The Economics of Bank Supervision

Thomas M. Eisenbach, David O. Lucca, and Robert M. Townsend*

August 2019

Abstract

We estimate a structural model of bank supervision on work hours of Federal Reserve staff supervising the consolidated universe of U.S. banks, with instrumental variables to account for the endogeneity of supervision to bank risk. Model estimates indicate that increased supervision lowers bank risk, but resources are only imperfectly reallocated to riskier banks. Needed supervisory resources grow less than proportionally with bank size and supervisors weight larger banks more than proportionally. We characterize times series variation in the aggregate shadow cost of resources, and show that it is not equalized across Federal Reserve districts.

Keywords: bank supervision, bank regulation, monitoring, time use
JEL Classification: D82, G21, G28

*Eisenbach and Lucca: Federal Reserve Bank of New York, thomas.eisenbach@ny.frb.org and david.lucca@ny.frb.org; Townsend: Massachusetts Institute of Technology, rtownsend@mit.edu. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. For valuable comments, we thank Doug Diamond, Richard Crump, Mark Flannery, Xavier Freixas, Paul Goldsmith-Pinkham, Daniel Lewis, and Gyongyi Loranth, as well as audience members at the New York Fed Conference on Supervising Large & Complex Financial Institutions, the Wharton Conference on Liquidity and Financial Crises, the BAFFI CAREFIN Annual International Banking Conference, the Jackson Hole Finance Conference, the Banco de España–CEMFI Conference on Financial Stability, the Western Finance Association Meeting, and the Chicago-Minnesota Theory Conference, as well as at Hong Kong University, Hong Kong University of Science and Technology, Singapore Management University, Nanyang Technological University, National University of Singapore, University of Mannheim, Bundesbank, Frankfurt School of Finance, and the European Central Bank. We thank Emily Eisner, Eilidh Geddes, Helene Hall, and Alice Liang for excellent research assistance. Any errors are our own.

1 Introduction

Resource constraints facing regulatory and supervisory agencies are often ignored when studying banking sector policies. As a result, periods of significant bank distress, such as the 2007–09 financial crisis, are often attributed to lax regulation or distorted incentives for supervisors (Admati and Hellwig, 2013; Barth, Caprio, and Levine, 2012; Carpenter and Moss, 2013). This paper shows that the level and allocation of supervisory resources are important determinants of banking sector outcomes. The scale and allocation of supervisory resources underwent large shifts in the past three decades, often following financial turmoil. The FDIC notes that "[it] faced severe challenges, such as the volatility of workload [and] fluctuating staffing levels," in managing the banking crisis of the 1980s and early 1990s (FDIC, 1997).[1] After the 2007–09 financial crisis, supervisory staff at the Federal Reserve increased roughly 50 percent, climbing to about 4,500 employees in 2015 (Figure 1, left panel). As resources increased, their allocation shifted — the share devoted to smaller banks decline by nearly half (right panel).

We study supervisory resources using a structural model estimated on a unique dataset of work hours spent by Federal Reserve staff supervising the universe of U.S. bank holding companies (BHCs) from 1998 to 2014.[2] In the model, supervisors minimize a weighted sum of banks' distress probabilities subject to a resource constraint. The first-order condition equalizes the marginal benefit of devoting additional resources to a given bank with the marginal cost, resulting in an optimal allocation that depends on several factors: the bank's size and level of risk, supervisors' preference weight for the bank, economies of scale in lowering risk, and the shadow cost of supervisory resources. With scarce overall resources, increased supervisory attention to one bank requires less attention be paid to other banks, as suggested by the post-2008 reallocation of resources from small to large banks (Figure 1). We use the model to estimate key structural parameters that measure the impact of supervision, technological constraints, and supervisory preferences. Parameter estimates suggest that the shadow cost of supervisory resources is not equalized across Federal Reserve districts and that the 50 percent increase in supervisory resources post-2008 only resulted in a 10 percent decline in the average shadow cost of resources because of changes in the size and composition of banks under Fed supervision. Parameter estimates also indicate that the post-2008 resource reallocation lowered risk at large banks but increased it at smaller banks and, on net, across the bank universe. Within Federal

[1] FDIC staff size started at 3,600 in the 1980s, reaching 14,000 by 1990, before dropping back to 11,000 in 2000.

[2] Our analysis is at the level of BHCs; we refer to them interchangeably as "banks."

Reserve districts, the frictionless model suggests that resources have not been sufficiently reallocated to riskier banks.

A key challenge in estimating the effect of supervisory work hours on bank distress is that hours are endogenous to risk. We use instrumental variables consistent with the model specification and show that, under a linear approximation of the first-order condition, the model can be expressed as an instrumental variable probit. We consider four different instruments and show robustness of the results across instruments.

The first instrument is the shadow cost of supervisory resources, which negatively affects the supervisory attention paid to a bank, conditional on its scale and level of risk. Under delegated authority from the Board of Governors, each of the twelve Reserve Banks in the Federal Reserve System supervises bank holding companies based in its own district with a dedicated supervisory staff. Consistent with resource constraints binding at the district level, we find that, after controlling for size and risk, a bank receives fewer hours of supervision when other banks in the same district become distressed and when there are fewer total district resources relative to total district bank assets.

