BANK OF ENGLAND

Staff Working Paper No. 922

Measure for measure: evidence on the relative performance of regulatory requirements for small and large banks

Austen Saunders and Matthew Willison

May 2021

BANK OF ENGLAND

Staff Working Paper No. 922

Measure for measure: evidence on the relative performance of regulatory requirements for small and large banks

Austen Saunders[1] and Matthew Willison[2]

Abstract

This paper compares the performance of regulatory thresholds as predictors of distress for large banks with their performance for small banks. Using a data set of capital and liquidity ratios for a sample of UK-focused banks in 2007, we apply simple threshold-based rules to assess how regulatory thresholds might have identified banks that subsequently became distressed. We compare results for large banks with results for small banks, optimising thresholds separately for the two groups. Our results suggest that the regulatory ratios we use are better aligned with risks which cause distress of large banks than with those which cause distress of small banks. We find that when thresholds are set to correctly identify a high proportion of banks which subsequently became distressed, they generate materially lower false alarm rates for large banks than for small. This result is robust to definitional choices and to resampling. We also test whether supervisors' judgements about the quality of banks' governance have predictive power with regard to distress. We find that adding supervisors' judgements to regulatory ratios improves predictions for small banks but not for large banks.

Key words: Banking regulation, Basel III, bank failure, global financial crisis, regulatory complexity.

JEL classification: G01, G21, G28.

(1) Bank of England. Email: austen.saunders@bankofengland.co.uk
(2) Bank of England. Email: matthew.willison@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are grateful to Marcus Buckmann, Peter Eccles, William Francis, Paul Grout, Alison Scott, and Joel Suss, and participants at a Bank of England seminar in March 2021 for their assistance and helpful comments.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Regulatory requirements constrain banks' risk taking in order to make them more stable. Different banks, however, have different characteristics and the effectiveness of regulatory requirements might vary with these characteristics. For example, some banks are bigger than others and it might be the case that requirements which are effective for large banks are less effective for small banks. Knowing whether this is the case would help regulators set requirements for different sorts of banks. In particular, evidence about the relative performance of requirements when applied to large and to small banks would aid policy makers considering how to apply to small banks regulatory reforms adopted since the Global Financial Crisis of 2007-8. Among other reforms, the package of international banking regulations known as Basel III introduced a new funding ratio known as the Net Stable Funding Ratio ('NSFR'), widened the application of a leverage ratio to global banks, and made changes to how risk based capital ratios are calculated.[1] Because Basel requirements apply directly to mostly large, internationally active banks, it may be the case that they have been designed in ways which make them better aligned with the risks faced by large banks than with those faced by small ones. That could imply that imposing the full set of Basel requirements on small banks which are not internationally active delivers fewer benefits than imposing them on large banks.

Establishing whether or not this is the case involves measuring the effectiveness of requirements when applied to different banks. One way to do this is to measure the ability of requirements to discriminate ex-ante between distressed and non-distressed banks when a banking system is hit by a systemic shock. This is the approach taken by both Aikman et al. (2018) and Buckmann et al. (2021) to test whether thresholds set for multiple regulatory ratios outperform thresholds set for single ratios as predictors of bank distress. They conclude that they do. But whilst Buckmann et al. (2021) consider differences between large and small banks by comparing results for banks with total assets above and below their sample median, they caution against drawing conclusions from that exercise given the relatively large size of all the banks in their sample (their sample is of 116 global banks each with assets of more than $100 billion).

In order to provide new insights into whether the effectiveness of regulatory requirements varies according to the size of the banks to which they apply, this paper applies the methodology used by Aikman et al. (2018) and Buckmann et al. (2021) to data for 118 UK-focused banks. Because these range in size from £3 million to over £1 trillion of total assets (with a median of £500 million), we are able to compare the performance of metrics between groups of banks which vary in size by orders of magnitude. When we use regulatory data from mid-2007 to assess thresholds set for the leverage ratio, risk weighted capital ratio, and NSFR as predictors of distress over the following 18 months, we find that the thresholds perform better for large banks than for small banks. We find that this result is robust to definitional choices regarding distress and size, and to random re-sampling of our data. We extend our analysis to consider the performance of supervisors' scores for the quality of banks' governance as a predictor of distress. We find that those scores provide additional predictive power for small banks, but not for large.

This paper is related to the literature on determinants of bank failure and early-warning models of bank distress. Rather than reviewing the literature in its entirety here, we will compare and contrast our paper with several representative papers from that literature. Our overall approach is based on the aforementioned papers by Aikman et al. (2018) and Buckmann et al. (2021). This approach differs from the approaches taken in other papers, which typically use probit or logit models to estimate the probability that a bank fails within a given interval of time (e.g. Cole and White (2012)) or estimate a bank's expected survival time (e.g. Cole and Gunther (1995), Whalen (1991)). Our approach of using supervisory assessments to measure bank distress and then seeking to predict those assessments using lagged bank-level variables compares to the approaches taken, for example, in Cole et al. (1995) and Suss and Treitel (2019). It contrasts with other approaches that classify banks as distressed based on actual failures or claims made on deposit insurance schemes (e.g., Cole and Gunther (1995) and Coen et al. (2019)) or reports of distress in the media (e.g. Poghosyan and Čihak (2011)). Our focus

[1] The NSFR measures the stability of a bank's funding profile. Considering a time horizon of one year, it is calculated as the ratio of 'available stable funding' (the portion of capital and liabilities expected to be reliable over the next year) to 'required stable funding' (calculated having taken into account the liquidity characteristics and residual maturity of the bank's assets). The ratio should be at least 100% on an ongoing basis. See Basel Committee on Banking Supervision (2014).

on using regulatory ratios as predictors also contrasts with the literature, which typically use in addition other relevant bank-level variables that may predict distress (e.g. Francis (2014), Cole and White (2012)). We do not consider a larger set of bank-level variables because we are not seeking to estimate an early-warning model to identify banks heading for difficulties. For the same reason, we do not include macroeconomic variables or measures of inter-bank contagion, which have been shown to predict bank distress (see Mare (2015) and Poghosyan and Čihak (2011), respectively). The differences between small and large banks has been considered in the literature before; Cole and White (2012) did not find that the predictors of bank distress changed significantly when they split their sample by size. Alzugaiby et al. (2020), however, find predictors of failure do vary between banks of different sizes; for instance, equity as a proportion of total assets performs well as a predictor of medium-sized and large banks, but not for small banks. Finally, several papers have investigated whether the quality of bank management or governance predict distress. Francis (2014) and Wheelock and Wilson (2000) both find less efficient banks are more likely to get into distress. Alzugaiby et al. (2020) find bank efficiency performs well as a predictor for small banks, but not for larger banks. Berger et al. (2016) find ownership of shares among management below the CEO-level is positively associated with the risk of bank failure. We use supervisory assessments of the quality of governance in banks rather than measures of bank efficiency or management remuneration.

