarlin, Eero Savolainen, Heli Snellman, Mervi Toivanen, Jukka Topi, Juuso Vanhala, Jukka Vauhkonen, Fabio Verona, Timo Virtanen, Milan Vojvonic, Ville Voutilainen, Tuomas Välimäki, listed here in alphabetical order.

Finally, I would like to thank so much my thesis supervisors Esa Jokivuolle and Antti Ripatti, for all their hard work and everything they bring to the table. I'm also most thankful to pre-examiners Steven Ongena and Siem Jan Koopman and defence opponent Paul Wachtel for both accepting their roles and for their excellent feedback.

Table of Contents

List of articles

This thesis consists of the introduction and the following four publications. Since some of the articles I-IV contain multiple authors, the author's contribution is disclosed (roundly) as 70, 100, 60, and 60 percent, respectively.

I. Tölö, E., Laakkonen, H., and Kalatie, S., 2018, "Indicators used in setting the countercyclical capital buffer," *International Journal of Central Banking*, Vol. 14, No. 2, pp. 52–111.

II. Tölö, E., 2020, "Predicting systemic financial crises with recurrent neural networks," *Journal of Financial Stability*, Vol. 49, 100746.

III. Tölö, E., Jokivuolle, E., and Virén, M., 2017, "Do banks' overnight borrowing rates lead their CDS Price?" *Journal of Financial Intermediation*, Vol. 31, pp. 93–106.

IV. Tölö, E., Jokivuolle, E., and Virén, M., 2021, "Have Too-Big-To-Fail Expectations Diminished? Evidence from the European Overnight Interbank Market," *Journal of Financial Services Research,* forthcoming.

1. Introduction

About a decade ago, we encountered a cluster of financial crises, including the 2007-08 global financial crisis, and afterward, the European debt crisis. The costs have been enormous. For example, the toll of the 2008 financial crisis in the US alone amount to $70,000 per American, according to Barnichon et al. (2018). The damages are not only financial but also political and social (see, e.g., Mukunda, 2018). Policymakers have implemented reforms to reduce the likelihood of future financial crises. A part of these reforms has been to strengthen the banking system through more stringent regulatory requirements.[1]

This thesis delivers four empirical studies related to financial crises (especially banking crises) and the post-crisis reforms that have followed the latest crises. Unsurprisingly, the recent predicaments have spurred research on the developments that precede typical financial crises. The reasoning is that by identifying the risk factors and devising targeted policies, the risk of financial crises could be mitigated proactively. Quantitative early-warning models can inform about the probability and scale of prospective crisis events and help in timing policy actions. The first two articles in this thesis are studies along these lines and contribute to the vast literature that analyses factors preceding financial crises using cross-country panel data. Early work in this field includes Demirgüc-Kunt and Detragiache (1998, 2000), Kaminsky and Reinhart (1999), Hardy and Pazarbasioglu (1998), Caprio and Klingebiel (1997), and Berg and Pattillo (1999).

The first article is a comprehensive study of early warning indicators. We study which indicators would have been most informative for predicting crises that took place in Europe during the past few decades. Here we focus on single variable indicators that, through their relative simplicity, can be readily implemented as charts to support financial stability decisions and public communication. Consistent with earlier literature, we find that besides credit growth and debt servicing costs, many price-based indicators are useful.

In the background of the first article is the implementation of the countercyclical capital buffer (CCyB). The CCyB was included in the Basel III regulatory pact that

[1] New regulations have also been introduced, for example, for investment funds and derivatives markets.

followed the 2008 financial crisis. It is a dynamically adjusted capital constraint imposed on banks' balance sheets, which is set to be higher during periods of high credit cycle intensity. This concept of cycles related to financial markets, as opposed to business cycles, is relatively new in the economics literature (related literature include Bernanke and Lown, 1991; Bernanke and Gertler, 1995; Holmström and Tirole, 1997; Kiyotaki and Moore, 1997; Eichengreen and Mitchenes, 2007; Geanakoplos, 2010; Borio, 2012). The bottom line is that the credit cycles amplify business cycle fluctuations, which increases the likelihood of financial crises. The CCyB seeks to counteract this amplification, and early warning indicators would be helpful in deciding when to increase the level of the CCyB.

Early warning models that go beyond single variables can give more accurate crisis predictions. Traditionally, they have been linear discrete choice models that use a cross-section of macro-financial variables for making predictions. The second article investigates how crisis predictions could be improved by plugging in time-series to modern neural nets. From the econometric perspective, the neural nets serve as universal function approximations (see Kuan and White, 2007). This could be helpful for crisis prediction as the crises are fundamentally non-linear events making linear models prone to misspecification.

We show that the predictions can be made more accurate by taking advantage of recent developments in sequence modeling, nowadays commonly used in speech recognition and other data science applications. In particular, the new models based on gated recurrent neural net architectures outperform more basic neural-nets and the benchmark logistic model. One of the reasons why neural nets have been used relatively little in economics is their black-box characteristics. To alleviate this problem, we demonstrate how the drivers of the neural net predictions can be characterized by decomposition methods introduced recently to the machine learning literature.

The third article continues the quest for risk indicators but takes the perspective of the interbank markets. The interbank market is where banks provide each other short-term liquidity to facilitate outgoing payments and compliance with the central bank's reserve requirement. In Europe, the interbank market transactions take place in the TARGET2 large-value payment system. We have utilized an algorithm (Arciero et al., 2016) that is able to identify interbank loans, among other payment system transactions that take place in TARGET2. Using the data of filtered interbank loan transactions, we construct a bank-specific measure based on the interest rate that the bank pays for its overnight interbank funding.

Previous literature originating from the seminal article by Furfine (2001) has shown that overnight loans reflect the credit risk of the borrower banks. To investigate the practical value of the information included in the overnight loans, we compare the bank-specific loan rates with the public information in the CDS market. Given the fact that TARGET2 has been flooded with liquidity due to the actions of the ECB related to the global financial crisis and the Euro crisis, and that price discovery is not a high priority in the money market to begin with (cf. Holmström, 2015), it is not ex-ante clear whether the information in money market rates would provide value-added. CDS market, on the other hand, has been shown to lead bond prices and stock prices, which suggests that the traders in the CDS market are well informed, and the quotes provide timely information relative to the other sources.

Against this background, we find that during calm periods the money markets are relatively uninformative, and the price discovery predominantly takes place in the CDS market. However, during stressed periods the overnight loans can lead the CDS price, especially for banks that are riskier. The results show that information in the interbank market rates can be useful, but the relative value of the information depends on the situation.

The final article is related to the issue of too-big-to-fail (TBTF) banks. It has been suggested that TBTF financial institutions made the 2008 financial crises significantly worse (Bernanke, 2010). A bank is called TBTF if its failure would endanger the stability of the financial system. If a bank's creditors expect the bank to be bailed-out, resulting in a lack of market discipline, there are fewer incentives for the bank to act prudently. As part of the post-crisis reform pact, the EU has implemented a Bank Resolution and Recovery Directive that ensures that banks' shareholders and creditors pay their share of resolution costs.

The fourth article investigates the issue of TBTF from the perspective of the interbank market. It contributes to a growing literature that analyses the issue of TBTF subsidies surrounding the recent financial crisis events (Acharya et al., 2016; Ahmed et al., 2015; Araten and Turner, 2013; among others). Using the same data source as in the previous article, we investigate how the interbank rates depend on bank size and other bank characteristics. We find that large banks consistently obtain cheaper funding than their smaller peers, which could be an indication of TBTF status. Since the cost differential could also be related to other factors, we try to measure, whether it has changed following the adoption of the new anti-bail-out resolution regime. Despite the short maturity of the overnight loans, we can't pin-down a change in the funding cost at the precise implementation dates. Instead, the magnitude of the size premium decreases when actual bail-in events occur during the sample period. Additionally, there is evidence of a gradual change in the funding costs that matches the timeline for the longer process of proposing and legislating the new regulation.

Hence, overall the four articles contribute to the literature on financial stability and financial crises. This literature will likely continue to grow in the future, and despite the new stability measures will be fostered by occasional crises. Alongside there will be research on proactive financial stability measures, a literature that will flourish in the coming years when different policies are tried, and the policymakers seek the optimal trade-off between growth and stability.

The rest of this thesis is structured as follows. Section 2 reviews the econometric methods used in the studies on a general level in so far as they are not already sufficiently covered in the articles. This leaves outside the neural net models, which are discussed at length in Article II and its annexes. Section 3 summarizes the articles and their findings. Section 4 concludes by discussing the contributions relative to the earlier literature and offers some suggestions for further research.

2. Methods

Econometrics is statistics applied to economic data. Articles I-IV make use of various econometric and statistical methods outlined in this chapter.

2.1 Binary classification problem in crisis prediction

Following the publication of various crisis datasets since the late 1990s, the early warning literature commonly poses the financial crisis prediction as a discrete classification problem. In a generic classification problem, we want to predict the discrete category y of an observation based on a set of covariates X. In Articles I and II, we seek to predict whether the data correspond to a pre-crisis or normal observation. That corresponds to a binary classification problem. In this context, a classification model assigns a probability for a period t to belong to one of the categories (pre-crisis/normal) given vector X_t (see Figure 1 for illustration).

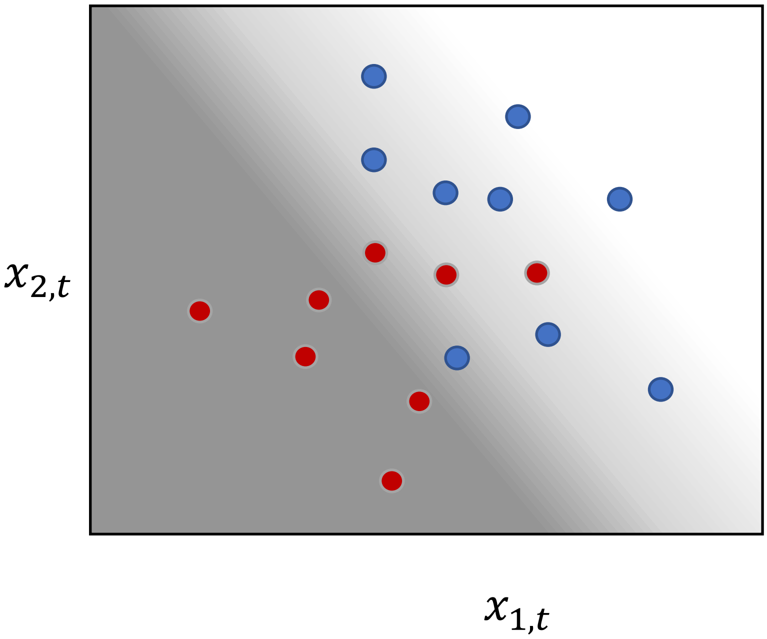


Figure 1. Binary classification based on two covariates $x_{1,t}$ and $x_{2,t}$. Darker background color corresponds to a higher probability of belonging to the red category.

Classification models can be split into two categories (see Murphy, 2012). The first category we call *generative models*. Generative models treat both X and y as random and specify their joint distribution. Since y is a categorical variable, it suffices to specify the distribution of X at each category $\pi(X|y = k)$. We assign prior probabilities (although this is not really Bayesian inference) to each category $p(y = k)$ and use Bayes theorem to obtain $p(y = k|X)$. Examples of generative models include linear and quadratic discriminant analysis. Generative models are called generative as they allow to generate new observations based on the distribution of X.

Generative models have been used in crisis prediction, but they are not as popular as the second class of models called discriminative models. Often we are not interested in the distribution of X, so discriminative models make fewer assumptions by not assuming distribution for X. The models directly specify the marginal likelihood $\pi(y = k|X)$ and perform statistical inference based on that. Examples include logistic and probit regression, which are the classical benchmark models for crisis prediction. Some authors also use multinomial logistic models, typically with three or four categories (see Bussiere and Fratzscher, 2006; Caggiano et al., 2014). Neural networks also fall within the class of discriminative models as neural nets directly specify the output without considering the distribution of X.

In the logistic regression model, we assume $y_i|X_i \sim Bernoulli(\sigma(X_i'\beta))$, where σ is the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$. The model is estimated by maximizing the conditional likelihood

$$f(y|X,\beta) = \prod_{i=1}^{n} \sigma(X_i'\beta)^{y_i}[1 - \sigma(X_i'\beta)]^{y_i - 1}. \qquad (1)$$

The predicted probability for a new observation X_{new} to belong to category one is obtained via

$$p\left(y_{new} = 1|X_{new}, \hat{\beta}\right) = \sigma\left(X_{new}'\hat{\beta}\right), \qquad (2)$$

where $\hat{\beta}$ is the maximum likelihood estimate.

Each of the early-warning models outputs an estimated probability for the pre-crisis state given the observed explanatory variables $\hat{\pi}(y_t = 1|x_t)$. Note that to avoid endogeneity problems caused by simultaneity, the information in the covariates has to be lagged relative to the information in the crisis-dummy. If the probability $\hat{\pi}$ is larger than some threshold h, then we ex-ante classify the state as a pre-crisis state, and a normal state otherwise. Afterward, we observe the ex-post state and calculate the classification error. Correctly predicted crisis is called a true positive (TP). Correctly predicted normal state is a true negative (TN). A false alarm is a false positive (FP), and a missed crisis is a false negative (FN). The TPs, FNs, TNs, and FPs can be accumulated to calculate various performance statistics. In Article I, we evaluate the EWMs using two alternative classification performance measures: the area under the ROC curve (AUC) and the policymaker's relative usefulness. Article II only uses the AUC statistics. They are both widely used measures in the context of financial crisis prediction, but AUC is used more widely across disciplines. The performance statistics make use of following error rates

$$\text{Sensitivity} \coloneqq \frac{TP}{TP+FN} = TPR = 1 - FNR\ , \tag{3}$$

$$\text{Specificity} \coloneqq \frac{TN}{TN+FP} = TNR = 1 - FPR\ . \tag{4}$$

AUC: The receiver operating characteristic (ROC) curve is obtained by plotting the true positive rate (TPR) on the vertical axis and the false positive rate (FPR) on the horizontal axis for all possible values of the threshold h (see Figure 2b in Article I). The area under the curve (denoted AUC or AUROC) is a scoring rule for the classification task. The highest value of the AUC, 1.0, is achieved for a model that is able to perfectly predict the crisis and normal states for some threshold h. A random guess obtains AUC = 0.5. AUC<0.5 indicates that the predictions are worse than a random guess. It is important to remember that the AUC statistic is agnostic to which threshold h the policymaker should actually use.

We obtain confidence intervals for the AUC using the bootstrapping algorithm by Pepe et al. (2009). In the algorithm, subjects that contribute several observations to the ROC curve, in our case countries, are identified as resampling clusters. In our application, adjusting for clustering is important, especially with shorter frequencies (quarterly or shorter) and when using more than one pre-crisis period, as in that case, the prediction errors at successive periods become correlated to a significant degree.

Relative usefulness: Consider a policymaker that chooses a threshold h. We can think the policymaker as having simple preferences over the error rates (Alessi and Detken, 2011) such that the policymaker's loss function is $L = \theta\, FNR + (1-\theta) FPR$. He can always achieve loss Min(θ,1 − θ) by classifying all 0 or all 1. Hence, we can define normalized relative usefulness as $U_r = \frac{Min(\theta,1-\theta)-L}{Min(\theta,1-\theta)}$. Usefulness is positive for an indicator that helps policymaker reduce loss beyond what he can always achieve, and it is one if the error rates are exactly zero. For the relative usefulness, a higher value is better (see Figure 2c in Article I for illustration).

2.2 Neural nets and Shapley values

The neural net models used in Article II are discussed extensively in the corresponding article and its supplement. Here, we provide additional remarks about the use of Shapley values in this context. It is helpful to contrast the interpretation of a logistic model and a neural net. In the case of the logit model, the contribution of each predictor variable can be easily understood from the associated component of the coefficient $\hat{\beta}$. In contrast, the neural nets come with a potentially much larger number of parameters, and it is hard to comprehend the contribution of the individual predictors from the set of parameters alone. Model users naturally want to understand what the predictions are based on. Fortunately, the contribution of different predictors to the predictions can be quantified using Shapley values (Shapley, 1953).

Shapley values are an old concept from game theory that has been recently applied to understanding drivers of machine learning models (see Lundberg and Su-In Lee, 2017). Bluwstein et al. (2020) are first to use them for understanding how machine learning methods predict financial crises. For more details on the method, see Lipovetsky and Conklin (2001) and Lundberg and Su-In Lee (2017).

Shapley value quantifies how much a player contributes to the payoff in a multiplayer coalitional game. The order in which the players enter the game matters. For example, in soccer, adding a new goal-keeper adds little value if the team already has a good goal-keeper. We denote the set of all potential players by N and the subset of players that participate in the game by $S \subseteq N$. The payoff associated with coalition S is denoted by f_S. Because the contribution of player k depends on which other players are already included in the game, the overall

Shapley value of player k, denoted by $\phi(k)$, is a combinatorial sum of differences given by

$$\phi(k) = \sum_{S \subseteq N-\{k\}} \frac{|S|!(|N|-|S|-1)!)}{|N|!} (f_{S\cup\{k\}} - f_S). \qquad (5)$$

In the neural net application, each predictor will correspond to a player in a game of out-of-sample prediction. This is a repeated game where each round corresponds to a country-time tuple (i,t). In the basic application, we calculate the Shapley value $\phi_{i,t}(k)$ separately for each round. In this case, the payoff for coalition S is $P_S(x_{i,t,S})$, the probability output from a neural net.

A fundamental choice is whether the neural net used for predicting with the subset of predictors $S \subseteq N$ is the same as the original neural net,[2] or if we train a new neural net for the input variables S. We choose the latter approach. While training a new neural net for each input combination is computationally heavy, it is reasonable for our relatively small neural nets. It also ensures that the included predictors do not underperform due to ignoring interaction with the excluded predictor.

Defining $\phi_0 = P_\emptyset(x_{i,t,\emptyset})$, an important additivity identity holds

$$P_N\left(x_{i,t,N}\right) = \sum_{k=1}^{N} \phi_{i,t}(k) + \phi_0. \qquad (6)$$

The above identity means that the Shapley values can be used to decompose the output into contributions of each predictor.

2.3 Vector autoregression and vector error-correction models

Article III makes use of vector autoregressions (VAR) and vector error correction models, which are powerful tools for summarizing intertemporal relationships in multivariate time series data. VARs and their descendants are extensively used in finance and macroeconomics following the work of Wiener (1956), Granger (1969), Sims (1980), and others (see also Granger, 2003). Technical advances continue to this day as VARs are extended to more complex model structures.

[2] Even if the neural net requires a preset number of input variables, we could in principle replace the excluded inputs with uninformative sample averages as is done by Bluwstein et al. (2020).

VARs are popular for macroeconomic forecasting. In finance, they are often used to study information content in prices in different markets; in the case of Article III, the two markets are the overnight loan market and the CDS market. The adopted framework treats the two variables on an equal basis and allows dependencies in both directions.

We define a VAR process by the equation

$$y_t = A_0 + A_1 y_{t-1} + \cdots + A_p y_{p-1} + u_t \,, \qquad (7)$$

where y_t is a d-dimensional vector of observations at time t; A_i are parameter matrices, and u_t is an error process. The error process is assumed to be white noise, i.e. serially uncorrelated with zero mean $E[u_t] = 0$, and finite covariance $E[u_t u_s'] = \Sigma_u$. It is reasonable to expect these conditions to be satisfied if y_t is a two-component vector consisting of the overnight spread and the CDS spread.

If we know the numbers in the parameter matrices A_i, we can use the zero-mean assumption of the error term and recursively calculate the forecasted future path for y_t. If the assumptions above are satisfied, the resulting forecast is unbiased and minimizes the mean squared forecast error. That makes VAR models attractive for forecasting.

The VAR model is convenient for summarizing the dynamics of two or more dependent time series. The parameter matrices A_i and error covariance Σ_u concisely summarize dynamic information. The parameter matrices reveal if known values in one series contain information that is useful for predicting the future value of other series. In Article III, we are especially interested in the cross-terms in A_i, which tell about the relationship between the two markets. We make use of the causality concept by Granger (1969), which is based on the observations that 1) The cause occurs before the effect; and 2) The cause contains information about the effect that is unique, and is in no other variable. In the case of Article III, we look at whether one of the market prices causes the price in the other market subject to a number of control variables. For example, in our case, we note that the A_i matrix has a significant off-diagonal term (such that past O/N rate affects current CDS price) that remains robust to all relevant control variables so that the definition of causality is satisfied.

So far, we did not discuss how to estimate a VAR. We can obtain the maximum likelihood estimates (MLE) of the coefficients A_i by performing least-squares regression on each equation (provided there are no restrictions on the parameters; see Kilian and Lütkepohl, 2017). We can then calculate the residuals $\widehat{u_i}$ and an estimate for the variance-covariance matrix $\hat{\Sigma}_u$. The MLE method also yields standard errors. However, these are not necessarily valid as such, especially in the panel setup. With the pooled panel, we use robust standard errors adjusted for clustering, as advocated by Petersen (2008).

We have concluded that VARs are useful for summarizing time-series data. However, something that we so far failed to mention is that economic time series often have non-standard properties. That is part of the reason why the current chapter is titled econometric methods instead of just statistical methods. An important feature of many economic time series is that they are relatively smooth and characterized by trend behavior; in other words, the series are non-stationary. Empirically, the CDS spread time-series over our sample period look non-stationary, which is further confirmed by unit roots tests (Dickey and Fuller, 1979).

Oftentimes, we have two time series, which both are non-stationary. Whether the two series bear any real relationship, if we regress one on another, we often find a statistically significant relationship. This is a problem of spurious regression. An important discovery in the 80s by Engle and Granger (1987) was that if the two or more time series are *cointegrated*, i.e. share a common trend, then suitably scaled a linear combination of the cointegrated series would be stationary. This leads us to the vector error correction (VECM) model, which can be written as

$$\Delta y_t = \lambda \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \, , \tag{8}$$

where $\Delta y_t = y_t - y_{t-1}$, Γ_i are the parameter matrices, u_t is an error process, and $\lambda \beta' y_{t-1}$ is the lagged error correction term. λ and β are $d \times r$ matrices with r the number of co-integration vectors. In Article III, there are two credit spread series ($d = 2$), and there is potentially one cointegration vector $r = 1$.

We estimate the VECM model using a maximum likelihood method by Johansen and Juselius (1995). We find mixed evidence in favor of a co-integration relationship between the two credit spreads. Hence, we report both VECM and VAR results.

Article III uses two measures of price discovery that are based on the VECM model. In an ideal setting, we have identical securities that are traded on multiple markets. The price discovery measures determine, to which extent each market produces new price information. In practice, the securities need not be identical as long as they are linked by arbitrage or approximate parity relations such that they are co-integrated and VECM model can be used to describe the relationship. In the case of CDS and overnight rates, the maturities differ, so there is no exact arbitrage relationship. However, the fact that both describe credit spreads of the same bank seems to be enough for a co-integration relationship to exist.

The information share by Hasbrouck (1995) presents the proportional contribution of a market's innovation to the innovation in the common efficient price. In our case, the "efficient common price" is an unobserved common credit risk factor that drives both CDS and overnight loan prices. The Gonzalo-Granger measure (Gonzalo and Granger, 1995) is simply the common factor component weight ($|\lambda_1|/[|\lambda_1| + |\lambda_2|]$). Both measures lie in the interval [0,1] and are closely related. However, the information share takes into account the variability of the innovations in each market, whereas the common factor component weight does not (cf. de Jong, 2002).

2.4 Panel data techniques - fixed effects model and differences in differences

Panel data consists of repeated cross-sections. Although all of the four articles deal with panel data, the techniques described here pertain mainly to Article IV. The fixed-effect model allows controlling for specific forms of omitted variable bias. The difference in differences is a technique used to infer the effect of a treatment in a natural experiment by observing the treatment group and the control group over time.

2.4.1 Fixed effects model

In Article IV, we use a fixed-effects model to analyze the determinants of the interest rate that a bank pays in the interbank market. In this application, the fixed-effect model helps to control a specific type of omitted-variables bias related to heterogeneity across countries and time. The model is written as

$$y_{i,j,t} = b_{0,c(i)} + b_{0,t} + b'x_{i,j,t} + \epsilon_{i,j,t} \ , \qquad (9)$$

where i and j denote the borrower and lender banks, respectively, $c(i)$ is the country of the borrower, t denotes day, y is the dependent variable (interest rate), and x is a vector of explanatory variables. $b_{0,c(i)}$ and $b_{0,t}$ are the two fixed effect terms. $b_{0,c(i)}$ is used to control potential unobserved country heterogeneity that would be constant over time. $b_{0,t}$ controls fixed time effects, i.e. potential unobserved time heterogeneity that would be constant over loans granted on a given day. These could be related to the market-wide liquidity conditions in the overnight market.

For the fixed effect model in Article IV, we use the least square coefficient estimates together with robust standard errors adjusted for clustering at the bank level (Rogers, 1993). The robust standard errors adjusted for clustering relax the assumption that the errors are independent and identically distributed. Instead, the errors only need to be independent across clusters. Because the observations are at the transaction level while some of the explanatory variables are at the annual level, using the clustered standard errors is crucial in order not to have grossly inflated t-values. An alternative approach would be to aggregate the data to annual level.

2.4.2 Difference in differences

Controlled experiments are important for causal inference. In economics, we typically do not have controlled experiments. Nevertheless, sometimes an experimental setup can arise naturally, as is the case in Article IV, where legislation is implemented in only some countries of a multi-country data set. Such natural experiments can be analyzed through the difference(s)-in-differences (DiD) method. For DiD analysis, we need to have panel data of the outcome y for a treatment group and a control group. The event, which should only affect the treatment group, should take place during a short period such that there are no confounding effects. The inference is based on observations before and after the event.

For simplicity, let us assume that there is one period before the event (t = 0) and one period after the event (t = 1). Then the DiD analysis can be implemented by a regression

$$y_{i,t} = \alpha t + \beta S_i + \gamma S_i t + \delta + \epsilon_{i,t}, \qquad (10)$$

where $S_i = 1$ for the treated group and $S_i = 0$ for the control group. If we only have two periods, the parameters $(\alpha, \beta, \gamma, \delta)$ can be estimated by OLS. Otherwise, it is appropriate to use the robust standard errors adjusted for clustering (cf. Bertrand et al. 2002). α is the assumed parallel trend if the treatment does not have an independent effect on the treated group. β is the difference in the outcomes for the treated group and the control group before the event. γ is the DiD estimate of the effect of the event on the treated group.

In Article IV, the event is the implementation of the Bank Resolution and Recovery Directive (BRRD). The treated group is formed by banks that reside in an implementing country, and other banks form the control group. Some countries do not implement the legislation, and other countries enforce the legislation at different dates, which provides a natural split into treatment and control groups. The outcome is the interest rate that the bank pays in the interbank market. We also introduce a further explanatory variable in (4) (the size of the bank), the coefficient of which could change following the event.

There are two factors that increase the risk that DiD analysis may not give a conclusive result for the effect of BRRD. First, it is not clear whether the effect should occur at the exact implementation date, or gradually over time. On the other hand, using a longer time-window would risk confounding effects from the crisis episodes and related ECB operations. Second, it's also possible that the effect for short maturity loans could be dampened due to their exclusion from the immediately bail-innable funds. For these reasons, we do additional analysis with dummy variables corresponding to developments in the legislative process and comment on additional checks with longer maturity loans.

3. Results

Article I: Indicators used in setting the countercyclical capital buffer

Articles I and II deal with systemic banking crisis prediction. There are many definitions of a systemic banking crisis. Still, quite generally, it can be thought of as a situation when a country's banking sector has bank runs, significant losses in the banking system, or bank liquidations (cf. Laeven and Valencia, 2012). In the first article (henceforth, *the indicator study*), we survey and test which indicators best predict the outbreak of a banking crisis. Thus, the focus is on the variables. In contrast, in the second article discussed in the next subsection, the focus will be on making most out of a set of variables by using advanced modeling techniques.

The motivation for the indicator study was to collect and test a broad set of early warning indicators to inform decisions regarding the so-called countercyclical capital buffer (CCB, or often CCyB for countercyclical buffer). The CCyB is a financial stability policy instrument whose purpose is to mitigate the harmful effects of credit cycles. The idea of the CCyB is to strengthen the banks' capital buffers during times when vulnerabilities in the banking system build-up. On the one hand, the incremental capital requirements can dampen down credit growth. On the other hand, the banks would be more resilient in the possible event of a crisis. In any case, the early warning indicators would be used to identify periods likely related to a build-up of banking crisis risk.

The article contributes to a vast literature of early warning indicators (reviewed in the article, so we do not repeat it here). Here, our goal is to find suitable indicators in several categories of risk,

1. credit developments,
2. the private-sector debt burden,
3. potential overvaluation of property prices,
4. external imbalances,
5. potential mispricing of risk, and
6. strength of bank balance sheets.

In the study, we first perform a literature survey that encompasses 30 articles that use panel data to predict banking crises. Owing to the different setups in the articles, the predictive power of the indicators could not be directly compared. However, we report whether the indicator is a statistically significant predictor in the main specifications reported in the articles. According to the literature survey, among the most generally used and consistently significant indicators are indicators like credit, credit relative to GDP, debt servicing costs, house prices, stock prices, interest rates, current account deficit, bank leverage, and non-core liabilities.

In the testing phase, we include those above and many other predictors. The methodology is similar to Detken et al. (2014). We collect an unbalanced quarterly panel of indicator data for EU-28 countries for the period 1970 to 2012. Adapting to a reasonable time frame for CCyB policy decisions, we set the prediction horizon to 1 to 3 years. We also primarily use the ESRB crisis dataset, which aims to capture credit booms. Among the robustness checks, we consider alternative prediction horizons and crisis datasets.

The results broadly align with the findings of the literature survey, but importantly allow to some extent rank the different indicators using performance scores (AUC and relative usefulness). Based on the performance scores, the credit-based indicators, debt-service ratio, and house prices are among the most informative indicators. However, the results for some other indicators such as current account deficit and bank balance sheet variables suggest less robust prediction performance; and the performance either did not carry to out-of-sample or were not robust across different crisis datasets. We also discover two new indicators (low) VIX index and (low) high-yield bond spread that were informative predictors across different datasets.

The main contributions of the article are the extensive literature survey and tests of the indicators, which should be useful for policymakers in Europe. We don't give numerical trigger values for the indicators, but instead, recommend that the policymakers use judgment when incorporating the information from the indicators to their decision making.

Article II: Predicting systemic financial crises with recurrent neural networks

Once we have a set of candidate early warning indicators, the predictions can typically be made more accurate by combining the indicators in a multivariate prediction model. Traditionally the early warning models are put together using logit or probit model. However, in the past 15 years, there has been increasing use of various machine learning methods in the field (as reviewed in Article II). In particular, neural nets have been successfully used in banking crisis prediction in many studies. From an econometric perspective, the neural nets can provide parsimonious function approximations that help in forecasting non-linear phenomena. Although the amount of crisis data is not comparable to the large datasets used with very deep neural nets in, say, image or speech recognition, the crisis prediction can well benefit from smaller neural nets.

In the past few decades, the increase in computing power has spurred a lot of new research with neural nets. Article II leverages on the recent advantages in dealing with sequence data with so-called recurrent neural nets (RNNs). The gating mechanisms introduced by (Hochreiter and Schmidhuber, 1997) and (Cho et al., 2014) have made it possible to estimate RNNs that are numerically stable and can retain past events in memory for extended periods of time. Such networks have been used successfully in forecasting applications, but generally, their application in economics has been quite limited. In Article II, we benchmark the modern RNNs against more basic neural nets and the logistic model. The competing neural network architectures are illustrated in Figures 1-4 in Article II (see Annex in the respective article for details of the neural net models). The RNNs considered here include a basic RNN, a Long-Short Term Memory RNN (Hochreiter and Schmidhuber, 1997), and a Gated Recurrent Unit RNN (Cho et al., 2014). Our main result is that the gated RNNs outperform both the basic neural nets and logistic models. The advantage derives from the recurrent neural nets' ability to make robust predictions with time-series data.

We evaluate the crisis prediction performance with one to five-year prediction horizons using an annual unbalanced panel dataset by Jordá et al. (2017) that covers 17 countries over the period 1870-2016. The same dataset also includes financial crisis dates. The indicators used include one-year growth in credit-to-

GDP ratio, GDP, house prices, and stock prices, and level of current account deficit (relative to GDP).

In machine learning, the in-sample predictions are typically irrelevant as models with enough parameters can achieve a perfect fit. Hence the results are evaluated out-of-sample. We consider two types of out-of-sample evaluations: cross-validation and sequential evaluation, which are both commonly used in the literature. In the country-by-country cross-validation, each country, in turn, is used as a test sample, while the other countries are used for estimation. In the sequential evaluation, the model is estimated for one time period and tested for a later period. We see a consistent performance advantage for the gated RNNs in cross-validation and sequential evaluation using various subsamples and prediction horizons.

Often the neural nets are seen as black-box prediction devices that would not offer much interpretation for their predictions. However, various methods to interpret the drivers of the neural net forecasts have been recently proposed (see Lundberg and Su-In Lee, 2017). We employ the method based on Shapley values, which decomposes the contribution of each predictor to the predicted probabilities (see the previous section for discussion). This approach is broadly similar to Bluwstein et al. (2020), apart from the fact that they use the prediction model with uninformative inputs instead of estimating a new model with a smaller set of explanatory variables.

The Shapley value decomposition reveals that the LSTM recurrent neural net still largely prefers the same variables as the logit model. However, stock prices seemed to be particularly important for performance improvement in cross-validation. In the sequential evaluation, the variables contributed more evenly. Also, we observed that the recurrent neural net needs to have a sufficiently rich set of predictor variables to outperform the traditional models significantly. Single variable models did not lead to statistically significant improvement.

In summary, Article II demonstrates that modern neural net models can be advantageous in financial crisis prediction. It further investigates the model drivers and finds them to be broadly consistent with earlier literature.

Article III: Do banks' overnight borrowing rates lead their CDS price?

Article III leaves the topic of financial crisis prediction but is still related to predicting bank stress. Every day banks borrow and lend money from each other in the interbank market. The associated interest rates are used to construct benchmark lending rate indices such as LIBOR and Euribor. However, the bank-specific interbank market data can also be a source of indicators of banking problems. We construct a bank-specific risk measure of a bank's relative creditworthiness based on the premium that a bank pays for its overnight interbank funding. To the extent that the measure reflects the credit quality of the borrower, it can be interpreted as a health indicator for the bank. We investigate how this indicator compares to the leading market-based indicator, the bank's CDS price.

The data comes from a proprietary dataset of TARGET2 money market loans, which is available for use within the Euro System. Although the dataset contains loans of all maturities, we focus on the overnight segment, which is the most liquid and most robust in terms of the data quality. The CDS was chosen as the benchmark because it is regarded as the leading public information of credit risk (see e.g. Blanco et al. 2005 or Arora et al. 2012). Moreover, quotes for large banks are available on a daily level, and the contracts are reasonably liquid.

The interbank loans arise from bilateral agreements, so in principle, the price can contain private information before it is incorporated into the price in public markets. Earlier research has already shown that interbank rates reflect borrower's credit risk characteristics but also lending relationships. Therefore, we focus on how fast credit risk is priced in the overnight interbank loans and to which extent they contain private information. To best of our knowledge, these have not been considered before.

The data covers the period 2008-2013. The transactions are filtered from TARGET2 data using an algorithm similar to Furfine (2001) implemented in Arciero et al. (2016). The overnight loan data, which is available by the transaction timestamp, is aggregated to a daily level and covers a period from June 2008 to the end of December 2013. The sample consists of 60 banks that have regular CDS quotes and frequent borrowing in the money market.

We use the VAR and VECM framework to summarize the dynamic relationship between the overnight loan price and the CDS price. The methodology is largely adopted from Blanco et al. (2005), who investigate the relationship between corporate bond yields and CDS prices. The VECM structure allows us to calculate the Hasbrouck and Granger-Gonzalo price discovery measures.

Consistent with the understanding that the purpose of the interbank market is liquidity provisioning and not the price discovery (cf. Holmström, 2015), we find that price discovery usually takes place in the CDS market. Figure 5 in Article I shows the evolution of price discovery measures. The figure is interpreted such that values closer to 1 mean that the price discovery takes place in the CDS market, and values closer to 0 indicate that the price discovery takes place in the overnight market

We also find evidence that the overnight rates sometimes contain private information, which is not immediately reflected in the CDS market; in other words, AOR Granger causes the CDS (see Table 4 in Article III). The private information in the overnight rates seems to be mostly present at times of market stress. It is primarily related to banks, which mainly borrow through a long-term relationship (Table 6 in Article III). These are typically smaller banks that have weak credit ratings and were often located in the GIIPS countries.

An implication of Article III is that the overnight borrowing rate can be a useful measure of bank risk, especially in stressed times. Although we cannot test this directly, the loan rate could be particularly helpful information for monitoring smaller banks that do not have public credit risk measures such as CDS or bond spreads. An obvious caveat is that the observed bank needs to be active in the unsecured overnight market. This may not be the case if the creditors require collateral or when the lenders refuse to lend money altogether.

Article IV: Have Too-Big-To-Fail Expectations Diminished? Evidence from the European Overnight Interbank Market

Article IV continues the interbank market theme and utilizes data of overnight loans filtered from the TARGET2 payment system for the period 2008–2016. [3] We investigate the determinants of overnight loan rates, especially from the perspective of whether they have been altered by the introduction of the new bank resolution rules in the EU (the Bank Recovery and Resolution Directive, BRRD). The new rules facilitate an orderly restructuring of a failing bank such that the equity and debt holders bear the costs. In other words, instead of bank bail-outs by public money, there would be bail-ins for larger banks and regular bankruptcy procedures for smaller banks. The change could make the interbank loans more sensitive to borrower risk characteristics and less sensitive to the borrower's systemic importance, which may be measured by the size of the balance sheet. A caveat is that short maturity interbank loans are exempted from the immediately bail-innable debt in the BRRD.

We match the loan data with each borrower bank's characteristics from BankScope and construct bank relationship variables based on the payment system's data. First, we analyze the determinants generally via a panel regression where the determinants are used to explain bank-specific interest rate premiums similarly as in Furfine (2001), see Table 2 in Article IV. We find that both credit risk and relationship characteristics help explain the observed rates. The supplementary material (see Table S1) also shows that the determinant that works best across different years is the bank size (measured by the logarithm of total assets). Thus, the results mean that cheaper funding for large banks is a pervasive feature in the money market data.

There are several explanations for why the bank size is such a strong predictor of the borrowing rates. 1) A large bank potentially has a significant market share in the overnight market and borrows from a larger number of lenders, which converts to bargaining power when negotiating the rates. 2) A large bank is potentially more diversified and may exhibit economies of scale, which justifies a lower borrowing rate (cf. Hughes and Mester, 2013). 3) A large bank may benefit from an implicit government guarantee (high expected bail-out probability).

[3] More precisely, the data period extends from June 2008 to September 2016. Following the ECB's decision to set the deposit facility interest rate to a negative value from 11 June 2014 onwards, a large portion of the overnight rates henceforth takes place at negative interest rates. Thus we had to amend the algorithm by Arciero et al. (2016) to allow for negative interest rate loans.

We consider a few alternative approaches when we investigate whether the large banks' funding cost advantage could be related to bail-out expectations. In the basic analysis, we aim to control for economies of scale, the diversification, and market power by including control variables ROE, a measure of geographic diversification, and the number of overnight market counterparties. The funding cost advantage is robust to the inclusion of the controls and seems to be most significant at times of market stress. In an auxiliary regression, we show that banks owned by governments with solid finances exhibit a funding advantage, but it does not depend on the bank size. These findings support the notion that part of the coefficient of bank size could be attributed to implicit government guarantees that depend on bank size. At the very least, they show that the large banks may be perceived as safe havens for overnight deposits at times of market stress. Despite the controls, it is still possible that the cost advantage related to bank size could be related to factors other than TBTF. Hence, we turn to investigate the behavior of the borrowing rates around the implementation of new resolution rules that should have affected the TBTF subsidies. A change in the size premium following the new rules would be seen as evidence of a change in the TBTF subsidy. Along similar lines, we investigate if the size premium reacts to actual bail-in events that took place during the sample period.

We carry out these investigations through a difference-in-difference (DiD) analysis (see Table 4 in Article IV). The DiD analysis shows no change in the size premium around the BRRD implementation dates.[4] Nevertheless, the size premium decreases significantly following the actual bail-in events, which supports the notion that at least part of the size premium is related to TBTF subsidies.

The non-result around the BRRD implementation could well result from more gradual pricing-in of the new resolution framework. Hence, we implement dummy regressions to investigate potential gradual changes during the period 2012–2016 (Table 5 in Article IV). We find that the funding cost advantage of large banks has reduced (especially in GIIPS countries) over the years that the

[4] Although the main analysis was conducted using overnight loans, we confirmed that there was also no change for longer maturity interbank loans.

new resolution framework was implemented. However, for the long-term analysis, the monetary policy operations of the ECB are a confounding factor.

The contribution of the article was to analyze the interbank rate determinants in Europe and investigate the issue of TBTF in some detail. While the data does not allow us to unambiguously attribute the funding cost advantage of the large banks to different sources, we come up with findings that can benefit future work in this area. The large banks in Europe have benefited from a consistent funding cost advantage, and especially the large banks in safe countries seem to have been perceived as safe havens at times of market stress. When considering the effects of the new resolution regulation, it may be more helpful to look for gradual change than a sudden change around the BRRD implementation dates.

4. Discussion and Conclusions

The four articles that we have discussed make overlapping contributions in the fields of applied time series analysis, banking crisis prediction, central banking, financial stability policy, money markets, and banking.

Articles I and II are contributions in the early warning literature of banking crises. This literature dates back to the pioneering work by Demirgüc-Kunt and Detragiache (1998, 2000), Kaminsky and Reinhart (1999), Hardy and Pazarbasioglu (1998), and Caprio and Klingebiel (1997). During past 20 years, this literature has grown by dozens of studies that investigate the causes of banking crises, provide new data sets of financial crisis events (e.g., Laeven and Valencia, 2012; Reinhart and Rogoff, 2009; Schularick and Taylor, 2012) or optimal indicators for policy use (e.g., Drehmann and Juselius, 2014; Detken et al., 2014), and other related issues. In layman terms, this literature demonstrates that unlike many natural disasters, banking crises are predictable. Naturally, we need to understand that the predictions are only probabilistic. Still, even if the crisis in a particular country are often triggered by external events, it is the underlying vulnerabilities that are measurable and can be taken into account in formulating pre-emptive policies.

What sets Article I apart in this literature is that it summarizes a broad set of early warning indicators identified by the previous studies and benchmarks them against each other in a much more comprehensive way than is usually done. Not only does the study consider more indicators than any earlier single study, but it also considers various transformations, alternative prediction horizons, alternative crisis datasets, and both in-sample and out-of-sample predictions.

Since the early days, the early warning literature has also gone forward in terms of models. Many recent articles have demonstrated the advantages of machine learning methods (e.g., Holopainen and Sarlin, 2017; Bluwstein et al., 2020), although some contrary evidence exists as well (see Beutel et al., 2019). The machine learning methods implemented in these articles are often not particularly new anymore. Article II shows that the recent advantages in the field of recurrent neural nets can be successfully harnessed to obtain more accurate predictions of banking crises.

The work in Articles I and II could be continued in many ways. Article I is restricted to European countries, but the banking crises remain an issue for a larger set of countries. It could be investigated whether the same indicators are useful for non-European advanced economies and non-advanced economies. The analysis could also be extended by linking the indicators with the level of the countercyclical capital buffer. The literature survey could be extended to a quantitative meta-analysis. In future work related to Article II, one could consider datasets with shorter frequency and other explanatory variables. It would also be interesting to try the technique for different types of crises, including currency crises and recessions.

Article III and IV dealt with the interbank money market. Part of this literature is also related to banking crises in so far as the crises can spread through the interbank linkages (see Upper and Worms, 2004). The money market can also react strongly to a banking crisis when the banks' trust toward each other is hampered (see Afonso et al., 2011; Angelini et al., 2011). The resulting scarcity for liquidity is why the ECB had to switch to fixed-rate full-allotment policy in its open market operations, to provide enough liquidity for the banks. Articles III and IV investigate pricing in the European money market during a period that encompasses the global financial crisis and the Euro area debt crisis.

The literature on interbank money markets is also vast but somewhat limited by the availability of data since the datasets are typically proprietary and often not available for research purposes. A seminal study with interbank loan data is Furfine (2001), who investigates the loan pricing in the Federal Funds market. Besides considering a much longer data set that covers the whole Euro area, Articles III and IV extend the analysis of Furfine (2001) in several directions as discussed below.

Article III investigates how fast credit risk is priced in the overnight rates. Our finding that overnight rates can lead CDS prices suggests that overnight rates can have private information. The lead seems to be related exclusively to times of market stress which is consistent with Dang et al. (2015), who describes how a money-like debt instrument can become sensitive to the borrower's risk when there sufficiently bad new concerning the items that are serving as implicit or explicit collateral. For comparison, Acharya and Johnson (2007) show that a firm's CDS prices typically lead their stock price, and Blanco et al. (2005) show that CDS prices lead corporate bond spreads. A potential extension of Article III, could make use of transaction level data of the CDS market. Such data could have an information advantage over the publically available CDS quotes (see Bilan et al. 2020).

Article IV extends the Furfine (2001) analysis of overnight rate determinants. It also analyses the effects of resolution reform on the cost of interbank funding for large banks. Our finding that large banks obtain cheaper interbank financing is consistent with the earlier studies by Furfine (2001) and Cocco et al. (2009). Also, Angelini et al. (2011) report that overnight loans become sensitive to borrower's size (among other things) following 2007-08 events, which is consistent with our findings. Also compatible with our results, Schäfer et al. (2016) find that bail-in expectations may depend on the sovereign's fiscal strength and market reactions may be more important than the implementation schedule of legal reforms.

With regard to future work with pricing determinants in the money market dataset, one could investigate alternative competition measures (such as Lerner or Boone type index). One could also examine the role of lender characteristics in loan pricing.

References

Acharya, V. V., Johnson, T. C., 2007, Insider trading in credit derivatives, Journal of Financial Economics 84: 110-141.
Available at: https://doi.org/10.1016/j.jfineco.2006.05.003.

Acharya, V.V., D. Anginer, A.J. Warburton, 2016, The end of market discipline? Investor expectations of implicit government guarantees. SSRN working paper.
Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1961656.

Afonso, G., Kovner, A., Schoar, A., 2011, Stressed, not frozen: the federal funds market in the financial crisis, Journal of Finance 66: 1109-1139.
Available at: https://doi.org/10.1111/j.1540-6261.2011.01670.x.

Ahmed, J., Anderson, C., Zarutskie, R., 2015, Are the borrowing costs of large financial firms unusual? Finance and Economics Discussion Series 2015-024. Board of Governors of the Federal Reserve System, FEDS Working Paper 24.
Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2503644.

Alessi, L., and K. Detken, 2011. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity. European Journal of Political Economy 27 (3): 520– 33.
Available at: https://doi.org/10.1016/j.ejpoleco.2011.01.003.

Angelini, P., Nobili, A., Picillo, C., 2011, The interbank market after August 2007: what has changed and why, Journal of Money, Credit and Banking 43: 923-958.
Available at: https://doi.org/10.1111/j.1538-4616.2011.00402.x.

Araten, M., Turner, C., 2013, Understanding the funding cost difference between global systemically important banks (GSIBs) and non-GSIBs in the USA, Journal of Risk Management in Financial Institutions: 387-410.
Available at: https://dx.doi.org/10.2139/ssrn.2226939.

Arciero, L., Heijmans, R., Heuver, R., Massarenti, M., Picillo, C., Vacirca, F., 2016, How to measure the unsecured money market? The Eurosystem's implementation and validation using TARGET2 data, International Journal of Central Banking 12: 247-280.
Available at: https://www.ijcb.org/journal/ijcb16q1a8.pdf.

Arora, N., Gandhi, P., Longstaff, F. A., 2012, Counterparty credit risk and the credit default swap market, Journal of Financial Economics 103: 280-293.
Available at: https://doi.org/10.1016/j.jfineco.2011.10.001.

Barnichon, R., Matthes, C., Ziegenbein, A., 2018, The Financial Crisis at 10: Will We Ever Recover? FRBSF Economic Letters, August 13, 2018.
Available at:
https://www.frbsf.org/economic-research/publications/economic-letter/2018/august/financial-crisis-at-10-years-will-we-ever-recover/.

Berg, A., Pattillo, C., 1999, Are Currency Crises Predictable? A Test, IMF Staff Papers 46(2): 1.
Available at: https://doi.org/10.2307/3867664.

Bernanke, B., 2010, Causes of the Recent Financial and Economic Crisis, Before the Financial Crisis Inquiry Commission, Washington, D.C., September 02, 2010.
Available at:
https://www.federalreserve.gov/newsevents/testimony/bernanke20100902a.htm.

Bernanke, B., Gertler, M., 1995, Inside the Black Box: The Credit Channel of Monetary Policy Transmission, Journal of Economic Perspectives, Vol. 9, No. 4, Fall 1995.
Available at: http://doi.org/10.1257/jep.9.4.27.

Bernanke, B. S., Lown, C., S., 1991, The Credit Crunch, Brooking Papers on Economic Activity, 2:1991, pp. 205-247.
Available at: https://www.brookings.edu/bpea-articles/the-credit-crunch/.

Bertrand, M., Duflo, E., Mullainathan, S., 2002, How much should we trust differences-in-differences estimates? NBER Working Paper 8841.
Available at: https://doi.org/10.3386/w8841.

Beutel, J., List, S., von Schweinitz, G., 2019, Does machine learning help us predict banking crises? Journal of Financial Stability 45: 100693.
Available at: https://doi.org/10.1016/j.jfs.2019.100693.

Bilan, A., Ongena, S., Pancaro, C., 2020, The Value of Bank Privileged Information: Evidence from the CDS Market. European Central Bank, Frankfurt. (see Bilan, 2020 for availability).

Bilan, A., 2020, Three Essays on the Consequences of Financial Market Frictions. 2020, University of Zurich, Faculty of Economics.
Available a: https://doi.org/10.5167/uzh-193077.

Blanco, R., Brennan, S., Marsh, I. W., 2005, An empirical analysis of dynamic relation between investment-grade bonds and credit default swaps, Journal of Finance LX: 2255-2281.
Available at: https://doi.org/10.1111/j.1540-6261.2005.00798.x.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., Simsek, Ö., 2020, Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach, Bank of England Staff Working Paper No. 848.
Available at:
https://www.bankofengland.co.uk/working-paper/2020/credit-growth-the-yield-curve-and-financial-crisis-prediction-evidence-from-a-machine-learning.

Borio, C., 2014, The financial cycle and macroeconomics: What have we learnt?, Journal of Banking and Finance 45: 182-198.
Available at: https://doi.org/10.1016/j.jbankfin.2013.07.031.

Bussiere, M. , Fratzscher, M., 2006, Towards a new early warning system of financial crises, Journal of International Money and Finance 25(6): 953-973.
Available at: https://doi.org/10.1016/j.jimonfin.2006.07.007.

Caggiano, G., Calice, P., Leonida, L., 2014, Early warning systems and systemic banking crises in low income countries: A multinomial logit approach, Journal of Banking and Finance 47: 258-269.
Available at: https://doi.org/10.1016/j.jbankfin.2014.07.002.

Caprio, G., Klingebiel, K., 1997, Bank Insolvencies: Bad Luck, Bad Policy, or Bad Banking?, Annual World, Bank Conference on Development Economics, pp. 79–04.

Cho, K.; van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014, Learning Phrase Representations using RNN Encoder-

Decoder for Statistical Machine Translation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734. Available at: http://dx.doi.org/10.3115/v1/D14-1179.

Cocco, J. F., Gomes, F. J., Martins, N. C., 2009, Lending relationships in the interbank market, Journal of Financial Intermediation 18(1): 24-48.
Available at: https://doi.org/10.1016/j.jfi.2008.06.003.

Dang, T. V., Gorton, G., Holmström, B., 2015, The Information Sensitivity of a Security, Working Paper.

de Jong, F., 2002, Measures of contributions to price discovery: a comparison, Journal of Financial Markets 5(3): 323-327.
Available at: https://doi.org/10.1016/S1386-4181(02)00028-9.

Demirgüc-Kunt, A., Detragiache, E., 1998, The Determinants of Banking Crises in Developed Countries, IMF Staff Papers, 45(1): 81–109.
Available at: https://www.imf.org/external/pubs/ft/wp/wp97106.pdf.

Demirgüc-Kunt, A., Detragiache, E., 2000, Monitoring Banking Sector Fragility: A Multivariate Logit Approach, World Bank Economic Review 14: 287-307.
Availalble at: https://doi.org/10.1093/wber/14.2.287.

Detken, K., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M., Casto, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N., Kakes, J., Klaus, B., Lang, J., Puzanova, N., Welz, P., 2014, Operationalizing the Countercyclical Capital Buffer: Indicator Selection, Threshold Identification and Calibration Options, ESRB Occasional Paper No. 5 (June).
Available at: https://ssrn.com/abstract=3723336.

Dickey, D. A., Fuller, W. A., 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74(366a): 427–431.
Available at: https://doi.org/10.1080/01621459.1979.10482531.

Drehmann, M., and M. Juselius, 2014, Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements, International Journal of Forecasting 30 (3): 759–80.
Available at: https://doi.org/10.1016/j.ijforecast.2013.10.002.

Engle, R. F, Granger, C. W. J., 1987, Co-integration and error correction: Representation, estimation and testing, Econometrica 55 (2): 251–276
Available at: https://doi.org/10.2307/1913236.

Furfine, C., H., 2001, Banks as monitors of other banks: Evidence from the overnight Federal Funds market, Journal of Business, 74(1): 33-57.
Available at: https://doi.org/10.1086/209662.

Geanakoplos, J., 2010, The Leverage Cycle, Chapter in NBER Macroeconomics Annual 2009, Vol. 24.
Available at: https://doi.org/10.1086/648285.

Granger, C. W. J., 1969, Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica. 37 (3): 424–438.
Available at: https://doi.org/10.2307/1912791.

Granger, C. W. J., 2003, Time series analysis, cointegration, and applications, Nobel lecture, December 8, 2003.
Available at: https://doi.org/10.1257/0002828041464669.

Gonzalo, J., Granger, C. W. J., 1995, Estimation of Common Long-Memory Components in Cointegrated Systems, Journal of Business & Economic Statistics, 13 (1): 27-35.
Available at: https://doi.org/10.2307/1392518.

Hardy, D. C., Pazarbasioglu, C., 1999, Determinants and Leading Indicators of Banking Crises: Further Evidence, IMF Staff Paper 46(3): 1.
Available at: https://www.jstor.org/stable/3867642.

Hasbrouck, J., 1995, One Security, Many Markets: Determining the Contributions to Price Discovery, Journal of Finance 1(4): 11751199.
Available at: https://doi.org/10.1111/j.1540-6261.1995.tb04054.x.

Hochreiter, S., Schmidhuber, J., 1997, Long short-term memory, Neural Computation, 9(8): 1735–1780.
Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

Holmström, B., 2015, Understanding the role of debt in the financial system, BIS Working Papers No. 479.
Available at: https://www.bis.org/publ/work479.pdf.

Holmström, B., Tirole, J., 1997, Financial Intermediation, Loanable Funds, and the Real Sector, The Quarterly Journal of Economics 112(3): 663-691.
Available at: https://doi.org/10.1162/003355397555316.

Holopainen, M., Sarlin, P. 2017, Toward robust early-warning models: a horse race, ensembles and model uncertainty, Quantitative Finance 17: 1933–1963.
Available at: https://doi.org/10.1080/14697688.2017.1357972.

Hughes, J. P., Mester, L. J., 2013, Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function, Journal of Financial Intermediation 22: 559-585.
Available at: https://doi.org/10.1016/j.jfi.2013.06.004.

Johansen, S, Juselius, K., 1990, Maximum likelihood estimation and inference on cointegration – with application to the demand for money, Oxford Bulletin of Economics and Statistics 52(2): 169-210.
Available at: https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x.

Jordà, Ò., Schularick, M., Taylor, A. M., 2017, Macrofinancial History and the New Business Cycle Facts, NBER Macroeconomics Annual 2016, volume 31, edited by Martin Eichenbaum and Jonathan A. Parker. Chicago: University of Chicago Press.
Available at: https://doi.org/10.1086/690241.

Kaminsky, G., Reinhart, C., 1999, The twin crises: the causes of banking and balance-of-payments problems, American Economic Review 89(3): 473-500.
Available at: https://doi.org/10.1257/aer.89.3.473.

Kilian, L., Lütkepohl, H., 2017, Structural Vector Autoregressive Analysis, Cambridge University Press, Themes in Modern Econometrics Series.

Kiyotaki, N., Moore, J., 1997, Credit Cycles, Journal of Political Economy, 105(2): 211248.
Available at: https://doi.org/10.1086/262072.

Kuan, C.-H., White, H., 2007, Artificial neural networks: an econometric perspective, Econometric Reviews 13(1): 1-91.
Available at: https://doi.org/10.1080/07474939408800273.

Laeven, L., Valencia, F., 2012, Systemic Banking Crises Database: An Update, IMF Working Paper No. 08/224.
Available at:
https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Systemic-Banking-Crises-Database-An-Update-26015.

Lipovetsky, S., Conklin, M., 2001, Analysis of regression in game theory approach, Stochastic Models in Business and Industry 17(4): 319–330.
Available at: https://doi.org/10.1002/asmb.446.

Lundberg, S. M., Lee, S.-I., 2017, A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems 30.
Available at: https://dl.acm.org/doi/10.5555/3295222.3295230.

McLeay, M., Radia, A., Thomas, R., 2014, Money creation in the modern economy, Bank of England Quarterly Bulletin, 14 March 2014.

Mukunda, G., 2018, The Social and Political Costs of the Financial Crisis, 10 Years Later, Harvard Business Review, September 25 Issue.

Murphy, K. P., 2012, Machine Learning: A Probabilistic Perspective, The MIT Press, Cambridge, Massachusetts, London, England.

Pepe, M., Longton, G., Janes, H., 2009, Estimation and Comparison of Receiver Operating Characteristic Curves, Stata Journal 9(1): 1.

Petersen, M. A., 2008, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, Review of Financial Studies 22(1) : 435–480.
Available at: https://doi.org/10.1093/rfs/hhn053.

Reinhart, C. M., Rogoff, K. S., 2009, This Time Is Different: Eight Centuries of Financial Folly, Princeton University Press.

Rogers, W. H., 1993, Regression standard errors in clustered samples, Stata Technical Bulletin 13: 19–23.

Schularick, M., Taylor, A. M., 2012, Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870-2008, American Economic Review, 102(2): 1029-1061.
Available at: https://doi.org/10.1257/aer.102.2.1029.

Schäfer, A., Schnabel, I., Weder di Mauro, B., 2016, Bail-in expectations for European banks: actions speak louder than words, CEPR Discussion Paper No. 11061.

Shapley, L. S., 1953, A value for n-person games, Contributions to the Theory of Games, 2(28): 307–317.

Sims, C. A., 1980, Macroeconomics and reality, Econometrica 48: 1–48.
Available at: https://doi.org/10.2307/1912017.

Upper, C., Worms, A., 2004, Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? European Economic Review 48(4): 827-849.
Available at: https://doi.org/10.1016/j.euroecorev.2003.12.009.

Wiener, N., 1956, The Theory of Prediction, Modern Mathematics for Engineers, Series 1 (Eds. B. F. Beckenback), Chapter 8.

Annex: Articles I-IV

Article I

Tölö, E., Laakkonen, H., and Kalatie, S., 2018, "Indicators used in setting the countercyclical capital buffer," *International Journal of Central Banking*, Vol. 14, No. 2, pp. 52–111.

Available at: https://www.ijcb.org/journal/ijcb18q1a2.htm.

Evaluating Indicators for Use in Setting the Countercyclical Capital Buffer*

Eero Tölö, Helinä Laakkonen, and Simo Kalatie
Bank of Finland

The European Systemic Risk Board (ESRB) recently issued a recommendation on the use of early warning indicators in macroprudential decisions involving the countercyclical capital buffer (Basel III framework). In addition to a primary indicator, deviation in the credit-to-GDP ratio from long-term trend, the ESRB advises the use of supplemental indicators to measure private-sector credit developments and debt burden, overvaluation of property prices, external imbalances, mispricing of risk, and strength of bank balance sheets. Based on empirical analysis of data for European Union countries, a large assortment of potential indicators, and comprehensive robustness checks, we propose specific suitable early warning indicators for each of the six risk categories set forth by the ESRB.

JEL Codes: G01, G28.

1. Introduction

The purpose of the countercyclical capital buffer proposed by the Basel Committee on Banking Supervision (BCBS 2011) is to

*The authors would like to thank Mikael Juselius and Tuomas Peltonen at the Annual Meeting of the Finnish Economic Association and the participants in the 2015 International Symposium of Forecasting in Riverside, California, for their valuable comments. Our gratitude also goes to Esa Jokivuolle, Karlo Kauko, Hanna Putkuri, Katja Taipalus, Jouni Timonen, Jouko Vilmunen, and Matti Virén for their insights, Gregory Moore for proofreading the manuscript, and Timo Virtanen for research assistance. Finally, we thank our anonymous referees for their help in greatly improving the manuscript. The views presented are those of the authors and do not necessarily represent the views of the Bank of Finland. Any remaining errors are solely ours. Corresponding author e-mail: eero.tolo@bof.fi.

mitigate credit booms and related procyclicality in the financial system. When there are signs of excessive credit growth and emerging vulnerabilities related to the credit cycle, the BCBS advises monetary authorities to raise bank capital requirements. The buffer requirement, which is intended to improve bank resilience against future losses, may also slow credit growth as capital requirements are adjusted to a higher level.[1] To properly time adjustments in the countercyclical capital buffer level, policymakers must have some certainty that they have correctly identified the emergence of cyclical vulnerabilities.

The countercyclical capital buffer requirement was implemented under the European Union's (EU's) 2013 Capital Requirements Directive.[2] In determining appropriate buffer requirements, national authorities are advised to follow the BCBS harmonized buffer guide[3] and the European Systemic Risk Board (ESRB) guidance and official recommendations,[4] as well as to take into consideration domestic conditions relevant to cyclical vulnerabilities. The ESRB's official recommendation (ESRB 2014), based on the results of the empirical study by Detken et al. (2014), instructs policymakers to use a set of indicators that encompasses six risk categories: credit developments, potential overvaluation of property prices, private-sector debt burden, external imbalances, potential mispricing of risk, and strength of bank balance sheets. Beyond that, however, there is little guidance on the specific indicators to apply in each of these risk categories. Given the tangible economic consequences of capital requirements

[1] There are not yet many empirical impact studies on the countercyclical capital buffer due to the limited amount of data on policy decisions. See Akinci and Olmstead-Rumsey (2015), Cerutti, Claessens, and Laeven (2017), and Cerutti et al. (2016) for some early empirical evidence.

[2] CRD IV 2013/36/EU.

[3] The buffer guide is based on the deviation of the ratio of credit to GDP from its long-term trend calculated following the methodology of the BCBS with a one-sided Hodrick-Prescott filter and smoothing parameter $\lambda = 400{,}000$ (i.e., credit-to-GDP gap). When this trend gap is below (above) or equal to 2 percent (10 percent), the buffer guide suggests a 0 percent (2.5 percent) countercyclical capital buffer. Within the gap band, the countercyclical capital buffer would depend linearly on the trend gap.

[4] Although characterized as recommendations, they are not taken lightly by national policymakers. Compliance is monitored via an "act or explain" mechanism.

(Van den Heuvel 2008) and the economic impacts of such indicators in decisionmaking, it would be valuable for policymakers to have the clearest possible grasp of these state-of-the-art indicators in each category before issuing a buffer rate decision.

This empirical work continues that of Detken et al. (2014) with the aim of identifying informative warning indicators for the six risk categories. Using an unbalanced quarterly panel of twenty-eight EU countries for the period 1970 to 2012 as our data set, we consider roughly fifty conceptually varied indicators from national accounts, financial accounts, balance of payments, financial markets, and bank balance sheets. When all transformations are included, the number of considered indicators rises to nearly 400. Our indicator set brings together indicators identified in earlier studies and examines them in a consistent setup. We also include several theoretically motivated indicators that, to our knowledge, have never been studied in this context: the VIX index, the ratio of cross-border loans to GDP or assets, the spread between high-yield and investment-grade corporate bonds, benchmark government bond yields, household interest expense burden, and balance sheet indicators based on liquidity and short-term funding.

Indicator performance is assessed with standard measures from the early warning literature. We apply receiver operating characteristic (ROC) and relative usefulness analyses, which are both based on the relative numbers of type I (false positive) and type II (false negative) errors of the warning signals. The indicators are examined using most parsimonious non-parametric and parametric methods full sample and out of sample in a large panel of countries. Different crisis-prediction horizons and alternative financial crisis data sets are considered.

This work contributes to the current policy discussion on the EU legislative framework for countercyclical capital buffers. Due to the huge diversity of possible indicators in the six risk categories, we are compelled to investigate simultaneously a set of indicators larger than in any previous study. Our common evaluation setup facilitates thorough robustness checks and equal treatment of predictor performance that would otherwise be difficult to compare among existing findings. While the earlier literature has shown that combining multiple indicators into a composite indicator can improve signaling power, we focus mainly on individual indicators in order to identify specific robust indicators for each prescribed category.[5]

In line with the earlier literature (see the literature review in section 2.2), we find that measures of credit developments, especially those based on the credit-to-GDP ratio, are historically among the best predictors of financial crises. We further note that measures of private-sector debt burden and overvaluation of property prices (e.g., debt-service ratios and relative house prices) are highly useful. To our best knowledge, this is also the first study to identify the VIX index, the high-yield corporate bond spread, and benchmark government bond yields as useful indicators in this context. We report evidence of statistically significant predictive power of many indicators in the external imbalances, mispricing of risk, and bank balance sheet categories, including the ratio of current account to GDP, the ratio of cross-border loans to GDP, various measures based on stock market prices, the leverage ratio, and the ratio of total bank assets to GDP. Drawing on these findings, we recommend a practical set of indicators that appear to be relatively good predictors of financial crises and that meet the provisions of the ESRB recommendation.

The robustness checks with the alternative prediction horizons reveal that the indicators have no unique ranking in terms of performance. Instead, the predictors work optimally at different prediction horizons, a feature that could be quite valuable in policy decisions. Moreover, changing the crisis data set sometimes has a large impact on evaluated performance, underscoring the challenge of predicting financial crises without a clear definition of what constitutes a crisis.

The paper is organized as follows. Section 2 discusses the operationalization of the countercyclical capital buffer (2.1), along with the early warning indicator literature and our list of potential indicators to be considered in each of the ESRB's proposed categories (2.2). The data and empirical techniques are discussed in section 3, which presents the data sources and transformations (3.1–3.2), and reviews the concepts of signal extraction (3.3) as well as ROC analysis, usefulness measures, and the evaluation process (3.4–3.5). Section 4 presents the main results and our recommended set of indicators (4.1), results with alternative crisis-prediction horizons (4.2),

[5]Aikman et al. (2014) suggest that simple indicators often outperform more complex alternatives when there is uncertainty.

and alternative crisis data sets (4.3). Section 4.4 discusses various frameworks on how indicators might be interpreted or embedded in a monitoring framework. Section 5 concludes.

2. Early Warning Indicators Identified in the Previous Literature

In this section, we review the ESRB recommendation on operationalizing the countercyclical capital buffer and recent literature seeking a similar goal to ours, i.e., identification of indicators to be considered when setting the countercyclical capital buffer.[6] We next discuss, based on empirical evidence presented in the literature or conceptual relevance, each indicator category and potential indicators to be analyzed in the empirical part of this work.

2.1 Operationalizing the Countercyclical Capital Buffer

The ESRB recommendation (ESRB 2014) says that level adjustments of the countercyclical capital buffer should be based primarily on deviation of the private-sector credit-to-GDP ratio from its long-term trend (credit-to-GDP gap). Indeed, a number of empirical studies support the view that the credit-to-GDP gap is the best single indicator in predicting a banking crisis.[7] However, as there are potentially large uncertainties for the signals given by any single indicator, the ESRB recommends that authorities base their decisions on a wide set of information that captures the vulnerabilities caused by excessive credit growth and note six categories of risk usually associated with excessive credit growth.[8]

[6]Kauko (2014) provides a comprehensive literature survey on early warning indicators.

[7]See, e.g., Babecký et al. (2014), Behn et al. (2013), Bonfim and Monteiro (2013), Detken et al. (2014), Drehmann, Borio, and Tsatsaronis (2011), Drehmann et al. (2010), and Drehmann and Juselius (2014). For criticism, see, e.g., Repullo and Saurina (2011).

[8]The ESRB recommendation has a seventh category of indicators that includes indicators that combine information on the credit-to-GDP gap and indicators from the six alternative indicator categories. We do not consider these seventh-category indicators in our empirical analysis for two reasons. First, selection of these indicators only occurs after the suitable indicators for the other six categories have been determined. Second, the ESRB recommendation provides no guidance on calculation or public disclosure of seventh-category indicators.