The other three instruments are separate preference shocks, or "shifters," that affect the amount of supervisory attention directed toward a given bank, conditional on its level of risk. The first shifter draws from Hirtle, Kovner, and Plosser (2018), who find that the largest banks in each district receive more supervisory attention than similar banks in other districts, indicating that regional supervisors are most concerned with the performance of the largest banks in their district. The second shifter is based on a difference-in-differences approach that compares supervisory hours for banks with assets above and below $10 billion, which is a key threshold in the Dodd-Frank Act, before and after 2008. Supervisory hours for banks above the $10 billion threshold increase 93 percent in the post-2008 period while hours for smaller banks decrease 24 percent, consistent with an increase of the preference weight that supervisors assign to the largest banks. Finally, we instrument hours with the number of supervisory examinations in the year prior for non-complex banks that have a satisfactory rating and less than $10 billion in assets. For these banks, examinations are required only every other year, so supervisory hours predictably cluster in some years as opposed to being smooth over time. As a result, there is a negative correlation between the number of examinations in the year prior and the number of hours allocated to the same bank in the current year.

As outcome variables, we consider three separate measures of bank distress increasing in severity: low return on assets (below the 10th percentile, which is about zero), "severe stress" (failure or a failing supervisory rating), and outright failure.[3] Based on our

[3]From a supervisory perspective, failures of supervised banks are the most meaningful events, but these

estimates, the probability of severe stress in the following year declines about 3 percentage points in response to a 100 percent increase in current supervisory hours, which is more than half of the baseline probability of 6 percent. This average marginal effect masks considerable heterogeneity; for banks with the worst possible risk ratings of 4 or 5, the marginal effect is 8 percentage points. Effects for the other outcome variables are similarly large.[4] We also find that the reduction in distress probability is associated with reduced likelihood of low non-interest income, high loan loss provisioning, and low realized gains on securities, consistent with bank de-risking in response to increased supervision.

The structural model enables us to decompose the empirical unit elasticity of supervisory hours to bank size into competing effects: supervisory technology and preferences. We uncover significant economies of scale in supervision, with an estimated elasticity of supervisory hours costs with respect to bank size of about 0.6 — that is, supervising a bank with twice as many assets requires only 60 percent more resources. Such strong scale economies suggest that banks have not grown to be "too large to be supervise." Regarding supervisors' preferences, we find an estimated weight on bank size that is greater than one, meaning that supervisors weight large banks at an increasing rate, consistent with systemic risk concerns. We further estimate supervisors' latent preference weights for banks with different risk levels. Although hours allocated to risky banks (rated 4 or 5) are more than three times greater than hours allocated to safe banks (rated 1), we find that preference weights discount riskier banks. Thus, fewer resources are allocated to risky banks than predicted by the cost-benefit tradeoff in the model, which may be evidence of frictions in reallocating resources across banks, of biases in supervisors' beliefs, or of binding regulation that mandates a minimum allocation of supervisory resources even at the safest banks.

Concurrent with the post-2008 increase in supervisory resources and their reallocation to large banks, average risk and total assets under Fed supervision increased (Federal Reserve, 2018). Absent the post-2008 increase in supervisory resources, model estimates indicate that the average distress probability would have increased about 15 percent relative to baseline levels. Although the post-2008 reallocation to large banks had the expected effect of decreasing the distress probability at large banks and increasing it at small banks, the increase at small banks outweighs the decrease at large banks. The model thus suggests

events are also rare (only 0.5 percent in our data) and we therefore consider the alternative outcomes as well. Severe stress carries negative consequences such as inclusion on the FDIC's list of "problem banks," and occurs with roughly 5 percent probability. Low ROA occurs, by definition, with 10 percent probability.

[4]For probability of failure, the average marginal effect of a 100 percent increase in hours is -0.4 percentage points, almost offsetting the baseline probability; for probability of low ROA, the marginal effect is -2.1 percentage points.

that the reallocation to large banks increased, on net, risk in the bank universe, a result that is consistent with supervisors trading off micro- and macro-prudential objectives.

To assess resource efficiency, we study counterfactual policy experiments in which we alter the allocation of supervisory resources, respecting the budget constraint, and then trace out implications on distress probabilities. With respect to the response of supervisory hours to risk, we estimate that, in terms of distress probabilities, the actual allocation is approximately equidistant from two stylized extremes in which resources are either not-at-all or perfectly responsive to current bank risk. Lastly, we show that the current decentralization of the resource allocation causes efficiency losses. Equalizing the shadow costs of supervisory resources across districts, through a completely centralized allocation scheme, for example, lowers average distress probabilities, with much larger effects for the riskiest banks. We conclude with a simple back-of-the-envelope calculation to gauge whether the current level of overall supervisory resources is appropriate.

Related literature. The 2007–09 financial crisis has spurred renewed attention to banking regulation and supervision and their role in the buildup of risk in the financial sector leading up to the crisis. Duffie (2019) argues that the financial system was "prone to fail" because of a combination of weak regulation and supervision. Some argue that the banking sector was too levered and relied on unstable short-term funding (Greenwood, Stein, Hanson, and Sunderam, 2017; Aikman, Bridges, Kashyap, and Siegert, 2019). Others stress that regulators placed too much faith in market discipline, which was distorted by expectations that some institutions were "too big to fail" (Admati and Hellwig, 2013), or underestimated the probability of a severe shock (Gennaioli and Shleifer, 2018). The role of supervisory resource availability and allocation has not been previously discussed, with the exception of Duffie (2019), who notes significant differences in staffing at the Securities and Exchange Commission (SEC) and the Federal Reserve. Lucca, Seru, and Trebbi (2014) show pro-cyclical net flows of staff from banking authorities to the private sector, consistent with scarce supervisory resources at the end of economic expansions.