The paper proceeds as follows. Section 2 describes the data we use. Section 3 outlines the methodology. The results are set out in section 4 whilst section 5 sets out results when we also use supervisor's scores for the quality of governance. Section 6 concludes.

2 Data

We use regulatory data submitted in 2007 by 118 UK-focused banks.[2] Data are taken from the Bank of England's Historical Banking Regulatory Database ('HBRD'), from which we take each bank's total assets, leverage ratio, risk weighted capital ratio, and a proxy for its NSFR estimated from regulatory data about its assets and liabilities.[3] For each variable we take the last reported value as of 1 July 2007. These are data reported on 30 June 2007 for the vast majority of banks, but reported earlier in 2007 for 11 banks and in 2006 for 4. We combine these regulatory data with records of supervisors' judgements about the quality of banks' governance. These judgements take the form of scores allocated by supervisors on a scale of 1 to 10 where one is best and 10 worst (the average allocated score was four).[4] Again, for each bank we take the last reported score as of 1 July 2007. These date from no earlier than 30 September 2006. Governance scores are available for 89 of the banks for which we have regulatory data from HBRD (the scoring system was relatively new in 2007, and supervisors had yet to allocate scores to all regulated banks). Observations for governance scores are available across the size distribution; 24 of the allocated scores were for banks in the upper size quartile whilst 65 were for banks in the lower three quartiles. Summary statistics for all five variables are shown in Table 1.

We identify those banks in our sample which became distressed during the Global Financial Crisis using supervisors' judgements from the period. As well as allocating scores for individual risk factors (such as governance), supervisors awarded an overall score where one meant least risk to the bank's safety and soundness (according to supervisors' judgement) and 10 the highest risk. We define as distressed banks which received a score of 10 (the worst possible supervisory rating) any time between 1 July 2007 and 31 December 2008.

[2]'UK-focused banks' are banks and building societies which were headquartered in the UK in 2007, as well as subsidiaries of overseas banks when those subsidiaries had significant retail or SME banking businesses in the UK.

[3]On HBRD, see de Ramon et al. (2017). The NSFR proxy is calculated as the ratio of a proxy for Available Stable Funding to a proxy for Required Stable Funding, where the proxy for Available Stable Funding is calculated as $0.7 \times Non\,Financial\,Deposits + Total\,Capital$ and the proxy for Required Stable Funding is $0.85 \times Loans + Other\,Non\,Liquid\,Banking\,Book\,Assets + 0.5 \times Non\,Liquid\,Trading\,Assets + 0 \times Liquid\,Assets$.

[4]See Suss and Treitel (2019) for a full account of how scores were allocated. In summary, scores were allocated by line supervisors with reference to the amount of risk posed to the FSA's objectives (a score of one meant low risk to these objectives). Supervisors carried out periodic assessments of each bank before setting scores which were then reviewed by an internal panel of senior management and risk specialists. These assessments took place every two or three years, but scores could be updated between reviews at the discretion of line supervisors. Our data include these interim updates, which were frequently made for both large and small banks (see Table 2).

Figure 1 summarizes the observation periods for the data we use to make predictions and for the data we use to measure distress.

Supervisors' overall scores allocated during this period are available for 102 banks, 22 of which are defined as distressed by our measure and 80 as not distressed. Scores for all banks were updated regularly. Table 2 shows, for each calendar quarter between the middle of 2007 and the end of 2008, the number of banks whose supervisors made a change to any of their risk scores (either the overall score or the scores for an individual risk factor). It shows that each calendar quarter the scores of between one third and three quarters of the banks in our sample were updated (this does not include cases where scores were reviewed by supervisors but no changes were made). This is true of both 'large' banks (those with more than £5 billion of assets) and 'small' banks (those with less than £5 billion of assets). It is also true of banks which we count as distressed and of those which did not become distressed according to our measure (the rows in Table 2 for distressed banks count updates to scores for banks which became distressed at any point during the observation period, not just those counted as distressed in a given quarter). In consequence, scores remained relatively fresh as supervisors took into account new information. This is important because our measure of distress uses the highest overall score allocated to each bank at any point between July 2007 and December 2008 (which might not necessarily be the last score allocated during that period). We might therefore undercount instances of distress if supervisors were not regularly updating their scores for some banks to take into account developments during the latter part of the observation period.

Partly in consequence of the regularity with which scores were updated, banks entered distress (i.e. first received an overall score of 10) throughout our observation period (see Table 3). Again, this is true of both large and small banks. This means that the balance sheet data we use to make our predictions are no more or less likely to be very recent for a large or a small distressed bank.

The distribution of banks in our sample by total assets is highly skewed with many small banks (the median bank has total assets of £520 million) and a small number of very large banks (the largest two of which have assets of over £1 trillion). Figure 2 shows the cumulative distribution by total assets, with banks which meet our distress definition shown separately from those which do not. Whilst distress observations occur at all sizes, distress observations are somewhat more concentrated among the bottom size quartile (banks with less than £160 million of total assets) and top quartile (banks with more than £4.7 billion of total assets). See Table 4.

Figure 3 shows the relationships between total assets and each ratio, and between the ratios themselves. Larger banks in our sample tend to have lower capital ratios, whether measured as a leverage ratio or risk weighted capital ratio. Leverage and risk weighted ratios are themselves clearly related, but they are not perfectly correlated. This reflects different risk weight densities across banks in our sample. Larger banks in our sample also tend to have lower NSFRs, meaning that banks with low capital ratios tend to also have low NSFRs. Table 5 shows correlation coefficients for the same relationships. Correlations between all pairs of ratios are highly significant. When distressed and not distressed banks are separated, almost all correlations remain significant with the same sign (the exception is the coefficient for leverage ratio and NSFR for distressed banks, which is not significant). Relationships between ratios are therefore similar for distressed and for not distressed banks.

Turning to supervisors' scores for governance, we find that the very best scores for governance were only allocated to banks with less than £10 billion total assets (Figure 4). Larger banks tended to be given worse scores. Having a low leverage ratio did not prevent a bank receiving a good score for governance whilst, conversely, few banks have both a high leverage ratio and a good score for governance. In contrast, high NSFRs are associated with good scores for governance. The distributions of governance scores for both distressed and not distressed banks is centred around four (Figure 5). The distribution for distressed banks has proportionally fatter tails on both sides, with not distressed banks being slightly more likely to have a governance score below four than above.

3 Methodology

We adopt a technique used by Aikman et al. (2018) and Buckmann et al. (2021) to measure the performance as predictors of distress of thresholds applied to regulatory ratios. Readers interested in detailed accounts of the methodology should refer to those papers.[5]

We make predictions by specifying minimum thresholds for each of three regulatory ratios: leverage ratio, risk weighted capital ratio, and NSFR. Distress is predicted for any bank which in 2007 would have failed to meet at least one of those thresholds. Predictions are compared with actual outcomes as defined by our definition of distress. Table 6 summarises how we describe the performance of thresholds by classifying predictions as 'hits', 'misses', 'false alarms', or 'correct rejections'. From this we calculate a 'hit rate' (the proportion of actually distressed banks for which distress is predicted) and a 'false alarm rate' (the proportion of banks which were not actually distressed but for which distress is predicted).