In addition to the credit-to-GDP gap, the recommendation calls on authorities to monitor and publicly disclose at least one other indicator per category to accompany a countercyclical capital buffer adjustment. The six indicator categories are measures of

(i) credit developments,
(ii) private-sector debt burden,
(iii) potential overvaluation of property prices,
(iv) external imbalances,
(v) potential mispricing of risk, and
(vi) strength of bank balance sheets.

With respect to the actual indicators that describe these six categories, the ESRB only offers suggestions based on an empirical analysis by Detken et al. (2014). It does not provide specific recommendations, and thus leaves the decision on which specific indicators to use to the national authorities.

2.2 The Literature and Candidate Indicators for the Six Categories

We provide an extensive survey table of early warning indicators studied in earlier empirical works (see table 1). We make an attempt to incorporate most of the published research articles and some relevant unpublished works that evaluate early warning indicators of banking crises using panel data. Studies that rely on data on a single country are not included. Due to disparate approaches of the papers, it is not possible to incorporate much detail or to do full justice to earlier findings.

Within the voluminous literature of financial crises, there are several recent studies that focus on identifying indicators for guiding decisions on the countercyclical capital buffer.

In addition to the above-mentioned study of Detken et al. (2014), Behn et al. (2013) evaluate a wide set of macrofinancial and banking-sector indicators using data for EU member states. In addition to domestic factors such as credit developments and equity and house prices, they suggest that global variables on house prices and credit

Table 1. Survey of Early Warning Indicators

	Ferrari and Pirovano (2015)	Holopainen and Sarlin (2015)	Jordà, Schularick, and Taylor (2015)	Detken et al. (2014)	Anundsen, Gerdrup, and Hansen (2014)	Babecký et al. (2014)	Drehmann and Juselius (2014)	Lainà, Nyholm, and Sarlin (2014)	Behn et al. (2013)	Bonfim and Monteiro (2013)	Hahm, Shin, and Shin (2013)	Lo Duca and Peltonen (2013)	Bordo and Meissner (2012)	Kauko (2012a)	Kauko (2012b)	Roy and Kemme (2012)	Schularick and Taylor (2012)	Alessi and Detken (2011)	Drehmann, Borio, and Tsatsaronis (2011)	Barrell et al. (2010)	Bunda and Ca' Zorzi (2010)	Büyükkarabacak and Valev (2010)	Joyce (2011)	Borio and Drehmann (2009)	Davis and Karim (2008)	Von Hagen and Ho (2007)	Domaç and Peria (2003)	Demirgüç-Kunt and Detragiache (2000)	Kaminsky and Reinhart (1999)	Hardy and Pazarbaşıoğlu (1998)
Crisis Data Set / Target Variable:	**B**	**L**	**C**	**D**	**C**	**B**	**L**	**C**	**B**	**D**	**O**	**FSI**	**C**	**NPL**	**NPL**	**R**	**C**	**O**	**C**	**L**	**R**	**C**	**CK**	**C**	**C**	**O**	**DD**	**DD**	**R**	**LI**
No. of Countries:	**15**	**15**	**17**	**28**	**16**	**40**	**26**	**11**	**23**	**9**	**30**	**28**	**14**	**25**	**34**	**14**	**14**	**18**	**36**	**14**	**76**	**37**	**20**	**18**	**105**	**47**	**88**	**34**	**20**	**38**
1. Credit Developments																														
Total Credit to Private Sector		x	x	x	x	x	x	x	x	x		o	x			x	x		x	x		x			~	o	x	x		x
Bank Credit to Private Sector				x											x							x								
Household Credit						~		x														x								
Mortgage Loans								x																						
Corporate Credit				x																		x								
Public Credit										x																				
Global Credit									x									x												
Credit-to-GDP Ratio	x	x		x	x		x	x	x	x	x								x		x			x	o	o	o		x	
Bank Credit-to-GDP Ratio	x			x													x						x							
Household Credit-to-GDP Ratio	x			x	x																									
Mortgage Loans-to-GDP Ratio								x																						
Corporate Credit-to-GDP Ratio					x																									
Public Credit-to-GDP Ratio		x		x		x				x						o					x									
Global Credit-to-GDP Ratio					o	o			x																					
Differenced Relative Total Credit		o		x											x															
Loans-to-Income Ratio																														
2. Private-Sector Debt Burden																														
Real Mortgage Interest Rate		x								x																				
Debt-Service Ratio			x		x				x																					
Household Debt-Service Ratio					x																									
Corporate Debt-Service Ratio					o																									

(*continued*)

Table 1. (Continued)

	Ferrari and Pirovano (2015)	Holopainen and Sarlin (2015)	Jordà, Schularick, and Taylor (2015)	Detken et al. (2014)	Anundsen, Gerdrup, and Hansen (2014)	Babecký et al. (2014)	Drehmann and Juselius (2014)	Lainà, Nyholm, and Sarlin (2014)	Behn et al. (2013)	Bonfim and Monteiro (2013)	Hahm, Shin, and Shin (2013)	Lo Duca and Peltonen (2013)	Bordo and Meissner (2012)	Kauko (2012a)	Kauko (2012b)	Roy and Kemme (2012)	Schularick and Taylor (2012)	Alessi and Detken (2011)	Drehmann, Borio, and Tsatsaronis (2011)	Barrell et al. (2010)	Bunda and Ca' Zorzi (2010)	Büyükkarabacak and Valev (2010)	Joyce (2011)	Borio and Drehmann (2009)	Davis and Karim (2008)	Von Hagen and Ho (2007)	Domaç and Peria (2003)	Demirgüç-Kunt and Detragiache (2000)	Kaminsky and Reinhart (1999)	Hardy and Pazarbaşıoğlu (1998)
Crisis Data Set / Target Variable:	**B**	**L**	**C**	**D**	**C**	**B**	**L**	**C**	**B**	**D**	**O**	**FSI**	**C**	**NPL**	**NPL**	**R**	**C**	**O**	**C**	**L**	**R**	**C**	**CK**	**C**	**C**	**O**	**DD**	**DD**	**R**	**LI**
No. of Countries:	**15**	**15**	**17**	**28**	**16**	**40**	**26**	**11**	**23**	**9**	**30**	**28**	**14**	**25**	**34**	**14**	**14**	**18**	**36**	**14**	**76**	**37**	**20**	**18**	**105**	**47**	**88**	**34**	**20**	**38**
3. Potential Overvaluation of Property Prices																														
House Price	x	x	x	x		o	x	x	x	x						x			x	x	x			x						
House Price / Income	x	x		x	x																									
House Price / Rent	x			x																										
Global House Prices									x																					
Global House Price / Income					x																									
Commercial Real Estate Price				x																										
4. External Imbalances																														
Current Account / GDP		x		x		x		x		x		x	o		x	x					x									o
Trade Balance						o																	o							
Trade / GDP						x																								
Exports															o														x	
Imports															o														x	x
Capital Flows / GDP																											o			
Foreign Assets																							o							
Foreign Liabilities																							x							x
Foreign Liabilities / Foreign Assets																											o			
Foreign Direct Investment (Decrease)																							x							
Foreign Portfolio Investment (Decrease)																							x							
Terms of Trade						x																o			x		o	o	x	o

(*continued*)

Table 1. (Continued)

	Ferrari and Pirovano (2015)	Holopainen and Sarlin (2015)	Jordà, Schularick, and Taylor (2015)	Detken et al. (2014)	Anundsen, Gerdrup, and Hansen (2014)	Babecký et al. (2014)	Drehmann and Juselius (2014)	Lainà, Nyholm, and Sarlin (2014)	Behn et al. (2013)	Bonfim and Monteiro (2013)	Hahm, Shin, and Shin (2013)	Lo Duca and Peltonen (2013)	Bordo and Meissner (2012)	Kauko (2012a)	Kauko (2012b)	Roy and Kemme (2012)	Schularick and Taylor (2012)	Alessi and Detken (2011)	Drehmann, Borio, and Tsatsaronis (2011)	Barrell et al. (2010)	Bunda and Ca' Zorzi (2010)	Büyükkarabacak and Valev (2010)	Joyce (2011)	Borio and Drehmann (2009)	Davis and Karim (2008)	Von Hagen and Ho (2007)	Domaç and Peria (2003)	Demirgüç-Kunt and Detragiache (2000)	Kaminsky and Reinhart (1999)	Hardy and Pazarbaşıoğlu (1998)
Crisis Data Set / Target Variable:	**B**	**L**	**C**	**D**	**C**	**B**	**L**	**C**	**B**	**D**	**O**	**FSI**	**C**	**NPL**	**NPL**	**R**	**C**	**O**	**C**	**L**	**R**	**C**	**CK**	**C**	**C**	**O**	**DD**	**DD**	**R**	**LI**
No. of Countries:	**15**	**15**	**17**	**28**	**16**	**40**	**26**	**11**	**23**	**9**	**30**	**28**	**14**	**25**	**34**	**14**	**14**	**18**	**36**	**14**	**76**	**37**	**20**	**18**	**105**	**47**	**88**	**34**	**20**	**38**
Exchange Rate	x			o		o						o				o					x				o	~		o	x	x
Foreign Exchange Reserves																														o
5. Potential Mispricing of Risk																														
Short-Term Interest Rate	x			x		x			x		o		~					x												
Long-Term Interest Rate	x	x		x					o	x																				
Yield Curve						x																								
Lending Rate / Deposit Rate																													x	
Stock Returns	x	x	~	x		~	o		o	x		x				x	o	x	x					x		o			x	
Global Stock Returns									x																					
Aggregate Asset Prices																		x						x						
LIBOR-OIS Spread																			x											
Corporate Bond Spread						x													o											
6. Strength of Bank Balance Sheets																														
Leverage Ratio				o	x				x			x								x										
Bank Profits									x										o											
Bank Deposits																													x	x
Loan / Deposits								x																						
Non-Core Liabilities					x		x				x																			
Banks Net Foreign Assets																														o
Bank Reserves / Assets																						o								
Bank Liquidiity															o					x						o	o			

(*continued*)

Table 1. (Continued)

	Ferrari and Pirovano (2015)	Holopainen and Sarlin (2015)	Jordà, Schularick, and Taylor (2015)	Detken et al. (2014)	Anundsen, Gerdrup, and Hansen (2014)	Babecký et al. (2014)	Drehmann and Juselius (2014)	Lainà, Nyholm, and Sarlin (2014)	Behn et al. (2013)	Bonfim and Monteiro (2013)	Hahm, Shin, and Shin (2013)	Lo Duca and Peltonen (2013)	Bordo and Meissner (2012)	Kauko (2012a)	Kauko (2012b)	Roy and Kemme (2012)	Schularick and Taylor (2012)	Alessi and Detken (2011)	Drehmann, Borio, and Tsatsaronis (2011)	Barrell et al. (2010)	Bunda and Ca' Zorzi (2010)	Büyükkarabacak and Valev (2010)	Joyce (2011)	Borio and Drehmann (2009)	Davis and Karim (2008)	Von Hagen and Ho (2007)	Domaç and Peria (2003)	Demirgüç-Kunt and Detragiache (2000)	Kaminsky and Reinhart (1999)	Hardy and Pazarbaşıoğlu (1998)
Crisis Data Set / Target Variable:	**B**	**L**	**C**	**D**	**C**	**B**	**L**	**C**	**B**	**D**	**O**	**FSI**	**C**	**NPL**	**NPL**	**R**	**C**	**O**	**C**	**L**	**R**	**C**	**CK**	**C**	**C**	**O**	**DD**	**DD**	**R**	**LI**
No. of Countries:	**15**	**15**	**17**	**28**	**16**	**40**	**26**	**11**	**23**	**9**	**30**	**28**	**14**	**25**	**34**	**14**	**14**	**18**	**36**	**14**	**76**	**37**	**20**	**18**	**105**	**47**	**88**	**34**	**20**	**38**
Banking-Sector CDS Spread																			x											
Financial-Sector Size																	x													
7. Real Economy Variables																														
GDP	o	x		x			o		o	o		x	~		o	x		x	x	o	o	o	x		x	x	o	x		x
Global GDP						x			o		o																			
Real GDP per Capita															x							x			~	o	o	o		
Output Gap					x																									
Unemployment				x						o																				
Industrial Production						x																							x	
Industry Share of GDP						x																								
Consumption						o												x												x
Investment						o					o							x												o
Capital / Output																														o
Government Consumption						o																								
Fiscal Deficit										~					o					o					o	x	o	o	x	
Net National Savings						o																								
Global Trade						o																								
8. Monetary Aggregates																														
Reserves																													x	o
M1						o				x								x								o			x	
Global M1																		x												
M2																			x											
M2 / Reserves																x				o		x	o		o		x	x	x	
M3					x	x									o															

(*continued*)

Table 1. (Continued)

	Ferrari and Pirovano (2015)	Holopainen and Sarlin (2015)	Jordà, Schularick, and Taylor (2015)	Detken et al. (2014)	Anundsen, Gerdrup, and Hansen (2014)	Babecký et al. (2014)	Drehmann and Juselius (2014)	Lainà, Nyholm, and Sarlin (2014)	Behn et al. (2013)	Bonfim and Monteiro (2013)	Hahm, Shin, and Shin (2013)	Lo Duca and Peltonen (2013)	Bordo and Meissner (2012)	Kauko (2012a)	Kauko (2012b)	Roy and Kemme (2012)	Schularick and Taylor (2012)	Alessi and Detken (2011)	Drehmann, Borio, and Tsatsaronis (2011)	Barrell et al. (2010)	Bunda and Ca' Zorzi (2010)	Büyükkarabacak and Valev (2010)	Joyce (2011)	Borio and Drehmann (2009)	Davis and Karim (2008)	Von Hagen and Ho (2007)	Domaç and Peria (2003)	Demirgüç-Kunt and Detragiache (2000)	Kaminsky and Reinhart (1999)	Hardy and Pazarbaşıoğlu (1998)
Crisis Data Set / Target Variable:	B	L	C	D	C	B	L	C	B	D	O	FSI	C	NPL	NPL	R	C	O	C	L	R	C	CK	C	C	O	DD	DD	R	LI
No. of Countries:	15	15	17	28	16	40	26	11	23	9	30	28	14	25	34	14	14	18	36	14	76	37	20	18	105	47	88	34	20	38
9. Inflation Variables																														
Commodity Prices						o																								
Inflation	o	x		x		o		x	~	o		x			o					o		~	x		~	o	x	x		x
Global Inflation						o																								
Real Interest Rate	x			o												x				o		x			~	x	o	x	x	o
10. Other Variables																														
Income Inequality													o			x														
Deposit Insurance Scheme																										x				
Finanical Deregulation																							x			o	x			
Capital Requirement																							x							
Contagion																							x							
Fixed Exchange Rate																							x				x			
Tax Burden						x																								

Notes: B = Babecký et al. (2014), C = compilation of various crisis databases, CK = Caprio and Klingebiel (1996), D = Detken et al. (2014), DD = Demirgüç-Kunt and Detragiache (1998), FSI = financial stress index as target variable, L = Laeven and Valencia (2012), LI = Lindgren et al. (1996), NPL = NPL ratio as target variable, O = author's own setup, R = Reinhart and Rogoff (2009). o = No significant predictive performance; x = At least some predictive performance; ∼ = Ambiguous predictive performance. Few studies consider the significance of individual indicator variables. For studies that only present multivariate regressions, we have sought to identify whether the relevant predictor tends to be a statistically significant predictor of banking crises. However, due to disparate methods taken by different authors, the predictive performance is not comparable across their works. This may result in inconsistencies in the survey table. Even so, the information about where different types of indicators have been considered should be helpful in itself. In order to keep the number of different indicators feasible, first, we do not distinguish between different transformations of the same underlying variable, and second, we typically merge real and nominal variables into a single line.

developments have good forecasting properties.[9] Importantly, their multivariate approach provides superior crisis prediction relative to the traditional univariate approach, i.e., policymakers are likely to benefit from using a wide range of indicators in setting the countercyclical buffer rate.

Following Behn et al. (2013), Anundsen, Gerdrup, and Hansen (2014) propose a set of multivariate early warning models to guide policymakers in adjustment of the countercyclical capital buffer. They find that indicators on household credit developments predict crises better than those of non-financial corporations and that global housing market imbalances may be useful in signaling a crisis. They also propose a novel measure of housing and credit market exuberance based on the time-series methods proposed by Phillips, Shi, and Yu (2013).

Bonfim and Monteiro (2013) discuss suitable indicators for implementation of the countercyclical capital buffer. Their empirical analysis of nine European countries suggests that policymakers need to carefully monitor indicators on house and stock prices and credit developments.

In addition, a number of authorities have published single-country studies to justify their choice of indicators. Using Spanish data, Castro, Estrada, and Martinez (2014) analyze a group of potential additional indicators. In their analysis of the United Kingdom, Giese et al. (2014) suggest several complementary indicators for use alongside the credit-to-GDP gap.

In the following subsections, we continue this literature review beyond the studies focused explicitly on application to countercyclical capital buffer indicator and propose candidate indicators for each of the six categories in the ESRB recommendation. Detailed data definitions are provided in section 3.1.

2.2.1 Credit Developments

Credit growth is probably the most-analyzed indicator measuring credit developments. It has been found to be a statistically significant

[9]They remind us that the success of these variables might at least partly be explained by the global financial crisis, which causes a strong clustering of crisis episodes in the data.

predictor of banking crises in numerous studies (see, e.g., Schularick and Taylor 2012 and the references in table 1).

Nevertheless, other potential indicators should not be ruled out. For starters, we should consider the scope of credit indicators. Do we define credit as total credit that incorporates all credit regardless of the creditor or just the credit provided by the banks? Do we consider long-term growth rates such as three-year growth or absolute changes in credit levels in lieu of yearly growth rates? Do we acknowledge that private-sector, household, and non-financial corporation credit growth may each possess different signaling power with respect to an emerging banking crisis?[10]

There are also indicators that are quite similar to the benchmark indicator (credit-to-GDP gap calculated with a one-sided Hodrick-Prescott (HP) filter) that may contain additional relevant information helpful in predicting crises. For example, the credit-to-GDP gap could be analyzed separately for households and non-financial corporations. These indicators can be seen as augmenting credit-to-GDP gap information with detailed information on what underlies the primary indicator signal.

A well-known weakness of the credit-to-GDP gap is that it tends to increase when GDP declines (Repullo and Saurina 2011). In a slowing real economy, it may even be counterproductive to raise buffers. Indeed, if credit growth has already come to a halt, higher capital requirements could induce a large negative shock to the economy. Kauko (2012a) proposes two credit development measures that compare the one-year change in credit to the five-year moving average of GDP. The first measure is

$$X_{1,t} = \frac{5L_t}{\sum_{i=0}^{4} Y_{t-i}} - \frac{5L_{t-1}}{\sum_{i=1}^{5} Y_{t-i}}, \tag{1}$$

where L_t is the outstanding debt and Y_t is the GDP in year t. The second measure is such that the differencing is applied only to the debt variable,

[10]For example, Anundsen, Gerdrup, and Hansen (2014), Büyükkarabacak and Valev (2010), and Detken et al. (2014) all find that indicators of household credit developments are better at predicting banking crises than indicators of non-financial corporations.

$$X_{2,t} = \frac{5(L_t - L_{t-1})}{\sum_{i=0}^{4} Y_{t-i}}. \tag{2}$$

Kauko (2012a) argues that using a five-year moving average of GDP instead of yearly GDP addresses the problem of large short-term declines in GDP that hamper the use of the benchmark indicator. Detken et al. (2014) confirm that the indicator in which the credit change is divided by the one-year moving average of the GDP is among the best indicators for describing credit developments that foreshadow systemic financial crises.

For measuring credit developments, we consider the real credit and credit-to-GDP ratios. In each case, we consider four definitions of credit: total credit to non-financial private sector, total credit to households, total credit to non-financial corporations, and bank credit to private non-financial sector. Total credit includes loans and debt securities, irrespective of the creditor sector as reported in the financial accounts. Bank credit only includes credit where the creditor belongs to the banking sector.

2.2.2 Private-Sector Debt Burden

Private-sector indebtedness is unsustainable when borrowers can no longer meet their debt-servicing obligations. High private-sector indebtedness generates credit risk for banks and may depress consumption and investment throughout the economy. Indeed, both the debt-to-income ratio and the debt-service ratio have been found useful in signaling financial crises (e.g., Detken et al. 2014; Drehmann and Juselius 2014; Giese et al. 2014).[11] Adverse trends in the household debt burden may matter more for financial stability than the debt burden trends of non-financial corporations. Detken et al. (2014) conclude that the non-financial corporate debt-service ratio has no predictive power for banking crises.

Public data sources do not typically provide data on debt-servicing ratios.[12] Here, we use the data set collected for Detken et al. (2014). We also construct proxy indicators of the interest expense

[11] The debt-service ratio measures the interest rate and amortization costs of the debt relative to income.

[12] The Bank for International Settlements (BIS) recently began to post debt-service ratio data on its website.

burden without amortization costs. The constructed indicators are relevant in countries where mortgages have floating rates that move with market interest rates.[13] The first indicator is calculated by multiplying the household credit-to-GDP ratio by the three-month money market rate. The second indicator is calculated similarly, but the ten-year government bond interest rate replaces the three-month money market rate.

2.2.3 Potential Overvaluation of Property Prices

Variables related to developments in the real estate sector have been found useful in predicting banking crises (e.g., Jordà, Schularick, and Taylor 2015; see table 1). In particular, the combination of strong credit growth and rising house prices has been identified as threatening to financial stability (Barrel et al. 2011; Behn et al. 2013; Borio and Drehmann 2009; Jordà, Schularick, and Taylor 2015).

Credit and house prices tend to move hand-in-hand. House purchases are typically financed with loans, and house value affects the decision to grant a loan through the collateral process. Mortgages also typically make up a large share of household and bank balance sheets, making both vulnerable to swings in housing prices. In a downturn, the substantial losses to banks caused by defaults on household mortgages and loans to construction companies may be exacerbated by losses on other corporate lending caused by contractions in output and consumption. Many banks use mortgages to secure their own market-based funding, so a sharp negative correction in house prices may also increase costs of funding for troubled banks.

The state of the housing market can be assessed by comparing house prices with household income or housing rents. Relative developments in house prices and income reflect the affordability of housing from the buyer's point of view, while the relationship between housing prices and rents is conceptually identical to the stock market price-to-earnings ratio. Detken et al. (2014) find that

[13]In Finland, for example, mortgage interest rates are typically tied to EURIBOR rates (plus a fixed spread). Prime rates of European banks also typically track EURIBOR rates.

relative house price measures perform better in crisis prediction than other market- or real economy-based indicators.

For measuring potential overvaluation of property prices, we consider real residential property prices, the residential property price-to-rent ratio, the residential property price-to-income ratio, and commercial real estate prices.

2.2.4 External Imbalances

Indicators that measure excessive credit growth indirectly have been found useful in predicting banking crises. It is well known that when credit growth is much higher than GDP growth, domestic savings are typically insufficient to finance the credit expansion and indebtedness is financed with foreign money. Excessive foreign borrowing appears as a deficit in the current account. Many studies have found a link between large external imbalances and the frequency of financial crises. For example, Laeven and Valencia (2008) found that out of forty-one banking crisis around the world, thirty-nine countries ran current account deficits in the year preceding the crisis. There are also several studies that find a statistically significant relationship between the current account deficit and the likelihood of a banking crisis (see table 1). Joyce (2011) studies banking crises in emerging countries and concludes that an increase in foreign debt liabilities contributes to an increase in the incidence of crises, but foreign direct investment and portfolio equity liabilities have the opposite effect.

It has been argued in the literature that money originating from abroad, especially portfolio investment, provides a less stable credit source than money from domestic providers. In other words, heavy foreign borrowing may constitute a vulnerability to the financial system. Kim and Wei (2002) suggest that part of this vulnerability stems from the difficulties foreign investors have in evaluating risks in another country. This low-information condition leads to herding behavior that may trigger panicked pull-outs if risks materialize. Such investor flight may also drive up external imbalances (Kim and Wei 2002).

A number of studies consider trade- and currency-related variables such as exports, terms of trade, and exchange rate overvaluation, which are sometimes found to be statistically significant predictors (see table 1). We did not examine such variables in this study in order to steer away from the currency crises literature, which is beyond the scope of this paper.

Hence, we consider the current account deficit (ratio of current account to GDP), capital account deficit, ratio of portfolio investments to GDP, and other investment-to-GDP ratios as indicators for external imbalances. We also consider separately cross-border loans in foreign currency and domestic currency (divided by GDP) coming from abroad.

2.2.5 Potential Mispricing of Risk

Credit and asset price booms are typically associated with times of positive economic developments. During long periods of good times, agents may become oblivious to certain types of risk, which may be reflected as banks loosening their credit standards or investors demanding lower risk premia for risky securities.

In the securities markets, one might look for trends in the stock and bond markets. Rapid price increases on the stock market or high stock valuations (e.g., share prices relative to dividend yields, i.e., price-earnings (P/E) ratios) or a rapid decrease in the required risk premiums between safe and risky corporate bonds might reflect increased risk appetite among investors that leads to a mispricing of risk. Moreover, low levels of asset return volatility typically lead to increased risk-taking, i.e., in times of low volatility, investors seek out riskier assets to get the same returns as in times of higher volatility. The results of the previous literature on equity market indicators are mixed. Some studies find a link between stock market trends and banking crises, while others do not (see table 1).

As for the bond market, it is difficult to find sufficiently long time series of country-specific corporate bond data. Since corporate bond risk premiums have significant correlations across European countries (Krylova 2016), however, it may be sufficient here to use an international corporate bond risk premium for all countries.[14]

[14]Babecký et al. (2014) use the U.S. BAA corporate bond spread and find it to be one of the best predictors of banking crisis within a nine- to twelve-quarter horizon.

Several studies suggest that global indicators such as global equity price growth (Behn et al. 2013), global liquidity measures, or the global credit gap (Alessi and Detken 2011) are useful in predicting local crises.

A potential indicator that banks are mispricing risk may be seen in changes in the interest rate margin banks require for loans to households or corporations. A rapid drop in margins on new bank loans could indicate that banks are mispricing risk, e.g., due to increased competition. Risk-management tools of banks such as the value-at-risk metric may also tolerate higher risk-taking in periods of low volatility.

For measuring potential mispricing of risk, we consider the following indicators: local stock market index and local bank stock index, stock market volatility, dividend yield, P/E ratio, price-to-book (P/B) ratio, VIX index, high-yield corporate bond risk premiums, long- and short-term interest rates of two major economies (the United States and Germany), lending margin of household loans, and lending margin of corporate loans.

2.2.6 Strength of Bank Balance Sheets

Although it is quite clear that the causes of a banking crisis are at least partly manifested in vulnerabilities in bank balance sheets, the identification of reliable warning indicators contained in bank balance sheets is rare (see table 1). This likely relates to data issues. Bank aggregate balance sheets tend to be short and published on a yearly basis. They may also contain structural breaks due to changes in the banking industry and accounting standards.

Detken et al. (2014) consider and reject the leverage ratio as a predictor for systemic banking crises, as it lacks predictive power. Behn et al. (2013) find that higher aggregate banking-sector capitalization decreases the probability of banking crisis, while higher banking-sector profits may lead to excessive risk-taking and tend to precede banking crises.

There is some empirical evidence that the indicators of a bank's funding structure might work as predictors. Bank funding can be divided into core liabilities (stable deposits) and non-core liabilities (e.g., unstable short-term wholesale funding). During periods of rapid lending growth, banks may finance their increased lending with market funding. While deposit guarantee schemes have generally made traditional bank deposit runs extremely rare, market-based funding can face a bank run if the bank's prospects deteriorate. Hence, a higher share of more unstable market funding makes banks more vulnerable. Kamin and DeMarco (2012) and Lainà, Nyholm, and Sarlin (2015) note evidence that a larger share of deposit funding has a stabilizing effect for the financial system. Betz et al. (2013) and Hahm, Shin, and Shin (2013), similarly, show that a high share of non-core liabilities is a good predictor of an impending banking crisis.

For measuring the strength of the bank balance sheet, we consider the following indicators: ratio of total assets to GDP, leverage ratio, loans-to-deposits ratio, ratio of non-core liabilities to total assets or GDP, (short-term liabilities – liquid assets)/total assets, and short-term liabilities/liquid assets.

3. Empirical Analysis

3.1 Indicator Data and Transformations

We compile quarterly indicator data from central banks, international organizations, and commercial data sources. Table 2 provides the full list of the examined indicators together with definitions, and data sources.

The unbalanced panel data cover twenty-eight EU member states for the period 1970 to 2012. The length and availability of economic time series still varies across EU countries (e.g., available data are scarce for new EU member states). Table 3 shows the descriptive statistics, where the number of countries, number of observations, and number of financial crises is highlighted for each indicator.

We consider various transformations of indicators such as differences, growth rates, and trend gaps for each indicator. This is because the indicator as such may be non-stationary—an undesirable feature for a good indicator. Indeed, Kauko, Vauhkonen, and Topi (2014) argue that if an indicator lacks an equilibrium level to which it tends to return, interpretation of the indicator becomes a non-trivial task. In any case, the application of transformations solves potential non-stationarity problems.

Table 2. List of Indicators and Data Sources

Indicator	Definition	Transformations	Data Source
1. Credit Developments			
1.1. Real Total Credit	Total credit to private non-financial sectors by all sectors divided by CPI.	Growth rates, trend gaps	BIS (credit), IMF (CPI)
1.2. Real Total Bank Credit	Credit to private non-financial sectors by domestic banks divided by CPI.	Growth rates, trend gaps	BIS (credit), IMF (CPI)
1.3. Real Household Credit	Total credit to households and non-profit institutions serving households by all sectors divided by CPI.	Growth rates, trend gaps	BIS (credit), IMF (CPI)
1.4. Real Corporate Credit	Total credit to non-financial corporations by all sectors divided by CPI.	Growth rates, trend gaps	BIS (credit), IMF (CPI)
1.5. Total Credit / GDP	Total credit to private non-financial sectors by all sectors divided by GDP.	Growth rates, differences, trend gaps	BIS
1.6. Total Bank Credit / GDP	Credit to private non-financial sectors by domestic banks divided by GDP.	Growth rates, differences, trend gaps	BIS
1.7. Total Household Credit / GDP	Total credit to households and non-profit institutions serving households by all sectors divided by GDP.	Growth rates, differences, trend gaps	BIS
1.8. Total Corporate Credit / GDP	Total credit to non-financial corporations by all sectors divided by GDP.	Growth rates, differences, trend gaps	BIS
2. Private-Sector Debt Burden			
2.1. Debt-Service Ratio	Ratio of interest payments plus amortizations divided by income; includes households and non-financial corporations. See ESRB (2015).	Growth rates, differences, trend gaps	ESRB
2.2. Corporate Debt-Service Ratio	Ratio of interest payments plus amortizations divided by income; includes non-financial corporations.	Growth rates, differences, trend gaps	ESRB

(*continued*)

Table 2. (Continued)

Indicator	Definition	Transformations	Data Source
2.3. Household Debt-Service Ratio	Ratio of interest payments plus amortizations divided by income; includes households and non-profit institutions serving households.	Growth rates, differences, trend gaps	ESRB
2.4. Total HH Credit × 10y Rate / GDP	Total HH credit / GDP indicator multiplied by the country-specific ten-year government bond yield.	Growth rates, differences, trend gaps	Bloomberg (rate), BIS
2.5. Total HH Credit × 3m Rate / GDP	Total HH credit / GDP indicator multiplied by the country-specific three-month money market rate.	Growth rates, differences, trend gaps	Bloomberg (rate), BIS
3. Potential Overvaluation of Property Prices			
3.1. Real House Price	Deflated using the private consumption deflator from the national account statistics.	Growth rates, trend gaps	OECD
3.2. House Price / Rent	Nominal house index divided by rent price index.	Growth rates, differences, trend gaps	OECD
3.3. House Price / Income	Nominal house price divided by nominal disposable income per head.	Growth rates, differences, trend gaps	OECD
3.4. Real Commercial Real Estate Price	Commercial real estate appraisal index divided by CPI.	Growth rates, trend gaps	ECB
4. External Imbalances			
4.1. CurrentAccount / GDP	Current account balance divided by GDP.	Growth rates, differences	ECB BOP
4.2. Capital Account / GDP	Capital account balance divided by GDP.	Growth rates, differences	ECB BOP
4.3. Portfolio Investments / GDP	Portfolio investments part of the financial account divided by GDP. Unadjusted amount at the end of period.	Growth rates, differences	ECB BOP
4.4. Other Investments / GDP	Other investments part of the financial account divided by GDP. Unadjusted amount at the end of period.	Growth rates, differences	ECB BOP

(*continued*)

Table 2. (Continued)

Indicator	Definition	Transformations	Data Source
4.5. Foreign Currency Cross-Border Loans / GDP	All foreign currency cross-border loans extended to foreign countries divided by GDP.	Growth rates, differences, trend gaps	ECB BSI
4.6. Own Currency Cross-Border Loans / GDP	All own currency cross-border loans extended to foreign countries divided by GDP.	Growth rates, differences, trend gaps	ECB BSI
5. Potential Mispricing of Risk			
5.1. Stock Market Volatility	Average quarterly volatility of the main national stock market index.	Growth rates, differences	Bloomberg
5.2. Stock Market Index	Level of the main national stock market index.	Growth rates	Bloomberg
5.3. Bank Stock Index	Level of the index formed by the domestic listed banks.	Growth rates	Bloomberg
5.4. Stock Market P/E Ratio	Price-to-earnings ratio of the main national stock market index.	Growth rates, differences	Bloomberg
5.5. Stock Market P/B Ratio	Price-to-book value ratio of the main national stock market index.	Growth rates, differences	Bloomberg
5.6. Stock Market Dividend Yield	Dividend yield of the main national stock market index.	Growth rates, differences	Bloomberg
5.7. Household Lending Spread	The average rate at which banks issue new loans to households and non-profit institutions serving households. Unconsolidated.	Growth rates, differences	ECB MIR
5.8. Corporate Lending Spread	The average rate at which banks issue new loans to non-financial corporations.	Growth rates, differences	ECB MIR
5.9. High-Yield Spread	Difference between the Bank of America Merrill Lynch euro non-financial high-yield bond index (HNE0) and euro non-financial investment-grade bond index (EN00).	Growth rates, differences, trend gaps	Bloomberg
5.10. VIX Index	Measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices.	Growth rates, differences, trend gaps	Chicago Board Options Exchange

(continued)

Table 2. (Continued)

Indicator	Definition	Transformations	Data Source
5.11. German 10y Bund	Yield of German ten-year bund.	Growth rates, differences, trend gaps.	Bloomberg
5.12. German 1y Bill	Yield of German one-year bill.	Growth rates, differences, trend gaps	Bloomberg
5.13. German 1m Bill	Yield of German one-month bill.	Growth rates, differences, trend gaps	Bloomberg
5.14. U.S. 10y T-Note	Yield of U.S. ten-year Treasury note.	Groth rates, differences, trend gaps	Bloomberg
5.15. U.S. 1y T-Bill	Yield of U.S. one-year Treasury bill.	Growth rates, differences, trend gaps	Bloomberg
5.16. U.S. 1m T-Bill	Yield of U.S. one-month Treasury bill.	Growth rates, differences, trend gaps	Bloomberg
6. Strength of Bank Balance Sheets			
6.1. Leverage Ratio	Total equity divided by total assets.	Growth rates, differences	ECB CBD2
6.2. Loans / Deposits	Bank loans to private non-financial sector divided by bank deposits from the private non-financial sector.	Growth rates, differences	ECB CBD2
6.3. Total Assets / GDP	Total assets divided by GDP.	Growth rates, differences	ECB CBD2
6.4. Non-core Liabilities / GDP	Non-core liabilities = Total assets – Deposits – Capital and reserves.	Growth rates, differences	ECB BSI
6.5. Non-core Liabilities / Total Assets	See above.	Growth rates, differences	ECB BSI
6.6. Net ST Liabilities Ratio = (ST Liabilities – Liquid Assets) / Total Assets	Short-term liabilities include debt securities issued with maturity less than one year, short-term deposits (euro-area private sector, non-euro-area and euro-area other general government), inter-MFI deposits. Liquid assets include holdings of cash, MMF shares, euro-area private-sector debt securities with maturity below one year, inter-MFI loans, and government debt securities.	Growth rates, differences	ECB BSI
6.7. ST Liabilities / Liquid Assets	Ratio of short-term liabilities and liquid assets. The components are defined as above.	Growth rates, differences	ECB BSI

Notes: ECB data are for all resident monetary financial institutions (MFIs), excluding money market funds (MMF). ECB balance sheet items (BSI), and MFI interest rates (MIR) statistics are reported on an unconsolidated basis. ECB Consolidated Banking Statistics (CBD2) is consolidated. BOP = balance of payments. HH = household.

Table 3. Descriptive Statistics

Indicator	$\bar{X}$	Sd(x)	Min.	p_{25}	p_{50}	p_{75}	Max.	N	N_c	N_f
1. Credit Developments										
1.1. Real Total Credit	10.48	12.56	0.27	1.90	5.25	15.83	80.46	1735	15	15
1.2. Real Total Bank Credit	6.64	7.61	0.17	1.11	2.55	10.61	42.83	1716	15	15
1.3. Real Household Credit	4.69	5.04	0.05	1.04	2.16	6.99	26.35	1434	15	15
1.4. Real Corporate Credit	7.67	8.47	0.36	1.80	5.12	11.31	54.72	1434	15	15
1.5. Total Credit / GDP	0.93	0.82	0.005	0.33	0.71	1.27	5.20	2746	18	22
1.6. Total Bank Credit / GDP	0.55	0.43	0.004	0.22	0.45	0.80	2.20	2715	18	22
1.7. Total Household Credit / GDP	0.41	0.32	0.01	0.16	0.35	0.59	1.60	2022	18	20
1.8. Total Corporate Credit / GDP	0.73	0.61	0.04	0.35	0.59	0.90	4.47	1998	18	19
2. Private-Sector Debt Burden										
2.1. Debt-Service Ratio	0.19	0.16	0.01	0.12	0.15	0.19	1.08	2899	27	27
2.2. Corporate Debt-Service Ratio	0.37	0.21	0.10	0.25	0.32	0.44	1.77	1713	26	19
2.3. Household Debt-Service Ratio	0.12	0.06	0.02	0.08	0.11	0.14	0.36	1701	26	19
2.4. Total HH Credit × 10y Interest Rate / GDP	2.93	1.67	0.38	1.80	2.56	3.67	12.83	1451	20	17
2.5. Total HH Credit × 3m Interest Rate / GDP	2.19	1.73	0.06	0.91	1.75	2.96	11.90	1923	25	21
3. Potential Overvaluation of Property Prices										
3.1. Real House Price	81.88	29.07	23.18	58.98	79.28	100.8	178.6	2241	21	22
3.2. House Price / Rent	82.98	27.81	23.88	61.35	81.71	101.1	178.6	2071	20	21
3.3. House Price / Income	86.63	25.49	32.75	66.93	87.89	100.9	189.4	2070	21	21
3.4. Real Commercial Real Estate Price	97.03	37.18	37.54	73.60	94.97	108.4	255.7	1209	15	14
4. External Imbalances										
4.1. Current Account / GDP	–0.33	1.74	–13.82	–1.07	–0.21	0.66	9.53	2472	26	16
4.2. Capital Account / GDP	0.001	0.003	–0.02	0	0	0.001	0.04	1491	21	16
4.3. Portfolio Investments / GDP	–0.46	2.00	–19.89	–0.30	–0.10	–0.01	0.90	1024	21	14
4.4. Other Investments / GDP	–0.46	2.00	–19.89	–0.30	–0.10	–0.01	0.90	1024	21	14
4.5. F.C. Cross-Border Loans / GDP	0.14	0.45	0	0.01	0.01	0.05	2.67	1303	15	12
4.6. D.C. Cross-Border Loans / GDP	0.10	0.34	0	0.004	0.01	0.04	2.13	1303	15	12

(continued)

Table 3. (Continued)

Indicator	$\bar{X}$	Sd(x)	Min.	p_{25}	p_{50}	p_{75}	Max.	N	N_c	N_f
5. Potential Mispricing of Risk										
5.1. Stock Market Volatility	0.18	0.11	0	0.11	0.15	0.22	1.19	3180	28	27
5.2. Stock Market Index	3869	6149	47.67	714.4	2017	4477	47803	1584	14	14
5.3. Bank Stock Index	1050	1544	11.66	185.7	434.7	1134	9288	1632	13	17
5.4. Stock Market P/E Ratio	41.81	337.01	0.32	11.66	15.12	21.19	9377	1350	23	15
5.5. Stock Market P/B Ratio	1.65	0.62	0.32	1.18	1.54	2.07	3.25	494	14	9
5.6. Stock Market Dividend Yield	3.33	1.33	0.89	2.44	3.12	3.99	10.95	720	14	9
5.7. Household Lending Spread	1.95	1.17	–5.36	1.20	1.85	2.61	8.32	1112	28	14
5.8. Corporate Lending Spread	1.92	0.84	0.03	1.26	1.76	2.47	5.51	1070	27	13
5.9. High-Yield Spread*	576.6	347.8	164.2	335.4	486.3	715.9	1744	1596	28	30
5.10. VIX Index*	21.39	6.97	12.50	14.73	20.91	25.92	40.60	2632	28	17
5.11. German 10y Bund*	6.06	2.33	0.77	4.24	6.17	7.93	11.10	5040	28	33
5.12. German 1y Bill*	4.37	2.43	0.33	2.36	4.09	5.65	9.90	3696	28	31
5.13. German 1m Bill*	4.44	2.84	0.01	2.62	3.95	5.68	12.38	4144	28	31
5.14. U.S. 10y T-Note*	6.95	2.81	1.61	4.85	6.75	8.32	15.32	4816	28	31
5.15. U.S. 1y T-Bill*	5.60	3.19	0.93	3.18	5.57	7.50	15.51	3472	28	31
5.16. U.S. 1m T-Bill*	6.01	4.07	0.30	3.21	5.62	8.26	20.58	3920	28	31
6. Strength of Bank Balance Sheets										
6.1. Leverage Ratio	8.80	3.39	2.54	6.21	8.38	10.53	21.33	1348	28	14
6.2. Loans / Deposits	134.0	52.3	47.1	100.9	123.9	151.6	327.1	1101	28	13
6.3. Total Assets / GDP	3.76	6.67	0.001	0.90	2.26	3.42	39.75	1274	21	14
6.4. Non-core Liabilities / GDP	2.54	5.61	0.001	0.30	1.13	1.94	33.92	844	20	10
6.5. Non-core Liabilities / Total Assets	0.48	0.15	0.17	0.35	0.50	0.58	0.82	1031	27	12
6.6. (ST Liabilities – Liquid Assets) / Total Assets	0.19	0.14	–0.03	0.06	0.20	0.31	0.48	554	13	7
6.7. Short-Term Liabilities / Liquid Assets	1.72	0.57	0.92	1.17	1.62	2.15	3.62	554	13	7

Notes: The sample statistics are calculated for the full sample, 1970–2012. $\bar{X}$ and Sd(x) are the sample mean and sample standard deviation. p_{25}, p_{50}, and p_{75} denote the first, second, and third quartiles, respectively. N, N_c, and N_f are the number of observations, countries, and financial crises, respectively. F.C. and D.C. refer to foreign currency and domestic currency, respectively. Indicators marked by * are understood as global indicators, so the data are repeated for each country.

The simplest transformations are the growth and difference. *n-year growth* is calculated as

$$100\frac{x_t - x_{t-4n}}{x_{t-4n}}. \tag{3}$$

n-year difference is calculated as

$$x_t - x_{t-4n}. \tag{4}$$

We simply apply the rates of growth and differences $n = 1$ (year) and $n = 3$ (years) that correspond to typical choices used by practitioners when monitoring macroeconomic and financial developments. Why do we consider both differences and rates of growth? Note that in the panel setup, the level values of some indicators (such as house price index or real credit stock) may not lead to an economically sensible model. In such cases, it is more appropriate to use relative measures such as rates of growth and relative trend gap (defined below).

Additionally, we consider four alternative trend gaps. Two alternative trend gaps utilize the trend calculated with a one-sided HP filter with smoothing parameter $\lambda = 400{,}000$. "One-sided" here means that the trend at time t is calculated using only values up to time t. Once the trend component is estimated, we form the trend gap (denoted *trend gap* in the tables) as

$$x_t - trend_t, \tag{5}$$

and the relative trend gap (denoted *relative gap*) as

$$100\left(\frac{x_t}{trend_t} - 1\right), \tag{6}$$

respectively. Because the one-sided trend makes little sense for the first few observations of the time series, the trend gaps are calculated only after the time series has five years of historical data. Hence, the trend-gap-transformed indicators have somewhat lower number of observations than the original series. Finally, we consider two more alternative definitions of the trend. First, a trend that is the historical average of the original indicator x_t is calculated as

$$average_t = \sum_{s=t_0}^{t} \frac{x_s}{t - t_0 + 1}. \tag{7}$$

Second, a trend that is the five-year moving average of the original indicator x_t is calculated as

$$5y\ ma_t = \sum_{s=0}^{19} \frac{x_{t-s}}{20}. \tag{8}$$

The corresponding trend gaps are denoted *ave. gap,* calculated as

$$x_t - average_t, \tag{9}$$

and *5y M.A. gap,* calculated as

$$\text{and } x_t - 5y\ ma_t. \tag{10}$$

As with the one-sided HP-filtered trends, these trend gaps are only calculated after five years of historical data are available.

3.2 Banking Crisis Variable

Our main results are reported for the systemic financial crisis variable published by Detken et al. (2014). At the time of writing, this was the most recent available financial crisis database. As our work extends that of Detken et al. (2014), their crisis data set (henceforth labeled Detken's) is a natural starting point. However, a variety of banking crisis data sets are provided in the earlier literature, with Babecký et al. (2014) and Laeven and Valencia (2012) among the newest (henceforth labeled Babecký's and Laeven's crisis data sets).[15]

The data sets use different definitions as to what constitutes a banking crisis. Therefore, table 4 lays out these alternative crisis definitions. Detken's data set, which is based on Babecký's data set, includes numerous modifications to align crisis episodes with policymakers' objectives. Crises that were not systemic banking crises or not associated with a domestic credit cycle are excluded, while

[15]Table 1 shows the crisis data sets used in some earlier studies. In addition to Babecký's, Detken's, and Laeven's crisis data sets, crisis dating of, e.g., Caprio and Klingebiel (1996), Demirgüç-Kunt and Detragiache (1998), Lindgren, Garcia, and Saal (1996), and Reinhart and Rogoff (2009) have been used.

Table 4. Information about Alternative Banking Crisis Data Sets: Banking Crisis Definitions

Label	Source	Banking Crisis Definition
Babecký's	Babecký et al. (2014)	They collect information about crisis occurrence from ten influential papers. They validate the coding of crises with the help of a comprehensive survey among country experts.
Detken's	Detken et al. (2014)	They amend Babecky's data set with the following changes: Non-systemic banking crises and crises not associated with the credit cycle are excluded. "Would-be crises" (i.e., periods where domestic developments related to the credit cycle could have caused a systemic banking crisis had it not been for policy action or an external event that dampened the financial cycle) are added.
Laeven's	Laeven and Valencia (2012)	A banking crisis is defined as systemic if two conditions are met: (i) significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations) and (ii) significant banking policy intervention measures in response to significant losses in the banking system.

periods where domestic developments related to the credit cycle that likely would have led to a systemic banking crisis in the absence of policy intervention or an external event that dampened the credit cycle are added.

As the differences can be quite significant, we consider the three separate crisis data sets, and provide a summary of our key findings with each alternative crisis data set.[16] The crisis periods for the three financial crisis data sets considered in this work are summarized in table 5.

3.3 Extracting Early Warning Signals

We follow a common approach to extracting warning signals from early warning indicators—the *signaling* approach. Basically, it is a non-parametric model suitable for single-variable warning indicators (Alessi and Detken 2011; Borio and Drehmann 2009; Borio and Lowe 2002; Drehmann, Borio, and Tsatsaronis 2011; Drehmann et al. 2010). The earlier literature considers other approaches to extract early warning signals, such as the discrete choice model (Barrell et al. 2010; Behn at al. 2013; Davis and Karim 2008; Demirgüç-Kunt and Detragiache 2000; Frankel and Rose 1996; Hardy and Pazarbaşıoğlu 1998; Lo Duca and Peltonen 2013; and Lund-Jensen 2012), decision trees, and machine-learning techniques (Holopainen and Sarlin 2015). For our work, the primary advantages of the signaling method are its transparency and ease of interpretation relative to the other, more complex techniques. It helps us keep the focus on identifying informative indicators rather than useful methods.

The idea of the signaling approach is simple. Below or above some *signaling threshold*, a warning signal of increased vulnerability is issued when that threshold is crossed. For example, a warning signal might be issued if one-year growth in real household credit exceeds 6 percent.

The rationale for specifying the thresholds is closely related to the performance evaluation of the warning indicators. If the threshold is too insensitive, so that it rarely gives alarms, the number of false alarms is likely to be low, but the indicator may also fail to warn on the cusp of most crises. Conversely, if the threshold is overly sensitive, false alarms are frequent, but few crises are missed.

3.4 Evaluating Early Warning Indicators

We use two early warning indicator evaluation statistics—area under the receiver operating characteristic (ROC) curve (this area is

[16] The full set of results calculated with alternative data sets is available from the authors on request.

Table 5. Information about Alternative Banking Crisis Data Sets: Crisis Periods in EU Countries for Different Data Sets

Country	Babecký's	Detken's	Laeven's
Austria	2008:Q1–2008:Q4		2008:M9–2010
Belgium	2008:Q1–2008:Q4		2008:M9–2010
Bulgaria	1971:Q1–1971:Q2		
	1994:Q1–1997:Q4	1995:Q2–1997:Q4	1996:M1–1997
		2004:Q4–2007:Q2*	
Croatia		1998:Q1–2000:Q2	1998:M3–1999
Cyprus		2012:Q2–2012:Q4	
Czech Republic	1991:Q1–1991:Q4		
	1994:Q1–2000:Q4	1998:Q1–2002:Q2	1996:M6–2000
Denmark	1987:Q1–1993:Q4	1987:Q1–1993:Q4	
	2008:Q1–2010:Q4	2008:Q3–2012:Q4	2008:M9–2010
Estonia	1992:Q1–1995:Q4		1992:M11–1994
	1998:Q1–1998:Q4	1998:Q2–1998:Q4	
Finland	1991:Q1–1995:Q4	1991:Q3–1995:Q4	1991:M9–1995
France	1994:Q1–1995:Q4	1993:Q3–1995:Q4	
	2008:Q1–2009:Q4	2008:Q3–2012:Q4	2008:M9–2010
Germany	1974:Q2–1974:Q4		
	1977:Q1–1979:Q4		
		2000:Q1–2003:Q4	
	2008:Q1–2008:Q4		2008:M9–2010
Greece	1991:Q1–1995:Q4		
	2008:Q1–2010:Q4	2008:Q1–2012:Q4	2008:M8–2010
Hungary	1991:Q1–1995:Q4		1991:M9–1995
	2008:Q1–2009:Q2	2008:Q3–2012:Q4	2008:M9–2010
Ireland	1985:Q1–1985:Q1		
	2007:Q1–2010:Q4	2008:Q3–2012:Q4	2008:M9–2010
Italy	1990:Q1–1995:Q4	1994:Q1–1995:Q4	
			2008:M9–2010
Latvia	1995:Q1–2003:Q4		1995:M4–1996
	2008:Q1–2008:Q4	2008:Q4–2010:Q3	2008:M9–2010
Lithuania	1995:Q1–1996:Q4	1995:Q1–1996:Q4	1995:M12–1996
	2009:Q1–2009:Q4	2008:Q4–2010:Q4	
Luxembourg	2008:Q1–2010:Q4		2008:M9–2010
Netherlands		2002:Q1–2003:Q4*	
	2008:Q1–2008:Q4	2008:Q3–2012:Q4	2008:M9–2010
Poland	1991:Q1–1994:Q4		1992–94
Portugal		1999:Q1–2000:Q1*	
	2008:Q1–2008:Q4	2008:Q4–2012:Q4	2008:M9–2010
Romania	1990:Q1–1999:Q4		1990–92
		1997:Q2–1999:Q1	

(*continued*)

Table 5. (Continued)

Country	Babecký's	Detken's	Laeven's
Slovak Republic	1991:Q1–2002:Q4		1998–2002
Slovenia	1992:Q1–1994:Q4	1992:Q1–1994:Q4	1992–92
	2008:Q1–2008:Q4	2008:Q1–2012:Q4	2008:M9–2010
Spain	1977:Q1–1985:Q4	1978:Q1–1985:Q3	1977–81
	2008:Q1–2008:Q4	2009:Q2–2012:Q4	2008:M9–2010
Sweden	1990:Q3–1995:Q4	1990:Q3–1993:Q4	1991:M9–1995
	2008:Q1–2008:Q4	2008:Q3–2010:Q4	2008:M9–2010
United Kingdom	1974:Q1–1976:Q4	1973:Q4–1975:Q4	
	1984:Q1–1984:Q4		
	1991:Q1–1995:Q4	1990:Q3–1994:Q2	
	2007:Q1–2007:Q4	2007:Q3–2012:Q4	2007:M9–2010

Note: For Detken's data set, the three events marked by * are not actual realized crises but domestic developments related to the credit cycle that could well have caused a systemic banking crisis had it not been for policy action or an external event that dampened the credit cycle.

henceforth denoted AUC) and relative usefulness (U_r). Both evaluation statistics have been quite popular in recent banking crisis early warning literature.[17] We provide only a short introduction to the methods, as detailed expositions of the measures are available elsewhere.[18]

Both AUC and relative usefulness consider the relative amounts of type I and type II errors produced by the early warning indicator (see figure 1A). The measures can be applied more generally to any situation where there is a trade-off between type I and type II errors. In our case, a type I error (false positive) corresponds to a false alarm, i.e., the indicator issues an early warning signal, but no crisis follows within the specified prediction horizon. A type II

[17] AUC is used in, e.g., Bonfim and Monteiro (2013), Buchholst and Rangvid (2013), Comelli (2014), and Drehmann and Juselius (2014). Both statistics are applied in Behn et al. (2013), Betz et al. (2013), and Detken et al. (2014). Relative usefulness is found in Alessi and Detken (2011), Babecký et al. (2014), Lainà, Nyholm, and Sarlin (2015), and Lo Duca and Peltonen (2013).

[18] See, e.g., Drehmann and Juselius (2014) and Sarlin (2013) for AUC and usefulness, respectively.

Figure 1. Correspondence of the Generic Confusion Matrix with the Early Warning Exercise

A. Generic Confusion Matrix

		True condition	
		Condition positive	Condition negative
Predicted condition	Predicted condition positive	True positive	False positive (Type I error)
	Predicted condition negative	False negative (Type II error)	True negative

B. Confusion Matrix for the Early Warning Exercise

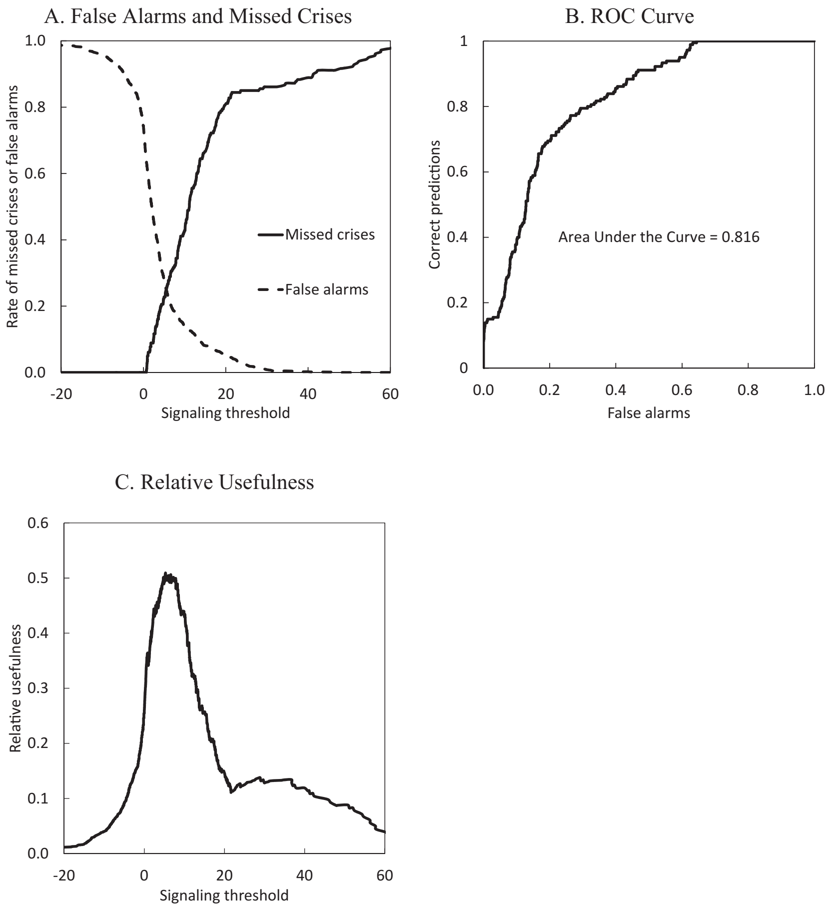
		True condition	
		Crisis	No crisis
Predicted condition	Signal	Correct alarm (A)	False alarm (B)
	No signal	Missed crisis (C)	Correctly no alarm (D)

error (false negative) corresponds to a missed crisis, i.e., the indicator does not give a signal, but a banking crisis occurs within the specified prediction horizon.

Figure 2A illustrates the tradeoff between false alarms and missed crises for the total credit-to-GDP trend gap indicator. If the signaling threshold is below the trend gap of 0 percent, there will be no missed crises, but the number of false alarms will be quite high. In contrast, if the threshold is above the trend gap of 20 percent, the share of missed crises is quite high, but the false alarm rate is very low. Thus, the policy-relevant threshold likely lies somewhere between 0 and 20 percent (the actual BCBS benchmark buffer guide applied in the EU legislation has triggers in the range of 2 to 10 percent of the trend gap).

In specifying the horizon for crisis prediction, we follow the conventions in Detken et al. (2014) and set the crisis-prediction horizon to three years. If the time to crisis is less than a year, the policymaker lacks sufficient lead time to react. Hence, we do not include in the evaluation observations that take place when the distance to the banking crisis is less than one year. A publication lag of one

Figure 2. False Alarms and Missed Crises for Different Signaling Thresholds, ROC Curve, and Relative Usefulness for Different Signaling Thresholds



quarter is applied to all indicators. As a robustness check and to gain further insight on the lead-lag structure of different indicators, i.e., when different indicators are expected to give signals, we also consider prediction horizons from six months to five years.

The ROC curve is the visual curve that shows the tradeoff between type I and type II errors. This is illustrated for the credit-to-GDP gap indicator in figure 2B. For a given rate of type I errors on the horizontal axis (false alarms), it would be desirable for the rate of correct alarms (vertical axis) to be as close to 1 as possible. Broadly speaking, the larger the area under the ROC curve (AUC), the better the indicator. For a completely uninformative indicator, AUC = 0.5, while for a perfect indicator AUC = 1. Thus, to be an informative indicator, we need AUC > 0.5. In our credit-to-GDP gap example, AUC = 0.82 would make it a very good indicator in this context.

The *relative usefulness* statistic uses a loss function that accounts for type I and type II errors. The weights of the loss function reflect the presumed preferences for the errors. The methodology goes back to the policy loss functions of Bussière and Fratzscher (2008) and Demirgüc-Kunt and Detragiache (2000), and the usefulness measure proposed by Alessi and Detken (2011) and later supplemented by Sarlin (2013).

The loss function of Alessi and Detken (2011) is defined as follows:

$$L_{AD}\left(\theta\right) = \theta T_2 + \left(1-\theta\right) T_1 = \theta \frac{C}{A+C} + \left(1-\theta\right) \frac{B}{B+D}, \tag{11}$$

where the right-hand side is a weighted average of type I and type II error rates, T_1 and T_2, respectively.[19] The correspondence of the right-hand-side alphabetic letters with the generic confusion matrix is illustrated in figure 1B. A is the number of periods in which an indicator provides a correct signal (crisis starts within one to three years of issuing the signal), and B is the number of periods in which a wrong signal is issued. C is the number of periods in which a signal is not provided although the crisis is starting within a reasonable number of periods (one to three years). At last, D denotes the number of periods in which correctly no signal is provided. In other words, $A = TP$, number of true positives; $B = FP$, number of false positives; $C = FN$, number of false negatives; and $D = TN$, number

[19]In the formula, the order of T_1 and T_2 differs from some of the earlier literature. It is just a matter of convention for forming the null hypothesis. Type I error (or false positive) is the incorrect rejection of a true null hypothesis H_0. We set the H_0: "no crisis within the next three years" so that a false positive means a false alarm. Type II error (false negative) is incorrectly retaining a false null hypothesis so that in our case false negative means failure to detect a crisis.

of true negatives. θ is the parameter revealing the policymaker's relative risk aversion to type I and type II errors. A higher parameter value θ means that the policymaker is more averse to missing a crisis than getting a false alarm.

Sarlin (2013) augments the loss function with the unconditional crisis probability such that

$$L_S(\mu) = \mu P T_2 + (1-\mu)(1-P) T_1, \tag{12}$$

where $P = \frac{A+C}{A+B+C+D}$ is the unconditional crisis probability as estimated from the sample. The advantage of the augmented loss function is that it is explicit with respect to the relative frequency of situations when type I or type II errors can occur. Yet, for each μ there exists θ such that the two alternative loss functions lead to equivalent policies.

For either loss function, the relative usefulness statistics is defined as

$$U_r = \frac{\omega - L}{\omega}, \tag{13}$$

where for Alessi and Detken (2011) $\omega = \min(\theta, 1-\theta)$ and for Sarlin (2013) $\omega = \min(\mu P, 1-\mu P)$. The normalization parameter α ensures that the maximum value of relative usefulness is 1, i.e. a perfect warning indicator. In theory, any indicator is useful to a policymaker if its usefulness is larger than 0 (the higher the better), and useless if usefulness is less than 0 (all useless indicators are equally useless). In practice, indicators with low positive usefulness would likely be ignored by a policymaker with access to more useful indicators.

We set $\theta = 0.5$ as the point at which the policymaker is indifferent to type I and type II errors. In our data the probability of crisis is $P = 0.1$, so our choice of θ is equivalent to our choice of $\mu = 0.9$. Whether these are the correct values for μ or θ is up to the policymaker's actual preferences. In any case, both parameter values are close to those previously used in the literature.[20]

[20] Babecký et al. (2014) and Lo Duca and Peltonen (2013) both use $\theta = 0.5$. For example, Detken et al. (2014) use $\theta = 0.5/0.6/0.7$. Behn et al. (2013) use $\mu = 0.9$ and Betz et al. also use $\mu = 0.9$ as the benchmark case.

Figure 2C illustrates the relative usefulness for credit-to-GDP gap for different signaling thresholds. As expected from previous discussion, the credit-to-GDP gap indicator is most useful for signaling thresholds between 0 percent and 20 percent. Following the curve from left to right, usefulness initially increases as the rate of false alarms goes down rapidly, while the rate of missed crises increases at a relatively slow pace (see figure 2A). At the peak of the usefulness curve, the rate of change of false alarms is exactly opposite to the rate of change of missed crises. From this point onwards, usefulness starts to decrease as the improvement in false alarms no longer offsets the increase in the missed crisis rate.

Note that it is possible in principle that both high and low indicator values might signal increased vulnerability. This turned out not to be much of an issue for the indicators considered in this paper, however.[21] Hence, we report the evaluation results for each indicator using the single directionality, which is based on economic reasoning and earlier literature discussed in section 2.2. The hypothesized directionality is indicated for each indicator together with the evaluation results. If the observed data goes against the hypothesized direction, the resulting usefulness values are expected to be low or negative and the AUC statistic below 0.5. Additionally, the tables in an online appendix (available at http://www.ijcb.org) report the statistical significance for the logit model coefficient β_1 for the model

$$\Pr(precrisis = 1) = F(\beta_0 + \beta_1 indicator), \tag{14}$$

where $F(z) = e^z/(1+e^z)$ is the cumulative logistic distribution, and the binary dependent variable is 1 for the pre-crisis quarters (one to three years before onset of crisis) and for the normal quarters (more than three years before crisis).[22] As is evident in the tables in the online appendix, with rare exceptions the logit coefficient either has the hypothesized sign or is not statistically significant.

[21] We initially evaluated the usefulness of each indicator in both directions—the direction hypothesized based on economic reasoning, and the opposite direction. Generally the opposite direction was not useful, but in a handful of cases the opposite direction was more useful than the hypothesized direction. These cases were the weakly performing indicators 4.1.4., 4.2.4., 4.3.4., 5.4.2., and 5.4.4.

[22] Note that late pre-crisis quarters (less than one year before crisis) and crisis quarters are excluded from all the evaluations because we are looking for *early* warning signals.

3.5 Full-Sample and Out-of-Sample Evaluation

We perform two types of performance evaluations for the indicators—full-sample and out-of-sample evaluation—following an approach which is by now common in the literature (see, e.g., Holopainen and Sarlin 2015). In our *full-sample* evaluation, the data extend from 1970 to 2012. While we have the data after 2012, it is excluded because our crisis-prediction horizon extends to three years, so we cannot say if signals after 2012 are correct or wrong. In the full-sample evaluation, the policymaker's choices are based on the same signaling threshold throughout all time periods. The full-sample usefulness statistics are then based on this threshold. The AUC measure is only reported for the full sample, as the methodology does not naturally accommodate a changing threshold.

For the relative usefulness metric, we also perform an *out-of-sample* evaluation. The out-of-sample evaluation is a recursive simulation for the period 2000 to 2012. In 2000:Q1, the policymaker uses information about a crisis data set for the periods 1970:Q1–1999:Q4 and about the previous indicator values. Because the policymaker does not yet know whether 1997:Q1–1999:Q4 are tranquil or pre-crisis periods, only the data within the period 1970:Q1–1996:Q4 are usable. The policymaker determines what is the optimal signaling threshold based on this history (and the policy parameter $\theta = 0.5$). This, combined with the indicator value for 2000:Q1, determines whether or not there is a warning signal in 2000:Q1. The signal is compared with the ex post information about 2000:Q1, and we record a true positive, false positive, true negative, or false negative. The same procedure is repeated for the next quarter 2000:Q2 (i.e., the signaling threshold now depends on the data for 1970:Q1–1997:Q1), and so on. This process continues until we reach 2012:Q4, our last evaluated quarter. The resulting out-of-sample relative usefulness is denoted $U_{r,o}$.

4. Results of the Empirical Analysis

4.1 The Set of Recommended Indicators

Recall that our objective is to identify a set of indicators that satisfies the criteria of high information content, simplicity, and robustness, and that we seek indicators relevant to each of the ESRB's six risk categories (credit developments, private-sector debt burden, overvaluation of property prices, external imbalances, mispricing of risk, and strength of bank balance sheets). The main result of the paper, of course, is the indicator set we present in table 6. The AUC and relative usefulness measures in table 6 are based on Detken's crisis data set,[23] with the crisis-prediction horizon set to one to three years. In subsequent subsections, we discuss how our results change as the prediction horizon or crisis definition is altered. The detailed performance numbers for the full set of indicators and transformation are available in the online appendix (tables A1–A7). Below we summarize the main findings from table 6 for each risk category (blocks 1–6).

4.1.1 Credit Developments

In line with findings of previous literature, we find that the ratio of credit to GDP (1.5.5.) tends to be more informative than credit alone (1.1.1.); see the first block in table 6. The result remains intact regardless of the definition of credit used. Alternative definitions include total private-sector credit (which includes, e.g., bank credit and market-based funding), total bank credit to private sector, total credit to households, and total credit to non-financial corporations. The *benchmark indicator* proposed by the Basel Committee, the total credit-to-GDP trend gap (1.5.5.) calculated using the broadest definition of credit, is clearly among the top-performing indicators in this category. However, various alternative transformations and credit concepts are found to be at least as informative. Using total bank credit to private sector or total credit to households in the numerator (1.6.5., 1.7.5.) generally leads to a slightly better AUC and higher full-sample and out-of-sample relative usefulness than the benchmark indicator. In contrast, calculating the trend gap using the prescribed HP filter does not seem to lead to improvement over the more practical transformations such as three-year difference or deviation from the five-year moving average (see table A1 in the online appendix for detailed results for the alternative transformations). The indicators proposed by Kauko (2012a) that relate credit to a

[23]Recall from section 3.2 our labeling of banking crisis data sets.