The model of resource allocation underlying our analysis is in the neoclassical tradition of Becker (1965) and Radner and Rothschild (1975), whereby the allocation of time or effort maximizes an objective function subject to a resource constraint. In estimating the effect of supervision on bank risk, we do not explicitly specify the channel through which supervision operates or why supervision by a banking authority is necessary. Prior contributions show, for example, that limited liability raises moral hazard issues leading to excessive risk taking (Jensen and Meckling, 1976), and that there are limits to the ability of markets to provide the necessary discipline for banks (Flannery, 1998; Rochet, 2004). Channels through which supervision can counteract these issues include auditing bank asset values to detect breaches of capital requirements (Rochet, 2007); preventing banks from taking observable but non-verifiable actions (Dewatripont and Tirole, 1994); incentivizing banks through punitive interference after verifiable outcomes (Marshall and Prescott, 2001; Harris and Raviv, 2014); and taking corrective action to affect banks' risk-return tradeoff before outcomes realize (Carletti, Dell'Ariccia, and Marquez, 2017). The analysis in this paper focuses on supervision as opposed to other pillars of banking policy, such as bank capital regulation, which have been the focus of an extensive literature (e.g. Repullo and Suarez, 2012).

Using instrumental variables, we find that there is a significant effect of supervision on future bank risk, similar to recent contributions by Granja and Leuz (2017), Bisetti (2018) and Hirtle, Kovner, and Plosser (2018). Earlier literature shows that supervision produces valuable information (Hirtle and Lopez, 1999; Peek, Rosengren, and Tootell, 1999). Our results are consistent with these prior findings but, by using observable measures of supervisory effort, we conduct a comprehensive structural analysis of the allocation of supervision and its impact and evaluate policy with counterfactuals.

To the best of our knowledge, our paper is the first to estimate supervisory preference weights and how they depend on bank characteristics, such as size and risk. A new element of post-crisis analysis has been macro-prudential regulation that considers externalities (e.g. Borio, 2011), consistent with our finding that large banks receive disproportionate preference weight. Starting with the rent-seeking theory of regulation of Stigler (1971), a large literature also considers incentive issues for the supervisors themselves. For example, Kroszner and Strahan (1999) investigate the role of rent-seeking in bank branching restrictions. Agarwal et al. (2014) find differences in supervision across federal and state supervisors. Similarly, Granja and Leuz (2017) find differences between the Office of Thrift Supervision (OTS) and the agencies that replaced it, the Office of the Comptroller of the Currency (OCC), and the Federal Deposit Insurance Corporation (FDIC). Finally, Kisin and Manela (2014) study the effect of fee structures on supervisory incentives.

Finally, our analysis of the allocation of supervisory hours is similar to the literature on time-use of private households (Aguiar and Hurst, 2007; Blundell, Pistaferri, and Saporta-Eksten, 2018), which also takes a neoclassical approach in the spirit of Becker (1965). Our use of a structural model and counterfactuals to analyze changes in supervision are similar to the analysis of deposit fragility and the effects of changes in capital regulation of Egan, Hortacsu, and Matvos (2017) or the analysis of stress tests of Corbae et al. (2018).

The rest of the paper is organized as follows. Section 2 provides brief institutional detail about the practice of bank supervision and then presents our economic model of supervision as well as our econometric specification. Section 3 discusses the data and basic determinants of supervisory hours, while Section 4 presents the instrumental variables used for supervisory hours. Section 5 provides the main results of our analysis: estimates for the effect of supervision and for supervisory preference and technology parameters. Section 6 uses the estimates to conduct counterfactual policy experiments, and Section 7 presents our conclusion.

2 Allocation of supervisory resources: economic model

This section presents the model of supervisory resource allocation and characterizes the optimal allocation. To estimate the model parameters, we linearize the first-order condition and use it as the first stage in an instrumental variable probit specification. The econometric specification explicitly accounts for the fact that supervisors have more information about the likelihood of future bank distress than the econometrician. We then estimate the model and obtain underlying structural parameters that measure the impact of supervision on bank distress, the economies of scale in the supervisory production function and supervisory preference weights for banks with different characteristics. Before turning to the model we provide a brief institutional overview of bank supervision at the Federal Reserve (for a more extensive overview, see Eisenbach et al., 2015).

2.1 Overview of supervisory practice

In the United States, the Federal Reserve Board of Governors is entrusted with the authority and responsibility for supervising BHCs on a consolidated basis. In practice, each of the twelve Reserve Banks in the Federal Reserve System supervise the BHCs that are located within its own district under delegated authority from the Board of Governors. While supervisory activities are coordinated at the Federal Reserve System level via committees, especially so in recent years and for the largest and most complex banks,[5] each Reserve Bank employs dedicated supervisory staff ("examiners") and determines its own hiring, performance assessments, and staff allocations. Our model studies the allocation of staff resources within a Federal Reserve district, taking the total district-level resources as given. In Section 4, we form instruments based on variation in district-level resources and the relative size and risk of banks within a district.