When different thresholds are specified, different hit rates are generated. Only certain hit rates are possible, there being a fixed number of distressed banks (22) according to our definition of distress. For each hit rate, we find the combination of thresholds which generates that hit rate whilst producing the fewest false alarms. Instead of using the mixed integer program employed by Buckmann et al. (2021) to implement the optimization for selected hit rates, we test *all* combinations of thresholds to find the lowest false alarm rate for each possible hit rate. This is efficient because, whilst in principle the threshold for each ratio could be set at any positive or negative value, in practice only values reported in our sample need to be tested. This is because thresholds applied between these values (or above and below the highest and lowest reported values) capture no additional banks. Our implementation is more easily scalable to large samples and makes robustness testing using large numbers of random samples more practical (see section 4.3).

We find optimal combinations of hit rates and false alarm rates using all three ratios, when thresholds are set for only one ratio, and when pairs of ratios are used. We use these results to plot receiver operating characteristic curves ('ROC curves'). A ROC curve connects optimal combinations to show the best hit rate that can be achieved for any given false alarm rate. Figure 6 shows ROC curves obtained when ratios are used singly and when all three ratios are used together. The best performing ROC curve is that closest to the top left corner (a 100% hit rate with for 0% false alarm rate), which in this case is the ROC curve produced when all three ratios are used.[6]

ROC curves can also be used to compare performance for different populations of banks. Figure 7 shows ROC curves obtained having divided our sample into subsamples of 'small' banks (those with less that £5 billion of total assets) and 'large' banks (those with more than £5 billion of total assets). We use £5 billion as a convenient size threshold which ensures that there are enough small and large banks to analyse (£5 billion is equivalent to the 75th percentile of the assets distribution). Figure 7 shows that false alarm rates are substantially lower for large than for small banks. We analyse differences between large and small banks further in the next section.

4 Results

4.1 Differences between large and small banks

We find that false alarm rates are consistently lower for large banks than for small banks for given hit rates, whichever ratios are used to make predictions (see Figure 7).

We also find that, for large banks, using the leverage ratio, risk weighted capital ratio, and NSFR together produces slightly lower false alarm rates at high hit rates (i.e. over 75%) than using any ratio on its own (see Figure 7, left-hand panel). Whilst the outperformance we detect in our sample of large banks is small, this finding is consistent with Aikman et al. (2018) and Buckmann et al. (2021) who find that this same combination of metrics performs better than any of the metrics on its own. We find in contrast that, for small banks, ratios in combination do not outperform the best performing single metrics when hit rates are above 70% (see Figure 7, right-hand panel).

[5] See especially Aikman et al. (2018), pp. 21-25; and Buckmann et al. (2021), pp. 9-12.

[6] ROC curves can also be compared by measuring the area under the curve ('AUC') of each ROC curve, with a higher AUC being characteristic of superior performance. We do not use this measure of performance. That is because we assume that as bank failures are costly, good predictors must have high hit rates. We therefore compare ROC curves primarily by comparing false alarm rates at high hit rates instead of across the whole of the ROC curve.

4.2 Robustness to different definitions of size and distress

To test the robustness of our results to different definitions of distress, we repeat our analysis using five alternative definitions of distress. Our first alternative definition is a judgement we make, using publicly available information, about which banks would have defaulted in the absence of government support or mergers undertaken under stressed conditions between 1 July 2007 and 31 December 2008. Our second alternative definition captures all banks classified as distressed using the primary definition and the first alternative definition. Our final three definitions are the same as our primary definition and our first two alternatives, but applied to a longer period of 1 July 2007 to 31 December 2009.

To test the robustness of our results to different definitions of 'large' and 'small', we also repeat our analysis using £1 billion and £5 billion as size thresholds separating large from small banks. We use both size thresholds in combination with each of our six definitions of distress.

Table 7 summarizes the different definitions we use and shows the number of distress observations when each combination of distress and size definitions is applied. All combinations of size and distress definitions include both distressed and not distressed banks. However, the proportion of distressed banks in our 'large' and 'small' subsamples varies greatly. For example, only two out of 71 banks with total assets less than £1 billion are included in our list of banks which would have defaulted before the end of 2008 without government support or a merger. This proportion rises very substantially to 11 out of 60 when supervisory scores are also used to define distress (n.b. supervisory scores are not available for 11 of the smallest banks). The definitions of distress which capture very few banks during what was a severe financial crisis may be overly strict, but we include them so that results which use supervisors' subjective judgements of risk can be tested against the more objective measure of (near) default.

Our finding that false alarm rates for given hit rates are higher for small banks than for large banks is robust to these definitional choices.

Results for all 12 combinations of size thresholds and definitions of distress are shown in Figure 8. This was created by repeating our analysis using each definition of distress and using first £5 billion and then £1 billion as the size threshold dividing large and small banks. This gives us, for both large and small banks, 12 sets of results (six definitions of distress combined with the two size thresholds). From these 12 sets of results we show (in the left-hand panel) all the optimal combinations of hit rates and false alarm rates we could achieve for large banks using using any ratio (or any combination of ratios), when applying any definition of distress, and when large is defined using either size threshold. The lines show all the ROC curves we could achieve using all three ratios in combination (there is therefore one line for each of our 12 sets of results, although some lines overlap). In the right-hand panel we show the equivalent for small banks.

The results are distributed differently. The results for small banks show higher false alarm rates than those for large banks for given hit rates, whilst the ROC curves for large banks are closer to the optimal point. The best-performing ROC curves for small banks (highlighted in red) represent results when distress is defined as meaning a bank would have defaulted without government support or a merger, but when supervisors' scores are not taken into account (definitions two and five in Table 7). Whilst false alarm rates for small banks are lower when these definitions are used, they are still higher than the equivalent false alarm rates for large banks (also highlighted in red to aid comparison). From this like-for-like comparison we conclude that our results are robust even when the strictest definition of distress is applied.

Our finding that false alarm rates are higher for small banks than for large banks remains robust when we test it with more size thresholds. Figure 9 shows differences between large and small banks when size thresholds of £1 billion to £15 billion are used (varying in increments of £1 billion). Differences are calculated by subtracting the false alarm rate for large banks from that for small banks, meaning that a positive difference indicates better performance for large banks. We calculate this difference several times to test whether differences are sensitive to definitional choices and to the hit rates for which we compare false alarm rates. Figure 9 shows distributions when we calculate differences using all six definitions of distress, in combination with all three ratios, at hit rates of 75%, 80%, 85%, and 90%. It shows that whichever choices we make, false alarm rates are consistently lower for large banks across all thresholds except for a small number of outlier cases when thresholds of £9 billion or above are used.