Table 6. Summary of the Recommended Indicators for Each Category

Indicator	Transformation	Sign	Full Sample							Out of Sample					
			AUC	U_r	FNR	FPR	N	N_c	N_f	$U_{r,o}$	FNR	FPR	N	N_c	N_f
1. Credit Developments															
1.1.1. Real Total Credit	1y Growth	+	0.69***	0.30	0.07	0.63	1262	15	14	0.24	0.01	0.75	471	15	9
1.2.1. Real Total Bank Credit	1y Growth	+	0.71***	0.37	0.14	0.49	1243	15	14	0.33	0.09	0.58	471	15	9
1.3.1. Real Household Credit	1y Growth	+	0.66***	0.27	0.23	0.50	961	15	14	0.17	0.03	0.81	450	15	9
1.4.1. Real Corporate Credit	1y Growth	+	0.69***	0.29	0.29	0.43	961	15	14	0.21	0.08	0.71	450	15	9
1.5.5. Total Credit / GDP	Trend Gap	+	0.82***	0.53	0.24	0.23	1787	18	20	0.32	0.07	0.61	558	18	11
1.5.8. Total Credit / GDP	KK_1	+	0.80***	0.53	0.31	0.16	2042	18	22	0.31	0.00	0.69	573	18	11
1.6.5. Total Bank Credit / GDP	Trend Gap	+	0.83***	0.55	0.26	0.19	1755	18	20	0.29	0.31	0.41	558	18	11
1.6.8. Total Bank Credit / GDP	KK_1	+	0.80***	0.55	0.22	0.24	2010	18	22	0.38	0.08	0.53	573	18	11
1.7.5. Total Household Credit / GDP	Trend Gap	+	0.83***	0.57	0.19	0.23	1135	17	18	0.42	0.03	0.54	516	17	11
1.7.8. Total Household Credit / GDP	KK_1	+	0.82***	0.55	0.29	0.16	1368	18	20	0.47	0.03	0.50	552	18	11
1.8.5. Total Corporate Credit / GDP	Trend Gap	+	0.66***	0.28	0.29	0.42	1115	17	18	0.11	0.33	0.56	516	17	11
1.8.8. Total Corporate Credit / GDP	KK_1	+	0.77***	0.42	0.20	0.39	1356	18	19	0.30	0.16	0.54	552	18	11
2. Private-Sector Debt Burden															
2.1.1. Debt-Service Ratio	1y Difference	+	0.78***	0.42	0.36	0.22	2161	26	26	0.26	0.09	0.65	764	26	16
2.2.1. Corporate Debt-Service Ratio	1y Difference	+	0.73***	0.39	0.28	0.33	967	25	17	0.20	0.03	0.77	599	25	13
2.3.1. Household Debt-Service Ratio	1y Difference	+	0.75***	0.37	0.29	0.34	952	25	18	0.21	0.00	0.79	599	25	13
2.4. HH Credit × 10y Rate / GDP		+	0.66***	0.33	0.35	0.32	951	18	17	0.41	0.42	0.17	539	18	11
2.4.2. HH Credit × 10y Rate / GDP	3y Difference	+	0.68***	0.34	0.28	0.38	788	17	14	0.23	0.34	0.44	481	17	11
2.5. HH Credit × 3m Rate / GDP		+	0.60**	0.20	0.27	0.52	1328	25	21	0.23	0.54	0.24	738	25	15
2.5.1. HH Credit × 3m Rate / GDP	1y Difference	+	0.71***	0.29	0.26	0.45	1240	25	21	0.14	0.06	0.81	696	25	15
3. Potential Overvaluation of Property Prices															
3.1.2. Real House Price	3y Growth	+	0.67***	0.30	0.43	0.27	1429	16	20	0.14	0.42	0.44	465	16	11
3.2.1. House Price / Rent	1y Difference	+	0.64**	0.27	0.57	0.16	1428	17	21	0.09	0.52	0.39	526	17	12
3.2.2. House Price / Rent	3y Difference	+	0.70***	0.34	0.42	0.24	1286	17	20	0.16	0.37	0.47	483	17	12
3.2.8. House Price / Rent	Avg. Gap	+	0.74***	0.45	0.38	0.16	1174	16	20	0.25	0.15	0.60	448	16	12
3.3.1. House Price / Income	1y Difference	+	0.69***	0.33	0.50	0.18	1410	20	21	0.30	0.44	0.26	563	20	12
3.3.2. House Price / Income	3y Difference	+	0.77***	0.45	0.38	0.18	1260	18	19	0.26	0.27	0.47	512	18	12
3.3.8. House Price / Income	Avg. Gap	+	0.81***	0.52	0.30	0.18	1148	17	19	0.31	0.09	0.61	470	17	12
3.4.1. Real Commercial Real Estate Price	1y Growth	+	0.73***	0.39	0.21	0.40	718	15	14	0.39	0.35	0.26	391	15	10

(*continued*)

Table 6. (Continued)

Indicator	Transformation	Sign	Full Sample AUC	U_r	FNR	FPR	N	N_c	N_f	Out of Sample $U_{r,o}$	FNR	FPR	N	N_c	N_f
4. External Imbalances															
4.1. Current Account / GDP		–	0.64*	0.30	0.45	0.25	1159	19	16	0.14	0.4	0.5	601	19	12
4.1.8. Current Account / GDP	Avg. Gap	–	0.70**	0.41	0.35	0.24	792	19	13	0.29	0.43	0.28	410	19	11
4.5.2. F.C. Cross-Border Loans / GDP	3y Difference	+	0.56	0.24	0.53	0.23	698	13	11	0.22	0.31	0.47	389	13	7
4.6.2. D.C. Cross-Border Loans / GDP	3y Difference	+	0.52	0.19	0.43	0.38	698	13	11	0.18	0.19	0.63	389	13	7
5. Potential Mispricing of Risk															
5.1. Stock Market Volatility		–	0.56*	0.13	0.44	0.43	2274	28	27	0.16	0.21	0.63	933	28	16
5.2.1. Stock Market Index	1y Growth	+	0.60***	0.28	0.16	0.57	1062	14	14	0.41	0.16	0.44	458	14	9
5.2.2. Stock Market Index	3y Growth	+	0.65***	0.33	0.21	0.47	958	14	14	0.34	0.22	0.44	453	14	9
5.9. VIX Index		–	0.71***	0.35	0.33	0.32	2205	28	30	0.51	0.13	0.36	947	28	16
5.10. High-Yield Spread		–	0.79***	0.49	0.17	0.33	1043	28	17	0.42	0.12	0.46	947	28	16
5.15.1. U.S. 1y T-Bill	1y Difference	+	0.63***	0.25	0.30	0.45	2612	28	31	0.21	0.19	0.61	947	28	16
5.15.2. U.S. 1y T-Bill	3y Difference	+	0.71***	0.39	0.38	0.24	2396	28	31	0.52	0.16	0.32	947	28	16
5.16.1. U.S. 1m T-Bill	1y Difference	+	0.63***	0.25	0.28	0.48	3044	28	31	0.23	0.06	0.71	947	28	16
5.16.2. U.S. 1m T-Bill	3y Difference	+	0.67***	0.35	0.37	0.28	2828	28	31	0.48	0.15	0.37	947	28	16
6. Strength of Bank Balance Sheets															
6.1.1. Leverage Ratio	1y Difference	–	0.61**	0.21	0.46	0.33	605	26	14	0.36	0.17	0.47	605	26	14
6.1.2. Leverage Ratio	3y Difference	–	0.67***	0.33	0.16	0.50	422	24	12	−0.02	0.85	0.17	422	24	12
6.3.1. Total Assets / GDP	1y Difference	+	0.64**	0.22	0.26	0.52	658	21	13	0.22	0.41	0.37	589	21	12
6.3.2. Total Assets / GDP	3y Difference	+	0.57	0.19	0.53	0.28	507	19	11	0.18	0.63	0.19	504	19	11

Notes: Sign + (–) indicates that larger (smaller) values of indicator signal a financial crisis. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent significance level, respectively, based on clustered bootstrap estimation. AUC($\leq$1) is area under the ROC curve; larger AUC is better. U_r and $U_{r,o}$($\leq$1) are the full-sample and out-of-sample relative usefulness with policy preference $\theta = 0.5$ (or equivalently $\mu = 0.9$); larger U_r is better. FNR and FPR are the false negative rate and false positive rate, respectively. N, N_c, and N_f are the number of observations, countries, and financial crises, respectively. Full-sample results are for 1970–2012; out-of-sample results are for 2000–12. All indicators are quasi-real time with a one-quarter publication lag. Detken's crisis data set is used; prediction horizon is one to three years. F.C. and D.C. refer to foreign currency and domestic currency, respectively. KK_1 is one of the indicators proposed Kauko (2012a); see equation (1) in section 2.2.

moving average of GDP perform very well. For space reasons, we only include the first version of the indicator (i.e., version that takes the ratio first, then the difference, KK_1, e.g., 1.5.8.). Even if the real credit growth rates (any definition of credit) and the corporate credit-to-GDP gap (1.8.5.) appear historically to be slightly worse predictors than other credit-to-GDP gaps, the authorities might benefit from using a broad range of credit development indicators such as those included in the table.

4.1.2 Private-Sector Debt Burden

Ratios that measure debt-servicing expenses relative to income are highly informative predictors of financial crises; see the second block in table 6. Furthermore, they are informative regardless of whether restricted to household or corporate debt-servicing costs. Our results indicate that authorities should make a special effort to monitor yearly changes in the debt-service ratio (2.1.1.). The approximations for interest rate burden (2.4., 2.4.2., 2.5., 2.5.2.), while informative, have slightly lower full-sample performance than the debt-service ratios. The difference disappears in the out-of-sample analysis.

4.1.3 Potential Overvaluation of Property Price

While all the indicators in this category are informative, the ratio of house price to rent (3.2.*) and the ratio of house price to income (3.3.*) generally outperform real house prices alone (3.1.2.); see the third block in table 6. Relating the house price to income rather than to rents apparently produces better signaling quality for the predictor. In both cases, the deviation from the long-term average and three-year differences were the highest-performing transformations. We also find evidence that growth in deflated commercial real estate prices (3.4.1.) increases the risk of a financial crisis.

4.1.4 External Imbalances

We find the ratio of current account to GDP (4.1) to be robust in full sample and out of sample; see the fourth block in table 6. Its deviation from the long-term average (4.1.8.) emerged as the most informative transformation. None of the other accounts in the balance of payments is particularly informative even full sample (details are available in table A4 in the online appendix). Changes in domestic and foreign currency cross-border loans-to-GDP ratios (4.5.2., 4.6.2.) are useful in the full sample and sometimes out of sample, but they still failed to produce statistically significant AUC with Detken's crisis data set. They perform better, however, with the alternative crisis data sets (see table 9). Given the paucity of indicators for external imbalances, we conclude that these cross-border loan ratios are worth monitoring.

4.1.5 Mispricing of Risk

Stock market volatility (5.1.) and growth in domestic stock price indexes (5.2.1., 5.2.2.) are informative predictors of risk of financial crises; see the fifth block in table 6. As global stock markets are highly interconnected, it is hardly surprising that the VIX index (5.9.) performs as well as or better than domestic stock market-based measures. We also find evidence of low (and subsequently increasing) interest rates (5.15.*, 5.16.*) and pricing of credit risk as an indicator of heightened risk of crisis. The lower spread between European high-yield and investment-grade corporate bonds (5.10.) shows very good performance in the full sample and out of sample, even when compared with the indicators in the credit developments category. Finally, both full-sample and out-of-sample metrics give some support to the predictive ability of lower household and corporate borrowing rate spreads. However, their usefulness values are quite low. For the high-yield spread, it is the lower value of interest rate spread that signals the risk. We also find that a rise in short-term U.S. interest rates (e.g., one-month and one-year maturities) signals increased vulnerability with high performance in the full sample and out of sample. These findings help explain why financial crises tend to cluster in time and affect multiple countries simultaneously.

4.1.6 Strength of Bank Balance Sheets

As noted above, the data series for bank balance sheets are generally quite short compared with our other indicator categories. We find that the leverage ratio (6.1.*) and the total assets-to-GDP ratio (6.3.*) are the only two indicators that have relatively robust performance both in the full sample and out of sample; see the last block in table 6. Using the leverage ratio in countercyclical capital buffer decisions could, however, prove problematic. We see that in the most recent financial crisis, banks built up excessive leverage while maintaining risk-based capital ratios. As a result, Basel III introduces a minimum requirement for bank leverage ratios to be implemented on January 1, 2017. Hence, it may be superfluous to use the current capital positions of banks in deciding whether banks need more capital or not. Also, due to the changes in the legislation, it is likely that this indicator will not work as well as it does here in the future. Regarding other bank balance sheet measures, we find evidence in the full sample that a large net short-term liabilities ratio,[24] large non-core-liabilities, and a large loans-to-deposits ratio signal increased risk of a financial crisis (see table A7 in the online appendix for details). None of these findings, however, extend to the out-of-sample evaluation (possibly due to short length of the data series).

As a general remark regarding the relationship among the performance measures in table 6, the observed AUC and relative usefulness have very high correlation (0.97). The out-of-sample relative usefulness has somewhat higher correlation with the full-sample relative usefulness (0.41) than with the AUC (0.29).

4.2 Robustness to Alternative Prediction Horizons

We now consider whether the choice of prediction horizon affects the quality of indicator warning signal. For example, some indicators might signal a banking crisis only six months before the onset of the crisis, while other indicators could be informative at longer prediction horizons. We follow the approach of Drehmann and Juselius (2014) and investigate the signaling quality when the prediction horizon is fixed at lengths extending from six months to five years. Similar to Drehmann and Juselius (2014), we focus here only on the AUC statistics because, as noted earlier, they are highly correlated with the relative usefulness statistics. Drehmann and Juselius

[24]Recall from table 1 that Net ST liabilities ratio = (Short-term liabilities – Liquid assets) / Total assets.

(2014) impose two additional stability conditions on policy-relevant indicators:

- interpretation of the signal should not reverse during the policy-relevant horizon,[25] and
- signaling quality should improve as the forecast horizon shortens.

Table 7 shows the AUC statistics at different prediction horizons for the recommended set of indicators introduced in the preceding subsection. We also highlight the interpretation of each indicator with a (+) sign if higher values of the indicator signal the crisis, and with a (–) sign if lower values of the indicator signal the crisis.

The recommended indicators generally satisfy the stability criteria of Drehmann and Juselius (2014) at the policy-relevant horizon, and most indicators become more informative as the crisis nears; see table 7. Two exceptions are the cross-border loans indicators (4.5.2., 4.6.2.) and the leverage ratio indicator (6.1.1.–6.1.2.). If the relevant policy horizon extends beyond three years, the cross-border loans indicators fail the first condition, as they have a reverse interpretation or are not informative at horizons longer than three years. The leverage ratio fulfills the first condition but fails to meet the second condition, as its signaling quality does not improve when the forecast horizon shortens. As noted by Behn et al. (2013), it may be that banks tend to be highly profitable in the years immediately preceding a financial crisis.

Indicators based on the credit-to-GDP ratio appear to signal crises from up to three and even five years; see the first block in table 7. Like other indicators with GDP in the denominator (e.g., current account to GDP, 4.1., and total assets to GDP, 6.3.1.), they are particularly informative in the late pre-crisis period (one or two quarters before the crisis); see the first, second, fourth, and sixth block in table 7. This is because a slowdown in GDP growth often precedes (and certainly follows) a financial crisis. Unfortunately, at

[25] In Drehmann and Juselius (2014), the relevant horizon for policy considerations is more than a year and less than five years. However, we assume the upper limit for countercyclical capital buffer considerations is only three years, as in Detken et al. (2014).

Table 7. AUC Statistics for Specific Prediction Horizons

Indicator	Transformation	Sign	Distance to Crisis (in Quarters)										
			2	4	6	8	10	12	14	16	18	20	Lag
1. Credit Developments													
1.1.1. Real Total Credit	1y Growth	+	0.65	0.72	0.69	0.68	0.69	0.66	0.58	0.57	0.58	0.52	8.7
1.2.1. Real Total Bank Credit	1y Growth	+	0.68	0.67	0.68	0.73	0.78	0.73	0.66	0.62	0.61	0.59	10.0
1.3.1. Real Household Credit	1y Growth	+	0.57	0.55	0.66	0.71	0.76	0.74	0.69	0.69	0.72	0.71	12.4
1.4.1. Real Corporate Credit	1y Growth	+	0.67	0.81	0.71	0.64	0.60	0.57	0.45	0.37	0.40	0.40	5.8
1.5.5. Total Credit / GDP	Trend Gap	+	0.86	0.89	0.86	0.85	0.84	0.82	0.79	0.77	0.78	0.81	10.4
1.5.8. Total Credit / GDP	KK_1	+	0.84	0.86	0.84	0.82	0.84	0.84	0.76	0.68	0.67	0.62	9.5
1.6.5. Total Bank Credit / GDP	Trend Gap	+	0.86	0.86	0.86	0.88	0.86	0.82	0.78	0.77	0.78	0.77	10.3
1.6.8. Total Bank Credit / GDP	KK_1	+	0.83	0.78	0.83	0.84	0.85	0.87	0.82	0.74	0.71	0.70	10.3
1.7.5. Total Household Credit / GDP	Trend Gap	+	0.80	0.85	0.88	0.88	0.87	0.85	0.82	0.79	0.77	0.79	10.6
1.7.8. Total Household Credit / GDP	KK_1	+	0.79	0.79	0.87	0.90	0.91	0.89	0.85	0.85	0.82	0.83	11.1
1.8.5. Total Corporate Credit / GDP	Trend Gap	+	0.79	0.81	0.69	0.66	0.60	0.57	0.53	0.55	0.58	0.68	8.5
1.8.8. Total Corporate Credit / GDP	KK_1	+	0.80	0.87	0.79	0.77	0.73	0.67	0.54	0.45	0.49	0.46	6.5
2. Private-Sector Debt Burden													
2.1.1. Debt-Service Ratio	1y Difference	+	0.82	0.86	0.83	0.81	0.78	0.69	0.64	0.60	0.61	0.56	8.3
2.2.1. Corporate Debt-Service Ratio	1y Difference	+	0.77	0.81	0.78	0.75	0.65	0.61	0.54	0.53	0.47	0.39	6.5
2.3.1. Household Debt-Service Ratio	1y Difference	+	0.59	0.71	0.80	0.81	0.76	0.75	0.65	0.65	0.67	0.64	10.5
2.4. HH Credit × 10y Rate / GDP		+	0.72	0.73	0.72	0.69	0.63	0.61	0.63	0.62	0.60	0.58	9.1
2.4.2. HH Credit × 10y Rate / GDP	3y Difference	+	0.79	0.76	0.74	0.76	0.60	0.54	0.61	0.64	0.60	0.61	8.7
2.5. HH Credit × 3m Rate / GDP		+	0.74	0.73	0.67	0.60	0.51	0.50	0.54	0.52	0.52	0.51	5.7
2.5.1. HH Credit × 3m Rate / GDP	1y Difference	+	0.77	0.83	0.82	0.74	0.57	0.55	0.55	0.48	0.47	0.36	5.8
3. Potential Overvaluation of Property Prices													
3.1.2. Real House Price	3y Growth	+	0.63	0.66	0.68	0.70	0.69	0.66	0.63	0.59	0.59	0.60	10.0
3.2.1. House Price / Rent	1y Difference	+	0.47	0.59	0.56	0.65	0.74	0.69	0.69	0.62	0.56	0.58	11.7
3.2.2. House Price / Rent	3y Difference	+	0.66	0.71	0.72	0.74	0.71	0.68	0.65	0.62	0.61	0.61	9.9
3.2.8. House Price / Rent	Avg. Gap	+	0.78	0.79	0.78	0.76	0.74	0.72	0.71	0.67	0.64	0.59	9.4
3.3.1. House Price / Income	1y Difference	+	0.43	0.60	0.64	0.74	0.75	0.76	0.74	0.60	0.52	0.67	11.5
3.3.2. House Price / Income	3y Difference	+	0.71	0.79	0.81	0.81	0.77	0.76	0.74	0.72	0.69	0.69	10.4
3.3.8 House Price / Income	Avg. Gap	+	0.83	0.85	0.84	0.84	0.82	0.81	0.77	0.74	0.71	0.69	10.0
3.4.1 Real Commercial Real Estate Price	1y Growth	+	0.53	0.67	0.77	0.76	0.69	0.64	0.52	0.39	0.33	0.33	7.7

(*continued*)

Table 7. (Continued)

Indicator	Transfor-mation	Sign	Distance to Crisis (in Quarters) 2	4	6	8	10	12	14	16	18	20	Lag
4. External Imbalances													
4.1. Current Account / GDP		–	0.71	0.68	0.64	0.64	0.63	0.61	0.58	0.55	0.56	0.57	8.6
4.1.8. Current Account / GDP	Avg. Gap	–	0.79	0.77	0.73	0.70	0.69	0.69	0.65	0.58	0.57	0.62	8.9
4.5.2. F.C. Cross-Border Loans / GDP	3y Difference	+	0.62	0.63	0.57	0.57	0.51	0.45	0.40	0.39	0.42	0.42	4.6
4.6.2. D.C. Cross-Border Loans / GDP	3y Difference	+	0.62	0.61	0.55	0.51	0.47	0.42	0.42	0.41	0.42	0.40	3.7
5. Potential Mispricing of Risk													
5.1. Stock Market Volatility		–	0.38	0.52	0.59	0.56	0.55	0.57	0.60	0.58	0.63	0.52	12.7
5.2.1. Stock Market Index	1y Growth	+	0.43	0.58	0.64	0.64	0.58	0.60	0.65	0.68	0.50	0.29	10.6
5.2.2. Stock Market Index	3y Growth	+	0.55	0.64	0.69	0.71	0.59	0.41	0.38	0.30	0.22	0.21	6.4
5.9. VIX Index		–	0.64	0.72	0.73	0.72	0.69	0.68	0.65	0.62	0.53	0.39	8.7
5.10. High-Yield Spread		–	0.66	0.82	0.87	0.80	0.74	0.69	0.79	0.74	0.72	0.62	10.3
5.15.1. U.S. 1y T-Bill	1y Difference	+	0.46	0.54	0.59	0.64	0.70	0.69	0.65	0.62	0.54	0.42	11.2
5.15.2. U.S. 1y T-Bill	3y Difference	+	0.62	0.70	0.73	0.75	0.70	0.63	0.56	0.44	0.33	0.29	7.4
5.16.1. U.S. 1m T-Bill	1y Difference	+	0.46	0.55	0.59	0.64	0.68	0.66	0.58	0.55	0.51	0.40	10.1
5.16.2. U.S. 1m T-Bill	3y Difference	+	0.64	0.70	0.70	0.71	0.66	0.58	0.54	0.42	0.36	0.32	6.9
6. Strength of Bank Balance Sheets													
6.1.1. Leverage Ratio	1y Difference	–	0.49	0.51	0.58	0.59	0.69	0.74	0.69	0.55	0.50	0.57	11.8
6.1.2. Leverage Ratio	3y Difference	–	0.64	0.67	0.69	0.65	0.73	0.75	0.69	0.61	0.67	0.59	10.6
6.3.1. Total Assets / GDP	1y Difference	+	0.71	0.69	0.65	0.63	0.61	0.59	0.53	0.48	0.46	0.49	6.3
6.3.2. Total Assets / GDP	3y Difference	+	0.66	0.64	0.59	0.54	0.54	0.51	0.48	0.53	0.54	0.59	8.2

Notes: Sign + (–) indicates that larger (smaller) values of indicator signal a financial crisis. AUC($\leq$1) is area under the ROC curve; larger AUC is better. All indicators are quasi-real time with a one-quarter publication lag. The time period is 1970–2012. Detken's crisis data set is used. F.C. and D.C. refer to foreign currency and domestic currency, respectively. KK_1 is one of the indicators proposed Kauko (2012a); see equation (1) in section 2.2. Lag is the weighted-average prediction horizon (in quarters) where the indicator is useful, and it is calculated as $Lag = \sum_{l=2}^{20} l * \max\ (AUC(l) - 0.5, 0) / \sum_{(k=2)}^{20} \max\ (AUC(k) - 0.5, 0)$.

that point it is likely too late for the policymaker to increase the countercyclical buffer without risking doing more damage than good.

The debt-service ratio (2.1.1., 2.2.1.) and interest burden indicators (2.4., 2.4.2., 2.5., 2.5.1.) are especially good in the short horizon; see the second block in table 7. In addition to the decline in income in the denominator of this ratio, the numerator of the debt-service ratio typically catches the rise in the interest rate that often triggers the recession in the economy.

Indicators based on asset prices such as stock index growth (5.2.1., 5.2.2.), real commercial real estate prices (3.4.1.), and the one-year change in the house price-to-income ratio (3.3.1.) typically start to fall already before the onset of a crisis; see the third and fifth block in table 7. Whether a decline in asset prices triggers the crisis or asset prices actually anticipate the future downturn may not matter, as these indicators are prone to change before the onset of a financial crisis.

The key take-away here is that policymakers should take into account the fact that the relevance of different indicators may depend on the remoteness from the crisis. Table 8 conveys the information in table 7 in a more practical format that could be useful for policymakers.[26] The recommended set of indicators are categorized into three categories according to the relevant policy horizon for that particular indicator. Short-term (one to two years) indicators tend to signal relatively late, giving the policymaker little time to react; see the first column in table 8. The medium-term indicators work best two to three years before the crisis. A few indicators, including some credit-based measures and low stock market volatility, appear informative even in the longer term (four to five years). Many indicators fall into several categories at the same time.

4.3 Robustness to Alternative Crisis Data Sets

The financial crisis data sets made available by various authors are of great benefit to early warning study, yet the definition of what constitutes a crisis colors every data set. This leads to considerable differences across the alternative crisis data sets. To fill the gap in

[26]In this categorization, we do not consider prediction horizons of less than a year, as the policymaker would essentially have no time to react.

Table 8. Summary Table of the Most Relevant Crisis-Prediction Horizons for Recommended Indicators

Indicator	Short Term 1–2 Years	Medium Term 2–3 Years	Long Term 4–5 Years
1. Credit Developments			
1.1. Real Total Credit		X	
1.2. Real Total Bank Credit		X	
1.3. Real Household Credit		X	X
1.4. Real Corporate Credit	X		
1.5. Total Credit / GDP	X	X	X
1.6. Total Bank Credit / GDP	X	X	X
1.7. Total Household Credit / GDP	X	X	X
1.8. Total Corporate Credit / GDP	X		
2. Private-Sector Debt Burden			
2.1. Debt-Service Ratio	X	X	
2.2. Corporate Debt-Service Ratio	X	X	
2.3. Household Debt-Service Ratio		X	
2.4. Total HH Credit × 10y Rate / GDP	X		
2.5. Total HH Credit × 3m Rate / GDP	X		
3. Potential Overvaluation of Property Prices			
3.1. Real House Price		X	
3.2. House Price / Rent	X	X	
3.3. House Price / Income	X	X	
3.4. Real Commercial Real Estate Price		X	
4. External Imbalances			
4.1. Current Account / GDP	X	X	
4.5. F.C. Cross-Border Loans / GDP	X		
4.6. D.C. Cross-Border Loans / GDP	X		
5. Potential Mispricing of Risk			
5.1. Stock Market Volatility		X	X
5.2. Stock Market Index		X	
5.9. VIX Index	X	X	
5.10. High-Yield Spread	X	X	
5.15. U.S. 1y T-Bill	X		
5.16. U.S. 1m T-Bill	X		
6. Strength of Bank Balance Sheets			
6.1. Leverage Ratio		X	
6.3. Total Assets / GDP	X	X	

Notes: Categorization is based on the AUC statistics for different prediction horizons reported in table 7. The prediction horizons where the indicator has relatively high performance, i.e., relative to its own performance at different prediction horizons, are marked with X.

the literature and further examine the stability of the indicators, we reproduce the performance measures in table 6 for the two additional (Babecký's and Laeven's) crisis data sets. The results of this robustness exercise are shown in table 9.

While all predictors remain informative, their rankings change depending on the crisis data set used. On average, the performance measures are significantly higher for Detken's and Laeven's data set compared with Babecký's data set, while there is on average no difference between Detken's and Laeven's data set.

The result that the crises in Detken's and Laeven's data sets are relatively easier to predict than the crises in Babecký's data set may derive from the fact that the two former data sets aim to include only systemic banking crises while the latter aims to include all banking crises. It is plausible that systemic banking crises emerge from larger economic imbalances than smaller banking crises, and the larger imbalances are then easier to detect with the early warning indicators.

In terms of average full-sample measures (AUC and U_r), the various credit-to-GDP measures are the best predictors only with Detken's data set; see the first block in table 9.[27] Using Babecký's data set, some measures of the overvaluation of property prices and mispricing of risk categories have better full-sample metrics and similar or better out-of-sample metrics; see the third block in table 9. Measures for the mispricing of risk also rank high using Laeven's data set (by any measure). The VIX index and U.S. Treasury bills perform especially good out of sample; see the fifth block in table 9. In contrast, the high-yield spread has lower out-of-sample performance for the two alternative data sets—even if its full-sample performance attains the highest numbers of all (AUC 0.88 and 0.89, U_r 0.71 and 0.70). As noted before, due to short length of time series, the indicators classified as strength of bank balance sheets have low

[27]The construction of Detken's data set from Babecký's data set implies that the improved performance of credit development indicators could result from the policy-oriented adjustments of crisis episodes that have been performed in deriving Detken's data set (see section 3.2 for the adjustments made). However, the near-identical successful performance of credit-related indicators for both Detken's and Laeven's data sets supports the notion that these indicators are good at predicting systemic banking crises (as opposed to non-systemic crises).

Table 9. Robustness of the Predictors against Alternative Crisis Variables

Indicator	Transformation	Sign	Babecký's Crisis Scheme			Detken's Crisis Scheme			Laeven's Crisis Scheme		
			AUC	U_r	$U_{r,o}$	AUC	U_r	$U_{r,o}$	AUC	U_r	$U_{r,o}$
1. Credit Developments											
1.1.1. Real Total Credit	1y Growth	+	0.50	0.11	0.17	0.69***	0.30	0.24	0.65***	0.29	0.03
1.2.1. Real Total Bank Credit	1y Growth	+	0.54	0.12	0.18	0.71***	0.37	0.33	0.68***	0.32	0.03
1.3.1. Real Household Credit	1y Growth	+	0.56	0.20	0.14	0.66***	0.27	0.17	0.64**	0.33	0.00
1.4.1. Real Corporate Credit	1y Growth	+	0.48	0.04	−0.05	0.69***	0.29	0.21	0.63***	0.22	0.05
1.5.5. Total Credit / GDP	Trend Gap	+	0.70***	0.32	0.20	0.82***	0.53	0.32	0.78***	0.53	0.25
1.5.8. Total Credit / GDP	KK_1	+	0.66***	0.35	0.24	0.80***	0.53	0.31	0.83***	0.64	0.35
1.6.5. Total Bank Credit / GDP	Trend Gap	+	0.66***	0.28	0.16	0.83***	0.55	0.29	0.74***	0.42	0.14
1.6.8. Total Bank Credit / GDP	KK_1	+	0.67***	0.36	0.25	0.80***	0.55	0.38	0.81***	0.56	0.30
1.7.5. Total Household Credit / GDP	Trend Gap	+	0.71***	0.39	0.25	0.83***	0.57	0.42	0.80***	0.53	0.31
1.7.8. Total Household Credit / GDP	KK_1	+	0.72***	0.42	0.35	0.82***	0.55	0.47	0.83***	0.62	0.27
1.8.5. Total Corporate Credit / GDP	Trend Gap	+	0.53	0.11	−0.12	0.66***	0.28	0.11	0.61*	0.25	0.04
1.8.8. Total Corporate Credit / GDP	KK_1	+	0.60**	0.20	0.28	0.77***	0.42	0.30	0.77***	0.46	0.21
2. Private-Sector Debt Burden											
2.1.1. Debt-Service Ratio	1y Difference	+	0.60**	0.24	0.12	0.78***	0.42	0.26	0.74***	0.40	0.15
2.2.1. Corporate Debt-Service Ratio	1y Difference	+	0.63***	0.25	0.09	0.73***	0.39	0.20	0.72***	0.38	0.13
2.3.1. Household Debt-Service Ratio	1y Difference	+	0.72***	0.36	0.13	0.75***	0.37	0.21	0.76***	0.45	0.12
2.4. HH Credit × 10y Rate / GDP		+	0.55	0.15	0.06	0.66***	0.33	0.41	0.51	0.12	−0.02
2.4.2. HH Credit × 10y Rate / GDP	3y Difference	+	0.54	0.12	−0.13	0.68***	0.34	0.23	0.62***	0.23	−0.03
2.5. HH Credit × 3m Rate / GDP		+	0.52	0.13	−0.03	0.60**	0.20	0.23	0.54	0.15	0.05
2.5.1. HH Credit × 3m Rate / GDP	1y Difference	+	0.70***	0.35	−0.07	0.71***	0.29	0.14	0.79***	0.45	0.32
3. Potential Overvaluation of Property Prices											
3.1.2. Real House Price	3y Growth	+	0.66***	0.32	0.18	0.67***	0.30	0.14	0.70***	0.38	0.08
3.2.1. House Price / Rent	1y Difference	+	0.67***	0.33	0.13	0.64**	0.27	0.09	0.68***	0.30	0.12
3.2.2. House Price / Rent	3y Difference	+	0.68***	0.32	0.19	0.70***	0.34	0.16	0.72***	0.42	0.12
3.2.8. House Price / Rent	Avg. Gap	+	0.71***	0.39	0.07	0.74***	0.45	0.25	0.79***	0.53	0.03
3.3.1. House Price / Income	1y Difference	+	0.69***	0.36	0.29	0.69***	0.33	0.30	0.70***	0.32	0.11
3.3.2. House Price / Income	3y Difference	+	0.71***	0.38	0.33	0.77***	0.45	0.26	0.75***	0.47	0.25
3.3.8 House Price / Income	Avg. Gap	+	0.77***	0.45	0.02	0.81***	0.52	0.31	0.80***	0.57	0.14
3.4.1 Real Commercial Real Estate Price	1y Growth	+	0.61**	0.20	0.07	0.73***	0.39	0.39	0.61**	0.19	0.11

(*continued*)

Table 9. (Continued)

Indicator	Transformation	Sign	Babecký's Crisis Scheme			Detken's Crisis Scheme			Laeven's Crisis Scheme		
			AUC	U_r	$U_{r,o}$	AUC	U_r	$U_{r,o}$	AUC	U_r	$U_{r,o}$
4. External Imbalances											
4.1. Current Account / GDP		–	0.51	0.14	−0.11	0.64*	0.30	0.14	0.52	0.17	−0.05
4.1.8. Current Account / GDP	Avg. Gap	–	0.56	0.21	0.10	0.70**	0.41	0.29	0.60	0.32	0.20
4.5.2. F.C. Cross-Border Loans / GDP	3y Difference	+	0.59	0.29	0.09	0.56	0.24	0.22	0.75***	0.48	0.25
4.6.2. D.C. Cross-Border Loans / GDP	3y Difference	+	0.60	0.29	0.06	0.52	0.19	0.18	0.75***	0.46	0.17
5. Potential Mispricing of Risk											
5.1. Stock Market Volatility		–	0.61***	0.21	0.29	0.56*	0.13	0.16	0.64***	0.25	0.28
5.2.1. Stock Market Index	1y Growth	+	0.61***	0.28	0.37	0.60***	0.28	0.41	0.66***	0.38	0.25
5.2.2. Stock Market Index	3y Growth	+	0.54	0.13	0.30	0.65***	0.33	0.34	0.67***	0.38	0.45
5.9. VIX Index		–	0.70***	0.38	0.45	0.71***	0.35	0.51	0.81***	0.52	0.67
5.10. High-Yield Spread		–	0.88***	0.71	0.06	0.79***	0.49	0.42	0.89***	0.70	0.08
5.15.1. U.S. 1y T-Bill	1y Difference	+	0.70***	0.38	0.61	0.63***	0.25	0.21	0.69***	0.36	0.14
5.15.2. U.S. 1y T-Bill	3y Difference	+	0.71***	0.40	0.39	0.71***	0.39	0.52	0.80***	0.55	0.64
5.16.1. U.S. 1m T-Bill	1y Difference	+	0.67***	0.38	0.62	0.63***	0.25	0.23	0.70***	0.37	0.30
5.16.2. U.S. 1m T-Bill	3y Difference	+	0.66***	0.35	0.29	0.67***	0.35	0.48	0.76***	0.52	0.66
6. Strength of Bank Balance Sheets											
6.1.1. Leverage Ratio	1y Difference	–	0.57*	0.20	−0.09	0.61**	0.21	0.36	0.51	0.07	−0.04
6.1.2. Leverage Ratio	3y Difference	–	0.58*	0.20	−0.18	0.67***	0.33	−0.02	0.51	0.14	−0.15
6.3.1. Total Assets / GDP	1y Difference	+	0.63***	0.23	−0.03	0.64**	0.22	0.22	0.66***	0.27	0.00
6.3.2. Total Assets / GDP	3y Difference	+	0.59*	0.22	−0.06	0.57	0.19	0.18	0.66**	0.29	0.04

Notes: Sign + (–) indicates that larger (smaller) values of indicator signal a financial crisis. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent significance level, respectively, based on clustered bootstrap estimation. AUC($\leq$1) is area under the ROC curve; larger AUC is better. U_r and $U_{r,o}$($\leq$1) are the full-sample and out-of-sample relative usefulness with policy preference $\theta = 0.5$ (or equivalently $\mu = 0.9$); larger U_r is better. Time period is 1970–2012. Full-sample results are for 1970–2012; out-of-sample results are for 2000–12. All indicators are quasi-real time with a one-quarter publication lag. F.C. and D.C. refer to foreign currency and domestic currency, respectively. KK_1 is one of the indicators proposed Kauko (2012a); see equation (1) in section 2.2. See section 3.2 for crisis data set labeling.

out-of-sample performance across the alternative crisis data sets; see the sixth block in table 9.

4.4 Interpreting Indicators for Policy Guidance

So far, we have identified indicators the policymaker should monitor to detect increased vulnerability ahead of a systemic banking crisis. Unfortunately, the policymaker must also correctly interpret signals or lack thereof from these indicators. While the interpretation ultimately depends on the policymaker's overall perception of financial stability and economic outlook, we offer some quantitative insights that may be helpful.

For most of these indicators, interpretation is straightforward in the sense that the higher (or lower) the value of the indicator, the more likely the risk of financial crisis. However, the policymaker also has to decide at which point the indicators have moved sufficiently to justify policy action. Within the EU, national policymakers consider the appropriateness of the countercyclical capital buffer every three months. If they find a need, for example, to raise the countercyclical capital buffer level, they can increase it gradually from 0 percent to the maximum 2.5 percent over a period of several years. The process involves a number of decisions that take place at different levels of the indicators. While the benchmark buffer guide readily suggests a value for the countercyclical buffer,[28] it is necessary for the policymaker to judge whether other relevant indicators comport with the benchmark story. While a comprehensive analysis of these issues is beyond the scope of the current paper, such comparison could at its simplest be achieved via descriptive analysis of historical values of indicators using, say, a logit or probit model to estimate the correspondence between crisis probabilities and indicator values.

In the online appendix, we report the statistical significance of logit-model coefficients as an additional robustness check for the warning indicators (see tables A1–A7 in the online appendix). Here, while we are reluctant to attach a specific crisis probability to a given value of the indicators, we offer a few insights that can be drawn from

[28]The benchmark buffer guide maps the value of the total credit-to-GDP trend gap into a value of countercyclical capital buffer.

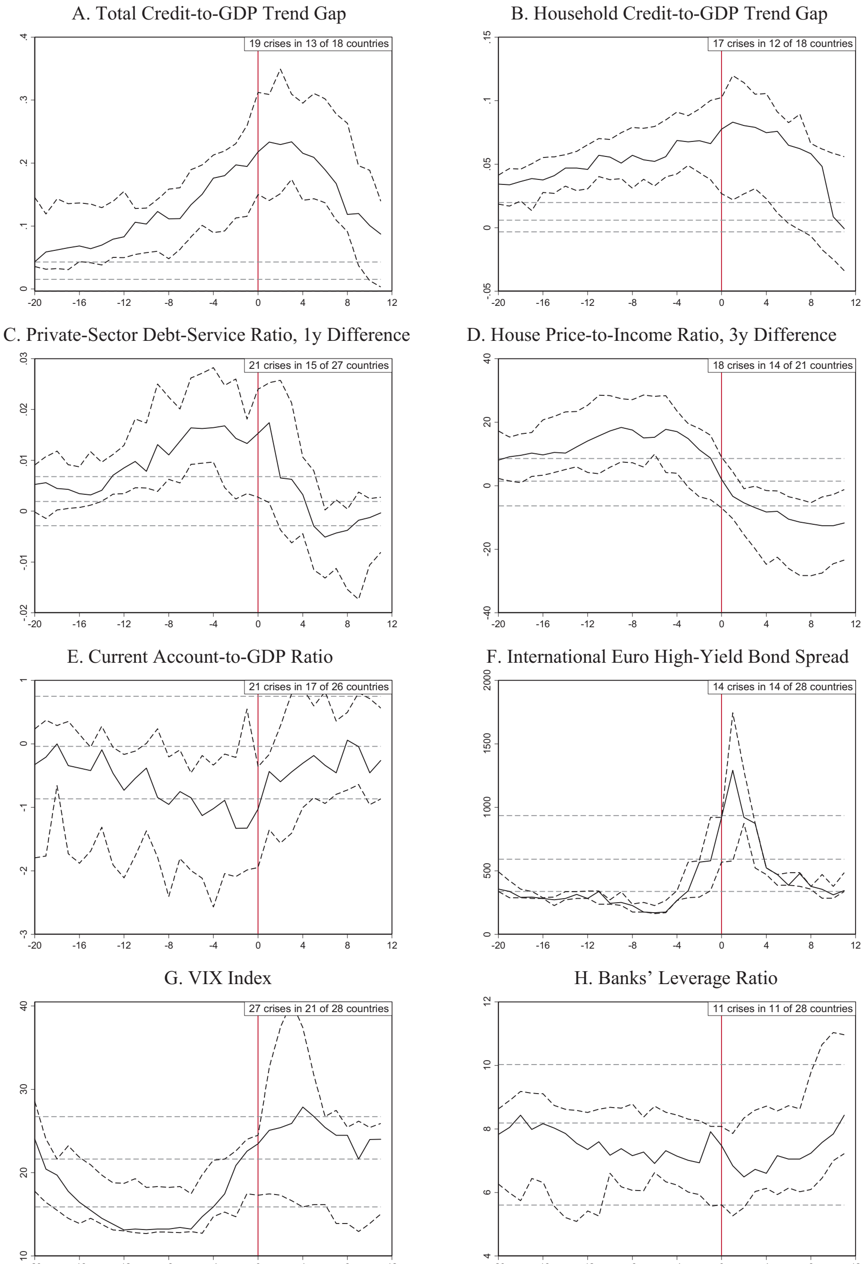
the logit estimates.[29] Regarding the credit-to-GDP trend gaps, we conclude that the crisis probability is more sensitive to the trend gap of total credit to households divided by GDP than the respective trend gaps that use total corporate credit or total credit. For example, if a 4 percent total credit-to-GDP trend gap corresponds to some probability of banking crisis, a 1 percent household credit-to-GDP trend gap would yield the same crisis probability. Regarding the new mispricing of risk indicators, VIX index values below 20 are associated with a significantly heightened probability of financial crisis. For the high-yield spread, values below 400 basis points are similarly associated with significantly heightened crisis probability.

Figure 3 visually draws out some of the top-performing indicators in each category. The horizontal axis shows the time to crisis such that negative values take place before the crisis, and zero, highlighted with a vertical line, corresponds to the first quarter of a financial crisis. The data are aggregated over all financial crises for which the indicator data is available for the five-year window prior to the crisis. The curves show first through third quartiles of the indicator data during this period and also during tranquil periods (the horizontal lines) for comparison.

If the aim is a specific threshold, perhaps the most common way of identifying threshold values for warning indicators is to derive them based on policymakers' preferences with respect to false alarms and missed crises (e.g., Alessi and Detken 2011; Behn et al. 2013; Detken et al. 2014; Drehmann, Borio, and Tsatsaronis 2011; Drehmann et al. 2010). In these methods, one makes an assumption about the preferences of policymakers in setting thresholds, e.g., the optimal noise-to-signal ratio or a specific formula for the policymaker's loss function with respect to missed crises and false alarms. Thus, it is not only difficult to assess the expected costs and

[29]As the logit model does not provide a good fit for some indicators, these insights are limited in scope. Also, due to sensitivity of the estimates to the data set at hand, the cited probabilities should not be treated unconditionally but conditionally on the given data set and the crisis variable used. Hence we seek to emphasize features that could remain robust in the wider context. Drehmann and Juselius (2014) note that a logit model can be employed to estimate the probabilities, but they emphasize that statistical properties of binary regression models are largely unknown under the high levels of persistency in their indicator variables.

Figure 3. Illustrations of Some Early Warning Indicators around Crisis Dates



Notes: Crisis dates are from Detken et al. (2014). The vertical line denotes the onset of crisis. First, second, and third quartile of indicator values are shown. The dashed horizontal lines show the first through third quartile of indicator values during tranquil periods. Only those crisis events for which the indicator data spans the whole thirty-two quarter period are included in the graphs. "19 crises in 13 of 18 countries" in the legend means that the corresponding graph is based on 19 crises that occurred in 13 countries, and the quartiles for tranquil periods employ data on 5 additional countries that did not have a crisis.

benefits of macroprudential policy, but about as daunting to specify an optimal trade-off. Hence, assumptions made about policymakers' preferences can be seen as somewhat arbitrary. To address this issue, Ferrari and Pirovano (2015) present a methodology for determining thresholds that is based on moments of an indicator's statistical distributions conditional on crisis periods and tranquil periods. Thresholds could also be country specific, as Ferrari and Pirovano (2015) show that their method works better when taking into account the country specificities. More complex methods try to derive thresholds based on multivariate models. Detken et al. (2014) show that this might be complicated, as there can be timing mismatch between different indicators and the data availability varies.

Given that there are significant uncertainties related to every potential method of determining thresholds, one should use them with care. Perhaps it would be wise to aim for a wider interpretation of the indicators than to aim for a single set of thresholds. One could use different methods to get a comprehensive picture of the information provided by various indicators.

5. Conclusions

The goal of this study has been to identify empirically a set of early warning indicators of banking crises that satisfy the policy requirements laid down in the EU legal framework. Specifically, we sought to identify suitable warning indicators for the ESRB's six categories for indicator measures: credit developments, private-sector debt burden, potential overvaluation of property prices, external imbalances, mispricing of risk, and strength of bank balance sheets. The results in general confirm earlier findings, but they also identify several new, highly useful predictors.

For the three most-studied categories (credit developments, private-sector debt burden, and potential overvaluation of property prices), we basically confirm earlier findings. Measures of credit-to-GDP, debt-service ratios, and measures of house price valuation and commercial real estate prices are all very good predictors of banking crises.

The previous literature reports mixed evidence for the remaining three categories (mispricing of risk, external imbalances, and strength of bank balance sheets). We propose several new predictors and subsequently report strong predictive performance for the following indicators in the category measures of potential mispricing of risk: the VIX index, the international credit spread between high-yield and investment-grade corporate bonds, and benchmark government bond yields. Our results hold firm in the full sample and out of sample, and for alternative crisis-prediction horizons and data sets. In addition, in agreement with Drehmann and Juselius (2014), we report some predictive success measures based on stock market price and stock market volatility.

In the external imbalances category, we find evidence in favor of the ratio of current account to GDP. None of the other examined items in the balance-of-payments accounts appear useful. We also propose a new predictor—the cross-border loans-to-GDP ratio—which shows some limited predictive performance.

Few of the bank balance sheet variables were robust predictors. This may have been hampered by the short time span of the available data. The strongest predictors, total banking assets-to-GDP ratio and leverage ratio, were statistically significant but otherwise showed weak performances. Several other indicators—such as a large net stable funding ratio, large non-core liabilities, and large loans-to-deposit ratios—are useful in the full sample, but that usefulness did not carry over to the out-of-sample results.

Our results contribute to the early warning literature of financial crisis and should help policymakers in selecting indicators for monitoring and making informed decisions on the countercyclical capital buffer. Our robustness checks are extensive compared with the earlier literature; we consider full-sample and out-of-sample estimations, many different transformations of the indicators, a range of prediction horizons, and three alternative financial crisis data sets. To the best of our knowledge, our robust findings on the informativeness of the VIX index and high-yield spread in predicting banking crises are new to literature.

A number of issues should be kept in mind when applying our results. First, we have selected the indicators based on evidence for the *average of all* countries.[30] Due to institutional or other country-specific features, some indicators might not work as well for some

[30] History shows that banking crises are caused by a group of fairly similar factors (Kauko 2014).

countries as others. Therefore, it might be optimal for some countries to select indicators other than those we propose when there is reason to believe that this country is not represented well in this average set of countries. Second, given that our aim has been to analyze data for as many countries as possible, we have relied mainly on public data sets. The national authorities monitoring these indicators in their own countries should avail themselves of the best available data. Nevertheless, we believe that our results hold for the indicators computed with different time series as long as they measure the same economic concepts.

References

Aikman, D., M. Galesic, G. Gigerenzer, S. Kapadia, K. Katsikopolous, A. Kothiyal, E. Murphy, and T. Neumann. 2014. "Taking Uncertainty Seriously: Simplicity versus Complexity in Financial Regulation." Financial Stability Paper No. 28, Bank of England.

Akinci, O., and J. Olmstead-Rumsey. 2015. "How Effective are Macroprudential Policies? An Empirical Investigation." International Finance Discussion Paper No. 2015-1136, Board of Governors of the Federal Reserve System.

Alessi, L., and K. Detken. 2011. "Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity." *European Journal of Political Economy* 27 (3): 520–33.

Anundsen, A., K. Gerdrup, and F. Hansen. 2014. "Bubbles and Crises: House Prices, Credit and Financial Market Turbulence." Mimeo.

Babecký, J., T. Havránek, J. Matěju, M. Rusnák, K. Šmídková, and B. Vašíček. 2014. "Banking, Debt, and Currency Crises in Developed Countries: Stylized Facts and Early Warning Indicators." *Journal of Financial Stability* 15 (December): 1–17.

Barrell, R., E. P. Davis, D. Karim, and L. Liadze. 2010. "Bank Regulation, Property Prices and Early Warning Systems for Banking Crises in OECD Countries." *Journal of Banking and Finance* 34 (9): 2255–64.

———. 2011. "How Idiosyncratic Are Banking Crises in OECD Countries?" *National Institute Economic Review* 216 (1): R53–R58.

Basel Committee on Banking Supervision. 2011. "Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems — Revised Version." (June). Bank for International Settlements.

Behn, M., C. Detken, T. Peltonen, and W. Schudel. 2013. "Setting Countercyclical Capital Buffers Based on Early Warning Models: Would It Work?" ECB Working Paper No. 1604.

Betz, F., S. Oprică, T. Peltonen, and P. Sarlin. 2013. "Predicting Distress in European Banks." ECB Working Paper No. 1597 (October).

Bonfim, D., and N. Monteiro. 2013. "The Implementation of the Countercyclical Capital Buffer: Rules versus Discretion." In *Financial Stability Report* (November): 87–110. Banco de Portugal.

Bordo, M. D., and C. M. Meissner. 2012. "Does Inequality Lead to a Financial Crisis?" *Journal of International Money and Finance* 31 (8): 2147–61.

Borio, C., and M. Drehmann. 2009. "Assessing the Risk of Banking Crises — Revisited." *BIS Quarterly Review* (March): 29–46.

Borio, C., and P. Lowe. 2002. "Assessing the Risk of Banking Crises." *BIS Quarterly Review* (December): 43–54.

Buchholst, B. V., and J. Rangvid. 2013. "Leading Indicators of Distress in Danish Banks in the Period 2008–12." Working Paper No. 85, Danmarks Nationalbank.

Bunda, I., and M. Ca' Zorzi. 2010. "Signals from Housing and Lending Booms." *Emerging Markets Review* 11 (1): 1–20.

Bussière, M., and M. Fratzscher. 2008. "Low Probability, High Impact: Policy Making and Extreme Events." *Journal of Policy Modeling* 30 (1): 111–21.

Büyükkarabacak, B., and N. T. Valev. 2010. "The Role of Household and Business Credit in Banking Crises." *Journal of Banking and Finance* 34 (6): 1247–56.

Caprio, G., and D. Klingebiel. 1996. "Bank Insolvencies: Cross-Country Experience." Policy Research Working Paper No. 1620, World Bank.

Castro, C., Á. Estrada, and J. Martínez. 2014. "The Countercyclical Capital Buffer in Spain: An Exploratory Analysis of Key Guiding Indicators." *Financial Stability Journal* (Bank of Spain) 27 (November): 31–59.

Cerutti, E. M., S. Claessens, and L. Laeven. 2017. "The Use and Effectiveness of Macroprudential Policies: New Evidence." *Journal of Financial Stability* 28 (February): 203–24.

Cerutti, E. M., R. Correa, E. Fiorentino, and E. Segalla. 2016. "Changes in Prudential Policy Instruments — A New Cross-Country Database." IMF Working Paper No. 16/110.

Comelli, F., 2014. "Comparing the Performance of Logit and Probit Early Warning Systems for Currency Crises in Emerging Market Economies." IMF Working Paper No. 14/65.

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

Demirgüc-Kunt, A., and E. Detragiache. 1998. "The Determinants of Banking Crises in Developed Countries." *IMF Staff Papers* 45 (1): 81–109.

———. 2000. "Monitoring Banking Sector Fragility: A Multivariate Logit Approach." *World Bank Economic Review* 14 (2): 287–307.

Detken. K, O. Weeken, L. Alessi, D. Bonfim, M. Boucinha, C. Castro, S. Frontczak, G. Giordana, J. Giese, N. Jahn, J. Kakes, B. Klaus, J. Lang, N. Puzanova, and P. Welz. 2014. "Operationalising the Countercyclical Capital Buffer: Indicator Selection, Threshold Identification and Calibration Options." ESRB Occasional Paper No. 5 (June).

Domaç, I., and M. S. Martinez Peria. 2003. "Banking Crises and Exchange Rate Regimes: Is There a Link? *Journal of International Economics* 61 (1): 41–72.

Drehmann, M., C. Borio, L. Gambacorta, G. Jiminez, and C. Trucharte. 2010. "Countercyclical Capital Buffers: Exploring Options." BIS Working Paper No. 317.

Drehmann, M., C. Borio, and K. Tsatsaronis. 2011. "Anchoring Countercyclical Capital Buffers: The Role of Credit Aggregates." *International Journal of Central Banking* 7 (4, December): 189–240.

Drehmann, M., and M. Juselius. 2014. "Evaluating Early Warning Indicators of Banking Crises: Satisfying Policy Requirements." *International Journal of Forecasting* 30 (3): 759–80.

European Systemic Risk Board. 2014. "Recommendation of the ESRB on Guidance for Setting Countercyclical Buffer Rates." Available at https://www.esrb.europa.eu/pub/pdf/recommendations/2014/140630_ESRB_Recommendation.en.pdf.

Ferrari, S., and M. Pirovano. 2015. "Early Warning Indicators for Banking Crises: A Conditional Moments Approach." MPRA Paper No. 62406.

Frankel, J., and A. Rose. 1996. "Currency Crashes in Emerging Markets: An Empirical Treatment." International Finance Discussion Paper No. 534, Board of Governors of the Federal Reserve System.

Giese, J., H. Andersen, O. Bush, C. Castro, M. Farag, and S. Kapadia. 2014. "The Credit-to-GDP Gap and Complementary Indicators for Macroprudential Policy: Evidence from the UK." *International Journal of Finance and Economics* 19 (1): 25–47.

Hahm, J.-H., H. S. Shin, and K. Shin. 2013. Noncore Bank Liabilities and Financial Vulnerability." *Journal of Money, Credit and Banking* 45 (s1): 3–36.

Hardy, D., and C. Pazarbaşıoğlu. 1998. "Leading Indicators of Banking Crises: Was Asia Different?" IMF Working Paper No. 98/91.

Holopainen, M., and P. Sarlin. 2015. "Toward Robust Early Warning Models: A Horse Race, Ensembles and Model Uncertainty." Discussion Paper No. 6, Bank of Finland.

Jordà, Ò., O. Schularick, and A. M. Taylor. 2015. "Leveraged Bubbles." *Journal of Monetary Economics* 76 (Supplement): S1–S20.

Joyce, J. 2011. "Financial Globalization and Banking Crises in Emerging Markets." *Open Economies Review* 22 (5): 875–95.

Kamin, S., and L. DeMarco. 2012. "How Did a Domestic Housing Slump Turn into a Global Financial Crisis?" *Journal of International Money and Finance* 31 (1): 10–41.

Kaminsky, G., and C. Reinhart. 1999. "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems." *American Economic Review* 89 (3): 473–500.

Kauko, K. 2012a. "Triggers for Countercyclical Capital Buffers." *BoF Online* 7 (2012).

———. 2012b. "External Deficits and Non-performing Loans in the Recent Financial Crisis. *Economics Letters* 115 (2): 196–99.

———. 2014. "How to Foresee Banking Crises? A Survey of the Empirical Literature." *Economic Systems* 38 (3): 289–308.

Kauko, K., J. Vauhkonen, and J. Topi. 2014. "How Should the Countercyclical Capital Buffer Requirement Be Applied?" *Bank of Finland Bulletin* 2 (2014): 57–67.

Kim, W., and S.-J Wei. 2002. "Foreign Portfolio Investors Before and During a Crisis." *Journal of International Economics* 56 (1): 77–96.

Krylova, E., 2016. "Determinants of Euro-Denominated Corporate Bond Spreads." ECB Working Paper No. 1912 (June).

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

———. 2012. "Systemic Banking Crises Database: An Update." IMF Working Paper No. 12/163.

Lainà, P., J. Nyholm, and P. Sarlin. 2015. "Leading Indicators of Systemic Banking Crises: Finland in a Panel of EU Countries." *Review of Financial Economics* 24 (January): 18–35.

Lindgren, C.-J., G. Garcia, and M. I. Saal, eds. 1996. *Bank Soundness and Macroeconomic Policy.* International Monetary Fund.

Lo Duca, M., and T. Peltonen. 2013. "Assessing Systemic Risks and Predicting Systemic Events." *Journal of Banking and Finance* 37 (7): 2183–95.

Lund-Jensen, K. 2012. "Monitoring Systemic Risk Based on Dynamic Thresholds." IMF Working Paper No. 12/159.

Phillips, P. C. B., S. Shi, and J. Yu. 2013. "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500." Discussion Paper No. 1914, Cowles Foundation.

Reinhart, C. M., and K. S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Repullo, R., and J. Saurina. 2011. "The Countercyclical Capital Buffer of Basel III: A Critical Assessment." CEPR Discussion Paper No. 8304.

Roy, S., and D. M. Kemme. 2012. "Causes of Banking Crises: Deregulation, Credit Booms and Asset Bubbles, Then and Now." *International Review of Economics and Finance* 24 (October): 270–94.

Online Appendix to Evaluating Indicators for Use in Setting the Countercyclical Capital Buffer

Eero Tölö,[a] Helinä Laakkonen,[a] and Simo Kalatie[b]
[a]Bank of Finland
[b]Savings Banks Group

Crisis-Signaling Performance Results for All Indicators and Their Transformations

The following information pertains to all tables (A1–A7) in this appendix.

The indicators and the transformations are defined in table 2 and equations (3)–(10) in the main part of the paper, respectively. Sign + (−) indicates that larger (smaller) values of indicator signal a financial crisis. $\mathrm{KK_i}$ are the indicators proposed by Kauko (2012a); see equations (1) and (2) in section 2.2.

The "AUC" column reports the area under the ROC curve ($0.5 \leq \mathrm{AUC} \leq 1$); larger AUC is better.

The "Logit" column reports the statistical significance of the respective coefficient in univariate logit model. *, **, and *** denote statistical significance with the expected sign at the 10 percent, 5 percent, and 1 percent significance level, respectively, based on clustered bootstrap estimation. † replaces * where the sign is unexpected.

The "$\mathrm{U_r}$" and "$\mathrm{U_{r,o}}$" columns report the full-sample and out-of-sample relative usefulness, respectively, with policy preference $\theta = 0.5$ or, equivalently, $\mu = 0.9$ ($\mathrm{U_r} \leq 1$); larger $\mathrm{U_r}$ is better. FNR and FPR are the false negative rate and false positive rate, respectively.

N, $\mathrm{N_c}$, and $\mathrm{N_f}$ are the number of observations, countries, and financial crises, respectively.

Full-sample results are for 1970–2012; out-of-sample results are for 2000–12. All indicators are quasi-real time with a one-quarter publication lag. Crisis data set is Detken et al. (2014); prediction horizon is one to three years.

Table A1. Performance Evaluation for Measures of Credit Developments

Indicator	Transformation	Sign	Full-Sample Evaluation AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	Out-of-Sample Evaluation $U_{r,o}$	FNR	FPR	N	N_c	N_f
1.1. Real Total Credit		+	0.70***	***	0.38	0.30	0.32	1312	15	15	0.18	0.35	0.47	475	15	9
1.1.1.	1y Growth	+	0.69***		0.30	0.07	0.63	1262	15	14	0.24	0.01	0.75	471	15	9
1.1.2.	3y Growth	+	0.67***	**	0.26	0.01	0.73	1172	15	12	–0.05	0.48	0.57	463	15	9
1.1.3.	Relative Gap	+	0.75***	***	0.41	0.01	0.58	1096	15	12	0.11	0.69	0.20	456	15	9
1.2. Real Total Bank Credit		+	0.71***	***	0.34	0.51	0.15	1293	15	15	0.12	0.42	0.46	475	15	9
1.2.1.	1y Growth	+	0.71***		0.37	0.14	0.49	1243	15	14	0.33	0.09	0.58	471	15	9
1.2.2.	3y Growth	+	0.73***	***	0.40	0.07	0.52	1153	15	12	0.23	0.13	0.64	463	15	9
1.2.3.	Relative Gap	+	0.79***	***	0.51	0.04	0.45	1077	15	12	–0.12	0.64	0.48	456	15	9
1.3. Real Household Credit		+	0.65***	**	0.25	0.16	0.59	1011	15	15	0.15	0.35	0.50	458	15	9
1.3.1.	1y Growth	+	0.66***	*	0.27	0.23	0.50	961	15	14	0.17	0.03	0.81	450	15	9
1.3.2.	3y Growth	+	0.71***	***	0.37	0.20	0.42	874	15	12	0.05	0.61	0.34	434	15	9
1.3.3.	Relative Gap	+	0.7***	***	0.41	0.06	0.53	798	15	12	0.18	0.17	0.65	417	15	9
1.4. Real Corporate Credit		+	0.61**	**	0.25	0.52	0.22	1011	15	15	0.04	0.53	0.43	458	15	9
1.4.1.	1y Growth	+	0.69***		0.29	0.29	0.43	961	15	14	0.21	0.08	0.71	450	15	9
1.4.2.	3y Growth	+	0.57*		0.14	0.17	0.69	874	15	12	0.02	0.17	0.81	434	15	9
1.4.3.	Relative Gap	+	0.58*		0.17	0.11	0.72	798	15	12	–0.24	0.66	0.58	417	15	9
1.5. Total Credit / GDP		+	0.72***	*	0.40	0.23	0.37	2114	18	22	0.04	0.00	0.96	577	18	11
1.5.1.	1y Difference	+	0.78***		0.49	0.28	0.23	2042	18	22	0.28	0.03	0.68	573	18	11
1.5.2.	3y Difference	+	0.80***		0.50	0.17	0.34	1899	18	20	0.20	0.06	0.74	565	18	11
1.5.3.	1y Growth	+	0.58*		0.15	0.19	0.66	2042	18	22	0.05	0.42	0.53	573	18	11
1.5.4.	3y Growth	+	0.55		0.12	0.01	0.87	1899	18	20	0.02	0.39	0.59	565	18	11
1.5.5.	Trend Gap	+	0.82***	***	0.53	0.24	0.23	1787	18	20	0.32	0.07	0.61	558	18	11
1.5.6.	Relative Gap	+	0.72***	***	0.35	0.13	0.53	1787	18	20	0.08	0.40	0.52	558	18	11
1.5.7.	5y M.A. Gap	+	0.81***		0.52	0.13	0.35	1787	18	20	0.25	0.01	0.74	558	18	11
1.5.8.	KK_1	+	0.80***		0.53	0.31	0.16	2042	18	22	0.31	0.00	0.69	573	18	11
1.5.9.	KK_2	+	0.79***		0.50	0.28	0.23	2054	18	22	0.18	0.00	0.82	573	18	11

(*continued*)

Table A1. (Continued)

Indicator	Transformation	Sign	Full-Sample Evaluation: AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	Out-of-Sample Evaluation: $U_{r,o}$	FNR	FPR	N	N_c	N_f
1.6. Total Bank Credit / GDP		+	0.74***	***	0.39	0.10	0.51	2082	18	22	0.13	0.00	0.87	577	18	11
1.6.1.	1y Difference	+	0.78***	***	0.53	0.31	0.16	2010	18	22	0.37	0.05	0.58	573	18	11
1.6.2.	3y Difference	+	0.84***	***	0.60	0.17	0.23	1867	18	20	0.32	0.11	0.57	565	18	11
1.6.3.	1y Growth	+	0.60**		0.16	0.27	0.56	2010	18	22	0.24	0.41	0.35	573	18	11
1.6.4.	3y Growth	+	0.60*		0.19	0.11	0.71	1867	18	20	0.25	0.26	0.49	565	18	11
1.6.5.	Trend Gap	+	0.83***	***	0.55	0.26	0.19	1755	18	20	0.29	0.31	0.41	558	18	11
1.6.6.	Relative Gap	+	0.74***	***	0.37	0.25	0.38	1755	18	20	0.25	0.29	0.45	558	18	11
1.6.7.	5y M.A. Gap	+	0.84***	***	0.60	0.17	0.23	1755	18	20	0.35	0.06	0.59	558	18	11
1.6.8.	KK_1	+	0.80***	***	0.55	0.22	0.24	2010	18	22	0.38	0.08	0.53	573	18	11
1.6.9.	KK_2	+	0.80***	***	0.55	0.20	0.25	2022	18	22	0.37	0.02	0.61	573	18	11
1.7. Total Household Credit / GDP		+	0.72***	***	0.37	0.54	0.09	1436	18	20	0.19	0.04	0.77	560	18	11
1.7.1.	1y Difference	+	0.80***	***	0.51	0.20	0.29	1368	18	20	0.38	0.02	0.60	552	18	11
1.7.2.	3y Difference	+	0.86***	***	0.63	0.12	0.26	1240	18	18	0.50	0.01	0.49	536	18	11
1.7.3.	1y Growth	+	0.58*		0.19	0.26	0.55	1368	18	20	0.13	0.32	0.56	552	18	11
1.7.4.	3y Growth	+	0.60**		0.26	0.05	0.69	1240	18	18	0.27	0.00	0.73	536	18	11
1.7.5.	Trend Gap	+	0.83***	***	0.57	0.19	0.23	1135	17	18	0.42	0.03	0.54	516	17	11
1.7.6.	Relative Gap	+	0.71***	***	0.37	0.16	0.47	1135	17	18	0.29	0.18	0.53	516	17	11
1.7.7.	5y M.A. Gap	+	0.85***	***	0.61	0.14	0.25	1135	17	18	0.50	0.02	0.47	516	17	11
1.7.8.	KK_1	+	0.82***	***	0.55	0.29	0.16	1368	18	20	0.47	0.03	0.50	552	18	11
1.7.9.	KK_2	+	0.82***	***	0.55	0.26	0.19	1372	18	20	0.39	0.00	0.61	552	18	11
1.8. Total Corporate Credit / GDP		+	0.71***		0.35	0.16	0.49	1424	18	19	0.13	0.00	0.87	560	18	11
1.8.1.	1y Difference	+	0.74***		0.41	0.31	0.28	1356	18	19	0.24	0.20	0.56	552	18	11
1.8.2.	3y Difference	+	0.65***		0.23	0.31	0.45	1227	18	18	–0.05	0.37	0.68	536	18	11
1.8.3.	1y Growth	+	0.62***	*	0.20	0.17	0.63	1356	18	19	0.11	0.59	0.30	552	18	11
1.8.4.	3y Growth	+	0.54		0.10	0.76	0.14	1227	18	18	–0.05	0.56	0.49	536	18	11
1.8.5.	Trend Gap	+	0.66***	**	0.28	0.29	0.42	1115	17	18	0.11	0.33	0.56	516	17	11
1.8.6.	Relative Gap	+	0.64***	***	0.27	0.11	0.62	1115	17	18	0.16	0.21	0.63	516	17	11
1.8.7.	5y M.A. Gap	+	0.67***	***	0.29	0.26	0.46	1115	17	18	0.06	0.28	0.66	516	17	11
1.8.8.	KK_1	+	0.77***		0.42	0.20	0.39	1356	18	19	0.30	0.16	0.54	552	18	11
1.8.9.	KK_2	+	0.77***	***	0.37	0.20	0.42	1360	18	19	0.05	0.61	0.34	552	18	11

See page 1 of this appendix for notes.