Hours spent by staff supervising banks can be classified into two main activities: (i)

[5]Specifically, starting in 2012, supervision of the largest and most complex banks has been coordinated across Reserve Banks and the Board of Governors through the Large Institution Supervision Coordinating Committee (LISCC) program. More generally, the Board of Governors provides oversight of operations and budgeting processes at each of the twelve Reserve Banks.

monitoring the risks a bank is exposed to and how the bank manages those risks, and (ii) intervening through corrective actions to change the bank's risk exposures or risk management. In their monitoring activities, supervisors meet with senior and business line management, with risk and control departments, and with members of a bank's board of directors. Supervisors conduct independent reviews of a bank's internal reports, such as those related to the bank's risk position, performance, budget, and strategy. The largest banks are typically monitored by a dedicated team that is assigned to the same bank on an ongoing basis. Assessments of these banks, which typically hold consolidated assets of more than $10 billion, take place under "continuous monitoring," and supervisors take stock of these continuous activities once a year in so-called "roll-ups." Smaller banks are monitored by a staff that rotates over a portfolio of banks, and periodic comprehensive assessments (typically yearly) are based on full-scope examinations. We exploit variation in mandated supervisory practices for econometric identification.

Full-scope examinations and annual roll-ups culminate in the assignment of a confidential supervisory rating. BHCs are assigned a 1-to-5 rating under the "RFI/C(D)" rating system, with lower numbers indicating fewer issues and thus a better rating. Banks with a rating of 1 or 2 are considered in satisfactory condition, presenting few significant supervisory concerns. Banks with a 3, 4, or 5 rating present moderate to extreme levels of supervisory concern.[6]

Finally, supervisory interventions are intended to modify a bank's behavior and typically reduce its risk positions. Interventions take place through corrective supervisory actions, which require banks to address unsafe or unsound practices and violations of regulations. These actions impose timelines and, until a bank remediates the issues, can impose restrictions on the bank's asset growth and set of activities as well as mandated divestitures of certain assets.

2.2 Model

We consider a model of bank supervision to study how the allocation of supervisory resources depends on bank characteristics, supervisory preferences, and the availability of overall resources. The setup is essentially static and is meant to capture the interaction of

[6] The letters in the rating system indicate different components considered in the rating assignment—"R" is for risk management, "F" is for financial condition, "I" is for potential impact of the non-depository entities in the holding company on the depository institution(s) in the holding company, "C" is for the composite rating (that is, the overall rating considering and weighting the ratings on "R", "F", and "I"), and "D" is the rating assigned to the depositories (for example commercial banks or thrifts) owned by the holding company. Prior to 2014, BHCs received ratings known as BOPECs, an acronym that stood for five areas of supervisory concern. Despite some differences, BOPECs and RFI/C(D) rating levels have similar supervisory interpretations and we splice these measures together in our analysis.

banks and supervisors over the course of a year, which is the frequency of our data. We first discuss the determinants of bank distress and how it is impacted by supervisory hours. Then we specify the supervisory objective function and derive the optimal allocation of supervisory hours.

Bank distress and impact of supervision. Let $y_{idt+1} \in \{0,1\}$ be an indicator variable for bank i in district d becoming distressed in year $t+1$. We use different measures of distress in the data: negative return on assets, supervisory rating of 4 or 5, or failure. Distress y_{idt} is determined by a continuous latent variable y^*_{idt}, such that $y_{idt+1} = \mathbb{I}\left[y^*_{idt} > 0\right]$ where $\mathbb{I}[\cdot]$ is the indicator function and $y^*_{idt} = D_{idt} + u_{idt}$. Here, $u_{idt} \sim \mathcal{N}\left(0, \sigma_u^2\right)$ is a shock realized at the end of year t and D_{idt} is a distress threshold determined by the bank's and the supervisor's actions in year t. The distress threshold is linear, $D_{idt} = \varrho_{idt} - \gamma s_{idt}$, where ϱ_{idt} denotes the riskiness of bank i due to its own actions in year t and s_{idt} denotes the intensity of supervision at bank i in year t which has impact γ. The resulting probability of distress is given by

$$\mathrm{PD}(\varrho_{idt}, s_{idt}) = \Phi\left(\frac{\varrho_{idt} - \gamma s_{idt}}{\sigma_u}\right), \tag{1}$$

where Φ denotes the c.d.f. of the standard normal distribution. The parameter $\gamma \geq 0$ captures the effectiveness of the supervisory intensity s_{idt} in reducing the probability of distress.