4.3 Robustness to resampling

We use jackknife resampling to examine whether our results are driven by specific individual banks. This involves removing one bank at a time from our sample to create 118 jackknifed samples. For each jackknifed sample, we generate results using each of our six definitions of distress and using both size cutoffs. This produces 12 set of results for each jackknifed sample which we combine as before (see Section 4.2). Figure 10 shows the distribution of results from all our jackknifed samples when the leverage ratio, risk weighted capital ratio, and NSFR are used together. (It is similar to a chart showing the ROC curves for this portfolio of metrics for each of our jackknifed samples, but presented as heat maps to make it easier to see where results for multiple jackknifed samples are the same). Whilst the distributions for large and for small banks overlap slightly they are clearly different, with false alarm rates tending to be lower for large banks than for small banks.

We also test whether our results are robust to random variability in our data by creating bootstrap samples. We do this by creating, for each definition of size and distress, random samples in which there are equal numbers of large and small distressed banks, and equal numbers of large and small non-distressed banks. We create up to 1000 of these samples for each definition of size and distress. Where there are not enough observations to produce 1000 unique samples, we use all possible samples. Figure 11 shows the bootstrapped distribution when the leverage ratio, risk weighted capital ratio, and NSFR are used together. As with jackknifed samples (see Figure 10), the distributions overlap but are clearly different with false alarm rates tending to be lower for large banks than for small banks.

We compare results using single definitions of size and distress in Figure 12. This shows the full range of results for each definition, along with the interval containing 95% of bootstrapped results (defined as the range between the 5th and 95th percentile of false alarm rates for each hit rate). We find that the 95% intervals do not overlap when hit rates are over 60%, except when distress observations are taken for the longer period of 2007-2009 and a £5 billion size threshold is used. They overlap in this case because results are more variable for small banks when our distress observation period is extended to the end of 2009 but predictions are still made using balance sheet data from mid-2007. This increased variability might suggest that the predictive power of regulatory ratios decreases over time following a single observation.

5 Quality of governance as a predictor of distress

The superior performance of regulatory thresholds as predictors of distress for large banks might be because distress of large banks was typically driven by balance sheet weaknesses, whilst distress of small banks was driven by other factors not well captured by the regulatory ratios we test. One possible other factor is quality of governance, as weak oversight of management and poor strategic decision making might create risks which crystalize when the economic environment deteriorates. We cannot directly observe quality of governance. But we can observe supervisors' scores which record their judgements about quality of governance (see section 2). In order to test whether supervisors' judgements about the quality of bank's governance predict bank distress, we repeat our analysis after adding supervisors' scores for quality of governance as a fourth metric.

Governance scores are numeric (integers from 1 to 10) and we use the same technique as before to make predictions using thresholds. The only change we make is to reverse the direction of predictions; because low governance scores represent low risk, we predict distress for any banks which have a governance score *above* a governance threshold, or *below* a threshold for any of the regulatory ratios. We then compare results when thresholds for governance are set and when they are not, having first disregarded cases in which the governance threshold would not be binding on any banks as this is in practice the same as not applying a threshold.

We find that for small banks, adding thresholds for governance scores reduces false alarm rates when hit rates of at least 75% are targeted. We find the opposite for large banks: adding thresholds for governance scores actually increases false alarm rates when the same hit rates are targeted and at least one bank is captured by the governance threshold we set (see Figure 13). We therefore conclude that adding governance scores to the three regulatory ratios does improve predictions for small banks (when measured by false alarm rates), but not for large banks.

This finding is robust to definitional choices. When we use our alternative definitions of size and distress, we continue to find that adding governance scores to regulatory ratios improves predictions for small banks more than it improves predictions for large banks. Figure 14 shows the reduction in false alarm rates when governance scores are added to the three regulatory ratios. Combined results are shown for all six definitions of distress and the £1 billion and £5 billion size cut-offs. The reduction in false alarm rates is calculated by subtracting the lowest false alarm rate achievable at a given hit rate when governance scores are not used from the lowest false alarm rate when they are. A positive difference therefore indicates improved performance when governance scores are used. Figure 14 shows that, across all definitions of distress, adding governance scores to regulatory ratios reduces false alarm rates more often – and by more – for small banks than for large ones.

6 Conclusion

This paper uses regulatory data for a sample of UK-focused banks to test whether the ability of capital and liquidity requirements to predict distress varies according to banks' size. We test how key elements of the Basel III framework would have performed as predictors of distress during the Global Financial Crisis by applying to banks' balance sheets in 2007 thresholds for the leverage ratio, risk weighted capital ratio, and the NSFR. We split our sample into sub-samples of large banks (those with over £5 billion of total assets) and small banks, and find that thresholds can be optimized to generate materially more accurate predictions for large banks than for small ones. Our findings are robust to five alternative definitions of distress and to different size cut-offs separating large from small banks and to jackknifed and bootstrapped resampling. They are also consistent with a hypothesis that these regulatory ratios are better aligned with risks which tend to cause large banks to become distressed than those which tend to cause small banks to become distressed.

Our findings suggest that the benefits of applying specific combinations of requirements to banks may vary according to banks' size when benefits are considered in terms of predictive power with regard to distress. In particular, the three regulatory ratios considered here might be more suited to large banks whilst other factors not well captured by these ratios may be especially important causes of distress for small banks. Regulators may want to take into account the differences between small and large banks when deciding which prudential requirements are effective for banks of all sizes. To look into this further, we test whether supervisors' judgements about the quality of banks' governance can be used to improve predictions made using regulatory ratios. We find that they do improve predictions for small banks, although they do not provide additional predictive power for large banks over and above that provided by the regulatory ratios we test. This finding confirms the importance of supervisory oversight to identify risks which cannot be identified from balance sheet data alone.

Our findings are relevant to questions about which requirements to apply to which banks, but not the level at which they should be calibrated. That would depend on regulators' risk appetite with regard to distress. This paper is agnostic about that appetite, merely assuming that when making predictions higher hit rates are better than lower hit rates.

When interpreting our findings, it should be noted that regulatory requirements for the leverage ratio and the NSFR were introduced after the Global Financial Crisis for some UK banks. Banks might have changed their balance sheets in response and this could have weakened the predictive power of these ratios by making it harder to use thresholds to discriminate between banks which are vulnerable to distress and those which are not (this could be the case if, for example, ratios have become more similar across banks after requirements were introduced). We do not test this possibility in this paper, but it should be borne in mind when drawing conclusions about how regulatory ratios could be used to predict distress since the Global Financial Crisis. However, this does not necessarily contradict the hypothesis that these regulatory ratios are better aligned with risks which tend to cause large banks to become distressed than with those which tend to cause small banks to become distressed. New requirements might have reduced the predictive power of ratios by addressing real vulnerabilities in banks' balance sheets which are more associated with large bank distress than with small bank distress. If those requirements were removed, banks might change their balance sheets again and those vulnerabilities might reappear.