Table A2. Performance Evaluation for Measures of Private-Sector Debt Burden

Indicator	Transformation	Sign	Full-Sample Evaluation: AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	Out-of-Sample Evaluation: $U_{r,o}$	FNR	FPR	N	N_c	N_f
2.1. Debt-Service Ratio		+	0.60**	**	0.22	0.57	0.21	2256	26	27	0.01	0.33	0.66	784	26	16
2.1.1.	1y Difference	+	0.78***	***	0.42	0.36	0.22	2161	26	26	0.26	0.09	0.65	764	26	16
2.1.2	3y Difference	+	0.76***	**	0.40	0.15	0.45	1966	24	23	0.11	0.19	0.70	724	24	15
2.1.3.	1y Growth	+	0.74***	**	0.38	0.11	0.51	2161	26	26	0.22	0.09	0.69	764	26	16
2.1.4.	3y Growth	+	0.73***	***	0.37	0.16	0.47	1966	24	23	0.08	0.20	0.72	724	24	15
2.1.5.	Trend Gap	+	0.72***	***	0.32	0.13	0.55	1817	22	22	0.20	0.15	0.65	694	22	14
2.1.6.	Relative Gap	+	0.70***	**	0.32	0.13	0.55	1817	22	22	0.15	0.17	0.68	694	22	14
2.1.7.	5y M.A. Gap	+	0.77***	**	0.41	0.20	0.39	1817	22	22	0.15	0.14	0.71	694	22	14
2.2. Corporate Debt-Service Ratio		+	0.60*		0.21	0.29	0.50	1058	25	19	0.14	0.40	0.46	649	25	13
2.2.1.	1y Difference	+	0.73***		0.39	0.28	0.33	967	25	17	0.20	0.03	0.77	599	25	13
2.2.2.	3y Difference	+	0.69***		0.30	0.41	0.29	798	21	14	0.19	0.06	0.75	499	21	11
2.2.3.	1y Growth	+	0.73***	***	0.37	0.33	0.30	967	25	17	0.22	0.03	0.75	599	25	13
2.2.4.	3y Growth	+	0.68***	***	0.26	0.43	0.30	798	21	14	0.10	0.35	0.55	499	21	11
2.2.5.	Trend Gap	+	0.64*	*	0.28	0.17	0.56	673	19	13	0.14	0.20	0.66	431	19	10
2.2.6.	Relative Gap	+	0.62*		0.26	0.19	0.55	673	19	13	0.01	0.37	0.62	431	19	10
2.2.7.	5y M.A. Gap	+	0.72***	***	0.35	0.34	0.30	673	19	13	0.10	0.04	0.86	431	19	10
2.3. Household Debt-Service Ratio		+	0.66**	***	0.29	0.31	0.40	1046	25	19	0.18	0.33	0.49	649	25	13
2.3.1.	1y Difference	+	0.75***	***	0.37	0.29	0.34	952	25	18	0.21	0.00	0.79	599	25	13
2.3.2.	3y Difference	+	0.77***	***	0.45	0.45	0.10	775	21	15	0.17	0.00	0.83	499	21	11
2.3.3.	1y Growth	+	0.68***	***	0.30	0.05	0.64	952	25	18	0.20	0.00	0.80	599	25	13
2.3.4.	3y Growth	+	0.68***	**	0.26	0.27	0.47	775	21	15	0.10	0.00	0.90	499	21	11
2.3.5.	Trend Gap	+	0.70***	***	0.38	0.04	0.58	650	19	14	0.30	0.04	0.66	428	19	10
2.3.6.	Relative Gap	+	0.64*		0.40	0.05	0.55	650	19	14	0.27	0.03	0.70	428	19	10
2.3.7.	5y M.A. Gap	+	0.78***	***	0.43	0.43	0.14	650	19	14	0.19	0.00	0.81	428	19	10

(*continued*)

Table A2. (Continued)

Indicator	Transformation	Sign	Full-Sample Evaluation: AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	Out-of-Sample Evaluation: $U_{r,o}$	FNR	FPR	N	N_c	N_f
2.4. Household Credit × 10y Rate / GDP		+	0.66***	**	0.33	0.35	0.32	951	18	17	0.41	0.42	0.17	539	18	11
2.4.1.	1y Difference	+	0.61***	***	0.23	0.56	0.21	893	18	15	–0.12	0.44	0.68	522	18	11
2.4.2.	3y Difference	+	0.68***	***	0.34	0.28	0.38	788	17	14	0.23	0.34	0.44	481	17	11
2.4.3.	1y Growth	+	0.59***	**	0.18	0.53	0.29	893	18	15	0.09	0.03	0.87	522	18	11
2.4.4.	3y Growth	+	0.65***	**	0.32	0.28	0.40	788	17	14	0.20	0.32	0.48	481	17	11
2.4.5.	Trend Gap	+	0.56		0.14	0.64	0.22	694	17	13	0.10	0.00	0.90	443	17	11
2.4.6.	Relative Gap	+	0.51		0.08	0.54	0.38	694	17	13	0.02	0.00	0.98	443	17	11
2.4.7.	5y M.A. Gap	+	0.67***	***	0.32	0.44	0.24	694	17	13	–0.14	0.56	0.58	443	17	11
2.5. Household Credit × 3m Rate / GDP		+	0.60**	*	0.20	0.27	0.52	1328	25	21	0.23	0.54	0.24	738	25	15
2.5.1.	1y Difference	+	0.71***	***	0.29	0.26	0.45	1240	25	21	0.14	0.06	0.81	696	25	15
2.5.2.	3y Difference	+	0.70***	***	0.31	0.35	0.34	1065	25	20	0.03	0.40	0.57	610	25	14
2.5.3.	1y Growth	+	0.70***	***	0.31	0.41	0.28	1240	25	21	0.20	0.08	0.72	696	25	15
2.5.4.	3y Growth	+	0.69***	***	0.31	0.25	0.44	1065	25	20	0.16	0.09	0.75	610	25	14
2.5.5.	Trend Gap	+	0.68**	**	0.31	0.41	0.28	906	25	19	0.13	0.16	0.72	528	25	14
2.5.6.	Relative Gap	+	0.66***	**	0.31	0.26	0.43	906	25	19	0.17	0.07	0.76	528	25	14
2.5.7.	5y M.A. Gap	+	0.74***	***	0.37	0.27	0.36	906	25	19	0.18	0.17	0.65	528	25	14

See page 1 of this appendix for notes.

Table A3. Performance Evaluation for Measures of Potential Overvaluation of Property Prices

Indicator	Transformation	Sign	Full-Sample Evaluation								Out-of-Sample Evaluation					
			AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	$U_{r,o}$	FNR	FPR	N	N_c	N_f
3.1 Real House Price		+	0.72***	*	0.39	0.25	0.36	1649	19	22	0.12	0.00	0.88	528	19	12
3.1.1.	1y Growth	+	0.61**	*	0.20	0.59	0.21	1573	19	22	0.00	0.71	0.30	507	19	12
3.1.2.	3y Growth	+	0.67***	**	0.30	0.43	0.27	1429	16	20	0.14	0.42	0.44	465	16	11
3.1.3.	Relative Gap	+	0.69***	***	0.26	0.60	0.14	1324	15	20	–0.03	0.22	0.81	437	15	11
3.2. House Price / Rent		+	0.74***	*	0.40	0.21	0.38	1501	19	21	0.15	0.00	0.85	547	19	12
3.2.1.	1y Difference	+	0.64**	**	0.27	0.57	0.16	1428	19	21	0.09	0.52	0.39	526	19	12
3.2.2.	3y Difference	+	0.70***	***	0.34	0.42	0.24	1286	17	20	0.16	0.37	0.47	483	17	12
3.2.3.	1y Growth	+	0.62**	**	0.21	0.58	0.21	1428	19	21	–0.03	0.67	0.35	526	19	12
3.2.4.	3y Growth	+	0.68***	***	0.32	0.50	0.18	1286	17	20	0.15	0.42	0.42	483	17	12
3.2.5.	Trend Gap	+	0.72***	***	0.37	0.54	0.09	1174	16	20	0.03	0.28	0.69	448	16	12
3.2.6.	Relative Gap	+	0.68***	***	0.26	0.54	0.20	1174	16	20	–0.03	0.28	0.75	448	16	12
3.2.7.	5y M.A. Gap	+	0.71***	***	0.36	0.46	0.18	1174	16	20	0.07	0.36	0.57	448	16	12
3.2.8.	Avg. Gap	+	0.74***	***	0.45	0.38	0.16	1174	16	20	0.25	0.15	0.60	448	16	12
3.3. House Price / Income		+	0.75***	**	0.39	0.39	0.23	1487	20	21	0.15	0.00	0.85	584	20	12
3.3.1.	1y Difference	+	0.69***	***	0.33	0.50	0.18	1410	20	21	0.30	0.44	0.26	563	20	12
3.3.2.	3y Difference	+	0.77***	***	0.45	0.38	0.18	1260	18	19	0.26	0.27	0.47	512	18	12
3.3.3.	1y Growth	+	0.67***	***	0.29	0.56	0.15	1410	20	21	0.24	0.55	0.21	563	20	12
3.3.4.	3y Growth	+	0.75***	***	0.40	0.34	0.26	1260	18	19	0.15	0.49	0.36	512	18	12
3.3.5.	Trend Gap	+	0.78***	***	0.48	0.38	0.14	1148	17	19	0.16	0.22	0.62	470	17	12
3.3.6.	Relative Gap	+	0.74***	***	0.37	0.39	0.23	1148	17	19	0.13	0.26	0.61	470	17	12
3.3.7.	5y M.A. Gap	+	0.78***	***	0.48	0.33	0.19	1148	17	19	0.33	0.31	0.36	470	17	12
3.3.8.	Avg. Gap	+	0.81***	***	0.52	0.30	0.18	1148	17	19	0.31	0.09	0.61	470	17	12
3.4. Real Commercial Real Estate Price		+	0.72***	***	0.39	0.18	0.43	775	15	14	0.08	0.00	0.92	416	15	10
3.4.1.	1y Growth	+	0.73***	***	0.39	0.21	0.40	718	15	14	0.39	0.35	0.26	391	15	10
3.4.2.	3y Growth	+	0.61**		0.22	0.68	0.09	599	14	14	0.04	0.90	0.07	330	14	10
3.4.3.	Relative Gap	+	0.65**	**	0.26	0.11	0.63	501	13	14	–0.16	0.45	0.71	274	13	10

See page 1 of this appendix for notes.

Table A4. Performance Evaluation for Measures of External Imbalances

Indicator	Transformation	Sign	Full-Sample Evaluation								Out-of-Sample Evaluation					
			AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	$U_{r,o}$	FNR	FPR	N	N_c	N_f
4.1 Current Account / GDP		–	0.64*		0.30	0.45	0.25	1159	21	16	0.14	0.35	0.50	601	21	12
4.1.1.	1y Difference	–	0.63**	**	0.26	0.43	0.32	1075	20	15	0.07	0.60	0.33	568	20	11
4.1.2.	3y Difference	–	0.64***	*	0.26	0.50	0.23	921	20	13	0.14	0.41	0.45	490	20	11
4.1.3.	1y Growth	–	0.4		0.03	0.07	0.90	1075	20	15	–0.01	0.03	0.98	568	20	11
4.1.4.	3y Growth	–	0.39	†††	0.02	0.05	0.93	921	20	13	–0.42	0.80	0.62	490	20	11
4.1.5.	Trend Gap	–	0.62**	**	0.20	0.36	0.44	792	19	13	0.19	0.36	0.45	410	19	11
4.1.6.	Relative Gap	–	0.41	**	0.04	0.19	0.77	792	19	13	–0.02	0.05	0.97	410	19	11
4.1.7.	5y M.A. Gap	–	0.66***	**	0.30	0.42	0.28	792	19	13	0.15	0.47	0.38	410	19	11
4.1.8.	Avg. Gap	–	0.7**	*	0.41	0.35	0.24	792	19	13	0.29	0.43	0.28	410	19	11
4.2. Portfolio Investments / GDP		–	0.51		0.12	0.03	0.85	622	21	14	–0.01	0.52	0.49	507	21	12
4.2.1.	1y Difference	–	0.58		0.14	0.67	0.18	542	21	14	0.08	0.54	0.38	447	21	12
4.2.2.	3y Difference	–	0.61		0.20	0.15	0.65	382	20	11	0.10	0.53	0.37	306	20	10
4.2.3.	1y Growth	–	0.54		0.18	0.68	0.13	542	21	14	0.07	0.29	0.64	447	21	12
4.2.4.	3y Growth	–	0.45		0.09	0.87	0.04	382	20	11	–0.06	0.31	0.75	306	20	10
4.3. Other Investments / GDP		–	0.51		0.12	0.03	0.85	622	21	14	–0.01	0.52	0.49	507	21	12
4.3.1.	1y Difference	–	0.58		0.14	0.67	0.18	542	21	14	0.08	0.54	0.38	447	21	12
4.3.2.	3y Difference	–	0.61		0.20	0.15	0.65	382	20	11	0.10	0.53	0.37	306	20	10
4.3.3.	1y Growth	–	0.54		0.18	0.68	0.13	542	21	14	0.07	0.29	0.64	447	21	12
4.3.4.	3y Gtowth	–	0.45		0.09	0.87	0.04	382	20	11	–0.06	0.31	0.75	306	20	10
4.4. Capital Account / GDP		–	0.4		0.01	0.00	0.99	1015	21	16	0.08	0.00	0.92	601	21	12
4.4.1.	1y Difference	–	0.5		0.07	0.67	0.26	931	20	15	0.25	0.17	0.58	568	20	11
4.4.2.	3y Difference	–	0.46		0.05	0.76	0.19	777	20	13	0.09	0.13	0.78	485	20	11
4.4.3.	1y Growth	–	0.54*		0.10	0.45	0.45	872	20	14	0.08	0.57	0.35	565	20	11
4.4.4.	3y Growth	–	0.53		0.13	0.50	0.36	718	20	12	0.06	0.51	0.43	482	20	11
4.5. Foreign Currency Cross-Border Loans / GDP		+	0.57		0.18	0.32	0.49	861	15	12	–0.08	0.26	0.82	434	15	8
4.5.1.	1y Difference	+	0.53		0.16	0.58	0.26	805	15	12	0.05	0.46	0.48	422	15	8
4.5.2.	3y Difference	+	0.56		0.24	0.53	0.23	698	13	11	0.22	0.31	0.47	389	13	7
4.5.3.	1y Growth	+	0.48		0.04	0.76	0.20	774	15	12	–0.02	0.69	0.33	422	15	8
4.5.4.	3y Growth	+	0.45		0.03	0.00	0.97	668	13	10	0.01	0.00	0.99	389	13	7
4.6. Own Currency Cross-Border Loans / GDP		+	0.55		0.14	0.08	0.77	861	15	12	–0.02	0.13	0.89	434	15	8
4.6.1.	1y Difference	+	0.51		0.12	0.62	0.26	805	15	12	–0.11	0.48	0.64	422	15	8
4.6.2.	3y Difference	+	0.52		0.19	0.43	0.38	698	13	11	0.18	0.19	0.63	389	13	7
4.6.3.	1y Growth	+	0.48		0.09	0.68	0.23	774	15	12	0.06	0.63	0.31	422	15	8
4.6.4.	3y Growth	+	0.43		0.03	0.00	0.97	668	13	10	–0.02	0.03	0.99	389	13	7

See page 1 of this appendix for notes.

Table A5. Performance Evaluation for Measures of Potential Mispricing of Risk: Country-Specific Indicators

			Full-Sample Evaluation								Out-of-Sample Evaluation					
Indicator	Transformation	Sign	AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	$U_{r,o}$	FNR	FPR	N	N_c	N_f
5.1. Stock Market Volatility		–	0.56*		0.13	0.44	0.43	2274	28	27	0.16	0.21	0.63	933	28	16
5.1.1.	1y Difference	–	0.49		0.08	0.20	0.72	2169	28	26	–0.12	0.86	0.27	917	28	16
5.1.2.	3y Difference	–	0.53		0.10	0.31	0.59	1957	28	24	0.25	0.34	0.42	885	28	15
5.1.3.	1y Growth	–	0.49		0.03	0.20	0.77	2168	28	26	–0.07	0.89	0.17	917	28	16
5.1.4.	3y Growth	–	0.53*	**	0.10	0.15	0.74	1956	28	24	0.22	0.11	0.67	885	28	15
5.2. Stock Market Index		+	0.53		0.11	0.37	0.53	1114	14	14	–0.24	0.42	0.82	458	14	9
5.2.1.	1y Growth	+	0.60***		0.28	0.16	0.57	1062	14	14	0.41	0.16	0.44	458	14	9
5.2.2.	3y Growth	+	0.65***	*	0.33	0.21	0.47	958	14	14	0.34	0.22	0.44	453	14	9
5.3. Bank Stock Index		+	0.63**	**	0.25	0.67	0.08	1150	13	17	0.18	0.12	0.70	420	13	9
5.3.1.	1y Growth	+	0.54		0.15	0.26	0.60	1098	13	17	–0.43	0.81	0.63	420	13	9
5.3.2.	3y Growth	+	0.62**	*	0.24	0.34	0.42	997	13	16	0.17	0.69	0.14	420	13	9
5.4. Stock Market P/E Ratio		+	0.47	††	0.15	0.07	0.78	755	23	15	0.09	0.08	0.83	580	23	13
5.4.1.	1y Difference	+	0.56***		0.16	0.23	0.61	675	23	14	0.21	0.33	0.46	536	23	13
5.4.2.	3y Difference	+	0.40	††	0.00	0.00	1.00	509	19	10	–0.17	0.95	0.22	434	19	10
5.4.3.	1y Growth	+	0.55**		0.14	0.34	0.52	675	23	14	0.16	0.37	0.47	536	23	13
5.4.4.	3y Growth	+	0.41	††	0.00	0.00	1.00	509	19	10	–0.14	0.97	0.18	434	19	10
5.5. Stock Market P/B Ratio		+	0.78***	**	0.52	0.16	0.33	159	14	9	0.00	1.00	0.00	159	14	9
5.5.1.	1y Difference	+	0.76***	***	0.40	0.11	0.49	103	13	8	–0.01	1.00	0.01	103	13	8
5.6. Stock Market Dividend Yield		–	0.56		0.19	0.01	0.80	369	14	9	0.07	0.86	0.07	369	14	9
5.6.1.	1y Difference	–	0.59**	***	0.34	0.05	0.61	321	14	9	–0.02	0.99	0.03	321	14	9
5.6.2.	3y Difference	–	0.72***	***	0.43	0.07	0.49	217	14	9	0.04	0.88	0.09	217	14	9
5.6.3.	1y Growth	–	0.58**	***	0.34	0.06	0.60	321	14	9	0.00	0.96	0.04	321	14	9
5.6.4.	3y Growth	–	0.72***	***	0.45	0.06	0.49	217	14	9	0.01	0.88	0.11	217	14	9
5.7. Household Lending Spread		–	0.55		0.16	0.39	0.45	462	27	14	–0.11	0.98	0.12	462	27	14
5.7.1.	1y Difference	–	0.53		0.12	0.26	0.62	361	27	14	–0.10	0.97	0.13	361	27	14
5.7.2.	3y Difference	–	0.6		0.26	0.08	0.66	168	18	9	0.00	1.00	0.00	168	18	9
5.7.3.	1y Growth	–	0.53		0.10	0.08	0.82	361	27	14	–0.09	0.98	0.11	361	27	14
5.7.4.	3y Growth	–	0.57		0.25	0.08	0.67	168	18	9	0.00	1.00	0.00	168	18	9
5.8 Corporate Lending Spread		–	0.5		0.08	0.06	0.86	452	26	13	–0.09	0.98	0.10	452	26	13
5.8.1.	1y Difference	–	0.51		0.06	0.18	0.76	355	26	13	–0.10	0.98	0.12	355	26	13
5.8.2.	3y Difference	–	0.68**	***	0.36	0.11	0.53	168	18	9	0.00	1.00	0.00	168	18	9
5.8.3.	1y Growth	–	0.49		0.07	0.16	0.77	355	26	13	–0.10	0.98	0.12	355	26	13
5.8.4.	3y Growth	–	0.63*	**	0.36	0.11	0.53	168	18	9	0.00	1.00	0.00	168	18	9

See page 1 of this appendix for notes.

Table A6. Performance Evaluation for Measures of Potential Mispricing of Risk: Global Indicators

			Full-Sample Evaluation								Out-of-Sample Evaluation					
Indicator	**Transfor-mation**	**Sign**	**AUC**	**Logit**	**U_r**	**FNR**	**FPR**	**N**	**N_c**	**N_f**	**$U_{r,o}$**	**FNR**	**FPR**	**N**	**N_c**	**N_f**
5.9. VIX Index		–	0.71***	***	0.35	0.33	0.32	2205	28	30	0.51	0.13	0.36	947	28	16
5.9.1.	1y Difference	–	0.57**		0.20	0.30	0.51	2097	28	30	–0.11	0.46	0.64	947	28	16
5.9.2.	3y Difference	–	0.7***	***	0.30	0.28	0.41	1883	28	28	0.20	0.07	0.74	947	28	16
5.9.3.	1y Growth	–	0.58**	**	0.18	0.32	0.50	2097	28	30	–0.13	0.53	0.60	947	28	16
5.9.4.	3y Growth	–	0.71***	**	0.31	0.27	0.42	1883	28	28	0.17	0.06	0.77	947	28	16
5.9.5.	Trend Gap	–	0.71***	***	0.35	0.44	0.21	1714	28	26	0.20	0.07	0.73	947	28	16
5.9.6.	Relative Gap	–	0.71***	***	0.37	0.50	0.12	1714	28	26	0.11	0.03	0.86	947	28	16
5.9.7.	5y M.A. Gap	–	0.69***	***	0.31	0.25	0.44	1714	28	26	0.18	0.06	0.76	947	28	16
5.9.8.	Avg. Gap	–	0.75***	***	0.38	0.28	0.34	1714	28	26	0.50	0.13	0.38	947	28	16
5.10. High-Yield Spread		–	0.79***	**	0.49	0.17	0.33	1043	28	17	0.42	0.12	0.46	947	28	16
5.10.1.	1y Difference	–	0.51	**	0.23	0.05	0.72	947	28	16	0.37	0.10	0.53	947	28	16
5.10.2.	3y Difference	–	0.61***	***	0.34	0.09	0.57	745	28	15	0.50	0.23	0.27	745	28	15
5.10.3.	1y Growth	–	0.56***	***	0.19	0.04	0.77	947	28	16	0.33	0.19	0.48	947	28	16
5.10.4.	3y Growth	–	0.62***	**	0.37	0.13	0.51	745	28	15	0.50	0.23	0.27	745	28	15
5.10.5.	Trend Gap	–	0.47		0.22	0.07	0.70	565	28	15	0.00	1.00	0.00	565	28	15
5.10.6.	Relative Gap	–	0.47	***	0.22	0.07	0.70	565	28	15	0.00	1.00	0.00	565	28	15
5.10.7.	5y M.A. Gap	–	0.46		0.22	0.07	0.70	565	28	15	0.00	1.00	0.00	565	28	15
5.10.8.	Avg. Gap	–	0.48	***	0.22	0.07	0.70	565	28	15	0.00	1.00	0.00	565	28	15
5.11. German 10y Bund		–	0.68***	***	0.29	0.56	0.15	4032	28	33	0.00	0.00	1.00	947	28	16
5.11.1.	1y Difference	+	0.52		0.09	0.06	0.85	3920	28	33	0.00	0.00	1.00	947	28	16
5.11.2.	3y Difference	+	0.51		0.14	0.10	0.76	3697	28	32	–0.04	0.20	0.84	947	28	16
5.11.3.	1y Growth	+	0.52		0.10	0.65	0.25	3920	28	33	–0.09	0.32	0.77	947	28	16
5.11.4.	3y Growth	+	0.48		0.09	0.33	0.57	3697	28	32	–0.07	0.33	0.74	947	28	16

(*continued*)

Table A6. (Continued)

Indicator	Transformation	Sign	Full-Sample Evaluation: AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	Out-of-Sample Evaluation: $U_{r,o}$	FNR	FPR	N	N_c	N_f
5.12. German 1y Bill		–	0.59**		0.20	0.37	0.43	2720	28	31	0.09	0.24	0.67	947	28	16
5.12.1.	1y Difference	+	0.61**	**	0.23	0.27	0.50	2612	28	31	0.28	0.18	0.54	947	28	16
5.12.2.	3y Difference	+	0.55		0.14	0.60	0.26	2396	28	31	–0.33	0.88	0.45	947	28	16
5.12.3.	1y Growth	+	0.62**	***	0.21	0.54	0.25	2612	28	31	0.23	0.25	0.52	947	28	16
5.12.4.	3y Growth	+	0.55		0.16	0.54	0.30	2396	28	31	0.01	0.46	0.53	947	28	16
5.13. German 1m Bill		–	0.62***	**	0.23	0.36	0.42	3152	28	31	0.16	0.01	0.82	947	28	16
5.13.1.	1y Difference	+	0.55		0.17	0.35	0.47	3044	28	31	0.05	0.02	0.93	947	28	16
5.13.2.	3y Difference	+	0.52		0.12	0.30	0.58	2828	28	31	–0.27	0.72	0.56	947	28	16
5.13.3.	1y Growth	+	0.58**		0.17	0.35	0.47	3044	28	31	0.11	0.04	0.85	947	28	16
5.13.4.	3y Growth	+	0.52		0.08	0.69	0.23	2828	28	31	–0.11	0.54	0.56	947	28	16
5.14. U.S. 10y T-Note		–	0.67***	***	0.27	0.38	0.35	4032	28	33	0.00	0.00	1.00	947	28	16
5.14.1.	1y Difference	+	0.49		0.09	0.07	0.84	3920	28	33	0.00	0.00	1.00	947	28	16
5.14.2.	3y Difference	+	0.52		0.15	0.50	0.35	3697	28	32	0.05	0.01	0.93	947	28	16
5.14.3.	1y Growth	+	0.51		0.11	0.09	0.81	3920	28	33	0.22	0.05	0.73	947	28	16
5.14.4.	3y Growth	+	0.54		0.16	0.50	0.35	3697	28	32	0.14	0.05	0.81	947	28	16
5.15. U.S. 1y T-Bill		–	0.57**	*	0.22	0.32	0.47	2720	28	31	0.04	0.01	0.94	947	28	16
5.15.1.	1y Difference	+	0.63***	***	0.25	0.30	0.45	2612	28	31	0.21	0.19	0.61	947	28	16
5.15.2.	3y Difference	+	0.71***	***	0.39	0.38	0.24	2396	28	31	0.52	0.16	0.32	947	28	16
5.15.3.	1y Growth	+	0.64***	***	0.26	0.30	0.44	2612	28	31	0.36	0.22	0.42	947	28	16
5.15.4	3y Growth	+	0.72***	***	0.39	0.35	0.26	2396	28	31	0.50	0.12	0.38	947	28	16
5.16. U.S. 1m T-Bill		–	0.63***	***	0.30	0.20	0.50	3152	28	31	0.09	0.03	0.88	947	28	16
5.16.1.	1y Difference	+	0.63***	***	0.25	0.28	0.48	3044	28	31	0.23	0.06	0.71	947	28	16
5.16.2.	3y Difference	+	0.67***	***	0.35	0.37	0.28	2828	28	31	0.48	0.15	0.37	947	28	16
5.16.3.	1y Growth	+	0.65***	***	0.24	0.34	0.42	3044	28	31	0.43	0.10	0.47	947	28	16
5.16.4.	3y Growth	+	0.71***	***	0.36	0.42	0.22	2828	28	31	0.47	0.10	0.43	947	28	16

See page 1 of this appendix for notes.

Table A7. Performance Evaluation for Measures of Strength of Bank Balance Sheets

Indicator	Transformation	Sign	Full-Sample Evaluation								Out-of-Sample Evaluation					
			AUC	Logit	U_r	FNR	FPR	N	N_c	N_f	$U_{r,o}$	FNR	FPR	N	N_c	N_f
6.1. Leverage Ratio		–	0.61		0.27	0.02	0.72	696	26	14	0.01	0.79	0.20	663	26	14
6.1.1.	1y Difference	–	0.61**	*	0.21	0.46	0.33	605	26	14	0.36	0.17	0.47	605	26	14
6.1.2.	3y Difference	–	0.67***		0.33	0.16	0.50	422	24	12	–0.02	0.85	0.17	422	24	12
6.1.3.	1y Growth	–	0.62***		0.22	0.49	0.29	605	26	14	0.38	0.13	0.49	605	26	14
6.1.4.	3y Growth	–	0.67***	**	0.36	0.18	0.46	422	24	12	–0.03	0.86	0.17	422	24	12
6.2. Loans / Deposits		+	0.66**	***	0.27	0.10	0.63	515	26	13	–0.12	0.86	0.26	512	26	13
6.2.1.	1y Difference	+	0.58		0.18	0.51	0.31	422	26	13	–0.10	0.86	0.24	422	26	13
6.2.2.	3y Difference	+	0.58		0.15	0.67	0.17	222	21	11	–0.14	0.79	0.34	222	21	11
6.2.3.	1y Growth	+	0.54		0.11	0.01	0.88	422	26	13	–0.10	0.85	0.25	422	26	13
6.2.4.	3y Growth	+	0.51		0.15	0.01	0.84	222	21	11	–0.26	0.79	0.47	222	21	11
6.3. Total Assets / GDP		+	0.55		0.14	0.45	0.42	739	21	14	0.05	0.39	0.56	613	21	12
6.3.1.	1y Difference	+	0.64**		0.22	0.26	0.52	658	21	13	0.22	0.41	0.37	589	21	12
6.3.2.	3y Difference	+	0.57		0.19	0.53	0.28	507	19	11	0.18	0.63	0.19	504	19	11
6.3.3.	1y Growth	+	0.68***	***	0.33	0.22	0.45	658	21	13	0.36	0.17	0.48	589	21	12
6.3.4.	3y Growth	+	0.64*	**	0.23	0.50	0.27	507	19	11	0.24	0.59	0.17	504	19	11
6.4. Non-core Liabilities / GDP		+	0.48		0.10	0.11	0.79	405	20	10	–0.10	0.98	0.12	405	20	10
6.4.1.	1y Difference	+	0.52		0.12	0.58	0.30	334	19	9	–0.07	0.97	0.10	334	19	9
6.4.2.	3y Difference	+	0.5		0.14	0.12	0.74	189	17	8	–0.31	1.00	0.31	189	17	8
6.4.3.	1y Growth	+	0.56		0.19	0.36	0.45	334	19	9	–0.18	0.97	0.21	334	19	9
6.4.4.	3y Growth	+	0.5		0.17	0.72	0.11	189	17	8	–0.25	1.00	0.25	189	17	8
6.5. Non-core Liabilities / Total Assets		+	0.57		0.19	0.40	0.41	475	25	12	–0.10	0.98	0.12	475	25	12
6.5.1.	1y Difference	+	0.49		0.06	0.91	0.03	386	24	11	–0.15	0.94	0.21	386	24	11
6.5.2.	3y Difference	+	0.5		0.13	0.75	0.12	202	20	10	–0.12	0.95	0.17	202	20	10
6.5.3.	1y Growth	+	0.49		0.06	0.94	0.00	386	24	11	–0.14	0.94	0.19	386	24	11
6.5.4.	3y Growth	+	0.48		0.07	0.91	0.02	202	20	10	–0.16	0.95	0.20	202	20	10
6.6. (ST Liabilities – Liquid Assets) / Total Assets		+	0.73**	*	0.37	0.60	0.03	286	13	7	–0.02	1.00	0.02	286	13	7
6.6.1.	1y Difference	+	0.70**		0.36	0.28	0.36	242	13	7	–0.05	1.00	0.05	242	13	7
6.6.2.	3y Difference	+	0.81**	**	0.49	0.21	0.29	138	12	6	–0.07	1.00	0.07	138	12	6
6.6.3.	1y Growth	+	0.70**	**	0.35	0.28	0.37	242	13	7	–0.08	0.98	0.10	242	13	7
6.6.4.	3y Growth	+	0.67**		0.35	0.19	0.46	138	12	6	–0.09	1.00	0.09	138	12	6
6.7. Short-Term Liabilities / Liquid Assets		+	0.77**	**	0.52	0.37	0.11	286	13	7	–0.01	0.98	0.03	286	13	7
6.7.1.	1y Difference	+	0.67**		0.37	0.28	0.35	242	13	7	–0.08	0.98	0.10	242	13	7
6.7.2.	3y Difference	+	0.86**	*	0.72	0.17	0.11	138	12	6	–0.05	1.00	0.05	138	12	6
6.7.3.	1y Growth	+	0.67**		0.33	0.35	0.32	242	13	7	–0.09	0.98	0.10	242	13	7
6.7.4.	3y Growth	+	0.86***	*	0.70	0.24	0.06	138	12	6	–0.04	1.00	0.04	138	12	6

See page 1 of this appendix for notes.

Article II

Tölö, E., 2020, "Predicting systemic financial crises with recurrent neural networks," Journal of Financial Stability, Vol. 49, 100746.

Available at: https://doi.org/10.1016/j.jfs.2020.100746.

Journal of Financial Stability 49 (2020) 100746

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfstabil

Predicting systemic financial crises with recurrent neural networks

Eero Tölö[a,b,c,*]

[a] Helsinki GSE, University of Helsinki, P.O. Box 17, 00014, Finland
[b] Department of Economics, London School of Economics and Political Science, Houghton Street, WC2A 2AE, London, United Kingdom
[c] Department of Financial Stability, Bank of Finland, P.O. Box 160, FI-00101, Helsinki, Finland

ARTICLE INFO

Article history:
Received 30 August 2019
Received in revised form 21 April 2020
Accepted 22 April 2020
Available online 31 May 2020

JEL classification:
G21
C45
C52

Keywords:
Early warning system
Systemic Banking crises
Neural networks
Validation

ABSTRACT

We consider predicting systemic financial crises one to five years ahead using recurrent neural networks. We evaluate the prediction performance with the Jórda-Schularick-Taylor dataset, which includes the crisis dates and annual macroeconomic series of 17 countries over the period 1870–2016. Previous literature has found that simple neural net architectures are useful and outperform the traditional logistic regression model in predicting systemic financial crises. We show that such predictions can be significantly improved by making use of the Long-Short Term Memory (RNN-LSTM) and the Gated Recurrent Unit (RNN-GRU) neural nets. Behind the success is the recurrent networks' ability to make more robust predictions from the time series data. The results remain robust after extensive sensitivity analysis.

1. Introduction

In their article, Schularick and Taylor (2012) describe two eras of finance capitalism. The latter period started around the mid-1900s and is characterized by unprecedented financial risk and leverage, where credit flows freely, and activist macroeconomic policies react when booms go bust. Following the 2008 financial crisis, proactive financial stability policies have gained ground, as authorities implement such policies to counteract the build-up of systemic risk. For evidence, see the growing list of active policies listed by the European Systemic Risk Board.[1] An example is the countercyclical capital buffer (Basel Committee on Banking Supervision (BCBS), 2011), an additional capital requirement enforced on banks when the authorities assess the credit growth as excessive. Timing and planning of financial stability policies require a timely view of the associated risks. If the authorities decide to curb lending at the wrong time, economic growth can be harmed without a commensurate benefit. Financial crisis prediction models help in timing policies by providing information about the likelihood of a crisis.

In this article, we investigate how well various types of artificial neural networks (ANNs) predict systemic financial crises using time series information. When we say a systemic financial crisis, we essentially mean a systemic banking crisis.[2] We are motivated by the developments in the ability of deep neural nets to handle sequential data that have taken place during the last few decades. From the econometric perspective (see Kuan and White, 2007), ANNs serve as universal function approximations, which avoid the problem of model misspecification, unlike parametric models commonly used in empirical studies. This should be useful for financial crisis prediction because the crises are fundamentally non-linear phenomena. We consider alternative neural net architectures and benchmark them against the logistic regression model, a standard model used in policy institutions at the time of writing (see e.g. Lang et al., 2019). Basic neural net models have often been looked down upon by economists for their lack of interpretability. Thanks to recent advances, drivers of neural net predictions can now be decomposed and understood on par with other econometric models.

* Correspondence to: Helsinki GSE, University of Helsinki, P.O. Box 17, 00014, Finland.
E-mail address: eero.tolo@helsinki.fi

[1] www.esrb.europa.eu.

[2] Following Schularick and Taylor (2012), we define systemic financial crises "as events during which a country's banking sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions."

https://doi.org/10.1016/j.jfs.2020.100746
1572-3089/

We postpone the discussion of existing financial crisis prediction literature to the next section. At this point, it suffices to know that these models usually make predictions based on a cross-section of macro-financial variables. The research agrees that neural nets outperform the logistic model in systemic financial crisis prediction in cross-validation. There remains some controversy about whether the benefits carry out to sequential out-of-sample evaluation that only uses information known at the time of forecasting.

We argue that there are gains in considering information in the economic time series beyond the latest cross-section. This should not be surprising since we understand that economic indicators in general exhibit rich dynamics. Moreover, in the context of crisis prediction, we have seen that each early warning indicator shows a distinct time pattern (see for example Drehmann and Juselius, 2014, or Tölö et al., 2018). While there are reasons to avoid multiple time lags in the context of logistic regression (mainly collinearity), for neural nets, the time series dimension presents an untapped source of information, whose potential remains so far mostly uncharted in the context systemic financial crisis prediction.

When it comes to choosing a neural net architecture for a multivariate time series prediction problem, we have some primary candidates. The so-called multilayer perceptron (MLP) is the most basic form of neural net that can produce universal function approximations. Compared to other non-parametric methods, such as Fourier or polynomial expansions, the MLP typically requires a lower number of components and is thus a more parsimonious approach (Barron, 1993). Earlier research has shown that MLPs can give accurate crisis predictions based on a cross-section of observations (Holopainen and Sarlin, 2017; Ristolainen, 2018; Bluwstein et al., 2020). Contrasting evidence is provided by Beutel et al. (2019), who find that the logit model outperforms basic machine learning methods, including the MLP in out-of-sample prediction in a sample of 15 countries. In any case, MLPs can be applied to time series using a time-window approach (Gers et al., 2001).

For sequence problems, there also exists a class of models called recurrent neural networks (RNN) that are designed for modeling temporal dynamic behavior. Binner et al. (2004, 2006) show that simple RNNs produce forecasts comparable to traditional Markov switching models. Since then, the RNNs have evolved to include gating mechanisms that allow them to retain a past event in memory for an extended time. The modern RNNs with Long-Short Term Memory (LSTM; Hochreiter and Schmidhuber, 1997) and Gated Recurrent Units (GRU; Cho et al., 2014) are state-of-the-art techniques for sequence learning. RNN-LSTMs have been used with great success for all kinds of sequence problems: text and speech recognition (e.g., Wu et al., 2016), video recognition, acoustic models, traffic, weather patterns, etc. but much less with economic data.[3] RNN-LSTMs have also been used successfully for demand forecasting and financial market predictions (e.g., Fischer and Krauss, 2018; Borovkova and Tsiamas, 2019; Minami, 2018). Siami-Namini et al. (2018) find RNN-LSTMs outperform ARIMA in time series forecasting. Cook and Smalter Hall (2017) improve forecasts of macroeconomic indicators with an LSTM and other neural nets.

Resolving whether the RNNs help in a systemic financial crisis prediction requires us to take the problem to data. This study uses the Jordá-Schularick-Taylor Macrohistory database (Jordà et al., 2017), which consists of annual time series for 17 economies over the period 1870–2016 and includes 66 systemic banking crises in the first era of finance capitalism (1870–1939) and 24 in the second era (1946–2016). The relatively small amount of data speaks towards less complicated models with fewer parameters. Therefore, we consider MLPs with one hidden layer and non-stacked RNNs. We include a small number of explanatory variables: loans-to-GDP, house prices and stock prices, current account ratio, and real GDP. The dependent variable is a zero/one dummy with 1 indicating a systemic financial crisis at a specific prediction horizon.[4] After establishing that the neural nets produce well-calibrated crisis probabilities, we proceed to test their prediction performance using the Area Under Curve (AUC) statistics.

First, we benchmark the MLP, RNN, and gated RNN neural nets against the logit model in a one-year ahead prediction. We evaluate the models using country-by-country cross-validation and sequential out-of-sample tests in various subsamples motivated by the two eras of finance capitalism and the end of Bretton-Woods. In the cross-validation, we estimate the model for all but one country and evaluate the model for the remaining country, such that we test the model for each country in turn. In the sequential evaluation, we estimate the model for an earlier time-period and test the model for a later time period. We find that the RNN neural nets, especially the gated RNNs, consistently outperform both the MLP neural nets and the logit model in all subsamples and evaluations. We attribute the performance advantage to three sources. First, including multiple lagged predictors (i.e. the time-window) adds useful information that benefits even the logit model in the cross-validation. Second, given the limited amount of data for estimation, the RNNs can make the best use of this information based on their temporal structure. Third, the gating mechanism in gated RNNs helps in extracting information from the time series.

The rest of the analysis focuses on the more recent 1970–2016 subsample. We consider forecast horizons of up to 5 years. For both cross-validation and sequential evaluation, we find that the RNNs, especially the gated RNNs, outperform the MLP and the logit model also at longer forecast horizons. We focus on the LSTM and demonstrate that it produces coherent predictions such that, on average, the predicted probability of crisis peaks at the intended distance to a crisis. Finally, we investigate which explanatory variables are mainly responsible for the performance improvements of the LSTM over the logit model. For this purpose, we analyze the prediction model using subsets of explanatory variables. We show that stock prices actively drive the cross-validated predictions, but also other variables contribute. The sequential predictions are driven more equally by the different predictors.

Overall, the improvement in prediction accuracy from the neural nets is substantial. For example, considering the three-year ahead forecast, when we fix the sensitivity such that the models detect more than 80% of the crises correctly, the LSTM produces less than half the amounts of false alarms in comparison to the logit model (about 20% vs. more than 40%). In the Appendix A, we report a sensitivity analysis. The results are found to be robust for changes in the number of neurons in the hidden units and the length of the time-window.

We contribute to the growing literature of using machine learning methods in crisis prediction (see Section 2 for a review). Our main contribution is to demonstrate that utilizing the time series information via gated RNNs improves the systemic financial crisis predictions. The consistent performance advantage in the sequential evaluation with different subsamples suggests that the gated RNNs are likely to provide out-of-sample predictions that are more robust than generally. Based on our findings, we expect similar techniques to be useful in predicting other related events such as

[3] Of course, some economic information comes in text form, so language models and economics are not mutually exclusive, see Apel et al. (2019) for an example.

[4] Treating each systemic financial crisis as a similar dependent dummy irrespective of the depth of the crisis and other characteristics fades the non-linear characteristics and likely favours the linear logistic regression.

recessions and currency crises. Exploratory results with neural nets could also help in devising better non-linear econometric models.

The rest of the article is organized as follows. Section 2 offers a survey of financial crisis prediction models. Section 3 presents the dataset. Section 4 reviews the neural nets considered in this study, MLPs and RNNs, respectively. Section 5 discusses the performance evaluation framework, which consists of defining the dependent variable for the forecast problem (Section 5.1), the validation and out-of-sample tests (Section 5.2), and performance measurement (Section 5.3). Section 6 presents the results. Section 7 concludes with discussion. Appendix A discusses in detail the neural net models and their estimation, and Appendix B presents the sensitivity analysis.

2. Early warning models for financial crisis prediction

Despite the long history of financial crises, early warning models based on cross-country panel data sets are a relatively new practice that emerged in the late 1990s through pioneering work including Demirgüc-Kunt and Detragiache (1998), (2000), Kaminsky and Reinhart (1999), Hardy and Pazarbasioglu (1999), Caprio and Klingebiel (1997), and Berg and Pattillo (1999). These articles collected datasets of crisis dates and investigated which macro-financial quantities were useful in predicting those events. In those days, the two standard methods were the so-called signaling method and the logit model. The former takes one time series (such as annual credit growth or some other risk measure), and a value that goes beyond a threshold is considered a warning signal. The logit model explains the crisis dummy (or pre-crisis dummy) directly using a set of variables. Numerous articles followed that investigated indicators that precede financial crises.[5] Following the 2008 episode, there was again a spur of interest, and datasets were extended (see Reinhart and Rogoff, 2009; Schularick and Taylor, 2012; Laeven and Valencia, 2012; Babecký et al., 2014; Detken et al., 2014). The findings from this literature are actively used in policy institutions. Tölö et al. (2018) include a convenient summary table of early warning indicators used in a broad set of studies, and whether they were significant in those studies: Numerous studies find that private sector loans and loans/GDP are significant banking crisis predictors. In fact, they are often the most robust predictor, although the result is somewhat dependent on the sample of countries and other characteristics of the dataset. A debt service burden indicator has been available in a smaller number of studies but is generally found to be informative. House prices, stock prices, and credit spreads are also found to be good predictors and have been included in many studies. Current account/GDP has been a significant predictor in many studies, but its performance has been somewhat less consistent. Various other measures of external imbalances tend to be significant about half the time that they have been included in published manuscripts. Variables related to real economy such as GDP, investment, and unemployment are occasionally significant in the prediction models, but never the strongest predictors. Other characteristics such as income inequality, fixed exchange rate, deregulation, and contagion have also been found to play a role.

Recent literature has already moved beyond the logit model; the new approaches include multinomial logistic model and machine learning methods. Caggiano et al. (2014), (2016) and (for currency crises) Bussiere and Fratzscher (2006) show that the multinomial logistic model outperforms the usual binomial logit model. Other articles seek to replace the logit model with machine learning methods such as decision trees, neural nets, random forests, support vector machines, and other classification methods.[6] Early examples are Duttagupta and Cashin (2011); Díaz-Martínez et al. (2011); Davis and Liadze (2011), and Manasse et al. (2013), who all employ classification trees. In these earlier studies, the focus is on identifying nonlinear rules for characterizing pre-crisis developments. For example, Duttagupta and Cashin (2011) find that a specific binomial tree condition that combines low bank profitability with modest export growth significantly increases the probability of a banking crisis in emerging/developing countries.

A fundamental question is whether the machine learning methods can robustly outperform the basic traditional econometric models given the limitations of the data. So far, the logit model has been a standard benchmark. Using a common reference increases comparability because, due to differences in datasets, the attained performance measures can't usually be compared directly with each other. A number of studies find that machine learning methods outperform the logit model: Joy et al. (2017) and Alessi and Detken (2018) use classification trees and random forests, Ristolainen (2018) artificial neural net, Casabianca et al. (2019) adaptive boosting (AdaBoost), Fouliard et al. (2019) model averaging; Holopainen and Sarlin (2017) and Bluwstein et al. (2020) benchmark various machine learning models against the logit model. In these two comparison studies, the MLP neural net performs roughly as well as the top-ranking methods. In contrast to these studies, Beutel et al. (2019) find no improvement for a decision tree, random forest, KNN, SVM, or MLP neural net in comparison to the logit model in sequential out-of-sample prediction. In the analysis of Beutel et al. (2019), the MLP neural net still comes closest to the logit model in terms of performance, and they call for further research with more sophisticated neural nets. Fricke (2017) also finds that various machine learning methods do not bring benefits when compared to the logit model in a sequential evaluation with univariate data.

Related work includes Qi (2001), Gogas et al. (2014), and Nyman and Ormerod (2017), who predict recessions using machine learning, and Fioramanti (2008) and Manasse and Roubini (2009) who predict sovereign debt crises. Suss and Treitel (2019) forecast bank distress in the UK with random forests. Nik et al. (2016) predict financial crises in emerging economies using neural nets.

A dominating practice in the above literature is that while the studies are based on panel data (multivariate time series for each country), the predictions are almost exclusively based on only a cross-section of variables. A few studies do consider the lag structure to a limited extent. While the early work often considered contemporaneous crisis determinants, the analysis soon turned to lagged predictors to avoid simultaneity issues. Importantly, we do not need to worry about endogeneity when we do prediction based on lagged variables, which are not influenced by the predicted event. Demirgüc-Kunt and Detragiache (1998) note that credit growth is significant if lagged by two periods. They also note that lagged GDP growth loses significance indicating that either the contemporaneous GDP causes the banking crisis very quickly or is itself caused by the crisis. Hardy and Pazarbasioglu (1999) note that the lag structure provides a rough indication of the distance to a crisis. Davis and Karim (2008) choose the optimal lag structure based on a grid search and conclude that transforming indicators and using lags and interaction terms improve the per-

[5] The list is long. Some examples are Alessi and Detken, 2011; Barrell et al., 2011; Bordo and Meissner, 2012; Borio and Lowe, 2002; Borio and Drehmann, 2009; Büyükkarabacak and Valev, 2010; Davis and Karim, 2008; Domaç and Martinez Peria, 2003; Drehmann and Juselius, 2014; Lo Duca and Peltonen, 2013; Behn et al., 2013; Jordà et al., 2015; Roy and Kemme, 2012; Kauko, 2012, 2014; Tölö et al., 2018; von Hagen and Ho, 2007.

[6] The multistate approach in Caggiano et al. (2014), (2016) and Bussiere and Fratzscher (2006) is not mutually exclusive with the machine learning methods. An example is Sarlin (2014), who considers four states of financial stability (normal, pre-crisis, crisis, post-crisis) for a visualization application.

formance of the early warning model. Indeed, a common approach in the reviewed literature has been to extract information from the time series through specific transformations such as two, three, or four-year growth, or extracting a cyclical component based on beliefs about the length of the financial cycle (cf. Drehmann et al., 2010; Kauko and Tölö, 2020a, 2020b). If the choice is based on full-sample information, that may have implications for out-of-sample predictions, however. Borio and Drehmann (2009) consider fine-tuning the lag structure of credit and equities such that they peak at the intended distance to the crisis. Schularick and Taylor (2012) and Bordo and Meissner (2012) consider five lags of credit growth and note that the first two lags have opposite signs. Fricke (2017) uses the same univariate five-lag specification in machine learning models. Drehmann and Juselius (2014) compare the performance of the credit-to-GDP gap and the debt-service ratio in banking crisis prediction and conclude that the former dominates at longer horizons and the latter at shorter horizons. Tölö et al. (2018) include a summary table that shows the prediction horizons for which different early-warning indicators were informative in the EU countries.

In summary, the literature has investigated which specific lags of some variables are informative. However, it has not really considered whether the information in the time series as a whole could be utilized in predictions using advanced techniques. An exception is Virtanen et al. (2018), who predict financial crises with unit root tests in an expanding window. Current article seeks to address this gap in the literature.

3. Data

All data for this study come from the Jórda-Schularick-Taylor macro history database (Jordà et al., 2017; and Knoll et al., 2016). The dataset includes 17 countries: Australia, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, UK, Italy, Japan, Netherlands, Norway, Portugal, Sweden, and the USA.

The dependent variable that we predict is based on the systemic financial crisis dummy variable, which takes value 1 for the year that marks the start of a systemic financial crisis in a given country. This pre-crisis dummy we discuss in detail in Section 5.1. The dataset classifies systemic financial crises as "events during which a country's banking sector experiences bank runs, sharp increases in default rates accompanied by large losses of capital that result in public intervention, bankruptcy, or forced merger of financial institutions" (Schularick and Taylor, 2012). Financial crises were fairly common in the first era of finance capitalism 1870–1939. The dataset lists 66 crises in this era (though all variables are not available for some of the old crises). The WW2 was followed by a relatively long absence of systemic financial crises until they started again starting from the 70s. The dataset includes 24 financial crises after WW2. The crisis pattern is associated with different economic policy regimes, as discussed more extensively in Schularick and Taylor (2012).

We consider up to five explanatory variables for the systemic financial crisis prediction model. The variables are

1. Loans to non-financial private sector divided by GDP, 1-year growth (*abbr.* l/gdp)
2. Real stock prices, 1-year growth (*abbr.* rsp)
3. Real house prices, 1-year growth (*abbr.* rhp)
4. Current account-to-GDP ratio, level (*abbr.* ca/gdp)
5. Real GDP, 1-year growth (*abbr.* gdp)

Note that we imposed the one-year growth to make the data stationary and comparable between countries. All the variables can potentially contain some relevant information that helps predict systemic banking crises. The five variables were chosen based

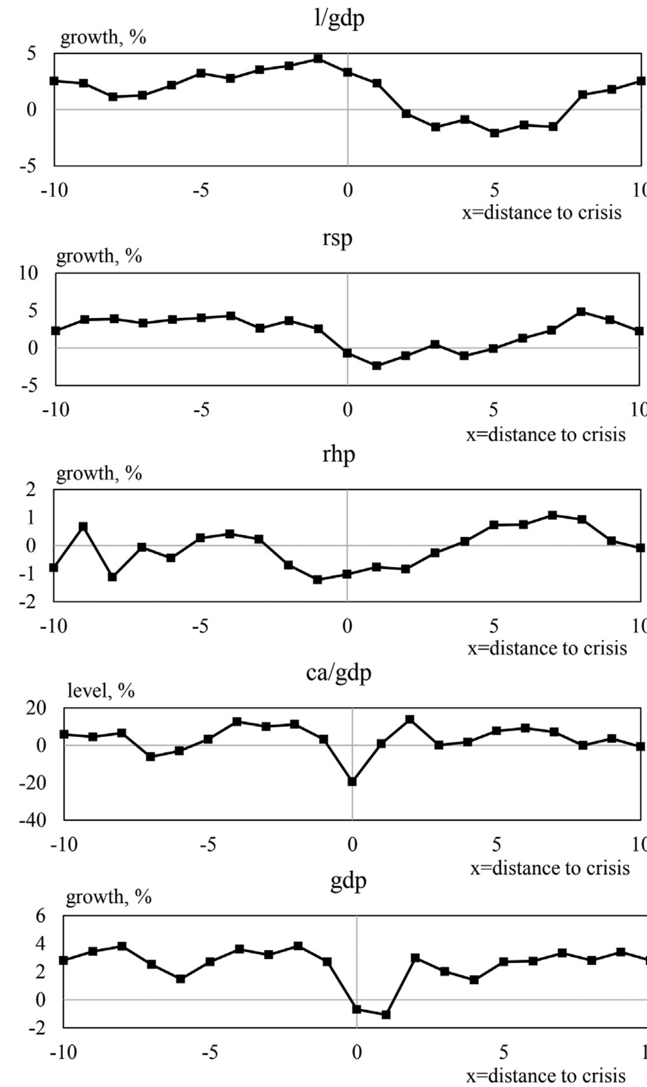


Fig. 1. Development of predictors around systemic financial crisis dates. The horizontal axis shows the distance in years to the systemic financial crisis that starts at x = 0.

on their generality and the fact that they describe distinct economic developments. As depicted in Fig. 1, the variables exhibit rich dynamics around systemic financial crises. Table 1 presents descriptive statistics for the variables. Based on the summary table in Tölö et al. (2018), which summarizes many of the early warning indicator studies in the literature, we expect loans/GDP, stock prices, and house prices to be the strongest predictors. Also, the current account-to-GDP ratio and real GDP have been found informative in some studies, but the evidence is not as strong.

In practice, apart from the stock prices, the explanatory variables have some information lags. Studies with quarterly data typically use one quarter publication lag. Like other studies with annual data, we do not introduce publication lags (see e.g. Borio and Lowe, 2002).

4. Neural nets for time series prediction

Artificial neural networks (ANNs), neural nets for short, are a class of nonlinear models that bear a resemblance to the biological neural structure. As discussed in the introduction, ANNs are interesting because they can provide parsimonious non-linear function approximations. ANNs typically consist of layers of nodes that are connected to subsequent layers of nodes through non-linear

Table 1
Descriptive statistics for predictors for the period 1870–2016.

Variable	Mean	Median	Std.	10th percentile	90th percentile	Observations
l/gdp	1.84	1.56	7.08	−5.17	8.63	1542
rsp	2.43	1.86	9.51	−7.26	11.70	1542
rhp	−0.06	−0.05	4.03	−4.36	4.39	1542
ca/gdp	3.84	3.41	19.64	−19.81	27.82	1542
gdp	3.16	3.07	4.48	−1.50	7.63	1542

Units are one-year percentage growth except for ca/gdp, which is a percentage level.

activation functions. In econometric terms, the nodes in the first layer present explanatory variables. The nodes in the middle layer present intermediate transformations of the explanatory variables. The node(s) in the final layer presents the predicted dependent variable. Each function is associated with a set of parameters called weights and biases. Training the neural net means estimating these parameters by minimizing a loss function that depends on the predicted dependent variable and its true values. Because neural nets typically have a large number of parameters and easily achieve close to perfect in-sample fit, it is crucial to estimate the model with one sample and test it with a different sample. In the following, we shortly describe the structure of the neural nets used in this study. Details of the models and their estimation are provided in Appendix A.

A feedforward neural network is the earliest and most straightforward type of artificial neural net. In this model, the explanatory variables are fed into the neural net in the input layer. They are transformed through activations functions as they pass through the net until they reach the output layer. Today, the most common example of a feedforward neural network is a multilayer perceptron (MLP). The MLP is characterized by all nodes in adjacent layers being connected with each other (see Appendix A.1 for details). The fundamental architectural question for the MLP is the number of hidden layers and the number of nodes in each hidden layer. According to a universal approximation theorem for neural nets (Hornik, 1991), every continuous function on a bounded domain can be approximated with an MLP with just one hidden layer. The problem is that the required size for such a network can be impractically large, making the system prone to overfitting. In large scale problems, empirical evidence suggests that depth can be beneficial. However, the systemic financial crisis is not a large-scale problem (in terms of data), and following the literature, we only include one hidden layer in the MLP.

We consider two alternative MLPs: MLP(1) and MLP(τ). MLP(1) makes predictions based on a cross-section of variables at time t. MLP(τ) makes predictions based on τ latest observations. The MLP sees all the predictor data within the time-window simultaneously. This can lead to predictions that do not appreciate the conceptual difference between current and past data, and the predictions might not generalize well outside the sample. In the following, we discuss a class of neural nets that seek to avoid this problem.

Recurrent neural networks (RNNs) are a family of neural nets designed for sequential data such as language processing and time series. An RNN takes in a time series of explanatory variables, which allows the RNN to use its hidden states (akin to memory) dynamically to process a sequence of input data. Importantly, the RNN preserves the temporal ordering of the time series, which can help reduce overfitting in an application where the temporal ordering is relevant. Conceptually, these two properties could make the RNNs ideal for systemic financial crisis prediction. By observing the time series, the RNN would identify when vulnerabilities are building up. The RNN would remember the accumulated level of vulnerability. Later, when a triggering event takes place (such as a domestic slowdown of the economy or an international shock), it would signal the alarm.

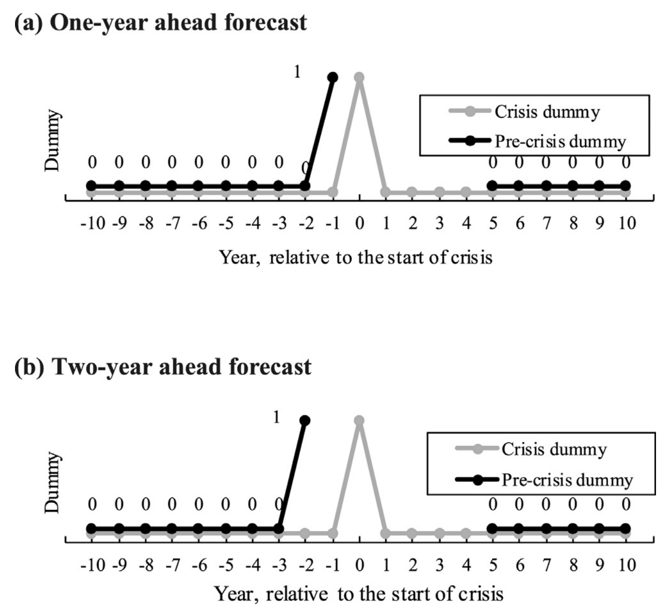


Fig. 2. Illustrations of the pre-crisis dummies around crisis dates.

In this study, we consider three different RNN architectures: basic RNN, RNN with Long-Short Term Memory (LSTM) cells, and RNN with GRU cells. Basic RNNs have few parameters but are sometimes susceptible to an exploding or vanishing gradient problem. LSTMs and GRUs avoid the stability problems of the basic RNN by using gating mechanisms (see Appendices A.3 and A.4 for details). That is why we refer to them together as gated RNNs. The gated RNNs have a great capacity to remember past developments in the time series.

The prediction models are summarized in Table 2. For MLPs, we set the number of nodes in the hidden-layer equal to 10. Similarly, for the RNNs, we set the dimension of the hidden recurrent state to 10. For the time-window MLP and the RNN nets, we use a time-window of 5 years, which corresponds to the five-lag specification used by Schularick and Taylor (2012), Bordo and Meissner (2012), and Fricke (2017). In principle, all these hyperparameters could be optimized using a separate validation. Since we will consider many different neural nets, estimation samples, and test setups, we prefer to calculate all the results with the same reasonable parameter values outlined above. Afterward, we confirm with robustness checks that the specific values of the hyperparameters are not important (see Appendix B).

5. Performance evaluation framework

5.1. Dependent variable

We aim to predict systemic financial crisis events. Hence, the dependent variable is the pre-crisis dummy related to a specific forecast horizon. This is best explained using an illustration. Fig. 2(a) shows the crisis dummy and the pre-crisis dummy for

Table 2
Summary of the prediction models.

Abbreviation	Model name	Hyperparameters	d	k
Logit(1)	Logistic regression model	none	5	6
Logit(5)	Logistic regression model	none	25	26
MLP(1)	Multilayer perceptron	hidden layers = 1, units in hidden layer = 10, L2 weight = 0.001	5	71
MLP(5)	Multilayer perceptron	hidden layers = 1, units in hidden layer = 10, L2 weight = 0.01	25	271
RNN	Recurrent Neural Network	time steps = 5, dimension of hidden state = 10, L2 weight = 0.001	25	171
RNN-LSTM	Long Short-Term Memory Recurrent Neural Network	time steps = 5, dimension of hidden state = 10, L2 weight = 0.001	25	691
RNN-GRU	Gated Recurrent Unit Recurrent Neural Network	time steps = 5, dimension of hidden state = 10, L2 weight = 0.001	25	521

d = number of explanatory variables.
k = number of parameters.

a one-year ahead forecast around the crisis event. Crisis dummy takes value 1 for the year that marks the start of the systemic financial crisis. The pre-crisis dummy is obtained by shifting the crisis dummy backward in time (according to the forecast horizon) and by removing observations corresponding to crisis and post-crisis periods. Excluding these years is a common approach in the literature and helps alleviate the post-crisis bias (see e.g. Drehmann and Juselius, 2014).

Fig. 2(b) illustrates the crisis dummy and the pre-crisis dummy for a two-year ahead forecast. In this case, in addition to the crisis and post-crisis years, we exclude the one year before the crisis. Conceptually this means that we do not care what the model predicts at the excluded observation. This handling of the pre-crisis and post-crisis periods is largely similar to other sources (see Drehmann and Juselius, 2014; Detken et al., 2014; Holopainen and Sarlin, 2017; or Ristolainen, 2018). As Ristolainen (2018) points out, the pre-crisis dummy is often set equal to one for multiple pre-crisis periods to avoid considering multiple lagged predictors.

5.2. Cross-validation and sequential out-of-sample evaluation

We consider two alternative out-of-sample performance evaluation frameworks: country-by-country cross-validation and sequential evaluation.

5.2.1. Country-by-country cross-validation

In the country-by-country cross-validation, we exclude each country in turn, estimate the network model, and then perform the out-of-sample prediction for each year for the country that was excluded.[7] Then we pool all the out-of-sample predictions together and evaluate the AUC statistics. The country-by-country cross-validation does not fully preserve temporal ordering in the sense that the full information about the other countries is used to estimate the network. This information from other countries is, of course, only limited to the estimation phase. For example, when the neural net makes its prediction for the UK in 2006, it does not explicitly know that a bunch of other countries is going to have a crisis in a few years. Also, because we do not use any information from the predicted country in the estimation, the test can be considered quite robust.

```
Cross-validation algorithm pseudocode:
[1]: Loop C over countries:
[2]:         Estimate the model excluding the data of country C
[3]:         Test the model using data of country C only
[4]:         Store the predicted probabilities for country C.
[5]: end loop
[6]: Calculate the out-of-sample AUC by pooling the predicted probabilities.
```

5.2.2. Sequential out-of-sample evaluation

In the sequential evaluation, we split the sample into two parts. The earlier part is used for estimating the parameters, and the latter part is used for testing. In this case, the temporal structure is fully preserved as we do not use any future information for prediction. We consider four alternative sample splits, as discussed further in the results. Similar sequential evaluation setups are used by Fricke (2017), Beutel et al. (2019), Bluwstein et al. (2020), Holopainen and Sarlin (2017), and Alessi and Detken (2018).

5.3. Performance measurement

We want to evaluate our prediction models using the area under the ROC curve (AUC), which is a common measure used in the systemic financial crisis prediction literature (for further discussion, see Drehmann and Juselius, 2014). As we will see in a moment, computing AUC makes sense because our prediction models output something that can be interpreted as a probability estimate.

Each of the models outputs a value that lies between 0 and 1. If the output is larger than some threshold h, then we say that the model predicts a crisis at the given forecast horizon. Otherwise, the prediction is that there won't be a crisis. Correctly predicted crisis is labeled a true positive (TP). Correctly predicted normal state is labeled a true negative (TN). A false alarm is labeled a false positive (FP), and a missed crisis is labeled a false negative (FN).

Sensitivity is defined as TP/(TP + FN), and specificity is defined as TN/(TN + FP). If we plot sensitivity vs. 1-specificity for all possible threshold values h, we get the receiver operating characteristic (ROC) curve. The area under the ROC curve is an approximately proper scoring rule for the classification task. Higher AUC corresponds to better predictions. The maximum value of the AUC, one, is achieved for a perfect model that can distinguish the two states perfectly. AUC = 0.5 corresponds to a random guess. AUC below 0.5 means that the predictions are worse than a random guess.

Let us now illustrate that the outputs from the prediction models correspond to probabilities. Earlier, Niculescu-Mizin and Caruana (2005) have demonstrated that simple neural net models produce well-calibrated probabilities, while many other machine learning methods do not.

[7] Country-by-country cross-validation is often replaced with random k-fold cross-validation. Note that k-fold cross-validation is not applicable in our setup because we would inevitably end up having overlapping observations in different folds.

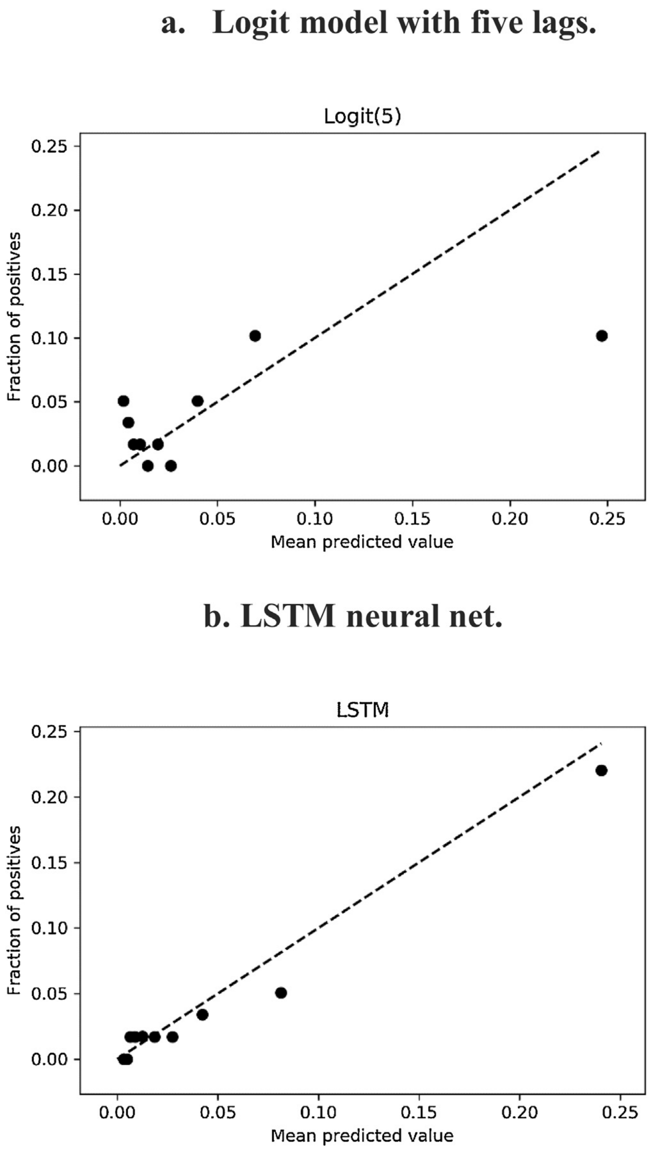


Fig. 3. Reliability diagrams. The horizontal axis shows the mean predicted value for each quantile bucket. The vertical axis shows the fraction of positives (pre-crisis observations) in the corresponding bucket. Both graphs correspond to country-by-country cross-validation in the 1970-2016 subsample. Well-calibrated probabilities should scatter close to the 45-degree line (dashed line).