Bank riskiness ϱ_{idt} is partly reflected in the supervisory rating $r_{idt} \in \{1, \ldots, 5\}$. If supervisors had no additional information on bank risk than is summarized in the rating, then γ could be identified with a standard maximum likelihood estimation of the probit model in (1), given data on confidential supervisory ratings and examiner hours. But, in practice, supervisors' information set is larger than the econometrician's because ratings are granular and not updated continuously. We formally account for this information asymmetry by positing that only the supervisor observes an additional variable η_{idt}, which is informative about future distress. From the econometrician's perspective, the latent distress variable is

$$y^*_{idt} = \varrho(r_{idt}) - \gamma s_{idt} + u_{idt}, \tag{2}$$

where $\varrho(r_{idt})$ is the component of bank risk ϱ_{idt} reflected in the rating r_{idt} and the error u_{idt} is

$$u_{idt} = \eta_{idt} + \varepsilon_{idt}, \tag{3}$$

with $\eta_{idt} \sim \mathcal{N}(0, \sigma_\eta^2)$, $\varepsilon_{idt} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and mutually independent. While ε_{idt} is observed neither by the supervisors nor the econometrician, η_{idt} is observed by the supervisors.

The latent variable y^*_{idt} and its parts do not directly depend on bank size. In the data, however, bank size varies across several orders of magnitude and strongly affects the allocation of supervisory resources. Banks with larger balance sheets have larger loan portfolios and engage in more activities, both of which require more supervisory resources. We therefore translate the scale-free supervisory intensity s_{idt} into scale-dependent supervisory hours with a cost function that yields the hours necessary to achieve the supervisory intensity s_{idt} at a bank of asset size A_{idt}:

$$h(s_{idt}, A_{idt}) = \exp(s_{idt})\, A_{idt}^{\alpha}. \tag{4}$$

Hours cost h is increasing and convex in the supervisory intensity s_{idt} as well as increasing in bank size A_{idt}, with α, the elasticity of hours costs with respect to bank size, measuring scale economies in the supervisory technology.

Supervisory objective and allocation of hours. We consider supervisors in a district d allocating resources to a set of banks $i \in \mathcal{I}_{dt}$ with an available budget of supervisory hours $\overline{H}_{dt}$. The supervisors' objective is to minimize a weighted sum of the banks' distress probabilities:

$$\min_{\{s_{idt}\}} \sum_{i \in \mathcal{I}_{dt}} \mathrm{PD}(\varrho_{idt}, s_{idt})\, W_{idt} \quad \text{subject to} \quad \sum_{i \in \mathcal{I}_{dt}} h(s_{idt}, A_{idt}) \leq \overline{H}_{dt} \tag{5}$$

The preference weight W_{idt} given to bank i tilts the allocation away from simply minimizing the average probability of distress. It can account for both micro-prudential objectives (the cost of distress only to the individual bank) and macro-prudential objectives (e.g. the spillover costs of distress on the wider economy) or statutory requirements. The weight W_{idt} can also be interpreted as the loss given default; then the objective in (5) is analogous to the credit risk framework used to calibrate regulatory capital requirements (Basel Committee on Banking Supervision, 2010). For example, in setting capital surcharges for global systemically important banks (GSIBs), the Federal Reserve considers a bank's "systemic loss given default" which explicitly includes externalities to the overall stability of the financial system (Board of Governors of the Federal Reserve System, 2015).

The supervisors' first-order condition in terms of the intensity s_{idt} is

$$-\frac{\partial}{\partial s}\mathrm{PD}(\varrho_{idt}, s_{idt}) \times W_{idt} = \frac{\partial}{\partial s} h(s_{idt}, A_{idt}) \times \Lambda_{dt}. \tag{6}$$

The left-hand side is the benefit of additional supervision at bank i: the reduction in distress probability multiplied by the preference weight. The right-hand side is the cost of additional supervision at bank i, which combines (i) the marginal hours cost given the bank's size and (ii) the shadow cost of hours, given by the Lagrange multiplier Λ_{dt} on the hours budget constraint, since additional hours at bank i mean reduced hours at other banks in district d. The first-order condition (6) makes clear predictions about the comparative statics of the allocation of supervisory hours with respect to bank size, riskiness, preference weight, and overall resource scarcity:

Proposition 1. *Supervisory hours at bank i are increasing in bank i's size A_{idt}, risk ϱ_{idt}, and preference weight W_{idt}, and decreasing in the shadow cost of hours Λ_{dt}.*

Proof. See Appendix A.1. □

In the main empirical specification we measure bank riskiness with supervisory rating dummies, $\varrho(r_{idt}) = \rho_1 + \sum_{r=2}^{5} \rho_r \mathbb{I}[r_{idt} = r]$, as of the end of year t; hours H_{idt} are total supervisory hours at bank i in year t and assets A_{idt} are total assets at the end of year t. After combining (1)–(4), the probability of distress conditional on the supervisor's information set, including η_{idt}, is

$$\begin{aligned}
\mathrm{PD}_{idt} &= \Pr[y^*_{idt} > 0 \mid r_{idt}, H_{idt}, A_{idt}, \eta_{idt}] \\
&= \Phi\left(\frac{\rho_1 + \sum_{r=2}^{5} \rho_r \mathbb{I}[r_{idt} = r] - \gamma \log H_{idt} + \alpha\gamma \log A_{idt} + \eta_{idt}}{\sigma_\varepsilon}\right),
\end{aligned} \tag{7}$$

and the first-order condition (6) is

$$\phi\left(\frac{\rho_1 + \sum_{r=2}^{5} \rho_r \mathbb{I}[r_{idt} = r] - \gamma \log H_{idt} + \alpha\gamma \log A_{idt} + \eta_{idt}}{\sigma_\varepsilon}\right) \frac{\gamma}{\sigma_\varepsilon} W_{idt} = H_{idt} \Lambda_{dt}. \tag{8}$$

We parameterize the preference weight W_{idt} as a log linear function of bank characteristics,

$$\log W_{idt} = \widetilde{\rho}_1 + \sum_{r=2}^{5} \widetilde{\rho}_r \mathbb{I}[r_{idt} = r] + \widetilde{\alpha} \log A_{idt} + w_{idt}, \tag{9}$$

where tildes denote preference parameters as opposed to the corresponding physical parameters in the probability of distress (7) and w_{idt} are preference shocks, or shifters, that affect the allocation of hours but are exogenous to the probability of distress.