Further work would be needed to establish what causes the difference we find between large and small banks. This is a question we do not address in this paper. Possible hypotheses include the suggestion that the regulatory ratios we test are better at predicting distress for banks which employ certain business models, and that those business models are more commonly used by large banks than by small banks. If this were the case, regulators would need to consider the merits of applying different requirements to banks with different business models, either as well of or instead of the merits of applying different requirements to banks of different size. Other hypotheses may also be plausible and further analysis would be needed to test them.

References

Aikman, D., A. G. Haldane, M. Hinterschweiger, and S. Kapadia (2018). Rethinking financial stability. *Bank of England Staff Working Paper No.712*.

Alzugaiby, B., J. Gupta, A. Mullineux, and R. Ahmed (2020). Relevance of size in predicting bank failures. *International Journal of Finance & Economics*.

Basel Committee on Banking Supervision (2014). *Basel III: the net stable funding ratio*. Bank for International Settlements.

Berger, A. N., B. Imbierowicz, and C. Rauch (2016). The roles of corporate governance in bank failures during the recent financial crisis. *Journal of Money, Credit and Banking 48*(4), 729–770.

Buckmann, M., P. Gallego Marquez, M. Gimpelewicz, K. Rismanchi, and S. Kapadia (2021). The more the merrier? Evidence from the Global Financial Crisis on the value of multiple requirements in bank regulation. *Bank of England Staff Working Paper No.905*.

Coen, J., W. B. Francis, and M. Rostom (2019). The determinants of credit union failure: Insights from the United Kingdom. *International Journal of Central Banking 15*(4), 207–240.

Cole, R. A., B. G. Cronyn, and J. W. Gunther (1995). FIMS: A new monitoring system for banking institutions. *Fed. Res. Bull. 81*, 1.

Cole, R. A. and J. W. Gunther (1995). Separating the likelihood and timing of bank failure. *Journal of Banking & Finance 19*(6), 1073–1089.

Cole, R. A. and L. J. White (2012). Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research 42*(1-2), 5–29.

de Ramon, S. J., W. B. Francis, and K. Milonas (2017). An overview of the UK banking sector since the Basel Accord: insights from a new regulatory database. *Bank of England Staff Working Paper No.652*.

Francis, W. B. (2014). UK deposit-taker responses to the financial crisis: what are the lessons? *Bank of England Staff Working Paper No.501*.

Mare, D. S. (2015). Contribution of macroeconomic factors to the prediction of small bank failures. *Journal of International Financial Markets, Institutions and Money 39*, 25–39.

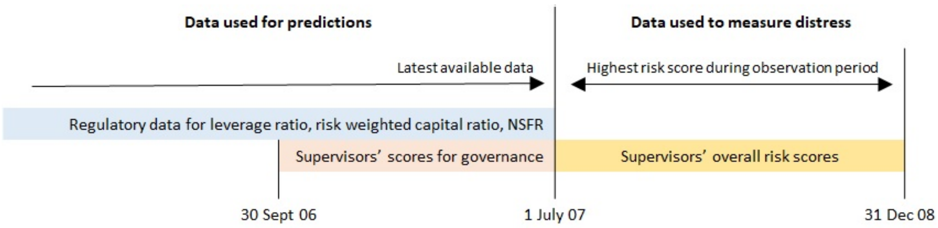
Poghosyan, T. and M. Čihak (2011). Determinants of bank distress in Europe: Evidence from a new data set. *Journal of Financial Services Research 40*(3), 163–184.

Suss, J. and H. Treitel (2019). Predicting bank distress in the UK with machine learning. *Bank of England Staff Working Paper No.831*.

Whalen, G. (1991). A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool. *Economic Review 27*(1), 21–31.

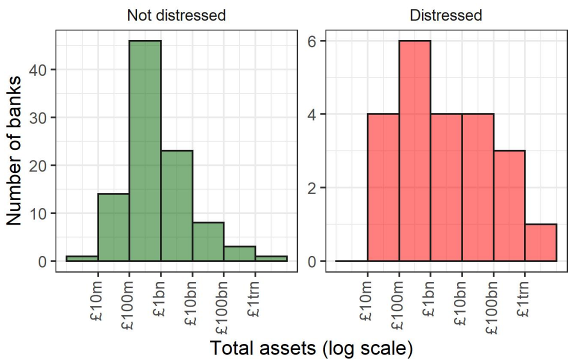
Wheelock, D. C. and P. W. Wilson (2000). Why do banks disappear? the determinants of US bank failures and acquisitions. *Review of Economics and Statistics 82*(1), 127–138.

Figure 1: Timeline of data



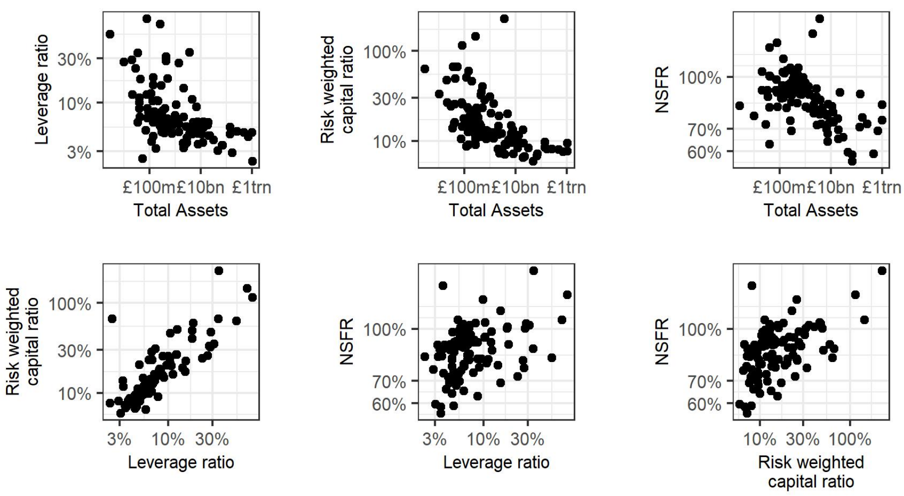
Notes: Regulatory data taken from historical regulatory returns (see de Ramon et al. (2017)). Supervisors' scores from from historic supervisory records (see Suss and Treitel (2019)).

Figure 2: Distribution of banks by total assets



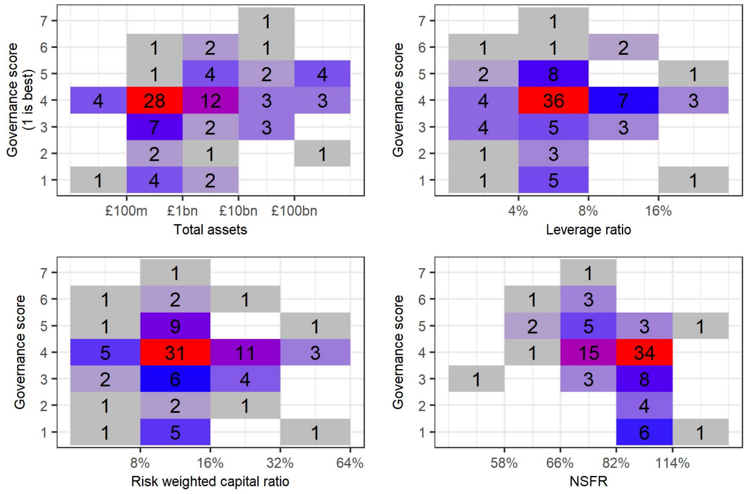
Notes: Total assets as of 1 July 2007, taken from historical regulatory returns (see de Ramon et al. (2017)). 'Distressed' banks are those which meet our definition of distress (receiving an overall risk score of 10 any time between 1 July 2007 and 31 December 2008). On risk scores, see Suss and Treitel (2019).