The relationship between the prediction model output and probabilities can be investigated via reliability diagrams (DeGroot and Fienberg, 1983; Niculescu-Mizin and Caruana, 2005). At first, the sample is ordered according to the predicted probabilities and split into ten bins. To get an equal number of observations in each bin, we use quantile bins.[8] For each bin, we calculate the true fraction of the positive states and plot it as a function of the mean value of the bin. If the probabilities are well-calibrated, the points should scatter around a 45-degree line.

We calculate reliability diagrams for the one-year ahead crisis prediction models using the country-by-country cross-validation for the period 1970–2016. Fig. 3(a) and (b) presents the reliability diagrams for the logit model and the LSTM neural net. The probabilities produced by the LSTM are scattered close to the 45-degree line, so they are well-calibrated. The probabilities produced by the logit model do not seem as reliable, but this is only because the logit model has lower explanatory power.

6. Results

6.1. Benchmark evaluation

In the following, we evaluate the prediction performance of neural net models by benchmarking against a logistic model with a single lag and multiple (5) lags, denoted Logit(1) and Logit(5), respectively. Recall the prediction models summarized in Table 2. In the first stage, we evaluate the models using a one-year forecast horizon and different subsamples. In the second stage, we extend the analysis to forecast horizons of up to five years. In both stages, we first consider performance in cross-validation and then in sequential out-of-sample evaluation.

Recall that in our cross-validation setup, each country, in turn, is used as a test sample while the other countries are used for estimating the model. We consider four alternative subsamples: (1) 1870–2016 (43 crises), (2) 1870–1939 (19 crises), (3) 1946–2016 (24 crises), and (4) 1970–2016 (23 crises). The subsamples are motivated by the first and second era of finance capitalism (Schularick and Taylor, 2012), and the end of the Bretton-Woods era in 1971.

The cross-validated AUC statistics are shown in Table 3(a), where each column corresponds to a different subsample. The obtained AUCs lie between 0.597 and 0.867. The corresponding standard errors adjusted for clustering at the country level range from 0.027 to 0.095. The standard errors are so large partly because the data is annual and partly because we only have one positive pre-crisis dummy for each crisis. Despite the relatively large standard errors, the ranking of the methods tends to hold across subsamples. The first two rows of Table 3(a) show that the logit model with multiple lagged predictors outperforms the logit model with a single lag in all subsamples. Using the test by Delong et al. (1988), we confirm that the differences in AUC are typically statistically significant at the 10% significance level. Thus, in cross-validation, the logit model gains from the information in the additional lagged predictors.

Similarly, the next two rows in Table 3(a) indicate that the time-window MLP(5) outperforms the static MLP(1) except for the short pre-WW2 sample, where the methods have similar performance. The lower rows in Table 3(a) reveal that the gated RNNs are among the top-performing models. They consistently outperform the logit models and the MLPs in all subsamples. The differences in AUC tend to be statistically significant at the 5% significance level. Also, the gated RNNs always outperform the basic RNN. In the cross-validation, all the prediction models tend to perform better in the post-WW2 era. This could be due to the larger number of observations, better data quality, or increased homogeneity among countries through globalization, trade, and development.

We move on to the sequential evaluation, where we again consider four alternative sample splits. In the pre-WW2 era, the sample is naturally split by the WW1. Here we estimate the model using pre-WW1 data and test it using the period between WW1 and WW2. Data during WW1 is excluded, as in Schularick and Taylor (2012). For the second era, we divide the sample into two parts, either at the year 1990 or 2000. Thus, the four alternative subsamples are:

(1) Estimation sample: 1870–1914 (8 crises), Test sample: 1920–1939 (11 crises).
(2) Estimation sample: 1946–1989 (7 crises), Test sample: 1990–2016 (17 crises).

[8] Sometimes the reliability diagrams use equally spaced bins. Since we have one class that is considerably underrepresented and limited amount of data, we use the quantile-based variant of the reliability diagram to avoid bins with too few observations.

Table 3
Performance for one-year ahead crisis prediction in different subsamples.

(a) Country-by-country cross-validation

Model	(1)		(2)		(3)		(4)	
Logit(1)	0.598	(0.044)	0.609	(0.095)	0.621	(0.062)	0.610	(0.058)
Logit(5)	0.662	(0.043)	0.624	(0.056)	0.673	(0.063)	0.642	(0.071)
MLP(1)	0.597	(0.044)	0.705	(0.064)	0.678	(0.046)	0.641	(0.052)
MLP(5)	0.698	(0.046)	0.652	(0.083)	0.710	(0.050)	0.735	(0.056)
RNN	0.736	(0.051)	0.691	(0.058)	0.807	(0.050)	0.782	(0.049)
RNN-LSTM	0.747	(0.035)	0.716	(0.049)	0.867	(0.031)	0.844	(0.039)
RNN-GRU	0.715	(0.039)	0.700	(0.056)	0.866	(0.027)	0.801	(0.039)
Period	1870–2016		1870–1939		1946–2016		1970–2016	
Crises	43		19		24		23	
N	1142		273		869		589	

(b) Sequential out-of-sample evaluation

Model	(1)		(2)		(3)		(4)	
Logit(1)	0.621	(0.111)	0.600	(0.065)	0.524	(0.070)	0.535	(0.068)
Logit(5)	0.656	(0.088)	0.521	(0.072)	0.395	(0.066)	0.374	(0.062)
MLP(1)	0.652	(0.092)	0.545	(0.059)	0.512	(0.072)	0.485	(0.063)
MLP(5)	0.601	(0.100)	0.490	(0.077)	0.493	(0.081)	0.474	(0.074)
RNN	0.701	(0.076)	0.576	(0.060)	0.652	(0.083)	0.651	(0.086)
RNN-LSTM	0.724	(0.077)	0.702	(0.054)	0.726	(0.045)	0.743	(0.055)
RNN-GRU	0.667	(0.095)	0.685	(0.069)	0.645	(0.063)	0.734	(0.066)
Training period	1870–1914		1946–1989		1946–1999		1970–1999	
Crises train	8		7		12		12	
N train	151		512		642		418	
Test period	1920–1939		1990–2016		2000–2016		2000–2016	
Crises test	11		17		12		12	
N test	122		357		227		227	

The numbers in the table are AUC. Inside parentheses are standard errors adjusted for clustering at the country level.
Panels (a) and (b) show the AUC statistics for cross-validation and sequential evaluation, respectively. A higher value is better. Columns correspond to different subsamples. The dependent variable is the pre-crisis dummy defined in Section 5.1. See Table 2 and Appendix A for details of the models. See Appendix A.5 for details of the neural net training. All the models use up to five lags of the same variables: real annual house price growth, real annual stock index growth, annual growth in credit-to-GDP ratio, current account-to-GDP ratio, and annual growth in real GDP.

(3) Estimation sample: 1946–1999 (12 crises), Test sample: 2000–2016 (12 crises).
(4) Estimation sample: 1970–1999 (12 crises), Test sample: 2000–2016 (12 crises).

The AUC statistics for the sequential evaluation are shown in Table 3(b), where the columns again correspond to different subsamples. The AUCs lie in the range of 0.374−0.743, and the standard errors are between 0.045 and 0.111.

In Table 3(b), the performance measures drop overall compared to the cross-validation in Table 3(a), reflecting how hard the sequential out-of-sample prediction is. The RNN based neural nets still perform best across all subsamples. The gated RNNs (LSTM and GRU) are the top performers. In subsamples (3) and (4), the GRU is a bit unlucky in timing the 2008 financial crisis and tends to give signals one year too early. Overall, the performance differences between RNN based models and the other models are clearly statistically significant at conventional significance levels. Even if the performance of the RNN based neural nets deteriorated when we moved to the sequential evaluation, this drop in performance is similar as for the Logit(1) and the MLP(1). In contrast, the Logit(5) and the MLP(5) suffered considerably more. Hence, it seems that the structure of the RNNs is advantageous for out-of-sample predictions at the one-year forecast horizon.

Note that in the sequential evaluation of Table 3(b), the MLP and the logit model are frequently no better than a random guess. We will soon see that the one-year prediction is a particularly tough problem compared to predictions at the longer horizons. The reason is that we are quite strict about the timing and require that the predicted crisis must take place no later than at the specified forecast horizon (see the definition of the pre-crisis dummy in Section 5.1).

This concludes the analysis for the one-year forecast horizon. To sum up, in both cross-validation and sequential evaluation using a one-year forecast horizon, we find that the RNN based models, especially gated RNNs, consistently outperform the usual MLP neural nets and logit models in all subsamples. For practical purposes, policymakers may need predictions with a longer forecast horizon, especially if they plan to put in place policy measures that are intended to slow down the build-up of vulnerabilities.

Next, we will consider prediction performance for forecast horizons extending up to five years. For brevity, we will focus on the 1970−2016 subsample because that is the currently relevant period. Table 4 presents the results for both cross-validation and sequential evaluation. Each column corresponds to one of the five alternative prediction horizons. Additionally, we show the average in-sample AUC for the different cross-validation folds. Comparing the in-sample AUC in Table 4 with the model descriptions in Table 2, we see that a higher number of model parameters correspond to a higher in-sample fit. All the models (even the most parsimonious Logit(1) model) have a considerably higher performance in the estimation sample than in the test sample.[9]

Let us, for a moment, concentrate on the cross-validated results. Looking at the columns in Table 4, we can see that the best performance is found at a forecast horizon of 2–3 years. Except for the Logit(1) and the MLP(1), which are quite sensitive to the forecast horizon, the models perform quite well even at the 4 to 5-year forecast horizons. As previously, the additional information in the time series is demonstrated by the fact that the cross-validated Logit(5) and MLP(5) outperform their single lag counterparts for all prediction horizons. While all neural nets generally outperform the

[9] The in-sample AUCs are comparable to the AUC reported in the in-sample horse race for various financial crisis prediction models in Alessi et al. (2015).

Table 4
Performance for different forecast horizons in the 1970-2016 subsample.

		Forecast horizon (years)									
Model	Evaluation	1		2		3		4		5	
	In-sample	0.689		0.711		0.726		0.762		0.615	
Logit(1)	Cross-validation	0.610	(0.058)	0.655	(0.064)	0.671	(0.061)	0.695	(0.065)	0.411	(0.045)
	Sequential	0.535	(0.068)	0.592	(0.094)	0.716	(0.080)	0.694	(0.073)	0.273	(0.077)
	In-sample	0.843		0.863		0.892		0.897		0.868	
Logit(5)	Cross-validation	0.642	(0.071)	0.689	(0.062)	0.724	(0.051)	0.746	(0.062)	0.722	(0.056)
	Sequential	0.374	(0.062)	0.243	(0.055)	0.251	(0.049)	0.517	(0.084)	0.566	(0.064)
	In-sample	0.842		0.892		0.908		0.815		0.822	
MLP(1)	Cross-validation	0.641	(0.052)	0.743	(0.067)	0.757	(0.050)	0.660	(0.063)	0.593	(0.076)
	Sequential	0.485	(0.063)	0.669	(0.087)	0.802	(0.070)	0.595	(0.081)	0.515	(0.083)
	In-sample	0.959		0.968		0.983		0.978		0.985	
MLP(5)	Cross-validation	0.735	(0.056)	0.776	(0.060)	0.829	(0.037)	0.784	(0.041)	0.793	(0.031)
	Sequential	0.474	(0.074)	0.485	(0.082)	0.715	(0.050)	0.668	(0.085)	0.703	(0.084)
	In-sample	0.920		0.956		0.972		0.978		0.983	
RNN	Cross-validation	0.782	(0.049)	0.780	(0.063)	0.814	(0.041)	0.845	(0.038)	0.806	(0.047)
	Sequential	0.651	(0.086)	0.692	(0.087)	0.769	(0.061)	0.724	(0.065)	0.777	(0.054)
	In-sample	0.998		0.998		0.998		0.997		0.997	
RNN-LSTM	Cross-validation	0.844	(0.039)	0.870	(0.044)	0.873	(0.033)	0.835	(0.037)	0.823	(0.035)
	Sequential	0.743	(0.055)	0.783	(0.064)	0.801	(0.071)	0.751	(0.057)	0.817	(0.065)
	In-sample	0.996		0.994		0.995		0.993		0.990	
RNN-GRU	Cross-validation	0.801	(0.039)	0.863	(0.052)	0.858	(0.032)	0.784	(0.041)	0.782	(0.050)
	Sequential	0.734	(0.066)	0.850	(0.062)	0.818	(0.058)	0.660	(0.060)	0.732	(0.069)
Period		1970–2016		1970–2016		1970–2016		1970–2016		1970–2016	
Crises		23		22		22		22		22	
N		589		566		544		522		500	

The numbers in the table are AUC. Inside parentheses are standard errors adjusted for clustering at the country level.
The table shows the AUC statistics for in-sample, cross-validation, and sequential evaluation. In-sample numbers correspond to average in-sample AUC across the cross-validation splits; hence, no standard error is available for that measure. Columns correspond to different forecast horizons. The dependent variable is the pre-crisis dummy for each forecast horizon defined in Section 5.1. Higher AUC is better. See Table 2 and Appendix A for details of the models. See Appendix A.5 for details of the neural net training. All the models use up to five lags of the same variables: real annual house price growth, real annual stock index growth, annual growth in credit-to-GDP ratio, current account-to-GDP ratio, and annual growth in real GDP.

logit models, the RNN based models have the highest AUC statistics. The RNN-LSTM has overall the most robust performance across prediction horizons. The test for equality of two AUCs by Delong et al. (1988) reveals that the difference to the MLP(5) is statistically significant at the 5% significance level only at the 2-year horizon.

Let us now turn to the sequential evaluation results for different forecast horizons, which are also included in Table 4. Again the predictions are most accurate for 2–3-year forecasts. Similar to Table 3(b), the important findings in this sequential evaluation are the drastic drop of performance for the Logit(5) and the MLP(5) relative to other models, and that the RNN based methods still perform well and generally reach the highest AUC statistics. The Logit(1) and the MLP(1) have their strongest performance at the 3-year forecast horizon. This is because, in this subsample, the explanatory variables often peak 2–3 years before the crisis. In general, carefully tailoring the lag structure or transformation for each prediction horizon would possibly improve model predictions for all the models.[10]

So far, we have demonstrated that RNN based models predict the crises more accurately, as measured by the AUC, than the logit model and the MLP neural nets. Differences in AUC as such are not very tangible, however, so we end this section with some illustrations.

Fig. 4 presents the ROC curves for the LSTM and the Logit(5) model for cross-validation in the 1970–2016 subsample. Let us first focus on Fig. 4(a), which is for the 1-year forecast horizon. The vertical axis is the sensitivity; in other words, the share of correctly predicted crises. The horizontal axis is one minus specificity; in other words, the frequency of false alarms. For sensitivities above 0.5, the ROC curve for the LSTM lies approximately halfway between the vertical axis and the ROC curve for the Logit(5). Hence, for a given level of sensitivity (>0.5), the LSTM produces about half the amount of false alarms compared to Logit(5). For example, at the sensitivity of 80% (vertical axis = 0.8), Logit(5) yields a false alarm 80 percent of the time (horizontal axis = 0.8), while the LSTM yields a false alarm less than 40 percent of the time (horizontal axis<0.4). At low sensitivities (<0.5), the distance between the ROC curve for the LSTM and the vertical axis is about one-half of the gap between the two ROC curves. Hence, for a given level of sensitivity (<0.5), the LSTM produces about 1/3 the number of false alarms compared to Logit(5). In Fig. 4(b), we have similar ROC curves for the 3-year forecast horizon. Here too, the ROC curve of the LSTM lies closer to the vertical axis than the ROC curve of Logit(5). So again, we conclude that the LSTM produces visibly less false alarms for a given level of sensitivity.

Fig. 5 illustrates the average predicted crisis probabilities around systemic financial crises for Logit(1) and the LSTM, again for the one and three-year forecast horizon. This time we choose to plot the static Logit(1) instead of the Logit(5) to illustrate the advantages of the time series predictors. The gray area in Fig. 5 presents the crisis and post-crisis period. Consistent with the high AUC statistics, the LSTM gives, on average, a correctly placed strong signal. The LSTM optimized for 1-year forecasts (black dashed line), starts to signal the crisis relatively late (as it should); a clear peak extends from one year before the crisis to one year after the crisis. In contrast, the logit model (gray dashed line) gives its maximum signal three years before the crisis, and the signal slightly decreases during the following two years. This serves to illustrate that the logit model does not automatically give signals at the intended horizon. Ex-post, the explanatory variables of the Logit(1) model could be lagged to produce the peak at the intended horizon. However, such tweaking is unnecessary with the LSTM.

When optimized for 3-year forecasts, both models (black and gray solid lines in Fig. 5) give their strongest signal three years

[10] The lag structure should then be optimized for each estimation sample separately and tested out of sample. This is outside the scope of present study.

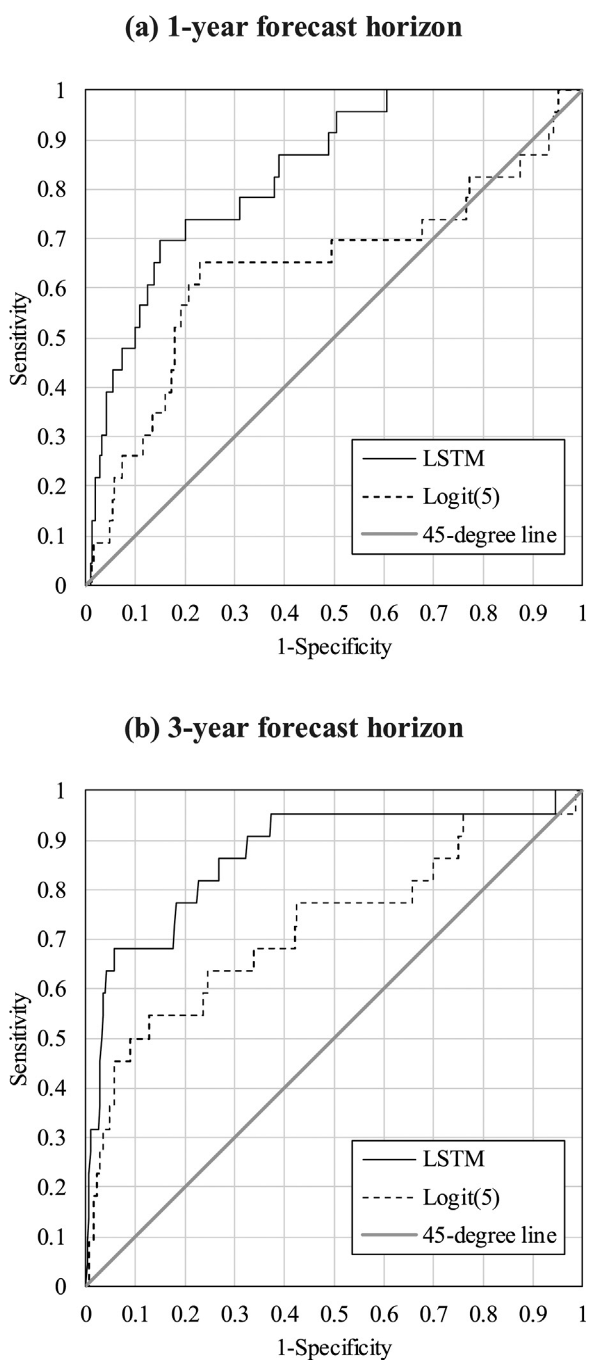


Fig. 4. ROC curves for the five-lag logit model and the LSTM neural net. Panels (a) and (b) show the ROC curves for the cross-validated one and three-year ahead crisis prediction in the 1970-2016 subsample, respectively. The notes of Table 4 apply.

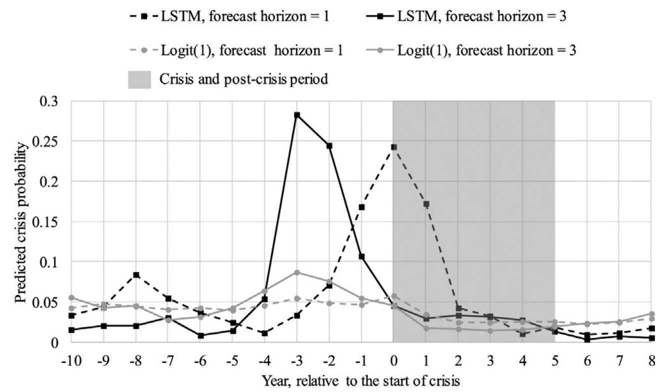


Fig. 5. Average predicted crisis probabilities around crisis dates for the one-lag logit model and the LSTM neural net. The line markers show the average predicted crisis probabilities at different years, relative to the year in which the crisis starts. The dependent variable is the pre-crisis dummy for the forecast horizon specified in the legend of each line. The optimized weights are for the 1970–2016 subsample. The notes of Table 4 apply.

before the crisis, and the signal fades after that. The fading signal is fine, but it has practical implications as the fading of the signal does not necessarily mean fading of the financial instability. It would be important to keep monitoring the probabilities for shorter forecast horizons as well.

6.2. Drivers of RNN predictions

Unlike the logit model, the neural nets include multiple layers of weights. Hence, we can't readily understand what drives their predictions. To determine each variable's contribution to the predictions, we calculate neural net predictions for subsets of input variables. A fundamental choice is whether we use the same neural net for predicting with a subset of predictors or estimate a new neural net. While training a new neural net for each input combination is computationally intensive, it is reasonable for our relatively small neural nets. It also ensures that the included predictors do not underperform due to ignoring interaction with excluded predictors.

We focus on the one and three-year forecast horizons. For brevity, we will focus on the LSTM, which had the most reliable performance in the previous section and compare it to the Logit(5). We train the LSTM neural net for each non-empty subset of the original predictor set {l/gdp, rsp, rhp, ca/gdp gdp}. Neural net hyperparameters remain the same as previously, and estimation is performed just as before.

The results for both the cross-validation and sequential evaluation are shown in Table 5. Each row shows the performance using a subset of predictors marked with x. Let us first discuss the cross-validated results (the left half in Table 5). We discuss the results for the one and three-year forecast horizons in parallel. Starting from the rows 1–5 of Table 5, we have the models with only one predictor. Here, the stock price (row 2) is the most informative predictor for both models and for both prediction horizons. Among the single predictor models, even if the LSTM has larger AUC statistics than the Logit(5) for both forecast horizons, the difference between the models is not statistically significant at conventional significance levels. We see this from the columns that show p-values for the Delong et al. (1988) test for the equality of the two AUCs.

Still discussing the cross-validated results, rows 6–12 in Table 5 show the models with two predictors. Again, the combinations that include the stock price generally perform the best, and the LSTM outperforms the Logit(5) by a statistically significant margin. The combinations of the stock price with other predictors also improve on the stock price alone (row 2). The same story continues to the three-variable models (rows 16–25) and the four-variable models (rows 25–30). The LSTM models that include the stock price perform best and are significantly better than the Logit(5). Hence, stock price seems to be an important variable among the selected predictors for the LSTM for this subsample. However, the LSTM model works without the stock price too. For the one-year forecast horizon, combinations of house price and GDP, and house price, current-account, and GDP also outperform the Logit(5) model by a statistically significant margin.

Table 5
Performance for different variable combinations in the 1970–2016 subsample.

Variables					Cross-validation						Sequential evaluation					
					1-year forecast			3-year forecast			1-year forecast			3-year forecast		
l/gdp	rsp	rhp	ca/gdp	gdp	LSTM	Logit(5)	p	LSTM	Logit(5)	p	LSTM	Logit(5)	p	LSTM	Logit(5)	p
x	–	–	–	–	0.667	0.652	0.482	0.670	0.636	0.167	0.645	0.584	0.183	0.576	0.531	0.511
–	x	–	–	–	0.744	0.674	0.341	0.846	0.788	0.313	0.632	0.449	0.000	0.829	0.460	0.000
–	–	x	–	–	0.694	0.656	0.465	0.655	0.621	0.346	0.664	0.635	0.730	0.712	0.613	0.192
–	–	–	x	–	0.694	0.659	0.601	0.591	0.538	0.551	0.590	0.605	0.862	0.571	0.480	0.474
–	–	–	–	x	0.653	0.550	0.062	0.686	0.627	0.354	0.462	0.324	0.029	0.617	0.570	0.429
x	x	–	–	–	0.807	0.686	0.013	0.891	0.798	0.075	0.695	0.484	0.000	0.785	0.428	0.000
x	–	x	–	–	0.671	0.659	0.794	0.668	0.641	0.715	0.650	0.572	0.307	0.680	0.504	0.044
x	–	–	x	–	0.681	0.654	0.483	0.628	0.604	0.606	0.695	0.569	0.032	0.592	0.534	0.360
x	–	–	–	x	0.684	0.644	0.536	0.711	0.688	0.526	0.625	0.533	0.005	0.562	0.585	0.743
–	x	x	–	–	0.777	0.671	0.094	0.896	0.787	0.054	0.674	0.440	0.000	0.887	0.348	0.000
–	x	–	x	–	0.821	0.698	0.041	0.819	0.784	0.538	0.693	0.503	0.000	0.710	0.389	0.000
–	x	–	–	x	0.792	0.613	0.004	0.871	0.748	0.013	0.582	0.350	0.000	0.869	0.390	0.000
–	–	x	x	–	0.737	0.662	0.177	0.680	0.612	0.225	0.717	0.632	0.376	0.741	0.538	0.027
–	–	x	–	x	0.736	0.603	0.015	0.707	0.699	0.880	0.649	0.423	0.006	0.721	0.638	0.118
–	–	–	x	x	0.660	0.634	0.550	0.569	0.612	0.476	0.593	0.521	0.078	0.663	0.560	0.316
x	x	x	–	–	0.836	0.677	0.011	0.857	0.777	0.054	0.698	0.428	0.000	0.834	0.288	0.000
x	x	–	x	–	0.835	0.693	0.013	0.891	0.786	0.021	0.766	0.474	0.000	0.763	0.385	0.000
x	x	–	–	x	0.810	0.649	0.009	0.876	0.748	0.007	0.695	0.422	0.000	0.752	0.380	0.000
x	–	x	x	–	0.728	0.646	0.068	0.677	0.619	0.464	0.747	0.565	0.017	0.718	0.486	0.000
x	–	x	–	x	0.709	0.633	0.168	0.745	0.697	0.312	0.645	0.439	0.005	0.675	0.503	0.026
x	–	–	x	x	0.674	0.651	0.609	0.682	0.660	0.588	0.701	0.530	0.000	0.537	0.584	0.538
–	x	x	x	–	0.813	0.678	0.013	0.869	0.775	0.075	0.717	0.504	0.002	0.791	0.317	0.000
–	x	x	–	x	0.808	0.622	0.001	0.890	0.753	0.033	0.632	0.303	0.000	0.920	0.329	0.000
–	x	–	x	x	0.804	0.645	0.007	0.849	0.741	0.034	0.672	0.419	0.000	0.692	0.367	0.001
–	–	x	x	x	0.741	0.653	0.043	0.707	0.688	0.773	0.682	0.550	0.161	0.761	0.562	0.022
x	x	x	x	–	0.832	0.673	0.013	0.861	0.763	0.023	0.752	0.426	0.000	0.786	0.271	0.000
x	x	x	–	x	0.831	0.641	0.000	0.878	0.730	0.002	0.688	0.355	0.000	0.843	0.263	0.000
x	x	–	x	x	0.832	0.652	0.007	0.897	0.738	0.002	0.783	0.412	0.000	0.705	0.346	0.000
x	–	x	x	x	0.680	0.646	0.556	0.684	0.676	0.781	0.698	0.481	0.005	0.732	0.468	0.001
–	x	x	x	x	0.829	0.649	0.011	0.897	0.747	0.014	0.707	0.411	0.000	0.810	0.316	0.000
x	x	x	x	x	0.833	0.645	0.004	0.871	0.726	0.008	0.748	0.374	0.000	0.801	0.251	0.000

The numbers in the table are AUC. p-values are for the Delong et al. (1988) test with $H_0 : AUC_1 = AUC_2$, where AUC_1 and AUC_2 are for the LSTM and Logit(5), respectively. p-values below 0.05 indicate that the AUCs are different at the 5% significance level.
The table shows the AUC statistics for the RNN-LSTM and the Logit(5) in cross-validation and sequential evaluation with one and 3-year forecast horizons. Higher AUC is better. Variables marked with x are included in the model specification of the corresponding row. The dependent variable is the pre-crisis dummy for each forecast horizon defined in Section 5.1. See Appendix A.5 for details of the neural net training. l/gdp = real annual house price growth, rsp = real annual stock index growth, rhp = annual growth in credit-to-GDP ratio, ca/gdp = current account-to-GDP ratio, and gdp = annual growth in real GDP.

Overall, we can conclude from Table 5 that the LSTM network outperforms the logistic model so long as we provide it with a sufficiently rich set of predictors.

Let us now briefly discuss the sequential evaluation results shown in the right half of Table 5. In the sequential evaluation, the stock price is relatively less important for the prediction at the 1-year forecast horizon, but still quite important for the prediction at the 3-year forecast horizon. The performance of the LSTM increases as we include more predictors. In contrast, the performance of the Logit(5) suffers as we include more predictors. Hence, the performance differences tend to be highly significant when we have multiple predictors.

The models calculated for Table 5 allow us to conveniently decompose the predictions of the original five variable LSTM into additive explanatory variable contributions, using a so-called Shapley value decomposition (see Lundberg and Lee, 2017; Bluwstein et al., 2020; Shapley, 1953). Fig. 6(a) and (b) present the average of this decomposition around systemic financial crisis dates for the one and three-year forecast horizon, respectively. Based on Fig. 6(a), the LSTM predictions at the one-year forecast horizon are driven on average most strongly by stock prices and then by house prices. The current account contributes strongly to the predicted probability at the crisis year, which is too late in terms of our performance statistics but may have practical implications. Fig. 6(b) shows that stock prices contribute strongly to the predictions at the three-year forecast horizon as well. Loans/GDP start to contribute two years before the crisis. House prices and GDP also help the predictions two to three years before the crisis.

Table 6
Decomposition of AUC for the RNN-LSTM neural net.

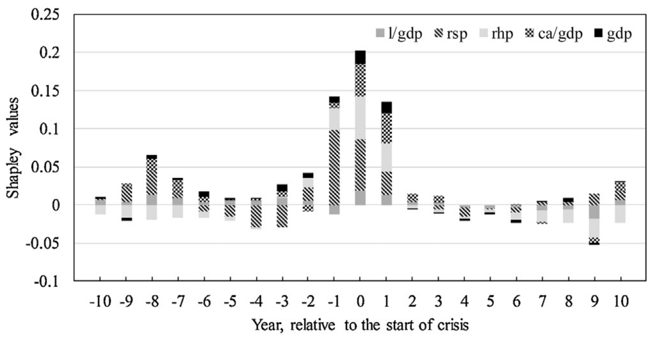
	LSTM			
	Cross-validation		Sequential evaluation	
Forecast horizon	1-year	3-year	1-year	3-year
l/gdp	0.040	0.042	0.075	−0.007
rsp	0.152	0.229	0.067	0.167
rhp	0.056	0.043	0.052	0.122
ca/gdp	0.051	0.005	0.072	−0.011
gdp	0.034	0.053	−0.018	0.030

The Shapley value decomposition of AUC is calculated directly from the formula $\phi_{AUC}(k) = \sum_{S \subseteq N-\{k\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(AUC_{S\cup\{k\}} - AUC_S)$, where $S \subseteq N$ is a subset of the set of five predictors and AUC_S is the AUC achieved by the corresponding neural net. The AUCs are the same as in Table 5 and $AUC_{\emptyset} = 0.5$. For example, the sum of the first column is 0.040 + 0.152 + 0.056 + 0.051 + 0.034 + 0.5 = 0.833 i.e. it decomposes the AUC of the 5 variable the LSTM into payoff contributions of each variable.

The Shapley value formula can also be used to calculate the contribution of each predictor to the AUC performance measure.[11] Table 6 conveniently summarizes the information in Table 5 using such a decomposition. From Table 6, we see that in the cross-

[11] An alternative measure would be to simply calculate the average AUC. However, we prefer the Shapley formula because it properly summarizes the value-added relative to other predictors.

(a) Cross-validation, 1-year forecast horizon



(b) Cross-validation, 3-year forecast horizon

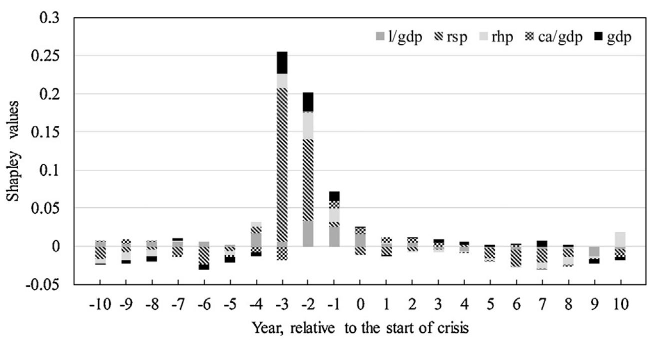


Fig. 6. Average Shapley values around crisis dates for the LSTM neural net. The bars show the average Shapley values of each predictor at different years relative to the year in which the crisis starts. The Shapley values are calculated directly from the formula $\phi(k) = \sum_{S \subseteq N-\{k\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(f_{S\cup\{k\}} - f_S)$, where $S \subseteq N$ is a subset of the set of five predictors and f_S is the output probability of the corresponding neural net. The optimized weights are for the 1970-2016 subsample, i.e. the same as in Table 5, and notes of Table 5 apply.

validated one-year forecast and three-year forecast, the stock price indeed emerges as the most important predictor, but also all other variables contribute. Only ca/gdp has a negligible contribution at the three-year horizon. For the sequentially evaluated one-year forecast, variables other than the gdp contribute roughly equally. In the corresponding three-year forecast, the stock price and house price are the most important predictors. These results are not surprising in relation to the earlier literature, which (as discussed in Section 2) also finds that, except for gdp, our variables are typically good predictors of banking crises. Nevertheless, it seems that price variables are particularly valuable inputs for the LSTM predictions, which to some extent reduces the role of loans/gdp.

7. Discussion and conclusions

We have investigated systemic financial crisis prediction with neural nets using the Jordá-Schularick-Taylor 1870–2016 dataset. We find that using time series input can lead to more accurate predictions. We also find that the RNNs, especially the gated RNNs (RNN-LSTM and RNN-GRU), outperform the logit model and the MLP neural nets. The results hold for both cross-validation and sequential evaluation across different subsamples. Thanks to their temporal structure, the RNN-based models show good performance in the sequential out-of-sample evaluation. The performance results hold for different forecast horizons of up to five years. We find that the LSTM neural net, on average, produces coherent signals at the intended forecast horizon.

We analyze the drivers of neural net predictions by considering subsamples of input variables. We find that at least two predictors are needed to realize a statistically significant benefit from the LSTM model. Among the considered set of predictors (loans/GDP, stock prices, house prices, current account/GDP, GDP), stock prices are found to be a key driver of the model predictions in cross-validation, but also other variables contribute. In the sequential evaluation, the variables contribute more evenly (except for the GDP, which is found to be less relevant).

Our results relate to the growing literature involved in predicting crises with machine learning and artificial intelligence. The findings for the RNN and the gated RNN neural nets are new since they have (to best of our knowledge) not been considered before in the crisis prediction literature. In general, our findings in cross-validation are consistent with studies that find that the new machine learning methods beat the logit model (Bluwstein et al., 2020; Holopainen and Sarlin, 2017; Ristolainen, 2018). Be that as it may, the logit model remains a useful policy tool for its ease of communication. In sequential evaluation, we do not find a substantial difference between the performance of MLP and the logit model. This is largely consistent with Beutel et al. (2019), who find that the logit model outperforms basic machine learning methods, including the MLP, in the sequential evaluation. In the cross-validation, we find the best forecast accuracy at the 3-year prediction horizon and using six years of time series input in the recurrent neural network. These numbers are consistent with the relatively long length of the financial cycle reported in the literature (8–20 years, see e.g. Borio, 2014 or Filardo et al., 2018). Findings on the minimum number of predictors are consistent with Fricke (2017), who finds that MLP neural nets based on only five lags of credit growth do not bring benefits in comparison to the logit model. The set of predictors that are found to be informative is largely consistent with the earlier literature. The findings for the contemporaneous informativeness of the current account are consistent with Davis and Karim (2008), who note that trade shocks play little part in the build-up of systemic risk, but a sudden deterioration in terms of trade could precipitate a crisis. Similarly, Bordo and Meissner (2012) found that the current account deficit bears no significant relationship with credit growth. However, due to endogeneity associated with simultaneous predictions, we cannot say whether the change in the current account is a cause or itself caused by the financial crisis. The issue of endogeneity is also important in the future if central banks or other financial stability authorities begin to exert influence on macro variables in anticipation of a crisis. Assessing whether a policy-induced reduction in risk indicators decreases the likelihood of a crisis is outside the scope of common financial crisis prediction models. Analysis of this important issue will require structural models and carefully controlled empirical studies.

In future work, a straightforward extension to ours would be to consider systemic financial crises prediction with RNNs using alternative predictors and datasets with quarterly or monthly frequency. Like Bussiere and Fratzscher (2006) and Caggiano et al. (2014), (2016), one could consider more than one class in the prediction task or a continuous dependent variable. The non-linearities associated with crises are expected to play a relatively larger role with a continuous dependent variable. One could also consider other types of crises or events, ranging from currency crises to recessions.

Acknowledgments

The author would like to thank Esa Jokivuolle, Helinä Laakkonen, Antti Ripatti, Matti Virén, Milan Vojnovic, seminar participants at the Helsinki GSE Time Series Econometrics Seminar, and the anonymous referees for useful comments and suggestions. The research was primarily conducted at the LSE in 2019 and supported by a HYMY travel grant from the University of Helsinki. The views expressed in this paper are those of the author and do not necessarily reflect the views of the Bank of Finland or the Eurosystem.

Appendix A. Neural net models

A.1 MLP

A multilayer perceptron (MLP) consists of three or more dense layers (see Fig. A1): an input layer, one or more hidden layers, and an output layer. "Dense layer" means that there is a connection between each node in successive layers. We consider two alternative MLPs, the MLP(1) and the MLP(τ). The MLP(1) makes predictions based on a cross-section of variables at time *s*. (We use *s* to denote the true time of the time series because *t* is conventionally reserved for the time-steps.) An MLP with one hidden layer can be defined recursively as:

$$\boldsymbol{h}(\boldsymbol{X}) = a_{relu}(\boldsymbol{W}\boldsymbol{X} + \boldsymbol{b}), \tag{A1}$$

$$o(\boldsymbol{h}) = a_{sigmoid}(\boldsymbol{V}\boldsymbol{h} + \boldsymbol{c}), \tag{A2}$$

where $a(.)$ are activation functions (applied elementwise). If we use a time-window of length τ, then $\boldsymbol{X}$ is the $d\tau$-dimensional input vector (d is the number of input time series and τ is the length of the time window). More precisely, for the prediction at time s, $\boldsymbol{X}$ is actually $vec([X_s X_{s-1} \ldots X_{s-\tau+1}])$, where X_s is the set of predictors at time *s,* and vec operator stacks the lagged values of the predictor into a one long column vector. The $\boldsymbol{h}$ is h-dimensional hidden layer vector, $o(.) \in [0, 1]$ is the output, $\boldsymbol{W}$ and $\boldsymbol{V}$ are $h \times d\tau$ and $h \times 1$-dimensional weight matrices, respectively, and $\boldsymbol{b}$ and $\boldsymbol{c}$ are h and one-dimensional bias vectors, respectively. We apply rectified-linear [relu, $a(x) = \max(0, x)$] activation function at the hidden nodes and sigmoid activation [$a(x) = 1/(1 + \exp(-x))$] at the single output node.

In our main results, the number of explanatory time series $d = 5$, the number of nodes in the hidden layer is $h = 10$, and the length of the time-window is $\tau = 1$ or $\tau = 5$. The latter corresponds to the number of lags used in Schularick and Taylor (2012) and Fricke (2017).

A.2 Basic RNN

In a basic RNN presented in Fig. A2, there is a hidden recurrent state $\boldsymbol{h}_t$ of dimensionality h, which evolves through time steps $t = 1, 2, \ldots \tau$. (Note that the index of the time series is $s = 1, 2, \ldots, T$). As previously τ corresponds to the window-length (say, five years), which is typically much less than the length of the whole sample T. The evolution of $\boldsymbol{h}_t$ depends on the previous hidden state $\boldsymbol{h}_{t-1}$ (if any) and the current explanatory variable $\boldsymbol{X}_{s-\tau+t}$. For ease of notation, let us consider the prediction at time $s = \tau$ such that $\boldsymbol{X}_{s-\tau+t} = \boldsymbol{X}_t$. At the final time step, the hidden state $\boldsymbol{h}_\tau$ is mapped to output prediction via a sigmoid function. The basic RNN can be defined recursively as

$$\boldsymbol{h}_t(\boldsymbol{h}_{t-1}, \boldsymbol{X}_t) = a_{tanh}\left(\boldsymbol{W}^{(t)}\boldsymbol{X}_t + \boldsymbol{U}^{(t)}\boldsymbol{h}_{t-1} + \boldsymbol{b}^{(t)}\right), \; t = 1,\ 2,\ 3,\ \ldots,\ \tau \tag{A3}$$

$$o(\boldsymbol{h}_\tau) = a_{sigmoid}(\boldsymbol{V}\boldsymbol{h}_\tau + \boldsymbol{c}). \tag{A4}$$

Here $a(.)$ are activation functions, $\boldsymbol{X}_t$ is the d-dimensional input vector of explanatory variables at the time step t, $\boldsymbol{h}_t$ is the h-dimensional recurrent state vector, $o(.) \in [0, 1]$ is the output (assumed to be one-dimensional in the binary classification task), $\boldsymbol{W}^{(t)}$, $\boldsymbol{U}^{(t)}$, and $\boldsymbol{V}$ are $h \times d$, $h \times h$ and $h \times 1$-dimensional weight matrices, respectively, and $\boldsymbol{b}^{(t)}$ and $\boldsymbol{c}$ are h and one-dimensional bias vectors, respectively. Usually, RNNs assume time-invariance[12] such that $\boldsymbol{W}^{(t)} = \boldsymbol{W}$, and $\boldsymbol{U}^{(t)} = \boldsymbol{U}$, and $\boldsymbol{b}^{(t)} = \boldsymbol{b}$. We assume time-invariance for the RNNs. In Appendix B, we verify that relaxing this assumption does not improve performance. We apply hyperbolic tangent activation function at the hidden nodes and sigmoid activation at the single output node (they correspond to the default setting in Keras).

In our main results, we consider a 5-dimensional time series of length $\tau = 5$, and recurrent state dimension $h = 10$. A sensitivity analysis is provided afterward. It should be noted that due to repeated multiplication of the hidden state by the same $\boldsymbol{U}$, estimating the basic RNN is susceptible to the problem of vanishing or exploding gradient when the number of time steps is large. RNN-LSTMs were originally developed to deal with this problem.

A.3 RNN-LSTM

The LSTM (Long Short Term Memory) Recurrent Network was proposed by Hochreiter and Schmidhuber (1997), and it has turned out to be quite popular. The idea is to make the recurrence going from $\boldsymbol{h}_t$ to $\boldsymbol{h}_{t+1}$ more subtle such that the network can accurately control what information propagates from one time step to another. Again, we consider the prediction at time $s = \tau$ such that $\boldsymbol{X}_{s-\tau+t} = \boldsymbol{X}_t$. To visualize the LSTM net, think of each hidden node in Fig. A2 being replaced by an LSTM cell depicted in Fig. A3.[13] The hidden state is now composed of two components: the recurrent hidden state $\boldsymbol{h}_t$ and the cell state $\boldsymbol{s}_t$, which both have dimensionality h. There is some optionality on how these two states are mapped to the output layer(s), as we will discuss momentarily.

Another new concept is the gating units σ, which are elementwise sigmoid functions that control the flow of information at points x in Fig. A3 (here x denotes Hadamard product $\odot$ i.e. elementwise multiplication). To understand the operation of the LSTM cell, let us start from the lower-left corner of Fig. A3. *Forget gate:* At time step *t*, the previous recurrent state and the input $(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ first feed into the forget gate. The forget gate outputs a vector of numbers $\boldsymbol{f}_t$ that lie in the interval [0,1]. At point x (above the forget gate in Fig. A3), we take a Hadamard product of this vector and the previous cell state $\boldsymbol{f}_t \odot \boldsymbol{s}_{t-1}$. In other words, the forget gate controls what information from the previous cell state is retained. The corresponding model equation reads

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}^f \boldsymbol{x}_t + \boldsymbol{U}^f \boldsymbol{h}_{t-1} + \boldsymbol{b}^f), \tag{A5}$$

where the $\boldsymbol{W}^f$ is $h \times d$-dimensional weight matrix, the $\boldsymbol{U}^f$ is $h \times h$-dimensional weight matrix, and the $\boldsymbol{b}^f$ is a h-dimensional bias vector.

Input gate: The input gate uses the information in $(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ to control what information from the $(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ themselves is stored to the cell state $\boldsymbol{s}_t$. The elementwise sum operation (at the center-top of Fig. A3), then combines the information that the forget gate retained from the previous cell state and the information that the input gate picked from the new input data and previous recurrent hidden state. The corresponding model equations read

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}^i \boldsymbol{x}_t + \boldsymbol{U}^i \boldsymbol{h}_{t-1} + \boldsymbol{b}^i), \tag{A6}$$

$$\boldsymbol{s}_t = \boldsymbol{f}_t \odot \boldsymbol{s}_{t-1} + \boldsymbol{i}_t \odot a_{tanh}(\boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{U}\boldsymbol{h}_{t-1} + \boldsymbol{b}), \tag{A7}$$

where the dimensions of $\boldsymbol{W}$, $\boldsymbol{U}$, and $\boldsymbol{b}$ (with and with superscripts) are the same as those in Equation (5).

[12] This is called parameter sharing in the neural net literature.

[13] There exists other variants of the LSTM such as a peephole LSTM (Gers et al., 2002).

Output gate: Finally, the output gate controls, again based on $(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$, to what extent the value in the new cell state is used to compute the new recurrent hidden state $\boldsymbol{h}_t$. The equations are

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}^o \boldsymbol{x}_t + \boldsymbol{U}^o \boldsymbol{h}_{t-1} + \boldsymbol{b}^o), \tag{A8}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot a_{tanh}(\boldsymbol{s}_t), \tag{A9}$$

where the dimensions of $\boldsymbol{W}^o$, $\boldsymbol{U}^o$, and $\boldsymbol{b}^o$ are the same as those in Equations ((5)–(7)).

The output of the LSTM: The final cell of the LSTM outputs a pair of vectors $(\boldsymbol{h}_\tau, \boldsymbol{s}_\tau)$. Because $\boldsymbol{h}_\tau$ already depends on $\boldsymbol{s}_t$ through Equation (9), the default approach is to discard the final cell state $\boldsymbol{s}_\tau$, and only use information in $\boldsymbol{h}_\tau$. However, there is no guarantee that discarding the cell state is optimal. Hence, while we discard $\boldsymbol{s}_\tau$ in the main results, we report the performance for the neural net that retains $\boldsymbol{s}_\tau$ in Appendix B. The output $(\boldsymbol{h}_\tau)$ [or $(\boldsymbol{h}_\tau, \boldsymbol{s}_\tau)$ in Appendix B] is connected to a single output unit using sigmoid activation (similarly as Eq. (A4)). The sigmoid and hyperbolic tangent activations in the recurrent layers correspond to default values in Keras.

In our main results, we consider inputs of dimension $d = 5$, time steps $\tau = 5$, and recurrent state dimension $h = 10$. A sensitivity analysis is provided afterward.

A.4 RNN-GRU

GRU is a gating mechanism proposed by Cho et al. (2014) with a similar purpose as the LSTM. It has only two gates - a reset gate and an update gate - and a single vector presents the hidden recurrent state. It has somewhat fewer parameters than the LSTM so that it can be computationally more efficient. While LSTM cells can do more complex tasks than GRU cells, GRUs have been shown to exhibit better performance in some relatively small datasets. Thus, in crisis prediction task, the GRU neural net could perform well for the same reasons as the LSTM.

In the GRU (see Fig. A4), the gates also have sigmoid activation and take $(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$ as the input. We denote by $\boldsymbol{u}_t$ and $\boldsymbol{r}_t$ the results from the update and reset gate. In the following, the $\boldsymbol{W}$s (with or without superscripts) are $h \times d$-dimensional weight matrices, the $\boldsymbol{U}$s (similarly) are $h \times h$-dimensional weight matrices, and the $\boldsymbol{b}$s (similarly) are h-dimensional bias vectors. The update controls to what extent information from the past is passed to the future. The gates have similar sigmoid structure as in the LSTM:

$$\boldsymbol{u}_t = \sigma(\boldsymbol{W}^u \boldsymbol{x}_t + \boldsymbol{U}^u \boldsymbol{h}_{t-1} + \boldsymbol{b}^u). \tag{A10}$$

The reset gate further helps in deciding what of the past information is forgotten. The formula is similar to the update gate, but the gate has different weights and connects differently to parts of the GRU cell.

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}^r \boldsymbol{x}_t + \boldsymbol{U}^r \boldsymbol{h}_{t-1} + \boldsymbol{b}^r) \tag{A11}$$

First, the reset gate is used to calculate an intermediate memory state $\boldsymbol{h'}_t$:

$$\boldsymbol{h'}_t = a_{tanh}(\boldsymbol{W}\boldsymbol{x}_t + \boldsymbol{U}(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}) + \boldsymbol{b}). \tag{A12}$$

Then we use the output of update gate, to produce the next recurrent state as a convex combination of the previous recurrent state and the intermediate memory state:

$$\boldsymbol{h}_t = \boldsymbol{u}_t \odot \boldsymbol{h}_{t-1} + (1 - \boldsymbol{u}_t) \odot \boldsymbol{h'}_t. \tag{A13}$$

Like RNNs and LSTMs, the hidden recurrent state from the last GRU cell feeds into a single unit dense output layer with sigmoid activation (similarly as Eq. (A4)). The sigmoid and hyperbolic tangent activations in the recurrent layers correspond to the default values of Keras at the time of writing. We consider 5-dimensional times series of length 5, hence $d = 5$ and $\tau = 5$. Like with other models, we set the recurrent state dimension $h = 10$. A sensitivity analysis is provided afterward.

A.5 Estimation of neural net parameters

The estimation of neural net parameters is normally called training the neural net. In the following, we review some neural net training concepts and summarize our training setup.

We train the neural nets by minimizing a loss function with Adam, an adaptive variant of the stochastic gradient descent algorithm. Training means optimizing the weights and biases such that the loss function is minimized based on a training data set. The prediction model is subsequently tested out-of-sample. For loss function, we use the cross-entropy given by

$$L\left(\{y_i, \hat{y}_i\}_{i=1}^N\right) = \sum\nolimits_{i=1}^N l\left(y_i, \hat{y}_i\right)$$

where the components are given by (A14)

$$l(y, \hat{y}) = -y \log\hat{y} - (1 - y)\log(1 - \hat{y}), \tag{A15}$$

where y is the outcome (1 for a pre-crisis period and 0 for a normal period) in the training data, $\hat{y}$ is the neural-net prediction, and N is the number of training observations.

As a typically non-convex optimization problem, training a neural net with a large number of weights is not trivial. Two optimization runs with different algorithms or different initial values hardly ever converge to the same local minimum. We overcome this issue by training the network several times, starting from random initial weights, and taking an average prediction. Of course, it is still important to choose suitable parameters for the optimization algorithm as in any other optimization problem. We experiment with different training parameters using cross-validation in the 1970–2016 subsample, as explained below.[14] We prefer parameters that lead to rapid convergence (to save computer resources) and high validated AUC statistics.

Key parameters in gradient descent based algorithms are the *learning rate*, the *batch size*, and the number of *epochs*. An epoch refers to one cycle through the full training dataset. The learning rate controls how much the neural weights change in each step of the gradient descent. The batch size is the number of observations used in each step of the gradient descent algorithm. Too large learning rate leads to non-convergence, while too small learning rate is computationally expensive and may cause the algorithm to converge too hastily to a local minimum. The batch size controls the randomness in training. Smaller batch size leads to more random paths for the parameters, which may help the algorithm to cover larger parameter space and find a better minimum. We considered learning rates in multiples of 10 (0.1, 0.01, 0.001, ...) and batch sizes in multiples of 2 (16, 32, 64, ...). We found the learning rate 0.01 and the batch size 16 to be a suitable combination for all the neural nets, although the batch size had little effect on the results. Somewhat lower learning rates also work, but the training would take a longer time. Note that we shuffle the data after every epoch, so each epoch uses different random batches.

To avoid overfitting, we use L2 regularization, which adds a penalty term proportional to the sum of the squared weights to the cross-entropy loss function, $\lambda\sum_j w_j^2$. We apply the regularization to weight matrices in all the dense layers (including the output layers) and the weight matrices in the RNN, LSTM, and GRU cells. We don't apply the regularization to the bias vectors, however. The

[14] Because we use a coarse grid for the training parameters, we expect that we would have ended-up with the same parameters had we used the full-sample.

strength of the regularization is controlled by the parameter λ. We consider values of λ in multiples of 10 (1, 0.1, 0.01,0.001, . . .). The time-window MLP is most sensitive to this parameter, and the preferred value is $\lambda = 0.01$. For other neural nets, we use $\lambda = 0.001$, although smaller values seem to work equally well.

The training time can impact the out-of-sample prediction performance of a neural net. If we train too little, the model performs poorly both in-sample and out-of-sample. If we train too much, the out-of-sample performance can (despite L2 regularization) start to deteriorate due to overfitting. To make sure that the neural nets are trained on par with each other, we apply a variant of so-called early-stopping algorithms adapted from Hansen et al. (1997). We train an ensemble of neural nets independently several times and stop at a time that is optimal for the average prediction made by the neural nets in the validation sample. Hence, the stopping time may not be optimal for the individual neural nets, but it is optimal for the ensemble of neural nets conditional on them being trained equal number of iterations. We set the maximum number of epochs to 100, which is enough for all our neural nets (results are largely unchanged if we set maximum epochs to only 20). The size of the ensemble is limited by computational resources because we need to train each neural net for each subsample and forecast horizon. To control the stochastic variation, we train a fifty-neural-net ensemble in the sequential evaluation. In cross-validation, we train an ensemble of 5 neural nets for each validation split in the main results and one in the driver analysis. For the cross-validation, the ensemble is largely unnecessary because we train an independent neural net for each validation split anyway; besides, the corresponding size of the validation sample is much larger than in the sequential evaluation. With such ensembles, the stochastic variation is already very small compared to the standard errors of our performance statistics.

In practice, we implement the neural nets and train them with Keras using a Tensorflow backend. For hardware, we use both Google Colab cloud service GPU runtime and a desktop computer with Nvidia GeForce GTX 1070. For example, training the LSTM neural net 50 times for 100 epochs using the 1970−1999 subsample takes approximately 20 minutes.

Appendix B. Sensitivity analysis and other robustness checks

In the following, we do a sensitivity analysis of the neural net parameters. Among the parameters, we consider the number of units in the hidden layer of the MLPs, the dimension of the hidden state in the RNNs, and the length of the time-window. We also verify that the time-invariance property of the RNNs is helpful. We consider both cross-validation and sequential evaluation. For brevity, we focus on the 1970−2016 subsample.

The number of units in the hidden layer or hidden state controls the complexity of the neural nets. Fig. B1(a) shows the AUC statistics for cross-validation as a function of the number of units (in the hidden layer for the MLP and in the hidden state in the RNNs). Consistently with the main text, the LSTM and the GRU rank at the top as long as there are at least five units in the hidden state. The third and fourth place generally goes to the basic RNN and the MLP(5), while the MLP(1) and the logit models have the weakest performance. Fig. B1(b) shows the same graph for the sequential evaluation. Again, the LSTM and the GRU generally perform the best. The RNN ranks third. The MLPs have poor performance in the sequential evaluation irrespective of the number of units. Hence, the results presented in the main article are robust for using a different number of units in the neural nets.

The length of the time-window controls how long periods of the time series the models use for each prediction. Fig. B2 presents the prediction performance (AUC) as a function of the length of the time-window for the neural net input data. For RNNs, the length of the time-window is usually called the number of time-steps. Panels (a) and (b) show the cross-validated and sequential AUC statistics, respectively. In both panels (a) and (b), we see that the optimal time-window is longer than 2. Also, the LSTM and the GRU generally rank the highest, followed by RNN, MLP, and then the logistic model. This confirms that the results in the main article do not depend on the specific choice of the time-window. Generally, the optimal window length seems to be six years for the cross-validation and four years for the sequential evaluation. With the MLP, long windows seem to result in overfitting. For the RNN based models, the longer window does not lead to overfitting. Thanks to the time-invariance, the longer window does not lead to a larger number of parameters in the RNNs.

Now, we consider relaxing the time-invariance assumption in the RNNs. Recall that in the RNN, we assume that $\boldsymbol{W}^{(t)} = \boldsymbol{W}$, and $\boldsymbol{U}^{(t)} = \boldsymbol{U}$, and $\boldsymbol{b}^{(t)} = \boldsymbol{b}$, for each time step $t = 1, \ldots, \tau$. Now, we relax this assumption and optimize each $\boldsymbol{W}^{(t)}$, $\boldsymbol{U}^{(t)}$, and $\boldsymbol{b}^{(t)}$ separately. This introduces many more parameters to the neural net, however. As an alternative, we consider partial time-invariance whereby $\boldsymbol{W}^{(t)} = \boldsymbol{W}$, and $\boldsymbol{U}^{(t)} = \boldsymbol{U}$, and $\boldsymbol{b}^{(t)} = \boldsymbol{b}$, for each time step $t = 1, \ldots, \tau - 1$, and $\boldsymbol{W}^{(\tau)}$, $\boldsymbol{U}^{(\tau)}$, and $\boldsymbol{b}^{(\tau)}$ are optimized separately. The motivation is that the additional parameters could be most valuable at the last time step just before the RNN outputs the prediction. We set $\tau = 5$ as in the main text.

Table B1 shows the prediction performance (AUC) for RNNs with different time-invariance assumptions. Panels (a) and (b) show

Table B1
Impact of alternative RNN assumptions.

Model	Time invariance	Cross-validation		Sequential evaluation	
RNN	yes	0.782	(0.049)	0.651	(0.086)
RNN	no	0.764	(0.056)	0.557	(0.083)
RNN	partial	0.782	(0.054)	0.645	(0.068)
RNN-LSTM	yes	0.844	(0.039)	0.743	(0.055)
RNN-LSTM+	yes	0.833	(0.046)	0.655	(0.062)
RNN-LSTM	no	0.756	(0.051)	0.526	(0.065)
RNN-LSTM	partial	0.801	(0.053)	0.708	(0.060)
RNN-GRU	yes	0.801	(0.039)	0.734	(0.066)
RNN-GRU	no	0.768	(0.054)	0.508	(0.079)
RNN-GRU	partial	0.767	(0.054)	0.705	(0.058)
Period		1970–2016		1970–2016	
N		589		418 + 227	

The numbers in the table are AUC. Inside parentheses are standard errors adjusted for clustering at the country level.
The table shows the AUC statistics for the RNN neural nets using different time-invariance assumptions and, in the case of LSTM, the use of the final cell state (LSTM+). The columns correspond to cross-validation and sequential evaluation with a one-year forecast horizon in the 1970−2016 sample. The dependent variable is the pre-crisis dummy for each forecast horizon defined in Section 5.1. Higher AUC is better. See Appendix A.5 for details of the neural net training. l/gdp = real annual house price growth, rsp = real annual stock index growth, rhp = annual growth in credit-to-GDP ratio, ca/gdp = current account-to-GDP ratio, and gdp = annual growth in real GDP.

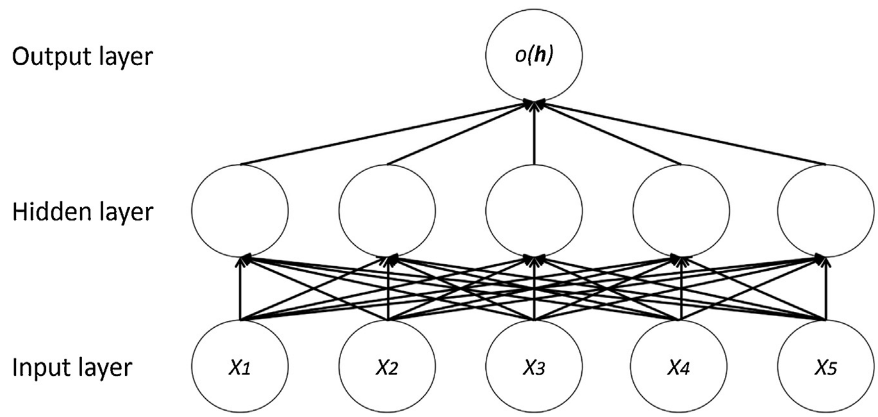


Fig. A1. A perceptron with one hidden layer.

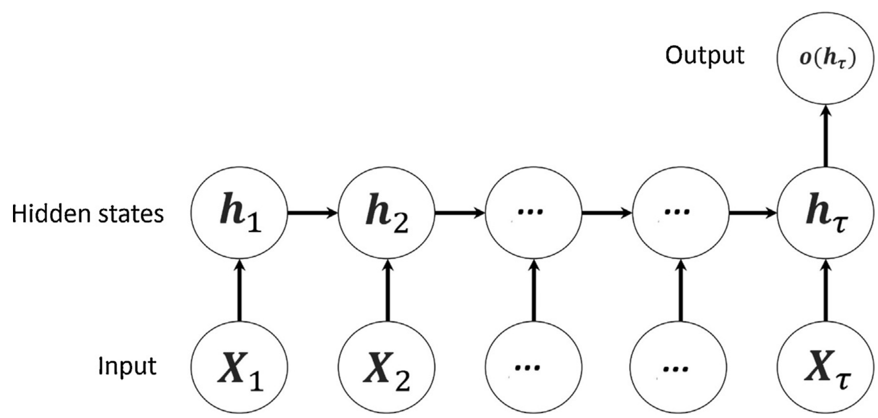


Fig. A2. A basic recurrent neural network.

the cross-validated and sequential results, respectively, using a one-year forecast horizon. We see that the time-invariant RNN models consistently outperform their more complex counterparts. The performance differences are larger in the sequential out-of-sample evaluation than in the cross-validation. In other words, we have found that the assumption of time-invariance helps make more robust predictions. Row 5, starting with "LSTM+..." shows the results for an LSTM neural net that additionally uses the final cell state output. As anticipated in the main text, using the final cell state output does not improve the result. Hence, the information in the final cell state is sufficiently captured in the other output state of the LSTM, which obtains the relevant information from the final cell state via the output gate (see Appendix A.3).

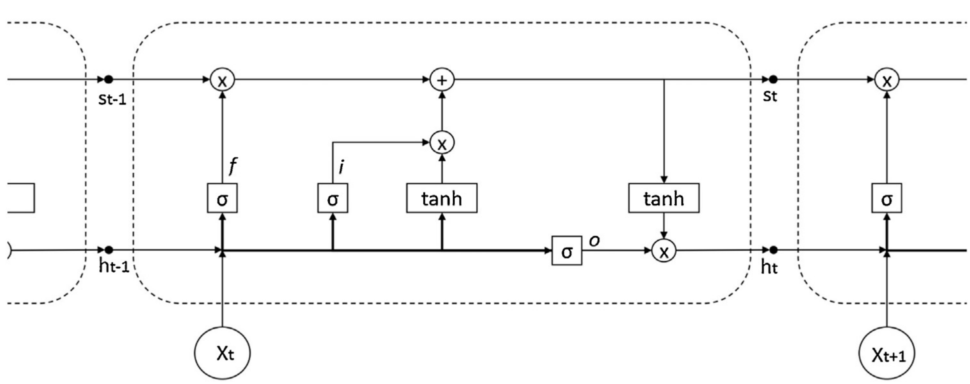


Fig. A3. A Long-Short Term Memory cell (adapted from illustrations based on Olah, 2015).

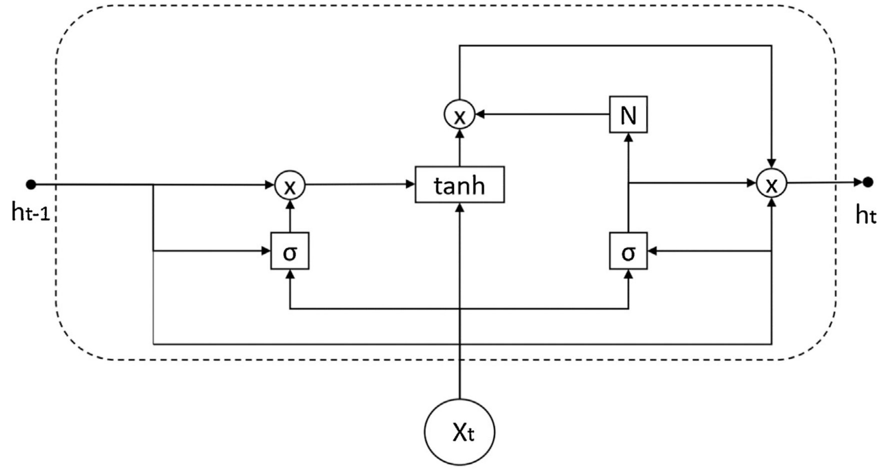
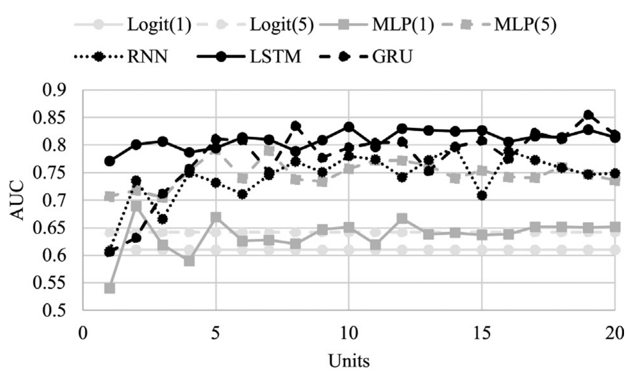


Fig. A4. A Gated Recurrent Unit (adapted from illustrations based on Olah, 2015).

(a) Cross-validation, the size of the ensemble = 17.



(b) Sequential evaluation, the size of the ensemble = 50.

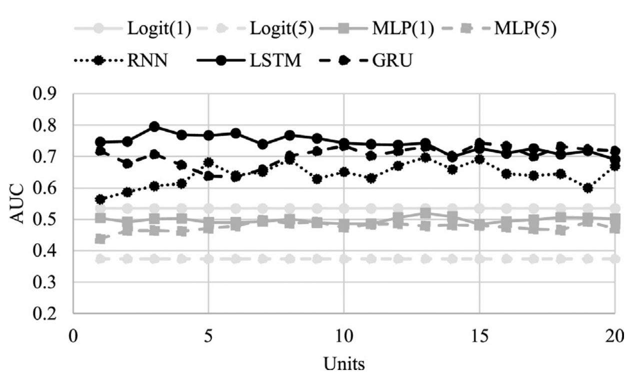
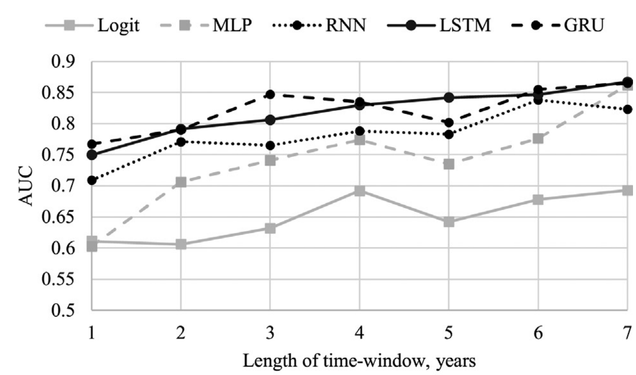


Fig. B1. Sensitivity analysis with respect to the number of units. Panels (a) and (b) show the cross-validated and sequential AUC statistics, respectively. The sample is 1970-2016. Higher AUC is better. The neural nets are estimated as explained in Appendix A.5.

(a) Cross-validation, the size of the ensemble = 85.



(b) Sequential evaluation, the size of the ensemble = 50.

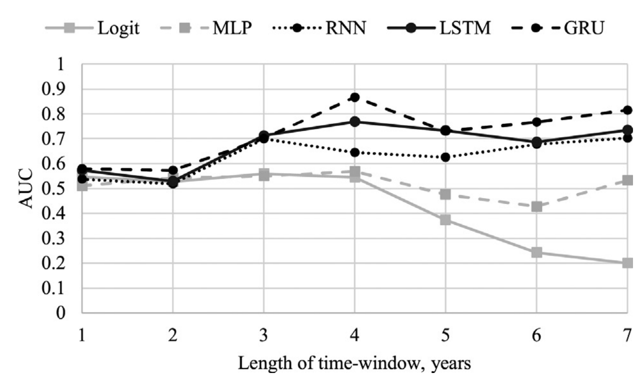


Fig. B2. Sensitivity analysis with respect to the length of the time-window. Panels (a) and (b) show the cross-validated and sequential AUC statistics, respectively. The sample is 1970-2016. Higher AUC is better. The neural nets are estimated as explained in Appendix A.5.