With our functional forms, the comparative statics of optimal hours with respect to bank characteristics takes a particularly simple form.

Proposition 2. *For any conditioning variable x_{idt} with loading κ in the distress threshold y^*_{idt} and loading $\widetilde{\kappa}$ in the supervisory preference weight $\log W_{idt}$, the local effect of x_{idt} on optimal hours*

$\log H_{idt}$ *is a convex combination of the loadings* $\kappa, \widetilde{\kappa}$ *and the impact of supervision* γ,

$$\frac{d \log H_{idt}}{dx_{idt}} = \pi_{idt} \frac{\kappa}{\gamma} + (1 - \pi_{idt})\, \widetilde{\kappa},$$

where the local weight π_{idt} *is given by*

$$\pi_{idt} = \frac{\Phi^{-1}(\mathrm{PD}_{idt})}{\Phi^{-1}(\mathrm{PD}_{idt}) - \frac{\sigma_\varepsilon}{\gamma}}.$$

Proof. See Appendix A.2. □

Proposition 2 implies that the average elasticity of hours with respect to bank size is a convex combination of the size elasticity α of technological hours cost (4) and the size elasticity $\widetilde{\alpha}$ of the preference weight (9):

$$E\left[\frac{d \log H_{idt}}{d \log A_{idt}}\right] = \overline{\pi}\alpha + (1 - \overline{\pi})\, \widetilde{\alpha}$$

with $\overline{\pi} = E[\pi_{idt}]$. Based on this expression, an estimated unit elasticity of hours with respect to size could arise from a combination of economies of scale, $\alpha < 1$, and preference weights increasing more than proportional with size, $\widetilde{\alpha} > 1$. Without the structural approach, the two effects could not be separately identified.

Econometric specification. To estimate the parameters of the model, we linearize the first-order condition (8) as shown in Appendix A.3. Together with the latent variable y^*_{idt} in (2) and the evolution of the binary variable y_{idt+1}, we obtain a probit model with an endogenous regressor, which can be estimated with maximum likelihood (Wooldridge, 2010):

$$\begin{cases} y_{idt+1} = \mathbb{I}[y^*_{idt} > 0] & (10) \\ y^*_{idt} = \beta_0 + \beta_H \log H_{idt} + \beta_A \log A_{idt} + \sum_{r=2}^{5} \beta_r \mathbb{I}[r_{idt} = r] + u_{idt} & (11) \\ \log H_{idt} = \delta_0 + \delta_A \log A_{idt} + \sum_{r=2}^{5} \delta_r \mathbb{I}[r_{idt} = r] + \delta_w w_{idt} + \delta_\Lambda \Lambda_{dt} + v_{idt} & (12) \end{cases}$$

The error terms in the second stage (11) and the first stage (12) are $u_{idt} = \eta_{idt} + \varepsilon_{idt}$ and $v_{idt} = \delta_\eta \eta_{idt}$, respectively, with $\delta_\eta > 0$ a coefficient of the linearization. The shock η_{idt} is the source of the econometric bias. Since $\mathrm{cov}(\log H_{idt}, \eta_{idt}) = \delta_\eta \sigma^2_\eta$ it is also the case that $\mathrm{cov}(\log H_{idt}, u_{idt}) = \delta_\eta \sigma^2_\eta$. The first stage (12) directly informs on variables that can be used as instruments to identify the parameters: supervisory preference shocks w_{idt} and the Lagrange multiplier Λ_{dt}, which measures the overall scarcity of supervisory resources.

As in standard probit models, the parameters in the second stage (11) cannot be separately identified from the variance of the error term and we therefore normalize $\mathrm{var}(u_{idt}) = 1$. The reduced-form coefficients of the instrumental-variable probit (10)–(12) are functions of the structural parameters in the model equations (7)–(9),

$$\begin{array}{llll} \beta_H = -\gamma, & \beta_A = \alpha\gamma, & \beta_r = \rho_r & \\ & \delta_A = \overline{\pi}\alpha + (1-\overline{\pi})\,\widetilde{\alpha}, & \delta_r = \overline{\pi}\frac{\rho_r}{\gamma} + (1-\overline{\pi})\,\widetilde{\rho}_r & \text{for } r \in \{2,\ldots,5\} \end{array} \quad (13)$$

with $\overline{\pi} = E[\pi_{idt}]$ from Proposition 2; β_0 and δ_0 constants; $\delta_w > 0$ and $\delta_\Lambda < 0$ coefficients of the linearization.