Figure 3: Relationships between total assets and regulatory ratios



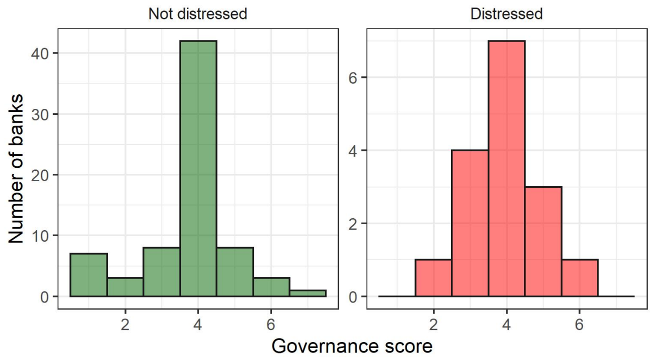
Notes: All axes on log scales. Balance sheet data are as of 1 July 2007, taken from historical regulatory returns (see de Ramon et al. (2017)). Distress from authors' calculations using historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 4: Relationships between ratios and governance scores



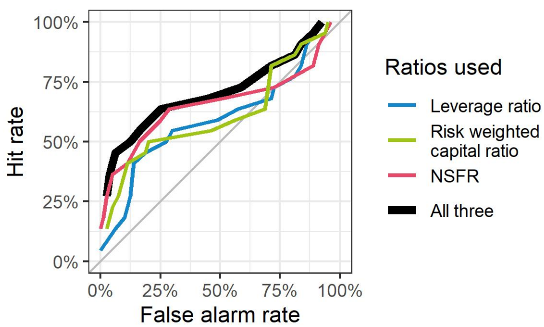
Notes: Axes for balance sheet data on log scales. All values are latest reported as of 1 July 2007. Balance sheet data are from historical regulatory returns (see de Ramon et al. (2017)). Governance scores are from from historic supervisory records (see Suss and Treitel (2019)). High outliers for total assets, leverage ratio, and risk weighted capital ratio collected into the highest buckets to preserve anonymity.

Figure 5: Distribution of governance scores for distressed and not distressed banks



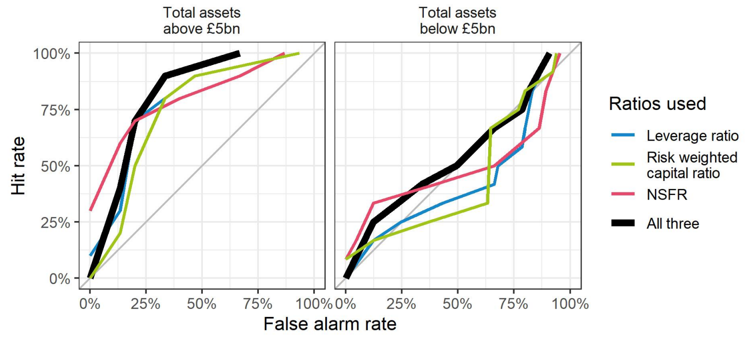
Notes: All values are latest reported as of 1 July 2007, taken from historic supervisory records (see Suss and Treitel (2019)). Distress from authors' calculations using historic supervisory records between 1 July 2007 and 31 December 2008.

Figure 6: ROC curve for all banks in our sample



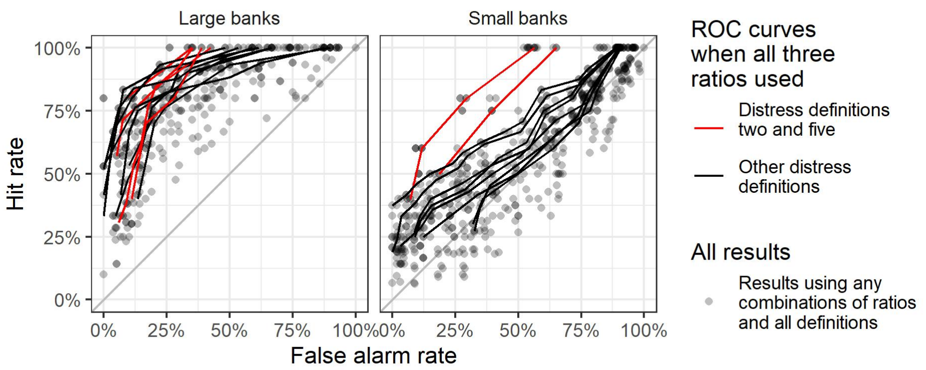
Notes: Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 7: ROC curves for large and small banks



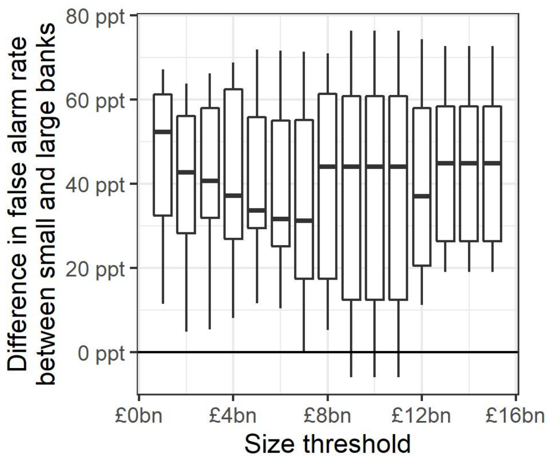
Notes: Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 8: Distribution of results using different definitions of distress and of size



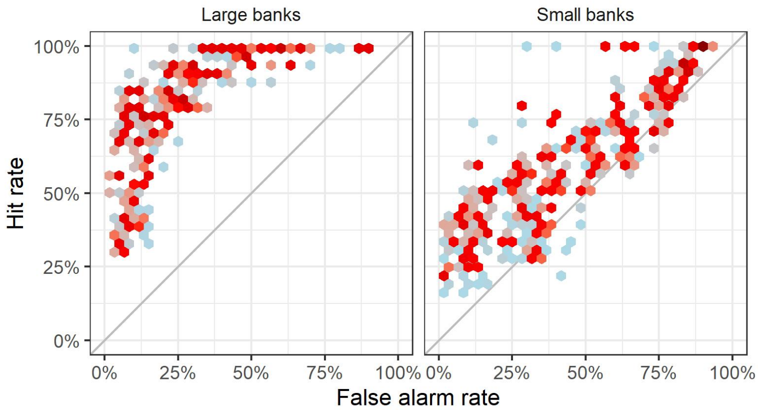
Notes: Combined results shown for all six definitions of distress listed in Table 7and for 'large' defined as both total assets greater than £1 billion and greater than £5 billion. Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 9: Differences between false alarm rates when size thresholds are varied