References

Alessi, L., Detken, K., 2011. Quasi real time early warning indicators for costly asset price boom/bust cycles: a role for global liquidity. Eur. J. Polit. Econ. 27 (3), 520–533.

Alessi, L., Detken, K., 2018. Identifying excessive credit growth and leverage. J. Financ. Stab. 35, 215–225.

Alessi, L., Antunes, A., Babecky, J., Baltussen, S., Behn, M., Bonfim, D., et al., 2015. Comparing Different Early Warning Systems: Results From a Horse Race Competition Among Members of the Macro-prudential Research Network, Available at SSRN: https://ssrn.com/abstract=2566165 or https://doi.org/10.2139/ssrn.2566165.

Apel, M., Grimaldi, M.B., Hull, I., 2019. How much information do monetary policy committees disclose? In: Evidence From the FOMC's Minutes and Transcripts, Sveriges Riksbank Working Paper Series 381.

Babecký, J., Havránek, T., Matěju, J., Rusnák, M., Šmídková, K., Vašíček, B., 2014. Banking, debt and currency crises in developed countries: stylized facts and early warning indicators. J. Financ. Stab. 15, 1–17.

Barrell, R., Davis, E.P., Karim, D., Liadze, L., 2011. How idiosyncratic are banking crises in OECD countries? Inst. Econ. Rev. 216, R53–R58.

Barron, A.R., 1993. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans. Inf. Theory 39 (3), 930–945.

Basel Committee on Banking Supervision (BCBS), 2011. Basel III: a Global Regulatory Framework for More Resilient Banks and Banking Systems — Revised Version. Bank for International Settlements.

Behn, M., Detken, C., Peltonen, T., Schuedel, W., ECB Working Paper No. 1603 2013. Setting Countercyclical Capital Buffers Based on Early Warning Models: Would It Work?

Berg, A., Pattillo, C., 1999. Are currency crises predictable? A test. IMF Staff Papers 46 (2), 1.

Beutel, J., List, S., von Schweinitz, G., 2019. Does machine learning help us predict banking crises? J. Financ. Stab. 45, 100693.

Binner, J.M., Elger, T., Nilsson, B., Tepper, J.A., 2004. Tools for non-linear time series forecasting in economics – an empirical comparison of regime switching vector autoregressive models and recurrent neural networks. Adv. Econometrics 19, 71–91.

Binner, J.M., Elger, T., Nilsson, B., Tepper, J.A., 2006. Predictable non-linearities in U.S. Inflation. Econ. Lett. 93 (3), 323–328.

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., Simsek, Ö., 2020. Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach. Bank of England Staff Working Paper No. 848.

Bordo, M.D., Meissner, C.M., 2012. Does inequality lead to a financial crisis? J. Int. Money Finance 31 (8), 2147–2161.

Borio, C., 2014. The financial cycle and macroeconomics: What have we learnt? J. Bank. Financ. 45, 182–198.

Borio, C., Drehmann, M., 2009. Assessing the Risk of Banking Crises – Revisited. BIS Quarterly Review, pp. 29–46, March.

Borio, C., Lowe, P., 2002. Assessing the Risk of Banking Crises. BIS Quarterly Review, December 2002.

Borovkova, S., Tsiamas, I., 2019. An ensemble of LSTM neural networks for high-frequency stock market classification. J. Forecast. 38 (6), 600–619.

Bussiere, M., Fratzscher, M., 2006. Towards a new early warning system of financial crises. J. Int. Money Finance 25 (6), 953–973.

Büyükkarabacak, B., Valev, N.T., 2010. The role of household and business credit in banking crises. J. Bank. Financ. 34, 1247–1256.

Caggiano, G., Calice, P., Leonida, L., 2014. Early warning systems and systemic banking crises in low income countries: a multinomial logit approach. J. Bank. Financ. 47, 258–269.

Caggiano, G., Calice, P., Leonida, L., Kapetanios, G., 2016. Comparing logit-based early warning systems: Does the duration of systemic banking crises matter? J. Empir. Finance 37, 104–116.

Caprio, G., Klingebiel, K., 1997. Bank insolvencies: Bad luck, Bad policy, or Bad Banking? Annual World, Bank Conference on Development Economics, pp. 79–04.

Casabianca, E.J., Catalano, M., Forni, L., Giarda, E., Passeri, S., 2019. An early warning system for banking crises: from regression-based analysis to machine learning techniques. Marco Fanno Working Papers 235.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1724–1734.

Cook, T.R., Smalter Hall, A., 2017. Macroeconomic indicator forecasting with deep neural networks, Federal reserve Bank of Kansas City. Research Working Paper 17-11, September, http://dx.doi.org/10.18651/RWP2017-11.

Davis, E.P., Karim, D., 2008. Comparing early warning systems for banking crises. J. Financ. Stab. 4 (2), 89–120.

Davis, E.P., Liadze, I., 2011. Should Multivariate Early Warning Systems for Banking Crises Pool Across Regions? Rev. World Econ. 147, 693–716.

DeGroot, M.H., Fienberg, S.E., 1983. The comparison and evaluation of forecasters. Statistician 32 (1/2), 12–22.

DeLong, E., DeLong, D., Clarke-Pearson, D., 1988. Comparing the areas under two or more correlated receiver operating characteristics curves: a nonparametric approach. Biometrics 44, 837–845.

Demirgüc-Kunt, A., Detragiache, E., 1998. The determinants of banking crises in developed countries. IMF Staff Pap. 45 (1), 81–109.

Demirgüc-Kunt, A., Detragiache, E., 2000. Monitoring banking sector fragility: a multivariate logit approach. World Bank Econ. Rev. 14, 287–307.

Detken, K., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M., Castro, C., et al., 2014. Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options. ERSB Occasional Paper Series No. 5 / June 2014.

Díaz-Martínez, Z., Sánchez-Arellano, A., Segovia-Vargas, M., 2011. Predicción de crisis financieras mediante conjuntos imprecisos (rough sets) y árboles de decisión. Innovar 21 (39), 83–100.

Domaç, I., Martinez Peria, M.S., 2003. Banking crises and exchange rate regimes: is there a link? J. Int. Econ. 61 (1), 41–72.

Drehmann, M., Juselius, M., 2014. Evaluating early warning indicators of banking crises: satisfying policy requirements. Int. J. Forecast. 30, 759–780.

Drehmann, M., Borio, C., Gambacorta, L., Jiménez, G., Trucharte, C., 2010. Countercyclical capital buffers: exploring options. BIS Working Papers 317.

Duttagupta, R., Cashin, P., 2011. Anatomy of banking crises in developing and emerging market countries. J. Int. Money Finance 30 (2), 354–376.

Filardo, A., Lombardi, M., Raczko, M., 2018. Measuring financial cycle time. BIS Working Papers 755.

Fioramanti, M., 2008. Predicting sovereign debt crises using artificial neural networks: a comparative approach. J. Financ. Stab. 4 (2), 149–164.

Fischer, T., Krauss, C., 2018. Deep learning with long-short term memory networks for financial marker predictions. Eur. J. Oper. Res. 270 (2), 654–669.

Fouliard, J., Howell, M., Rey, H., 2019. Answering the queen: online machine learning and financial crises. Presentation at the BIS Annual Conference.

Fricke, D., 2017. Financial Crisis Prediction: A Model Comparison (November 29, 2017). Available at SSRN: https://ssrn.com/abstract=3059052 or https://doi.org/10.2139/ssrn.3059052.

Gers, F.A., Eck, D., Schmidhuber, J., 2001. Applying LSTM to time series predictable through time-window approaches. International Conference on Artificial Neural Networks ICANN 2001, 669–676.

Gers, F.A., Schraudolph, N.N., Schmidhuber, J., 2002. Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. 3, 115–143.

Gogas, P., Papadimitriou, T., Matthaiou, 2014. Yield curve and recession forecasting in a machine learning framework. Comput. Econ. 45 (4), 635–645.

Hansen, L.K., Larsen, J., Fog, T., 1997. Early stop criterion from the bootstrap ensemble. 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing 4, 3205–3208.

Hardy, D.C., Pazarbasioglu, C., 1999. Determinants and leading indicators of banking crises: further evidence. IMF Staff Paper 46 (3), 1.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Holopainen, M., Sarlin, P., 2017. Toward robust early-warning models: a horse race, ensembles and model uncertainty. Quant. Finance 17 (12), 1933–1963.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.

Jordà, Ò., Schularick, M., Taylor, A.M., 2015. Leveraged bubbles. J. Monet. Econ. 76, S1–S20.

Jordà, Ò., Schularick, M., Taylor, A.M., 2017. Macrofinancial history and the New business cycle facts. In: Eichenbaum, Martin, Parker, Jonathan A. (Eds.), NBER Macroeconomics Annual 2016, vol. 31. University of Chicago Press, Chicago.

Joy, M., Rusnák, M., Šmídková, K., Vašíček, B., 2017. Banking and currency crises: differential diagnostics for developed countries. Int. J. Financ. Econ. 22 (1), 44–67.

Kaminsky, G., Reinhart, C., 1999. The twin crises: the causes of banking and balance-of-payments problems. Am. Econ. Rev. 89 (3), 473–500.

Kauko, K., 2012. External deficits and non-performing loans in the recent financial crisis. Econ. Lett. 115 (2), 196–199.

Kauko, K., 2014. How to foresee banking crises? A survey of the empirical literature. Econ. Syst. 38, 289–308.

Kauko, K., Tölö, E., 2020a. On the long-run calibration of the credit-to-GDP gap as a banking crisis predictor. Finnish Economic Papers (forthcoming).

Kauko, K., Tölö, E., 2020b. Banking crisis prediction with differenced relative credit. Appl. Econ. Q. (forthcoming).

Knoll, K., Schularick, M., Steger, T., 2016. No price like home: global house prices 1870-2012. Am. Econ. Rev. 107 (2), 331–353.

Kuan, C.-H., White, H., 2007. Artificial neural networks: an econometric perspective. Econom. Rev. 13 (1), 1–91.

Laeven, L., Valencia, F., 2012. Systemic Banking crises database: an update. Working Paper No.12/163.

Lang, J.H., Cosimo, I., Fahr, S., Ruzicka, J., 2019. Anticipating the bust: a new cyclical systemic risk inidicator to assess the likelihood and severity of financial crises. ECB Occasional Paper Series No. 219.

Lo Duca, M., Peltonen, T., 2013. Assessing systemic risks and predicting systemic events. J. Bank. Financ. 37 (7), 2183–2195.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems, 4768–4777.

Manasse, P., Roubini, N., 2009. "Rules of thumb" for sovereign debt crises. J. Int. Econ. 78 (2), 192–205.

Manasse, P., Savona, R., Vezzoli, M., 2013. Rules of thumb for Banking crises in emerging markets. Universita' di Bologna Working Papers No. 872.

Minami, S., 2018. Predicting equity price with corporate action events using RNN-LSTM. J. Math. Financ. 8 (1), 58–63.

Niculescu-Mizin, A., Caruana, R., 2005. Predicting Good probabilities with supervised learning. ICML 05 Proceedings of the 34th International Conference on Machine Learning, 625–632.
Nik, P., Jusoh, M., Shaari, A.H., Sarmdi, T., 2016. J. Econ. Cooperation Dev. 37 (1), 25–40.
Nyman, R., Ormerod, P., arXiv:1701.01428 2017. Predicting Economic Recessions Using Machine Learning Algorithms.
Olah, C., Available online: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ 2015. Understanding LSTM Networks.
Qi, M., 2001. Predicting US recessions with leading indicators via neural network models. Int. J. Forecast. 17 (3), 383–401.
Reinhart, C.M., Rogoff, K.S., 2009. This Time Is Different: Eight Centuries of Financial Folly. Princeton University Press.
Ristolainen, K., 2018. Predicting banking crises with artificial neural networks: the role of Nonlinearity and heterogeneity. Scand. J. Econ. 120 (1), 31–62.
Roy, S., Kemme, D.M., 2012. Causes of banking crises: deregulation, credit booms and asset bubbles, then and now. Int. Rev. Econ. Financ. 24, 270–294.
Sarlin, P., 2014. Mapping financial stability. Ai Commun. 27 (3), 285–297.
Schularick, M., Taylor, A.M., 2012. Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870-2008. Am. Econ. Rev. 102 (2), 1029–1061.
Shapley, L.S., 1953. A value for n-person games. Contributions to the Theory of Games 2 (28), 307–317.
Siami-Namini, S., Tavakoli, N., Namin, A.S., 2018. A comparison of ARIMA and LSTM in forecasting time series. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 1394–1401.
Suss, J., Treitel, H., 2019. Predicting bank distress in the UK with machine learning. Bank of England Staff Working Paper No. 831.
Tölö, E., Laakkonen, H., Kalatie, S., 2018. Evaluating Indicator for use in setting the countercyclical capital buffer. Int. J. Cent. Bank. 14 (2), 51–111.
von Hagen, J., Ho, T.-K., 2007. Money market pressure and the determinants of banking crises. J. Money Credit Bank. 39 (5), 1037–1066.
Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., Dean, J., arXiv:1609.08144 2016. Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation.

Article III

Tölö, E., Jokivuolle, E., and Virén, M., 2017, Do banks' overnight borrowing rates lead their CDS Price? *Journal of Financial Intermediation*, Vol. 31, pp. 93–106.

Available at: https://doi.org/10.1016/j.jfi.2017.05.006.

J. Finan. Intermediation 31 (2017) 93–106

Contents lists available at ScienceDirect

J. Finan. Intermediation

journal homepage: www.elsevier.com/locate/jfi

Do banks' overnight borrowing rates lead their CDS price? Evidence from the Eurosystem[☆]

Eero Tölö[a], Esa Jokivuolle[a,*], Matti Virén[b]

[a] Bank of Finland, PO Box 160, 00101 Helsinki, Finland
[b] Bank of Finland, PO Box 160, 00100 Helsinki, Finland and University of Turku, Finland

ARTICLE INFO

Article history:
Received 12 February 2016
Revised 15 March 2017
Accepted 17 May 2017
Available online 25 May 2017

JEL codes:
G01
G14
G21

Keywords:
Private information
Money markets
Overnight borrowing rates
Credit default swaps (CDS)
Lead-lag relationship
TARGET2
Eurosystem
Early-warning indicators

ABSTRACT

We construct a measure of a bank's relative creditworthiness from the Eurosystem's proprietary interbank loan data: average overnight borrowing rate relative to an overnight rate index (AOR). We then investigate the dynamic relationship between AOR and the credit default swap price relative to the corresponding market index of 60 banks during 2008–2013. Price discovery mainly takes place in the CDS market, but AOR also contributes to it. The lagged daily changes of AOR help predict CDS. This indicates that AOR includes private information, which the CDS market does not immediately incorporate. We further show that the private information advantage is concentrated on days of market stress and on banks, which mainly borrow from relationship lender banks. Such borrower banks are typically smaller, have weaker ratings, and are likely to reside in crisis countries. Competent authorities can use AOR as a complementary indicator of banks' concurrent health.

1. Introduction

In the wake of the recent financial crises, the need to understand the functioning of inter-bank money markets has grown considerably. Money market data may also be a source of early-warning indicators for future banking problems. We contribute to the quest for early-warning indicators by forming a measure of a bank's creditworthiness: its average overnight money market borrowing rate relative to an overnight rate index (henceforth AOR). We then investigate whether this spread provides timely information of changes in the bank's creditworthiness in addition to the leading market-based indicator, the bank's CDS price relative to a market-wide CDS index for European financial institutions (henceforth CDS).[1]

We use the proprietary database of the Eurosystem's overnight money market, which operates in the so-called TARGET2 large value payment system (Trans-European Automated Real-time Gross Settlement Express Transfer System 2). The overnight market is the shortest-term component of the interbank money market through which banks manage their liquidity. It is the key transmission channel for monetary policy in major central banks including the European Central Bank (ECB) and the US Federal Reserve. At the

☆ Eero Tölö, Esa Jokivuolle and Matti Virén are employed by the Bank of Finland. Eero Tölö is a member/alternate of one of the user groups with access to TARGET2 data in accordance with Article 1(2) of Decision ECB/2010/9 of 29 July 2010 on access to and use of certain TARGET2 data. The Bank of Finland and the MIPC (former PSSC) have checked the paper against the rules for guaranteeing the confidentiality of transaction-level data imposed by the MIPC pursuant to Article 1(4) of the above mentioned issue. The views expressed in the paper are solely those of the authors and do not necessarily represent the views of the Eurosystem or the Bank of Finland. We thank Simone Giansante, Iftekhar Hasan, Antti Lehtoranta, Ian Marsh, Niko Herrala, Jouko Vilmunen, Tuomas Välimäki, seminar and workshop participants at the Bank of Finland, Deutsche Bundesbank, and Universitat Pompeu Fabra, and, especially, an anonymous referee for valuable comments. All remaining errors are ours.

* Corresponding author.
E-mail addresses: eero.tolo@bof.fi (E. Tölö), esa.jokivuolle@bof.fi (E. Jokivuolle), matti.viren@bof.fi (M. Virén).

[1] Specifically, we form AOR and CDS by deducting the Euro OverNight Index Average (EONIA) from a bank's average overnight borrowing rate and, respectively, the iTraxx-index for European financials from the bank's CDS price.

http://dx.doi.org/10.1016/j.jfi.2017.05.006
1042-9573/

shortest maturity, the money market is a liquid credit market with high frequency of observations.[2]

Earlier research has already considered whether interest rates on overnight loans taken by a bank, typically from a number of other banks, reflect the borrower bank's creditworthiness. Furfine (2001) has shown with the Fed Funds data that the overnight borrowing rates do indeed reflect accounting measures of the bank's credit risk. However, to the best of our knowledge previous research has not considered how efficiently and fast these markets react to changes in credit risk.

We choose CDS as a benchmark because it is a leading *public* indicator of the credit risk of both corporations and banks (see e.g. Blanco et al., 2005, Longstaff et al., 2005, Forte and Peña, 2009, Norden and Weber, 2009, and Annaert et al., 2012; see also Arora et al., 2012 and Berg 2010). In spite of their maturity mismatch and the effect of the term structure of credit risk, new information about bank creditworthiness should push both AOR and CDS in the same direction.[3] Note that by defining AOR and CDS in relation to the respective market indices, we control for the term structure of interest rates and effectively separate the bank-specific part of overnight rates and CDS prices from general market conditions such as liquidity conditions.[4]

AOR is not publicly observable (other than to the borrower bank itself and the competent authorities of the Eurosystem). Moreover, many of the overnight interbank loans are results of longer-term lending relationships (cf. e.g. Cocco et al., 2009, Bräuning and Fecht 2012, Abbassi et al. 2014, and Affinito 2012) in which the lender may have acquired *private* information of the borrower. It is hence possible that AOR aggregates private information of the borrower bank's condition, which CDS has not yet incorporated. Such a situation can prevail so long as the informed lenders choose not to trade on their private information in the CDS market in such a way that their information would be immediately and fully revealed via CDS.[5] Moreover, CDS prices are quotes rather than actual transactions, which is another reason why AOR may reflect changes in a bank's creditworthiness faster than CDS given that the bank is willing and able to borrow in the interbank market.[6]

Our data cover the period from the beginning of June 2008 to the end of June 2013, comprising 60 banks, 1300 business days, and around 470,000 loan transactions with average value of about 100 million EUR. These yield approximately 46,000 daily AOR observations.

To investigate AOR's contribution to information concerning a bank in addition to CDS, we use two conventional price discovery measures, which are based on the vector error correction (VEC) framework (Hasbrouck 1995 and Gonzalo and Granger 1995), and Granger causality tests in a standard VAR model. We use the VEC and the VAR models as complementary approaches because we find that AOR is stationary and the evidence of co-integration between AOR and CDS is mixed. We use daily changes of AOR and CDS and estimate the models both for the panel of 60 banks and for individual banks.

During tranquil times, the overnight lenders of a bank may be less concerned about changes in the borrower bank's creditworthiness. However, because overnight loans are typically quite large and uncollateralized, AOR may become more informative of the borrower's credit risk in times of stress when lender banks become concerned of the asset quality and liability structure of the borrower bank.[7] As described, e.g., by Dang et al., (2015), a money-like debt instrument (overnight interbank loans in our case) can become sensitive to the issuing institution's asset quality when there are sufficiently bad public news concerning the asset quality. This can trigger private information acquisition among investors (the lender banks in our case). We investigate whether AOR is more informative when the underlying overnight loans are arguably more information sensitive. We do this by using conditioning variables, such as an indicator for days of market stress, which proxy for intensified information sensitivity of the overnight loans.

Our empirical results from the Granger causality framework show that in daily differences, using the panel of 60 banks, AOR leads CDS by up to two lags while there is no similar lead for CDS over AOR. The price discovery measures obtained from the VEC model indicate that although price discovery mainly takes place in the CDS market, AOR also contributes to it, and its contribution appears to intensify during periods of market stress. Bank-specific results vary considerably, and for individual banks, CDS may Granger cause AOR.

A further investigation in the Granger causality framework using the panel of banks reveals that AOR helps to predict CDS mainly during periods of market stress in the case of banks that are relatively dependent on relationship lender banks. These banks tend to have poorer ratings and come from crisis countries. These findings are consistent with the information sensitivity hypothesis, suggesting that private information, which makes AOR useful in predicting CDS, is most plentiful when the information sensitivity of overnight loans is elevated.

Our results have the following implications. First, by using the proprietary interbank overnight loan interest rate data, the Eurosystem authorities can extract information concerning banks' current condition, which complements the information obtained from banks' CDS prices.[8]

Second, our results provide rare evidence on the value of private information. It is not common to have data on private information signals, which bilaterally negotiated overnight loan rates

[2] Money market transaction data are available for longer maturities as well but we will focus on the overnight data because of the far bigger market size and liquidity, and because the accuracy of identifying interbank loans out of the entire population of large value payment transactions in the data base is highest in case of the overnight loans.

[3] We use the five-year CDS contract, which is the most liquid of the CDS contracts.

[4] Schwarz (2014) argues that during the financial crisis liquidity risk has explained the major part of the general rise in Euribor and sovereign interest rate spreads.

[5] The seminal paper on the theory of privately informed trading is Kyle (1985). We can consider the overnight loans market as a fragmented market whereas the CDS market is relatively more centralized. Our setting corresponds to a situation where both types of markets are open at the same time on the same asset but where prices are public knowledge only in the centralized market (the CDS market) while they are private knowledge in the fragmented market (overnight loans). As a result, information flows between the two markets may be asymmetric. We are not aware of theoretical papers which would exactly consider a setting of this kind although price formation in fragmented vs centralized markets has been studied e.g. by Wolinsky, 1990, and Biais, (1993). Studies on the upstairs and downstairs markets on stocks may also provide some guidance (see e.g. Booth et al., 2002). As Biais (1993, p. 175) puts it, "(a)n issue is whether inside traders can use the lack of transparency of fragmented markets to exploit their private information." Though compared to the stock market the CDS market is more of an insider market; see e.g. Acharya and Johnson, (2007), the quotes available in Bloomberg are in principle public. On strategic behavior of informed and uninformed traders, see also O'Hara (1997; chapters 4 and 5).

[6] In stressful times a bank may be denied credit for a "fair" rate at least by some banks in the interbank market so that the bank may opt for the central bank's liquidity facilities instead (provided it has sufficient collateral). Such unrealized overnight loan transactions could in themselves be quite informative. This phenomenon may hence create a bias against finding that AOR is more informative than CDS.

[7] Afonso, Kovner and Schoar (2011) using US overnight money market data find that "the day after Lehman Brothers' bankruptcy, loan terms become more sensitive to borrower characteristics". See also Angelini et al. (2011). Further, Covitz and Downing (2007) provide evidence from commercial paper spreads of non-financial companies that credit risk dominates liquidity risk even at very short maturities.

[8] Earlier literature which may have lacked access to sufficient CDS data, has also studied the role of bond and equity prices as leading indicators of bank fragility; see e.g. Gropp et al. (2004) and (2006).

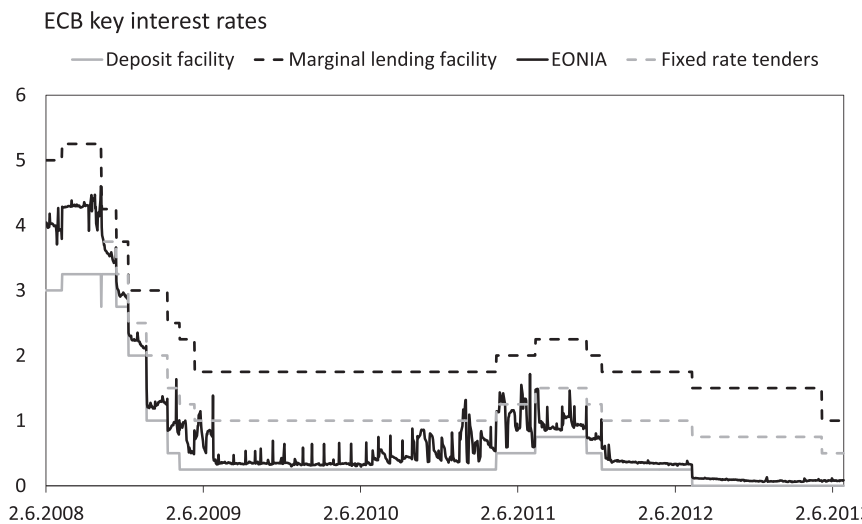


Fig. 1. ECB key interest rates for the sample period (beginning of June 2008 to end of June 2013).
The visible spikes in EONIA take place at the end of reserve maintenance periods and occasionally at the end of months. These spikes disappear starting from 2012 when the European banking system was flooded with unprecedented amounts of liquidity in the large long-term refinancing operations.

arguably reflect.[9] For comparison, Acharya and Johnson (2007) show that a firm's CDS price leads its stock price, and attribute this to private information in the CDS market. In a sense, our findings (using bank CDS data) take this a step further by showing that an aggregate of private information signals reflected in AOR can lead even CDS. Similarly, private information in AOR can explain our results with respect to Blanco et al., (2005) who show that CDS prices lead corporate bond spreads in the price discovery process.

We have organized the paper as follows. Section II describes the European interbank market and introduces the variables used. Section III presents the hypotheses, methodologies, co-integration tests, and empirical results. It also considers the economic significance of the results. Section IV concludes.

2. The data

2.1. Structure of the European interbank market

The Euro area monetary policy operations as well as the majority of transactions in the Euro area interbank market are settled in the TARGET2 system, which is the large value payment system of the Eurosystem.[10] Money market transactions are a subset of bank-to-bank large value payments. Bilateral loans are negotiated over-the-counter and are known only to the two parties involved in each transaction. The ECB's marginal lending facility and the deposit facility effectively set an upper bound and a lower bound, respectively, on the uncollateralized overnight loan rates.[11] The ECB's open market operations during the financial crisis and the subsequent European sovereign debt crisis have significantly increased the amount of excess reserves in the TARGET2, and have moved the average overnight rate (EONIA, see the description below) closer to the deposit facility rate. Further, the fixed tender rate of weekly main refinancing operations acts as a soft upper bound on the interbank rates (see Fig. 1). This is because a bank would be willing to borrow at a rate above the fixed tender rate only if it has no access to the ECB facilities, it has no available collateral, or it is concerned of reputational costs that could arise from borrowing from the central bank.

Regular "spikes" occur in the overnight rate series both on the last day of the reserve-maintenance period and typically on the last day of month. Spikes on the last day of the reserve-maintenance periods are caused by banks, which have a deficit in average reserves and borrow the gap for a premium from banks, which have accumulated an excess of reserves. Similar spikes on the last day of each month (especially quarter-ends and year-ends) apparently result from window dressing. However, because of the large excess reserves since the ECB's first long-term refinancing operations, these spikes no longer appear in the sample after December 2011. Unless noted otherwise, we report results with the special dates excluded from the data.

2.2. Panel and variables

2.2.1. 60 banks panel

Arciero et al. (2016) have provided the Eurosystem with a database of euro area money market transactions. They have identified money market loans from all TARGET2 transactions by an improved version of the algorithm originally suggested by Furfine (2001). The Arciero et al. (2016) algorithm is able to identify loan transactions with fair accuracy up to 3-month maturities, while the reliability is best for the overnight segment considered in this article.[12] The period of the dataset considered is from the beginning of June 2008, when the TARGET2 was fully operational, to the end of June 2013.

We identify the borrower and the lender with Business Identifier Codes (BICs). As one banking group may consist of several entities with their own BICs, we use information from the Swift BIC directory in order to consolidate the different entities under the common banking group and discard any loan transactions

[9] According to Garmaise and Moskowitz (2004), "(t)here are relatively few direct tests of the economic effects of asymmetric information because of the difficulty in identifying exogenous information measures".

[10] A number of European countries which do not belong to the Euro area are also connected to TARGET2 (for further details, see e.g. Arciero et al. 2016).

[11] These bounds are not strict because of two reasons. First, there may be banks without sufficient collateral who hence cannot access the ECB's liquidity facilities. Second, not all parties in the interbank market are eligible to access the central bank's facilities in the first place.

[12] We use a further improved version of the Arciero et. al. (2016) algorithm, which takes into account issues discussed in Armantier and Copeland (2012). We thank Arciero et. al. (2016) for providing this update.

Table 1
Descriptive statistics.
This table reports the number of observations each year that are used in the regressions and the mean and standard deviation for the key variables, AOR (EONIA subtracted) and CDS (iTraxx subtracted), and the bank relationship variables: Herfindahl–Hirschman Index (HHI), Borrower Preference Index (BPI), and bank size. AOR is defined as the average overnight borrowing rate minus the Euro OverNight Index Average (EONIA) and CDS is defined as the CDS price minus the iTraxx-index for European financials.

Year	Observations	AOR		CDS		HHI		BPI		Total assets	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
2008	6033	−0.141	0.166	0.311	0.807	0.069	0.128	0.083	0.135	639	604
2009	11,491	−0.151	0.119	0.542	1.040	0.089	0.142	0.101	0.138	648	623
2010	12,295	−0.078	0.113	0.558	1.575	0.098	0.156	0.113	0.155	635	597
2011	11,453	−0.081	0.176	1.062	3.072	0.083	0.152	0.100	0.150	670	595
2012	8983	−0.062	0.150	1.385	3.214	0.136	0.201	0.165	0.204	672	635
2013	3732	0.002	0.114	1.214	2.572	0.173	0.230	0.210	0.233	645	638
Total	53,987	−0.093	0.148	0.817	2.277	0.101	0.167	0.120	0.168	652	612

Observations for year 2008 start at the beginning of June and for 2013 end at the end of June. Total assets are in billions of US dollars.

that have taken place within banking groups. We match the BIC codes with Bloomberg CDS data and leave out banks with insufficient CDS data. As a result, our dataset has 60 borrower banks (domiciled in 19 different countries), 984 lender banks and 470,160 loan transactions. Taking the daily average yields 53,987 daily observations of AOR. Because of differencing and excluding the special dates with peaks, the final number of observations used in the regressions reduces to 33,823.

Table 1 includes descriptive statistics for the panel of 60 banks. For the period mid-2008 to mid-2012, there were around 12,000 observations per year. After mid-2012, the inter-bank overnight money market activity diminished because the ECB stepped up its liquidity providing operations and did not recover until the end of the data period. The decrease in money market activity is also accompanied with a change towards more concentrated markets with fewer counterparties, as measured by the bank relationship variables (see below for their precise definitions).

2.2.2. Average overnight rate with respect to EONIA (AOR)

For each business day, a bank may have borrowed from several lenders so we aggregate the daily rate from the multiple borrowings.[13] The loan issues generally take place between 7 a.m. and 6 p.m. Central European Time (CET) during the TARGET2 Day Trade Phase. Transactions towards the end of the day should contain the most recent information so we could use the time stamp as a weight in the aggregation.[14] The informativeness of a single transaction rate could also depend on the value of the loans or the intensity of the borrower-lender relationship. One could consider giving accordingly more weight to lenders that have close relationship with the bank (measured by past lending volume) or to loans that are of higher value or to use some percentile instead of average overnight rate. However, we found that different weighing schemes, or using percentiles of overnight rates, have only minor effect on the results so we simply use uniform weights in the daily rate aggregation per bank.

As already discussed above, to facilitate comparison with the CDS price, we transform the average overnight rates into average overnight rate spreads with respect to a suitable loan rate index. We find Euro OverNight Index Average (EONIA) a natural choice since it helps to account for general conditions in the euro money markets (e.g. the effects of policy rate changes, open market operations, and seasonal effects due to maintenance periods).[15] Hence, we define a bank's AOR as:

$$AOR_{B,t} = \frac{1}{N_{B,t}} \sum_{L} R_t^{L \to B} - EONIA_t. \tag{1}$$

Here $R_t^{L \to B}$ is the rate of an overnight loan from lender bank L to borrower bank B on day t. $N_{B,t}$ is the number of lender banks that lend to bank B on day t. For example, if on a given day EONIA is 1.016 % and banks A, B, and C, lend overnight to bank D at annualized rates of 1 %, 1.05 %, and 1.1 %, respectively, then AOR for bank D is calculated as $(1\% + 1.05\% + 1.1\%)/3 - 1.016\% = 0.034\%$.

2.2.3. Euro OverNight Index Average (EONIA)

The European Central Bank (ECB) calculates the EONIA rate every day based on the actual overnight loan transactions reported by a set of contributing banks. All overnight loans granted by the contributing banks before the close of TARGET2 at 6 p.m. CET are included and weighted according to their value. At the time of writing, the EONIA panel consists of 34 contributing banks.

2.2.4. Credit default swap price with respect to iTraxx

We obtain banks' CDS price data from Bloomberg. They are quotes rather than actual transaction prices. We use the last price field, which corresponds to the mid-quote at the end of trading. Because of time zone differences, the end of trading time may vary across the banks. Typically, the quotes take place 5 p.m. London time or 5 p.m. New York time and thus the price is quoted at the same time or later than the time at which the TARGET2 Day Trade Phase ends (5 p.m. London time).[16] Most of the overnight loans also take place well before closing; the average time a loan is advanced is 2 p.m. The CDS price is hence quoted somewhat more recently than the average money market transaction. This gives CDS price a small timing advantage, which creates a bias against finding a lead for AOR over CDS in the price discovery process. We only consider the most liquid CDS contract, the one with maturity of 5 years. To facilitate a comparison with AOR, we deduct the corresponding market index, the iTraxx Europe Financials CDS index (varying composition) from the bank's CDS price. We simply refer to the resulting "spread" as CDS, as already defined earlier.

Fig. 2 plots the daily AOR against CDS. Because of the much longer maturity of CDS (5 years) than of AOR (overnight), their re-

[13] In the case that a bank has not borrowed at all overnight on a given day, we treat this as a missing observation in our regressions.

[14] We find some evidence that "late" (after 12:00 CET) overnight loan transactions have a stronger impact on CDS than "early" (before 12:00 CET) overnight loan transactions but the difference is only marginally significant.

[15] Since the EONIA itself is not a risk-free rate, we transform the CDS prices in a corresponding manner (see the separate subsection below). The credit risk of EONIA is the value weighted credit risk of those who borrow from the EONIA panel banks.

[16] In other words, our main sample takes the most recent observation from either London or New York. In this way, we also maximize our sample size. However, as a robustness check, we have also run our main results with restricted London or New York samples and obtained similar results.

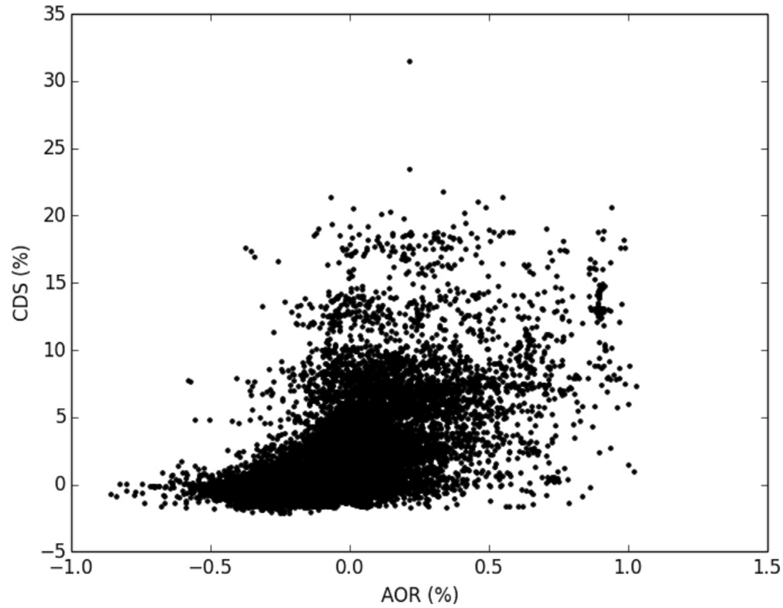


Fig. 2. Scatter plot of the daily AOR and CDS observations.

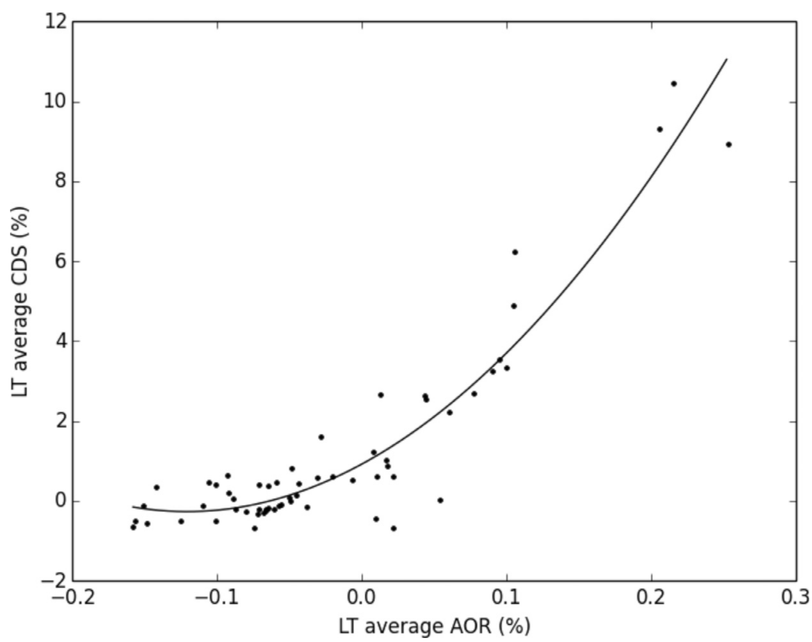


Fig. 3. Long-term average of AOR and CDS.
Each point corresponds to one of the 60 banks and the data are averaged over the whole period from begin of June 2008 to end of June 2013.

lationship is non-linear: higher values of AOR tend to be coupled with increasingly higher values of CDS. A similar pattern appears in Fig. 3 in which we have plotted the time-series averages of daily AOR and CDS over the entire sample against one another for each bank.

Fig. 4 provides an interesting insight by showing how the cross-sectional correlation between AOR and CDS varies in time. Before September 2008 and the ensuing global financial crisis the correlation was rather limited. In the aftermath of the Lehmann Brothers bankruptcy the correlation increased, and by the time the correlation reached its highest level, the Euro crisis had initiated from Greece. The change in the correlation would be consistent with AOR becoming more information sensitive during the sample period.

2.2.5. Markit iTraxx Europe senior financial sub index

The iTraxx Europe index is composed of the 125 most liquid CDS contracts of European entities. We use its sectoral sub index for financials, which consists of 25 equally weighted names most of which are direct participants in TARGET2. The iTraxx index has a high correlation (0.93) with the mean CDS price in our 60 banks panel.

2.2.6. Credit default swap bid-ask spread

The CDS bid-ask spread is used to proxy for the liquidity of CDS. We obtain the daily bid and ask CDS quote data from Bloomberg for 57 of the 60 banks (for three of the banks the data was unavailable) and calculate the bid-ask spread for each day. The bid-ask spread has a strong correlation (0.84) with the CDS price itself.

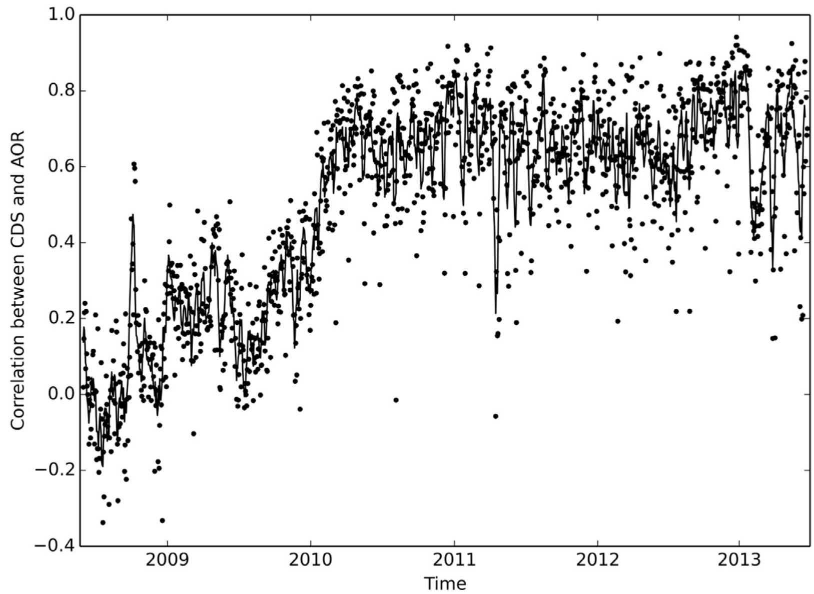


Fig. 4. The cross-sectional correlation between CDS and AOR.
The dots are the daily cross-sectional correlation values. The line shows 5 day moving averages. The correlation is calculated for those of the 60 panel banks that participate in the money market in the corresponding business day. The short-term variation may reflect different samples in different days.

2.2.7. Borrower Preference Index (BPI)

Similar to lending relationships between banks and corporates, there exist lending relationships between banks in the overnight loan market. Following Cocco et al., (2009), we measure the intensity of an interbank relationship by calculating how large a share that relationship contributes to the borrower's total borrowing during a certain period. The Borrower Preference Index (BPI) is the ratio of funds, F, that bank B has borrowed from bank L over a given time period Q_t, denoted by $F_{Q_t}^{L\to B}$, as a fraction of the total amount of funds that B has borrowed in the market in that period denoted by $F_{Q_t}^{ALL\to B}$:

$$BPI_{L,B,t} = \frac{F_{Q_t}^{L\to B}}{F_{Q_t}^{ALL\to B}}. \tag{2}$$

For each business day, t, we take period Q_t to be the last 62 business days which correspond to one quarter. Next, we define the average BPI as

$$\overline{BPI}_{B,t} = \frac{1}{N_{B,t}} \sum_L BPI_{L,B,t} \tag{3}$$

where the lender index, L, runs from 1 to $N_{B,t}$, which is the total number of lenders to bank B on day t. The higher the value of the average BPI on day t, the stronger is the relationship of the average lender to bank B on that day.

2.2.8. Herfindahl–Hirschman Index (HHI)

As an alternative proxy for the market structure and relationships we develop an application of the Herfindahl–Hirschman Index (HHI) to measure how concentrated the overnight borrowing activities of a given bank are on a given day. HHI is the total of squared daily market shares of each lender bank in the market of "all overnight lending to borrower bank B". If $F_t^{L\to B}$ is the amount of overnight funds bank B borrowed from bank L on day t, and $F_t^{ALL\to B}$ is the total amount of overnight funds bank B borrowed on day t, the HHI is then

$$HHI_{B,t} = \sum_L \left(\frac{F_t^{L\to B}}{F_t^{ALL\to B}} \right)^2. \tag{4}$$

Similar to BPI, the HHI index takes a value between 0 and 1. Generally, when the HHI is larger, the market is more concentrated. During times of financial market stress (as proxied by the iTraxx index) the average BPI and HHI tend to obtain higher values indicating more concentrated credit lines and more reliance on relationship lending.

2.2.9. Credit rating

As a bank credit rating, we use the Standard & Poor's long-term foreign currency issuer credit ratings. Following Covitz and Downing (2007), we convert the ratings to numbers from zero to 21.

2.2.10. TARGET2 excess reserves

As a response to the crisis, the ECB provided large amounts of liquidity to the banking system. We measure liquidity conditions by the amount of central bank money in the current account plus the deposit facility.

3. Dynamic relationship between AOR and CDS

3.1. Hypotheses

The following list summarizes our empirical hypotheses:

Hypothesis 1 (H1): Because of private information in the overnight loan market, AOR helps to predict CDS.

Hypothesis 2 (H2): The importance of AOR's private information in predicting CDS is greater

(i) *during financial market stress (crisis periods),*
(ii) *for relatively weaker banks,*
(iii) *for banks in countries with a sovereign debt crisis,*
(iv) *for banks which are relatively more dependent on relationship lenders,*
(v) *for banks whose CDS is relatively illiquid.*

H1 simply states our basic hypothesis that private information makes AOR a useful measure to predict CDS. AOR's predictive ability offers an implicit measure of the value of private information. Hypotheses *H2(i)–(iv)* refine *H1* and are motivated by potential

Table 2
Results from the panel vector error-correction analysis.
Panel (a) reports the panel tests for non-stationarity and co-integration. Panel (b) reports the maximum likelihood estimates for the vector error-correction (VEC) model. Panel (c) presents the price discovery measures that we compute based on the VEC-estimates.

(a) Tests for non-stationarity and cointegration.(i)

Hypothesis	Statistic	p
H_0: CDS is I(1), LLC	5.1	1.00
H_0: CDS is I(1), IPS	4.6	1.00
H_0: AOR is I(1), LLC	−12.2	0.00
H_0: AOR is I(1), IPS	−38.9	0.00
H_0: Zero cointegration vectors	1670.5	1.00

(b) Vector error-correction estimates.(ii)

Variable	ΔCDS_t	ΔAOR_t
ΔCDS_{t-1}	0.030	−0.008
	[0.71]	[0.86]
ΔAOR_{t-1}	0.031	−0.296
	[1.37]	[16.58]
$\lambda_{CDS/AOR}$	−0.001	0.004
	[1.86]	[15.58]
Constant	0.002	0.001
	[1.09]	[0.93]
Error-correction terms:		
CDS_{t-1}	1	1
AOR_{t-1}	−28.542	−28.531
Constant	−3.391	−3.260
R^2	0.002	0.168
N	33,823	33,823

(c) Price discovery measures.(iii)

Price discovery measures	
H_{LB}	0.99
H_{UB}	0.99
H_{MID}	0.99
GG	0.84

(i) LLC is Levin, Lin & Chu t* panel test for a common unit root process. IPS is Im, Pesaran, and Shin W-stat panel test for an individual unit root process. Co-integration test is the Johansen trace test.
(ii) Inside the brackets are robust t-statistics adjusted for clustering. N is the number of observations. The corresponding panel VEC equations are written as

$\Delta CDS_{i,t} = a_{11}\Delta CDS_{i,t-1} + a_{12}\Delta AOR_{i,t-1} + \lambda_{CDS}(CDS_{i,t-1} + b\ AOR_{i,t-1} + c) + \varepsilon_{1,i,t}$ and

$\Delta AOR_{i,t} = a_{21}\Delta CDS_{i,t-1} + a_{22}\Delta AOR_{i,t-1} + \lambda_{AOR}(CDS_{i,t-1} + b\ AOR_{i,t-1} + c) + \varepsilon_{2,i,t}$.

(iii) GG is the Gonzalo–Granger γ-measure, $\gamma = \lambda_{AOR}/(\lambda_{AOR} - \lambda_{CDS})$.
H is the Hasbrouck information share with $H_{MID} = (H_{UB} + H_{UB})/2$, where

$H_{UB} = (\gamma\sigma_{CDS} + \rho(1-\gamma)\sigma_{AOR})/[(\gamma\sigma_{CDS} + \rho(1-\gamma)\sigma_{AOR})^2 + (1-\gamma)^2\sigma_{AOR}{}^2(1-\rho^2)]$, and

$H_{LB} = \gamma^2\sigma_{CDS}{}^2(1-\rho)^2/[\gamma^2\sigma_{CDS}{}^2(1-\rho^2) + (\rho\gamma\sigma_{CDS} + (1-\gamma)\sigma_{AOR})^2]$,

Here σ^2 is the residual variance and ρ the respective correlation.

variation in a money-market debt contract's information sensitivity (see e.g. Dang et al., 2015). Hypothesis *H2(v)* can be motivated by the findings of Blanco et al., (2005) who argue that CDS leads the corresponding bond partly due to better liquidity. By the same logic, we can postulate that the role of private information in AOR in predicting CDS movements is bigger if the CDS contract is relatively illiquid.

3.2. Testing for co-integration between AOR and CDS

To test our hypotheses, we consider two alternative frameworks to model the dynamic relationship between AOR and CDS, which similar studies have used (see e.g. Blanco et al., 2005). The first of them, the vector error correction (VEC) model, establishes the long-term relationship between AOR and CDS by assuming they are co-integrated. The second is the Granger causality framework, using the standard VAR model, which does not include the co-integration relationship. Because AOR and CDS are essentially interest rate spreads, one can put forward an economic argument that they are stationary variables. However, similar studies have often found a co-integration relationship between the variables of interest (see e.g. Blanco et al., 2005), perhaps due to finite samples. Before proceeding, we test for stationarity and possible co-integration between AOR and CDS.

The co-integration analysis in the panel form is based on estimating the following VEC model:

$$\begin{aligned}\Delta CDS_{i,t} &= a_{11}\Delta CDS_{i,t-1} + a_{12}\Delta AOR_{i,t-1} \\ &\quad + \lambda_{CDS}(CDS_{i,t-1} + b\ AOR_{i,t-1} + c) + c_1 + \varepsilon_{1,i,t} \\ \Delta AOR_{i,t} &= a_{21}\Delta CDS_{i,t-1} + a_{22}\Delta AOR_{i,t-1} \\ &\quad + \lambda_{AOR}(CDS_{i,t-1} + b\ AOR_{i,t-1} + c) + c_2 + \varepsilon_{2,i,t} \qquad (5)\end{aligned}$$

OLS estimation of (5) assumes that the error term in both the CDS and AOR equation is independent over time and across banks. However, we find that while the autocorrelations of bank-specific error terms for both AOR and CDS are close to zero, the average pairwise correlation across banks is 17% for AOR errors and 26% for CDS errors. To account for the cross-sectional correlations we use robust t-statistics adjusted for clustering in all subsequent panel models. In our case, the sizes of the resulting robust t-statistics

Table 3
Results from the bank-specific vector error-correction analysis.
The banks are ordered by ascending Granger–Gonzalo price discovery measure.

Bank	H_0: CDS is I(1).	H_0: AOR is I(1).	Trace test	N	λ_{CDS}	t_{CDS}	λ_{AOR}	t_{AOR}	H_{LB}	H_{UB}	H_{MID}	GG
1	N	R	R	929	−12.5	4.23	−6.6	7.77	0.00	0.00	0.00	0.00
2	N	N	N	181	−23.9	3.10	3.8	1.63	0.21	0.23	0.22	0.14
3	N	R	R	632	−3.9	1.65	4.3	5.41	0.89	0.92	0.90	0.52
4	N	R	R	659	−6.2	2.58	7.4	6.68	0.83	0.88	0.85	0.54
5	N	R	R	634	−2.4	3.18	3.0	8.26	0.86	0.87	0.86	0.56
6	N	R	R	402	0.7	3.50	−0.9	12.87	0.95	0.93	0.94	0.57
7	N	R	R	882	−1.9	1.93	2.8	7.99	0.89	0.95	0.92	0.59
8	N	N	N	135	−20.3	1.86	34.6	3.32	0.79	0.75	0.77	0.63
9	N	R	R	866	−2.0	2.26	3.5	7.34	0.90	0.91	0.91	0.63
10	N	R	R	849	−14.4	2.59	25.1	5.89	0.85	0.84	0.84	0.64
11	N	R	R	788	7.2	2.16	−13.0	11.43	1.00	0.96	0.98	0.64
12	N	R	R	935	−8.9	2.73	18.1	8.41	0.93	0.90	0.92	0.67
13	N	R	R	757	0.7	1.20	−1.5	6.36	0.98	0.97	0.97	0.70
14	N	R	R	849	−2.6	1.38	7.7	5.63	0.94	0.94	0.94	0.75
15	N	R	R	791	−5.9	1.36	19.7	5.71	0.94	0.95	0.95	0.77
16	N	R	R	83	4.6	0.56	−17.4	5.46	1.00	0.99	0.99	0.79
17	N	R	R	361	3.3	1.17	−12.7	5.48	0.92	0.96	0.94	0.79
18	N	R	R	541	−0.7	0.79	2.7	6.04	1.00	0.98	0.99	0.79
19	N	R	R	689	1.0	1.58	−4.2	5.70	0.97	0.93	0.95	0.81
20	N	R	R	860	−8.1	2.40	35.7	9.96	0.95	0.94	0.95	0.82
21	N	R	R	920	−0.6	1.16	3.1	7.66	0.98	0.98	0.98	0.83
22	N	R	R	497	−0.2	0.83	0.9	4.97	0.97	0.97	0.97	0.84
23	N	R	R	563	−1.1	0.82	5.6	7.28	0.99	0.99	0.99	0.84
24	N	R	R	668	−1.3	0.78	8.3	7.20	0.97	0.99	0.98	0.87
25	N	R	R	178	5.2	1.02	−54.8	6.86	0.99	0.98	0.98	0.91
26	N	R	R	562	−1.3	0.66	14.5	6.26	0.99	0.99	0.99	0.92
27	N	R	R	892	−2.5	0.77	27.5	9.89	1.00	0.99	1.00	0.92
28	N	R	R	667	−0.7	1.25	8.8	10.88	0.99	0.99	0.99	0.92
29	N	R	R	307	−1.2	0.12	21.3	5.31	0.99	1.00	1.00	0.95
30	N	R	N	252	−0.1	0.05	4.0	4.17	1.00	1.00	1.00	0.97
31	N	R	N	166	−0.1	0.01	16.7	3.33	1.00	1.00	1.00	0.99
32	N	R	R	756	0.0	0.04	−5.9	5.76	1.00	1.00	1.00	0.99
33	N	N	R	434	0.3	0.06	13.0	7.22	0.99	1.00	0.99	1.00
34	N	R	R	365	−4.8	0.86	−18.1	4.31	0.98	1.00	0.99	1.00
35	N	R	N	62	0.2	0.09	1.4	2.77	0.92	1.00	0.96	1.00
36	N	R	R	363	−0.8	0.89	−1.6	4.29	1.00	1.00	1.00	1.00
37	N	R	R	266	−0.5	0.59	−9.1	7.67	0.99	1.00	1.00	1.00
38	N	R	R	730	−0.2	0.20	−4.5	8.76	0.99	1.00	1.00	1.00
39	N	R	R	848	−1.5	1.33	−8.2	8.92	1.00	1.00	1.00	1.00
40	N	R	R	602	0.8	1.31	2.8	6.05	1.00	1.00	1.00	1.00
41	N	R	R	171	2.0	0.58	9.6	5.48	0.98	1.00	0.99	1.00
42	N	R	R	875	−15.1	3.32	−22.6	9.06	1.00	1.00	1.00	1.00
43	N	R	R	806	−0.8	0.97	−3.7	5.23	1.00	1.00	1.00	1.00
44	N	R	R	456	3.8	0.97	4.4	4.83	0.99	1.00	1.00	1.00
45	N	R	N	142	−0.1	0.32	−1.0	4.43	0.99	1.00	1.00	1.00
46	N	R	R	901	−8.5	2.04	−13.6	7.20	1.00	1.00	1.00	1.00
47	N	R	R	647	−0.7	0.55	−3.8	6.09	1.00	1.00	1.00	1.00
48	N	R	R	575	0.0	0.02	−6.9	9.27	1.00	1.00	1.00	1.00
49	R	R	R	930	−4.1	0.97	−12.1	4.97	1.00	1.00	1.00	1.00
50	N	R	R	804	−1.0	1.07	−6.1	10.09	1.00	1.00	1.00	1.00
51	N	R	R	460	0.2	0.76	1.1	4.86	1.00	1.00	1.00	1.00
52	N	R	R	799	0.1	0.30	1.2	5.80	0.99	1.00	1.00	1.00
53	N	R	N	411	2.4	0.93	5.4	4.03	0.99	1.00	1.00	1.00
54	N	R	N	52	0.1	0.02	8.1	2.81	1.00	1.00	1.00	1.00
55	N	R	R	505	−0.1	0.20	−1.7	4.95	1.00	1.00	1.00	1.00
56	N	R	R	789	−2.6	1.86	−5.0	6.53	1.00	1.00	1.00	1.00
57	N	R	R	262	−0.1	0.28	−2.4	4.78	0.99	1.00	0.99	1.00
58	N	R	R	812	−1.1	0.93	−2.1	6.01	1.00	1.00	1.00	1.00
59	N	R	N	237	−3.0	0.92	−4.2	3.83	1.00	1.00	1.00	1.00
60	N	R	R	268	−1.5	1.80	−4.9	7.26	1.00	1.00	1.00	1.00
Mean/R	R (1)	R (57)	R (51)	564	− (43)	* (14)	−(28)	* (31)	0.94	0.94	0.94	0.86
Median/N	N (59)	N (3)	N (9)	617	+ (17)	** (8)	+ (32)	** (31)	0.99	1.00	0.99	0.98

R = H_0 rejected, N = H_0 not rejected. In last two rows R (.) and N (.) summarize the numbers of rejections and non-rejections (at the 0.05 significance level), respectively; **(.) and *(.) denote the number of significant coefficients in one-tailed *t*-test at 0.01 and 0.05 significance level, respectively; and − (.) and + (.) denote the number of negative and positive coefficients, respectively. N, λ, H, and GG are defined as in Table 2.

are about half of the corresponding unadjusted *t*-statistics. Table 2a presents the panel test results and Table 3 presents the bank-specific results.[17]

The evidence of stationarity of AOR and CDS and co-integration between them is mixed. We reject the non-stationarity of AOR in

[17] Note that the panel form model assumes that the coefficients are the same for all banks, whereas the bank-specific models naturally allow for different coefficients across banks. Hence, we may view the panel estimates as capturing the “average” relationships in the bank-specific AOR and CDS dynamics. Because of the high confidentiality of the individual bank data, individual bank results appear in a random order with no link to actual bank identities or bank attributes.

the panel model with two alternative tests (Table 2a). In bank-specific results (Table 3), we find AOR to be stationary in all other than three cases. In contrast, CDS is non-stationary in the panel model and in almost all cases in bank-specific results.

Despite the evidence of stationarity of AOR, we also conduct the co-integration tests. We reject the null hypothesis of no co-integrating vectors in the panel setting (see Table 2a). With individual banks in Table 3, we reject no co-integration for 51 out of 60 banks (at the 5% significance level).

As Kremers et al., (1992) note estimation of the error-correction coefficients in the VEC model provides a further test of the existence of co-integration. As Table 2b shows, the panel coefficients are significant and obtain the expected sign. However, the bank-specific results in Table 3 are more mixed. The error correction coefficient of the AOR equation (λ_{AOR}) is significant and of the expected (positive) sign for about half of the sample banks. The error correction coefficient of the CDS equation (λ_{CDS}) obtains the expected (negative) sign for two thirds of the sample banks but is significant only in 13 cases.

In summary, as AOR appears stationary and the results of co-integration tests are somewhat mixed, we apply both methodologies, the VEC model and the Granger causality framework, as complementary approaches in studying the price discovery process.

3.3. Measuring price discovery using VEC model

We consider two conventional price discovery measures computed from the VEC model parameters: the Hasbrouck (1995) measure and the Granger–Gonzalo (1995) measure.[18] We compute them for both the panel (Table 2c) and each individual bank (Table 3).

The price discovery measures obtained from the panel model in Table 2c indicate that price discovery mainly takes place in the CDS market, although the Gonzalo–Granger (1995) measure indicates that AOR also has a role.

Fig. 5 depicts the series of the price discovery measures from a rolling panel estimation with a six-month window. The rolling measures suggest that AOR's contribution to price discovery varies over time and seems to intensify during periods of market turbulence. This suggest, in accordance with our second hypothesis (*H2*), that the information potential of AOR may lie dormant for much of the time, apparently due to its short maturity, but comes to life when private information acquisition intensifies, as during a heightened financial crisis.

Table 3 reports the price discovery measures for individual banks, which vary considerably across banks. They indicate that on average price discovery mainly takes place in the CDS market. However, for some banks AOR strongly contributes to price discovery.

Overall, the results suggest that the value of private information in AOR in predicting CDS is rather specific to certain banks and periods. This is consistent with a varying degree of information sensitivity of overnight loans, as suggested by our second hypothesis, *H2*. We will investigate this further in section III.E below.

3.4. Granger causality

In this section, we test for Granger causality between AOR and CDS in the standard VAR framework, which does not include a co-integration relationship. Again, we report both panel estimation results and bank-specific results.

We readily obtain the VAR model with one lag in the panel form by setting parameters λ_{CDS} and λ_{AOR} equal to zero in Eq. (6):

$$\begin{aligned} \Delta CDS_{i,t} &= a_{11}\Delta CDS_{i,t-1} + a_{12}\Delta AOR_{i,t-1} + c_1 + \varepsilon_{1,i,t} \\ \Delta AOR_{i,t} &= a_{21}\Delta CDS_{i,t-1} + a_{22}\Delta AOR_{i,t-1} + c_2 + \varepsilon_{2,i,t} \end{aligned} \tag{6}$$

Regarding the lag length of the VAR, we start by considering only the first lag for both AOR and CDS, as specified in Eq. (6), but then extend the analysis to multiple lags. Table 4 reports results. Considering first the results in the first two columns with the one-lag model, the positive and significant coefficient on the lagged AOR in the CDS equation implies that AOR causes CDS. This supports the first hypothesis (*H1*) that private information in AOR has value in predicting CDS. The corresponding coefficient on the lagged CDS in the AOR equation is not significant.[19]

The Granger causality results with the one-lag VAR model hold with respect to the following robustness checks.[20] First, they hold even if we do not subtract the respective indices from AOR and CDS but include them as control variables. We further find that EONIA and iTraxx do not cause one another, which indicates that the lead for AOR over CDS is due to idiosyncratic (bank-specific) rather than system-wide shocks. Second, the results are robust with respect to the following exogenous control variables. We add the lagged stock return in both the CDS and AOR equation, motivated by Acharya and Johnson (2007) who have studied the dynamic relationship between CDS and equity returns (see also Fung et al., 2008, Marsh and Wagner 2011, and Giannikos et al., 2013). Further, we add the lagged sovereign CDS because of the intensified bank-sovereign loop during the Euro crisis. Lastly, we include the CDS bid-ask spread, which aims to control for the possibility that AOR's ability to predict CDS relates to CDS market liquidity conditions.

Consider next the VAR results with multiple lags in Table 4. With multiple lags we run into a problem that especially many smaller banks in both the CDS and AOR data have missing observations. As a result, the sample size shrinks considerably as we add lags. This does have an effect on our results, which are largely driven by the smaller and weaker banks. We use two alternative approaches to deal with this issue. In the first approach, we simply fill the "holes" in the data by the most recent observation. The second uses the Kalman filter to estimate the entire VAR system in the presence of missing observations.[21] We report results with both approaches. For the panel VAR model, the value of the SBIC information criterion starts stabilizing after the third lag. However, already the model with two lags appears to be sufficient in terms of capturing the Granger causality between AOR and CDS. For comparison, we report the model with two lags and with four lags.[22]

Columns 3 and 4 in Table 4 show results for the model with two lags in both the AOR and CDS equation using the Kalman

[18] With the Hasbrouck (1995) measure we consider the average of the lower and upper bound of that measure, which we denote in Table 2c by H_{MID}, H_{LB}, and H_{UB}, respectively (see e.g. Blanco et al. 2005). Both Hasbrouck (1995) and Gonzalo-Granger (1995) measure in theory obtain values between zero and one. In the current specification, values closer to one indicate a higher share of price discovery taking place in the CDS market.

[19] These estimated lag coefficients of the VAR system are in line with those obtained from the VEC model in Table 2 although the VEC model-based estimates are not significant with the robust t-ratios. Note also that the AOR's own lag in the AOR equation is negative even though we have excluded the special dates causing spikes in the data (see section II.A).

[20] The results are not reported but available from the authors upon request.

[21] Following Commandeur et al. (2011), we formulate the VAR system as a state-space model and estimate it using maximum likelihood. In the presence of missing observations, the log-likelihood function reduces to the usual VAR log-likelihood function, where any missing observations are replaced by states predicted by the Kalman filter. The maximization is done using the EM algorithm, see Dempster et al. (1977).

[22] We study the optimal lag length up to ten lags by using both approaches to deal with the missing observations (results are available from the authors upon request). The value of the SBIC criterion declines (almost) monotonously but does not reach a local minimum. This is not uncommon in studies using very large samples like ours (cf. Granger 1998).

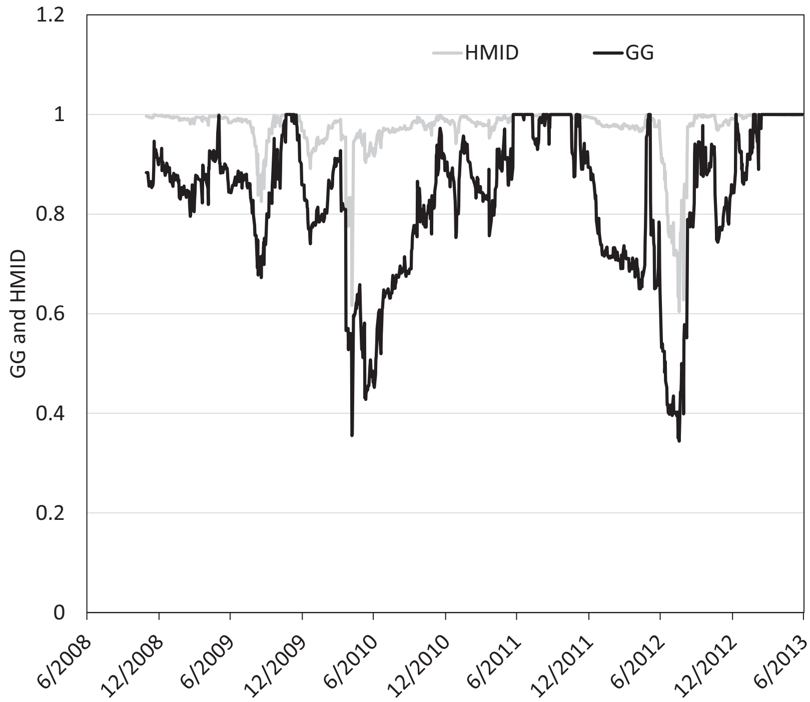


Fig. 5. The time-variation of price discovery measures.
The black line shows the Granger–Gonzalo (GG) price discovery measure and the gray line shows Hasbrouck information share mid-point value (HMID). Both measures are calculated with a six-month moving window with the model in Eq. (5) using daily observations.

Table 4
Estimates of a panel VAR with different lag structures.
In the models with 2 or 4 lags, we use two alternative approaches to deal with missing observations. In "Last price" we replace the missing observations by the most recent observation. Alternatively, we use Kalman filter to estimate the VAR system in the presence of missing observations.

Variable	1 daily lag		2 daily lags Kalman filter		2 daily lags Last price		4 daily lags Kalman filter		4 daily lags Last price	
	ΔCDS_t	ΔAOR_t	ΔCDS_t	ΔAOR_t	ΔCDS_t	ΔAOR_t	ΔCDS_t	ΔAOR_t	ΔCDS_t	ΔAOR_t
ΔCDS_{t-1}	0.030	−0.006	0.029	−0.005	0.029	−0.005	0.029	−0.003	0.029	−0.003
	[0.70]	[0.63]	[0.68]	[0.48]	[0.68]	[0.51]	[0.68]	[0.30]	[0.69]	[0.30]
ΔCDS_{t-2}			0.015	−0.005	0.015	−0.004	0.017	−0.003	0.017	−0.003
			[0.50]	[0.78]	[0.50]	[0.75]	[0.59]	[0.57]	[0.59]	[0.53]
ΔCDS_{t-3}							−0.027	−0.005	−0.027	−0.004
							[1.31]	[0.78]	[1.31]	[0.74]
ΔCDS_{t-4}							−0.025	0.003	−0.025	0.003
							[1.52]	[0.56]	[1.52]	[0.52]
ΔAOR_{t-1}	0.042	−0.354	0.056	−0.442	0.058	−0.437	0.054	−0.482	0.058	−0.474
	[2.02]	[18.99]	[2.51]	[23.57]	[2.60]	[23.37]	[2.26]	[25.48]	[2.40]	[25.25]
ΔAOR_{t-2}			0.037	−0.221	0.040	−0.207	0.034	−0.293	0.041	−0.274
			[1.62]	[14.08]	[1.90]	[13.61]	[1.29]	[17.12]	[1.66]	[16.65]
ΔAOR_{t-3}							−0.006	−0.189	−0.002	−0.176
							[0.27]	[11.30]	[0.10]	[11.11]
ΔAOR_{t-4}							−0.008	−0.085	0.004	−0.086
							[0.33]	[6.11]	[0.16]	[6.21]
Constant	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001
	[1.09]	[0.98]	[1.09]	[0.99]	[1.09]	[1.01]	[1.15]	[0.90]	[1.15]	[0.94]
R^2	0.001	0.127	0.002	0.174	0.002	0.171	0.003	0.196	0.004	0.193
N	33,823	33,823	33,781	33,781	33,781	33,781	33,700	33,700	33,700	33,700

Inside the brackets are robust t-statistics adjusted for clustering. N is the number of observations.

filter approach. Columns 5 and 6 show the corresponding results with the simple approach to fill in missing observations. Further, columns 7 to 10 show results with the four-lag model for both approaches. Overall, all models confirm the result that AOR Granger causes CDS but not the other way round. The coefficient estimates are quite stable across different models. Compared to the one-lag model, the coefficient on the first lag of AOR is somewhat bigger in models with two or four lags. The second lag also obtains a positive coefficient but is significant only in the two-lag model of column 5, which uses the simple approach to replace the missing observations. In the models with four lags, coefficients on both the third and the fourth lag are not significantly different from zero. In sum, the VAR models with two lags (in both the AOR and CDS equation) in columns 3 to 6 seem sufficient to capture the joint

Table 5
Results from the bank-specific vector autoregression analysis.
Lag length is selected using the Bayesian information criterion (SBIC). Missing observations are filled using the last observed price. Banks are ordered first according to the sign of the sum of AOR coefficients in the CDS equation, second according to the *p*-value of the corresponding Granger causality test.