Note that the second-stage coefficients on log hours, β_H, and ratings, β_r, yield parameters that describe the "physical" evolution of the probability of distress through their effect on the distress threshold D_{idt}. Specifically, γ is the loading of D_{idt} on the intensity of supervision s_{idt} while ρ_r is the loadings on the dummy for rating r. The marginal effect of supervision on the probability of distress is then equal to the product of γ and the density function, $\gamma\phi(\cdot)$. For a given γ, the second-stage coefficient on log assets, β_A, yields information on the elasticity of hours with respect to size, α. This parameter measures economies (or diseconomies) of scale of supervision and is identified under the assumption that D_{idt} is not directly affected by size. We provide support for this exclusion restriction in the results section.

In contrast, the first-stage coefficients on log assets, δ_A, and ratings, δ_r, yield linear combinations of the respective preference parameters — $\widetilde{\alpha}$ and $\widetilde{\rho}_r$ — with the corresponding physical parameters — α and ρ_r (Proposition 2). We conduct inference on the preference parameters relative to the physical parameters by comparing first- and second-stage coefficients and estimate the preference parameters using the sample analog of $\overline{\pi}$. The Lagrange multiplier Λ_{dt} enters the first stage (12) with a coefficient $\delta_\Lambda < 0$ due to the linearization. However, since δ_Λ is constant across districts and time, variation in Λ_{dt} is the same as variation in $\lambda_{dt} \equiv -\delta_\Lambda \Lambda_{dt}$; for brevity, we also refer to λ_{dt} as the "Lagrange multiplier."

3 Data and basic determinants of supervisory hours

Data. We use four types of data: (i) institution-level work hours for supervisory staff at the Federal Reserve; (ii) examination information and supervisory ratings from the National Examination Database (NED); (iii) balance sheet information, asset quality, and profitability from Y-9C regulatory filings; and (iv) bank structure information from the National Information Center (NIC). We discuss each data source and then present summary statistics for the variables included in the regressions.

Hours spent by Federal Reserve supervisory staff are from an internal database. The Federal Reserve supervises state member banks (SMBs) as well as all bank holding companies on a consolidated basis. Because we do not observe hours at non-SMB banks, which are supervised by other agencies, we focus on BHCs, which are exclusively under the purview of Fed supervisors. The hours data starts in 1998 and ends in 2014. The information is reported by supervisory employees on a weekly basis and includes information on the monitored BHC through its regulatory entity number (RSSD ID). Supervisory work at the smallest institutions is often recorded using a generic bank portfolio assignment, as opposed to an institution RSSD ID. By cross-checking hours information with independent information on the timing of supervisory inspection from NED, we find that consistent monitoring information with valid supervised-entity information is only available for institutions with assets of about $750 million or more; we therefore exclude institutions with less than $1 billion in assets. For each institution, we aggregate data by year, so that the resulting supervisory hours data is a dataset uniquely identified by a year and the supervised institution's RSSD ID.[7]

We match hours information to two other data sources. First, we obtain information on bank characteristics, including balance sheet and income statement, from public Y-9C reports, which are used to assess and monitor the financial condition of BHCs on a consolidated basis. In addition, we match supervisory hours to confidential information on supervisory ratings and exams from NED. Bank holding companies are assigned a rating from 1 to 5 under the "RFI/C(D)" rating system. As noted in Section 2.1, the acronym indicates the different components considered in constructing the rating. The highest rating is a 1, indicating the strongest performance and practices and least amount of supervisory concern, while a rating of 5 indicates the lowest performance and a very high degree of supervisory concern. We also obtain the yearly count of examinations from NED.

In terms of outcomes, we use three variables to measure the distress of a BHC by degree of severity. We count as outright failures whenever a termination of a BHC is recorded in the regulatory National Information Center (NIC) data due to a failure of the holding company, or when a subsidiary fails within one quarter of a BHC termination, for example because the holding company is acquired or merged. Because of the low incidence of actual failures during normal times, we additionally identify banks under "severe stress" that fail or have a rating of 4 or 5 at some point over the course of a year (officially considered "problem banks"). Finally, we use a realization of the return on assets (ROA) below the 10th percentile of the pooled distribution (precisely it is 10 basis points).

[7] We have information on pre-2000 hours for only a handful of districts and have information on all districts only starting in 2006.

Summary statistics. Table 1 provides summary statistics for the variables included in the regression specifications, split into small and large banks ($10 billion threshold).[8] The (unbalanced) panel is composed of about 750 unique BHCs located in the twelve Federal Reserve districts over the 1998–2014 time interval. As shown in panel (a), the average probability of failure is about 0.5 percent, severe stress occurs about 5 percent of the time, and, consistent with its definition, the ROA falls below the 10th percentile about 10 percent of the time. As shown in the first column of the panel (b), Fed supervisors allotted about 1,500 hours per year, on average, to supervising a holding company in our sample. Based on an eight-hour work day and 48 work weeks per year, a full-time supervisor works about 1,900 hours per year, so these recorded hours can be converted into about three supervisors for every four bank holding companies. However, large institutions have on average three times as many hours assigned, while smaller ones about a fifth.[9] About 20 percent of our sample is composed of BHCs with assets greater than $10 billion and 15 percent are also among the five largest BHCs within their respective districts (as defined in Hirtle et al., 2018). The average supervisory rating in our sample is 2 and about 15 percent of the sample is composed of banks with a "stressed" rating of 3 or worse.