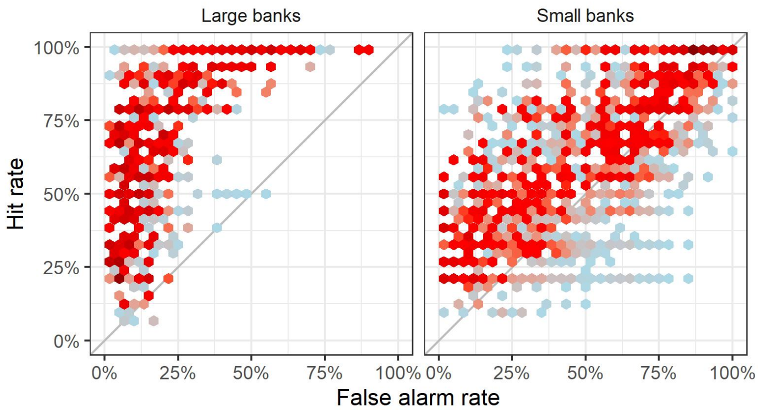
Notes: Combined results shown for all six definitions of distress. 'Large' is defined as total assets greater than the threshold specified on the horizontal axis. 'Difference in false alarm rate' is calculated by subtracting the false alarm rate for large banks from the false alarm rate for small banks. Plot shows median (central line), 25th and 75th percentiles (box), and all outliers (whiskers). Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 10: Distributions of results from jackknifed samples



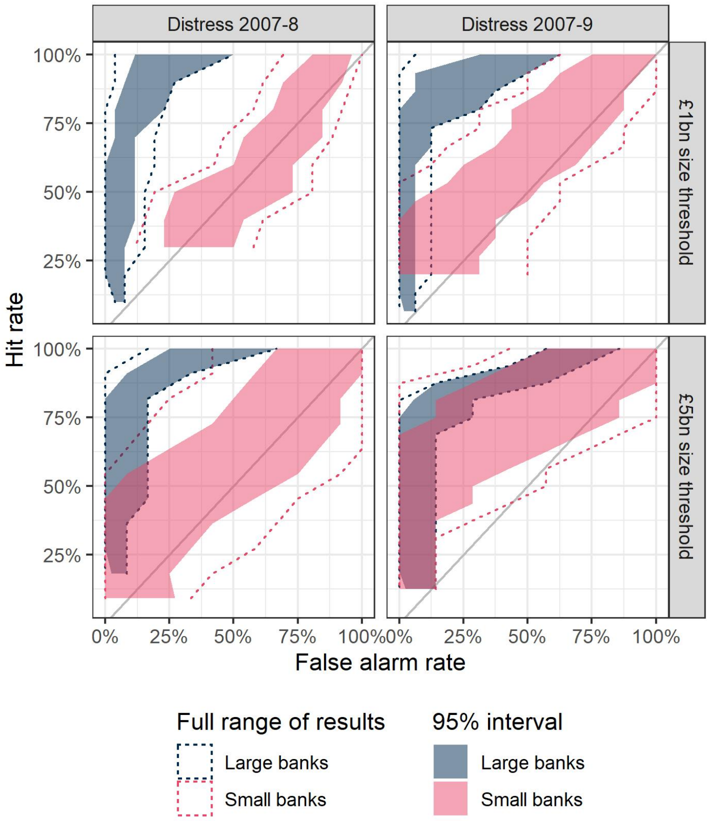
Note: Combined results are shown from all jackknifed samples, using leverage ratio, risk weighted capital ratio, and NSFR. Results shown are for all six definitions of distress listed in Table 7 and for 'large' defined as both total assets greater than £1 billion and greater than £5 billion. Colours represent the density of the distributions. Light blue represents low density, means that few jackknifed samples produce that combination of false alarm rates and hit rates. Dark red represents high density, meaning that many jackknifed samples produce that combination of false alarm rates and hit rates. Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 11: Distributions of results from bootstrap samples



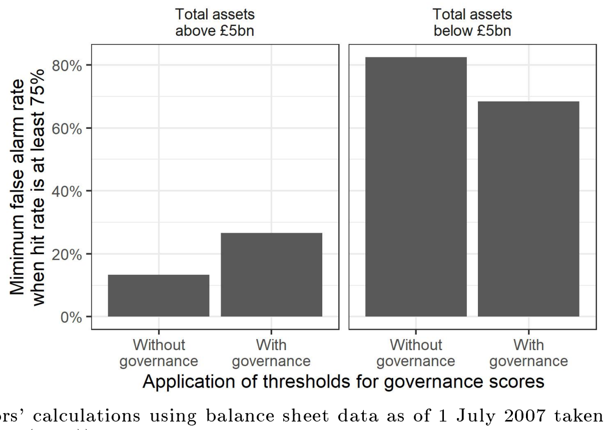
Note: Combined results are shown from all jackknifed samples, using leverage ratio, risk weighted capital ratio, and NSFR. Results shown are for all six definitions of distress listed in Table 7 and for 'large' defined as both total assets greater than £1 billion and greater than £5 billion. Colours represent the density of the distributions. Light blue represents low density, means that few bootstrapping samples produce that combination of false alarm rates and hit rates. Dark red represents high density, meaning that many bootstrapping samples produce that combination of false alarm rates and hit rates. Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 12: Ranges and 95% intervals for bootstrapped results



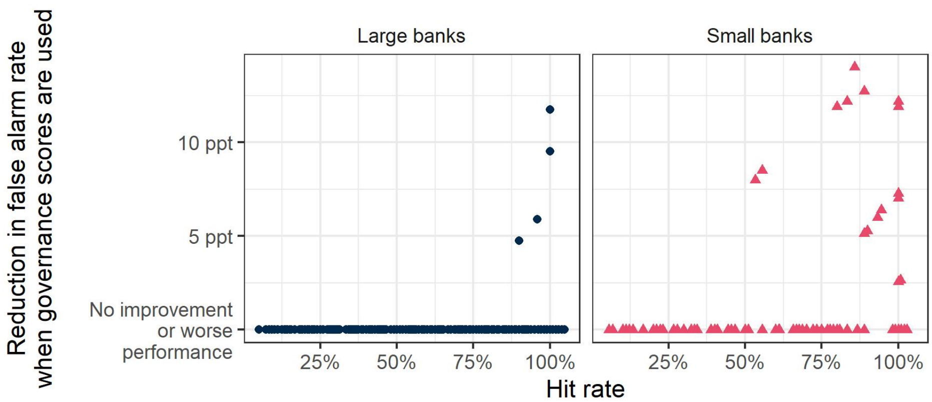
Note: Results using leverage ratio, risk weighted capital ratio, and NSFR. Distress defined as received worst supervisory rating OR would have defaulted without government support or merger in the period 2007-8 or 2007-9 (definitions of distress three and six as defined in Table 7), and for 'large' defined as both total assets greater than £1 billion and greater than £5 billion. Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2009 (see Suss and Treitel (2019)). The full range and 95% intervals can coincide in some cases because there are a discrete number of false alarm rates possible for each set of samples (these are determined by the number of non-distressed banks). When the number of possible false alarm rates is low, many samples will generate the same false alarm rate for a given hit rate. In these cases, when results for a given hit rate are ranked by false alarm rate the 0th and 5th percentile can be the same.