Bank	H_0: AOR G-causes CDS		H_0: CDS G-causes AOR				Bank	H_0: AOR G-causes CDS		H_0: CDS G-causes AOR			
	Sum of coefficients	*p*	Sum of coefficients	*p*	*N*	Lags		Sum of coefficients	*p*	Sum of coefficients	*p*	*N*	Lags
1	0.419	0.000	0.314	0.000	784	5	31	0.038	0.843	0.053	0.610	601	2
2	1.807	0.000	1.815	0.863	181	3	32	0.098	0.856	0.107	0.446	811	2
3	0.190	0.000	0.217	0.615	632	3	33	0.006	0.869	0.047	0.169	806	1
4	0.353	0.000	0.418	0.287	497	2	34	0.012	0.906	−0.007	0.729	166	1
5	0.079	0.004	0.006	0.000	726	5	35	0.069	0.909	0.049	0.347	631	3
6	0.207	0.007	0.205	0.923	411	1	36	–	–	–	–	588	0
7	0.224	0.010	0.201	0.086	401	2	37	–	–	–	–	672	0
8	0.356	0.048	0.334	0.065	456	1	38	−1.269	0.001	−1.261	0.060	927	5
9	0.158	0.068	0.137	0.105	882	1	39	−0.150	0.017	−0.157	0.710	866	1
10	0.148	0.078	0.134	0.690	857	4	40	−0.321	0.027	−0.393	0.000	363	4
11	0.211	0.088	0.233	0.568	562	2	41	−0.302	0.040	−0.369	0.002	252	1
12	0.064	0.129	0.039	0.727	665	3	42	−0.239	0.160	−0.163	0.231	268	2
13	0.145	0.148	0.129	0.644	788	2	43	−0.076	0.207	−0.076	0.261	306	2
14	0.183	0.155	0.157	0.268	932	4	44	−0.512	0.211	−0.554	0.610	52	2
15	0.223	0.200	0.240	0.198	434	1	45	−0.557	0.254	−0.591	0.199	871	5
16	0.066	0.208	0.030	0.418	668	2	46	−0.248	0.255	−0.201	0.010	502	4
17	0.098	0.245	0.032	0.664	262	2	47	−0.036	0.263	0.144	0.186	178	1
18	0.076	0.310	0.011	0.000	659	1	48	−0.026	0.346	−0.045	0.738	846	3
19	0.077	0.318	0.146	0.421	142	1	49	−0.072	0.357	−0.088	0.666	646	2
20	0.100	0.319	0.119	0.027	929	2	50	−0.112	0.402	−0.112	0.898	459	2
21	0.020	0.329	−0.004	0.134	847	3	51	−0.039	0.436	−0.055	0.130	918	3
22	0.127	0.342	0.130	0.968	900	2	52	−0.063	0.437	−0.129	0.167	364	2
23	0.074	0.355	−0.003	0.204	688	2	53	−0.164	0.540	−0.140	0.028	802	3
24	0.028	0.377	0.025	0.938	562	1	54	−0.049	0.710	−0.040	0.913	790	2
25	0.141	0.394	0.116	0.290	237	1	55	−0.094	0.833	−0.081	0.737	797	3
26	0.155	0.463	0.068	0.081	171	2	56	−0.027	0.867	−0.176	0.091	265	2
27	0.135	0.537	0.103	0.031	575	2	57	−0.011	0.915	−0.063	0.450	755	2
28	0.044	0.554	0.053	0.703	541	1	58	−0.029	0.925	−0.076	0.138	848	2
29	0.104	0.655	0.086	0.354	756	2	59	−0.050	0.950	0.065	0.228	135	2
30	0.027	0.697	−0.017	0.485	360	2	60	−0.006	0.993	0.021	0.611	890	3

N is the number of observations, *t* denotes the *t*-statistic, and *F*-statistic and *p*-values are those for the Granger causality test.

dynamics of AOR and CDS, regardless of the approach adopted to deal with missing observations.[23]

Table 5 reports results for each bank separately with the lag-length selected by the SBIC information criterion.[24] As with bank-specific VEC model results in Table 3, there is a large variation across banks. In only a few banks AOR Granger causes CDS, with a positive sum of the lagged AOR coefficients in the CDS equation. The lag-length also varies with some banks having a higher lag-order than two. As already discussed, the heterogeneity of the bank-specific results suggests that the AOR's role in providing a measure of private information may not be a general phenomenon but concentrated in certain periods and certain banks, according to the information sensitivity hypothesis (*H2*). Therefore, we next refine our analysis by conditioning the Granger causality tests on market conditions and bank characteristics.

[23] As a further alternative approach, we also estimate the one-lag VAR system using weekly or monthly data frequency. We do this by forming weekly and monthly average observations, respectively, of the daily data. In this procedure, even if we simply ignore the missing daily observations, we obtain a sample, which is balanced across banks of all size. Especially with the monthly data, we get results, which are consistent with those obtained with daily data although with robust t-ratios the AOR's (monthly) lead over CDS is not significant. Moreover, the monthly results suggest that there is also a lead for CDS over AOR. The results are available from the authors upon request.

[24] In order to facilitate the bank-specific analysis with multiple lags, we replace missing observations with the most recent observation. Table 4 suggests that multiple-lag models estimated with the help of this simple method are similar to those obtained with the Kalman filter.

3.5. Testing for the information sensitivity hypothesis

Table 6 extends the panel results of Table 4 by focusing on the CDS equation in (6), and conditioning the coefficient on the lagged AOR on a number of dummy variables, D_i^j, which can also be interacted with one another (cf. Acharya and Johnson, 2007). The dummy variables proxy for the factors listed in hypotheses *H2: i) – v)* plus some additional controls. This gives us the following equation for CDS within the VAR system:

$$\Delta CDS_{i,t} = a_{11}\Delta CDS_{i,t-1} + a_{12}\Delta AOR_{i,t-1} + \sum_j b_{11}^j \Delta AOR_{i,t-1} D_{i,t-1}^j + \sum_j b_{12}^j D_{i,t-1}^j + c_1 + \varepsilon_{1,i,t} \tag{7}$$

where index *j* refers to the different dummies and index *i* to individual banks, as earlier.

Table 6a focuses on conditioning the coefficient on the lagged AOR on different phases of the crises. We use the following dummy variables:

"Pre Lehman (before 15 Sep 2008)": equal to one from the beginning of our sample period until 15 September 2008;

"Post Lehman (before 2010)": equal to one from 16 September 2008 until the end of 2009; and

"Sovereign Crisis (2010 onwards)": equal to one from 1 January 2010 until the end of our sample period.

In addition, we also consider the following two dummies indicating conditions of relatively stressed markets, measured by the iTraxx index, and the intensity of the ECB's liquidity operations, respectively:

Table 6
Conditionality of the daily lead-lag relationship.
The table reports results for the following CDS equation of panel VAR regressions for CDS, AOR and interactions of the lagged AOR with a variety of dummy variables (see Section III.E for definitions of the dummies):

$$\Delta CDS_{i,t} = a_{11}\Delta CDS_{i,t-1} + a_{12}\Delta AOR_{i,t-1} + \sum_j b_{11}^{j}\Delta AOR_{i,t-1}D_{i,t-1}^{j} + \sum_j b_{12}^{j}D_{i,t-1}^{j} + c_1 + \varepsilon_{1,i,t}.$$

We report coefficient estimates for a_{12} and $b^j{}_{11}$ ($a_{11}, b^j{}_{12}$ and c_1 are not reported). For brevity, the name of the dummy, e.g., “Pre Lehman (before 15 Sep 2008)”, stands for the respective interactive term, $\Delta AOR_{i,t-1}D^j{}_{i,t-1}$. The number of observations is 46,729 (44,398 if credit rating is used). Special dates (see Section II.A) are included. Inside the brackets are robust t-statistics adjusted for clustering.

(a) Conditioning on stressed time periods.

Variable	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t
ΔAOR_{t-1}	0.048	0.064	0.018	0.004	0.031	−0.009
	[3.01]	[2.98]	[1.21]	[0.29]	[1.67]	[0.55]
Pre Lehman (before 15 Sep 2008)	−0.005	–	–	–	–	–
	[0.18]					
Post Lehman (before 2010)	–	−0.0491	–	–	–	–
		[1.82]				
Sovereign Crisis (2010 onwards)	–	–	0.048	–	–	–
			[1.76]			
Higher iTraxx	–	–	–	0.081	–	0.078
				[2.83]		[2.71]
Higher excess reserves	–	–	–	–	0.041	0.035
					[1.33]	[1.15]

(b) Conditioning on bank characteristics.

Variable	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t
ΔAOR_{t-1}	0.018	0.002	0.025	0.031	0.024	0.025	−0.035
	[1.20]	[0.13]	[2.37]	[2.67]	[2.37]	[2.36]	[1.37]
Higher BPI	0.053	–	–	–	–	–	0.030
	[2.03]						[1.37]
Higher HHI	–	0.068	–	–	–	–	0.061
		[2.85]					[2.68]
Worse rating	–	–	0.052	–	–	–	0.032
			[1.56]				[1.15]
Domicile in GIIPS	–	–	–	0.038	–	–	-0.004
				[1.16]			[0.16]
Smaller bank	–	–	–	–	0.051	–	0.041
					[1.62]		[1.59]
Larger CDS bid/ask	–	–	–	–	–	0.046	−0.012
						[1.56]	[0.44]

(c) Conditioning simultaneously on stressed time periods and bank characteristics.

Variable	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t	ΔCDS_t
ΔAOR_{t-1}	−0.033	0.018	0.002	0.025	0.031	0.024	0.025
	[1.28]	[1.21]	[0.14]	[2.38]	[2.68]	[2.38]	[2.36]
Higher BPI	−0.009	−0.019	–	–	–	–	–
	[0.40]	[0.88]					
Higher HHI	0.050	–	0.011	-	–	–	–
	[1.98]		[0.51]				
Worse rating	0.018	–	–	−0.026	–	–	–
	[0.49]			[0.87]			
Domicile in GIIPS	−0.012	–	–	–	−0.035	–	–
	[0.44]				[1.32]		
Smaller bank	0.046	–	–	–	–	−0.019	–
	[1.51]					[0.74]	
Larger CDS bid/ask	−0.045	–	–	–	–	–	-0.030
	[1.02]						[1.12]
Higher BPI x Higher iTraxx	0.072	0.130	–	–	–	–	–
	[1.92]	[3.06]					
Higher HHI x Higher iTraxx	0.025	–	0.104	–	–	–	–
	[0.83]		[2.84]				
Worse rating x Higher iTraxx	0.009	–	–	0.128	–	–	–
	[0.17]			[2.29]			
Domicile in GIIPS x Higher iTraxx	0.008	–	–	–	0.123	–	–
	[0.18]				[2.24]		
Smaller bank x Higher iTraxx	0.012	-	–	–	–	0.123	-
	[0.25]					[2.24]	
Larger CDS bid/ask x Higher iTraxx	0.042	–	–	–	–	–	0.127
	[0.71]						[2.54]

"Higher iTraxx": equal to one on days when the iTraxx index is above its time-series median;[25] and

"Higher excess reserves": equal to one on days when the TARGET2 excess reserves exceed their time-series median (see section II.B).

Consistent with hypothesis *H2(i),* there is clear evidence in regressions 4 and 6 in Table 6a that AOR's ability to predict CDS depends on market stress, measured by the "Higher iTraxx" dummy. There is also some evidence that the effect is weaker during the period after Lehman's bankruptcy but before the escalation of the European sovereign debt crisis in 2010. This can be seen from the effective coefficient on the lagged AOR, which is smaller in regression 2 (conditioning on "Post Lehman (before 2010)") than in either regression 1 (conditioning on "Pre Lehman (before 15 Sep 2008)") or regression 3 (conditioning on "Sovereign Crisis (2010 onwards)") in Table 6a. This is consistent with the fact that soon after Lehman's bankruptcy the EU governments provided various forms of state guarantees to their banks (see e.g. Panetta et al., 2009) but that the subsequent European sovereign debt crisis questioned the solidity of these guarantees in many countries.

In Table 6b, we test hypotheses *H2(ii–v)* and condition the coefficient on lagged AOR on the following bank-specific dummy variables:

"Higher BPI": relative dependence on relationship lending, which is equal to one on days when a bank's BPI measure, defined in section II.B above, is above the daily cross-sectional median BPI;

"Higher HHI": relative concentration of borrowing, equal to one on days when a bank's HHI measure, also defined in section II.B, is above the daily cross-sectional median HHI;

"Worse rating": bank credit quality, equal to one on days when a bank's public credit rating is weaker than the daily cross-sectional median rating;

"Domicile in GIIPS": bank domicile in a crisis country;[26]

"Smaller bank": bank size, equal to one for banks whose total assets are below the cross-sectional median total assets; and

"Larger CDS bid/ask": illiquidity of a bank's CDS contract, equal to one on days when a bank's CDS bid-ask spread is above the daily cross-sectional median spread.

The interaction term between any of these bank-specific dummies and the lagged AOR obtains a positive coefficient but only the ones indicating relatively strong dependence on relationship lending ("Higher BPI") and relative concentration of borrowing ("Higher HHI") are significant, rendering the base coefficient insignificant (see regressions 1 to 6 in Table 6b). When we condition the coefficient on the lagged AOR on all the bank-specific dummies at the same time (see regression 7 in Table 6b), the interactive term with "Higher HHI" is the only statistically significant one.[27]

Finally, we consider the coefficient on the lagged AOR under double-interaction dummies. We multiply each dummy from Table 6b by the market stress dummy, "Higher iTraxx", from Table 6a. We then augment the regressions in Table 6b by these double-interaction terms. The results appear in Table 6c. Regressions 2–7 in Table 6c show that conditioning on the market stress dummy strongly increases the impact of the other conditioning dummy variables on the coefficient of the lagged AOR. All double-interaction terms are significant and render the "single-interaction" terms, introduced in Table 6b, insignificant. However, when comparing with the effective size of the coefficient on the lagged AOR in Table 6a, it is also clear that the other conditioning variables strengthen the effect of market stress. In regression 1 of Table 6c where all variables appear simultaneously, the double-interaction between market stress "Higher iTraxx" and the relative dependence on relationship borrowing, "Higher BPI", together with single-interaction term of relatively concentrated borrowing, "Higher HHI", emerge as the conditions under which the lagged AOR obtains a significant positive coefficient.

An (unreported) auxiliary regression shows that a weak bank rating and small bank size relate to a high value of the bank's BPI index. Further, we also find that the BPI index is on average higher for banks in crisis countries. Hence, although the BPI index helps explain much of the variation in information sensitivity of AOR, the more fundamental bank characteristics, such as credit quality, domicile, and size, in turn explain a bank's reliance on relationship lenders. This is in line with Cocco et al., (2009) who find that "smaller banks and banks with more nonperforming loans tend to have limited access to international markets, and rely more on relationships".

Overall, results in Tables 6a–c clearly support the information sensitivity hypothesis, *H2*.

3.6. Economic significance

In order to assess the economic significance of our results further, we provide a numerical example of the impulse response of CDS to a shock in AOR. One way to interpret this is that it demonstrates how much we can improve a forecast for the future CDS price, based on the current CDS price, by incorporating the private information in AOR and its dynamic relationship with CDS.

We use the panel VAR model with two lags reported in Table 4. The long-run impulse response of CDS for a 10 basis point shock in AOR is 0.6 basis points.[28] The VEC model in Eq. (5), which establishes a long-run relationship between AOR and CDS via the error-correction term, would give a much stronger impulse response. In this respect, the above numerical example based on the VAR model, demonstrating the economic significance of using AOR to predict CDS is conservative. Note also that impulse responses based on the panel coefficients hide the large variation in the dynamic relationship between AOR and CDS across banks and periods, which we have documented in Tables 3 and 5 and especially Table 6. Based on the conditional coefficients in Table 6, the impulse responses in the panel VAR would be 1.5–2 times bigger during periods of market stress for relationship-oriented borrower banks.

4. Conclusions

We have constructed a measure of a bank's relative creditworthiness, the average overnight borrowing rate in relation to EONIA index (AOR), by using Eurosystem's proprietary inter-bank overnight loan data. Because banks bilaterally agree the overnight loan rates, and any bank can have several lending relationships of this kind, the average overnight rate arguably aggregates lender banks' private information signals concerning the borrower bank's

[25] The median value for iTraxx turned out to be 140 bps, which is quite high because the period contains the financial crisis and the Euro crisis.

[26] A crisis country is defined as being one of the so-called GIIPS countries; Greece, Ireland, Italy, Portugal or Spain.

[27] As a measure of bank quality, we also consider a dummy for overnight loans for which the borrower bank pays a rate that exceeds the ECB's fixed rate tender rate. There are 1284 such observations in our data, and these probably indicate heightened bank risk (the filter used to extract overnight loan transactions from the TARGET2 raw data essentially leaves out overnight loans with a rate higher than in the ECB's marginal lending facility, which in principle could be another similar bank quality measure). We find (in unreported results) that the coefficient on lagged AOR is much larger for these banks.

[28] The standard deviation of $\varepsilon_{1,i,t}$ (see Eq. 6) is 6 basis points. The impulse response stabilizes after five days. The impulse response estimates are similar regardless of whether we use the two-lag model in Table 4 estimated with Kalman filter or by filling in the most recent observation for missing observations.

creditworthiness. Our main hypothesis is that because of the private information, a bank's AOR helps to predict its CDS.

Standard price discovery measures suggest that price discovery mainly takes place in the CDS market, but AOR also contributes to it. Granger causality tests in a panel framework further show that the first two lags of the differenced AOR help predict the differenced CDS. These results support the hypothesis that AOR includes private information, which the CDS market does not immediately incorporate.

We further identify periods and banks for which overnight loans have heightened information sensitivity, indicated by how strongly AOR Granger causes CDS. This is the case for periods of market stress and for banks, which mainly borrow from relationship lender banks. Such borrower banks are typically smaller, have weak ratings, and are likely to reside in crisis countries.

The paper provides rare evidence of the value of private information in predicting future market prices. The results should also be interesting to competent authorities who have access to the interbank rate data. They indicate that inter-bank overnight loan rates are a valuable source of information in monitoring changes in banks' health in addition to the information contained in the CDS market. Unlike CDS quotes, overnight borrowing rates data are available for practically all banks.

References

Abbassi, P. et al., 2014, Liquidity and relationship lending: evidence from the European interbank market, Discussion Paper No 45, Bundesbank.

Acharya, V.V., Johnson, T.C., 2007. Insider trading in credit derivatives. J. Financial Econ. 84, 110–141.

Affinito, M., 2012. Do interbank customer relationships exist? And how did they function in the crisis? Learning from Italy. J. Bank. Finance 36, 3163–3184.

Afonso, G., Kovner, A., Schoar, A., 2011. Stressed, not frozen: the federal funds market in the financial crisis. J. Finance 66, 1109–1139.

Angelini, P., Nobili, A., Picillo, C., 2011. The interbank market after august 2007: what has changed, and Why. J. Money Credit Bank. 43, 923–958.

Annaert, J., De Ceuster, M., Van Roy, P., Vespro, C., 2012. What determines Euro area bank CDS spreads? Financial Stab. Rev. 7, 153–169.

Arciero, L., Heijmans, R., Heuver, R., Massarenti, M., Picillo, C., Vacirca, F., 2016. How to measure the unsecured money market? The Eurosystem's implementation and validation using TARGET2 data. Int. J. Central Bank. 12, 247–280.

Arora, N., Gandhi, P., Longstaff, F.A., 2012. Counterparty credit risk and the credit default swap market. J. Financial Econ. 103, 280–293.

Berg, T., 2010. The term structure of risk premia - new evidence from the financial crisis. ECB Working Paper Series 1165 No.

Biais, B., 1993. Price formation and equilibrium liquidity in fragmented and centralized markets. J. Finance 48, 157–185.

Blanco, R., Brennan, S., Marsh, I.W., 2005. An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. J. Finance LX 2255–2281.

Bräuning, F. and F. Fecht, 2012, Relationship lending in the interbank market and the price of liquidity, Deutsche Bundesbank Discussion Paper No 22/2012, Deutsche Bundesbank.

Cocco, J.F., Gomes, F.J., Martins, N.C., 2009. Lending relationships in the interbank market. J. Financial Intermed. 18 (1), 24–48.

Commandeur, J.F., Koopman, S.J., Ooms, M., 2011. Statistical software for state space methods. J. Stat. Softw. 41 (1), 1–18 Volume Issue.

Covitz, D., Downing, C., 2007. Liquidity or credit risk? The determinants of very short-term corporate yield spreads. J. Finance LXII 2303–2328.

Dang, T.V., G. Gorton and B. Holmström, 2015, Ignorance, debt and financial crises. Available at: http://www.columbia.edu/~td2332/Paper_Ignorance.pdf

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm". J. Royal Stat. Soc. Series B 39 (1), 1–38.

Forte, S., Peña, J.I., 2009. Credit spreads: An empirical analysis on the informational content of stocks, bonds, and CDS. J. Bank. Finance 33, 2013–2025.

Fung, H-G, Sierra, G.E., Yau, J., Zhang, G., 2008. Are the U.S. stock market and credit default swap market related. J. Altern. Invest. 11, 43–61.

Furfine, C.H., 2001. Banks as monitors of other banks: Evidence from the overnight Federal Funds market. J. Bus. 74 (1), 33–57.

Garmaise, M.J, Moskowitz, T.J., 2004. Confronting information asymmetries: evidence from real estate markets. Rev. Financial Stud. 17 (2), 405–437.

Giannikos, C.I., Guirguis, H., Suen, M., 2013. The 2008 financial crisis and the dynamics of price discovery among stock prices, CDS spreads, and bond spreads for US financial firms. J. Deriv. 21 (1), 27–48.

Granger, C.W.J., 1998. Extracting information from mega-panels and high-frequency data. Statistica Neerlandica 52 (3), 258–272.

Gropp, R., Vesala, J., Vulpes, G., 2004. Market indicators, bank fragility, and indirect market discipline. Econ. Policy Rev. 22, 53–62.

Gropp, R., Vesala, J., Vulpes, G., 2006. Equity and Bond market signals as leading indicators of bank fragility. J. Money Credit Bank. 38, 399–428.

Kremers, J.J.M, Ericsson, N., Dolado, J.J., 1992. The power of cointegration tests. Oxf. Bull. Econ Stat. 54 (3), 325–348.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53 (6), 1315–1335.

Longstaff, F.A., Mithal, S., Neis, E., 2005. Corporate yield spreads: default risk or liquidity? New evidence from the credit. J Finance LX 2213–2253.

Norden, L., Weber, M., 2009. The co-movement of credit default swap, bond and stock markets: an empirical analysis. Eur. Financial Manag. 15, 529–562.

Panetta, F., Faeh, T., Grande, G., Ho, C., King, M., Levy, A., Signoretti, F., Taboga, M., Zaghini, A., 2009. An Assessment Of Financial Sector Rescue Programmes, Bank for International Settlements, BIS Papers No. 48.

Schwarz, K., 2014. Mind The Gap: Disentangling Credit And Liquidity In Risk Spreads. Department of Finance, The Wharton School, University of Pennsylvania.

Article IV

Tölö, E., Jokivuolle, E., and Virén, M., 2019, "Have Too-Big-To-Fail Expectations Diminished? Evidence from the European Overnight Interbank Market," *Journal of Financial Services Research*, forthcoming.

Working paper version available at:
https://helda.helsinki.fi/bof/handle/123456789/16351.

Have Too-Big-To-Fail Expectations Diminished?

Evidence from the European Overnight Interbank Market*

Eero Tölö[a,b], Esa Jokivuolle[a], and Matti Viren[a,c]

[a] Bank of Finland, Snellmaninaukio, P.O. Box 160, 00101 Helsinki, Finland

[b] University of Helsinki, Helsinki GSE, P. O. Box 17, 00014 Helsinki, Finland

[c] University of Turku, Turku School of Economics, 20014 Turku, Finland

Current version 9 Mar 2021

Abstract

Using the Eurosystem's proprietary interbank loan data from June 2008-June 2020, we show that larger European banks have had a lower cost of overnight borrowing than smaller banks. The size premium remains significant after controlling for a large set of other factors but has decreased over time, especially in countries that were stricken by the Sovereign Debt Crisis. A difference-in-differences analysis suggests that the decline in the size premium is related to the actual bail-in events, not to the implementation dates of the Bank Recovery and Resolution Directive as such. This finding is robust to controlling for the effect of the ECB's long-term refinancing operations. Overall, the results suggest that the regulatory move towards bail-in rather than bailout policies to deal with financially distressed banks has reduced the too-big-to-fail expectations concerning large banks.

JEL classification: G21, G22, G24, G28
Keywords: overnight rates, too-big-to-fail, bail-in, bailouts, implicit government guarantee, interbank borrowing costs, Bank Recovery and Resolution Directive

*Esa Jokivuolle (corresponding author, esa.jokivuolle@bof.fi), Eero Tölö (eero.tolo@bof.fi) and Matti Virén (matti.viren@bof.fi) are employed by the Bank of Finland. Eero Tölö is a member/alternate of one of the user groups with access to TARGET2 data in accordance with Article 1(2) of Decision ECB/2010/9 of 29 July 2010 on access to and use of certain TARGET2 data. The Bank of Finland and the MIPC have checked the paper against the rules for guaranteeing the confidentiality of transaction-level data imposed by the MIPC pursuant to Article 1(4) of the above-mentioned issue. The views expressed in the paper are solely those of the authors and do not necessarily represent the views of the Eurosystem or the Bank of Finland. We would like to thank George Pennacchi, Iftekhar Hasan, Jouko Vilmunen, an anonymous referee, and seminar participants at the Bank of Finland for valuable comments. All remaining errors are ours. Earlier version of this work has been circulated as "Are too-big-to-fail banks history in Europe? Evidence from overnight interbank loans", Bank of Finland Discussion Paper No 29/2015.

1. Introduction

We study pricing in the European interbank overnight loan market with a particular focus on whether larger banks still benefit from the implicit ”too-big-to-fail” subsidy after the regulatory move towards “bail-in” instead of “bailout” policies to deal with financially distressed banks, especially through the implementation of the Bank Recovery and Resolution Directive (BRRD). We use an extensive confidential dataset from Europe, the Eurosystem’s proprietary transactions data from which the volumes and counterparties as well as interest rates on interbank loans can be filtered out for a large representative set of European banks in our sample period June 2008-June 2020. The representative size distribution of sample banks is a central advantage of this data as we evaluate importance of the too-big-to-fail expectations, measuring them by banks’ size premia in the interbank borrowing rates. We focus on overnight loans which are by far the most liquid segment of the market.

We study three sets of factors that may drive the interbank loan pricing. First, as controls we include variables related to banks’ risks as pioneered in overnight loans context by Furfine (2001), banks’ liquidity positions, and banks’ use of the ECB’s standing facilities in their liquidity management. Second, our confidential data set allows us to directly observe the history of loan sizes as well as the number and identity of interbank counterparties of a bank. In addition to being potential explanatory variables of interbank loan rates as such, we also use these variables to construct measures of interbank lending relationships (see Cocco et al., 2009 and Tölö et al., 2017) as well as measures of bank connectedness and competitive position in interbank markets. Third, we use bank size (as well as a bank’s interbank connections) as a simple measure of a bank’s systemic risk. We then study changes in its role in explaining overnight loan rates around the legal implementation of the BRRD in the European Banking Union’s member countries, taking advantage of the variation in the country-specific implementation dates. Similarly, we study the pricing effect of actual bail-in decisions. In particular, we test whether the introduction of the BRRD legislative framework per se and the actual bail-in events have reduced market expectations of the largest banks’ alleged too-big-to-fail status. As in much of the relevant literature, we measure the strength of too-big-to-fail expectations by magnitude of the coefficient of the bank size variable. Furfine (2001), who refers to Rochet and Tirole (1996), was early to touch upon the too-big-to-fail issue in the context of overnight loan pricing: “if the largest banks were viewed as immune from failure, then these banks' creditors, including other banks, would have little incentive to monitor their exposures."

A central aim of the BRRD is to facilitate bank restructurings without public bailouts. Hence, if the BRRD framework and its implementation is considered credible by the markets, then we should expect a reduction in the role of bank size as a factor related to bank systemic risk in explaining interbank loan rates. One caveat in our setting is that overnight interbank loans are not included in the first line of defense of bank liabilities in a bank's restructuring; i.e., they are not immediately available for "bail-in" (see the European Banking Authority's Interactive Single Rulebook on Bank Recovery and Resolution Directive). However, even the safest bank liabilities could become subject to "runs" and roll-over risk when uncertainty is high and creditors want to avoid any delays in getting their money back. They may also want to avoid reputational losses from doing business with counterparties which the market suddenly views excessively risky. Beside outright rationing lenders may require collateral to reduce the cost of monitoring. However, there is evidence that even the repo market is subject to runs (Gorton and Metric, 2009), and that secured securities may become sensitive to borrower's credit risk (Dang et al., 2015). Hence, it is possible that the introduction of the BRRD indirectly affects all interbank liabilities, also perhaps partly due to imperfect information among market participants of the detailed institutional features of the resolution framework. For these reasons, we find it well-grounded to test the effect of the BRRD implementation and actual bail-in decisions on overnight loan pricing.

Further, because of their maturity of "one day", overnight loan rates should not reflect any market expectations of the effects of BRRD even just before the date when the BRRD is implemented in each banking union member country. Therefore, the daily overnight loan rates are in principle well-suited to a difference-in-differences analysis of the effect of BRRD on treated and non-treated banks in our sample around a tight event window and can therefore well avoid potential simultaneity problems.

However, one may well argue that implementation of the BRRD is a longer process which constitutes a paradigm shift rather than a clear-cut legislative change. After the formal introduction of the BRRD there may have still been considerable uncertainty as to how the new regulation will be implemented in case of bank failures. Therefore, we also include the actual regulatory bail-in decisions in our difference-in-differences analysis, using the list of events and their dates from Schäfer et al. (2016) and Bellia and Maccaferri (2020). This is further motivated by findings such as those of Schäfer et al. (2016) that market expectations regarding implicit public support to too-big-to-fail institutions are mainly affected by actual bail-in decisions rather than the legal changes that make bail-in the norm. We also include the ECB's major long-term refinancing operations (LTRO) as controls in the difference-in-differences analysis as they may have had a significant impact on pricing in the overnight market in the period under study.

To measure banks' borrowing costs, we use European banks' interbank overnight (uncollateralized) borrowing rates from June 2008 to June 2020, which includes both EU and non-EU banks and other financial institutions (henceforth we refer to all institutions in the data as "banks", for brevity, unless noted otherwise). There are more than 2,500 borrower banks in our original sample. We use the difference between the loan-level overnight borrowing rate, constructed from the transactions data, and the EONIA rate as our measure of a bank's *relative* borrowing cost (for simplicity, we henceforth refer to this measure as the overnight borrowing cost of a bank).

We find that larger European banks have had a lower cost of overnight borrowing than smaller banks. The size premium remains significant after controlling for relationship lending, competitive environment of lenders, loan size, bank risk characteristics, ECB standing facilities, and various market factors. It has decreased over time for large banks especially in countries that were stricken by the Sovereign Debt Crisis. These results suggest that the BRRD may have helped to reduce the too-big-to-fail expectations concerning large banks. However, in the difference-in-differences analysis we find that the change in the size premium cannot be timed to the implementation dates of the BRRD legal framework in different member countries. Yet, we find that on average the actual bail-in events, even when controlling for the ECB's LTRO implementation events, have significantly reduced the size premium. These findings are in line with those of Schäfer et al. (2016) that indeed, "actions speak louder than words" in impacting the too-big-to-fail expectations.

The rest of the paper is organized as follows. Section 2 provides a brief review of the related previous literature. Section 3 provides an overview of the Eurosystem's interbank market. Section 4 introduces the data and section 5 presents the results. Section 6 concludes.

2 Literature review

Prior empirical work on the effect of introducing resolution frameworks such as the BRRD, has been reviewed e.g. in Berger et al. (2019); see also Dell'Ariccia et al. (2018). Berger et al. (2019) themselves find empirical evidence, based on a theoretical model, of changes in the US banks' capital ratios, which is consistent with the move towards more bail-in rather than bailout oriented policies. They also cite several studies finding that bail-ins strengthen market discipline and may reduce systemic risk but may also cause undesirable effects such as increased stock market volatility (see Berger et al., 2019, footnote 7). Schäfer et al. (2016) find that actual bail-in decisions have had more impact on market expectations than the legal implementation of the BRRD. More recent evidence in the European context is presented by Bellia and Maccaferri (2020) whose list of bail-in related events we also utilize in our empirical analysis. Using structural Merton-style models to estimate market-implied bailout and bail-in probabilities for banks Berndt et al. (2018) estimate that there has been a significant decline in the market-based too-big-to-fail expectations for the US systemically important banks (similar work in the international context is in progress by Guennewig and Pennacchi, 2019).

The issue with the too-big-to-fail (henceforth TBTF) expectations as a possible explanation of large banks' size premium in their borrowing costs dates from long before the recent financial crisis. It has been subject to intensive research especially in the United States (see e.g. Brewer III and Jagtiani (2013), Hughes and Mester (2013), Noss and Sowerbutts (2012), Ueda and Weder di Mauro (2013), Wheelock and Wilson (2012), and Santos (2014).

The previous literature has used several alternative measures to study the relationship between banks' borrowing costs and bank size such as deposit rates, bond returns, CDS rates, or the weighted average of these various borrowing costs. Combinations of rating and rate-based measures as well as equity returns have also been used in studying the effect of bank size on its financing costs more generally. For example, Bassett (2014) examines differences in the cost of deposits of large and small banks, Demirguc-Kunt and Huizinga (2013) use CDS rates, and Santos (2014) uses bond spread differences between small and large banks. Kroszner (2013) provides a comparative review of results obtained with alternative measures.

Empirical evidence on the relationship between bank size and bank borrowing costs seems to depend on the time period and the measure of the borrowing cost. The CDS data which are perhaps most often used suggest that prior to the global financial crisis financial firms generally had lower spreads but that they were less sensitive to firm size than spreads for several other industries (see Ahmed et al., 2015). This result appears to be consistent with the theory of Farhi and Tirole (2012) that many of the public sector's financial crisis fighting measures, including monetary policy operations, help the financial sector as a whole, not just the largest financial institutions. So, measuring the TBTF advantage of large banks from borrowing cost differentials of large and small banks may underestimate the full size of implicit government support to the financial sector.

Large banks seem to have enjoyed a borrowing cost advantage especially prior to the Lehman Brothers bankruptcy (cf. Ahmed et al., 2015). In contrast, the study with deposit rates of large and small banks (Bassett, 2014) fails to show very strong evidence of the perceived TBTF subsidy (cf. also O'Hara and Shaw, 1990). The result applies not only to ordinary deposits but also to interest bearing liquid deposits. Overall, the differences in deposit rates between small and large banks appeared to be small. This may well be explained by deposit insurance. Acharya et al. (2016) provide recent evidence that the largest banks' bond spreads are less risk-sensitive than those of smaller banks, which is consistent with the TBTF hypothesis.

One potential difficulty in measuring the TBTF advantage from borrowing cost differentials is that bank size and its borrowing costs may be related also for other reasons than TBTF expectations. Most recent studies suggest that there are scale economies in banking even for the largest group of banks for various reasons (cf. e.g. Hughes and Mester,2013 and Araten and Turner, 2013). However, it is possible to try to control for the effect of efficiency via bank profitability measures such ROE (as we do in the current paper). Moreover, Davies and Tracey (2014) show that what may seem like cost efficiency may be a cost advantage arising from a TBTF status: after controlling for the TBTF factor, they find that the cost efficiency of the largest banks vanishes. Interestingly, Penas and Unal (2004) provide evidence that gains to merging banks' bond holders primarily stem from diversification benefits, strengthening too-big-to-fail status, and only to a lesser degree, synergy gains.

Another potentially important factor that affects the overnight borrowing cost of a bank is the competitive environment in the interbank market. For instance, a bank may be relatively dependent on a few relationship lenders, or it can be a hub in a network of banks with a multitude of potential sources of interbank funds (cf. e.g. Cocco et al., 2009, and Brauning and Fecht, 2017). The advantage of our data is that it allows us to control for the number of interbank relationships. However, as already discussed above, the number of interbank relationships that we use could also proxy for a bank's systemic risk arising from its interconnectedness in the financial network. Interconnectedness can hence give rise to an additional source of market expectations of government support, not fully captured by bank size. Ringe and Patel (2019) argue that banks' interconnectedness can make almost the entire banking system too risky to fail. They also point out that the BRRD itself can increase de facto interconnectedness of banks over time and have adverse consequences in terms of the market structure (this issue is further discussed in Martino, 2020).

Various studies, including Furfine (2011) and Cocco et al. (2009), report that large banks obtain interbank funding at a lower cost (see also Afonso et al., 2011). Based on a dataset of around 100 banks that trade on the Italian based money market trading facility (E-MID), Angelini et al. (2011) report that the overnight loans become sensitive to borrowers' creditworthiness and size following August of 2007 (also later after the failure of Lehman Brothers) and suggest that moral hazard risks related to TBTF have increased. Akram and Christophersen (2017) use Norwegian overnight loan data of 28 banks over the period 2006–2009 and find relatively lower funding costs for banks of systemic importance, but only in some periods. Relative to these studies, our contribution is a substantially more extensive dataset in terms of banks and time span and a detailed analysis of the BRRD implementation.

Finally, Jacewitz and Pogach (2018) resembles our study in certain important respects. They show that larger US banks have enjoyed lower Money Market Deposit Account (MMDA) rates and risk premia than smaller banks for the uninsured part of these deposits and attribute this difference to the TBTF advantage of the larger banks. They further show that the advantage largely vanished when the limit on the deposit insurance coverage was raised in the midst of the 2008 financial crisis. Hence, like in our study, their focus is on bank funding with effectively very short maturity (presuming that MMDAs can be withdrawn relatively quickly), and like us, they demonstrate the funding cost advantage of larger banks and attribute that to TBTF expectations by showing the sensitivity of the cost advantage to a regulatory change affecting these expectations.

3 Recent developments in the Eurosystem's interbank money market

In this section we give a brief overview of the functioning of the Eurosystem's interbank money market. The euro area monetary policy operations as well as the majority of transactions in the euro area interbank market are settled in the TARGET2 system, which is the large value payment system of the Eurosystem. Access to TARGET2 is granted primarily to EU central banks and their national communities of commercial banks. Money market transactions are a subset of bank-to-bank large value payments. The great majority of bilateral loans are negotiated over the counter and hence (in the absence of any transparency regulation for these loans) are known only to the two parties involved in each transaction. Payments are settled in central bank money with immediate finality (i.e., in real time). In 2012, TARGET2 had a 92% market share in value terms of all large value payments in euro (see European Central Bank 2013). TARGET2 and Fedwire Funds for the US dollar are the two largest real-time gross settlement systems in the world.

The overnight maturity is by far the most traded segment in the unsecured money market. The Euro money market has the following specific features with potential implications for the analysis of the overnight rates data. First, eligible counterparties have everyday access to the ECB's marginal lending facility and the deposit facility. These banks need central bank money to fulfil their minimum reserve requirement at the Eurosystem. The marginal lending facility offers relatively expensive overnight funding against eligible collateral, and the marginal lending facility rate set by the ECB governing council acts as an (not exactly strict) upper bound for the uncollateralized overnight loan rates. The deposit facility offers a minimum return for overnight deposits, and the deposit facility rate set by the ECB acts as a (not exactly strict) lower bound for the overnight loan rates. The third important rate is that of the ECB's weekly main refinancing operations which also may have an effect on the interbank overnight rates through general liquidity conditions. This rate is typically half-way between the deposit and marginal lending rate (the "corridor"), and until October 2008 the average overnight rate (Euro Over Night Index Average, i.e., EONIA) was very close to this rate (see Figure 1). Since 15 October 2008, the weekly main refinancing operations have been carried out through a fixed rate tender procedure with full allotment at the interest rate on the main refinancing operations. This and the ECB's subsequent Longer-Term Refinancing Operations (LTRO) (notably 3-year operations in December 2011, scant €500bn, and in February 2012, over €500bn) as well as the asset purchase program since 2015 have significantly increased the amount of excess central bank liquidity in the TARGET2, and have moved EONIA closer to the deposit facility rate (see Figure 1). Following the large-scale asset purchase program, substantial part of unsecured overnight loan transactions is related to non-banks and foreign banks that make overnight deposits to banks in the euro area banking system who have access to the Eurosystem's deposit facility. Such transactions were often priced below the deposit facility rate (ECB, 2019). This is reflected in the TARGET2 filtered average overnight rate (see Figure 1), which starting from 2016 is slightly below EONIA (average lending rate reported by panel of euro area banks). The fixed tender rate of weekly main refinancing operations acts as a soft upper bound on the interbank rates and a large proportion of overnight loans take place between deposit facility rate and the fixed rate tender rate. A bank would be willing to borrow at a rate above the fixed tender rate only if it has no access to the ECB facilities, it has no available collateral, or it is concerned of reputational costs that could arise from borrowing from the central bank.

For three years following the 2008 financial crisis, the volume of unsecured overnight lending has followed a declining trend. According to Euro money market study (ECB, 2019) the declining trend is caused by aversion to counterparty credit risk, ample liquidity provided by the Eurosystem, and the new Basel III net stable funding ratio and liquidity coverage ratio requirements. At the same time secured overnight lending has remained at the 2007–2008 levels. Along with this development, bilateral repos increasingly take place through Central Counter Party (CCP) clearing while the share of triparty repos has remained largely unchanged (ICMA, 2018).

The drop in unsecured lending volume is observed also in our transaction level money market dataset (to be discussed in next section). Figure 2(a) shows the total transaction volume of the filtered overnight loans where periods of low volume match well with periods of high excess liquidity. Similar decline is observed for the number of daily borrowers and lenders in Figure 2(b). Reduction in market and excess liquidity as such should not affect banks' assessment of their counterparty credit risks. However, if small banks that pay higher rates have a higher propensity to exit the overnight market, this could have a gradual impact on the size premium. We discuss this possibility further in robustness checks (Section 5.4). In our empirical analysis (Section 5.1), we control for the possible effect that declining volumes coincide with reduced incentives to monitor and price in counterparty credit risk by subsample analysis.

Moreover, it is important to note that information sensitivity of ultra-short credit contracts like overnight loans are likely to be tied with general market conditions. Tölö et al. (2017), using the same overnight market data from 2008 until 2013, find that in periods of higher financial market stress the relative contribution of overnight markets to price discovery concerning counterparty risk is strengthened in relation to the CDS market.

The above discussed developments in euro money market could also have affected other specific aspects of the market such as relationship lending and roll-overs. About 10 percent of transactions are related to relationships that continue uninterrupted for 10 or more days. If we split the sample at the end of 2012 and compare the first half to the second, this share has gone up from 9 percent to around 14 percent. Hence, the relationships have on average become more persistent. We have further estimated the share of loans that are rolled over for the next N trading days. Roll-overs are interesting to monitor as they could lead to lower speed of price discovery. In the data overnight roll-overs are relatively rare and typically short-lived. If we compare again the first half of our sample to the second half (the results are available upon request), the share of estimated one-day roll-overs has increased from 2 to 5 percent. However, the share of transactions that are rolled over more than twice remains low; below 0.5 percent. All in all, these observations further corroborate the view that the overnight interbank market has continued to work quite normally despite banks' dramatically eased liquidity conditions thanks to central bank crisis measures.

4 Data

We use a rather novel dataset that comprises most of the global euro-denominated interbank overnight market for our sample period from June 2008 to June 2020. A central reason for focusing on the shortest (overnight) interbank loan maturity is the high liquidity of this market segment. The data are based on identification of transactions from the TARGET2 payment system, using improved version of the method described in Arciero et al. (2016). The data quality is analyzed in detail in Arciero et al. (2016), which also considers longer maturities. The quality is found to be very good for the overnight loans with false rates below 2 % for the overnight segment (cf. Armantier and Copeland, 2012). In the improved version, the identification of the counterparties is further improved based on the originator and beneficiary fields of the transaction message. As a data quality check, Figure 1 shows the EONIA rate and the average overnight rate calculated from the identified transactions. Most of the difference between the two series is explained by the differing sample and our filtered overnight rate is in fact more closely related to the preliminary Euro Short-term Rate reported in Euro Money Market Study (ECB, 2019). Further differences could potentially result from some bilateral non-CCP-cleared overnight repo transactions still remaining in the data. To ensure robustness of our results in this regard, we have identified the subset of loans that are relatively more likely to be secured by collateral. The results are robust against excluding such transactions.

Overall, the overnight loan data are quite comprehensive especially in terms of the number of banks, and it includes banks of all sizes. For example, the sample is much larger than in studies that use CDS data which are usually available only for large banks, or in studies based on banks with public credit ratings. The representative size distribution of sample banks is a central advantage of this data as we evaluate importance of the TBTF expectations, measuring them by banks' size premia in the overnight loan rates. The sample period includes the key episodes of the recent crises in Europe. This allows us to assess potentially different regimes of perceived public support to large banks.

Besides the borrowing rate and transaction size, the transaction data provide us with a number of other variables related to the bank's status in the overnight lending market. We calculate an activity measure based on how many days the bank borrows during the on-going reserve maintenance period. We use the number of lenders to proxy for a bank's bargaining position. We adapt from Cocco et al. (2009) the two relationship lending proxies Borrower Preference Index (BPI) and Lender Preference Index (LPI), which measure the strength of a borrower's relationship with certain lender and a lender's relationship towards certain borrower, respectively. As an alternative relationship lending proxy, we use the number of days within a given period that the lending relationship was active adapted from Furfine (2001). We also consider fixed effects at the borrower, lender, and the borrower-lender pair level.

Further, we also include the following three transactions-based measures. The first probes the geographic reach of a given bank. This index is calculated based on a bank's country distribution of incoming customer payments. We calculate the value share of customer payments originating from each country and aggregate the shares into a Herfindahl index. It can be interpreted as the geographic diversification of the bank's incoming liquidity stream related to customers (often large corporations). The two other measures are based on banks' transactions with the marginal lending facility and the deposit facility. They measure how actively a bank uses each of these facilities during a given reserve maintenance period calculated as the fraction of days (over a reserve maintenance period) that the bank uses the facility. Additionally, we include a dummy that distinguishes which banks have access to the ECB facilities. In robustness checks, we also consider various overnight market level variables including daily number of borrowers, daily concentration of borrowers measured by the Herfindahl index, and excess liquidity.

The transactions-based data are combined with data from Orbis Bank Focus using the bank identifier codes (BICs). We use the Orbis specialization information to restrict the sample of borrowers to bank holdings, holding companies, commercial banks, cooperative banks, investment banks, real estate and mortgage banks, and savings banks. The Orbis data includes typical proxies for the bank's credit risk: ROA, Tier 1 ratio and NPL ratio. Additional controls include revenue share of non-interest income, ratio of liquid assets to deposits and loans-to-deposits ratio. As a further control for whether bank size could be a proxy for other characteristics than systemic importance, such as informational transparency, we also include indicators for G-SIFI banks and listed banks as alternative measures of banks' systemic risk. We also obtain the iTraxx CDS index and sovereign CDS spreads from Bloomberg and Thomson Reuters, respectively. The final number of banks in various model specifications ranges from a few hundred up to 1,397 depending on which geographic regions and control variables are included.

Our (yearly) balance sheet data is of much lower frequency than the market-based variables. Therefore, we use values reported for the previous year when matching it with the daily data.[1] In the empirical analysis, we use robust standard errors clustered at the bank level to ensure more reliable estimates with the mixed frequency data. In analyses in which the time period is short (a few weeks) as in the difference-in-differences analysis, we use time-series averages of the daily data to avoid overweighting any single bank. In addition, when the short time window of analysis includes a new year, we use constant values for balance sheet data corresponding to that end of the year that is changing (for example from 20 December 2015 to 10 January 2016, we use end of 2015 balance sheet data throughout). Table 1 provides the descriptive statistics. The data sources and definitions are available in Table 7 in the Appendix.

[1] The data are nominal but deflating the series by the CPI does not make any noticeable difference in results.

5 Empirical results

5.1 Basic results

We start by studying the effect of bank size and other determinants on the bank's borrowing cost in the overnight interbank market using the following baseline panel equation:

$$r_{ijt} = \alpha_t + \beta_1 \log Z_{it} + \sum_{k=2}^{K} \beta_k X_{ijt}^{(k)} + \varepsilon_{ijt} \qquad (1)$$

The dependent variable r_{ijt} is the difference of the interest rate on an overnight loan, borrowed by bank *i* from bank *j*, and the EONIA rate on the same day, *t*. Note that the number of loans can vary across banks and also across time for each bank. The explanatory variable of our main interest is $\log Z_{it}$; the natural logarithm of the book value of total assets of bank *i*. $X_{ijt}^{(k)}$ are control variables, which may depend both on the borrower and the lender, and α_t captures fixed time effects. Equation (1) is estimated in the standard panel setting using an unobserved cluster effect so the standard error estimates are adjusted according to the method pioneered by Huber (1967) and Rogers (1993). All reported regressions except for the difference–in–differences analysis are based on Equation (1).

One could also consider using the difference between the overnight loan's interest rate and the overnight market interest rate in relative terms as the dependent variable. Here we cannot really do that because the market rate (the denominator) is close to zero or even negative for a large part of the sample. However, we estimated equation (1) for the pre-2012 part of the data so that the dependent variable was the relative difference of the overnight loan interest rate with respect to EONIA. The results (available from the authors upon request) are qualitatively very similar so for the rest of the paper we use the absolute difference specification. It should also be noted that in many standard theoretical models such as Duffie and Singleton (1999) the credit spread, which we also aim to measure, is defined as the absolute difference of the credit risky rate and the benchmark rate. The credit spread could depend on the level of the benchmark rate (see e.g. Duffee, 1998) but their theoretical relationship is not necessarily straightforward.

Table 2 provides the basic results, considering nine different model specifications. The number of banks (and hence observations) varies somewhat between specifications depending on the availability of data on the control variables. The main result in Table 2 is that bank size (the log of total assets) is negatively related to a bank's overnight borrowing cost. The coefficient is significant and stable across all model specifications as we add more controls and vary the type of fixed effects specification both in time and cross-section respect. Specifications (7) to (9) include fixed effects at the level of borrower, lender, or the bank relationship. The results are robust in terms of the fixed effects' menu suggesting that the outcome does not only reflect some common trend or level differences. With few exceptions, the control variables obtain expected signs and are stable across the different model specifications (and consequently, sample sizes). The key results remain qualitatively unchanged also if we take annual time-series averages of the daily data to match the true updating frequency of variables observed at lower-frequency such as bank balance sheet variables (the results are available from the authors upon request). In sum, our base results for the whole sample period in Table 2 strongly suggest that larger banks have had a cost advantage in overnight borrowing in interbank markets, which could partly be explained by the implicit government subsidy they enjoy. In what follows, we study how this relationship has evolved over time and whether it has responded to regulatory changes or actions aiming to limit that subsidy.

Before that, we will first gauge the role of control variables appearing in Table 2 in more detail. Bank profitability measured by return on assets (ROA) is negatively related with the overnight borrowing cost (we have also checked that return on equity, ROE, would yield a similar result). Bank solvency measure, the Tier 1 ratio (Tier 1 equity divided by risk weighted assets) is also negatively related with the overnight borrowing cost. The share of non-interest income of total revenue of a bank is negatively related to the borrowing cost, which suggests that more diversified revenue sources tend to reduce bank asset risks. All these relationships are quite stable and largely maintain their statistical significance across different model specifications.

The sovereign CDS spread of a bank's home country is positively related to the bank's overnight borrowing cost. In other words, a higher creditworthiness of the government (i.e., lower CDS spread) tends to lower the borrowing cost of banks domiciled in that country. The country fixed effects also explain a significant amount of the variation in overnight borrowing costs. However, the ratio of non-performing loans (NPL ratio), which is included as a measure of bank asset risk does not obtain a significant coefficient. Coefficient on the ratio of liquid assets to deposits has the expected sign but is statistically significant only in specification (4).

We also include a dummy for listed banks to control for the possibility that bank size could be a proxy for bank transparency. In other words, more transparent banks, such as those listed arguably are, might suffer less from problems of asymmetric information and hence obtain cheaper overnight funding. Counter to expectations, however, we find the opposite effect which is significant in three of the four specifications in which the indicator for listed banks is included. This result is probably explained by the fact that large size and being listed partly proxy for the same TBTF expectations: in an unreported regression we drop the bank size variable from the model, in which case the dummy for listing status obtains the expected negative coefficient that is significant.

Further, assuming that banks in the same country are better informed of one another than of their cross-border counterparties, we include a dummy for cross-border loans. It obtains a significant positive coefficient which is stable over different specifications, indicating that cross-border counterparties are indeed charged more, perhaps because of their informational disadvantage relative to domestic counterparties.

We also include a dummy for global systemically important financial institutions (G-SIFI), which obtains a positive coefficient, although it is insignificant in half of the specifications considered (see models 4-8 in Table 2). The positive coefficient is counter to the notion that the G-SIFI status would imply higher expected government support in a crisis and hence lower a G-SIFI bank's borrowing costs. However, like in the case of the dummy for listed banks earlier, the probable explanation to this result is that if there are such market expectations concerning G-SIFI banks, they are already captured by the bank size variable. In an unreported regression this is what we find: after dropping the bank size variable the G-SIFI dummy obtains a statistically significant negative coefficient, also when the dummy for listed banks is included in the regression. The results suggest that care should be taken when using the G-SIFI status as a proxy for TBTF expectations as bank size may subsume the effect.

Our data also allow us to include (the log of) transaction size as a control; see models 4-9 in Table 2. It obtains a significantly positive coefficient which is quite stable across the different models. The positive sign suggests that lender banks can charge a higher interest rate on bigger loans (recall that borrower bank size is the key explanatory variable of interest and hence controlled for). This possibly reflects expectations of increased credit risk or acute liquidity need of the borrower bank.

Further, the variable (log of) loan count measures how many loans the borrower took during the on-going reserve maintenance periods. This variable also obtains a positive and significant coefficient, which suggests that banks that are often short of liquidity are willing to pay a higher price. The same effect is observed for determinants that measure a borrower's use of the ECB standing facilities. We find that banks with excess liquidity that deposit to the ECB deposit facility pay a lower rate for their overnight loans. In contrast, those who more frequently borrow from the marginal lending facility pay a higher rate. Both coefficients are statistically significant for all specifications considered; see model specifications 5-9 in Table 2. In these same specifications we also observe that banks that are subject to reserve requirement and thus have access to the ECB facilities (indicated by variable "Subject to RR" in Table 2) get a discount as the respective coefficient is negative and significant. Such a discount may be reflecting the fact that the data includes non-banks and foreign depositors as discussed in Section 3.

Table 2 also indicates that the overnight borrowing cost of a bank is lower, the higher is the number of its counterparties in the overnight market (this is the case in most of specifications 5-9). This variable might partly proxy for a bank's systemic risk as does bank size, as it provides a simple measure of the bank's interconnectedness in the financial network. However, the coefficient of bank size is not much affected by its inclusion. Therefore, the number of counterparties may primarily capture the degree of competition in a bank's interbank lending relationships; more intense competition between lenders implying a lower overnight borrowing cost. As an alternative measure of a bank's interconnectedness, we consider the geographic concentration measure calculated from the bank's customer payments. In the absence of borrower level fixed effects, this variable obtains a positive and significant coefficient, which implies that more geographically diversified (or interconnected) banks obtain funding at a lower cost (see specification 5-9 in Table 2). However, as bank size maintains its importance as an explanatory variable in these specifications, the geographic concentration variable may rather capture the degree of diversification in banks' asset risks; lower concentrations contributing to a lower overnight borrowing cost.

Regarding the potential effect of lender-borrower relationships on the cost of overnight borrowing, we consider two relationship measures, the Borrow Preference Index (BPI) and Lender Preference Index (LPI) (see Cocco et al., 2009, for details) in model specifications 6-9 of Table 2. The relationship variables obtain significant coefficients when fixed effects for the borrower, lender, or the relationship-pair are included. The negative sign of BPI suggests that borrowers that rely on a relationship obtain cheaper overnight funding. The LPI variable has some explanatory power but the sign of its coefficient varies, possibly due to multicollinearity vis-a-vis the BPI.

Next, we turn to the more detailed analysis of behavior of the bank size premium in overnight markets. As discussed in Section 3, the overnight market has gone through significant changes during the sample period, e.g., conversion of the ECB's money market operations from partial allotment to fixed rate full allotment as well as the ECB's long-term refinancing operations and asset purchase programs. Therefore, in supplementary material we have analyzed the determinants of overnight borrowing rates separately for each sample year (the results are available from the authors upon request). The coefficient on bank size is generally the most consistent determinant of overnight borrowing costs, having the expected sign in all years and being statistically significant in almost all of them.

In Table 3 we provide further tests for whether the role of bank size indeed stems from the TBTF expectations. The idea is that if a bank is directly owned by a solid government, it would more likely receive government support in a crisis regardless of its size. For all other banks a large size would matter more as a "guarantee" of government support in a crisis. We test this in models 2 and 3 of Table 3. We form two dummy variables, the first of which ("State owned") takes the value one if a bank is owned by any government in our sample, and the second ("Core state owned") obtains value one if a bank is owned by a government with high credit worthiness (see Table 7 in the Appendix for the list of countries used in the definition). We focus on the interaction of these dummies with bank size. In models 2 and 3 of Table 3, it turns out that the interactions with both dummies are statistically significant while in model 3 the effect is somewhat stronger. Hence, bank size has a weaker negative effect on the overnight borrowing cost of banks which are government owned, and the effect is stronger if the government has a relatively high credit quality. In model 3 of Table 3, the effective constant term is also significantly affected by the second dummy, indicating that banks owned by the core states have lower average overnight borrowing costs. Overall, Table 3 provides further evidence that bank size serves as a proxy for the strength of the expected government support to TBTF institutions.

5.2 Economic significance of the bank size premium

The coefficient on the bank size variable implies an overnight borrowing rate differential between two banks of size *S* and *L*, respectively, according to the following formula:

$$\Delta r = 100\, \beta_1 log(Z_S/Z_L)\ , \tag{2}$$

where Δr is the interest rate differential in basis points, Z_S and Z_L are sizes of the two representative banks (where Z_S can be taken as larger than Z_L), and β_1 is the coefficient on bank size in Equation (1). When interpreting the interest rate differential, we must keep in mind which control variables are included.

The magnitude of the large banks' cost advantage is economically significant. In Table 2 with controls, the coefficient on bank size is approximately –0.03. This suggests that a bank, ten times the size of its peer, holding other factors constant, has a borrowing cost advantage of 7 bps. Similarly, a bank, fifty times the size of its peer, pays 12 bps less interest. For comparison, the coefficient of ROA (roughly 0.006 in Table 2) translates to a 0.6 bps change in interest for a 1 percentage point change in ROA.

5.3 The effect of the BRRD and actual bail-in decisions on the bank size premium

We now analyze whether the effect of bank size on banks' overnight borrowing costs has changed after the introduction of the new bank recovery and resolution framework (the BRRD) in the EU, the aim of which is to restrict government subsidies to banks. We also analyze the effect of actual bail-in decisions, using the same list of events as Bellia and Maccaferri (2020) who extend the list of events included in Schäfer et al. (2016). The analysis is carried out using the difference-in-differences methodology around event dates when either the legal framework has changed, making use of the different BRRD implementation dates in various EU countries and other related events, or when actual bail-in decisions have taken place.[2] We also include announcements of the ECB's LTROs in order to control their potential effects on banks' overnight borrowing costs. See in the Appendix Table 8 for the BRRD implementation dates and Table 9 for the related regulatory events, bail-in events, and LTRO announcement dates. Because the number of bail-in events is still rather limited, we will not restrict ourselves only to post-BRRD bail-in events but include all bail-in events during our sample period since mid-2008. We will comment on possibly different results in the pre-BRRD and post-BRRD subsamples. As supporting analysis, we also investigate changes in Bank Support Ratings published by Fitch Ratings (Fitch), which directly measure government support expectations. Finally, we also assess the potential longer-term impact of the BRRD on overnight borrowing costs in a regular panel regression.

Starting with Fitch's Support Ratings (SRs), they reflect the agency's view on the likelihood that a bank will receive extraordinary support to prevent it defaulting on its senior obligations. The scale is from 1 to 5 such that "1" corresponds to an extremely high probability of external support while "5" corresponds to a possibility of external support, which cannot be relied upon. Figure 3 shows evolution of the average SR for European banks from 2005 to 2017. Until mid-2015 the average SR hovers around "2" (high probability of external support). There is a minor increase in support expectations before the global financial crisis and a minor decrease afterwards. In May 2015 the average SR jumps to around 3.5, which corresponds to a moderate to limited probability of government support. Given the timing of the change, it can be attributed to the progress in implementing the BRRD in the EU as documented by Fitch. Although these are only the view of a single ratings agency, the changes in Fitch Support Ratings clearly support the notion that the BRRD has reduced government support expectations (see also Fitch Ratings 2014 and Fitch Ratings 2015).

5.3.1 Difference-in-differences analysis of the effects of the BRRD and bail-in decisions

[2] The BRRD implementation dates may not be exogenous in the sense that member states may have considered the state of their systemically relevant banks when deciding about the implementation date. However, most sample countries belong to the banking union which implied a time frame within which the implementation was expected to be done. The great majority of the sample countries implemented the framework either in 2015 or 2016. Hence, we do not expect possible discretion as to the exact timing of the BRRD implementation in each country to materially affect our estimation results.

Next, we perform the difference-in-differences analysis (DD) to investigate whether the adoption of the BRRD's legal framework of bail-in provisions at different times in different countries as well as the actual bail-in decisions have had an immediate impact on banks' overnight borrowing costs. Consider first the BRRD implementations. The dates of the BRRD implementation in different countries are gathered by ISDA (2016) and reported in Table 8 in the Appendix. We consider a 20-day window around each event (10 days before and 10 days after the event). We take the treatment group to be banks in countries that adopted the bail-in provisions at date t_I, and the untreated (control) group are the banks whose country either had so far not adopted the bail-in provisions or adopted the bail-in provisions at a different date (outside the 20 day window). We let S_i denote the treatment dummy for bank *i* (=1 for treated banks) and T denote the event dummy (=1 after implementation of the bail-in provisions). We are interested in the effect of the treatment on the coefficient on bank size, given by $\hat{\beta}_7$ in the following equation:

$$r_{i,t} = c_t + \beta_1 T_t + \beta_2 S_i + \beta_3 (T_t \cdot S_i) + [\beta_4 + \beta_5 T_t + \beta_6 S_i + \beta_7 (T_t \cdot S_i)] \log Z_{i,t} + \varepsilon_{i,t} \quad , \tag{3}$$

where $\log Z_{i,t}$ is again the log of total assets of bank *i* and c_t denote the time fixed effects. Note that the time fixed effects render the term $\beta_1 T_t$ redundant.

When we estimate the model in Equation (4) with the country-specific BRRD implementations (mostly either on 2015-01-01 or 2016-01-01), we pool all implementation dates into one regression). Therefore Equation (3) generalizes to the following form:

$$r_{i,t} = c_t + \sum_e \beta_{0,e} D_{t,e} + \beta_1 T_t + \sum_e \beta_{2,e} D_{t,e} S_{i,e} + \beta_3 (T_t \cdot S_i) + [\sum_e \beta_{4,e} D_{t,e} + \beta_5 T_t + \sum_e \beta_{6,e} D_{t,e} S_{i,e} + \beta_7 (T_t \cdot S_i)] \log Z_{i,t} + \varepsilon_{i,t} \ , \tag{4}$$

where index *e* denotes summation over events and $D_{t,e}$ is a dummy for event *e* (taking value "one" within the 20-day window). Here $S_i = \min(1, \sum_e S_{i,e} D_{t,e})$ and $T_t = \min(1, \sum_e T_{t,e})$ where $S_{i,e}$ and $T_{t,e}$ are the usual DD dummies for a single event *e*. Duplicating the terms in equation (4), the approach generalizes to controlling for competing event types. For brevity, we drop the event specific coefficients from our tables with the understanding that they are included in the event controls. The bail-in events and LTRO events are tested using a similar equation, but with S always equal to one so that the equation reduces to

$$r_{i,t} = c_t + \sum_e \beta_{0,e} D_{t,e} + \beta_1 T_t + [\sum_e \beta_{4,e} D_{t,e} + \beta_5 T_t] \log Z_{i,t} + \varepsilon_{i,t} \quad (5)$$

The results from model (2) in Table 4 show that the introduction the BRRD has not had any statistically significant effect on the bank size premium. So, there is no direct effect but as discussed earlier, we suspect that the effect may show up via the actual bail-in decisions. Hence, we next consider the effect of actual bail-in decisions, using the list of events in Bellia and Maccaferri (2020), and further relying on the specification in Equation (5) so that we pool all bail-in event dates in the same regression. The results in model (1) in Table 4 show that the bank size premium has markedly declined, though not vanished. Note that some of the actual bail-in decisions may well have “disappointed” in the sense that bail-in rules in those cases have been applied in a watered-down manner. Hence, the estimated coefficient on the bail-in event-dummy variable times bank size, which is positive and highly significant, should be interpreted such that on average the actual bail-in events have reduced market expectations of government support to larger banks.[3] Some of the events in the list of Bellia and Maccaferri (2020) are regulatory decisions or changes to legislative frameworks (outside the actual BRRD implementation). They are not actual bail-in decisions, so we do not include them in the bail-in events dummy used in Table 4. Model (3) in Table 4 gauges their potential impact on bank size premium separately but finds none.

The ECB announcements of its long-term refinancing operations (LTROs) may have had an impact not only on the level but also relative pricing of overnight loans for larger and smaller banks. In model (4) of Table 4 we investigate this with the DD analysis and find that the effective bank size premium has indeed considerably declined in conjunction with the LTRO announcements.[4] The effect can be due to LTROs reducing the market stress or causing some market participants to leave the overnight market as they shift to central banks financing, as discussed further in the robustness checks. Finally, in model (5) of Table 4 we pool all the

[3] The list of Bellia and Maccaferri (2020) that we use includes bail-in events both from before and after the implementation of the BRRD, and our results in Table 4 are based on including all of them in the DD analysis. In unreported results we find that if we include only bail-in events that took place after implementation(s) of the BRRD, the estimated impact on the effective bank size premium is still negative but not statistically significant.

[4] The history of all ECB open market operations is available at https://www.ecb.europa.eu/mopo/implement/omo/html/top_history.en.html.

previous “events” together. We find that the negative impact of bail-in events and the LTRO announcements on the effective bank size premium maintains its magnitude and high statistical significance in both cases.

Taken together with the evidence from Fitch ratings discussed earlier, the DD results suggest that while a rating agency appears to be attaching most weight on the legal implementation of a resolution framework as a decisive event, the markets rather react when they see actual regulatory bail-in decisions. We have additionally considered the events on an individual basis. Unreported results show that in none of the four bail-in events that took place after 2016, was there a significant impact on the size premium. Relying only on four events, this can be interpreted as casual evidence of these events being anticipated, perhaps because BRRD has aligned the bail-in expectations.

5.3.2 Long-term effects

We also investigate the long-term impact on bank size premium of the BRRD implementation and the respective policy announcement preceding, which may have taken place more gradually, and would be consistent with the evolution of Fitch Support ratings. We return to the basic specification (Equation 1) and multiply bank size by three alternative time dummy variables. The first of them equals one for the latter part of our sample starting from 1 July 2012, which indicates the European Commission’s proposal for new recovery and resolution tools for banks in crisis. Similarly, the second dummy equals one starting from 16 March 2013 (until the end of sample period), when the multilateral agreement was reached of a partial bail-in of bank debtors as part of the rescue package for Cyprus. James et al. (2016) argue that the resolution procedure applied in the case of Cyprus may already have fundamentally affected expectations regarding future crisis resolution policies in Europe. So, the first two dummies capture the various stages when concrete market expectations regarding future bank resolution legislation may have started to take shape. The third time dummy equals one starting from 1 January 2016 (until the end of sample) when the BRRD came into force.