Figure 13: False alarm rates before and after applying thresholds for governance scores



Notes: Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Figure 14: Effect of adding governance scores for different definitions of distress and size



Notes: Combined results shown for all six definitions of distress listed in Table 7and for 'large' defined as both total assets greater than £1 billion and greater than £5 billion. 'Reduction in false alarm rate when governance scores are used' calculated by subtracting the lowest false alarm rate achievable at a given hit rate when governance scores are not used from the lowest false alarm rate when they are. Results from authors' calculations using balance sheet data as of 1 July 2007 taken from historical regulatory returns (see de Ramon et al. (2017)) and historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Table 1: Summary statistics for regulatory data and governance scores

	Minimum	Maximum	Median	Mean	Standard Deviation
Total assets (£bn)	0.003	1,044	0.52	37.6	155.8
Leverage ratio (%)	2.3	80.5	6.1	9.8	11.7
Risk weighted capital ratio (%)	6.0	228.8	13.1	20.9	27.4
NSFR (%)	56.1	474.4	89.1	90.8	38.5
Governance score	1	7	4	3.8	1.2

Notes: All values are latest reported as of 1 July 2007. Balance sheet data are from historical regulatory returns (see de Ramon et al. (2017)). Governance scores are from from historic supervisory records (see Suss and Treitel (2019)).

Table 2: Frequency of updates to supervisors' scores

(a) Number of banks updated each quarter

Size of banks	Performance	2007Q3	2007Q4	2008Q1	2008Q2	2008Q3	2008Q4
Large	Not distressed	12	7	9	10	6	6
	Distressed	9	5	8	7	4	8
Small	Not distressed	42	36	25	53	18	22
	Distressed	3	6	10	8	6	7
All banks		**66**	**54**	**52**	**78**	**34**	**43**

(b) Number of banks updated each quarter (as percentage of total)

Size of banks	Performance	2007Q3	2007Q4	2008Q1	2008Q2	2008Q3	2008Q4
Large	Not distressed	80%	47%	60%	67%	40%	40%
	Distressed	90%	50%	80%	70%	40%	80%
Small	Not distressed	65%	55%	38%	82%	28%	34%
	Distressed	25%	50%	83%	67%	50%	58%
All banks		**65%**	**53%**	**51%**	**76%**	**33%**	**42%**

Notes: Updates identified from historic supervisory records (see Suss and Treitel (2019)). A bank's scores are classed as 'updated' if any score (either a score for an individual risk factor or the overall score) differs from that recorded for the previous quarter. 'Distressed' banks are those which meet our definition of distress (receiving an overall risk score of 10 any time between 1 July 2007 and 31 December 2008). 'Large' banks are those with total assets of more than £5 billion as of 1 July 2007.

Table 3: Calendar quarter in which banks first became distressed

Calendar quarter	Number of banks becoming distressed for first time	
	Small banks	Large banks
2007Q4	1	0
2008Q1	1	2
2008Q2	3	2
2008Q3	2	2
2008Q4	5	4

Notes: Table shows the number of banks which each quarter entered distress according to our definition (i.e. received an overall risk score of 10 for the first time). 'Large' banks are those with total assets of more than £5 billion as of 1 July 2007.

Table 4: Distribution of distress observations across size quartiles

Size quartile when banks ranked by total assets	Smallest bank in quartile (total assets)	Largest bank in quartile (total assets)	Number of distressed banks
1	£3m	£159m	8
2	£174m	£528m	2
3	£553m	£4,674m	2
4	£4,723m	£1,043,817m	10

Notes: Total assets as of 1 July 2007, taken from historical regulatory returns (see de Ramon et al. (2017)). Distress from authors' calculations using historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)).

Table 5: Correlations between total assets and regulatory ratios

Variables	All banks		Not distressed banks		Distressed banks	
Total assets and leverage ratio	-0.59	***	-0.83	***	-0.41	***
Total assets and risk weighted capital ratio	-0.68	***	-0.79	***	-0.56	***
Total assets and NSFR	-0.52	***	-0.56	**	-0.54	***
Leverage ratio and risk weighted capital ratio	0.84	***	0.93	***	0.81	***
Leverage ratio and NSFR	0.34	***	0.64	**	0.19	
Risk weighted capital ratio and NSFR	0.47	***	0.72	***	0.35	**

Notes: Correlation estimates are Spearman's rank correlation coefficients. Significance shows p-values for the null hypothesis that true correlation coefficient is equal to 0: *** means $p < 0.001$, ** means $0.001 \leq p < 0.01$, * means $0.01 \leq p < 0.05$. Calculations by authors. Balance sheet data are as of 1 July 2007, taken from historical regulatory returns (see de Ramon et al. (2017)). Distress from authors' calculations using historic supervisory records between 1 July 2007 and 31 December 2008 (see Suss and Treitel (2019)). Banks for which balance sheet data, but not supervisory records, are available are included in the 'All banks' column, but in neither the 'Not distressed' nor the 'Distressed' column.

Table 6: Confusion matrix used to describe performance of thresholds

	Distressed bank	Not distressed bank
Bank breaches threshold(s)	hit	false alarm
Bank meets all thresholds	miss	correct rejection

Notes: 'Distressed' and 'Not distressed' describe whether the bank for which predictions are being made meets our definition of distress (receiving an overall risk score of 10 any time between 1 July 2007 and 31 December 2008).

Table 7: Distress observations for different definitions of distress and size

Size threshold	Distress definition	Criteria	Time period	Large banks		Small banks	
				not distressed	distressed	not distressed	distressed
£1bn	1	Received worst supervisory rating	1 July 2007 to 31 December 2008	30	12	50	10
	2	Would have defaulted without government support or merger		40	7	69	2
	3	Received worst supervisory rating OR would have defaulted without government support or merger		27	15	49	11
	4	As 1	1 July 2007 to 31 December 2009	21	21	46	15
	5	As 2		34	13	69	2
	6	As 3		17	25	45	16
£5bn	1	As above	As above	15	10	65	12
	2			23	5	86	4
	3			13	12	63	14
	4			10	15	57	21
	5			18	10	85	5
	6			8	17	54	24