If the new legislation, or anticipation of it, has had an intended effect in reducing the implicit TBTF subsidies to banks, then our hypothesis is that the subsequent impact of bank size on bank overnight borrowing costs would have been weakened. Because the central aim of the BRRD is to weaken the bank-sovereign loop, we also apply the time dummies on the sovereign

CDS spread. Moreover, since it is possible that the intended effects are largest in the crisis stricken GIIPS countries, we also add a dummy variable for them.

The first four columns of Table 5 show the results, considering three model versions. In regressions 1 and 2 of Table 5, the effect of bank size on the bank's overnight borrowing cost has diminished during the planning and implementation of the BRRD. In case of regression 1 this is seen from the joint test for the three dummy interaction coefficients reported at the bottom of the table. However, regression 2 shows that the effect of sovereign CDS price on banks' overnight borrowing costs in the respective country has not changed after the BRRD. Regression 3 and especially regression 4 in Table 5 indicate that the BRRD has considerably weakened the effect of bank size on bank overnight borrowing cost for banks that are domiciled in the GIIPS countries.[5] In regression 4, we include only the dummy starting from 1 January 2016 when the BRRD came into effect, by interacting it with the dummy for GIIPS countries and the log of total assets. The effective coefficient on the log of total assets of banks in the GIIPS countries after the BRRD is reduced (in absolute terms) by two thirds compared to the base coefficient. Thus, the sum of coefficients of log of total assets in the GIIPS country group is -0.026 prior to January 1, 2016 and -0.009 after that. In contrast, we cannot reject the null that the BRRD has *not* changed the effect of sovereign CDS price on overnight borrowing cost of banks in the GIIPS countries either.

These results might well reflect the impact of unconventional monetary policy programs of the ECB; the Outright Monetary Transactions program announced in 2012 and the quantitative easing started in 2015, which may have indirectly affected banks' creditworthiness, especially in the GIIPS countries. At the same time the results are not inconsistent with the hypothesis that the BRRD has over time reduced the overnight borrowing cost advantage of larger banks in these countries.

5.4 Further robustness checks

[5] We have also tested whether the interaction terms of different dummy variables with the size are jointly statistically significant. F-test indicates that in regression (1) the interaction terms D(T) x log(Total assets) are not jointly significant (p=0.1731), but in regression (2) the interaction terms D(T) x GIIPS x log(Total assets) are jointly significant (p=0.0008).

In Table 6 we further examine the relationship between bank size and borrowing cost by conditioning this relationship on market turbulence. We include the product of bank size (the log of total assets) and the iTraxx Financials index for Europe, based on 5-year CDS spreads which are the most liquid CDS contracts. Models 3 and 5 of Table 6 show that the coefficient of the multiplicative term is negative and clearly significant. On balance, the coefficient of the log of total assets drops roughly by half (cf. model 1 in Table 6) but is still significant. The significant interaction term indicates that although overnight borrowing costs are expected to rise for all banks during market stress, they do less so for larger banks. This safe-haven effect may be taken as further evidence that bank size indeed serves as a proxy for the TBTF expectations.

In models 4 of Table 6 we consider the possibility that the European Central Bank's liquidity provisioning benefits banks differently depending on their size and might hence explain part of the effect of bank size on the overnight borrowing cost. The coefficient on the product of bank size and TARGET2 excess liquidity is statistically significant. Hence, central bank liquidity measures may overtime have had a negative effect on the size premium. This could be the case, if smaller banks that pay higher interest rates have higher propensity to exit the overnight market. Model 5 of Table 6 suggests that this effect is weak compared to the safe-haven effect so that the ECB's monetary policy and the ensuing changes in the overnight market should not contaminate our results. However, to the extent that ECB's monetary policy has calmed the markets reflected as lower levels of iTraxx, they may impact our quantification of TBTF expectations.

In Table 6 we also consider potential non-linearity in the bank size-borrowing cost relationship. It is possible that TBTF expectations primarily concern the very largest banks (cf. Acharya et al., 2016). In model 2 of Table 6 an additional quadratic bank size term increases the explanatory power of the model according to the F-test. The sign of the quadratic term is negative indicating that the TBTF effect indeed matters more for the very large banks. Evidence on nonlinearity may also explain why the size variable performs better than the G-SIFI dummy.

6 Conclusions

Using a comprehensive data set of European banks' interbank borrowing rates, derived from the TARGET2 transactions data, we have investigated whether large European banks have a cost advantage in unsecured overnight borrowing and whether the introduction of the Bank Recovery and Resolution Directive (BRRD) has reduced implicit public guarantees to larger banks in the light of this alleged bank size-borrowing cost relationship. We find that the overnight borrowing cost advantage of large banks does exist, and it is very robust in terms of both time periods, control variables and selection of countries and banks. Though we do not find an immediate reaction to implementation of the BRRD in the bank size premium in the overnight market, we do find that the actual bail-in events have had a dampening effect on the size premium. Thus, the regulatory effort to "level off" the playing field for banks in terms of reducing the need for implicit public support for the largest institutions seems to have been partly if not entirely successful in the eyes of the interbank overnight market.

References

Acharya, V.V., D. Anginer, A.J. Warburton (2016): The end of market discipline? Investor expectations of implicit government guarantees. SSRN working paper. https://dx.doi.org/10.2139/ssrn.1961656

Afonso, G., A. Kovner, A. Schoar (2011): Stressed, not frozen: The federal funds market in the financial crisis. *The Journal of Finance*, 66:4, 1109-1139. https://doi.org/10.1111/j.1540-6261.2011.01670.x

Ahmed, J., C. Anderson and R, Zarutskie (2015): Are the borrowing costs of large financial firms unusual? Finance and Economics Discussion Series 2015-024. Washington, Board of Governors of the Federal Reserve System, FEDS Working paper 24. http://dx.doi.org/10.17016/FEDS.2015.024

Akram, Q. F. and C. Christophersen (2017): Pricing in the Norwegian interbank market – the effects of liquidity and implicit government support, *Oxford Bulletin of Economics and Statistics*, 79, 165-204. https://doi.org/10.1111/obes.12147

Angelini, P., A. Nobili, C. Picillo (2011): The interbank market after august 2007: What has changed, and why? *Journal of Money, Credit and Banking*, 43:5, 923-958. https://doi.org/10.1111/j.1538-4616.2011.00402.x

Araten, M., and C. Turner (2013): Understanding the funding cost differences between global systemically important banks (GSIBs) and non-GSIBs in the USA. *Journal of Risk Management in Financial Institutions* 387-410. https://dx.doi.org/10.2139/ssrn.2226939

Arciero, L., R. Heijmans, R. Heuver, M. Massarenti, C. Picillo and F. Vacirca (2016): How to measure the unsecured money market? The Eurosystem's implementation and validation using TARGET2 data. *International Journal of Central Banking* Vol. 12, No 1, pp. 247–280. https://www.ijcb.org/journal/ijcb16q1a8.htm

Armantier, O. A. Copeland (2012): Assessing the quality of Furfine-based algorithms. Staff Reports No. 575, Federal Reserve Bank of New York. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr575.pdf

Bassett, W. F. (2014): Using insured deposits to refine estimates of the large bank funding advantage. Working paper, Board of Governors of the Federal Reserve System (FRB). https://dx.doi.org/10.2139/ssrn.2447997

Bellia, M., Maccaferri, S., 2020, Banks' bail-in and the new banking regulation: an EU event study, European Commission, JRC Technical Reports. http://dx.doi.org/10.2760/53487%20

Berger, A.N., C.P. Himmelberg, R.A. Roman, S. Tsyplakov (2019): Bank bailouts, bail-ins, or no regulatory intervention? A dynamic model and empirical tests of optimal regulation. SSRN working paper. https://dx.doi.org/10.2139/ssrn.3179226

Berndt, A., D. Duffie, Y. Zhu (2020): The decline of too big to fail. Available at SSRN: https://ssrn.com/abstract=3497897 or http://dx.doi.org/10.2139/ssrn.3497897

Brauning, F. and F. Fecht (2017): Relationship lending in the interbank market and the price of liquidity. *Review of Finance* 21, 33-75. https://doi.org/10.1093/rof/rfw042

Brewer III, E., and J. Jagtiani (2013): How much did banks pay to become too-big-to-fail and to become systemically important? *Journal of Financial Services Research* 43, 1-35. https://doi.org/10.1007/s10693-011-0119-6

Cocco, J.F., F.J. Gomes, and N. C. Martins (2009): Lending relationships in the interbank market, *Journal of Financial Intermediation 18:1, 24-48.* https://doi.org/10.1016/j.jfi.2008.06.003

Dang, T. V., Gorton, G., Holmström, B., (2015): The information sensitivity of a security, Working Paper. Available at: http://www.columbia.edu/~td2332/Paper_Sensitivity.pdf

Davies, R., B. Tracey (2014): Too big to be efficient? The impact of implicit subsidies on estimates of scale economies for banks. *Journal of Money, Credit and Banking*, Volume 46, Supplement 1, pp. 219-253(35). https://doi.org/10.1111/jmcb.12088

Dell'Ariccia, G., M. Soledad Martinez Peria, D. Igan, E. Addo Awadzi, M. Dobler, D. Sandri (2018): Trade-offs in bank resolution. IMF Staff Discussion Note 18/02. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2018/02/09/Trade-offs-in-Bank-Resolution-45127

Demirguc-Kunt, A., and H. Huizinga (2013): Are banks too big to fail or too big to save? International evidence from equity prices and CDS spreads. *Journal of Banking and Finance* 37, 875-894. https://doi.org/10.1016/j.jbankfin.2012.10.010

Duffee, G. (1998): The relation between treasury yields and corporate bond yield spreads. *Journal of Finance* 53, 2225–2242. https://doi.org/10.1111/0022-1082.00089

Duffie, D., K.J. Singleton (1999): Modeling term structures of defaultable bonds. *Review of Financial Studies* 12: 4, 687–720. https://doi.org/10.1093/rfs/12.4.687

ECB (2019): Euro money market study 2018, European Central Bank. https://www.ecb.europa.eu/pub/euromoneymarket/html/ecb.euromoneymarket201909_study.en.html

Farhi, E., J. Tirole (2012): Collective moral hazard, maturity mismatch and systemic bailouts, *American Economic Review*, vol. 102, pp. 60–93. https://doi.org/10.1257/aer.102.1.60

Fitch Ratings (2014): Sovereign support for banks – rating path expectations, Fitch Special Report 2014/03/27.

Fitch Ratings (2015): Sovereign support for European banks – bank resolution progress lowers likelihood of sovereign protection for bondholders, Fitch Special Report 2015/05/22.

Furfine, C.H. (2001): Banks as monitors of other banks: Evidence from the overnight Federal Funds market. *Journal of Business* 74: 1, 33-57. https://doi.org/10.1086/209662

Gorton, G., Metrick, A. (2009): Securitized banking and the run on repo, Yale ICF Working Paper 09–14. Available at SSRN https://ssrn.com/abstract=1440752

Guennewig, M. and G. Pennacchi (2019): Measuring bail-in credibility, Discussion note. https://drive.google.com/file/d/15JYZv244frg9ANfx1CSHynwPffiMYhwc/view?usp=sharing

Huber, P. J. (1967): The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 221–233. http://www.personal.psu.edu/~aza12/huber_s08.pdf

Hughes, J. P., and L. J. Mester (2013): Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation* 22, 559-585. https://doi.org/10.1016/j.jfi.2013.06.004

ICMA (2018): European repo market survey, No. 35, International Capital Market Association. https://www.icmagroup.org/assets/documents/Market-Info/Repo-Market-Surveys/No-35-June-2018/ICMA-European-repo-market-survey-number-35-conducted-June-2018.pdf

ISDA (2016): ISDA BRRD Implementation Monitor (5th Edition) – status of national implementation as of 1 December 2016, International Swaps and Derivatives Association Report. https://www.isda.org/a/YGiDE/icm-24590303-v6-isda-brrd-implementation-monitor-5th-edition.pdf

Jacewitz, S., Pogach, J. (2018): Deposit rate advantages at the largest banks. *Journal of Financial Services Research* 53, 1–35. https://doi.org/10.1007/s10693-016-0261-2

James, H., J-P. Landau, M.K. Brunnermeier (2016): The euro and the battle of ideas. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691172927/the-euro-and-the-battle-of-ideas; http://press.princeton.edu/chapters/s10828.pdf (First Chapter)

Kroszner, J. (2013): A review of bank funding cost differentials https://www.stern.nyu.edu/sites/default/files/assets/documents/confi044532.pdf

Martino, E. (2020): The bail-in beyond unpredictability: Creditors' incentives and market discipline. *European Business Organization Law Review*, May 2020, 3-40. https://doi.org/10.1007/s40804-020-00188-7

Noss, J., and R. Sowerbutts (2012): The implicit subsidy of banks. Bank of England Financial Stability Paper 15. https://www.bankofengland.co.uk/financial-stability-paper/2012/the-implicit-subsidy-of-banks or https://ssrn.com/abstract=2071720

O'Hara, M., and W. Shaw (1990): Deposit insurance and wealth effects, the value of being too big to fail. *The Journal of Finance* 45, 1587-1600. https://doi.org/10.1111/j.1540-6261.1990.tb03729.x

Penas, M.F., H. Unal (2004): Gains in bank mergers: Evidence from the bond markets, Journal of Financial Economics, Vol. 74, No. 1, pp. 149-179. https://doi.org/10.1016/j.jfineco.2003.05.004

Ringe, W-G, J. Patel (2019): The dark side of bank resolution: Counterparty risk through bail-in. European Banking Institute Working Paper Series 2019:31. https://dx.doi.org/10.2139/ssrn.3314103

Rochet, J.-C., Tirole, J. (1996): Interbank lending and systemic risk, *Journal of Money, Credit, and Banking*, Vol. 28, No. 4, pp. 733–762. https://doi.org/10.2307/2077918

Rogers, W. H. (1993): Regression standard errors in clustered samples. Stata Technical Bulletin 13: 19–23. https://www.stata-press.com/journals/stbcontents/stb13.pdf

Santos, J. (2014): Evidence from the bond market on banks' too-big-to-fail" subsidy. *Economic Policy Review* 20, 29-39. https://www.newyorkfed.org/medialibrary/media/research/epr/2014/1412sant.pdf

Schäfer, A., I. Schnabel, B. Weder di Mauro (2016): Bail-in expectations for European banks: Actions speak louder than words. European Systemic Risk Board Working Paper 7/2016. http://dx.doi.org/10.2849/768515

Tölö, E., E. Jokivuolle and M. Virén (2017): Do banks' overnight borrowing rates lead their CDS price? Evidence from the Eurosystem, *Journal of Financial Intermediation*, Volume 31, 93-106. https://doi.org/10.1016/j.jfi.2017.05.006

Ueda, K., and B. Weder di Mauro (2013): Quantifying structural subsidy values for systemically important financial institutions. *Journal of Banking & Finance* 37, 3830-3842. https://doi.org/10.1016/j.jbankfin.2013.05.019

Wheelock, D. C., and P. W. Wilson (2012): Do large banks have lower costs? New estimates of returns to scale for US banks. *Journal of Money, Credit and Banking* 44, 171-199. https://doi.org/10.1111/j.1538-4616.2011.00472.x

Table 1. Descriptive statistics.

The sample is restricted to bank holdings, holding companies, commercial banks, cooperative banks, investment banks, real estate & mortgage banks, and savings banks, as identified using Orbis Bank Focus. Note that RMP refers to Reserve Maintenance Period.

Variable	Observations	Mean	Std. Dev.	Median	IQR	Units	Aggregation level
Overnight rate spread	863,591	-0.049	0.173	-0.054	0.110	%	Pair-Day
Borrower preference index (BPI)	807,547	0.193	0.294	0.052	0.204	Fraction	Pair-Day
Lender preference index (LPI)	807,562	0.275	0.321	0.128	0.368	Fraction	Pair-Day
No. of days pairs	813,542	20.4	18.4	15.0	28.0	Number	Pair-Day
Transaction size	863,591	84.3	182	25	70	Million Euro	Pair-Day
Transaction size / Equity	738,248	1.91	17.42	0.28	0.96	%	Pair-Day
Cross-border dummy	863,591	0.558	0.497	1	1	0 or 1	Pair-Day
Geographic concentration	204,900	0.554	0.238	0.529	0.424	Fraction	Borrower-Day
No. of lenders	311,935	15.1	20.7	7.0	17.0	Integer	Borrower-Day
ECB deposit facility activity	38,208	0.043	0.186	0	0	Fraction	Borrower-RMP
ECB marginal lending facility activity	38,208	0.002	0.026	0	0	Fraction	Borrower-RMP
Number of loans in ON market	38,208	24.7	52.9	6.0	20.0	Integer	Borrower-RMP
Total assets	6,526	334,839	584,113	40,205	399,578	Million euro	Borrower-Year
ROA	6,519	0.53	1.33	0.46	0.80	%	Borrower-Year
Tier 1 ratio	5,520	13.00	6.22	12.08	4.70	%	Borrower-Year
NPL ratio	5,916	5.97	7.91	3.73	5.04	%	Borrower-Year
Liquid assets / Deposits	6,521	43.53	39.84	34.46	34.56	%	Borrower-Year
Non-interest income / Revenue	6,505	38.55	29.94	36.67	21.32	%	Borrower-Year
G-SIFI dummy	1,574	0.118	0.323	0	0	0 or 1	Borrower
State owned bank dummy	1,574	0.119	0.324	0	0	0 or 1	Borrower
Core state owned dummy	1,574	0.025	0.157	0	0	0 or 1	Borrower
Subject to reserve requirement dummy	1,574	0.164	0.370	0	0	0 or 1	Borrower
Exchange traded dummy	1,574	0.417	0.493	0	1	0 or 1	Borrower
Country CDS	67,394	217.1	1194.6	65.3	113.6	bps	Country-Day
iTraxx Financials CDS index	3,088	113.0	58.4	92.3	66.2	bps	Day
ON market concentration	3,088	0.045	0.023	0.040	0.032	Fraction	Day
ON market number of borrowers	3,088	134.9	74.7	116	156	Count	Day
TARGET2 excess reserves	3,083	745,144	715,397	429,087	1,504,994	Million euro	Day

Table 2. Determinants of overnight interbank loan rates.

The estimating equation is Eq. (1). The dependent variable is interest rate spread defined for every overnight loan as the difference between the interest rate on that loan and the EONIA interest rate on the same day. Numbers inside the brackets are t-values adjusted for clustering at the borrower-lender-pair level. The results are based on the full sample of banks. Estimated coefficients of variables are multiplied by 100 except for CDS, the coefficient of which is multiplied by 10,000. The sample is restricted to bank holdings, holding companies, commercial banks, cooperative banks, investment banks, real estate & mortgage banks, and savings banks, as identified using Orbis Bank Focus. Size = log (Total assets), Tier = Tier 1 ratio, Non-interest income = Non-interest income/ total income.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Size	-2.746	-2.582	-2.455	-2.787	-2.465	-2.240	-3.600	-1.724	-4.048
	[21.5]	[24.7]	[14.8]	[21.4]	[11.3]	[9.1]	[8.9]	[16.0]	[8.1]
ROA		-0.664	-0.607	-0.739	-0.475	-0.497	-1.193	-0.226	-1.048
		[2.2]	[2.0]	[2.3]	[1.4]	[1.4]	[4.6]	[1.0]	[3.2]
Tier 1		-0.211	-0.186	-0.148	-0.261	-0.252	-0.269	-0.245	-0.272
		[5.3]	[5.7]	[3.6]	[5.7]	[5.3]	[5.5]	[8.1]	[4.6]
CDS			0.108						
			[3.0]						
Non-interest income				-0.021	-0.019	-0.019	-0.010	-0.014	-0.007
				[6.7]	[6.4]	[6.3]	[5.7]	[9.4]	[4.3]
NPL ratio				-0.017					
				[0.5]					
Liquid assets/Deposits				-0.017	-0.003	-0.004	-0.001	0.003	-0.008
				[4.0]	[0.5]	[0.6]	[0.1]	[0.8]	[0.7]
G-SIFI dummy				2.502	0.446	0.438		1.642	
				[5.1]	[0.7]	[0.6]		[4.2]	
Listed bank dummy				0.077	0.753	0.831		0.482	
				[0.2]	[2.1]	[2.4]		[2.3]	
Subject to RR					-2.630	-2.663		-2.334	
					[6.6]	[6.8]		[8.4]	
ECB deposit facility activity					-3.445	-3.236	-2.539	-2.458	-2.539
					[6.6]	[6.3]	[6.6]	[7.7]	[6.8]
ECB lending facility activity					22.679	22.436	17.211	21.324	15.386
					[6.7]	[6.9]	[6.8]	[7.1]	[6.0]
Log (No. of lenders)					-3.650	-3.678	0.235	-2.544	0.596
					[7.7]	[11.4]	[1.0]	[13.9]	[2.5]
Log (Loan value)					1.441	1.431	2.065	0.326	1.057
					[11.5]	[9.8]	[24.3]	[3.3]	[10.1]
Log (Loan count)					2.754	2.844	1.235	1.935	0.884
					[10.3]	[12.0]	[10.9]	[17.7]	[9.1]
Geographic concentration					4.087	3.486	0.048	2.178	-0.046
					[4.3]	[3.7]	[0.1]	[4.9]	[0.1]
Cross-border dummy						2.484	1.163	0.696	
						[4.9]	[3.6]	[1.9]	
BPI						-0.720	-2.993	-1.265	-1.190
						[0.4]	[4.6]	[2.1]	[1.9]
LPI						-1.104	-1.002	3.305	0.671
						[1.5]	[2.6]	[8.7]	[1.7]
Fixed effects	Time	Time	Time + Country	Time	Time + Country	Time + Country	Time + Borrower	Time + Lender + Country	Time + Pair
N	766,600	726,829	688,101	704,501	541,795	537,669	537,669	537,669	537,669
Pairs	22,603	20,936	18,932	20,431	12,801	12,179	12,179	12,179	12,179
Borrower banks	1,397	1,222	1,036	1,180	381	352	352	352	352
R^2	0.246	0.244	0.362	0.253	0.410	0.416	0.469	0.469	0.580

Inside brackets are robust t-values adjusted for clustering at the bank level.

Table 3. Further investigation of bank size premium in overnight loan rates by conditioning bank size on bank ownership.

The estimating equation is Eq. (1). The dependent variable is the rate spread per overnight loan. Same notes as in Table 2 apply. “State owned” dummy captures banks owned by public authority, state or government. CDS denotes the country CDS. “Core state owned” dummy additional requires the country to be among AT, BE, DE, DK, CH, FI, FR, GB, LU, NL, NO, or SE.

Variable	(1)	(2)	(3)
Size	-2.455	-2.494	-2.501
	[14.8]	[14.8]	[15.0]
ROA	-0.607	-0.452	-0.401
	[2.0]	[1.5]	[1.3]
Tier 1	-0.186	-0.206	-0.221
	[5.7]	[6.7]	[6.9]
CDS	0.108	0.113	0.114
	[3.0]	[3.2]	[3.2]
D(State owned)*Size		1.101	
		[2.2]	
D(State owned)		-7.025	
		[1.2]	
D(Core state owned)*Size			1.755
			[3.1]
D(Core state owned)			-13.582
			[2.2]
Fixed effects	Country + Time	Country + Time	Country + Time
N	688,101	688,101	688,101
Pairs	18,932	18,932	18,932
Banks	1,036	1,036	1,036
R^2	0.362	0.370	0.370

Inside brackets are robust t-values adjusted for clustering at the bank level.

Table 4. Effects on bank size premium in overnight loan rates of changes in the EU bank resolution framework and bail-in events.

The estimating equation is Eq. (5) except for column (2), which is based on Eq. (4). The dependent variable is the overnight rate spread per loan. Notes of Table 2 apply but sample is further restricted to the European Economic Area (EEA) and the 20-day window around event dates.

	(1)	(2)	(3)	(4)	(5)
Size	-1.85	-2.51	-2.12	-2.72	-2.11
	[9.5]	[8.9]	[5.9]	[7.0]	[4.8]
T(BAIL-IN)*Size	0.58				0.58
	[5.6]				[5.6]
ST(BRRD)*Size		0.44			0.49
		[0.8]			[0.8]
T(BRRD)*Size		-0.01			-0.05
		[0.1]			[0.9]
T(LEGAL)*Size			0.05		0.05
			[0.6]		[0.6]
T(LTRO)*Size				0.61	0.62
				[9.9]	[9.9]
+EVENT CONTROLS					
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Events	12	20	12	23	67
N	33,404	46,048	29,875	154,844	259,243
Pairs	4,809	3,304	3,440	11,722	13,571
Banks	424	373	363	540	605
R^2	0.206	0.179	0.255	0.296	0.306

Inside brackets are robust t-values adjusted for clustering at the bank level.

Table 5. Long-term impact of the EU bank resolution regime on the bank size-dependence of banks' overnight borrowing costs.

The estimating equation is Eq. (1). The dependent variable is the overnight rate spread. Same notes as in Table 2 apply. The sample is restricted to EEA banks. $D(T) = 1$ if $t \geq T$ and 0 otherwise.

Variable	(1) ON rate	(2) ON rate	(3) ON rate	(4) ON rate
Size	-2.811	-2.629	-2.451	-2.515
	[16.7]	[14.5]	[9.4]	[9.8]
GIIPS*Size			-0.443	-0.120
			[1.5]	[0.4]
D(2012-07-01)*Size	0.698			
	[1.9]			
D(2013-12-19)*Size	0.094			
	[0.3]			
D(2016-01-01)*Size	0.062	0.670		
	[0.2]	[3.0]		
D(2012-07-01)*GIIPS*Size			0.990	
			[1.5]	
D(2013-12-19)*GIIPS*Size			0.009	
			[0.0]	
D(2016-01-01)*GIIPS* Size			0.311	1.010
			[0.7]	[2.8]
ROA	-0.923	-0.834	-0.931	-0.848
	[2.4]	[2.2]	[2.3]	[2.3]
Tier 1	-0.107	-0.112	-0.100	-0.104
	[3.2]	[3.4]	[2.9]	[3.1]
CDS	0.442	0.450	0.455	0.450
	[2.8]	[2.9]	[2.8]	[2.9]
D(2012-07-01)*CDS	-0.797			
	[1.3]			
D(2013-12-19)*CDS	-2.956			
	[4.3]			
D(2016-01-01)*CDS	2.824	-0.195		
	[2.9]	[0.2]		
D(2012-07-01)*GIIPS*CDS			-0.096	
			[0.1]	
D(2013-12-19)*GIIPS*CDS			-4.900	
			[2.3]	
D(2016-01-01)*GIIPS*CDS			6.982	2.189
			[1.6]	[0.6]
D(2012-07-01)*GIIPS			-0.140	
			[1.9]	
D(2013-12-19)*GIIPS			0.049	
			[1.1]	
D(2016-01-01)*GIIPS			-0.079	-0.129
			[1.0]	[2.3]
Fixed effects	Country + Time	Country + Time	Country + Time	Country + Time
N	639,923	639,923	639,923	639,923
Pairs	15,914	15,914	15,914	15,914
Banks	579	579	579	579
R^2	0.364	0.360	0.365	0.360
F-test for all D(T)*Size = 0	F(3,15913)=9.52		F(3,15913)=7.39	
or for all D(T)*GIIPS*Size = 0	p=0.0000		p=0.0001	

Inside brackets are robust t-values adjusted for clustering at the bank level.

Table 6. Testing for non-linearity of the bank size-dependence of banks' overnight borrowing costs.

The estimating equation is Eq. (1). The dependent variable is the overnight rate spread. Same notes as in Table 2 apply. Estimated coefficients of Size, $Size^2$, ROA, and Tier 1 are multiplied by 100. Coefficients of CDS and iTraxx*Size are multiplied by 10,000, and coefficient of Excess reserves*Size is multiplied by one 100 million.

Variable	(1)	(2)	(3)	(4)	(5)
Size	-2.455	-0.743	-0.307	-2.590	-0.411
	[14.8]	[0.6]	[1.6]	[14.9]	[2.1]
$Size^2$		-0.079			
		[1.4]			
iTraxx*Size			-1.596		-1.580
			[17.0]		[16.9]
Excess reserves*Size				0.420	0.258
				[4.0]	[2.5]
ROA	-0.607	-0.644	-0.638	-0.669	-0.676
	[2.0]	[2.1]	[2.1]	[2.2]	[2.2]
Tier 1	-0.186	-0.168	-0.120	-0.163	-0.106
	[5.7]	[4.9]	[4.3]	[5.3]	[3.9]
CDS	0.108	0.108	0.082	0.109	0.083
	[3.0]	[3.0]	[2.4]	[3.0]	[2.4]
Fixed effects	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time
N	688,101	688,101	688,101	687,895	687,895
Pairs	18,932	18,932	18,932	18,930	18,930
Banks	1,036	1,036	1,036	1,036	1,036
R^2	0.362	0.363	0.378	0.363	0.378

Inside brackets are robust t-values adjusted for clustering at the bank level.

Figure 1. Euro overnight index average (EONIA), the filtered TARGET2 rate, and the excess liquidity in TARGET2

The upper panel shows Euro overnight index average (EONIA), the filtered (derived) TARGET2 rate, and the corridor between marginal lending rate and deposit facility rate. The lower panel show the excess liquidity in TARGET2.

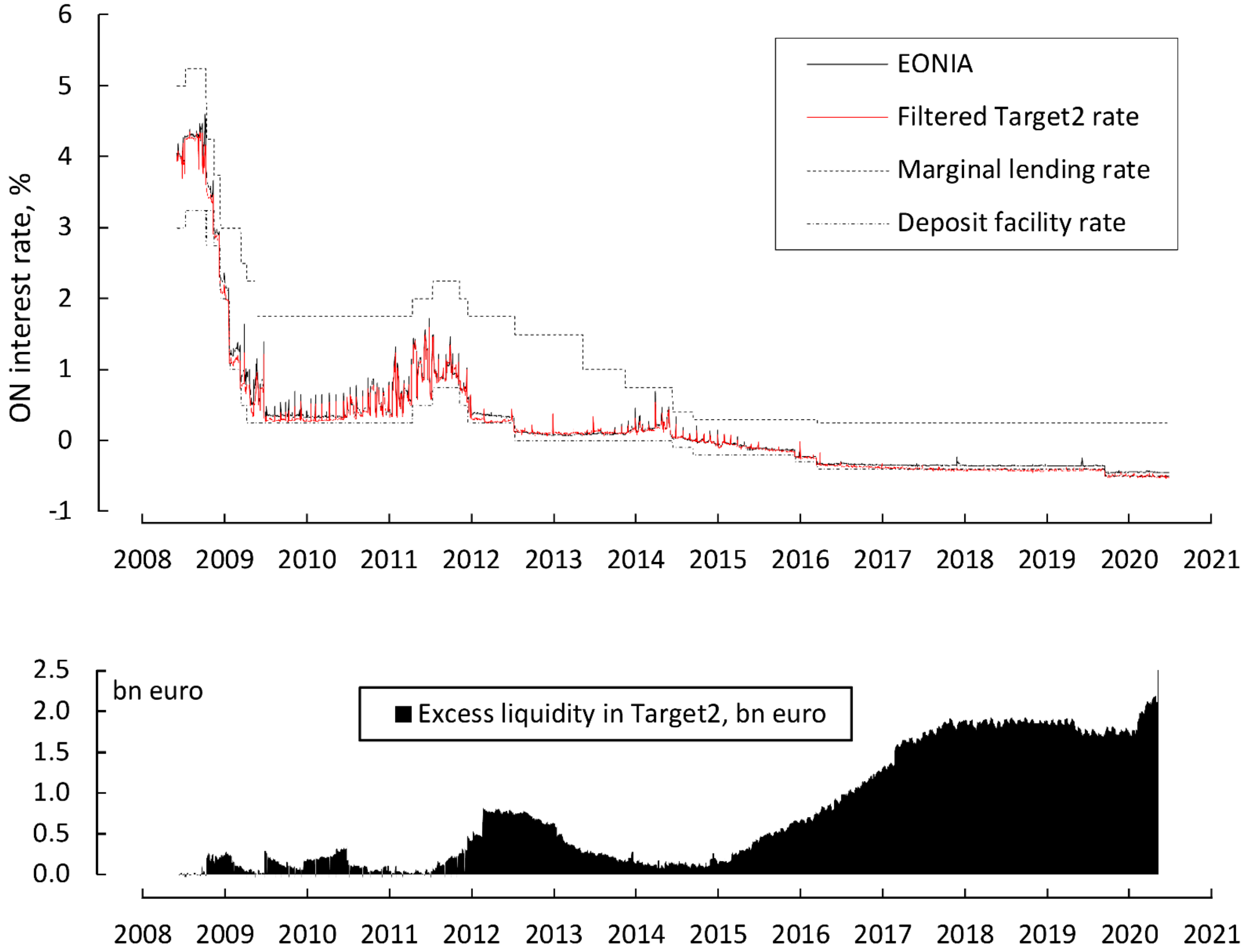
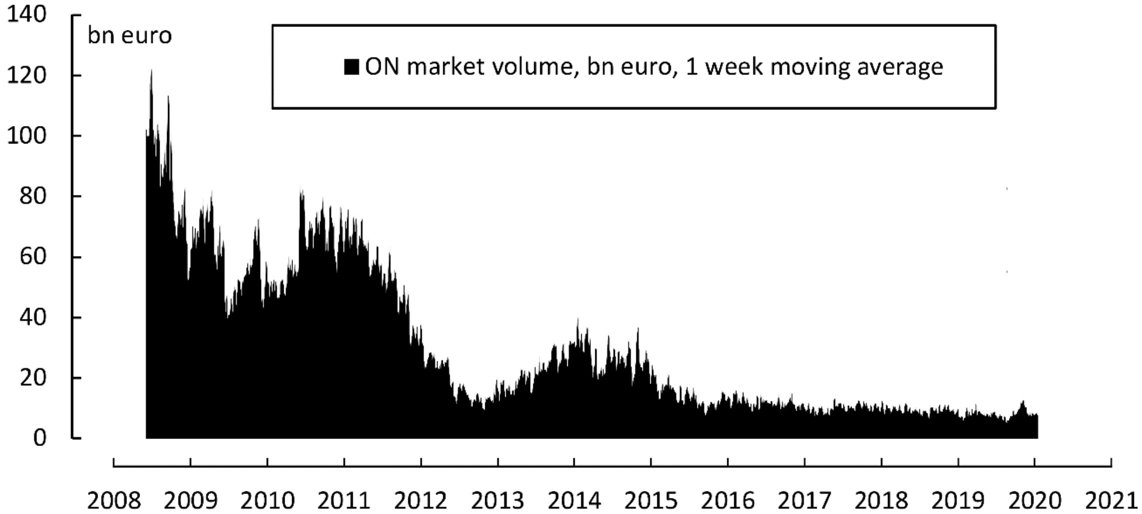


Figure 2. Overnight market volumes and number of participants.

(a) ON market volume as identified from TARGET2 transactions.



(b) Number of borrowers and lenders as identified from TARGET2 transactions.

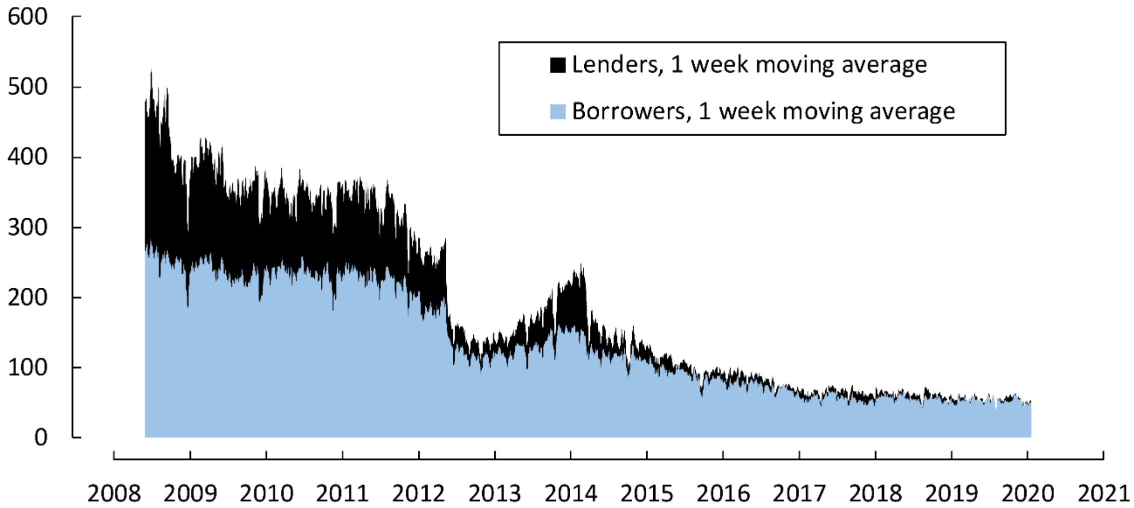
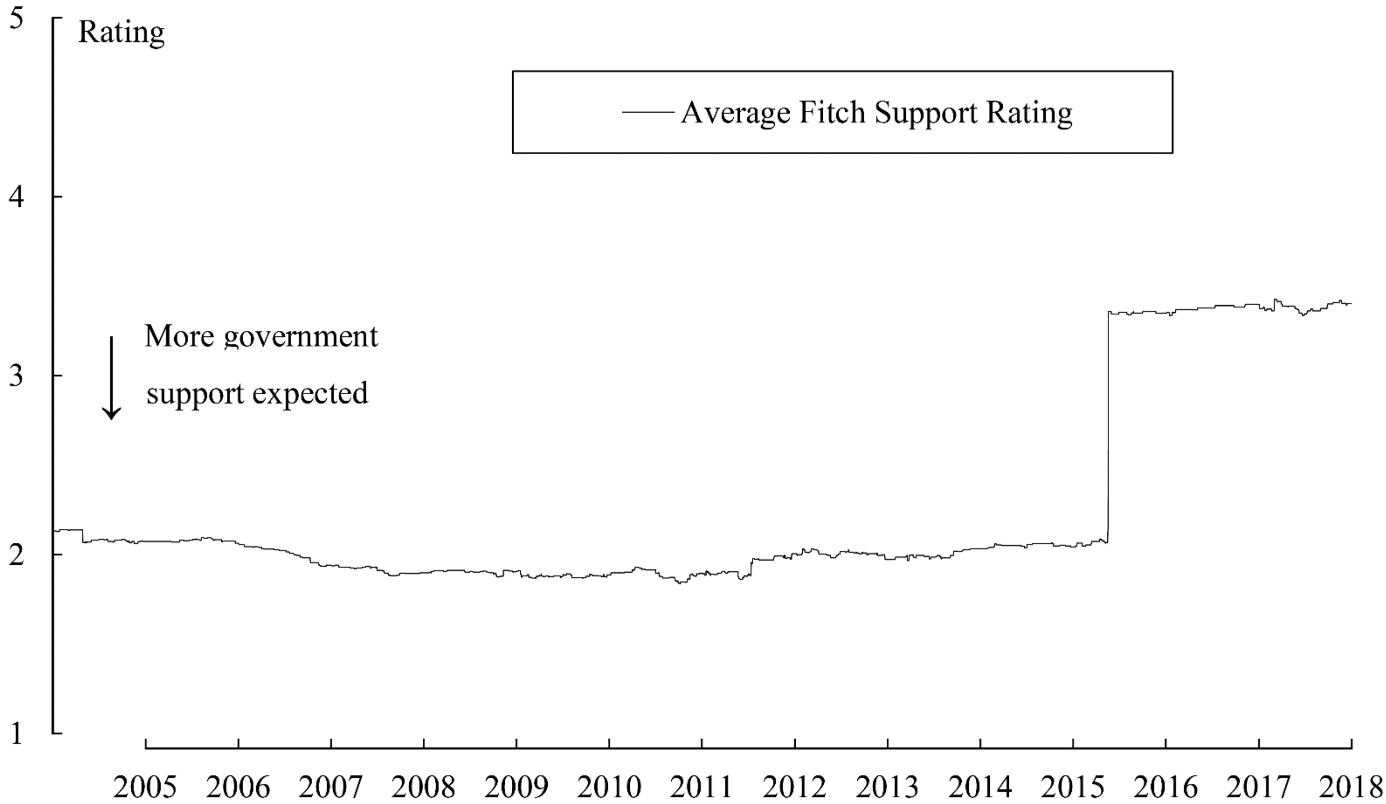


Figure 3. Historical development of Fitch Support Ratings for European Banks.

The different rating categories are 1: A bank for which there is an extremely high probability of external support. 2: A bank for which there is a high probability of external support. 3: A bank for which there is a moderate probability of support because of uncertainties about the ability or propensity of the potential provider of support to do so. 4: A bank for which there is a limited probability of support because of significant uncertainties about the ability or propensity of any possible provider of support to do so. 5: A bank for which there is a possibility of external support, but it cannot be relied upon.



Appendix:

Table 7. Data sources and definitions

Variable	Definition	Source	Type
Overnight rate spread	Loan specific borrowing rate minus EONIA rate.	ECB TARGET2	Pair-Day
Borrower preference index (BPI)	Total funds B has borrowed from L divided by total funds B has borrowed in the market. Based on past 90 calendar days.	ECB TARGET2	Pair-Day
Lender preference index (LPI)	Total funds L has lent to B divided by total funds L has lent in the market. Based on past 90 calendar days.	ECB TARGET2	Pair-Day
No. of days pairs	Number or days on which funds were sold by the given lender to during past 90 calendar days.	ECB TARGET2	Pair-Day
Transaction size	Amount borrowed by bank B from bank L on day t	ECB TARGET2	Pair-Day
Transaction size / Equity	Transaction size (above) divided by the borrower's total equity.	ECB TARGET2/Orbis	Pair-Day
Cross-border dummy	Used to identify home bias. 1, cross-border loan. 0, domestic loan.	ECB TARGET2	Pair-Day
Geographic concentration	Proxy for bank's geographic concentration calculated from the customer payments (payment type 1.1 in TARGET2) by forming a Herfindahl index from the shares of volume to each country within a year.	ECB TARGET2	Borrower-Day
No. of lenders	Number of lenders that sold funds to the given borrower during a year	ECB TARGET2	Borrower-Day
ECB deposit facility activity	Fraction of days in the on-going reserve maintenance period, the bank B made deposit to the ECB's deposit facility.	ECB TARGET2	Borrower-RMP
ECB marginal lending facility activity	Fraction of days in the on-going reserve maintenance period, the bank B withdraw a loan from the ECB's marginal lending facility.	ECB TARGET2	Borrower-RMP
Number of loans in ON market	Number of times that the bank B borrowed in the unsecured overnight market in the on-going reserve maintenance period.	ECB TARGET2	Borrower-RMP
Size	Logarithm of borrower’s total assets (see below).	Orbis Bank Focus	Borrower-Year
Total assets	Borrower's total assets	Orbis Bank Focus	Borrower-Year
ROA	Net income divided by total assets	Orbis Bank Focus	Borrower-Year
Tier 1 ratio	Total (Tier 1) capital as a fraction of risk-weighted assets	Orbis Bank Focus	Borrower-Year
NPL ratio	Impaired loans (i.e. non-performing loans) divided by gross loans	Orbis Bank Focus	Borrower-Year
Liquid assets / Deposits	Liquid assets divided by deposits	Orbis Bank Focus	Borrower-Year
Non-interest income / Revenue	Non-interest income divided by gross revenue	Orbis Bank Focus	Borrower-Year
G-SIFI dummy	1, institution is G-SIFI (global systemically important financial institution) or G-SIFI subsidiary. 0, otherwise.	Orbis Bank Focus	Borrower
State owned dummy	1, if global ultimate owned is public authority, state, or government. 0, otherwise.	Orbis Bank Focus	Borrower
Core state owned dummy	1, if state owned (as above) and additionally the state is one of the following: AT, BE, DE, DK, CH, FI, FR, GB, LU, NL, NO, SE. 0, otherwise.	Orbis Bank Focus	Borrower
Subject to reserve requirement dummy	1, if the borrower is subject to ECB's reserve requirement, and hence has access to ECB facilities. 0, otherwise.	ECB	Borrower
Exchange traded dummy	0, unlisted company. 1, listed company.	Orbis Bank Focus	Borrower
Country CDS	Index for 5-year, senior bonds	Thomson Reuters	Country-Day
iTraxx Financials CDS index	Equally weighted sub-index for financials of the iTraxx Europe index.	Bloomberg	Day
ON market concentration	Herfindahl index of the borrowing volumes on a given day.	ECB TARGET2	Day
ON market number of borrowers	Number of borrowers in the unsecured overnight market on a given day,	ECB TARGET2	Day
TARGET2 excess reserves	System level excess reserves in TARGET2. Calculated as Current account holdings + Overnight Deposits - Reserve requirement.	ECB	Day

Table 8. Country-specific BRRD implementation dates.

Source: ISDA (2016).

Country	Date bail-in provisions come into force
Switzerland	2012-11-01
Hungary	2014-09-16
Austria	2015-01-01
Germany	2015-01-01
Gibraltar	2015-01-01
Slovakia	2015-01-01
Croatia	2015-02-26
Estonia	2015-03-29
Denmark	2015-06-01
Latvia	2015-07-16
Bulgaria	2015-08-14
Greece	2015-11-01
Netherlands	2015-11-26
Lithuania	2015-12-03
Czech Republic	2016-01-01
Belgium	2016-01-01
Finland	2016-01-01
France	2016-01-01
Ireland	2016-01-01
Italy	2016-01-01
Luxembourg	2016-01-01
Malta	2016-01-01
Portugal	2016-01-01
Romania	2016-01-01
Spain	2016-01-01
United Kingdom	2016-01-01
Sweden	2016-02-01
Cyprus	2016-03-18
Slovenia	2016-06-25
Poland	2016-10-09
Liechtenstein	2017-01-01
Norway	2019-01-01
Iceland	2020-09-01

Table 9. List of bail-in and legal events and LTRO announcement dates.

Source: Bellia and Maccaferri (2020) and the ECB.

Event type	Event description	Date
BAIL-IN	Bail-in of senior debt of Amagerbanken	2011-02-06
BAIL-IN	Spanish bank rescue plan implies junior debt bail-in	2012-07-10
BAIL-IN	Germany backs Spanish bank rescue plan	2012-07-19
BAIL-IN	SNS Reeal nationalization	2013-02-01
BAIL-IN	Eurofi consider Cyprian senior debt bail-in	2013-02-11
BAIL-IN	Cyprian bail-in proposal	2013-03-18
BAIL-IN	Bail-in of senior debt of Cyprian banks	2013-03-25
BAIL-IN	Creditor bail-in Banco Espirito Santo	2014-08-04
BAIL-IN	MPS precautionary recapitalization	2016-12-29
BAIL-IN	Banco Popularresolution	2017-06-06
BAIL-IN	EC approves MPS recapitalization	2017-07-04
BAIL-IN	NordLB State Aid approved by EC	2019-12-05
LEGAL	Spanish authories suggest national bank resolution law	2012-08-23
LEGAL	Eurofi agrees on BRRD rules	2013-06-28
LEGAL	SRM proposal	2013-07-09
LEGAL	EU council accepts SRM approach	2013-12-18
LEGAL	EU parliament preliminary accepts SRM approach	2014-03-20
LEGAL	EU parliament accepts SRM approach	2014-04-15
LEGAL	EC proposes banking reform package	2016-11-23
LEGAL	EU council reaches agreement on banking package	2018-05-25
LEGAL	EP votes banking package	2018-06-19
LEGAL	Final agreement on banking package	2019-04-16
LEGAL	Publication of banking package legislation	2019-06-07
LEGAL	Applications of TLAC requirements	2019-06-27
LTRO	ECB's LTRO (allotted amount = 120 bn)	2008-09-30
LTRO	ECB's LTRO (allotted amount = 103.1 bn)	2008-10-30
LTRO	ECB's LTRO (allotted amount = 134.9 bn)	2008-12-10
LTRO	ECB's LTRO (allotted amount = 113.4 bn)	2009-01-21
LTRO	ECB's LTRO (allotted amount = 104.7 bn)	2009-02-11
LTRO	ECB's LTRO (allotted amount = 120.2 bn)	2009-03-11
LTRO	ECB's LTRO (allotted amount = 131.8 bn)	2009-04-08
LTRO	ECB's LTRO (allotted amount = 116.1 bn)	2009-05-13
LTRO	ECB's LTRO (allotted amount = 442.2 bn)	2009-06-25
LTRO	ECB's LTRO (allotted amount = 131.9 bn)	2010-07-01
LTRO	ECB's LTRO (allotted amount = 104 bn)	2010-09-30
LTRO	ECB's LTRO (allotted amount = 149.5 bn)	2010-12-23
LTRO	ECB's LTRO (allotted amount = 129.5 bn)	2011-03-31
LTRO	ECB's LTRO (allotted amount = 132.2 bn)	2011-06-30
LTRO	ECB's LTRO (allotted amount = 140.6 bn)	2011-09-29
LTRO	ECB's LTRO (allotted amount = 489.2 bn)	2011-12-22
LTRO	ECB's LTRO (allotted amount = 529.5 bn)	2012-03-01
LTRO	ECB's LTRO (allotted amount = 129.8 bn)	2014-12-17
LTRO	ECB's LTRO (allotted amount = 399.3 bn)	2016-06-29
LTRO	ECB's LTRO (allotted amount = 233.5 bn)	2017-03-29
LTRO	ECB's LTRO (allotted amount = 109.1 bn)	2020-03-18
LTRO	ECB's LTRO (allotted amount = 115 bn)	2020-03-25
LTRO	ECB's LTRO (allotted amount = 1308.4 bn)	2020-06-24

Supplement to Tölö et al. (2021), Have Too-Big-To-Fail Expectations Diminished? Evidence from the European Overnight Interbank Market

1. Yearly evaluation of the determinants of overnight rates

Table S1 shows that there is significant variation from year to year in the estimated coefficients of the key explanatory variables. For the central parts, the yearly results however follow a similar pattern as those in Table 2 of the main article. The coefficient on bank size is generally the most consistent determinant of overnight borrowing costs, having the expected sign in all years and being statistically significant in almost all of them. Among the borrower credit risk variables, the ROA, Tier 1 ratio, and relative transaction size (measured by the log of the ratio of loan value to equity) are statistically significant in the full sample but often not in individual years. To some extent the magnitude of the bank size coefficient compared to that of the relative transaction size are inversely related. Further, the transaction size (log of loan value) dummies are not very informative when the relative transaction size is accounted for. Note that in the yearly models of Table S1 we control for the effect of lending-relationships with simpler measures than in Table 2, given only one year of data per model.

Year 2008 differs markedly from the rest of the sample years in Table S1. The data span only the latter half of that year so some of the balance sheet variables may not fully reflect the developments in that turbulent year. Moreover, the rapid repricing of risk starting from the sudden collapse of investment bank Lehman Brothers is likely behind the high importance of the Tier 1 ratio and the relative transaction size variables in that year.[1]

At the bottom of Table S1a, we include joint significance tests for groups of credit risk and relationship variables. The tests show that borrower credit risk and relationship characteristics are important determinants of overnight borrowing costs throughout the sample years.

We conclude from the above considerations that fixed country effects could enhance the estimation of some of the determinants, and hence we replicate the models of Table S1a in Table S1b with country fixed effects (not present in Furfine (2001), which was based on Fedwire data). Compared to Table S1a, the bank size premium becomes somewhat less precise and smaller. This is largely explained by the fact that the fixed effects absorb the effect of bank size in countries where there is less variation in banks' size.

Table S1. Replication of the analysis by Furfine (2001) for different years

a) Estimating equation includes time fixed effects only

Variable	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2008-2020
Size	-1.002	-1.814	-1.655	-1.356	-1.010	-2.283	-1.276	-1.454	-2.270	-2.334	-1.843	-1.241	-0.839	-1.421
	[1.9]	[4.1]	[5.4]	[4.7]	[1.9]	[3.4]	[4.0]	[5.3]	[5.8]	[5.1]	[4.7]	[3.1]	[1.5]	[7.1]
Borrower's credit risk:														
NPL ratio	-0.197	-0.345	0.0908	0.075	-0.033	0.213	0.085	0.036	-0.075	-0.130	0.005	-0.151	-0.0629	-0.125
	[1.1]	[2.1]	[1.6]	[0.9]	[0.3]	[1.5]	[1.0]	[0.7]	[1.5]	[3.2]	[0.1]	[1.1]	[0.2]	[3.3]
ROA	-0.380	-1.848	1.210	-0.102	-0.753	0.486	-0.441	-0.493	0.291	-0.861	5.005	3.561	3.584	-0.973
	[0.5]	[4.3]	[3.0]	[0.1]	[0.7]	[0.6]	[1.3]	[1.0]	[0.8]	[0.8]	[3.3]	[1.6]	[1.9]	[3.1]
Tier 1	-0.774	-0.568	-0.304	-0.209	-0.060	0.095	0.185	0.259	-0.273	-0.1852	0.014	-0.085	0.1444	-0.113
	[3.9]	[3.5]	[3.2]	[2.0]	[0.3]	[0.7]	[2.7]	[1.2]	[2.6]	[1.8]	[1.3]	[1.0]	[1.1]	[3.1]
Log (Loan value/equity)	3.620	1.275	0.959	2.446	1.342	0.102	0.486	-0.346	-0.483	-0.074	-0.073	0.161	0.331	1.289
	[7.0]	[4.6]	[5.0]	[9.0]	[2.8]	[0.3]	[2.6]	[1.4]	[1.3]	[0.2]	[0.3]	[0.6]	[1.0]	[7.8]
Transaction characteristics:														
Log (Loan value) <10	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base
10<Log (Loan value) <100	1.398	1.131	1.343	-0.656	-0.367	1.572	-0.654	-0.323	-0.413	1.163	0.365	-1.200	-2.962	0.183
	[1.4]	[1.8]	[3.4]	[1.1]	[0.4]	[1.5]	[1.4]	[0.5]	[0.5]	[1.1]	[0.5]	[1.3]	[2.8]	[0.5]
Log (Loan value)>100	1.700	2.955	1.843	-0.480	-0.354	1.099	-0.703	-0.924	-0.044	1.514	0.620	-0.564	-1.883	0.803
	[1.0]	[3.2]	[2.9]	[0.5]	[0.2]	[0.7]	[1.0]	[0.9]	[0.0]	[1.1]	[0.6]	[0.5]	[1.4]	[1.4]
Relationship characteristics														
Log (no. of days pair)	-16.238	-7.389	-6.314	-11.469	7.944	-10.486	-3.281	-8.075	1.053	0.104	1.882	-1.938	4.988	-2.282
	[20.3]	[12.9]	[7.2]	[4.8]	[2.6]	[3.2]	[4.0]	[1.7]	[0.3]	[0.1]	[0.6]	[0.8]	[0.4]	[0.6]
Log (no. of days pair)	15.758	7.500	6.436	10.897	-7.696	10.994	3.488	8.501	-0.233	-0.486	-2.818	1.290	-5.819	2.296
*(small borrower)	[20.5]	[12.3]	[6.9]	[4.5]	[2.5]	[3.2]	[3.9]	[1.8]	[0.1]	[0.3]	[0.8]	[0.5]	[0.4]	[0.6]
Log (no. of lenders)	-1.223	-0.902	-1.639	-1.308	-3.385	-0.887	-0.501	-2.224	-1.574	0.404	0.035	-0.349	0.012	-1.136
	[2.5]	[2.0]	[3.5]	[3.3]	[5.6]	[1.8]	[1.4]	[4.2]	[2.2]	[0.8]	[0.1]	[0.8]	[0.0]	[4.4]
Fixed effects	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time
Banks	376	460	484	503	514	464	515	422	355	323	272	269	186	1,145
Pairs	5,989	7,450	7,102	7,193	4,461	3,590	4,440	2,760	1,878	1,232	965	971	573	19,028
Observations	43,837	109,597	107,495	108,624	61,628	46,222	57,228	40,317	31,893	20,616	15,965	14,819	7,045	665,286
R-squared	0.325	0.406	0.445	0.391	0.563	0.526	0.466	0.582	0.528	0.635	0.619	0.615	0.457	0.382
Joint significance of the four credit risk variables														
Wald statistics	23.85	19.49	11.57	24.20	2.74	0.80	5.41	1.17	2.52	4.08	3.87	2.03	1.93	21.34
p-value	0.000	0.000	0.000	0.000	0.027	0.523	0.000	0.321	0.039	0.003	0.004	0.088	0.104	0.000
Joint significance of the three relationship characteristics														
Wald statistics	145.06	56.29	19.47	19.37	24.15	4.13	5.99	6.87	1.89	0.65	8.73	5.85	4.14	10.94
p-value	0.000	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.129	0.582	0.000	0.001	0.006	0.000

The estimating equation is Eq. (1). The dependent variable is the overnight rate spread. Same notes as in Table 2 apply. The control variables are adapted from Furfine (2001), and hence different from Table 2.

Variable	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2008-2020
Log (Total assets)	-2.122	-2.412	-2.746	-3.424	-2.970	-2.939	-1.877	-1.541	-1.345	-1.298	-0.981	-1.452	-1.610	-2.469
	[12.4]	[9.1]	[11.2]	[16.2]	[8.3]	[6.1]	[9.1]	[6.3]	[3.9]	[3.1]	[2.8]	[6.0]	[4.2]	[14.8]
Fixed effects	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time	Country + Time
Banks	610	614	645	668	687	584	633	484	406	372	315	300	208	1,397
Pairs	9,551	8,312	7,920	8,088	5,056	3,987	4,842	2,921	1,997	1,327	1,066	1,026	605	22,603
Observations	91,202	119,184	116,914	119,103	68,193	51,425	62,302	42,119	33,510	21,524	18,126	15,718	7,280	766,600
R^2	0.262	0.369	0.452	0.371	0.484	0.532	0.468	0.540	0.475	0.607	0.549	0.594	0.478	0.365

In the lower panel, all other explanatory variables except the bank size are left out.

b) Estimating equation includes both country and time fixed effects

Variable	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2008-2020
Size	-0.644	-2.057	-1.995	-1.984	-1.546	-1.999	-1.076	-2.166	-2.873	-3.545	-2.272	-2.251	-1.037	-1.788
	[1.1]	[6.3]	[7.2]	[6.4]	[2.5]	[4.0]	[5.3]	[5.8]	[6.8]	[8.6]	[4.1]	[3.6]	[1.6]	[9.6]
Borrower's credit risk:														
NPL ratio	-0.173	-0.229	0.1425	0.248	0.057	0.162	0.249	0.177	-0.086	-0.091	0.160	0.310	0.7322	0.036
	[1.0]	[1.7]	[2.3]	[2.9]	[0.6]	[1.2]	[3.3]	[3.1]	[1.3]	[1.4]	[1.9]	[1.9]	[3.0]	[1.1]
ROA	2.853	-1.062	1.177	-0.530	-2.016	-0.255	0.646	-0.069	0.978	0.757	2.975	4.119	1.820	-0.448
	[3.2]	[2.8]	[2.4]	[0.8]	[2.1]	[0.3]	[1.5]	[0.1]	[2.2]	[0.8]	[3.1]	[2.5]	[1.3]	[1.6]
Tier 1	-0.767	-0.345	-0.322	-0.314	-0.023	0.271	0.234	0.150	-0.617	-0.4993	-0.034	-0.389	-0.0106	-0.118
	[4.1]	[2.7]	[3.7]	[3.0]	[0.1]	[1.7]	[2.4]	[0.6]	[3.8]	[4.4]	[1.6]	[2.9]	[0.1]	[3.1]
Log (Loan value/equity)	4.041	0.983	0.915	2.757	1.020	0.165	0.545	-0.405	-1.042	-1.039	-0.540	-0.666	0.342	1.219
	[6.5]	[3.3]	[3.3]	[8.1]	[1.8]	[0.3]	[2.8]	[0.9]	[2.0]	[2.0]	[0.9]	[0.9]	[0.8]	[6.4]
Transaction characteristics:														
Log (Loan value) <10	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base	Base
10<Log (Loan value) <100	2.260	1.174	1.135	-1.872	-2.119	-0.744	-1.589	-1.104	-2.277	1.297	-1.070	-2.589	-3.294	-0.654
	[1.9]	[1.7]	[1.9]	[2.4]	[1.6]	[0.5]	[2.1]	[1.1]	[2.1]	[1.0]	[0.8]	[1.8]	[2.6]	[1.4]
Log (Loan value)>100	1.662	3.029	1.999	-1.178	0.105	-0.628	-1.790	-1.982	-0.224	3.763	0.858	-0.711	-2.384	0.099
	[0.8]	[2.9]	[2.0]	[0.9]	[0.1]	[0.3]	[1.8]	[1.3]	[0.1]	[2.1]	[0.5]	[0.4]	[1.5]	[0.1]
Relationship characteristics														
Log (no. of days pair)	-7.989	-5.569	-5.901	-10.828	10.066	-10.350	-3.991	-1.386	5.015	14.628	0.015	11.482	30.303	-0.219
	[9.7]	[12.8]	[7.4]	[5.0]	[3.1]	[3.9]	[4.1]	[0.7]	[1.6]	[5.8]	[0.0]	[11.8]	[4.9]	[0.1]
Log (no. of days pair)	8.087	5.767	6.099	10.433	-9.775	11.555	4.596	2.364	-4.448	-15.509	-1.457	-12.188	-31.758	0.399
*(small borrower)	[10.0]	[12.0]	[6.7]	[4.7]	[2.9]	[3.7]	[4.4]	[1.1]	[1.4]	[6.1]	[0.3]	[10.8]	[5.2]	[0.1]
Log (no. of lenders)	-0.676	-0.434	-1.579	-0.999	-3.118	-3.764	-1.781	-4.257	-2.957	-0.773	-0.255	-1.389	-1.920	-1.621
	[1.2]	[1.2]	[3.3]	[2.8]	[4.4]	[4.4]	[5.5]	[7.0]	[3.5]	[1.6]	[0.6]	[3.0]	[4.4]	[6.4]
Fixed effects														
	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time
Banks	376	460	484	503	514	464	515	422	355	323	272	269	186	1,145
Pairs	5,989	7,450	7,102	7,193	4,461	3,590	4,440	2,760	1,878	1,232	965	971	573	19,028
Observations	43,837	109,597	107,495	108,624	61,628	46,222	57,228	40,317	31,893	20,616	15,965	14,819	7,045	665,286
R^2	0.216	0.282	0.291	0.259	0.319	0.208	0.270	0.323	0.241	0.308	0.265	0.275	0.227	0.274
Joint significance of the four credit risk variables														
Wald statistics	20.96	6.22	8.39	27.62	2.71	0.83	5.32	2.82	4.98	6.44	3.45	3.93	2.65	16.26
p-value	0.000	0.000	0.000	0.000	0.028	0.509	0.000	0.024	0.001	0.000	0.008	0.004	0.032	0.000
Joint significance of the three relationship characteristics														
Wald statistics	34.10	58.65	23.43	13.73	25.32	13.06	18.05	24.73	9.67	21.76	13.58	60.71	30.72	26.72
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The estimating equation is Eq. (1). The dependent variable is the overnight rate spread. Same notes as in Table 2 apply.

Variable	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2008-2020
Size	-2.198	-2.551	-3.167	-4.060	-2.549	-2.872	-2.189	-1.861	-1.494	-1.983	-1.768	-2.313	-2.440	-2.746
	[15.8]	[12.3]	[16.7]	[26.9]	[9.0]	[5.8]	[8.8]	[8.1]	[5.3]	[5.6]	[4.6]	[9.1]	[6.5]	[21.5]
Fixed effects														
	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time
Banks	610	614	645	668	687	584	633	484	406	372	315	300	208	1,397
Pairs	9,551	8,312	7,920	8,088	5,056	3,987	4,842	2,921	1,997	1,327	1,066	1,026	605	22,603
Observations	91,202	119,184	116,914	119,103	68,193	51,425	62,302	42,119	33,510	21,524	18,126	15,718	7,280	766,600
R^2	0.184	0.253	0.275	0.226	0.198	0.146	0.221	0.167	0.108	0.192	0.158	0.182	0.158	0.246

Except for the fixed effects, the whole table is the same as Table S1a.

2. Alternative difference-in-differences setup

It is possible to come up with an alternative formulation whereby the treatment group is "large banks" and the non-treated group is "small banks". Table S2 shows the result of the event study remain essentially unchanged if we base it on G-SIFI dummy instead of the bank size. Bail-in and LTRO events cause a change in the coefficient of size while legal events do not.

Table S2. Estimates of the impact of grouped events using bank size or G-SIFI dummy.

Variable	X = G-SIFI dummy ON rate	X = Log(Total assets) ON rate
X	-0.062	-0.026
	[17.0]	[19.2]
BAILIN_EVENT*X	0.013	0.006
	[3.4]	[5.7]
LEGAL_EVENT*X	0.006	0.001
	[1.5]	[0.6]
LTRO_EVENT*X	0.012	0.006
	[4.1]	[9.4]
+ EVENT CONTROLS		
Fixed effects	None	None
N	767,574	701,668
Pairs	20,987	18,347
Banks	743	680
R^2	0.087	0.199

G-SIFI denotes a global systemically important financial institution or subsidiary. The estimating equation is Eq. (5). The dependent variable is the overnight rate spread per loan. Notes of Table 2 apply but sample is further restricted to the European Economic Area (EEA) and the 20-day window around event dates.

3. Alternative dependent variable

Table S3 shows the basic specifications of Table 2 from the main article using two alternative dependent variables, nominal overnight rate spread (R) and relative overnight rate spread R/EONIA, when restricted to the sample with EONIA > 0.5 %. The coefficients have largely similar signs and statistical significance. Coincidentally, these two dependent variables have roughly the same standard deviations (0.210 for R and 0.205 for R/EONIA), and hence the magnitude of the coefficients is also very similar.

References:

Furfine, C.H. (2001): Banks as monitors of other banks: Evidence from the overnight Federal Funds market. Journal of Business 74: 1, 33-57. Available at: https://doi.org/10.1086/209662 .

Table S3. Estimates with alternative dependent variable.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variable	R	R/EONIA	R	R/EONIA	R	R/EONIA	R	R/EONIA	R	R/EONIA
Size	-3.335	-3.312	-3.318	-3.272	-3.345	-3.170	-3.487	-3.433	-3.176	-2.962
	[26.5]	[27.6]	[29.6]	[29.4]	[20.5]	[20.3]	[23.8]	[22.8]	[12.9]	[11.8]
ROA			-1.000	-0.709	-0.053	0.024	-0.616	-0.531	0.002	0.464
			[2.9]	[2.1]	[0.1]	[0.0]	[1.6]	[1.4]	[0.0]	[1.0]
Tier 1			-0.637	-0.547	-0.807	-0.699	-0.734	-0.602	-0.637	-0.604
			[9.0]	[8.0]	[9.6]	[9.2]	[8.5]	[7.2]	[7.3]	[7.4]
CDS					0.812	0.442				
					[4.0]	[2.9]				
Non-interest income							-0.020	-0.022	-0.018	-0.018
							[7.1]	[8.1]	[7.4]	[8.6]
NPL ratio							0.069	-0.005		
							[0.9]	[0.1]		
Liquid assets/Deposits							-0.001	-0.007	0.017	0.017
							[0.1]	[1.0]	[1.9]	[1.9]
G-SIFI dummy							2.381	1.907	-0.676	-0.897
							[4.1]	[3.0]	[0.8]	[1.0]
Listed bank dummy							-0.967	-0.582	1.795	1.431
							[2.1]	[1.2]	[3.4]	[2.7]
Subject to RR									-5.925	-6.332
									[8.7]	[9.8]
ECB deposit facility activity									-11.305	-11.490
									[11.4]	[12.0]
ECB lending facility activity									18.850	24.806
									[5.0]	[5.4]
Log(No. of lenders)									-2.245	-4.163
									[4.9]	[9.4]
Log(Loan count)									2.098	3.996
									[7.5]	[14.5]
Log(Loan value)									2.482	2.742
									[13.6]	[14.9]
Geographic concentration									12.306	10.001
									[10.7]	[8.0]
Fixed effects	Time	Time	Time	Time	Time + Country	Time + Country	Time	Time	Time + Country	Time + Country
N	289,565	289,565	274,029	274,029	247,962	247,962	265,416	265,416	184,251	184,251
Pairs	16,421	16,421	15,087	15,087	13,363	13,363	14,617	14,617	9,383	9,383
Borrower banks	923	923	758	758	653	653	719	719	264	264
R^2	0.241	0.247	0.251	0.246	0.361	0.355	0.261	0.254	0.411	0.429

The estimating equation is Eq. (1) but with two alternative dependent variables R and R/EONIA, which correspond to nominal and relative overnight rate spreads. Notes of Table 2 apply. Additionally, the sample is restricted to periods when EONIA>0.5 to avoid division by a very small denominator.