Bank supervisors make a complexity assessment annually for each BHC (RSSD 9057) using a number of criteria, such as material credit-extending activities; significant and risky nonbank activities, such as securities broker-dealer activities or insurance underwriting; and subsidiaries that issue significant debt to the general public. About 30 percent of our sample is composed of complex BHCs, most of which fall into the large BHC category. In Section 4, we use information on rating, size, and complexity to construct a "high examination frequency" dummy variable that identifies BHCs that receive more examinations each year. On average, BHCs receive about 1.5 examinations each year; large BHCs, which all have "high examination frequency," receive about four examinations per years, while smaller BHCs receive less than one per year.

Reduced form relation between supervisory hours, bank size, and risk. We characterize pooled and within-bank variation in supervisory hours at banks in terms of basic explanatory variables, specifically bank size and risk, which appear prominently in guidance to examiners in the BHC supervisory manual.[10] We measure risk in terms of bank profitability, bank asset quality, and regulatory capital, or using confidential supervisory ratings. Table 2 presents parameter estimates from a linear regression specification of log supervi-

[8]Appendix E provides exact variable definitions and constructions.

[9]This calculation excludes hours that have not been booked by the supervisor to a specific institution. In addition, the day-count translation would underestimate an actual headcount because it doesn't account for other administrative or training activities that a supervisor may be involved in when not assigned to a bank.

[10]Available at `https://www.federalreserve.gov/publications/supervision_bhc.htm`.

sory hours on log bank assets and these measures of bank risk. The model specifications shown in the table differ in terms of bank risk controls and by the inclusion of bank fixed effects (even columns). In all regression tables, standard errors, which are reported in square brackets, are clustered at the bank level.

The coefficient on log assets (expressed in constant 2012 dollars) captures the elasticity of supervisory hours with respect to bank size. If hours increase proportionally with assets, the coefficient is equal to 1, and if less than proportionally, less than 1. We find an elasticity of supervisory hours to bank assets of slightly less than 1 without bank fixed effects (columns 1, 3, & 5) and of about 0.7 when including bank fixed effects (columns 2, 4, & 6). Only the estimates with bank fixed effects are significantly less than 1. This size-elasticity estimate corresponds to δ_A in (13), which commingles the economies of scale of supervision, α, as well as the size elasticity of the supervisor's preference weight, $\widetilde{\alpha}$. The structural estimation in the next section separates these different effects.

With respect to bank risk, we measure profitability as ROA; asset quality as the fraction of nonperforming loans (NPL ratio), i.e., the fraction of loans that are either delinquent for 90+ days or in non-accrual status; and regulatory capital as the ratio of tier 1 capital to risk-weighted assets.[11] As shown in columns (1)–(2), when supervisory ratings are omitted, risk measures from regulatory filings enter with the expected signs and are generally statistically significant at conventional levels: Higher supervisory hours are associated with lower ROA, lower tier 1 capital, and a higher fraction of nonperforming loans.

Because ratings are ordinal rather than cardinal measures, we separate dummy indicators for each rating, leaving out the best rating category of 1. Once we include ratings, the other risk measures lose most of their significance (columns 3 & 4), suggesting that the supervisory rating indicators span the information contained in regulatory filings. In the main specifications of the paper, we largely focus on measuring bank risk using only supervisory ratings (as in columns 5 & 6). The effect of bank risk, as measured by the rating, is very significant both statistically (p-vals < 0.01) and economically. For example, as compared to a bank with the best possible rating of 1, which is the omitted category in the regression, a bank of the same size but with a rating of 3 receives about two to three times more hours (columns 5 & 6), an effect comparable to more than doubling the size of the bank.[12] Similar to the coefficient on size, the estimated coefficients on bank risk commingle the loading ρ_r of future bank distress on current risk with the loading $\widetilde{\rho}_r$ of supervisor

[11] Information is as of the quarter-end of the exam in which the rating used in specifications (3)–(6) is assigned.

[12] Note that in regressions with log hours as the dependent variable, the coefficient δ_d on a dummy variable d has to be transformed as $\exp(\delta_d) - 1$ to calculate the percentage change in hours for a dummy value of 1 vs. 0: $\delta_d = \log H|_{d=1} - \log H|_{d=0}$ implies $\exp(\delta_d) - 1 = (H|_{d=1} - H|_{d=0}) / H|_{d=0}$.

preference weights on banks of different risk.

As an alternative measure of supervisory efforts, in Appendix B, we use supervisory fees assessed on federally chartered commercial banks by the OCC, which reflect supervisory costs at institutions as a function of risk and size. Overall, the sensitivities of Fed supervisory hours and OCC assessment fees to size and risk are similar.

4 Identification of the model parameters

The preference shifters w_{idt} and the Lagrange multiplier λ_{dt} are valid instruments "z_{idt}" in the structural model described by equations (10)–(12). Because they enter the first but not the second stage, they satisfy the exclusion restriction, $\text{cov}(\eta_{idt}, z_{idt}) = 0$, and the relevance condition, $\text{cov}(\log H_{idt}, z_{idt}) \neq 0$. We measure λ_{dt} as resource scarcity for a given Federal Reserve district and year and consider three separate proxies for the preference shifter w_{idt}.

4.1 Resource scarcity within districts

From Proposition 1, supervisory resources allocated to a bank not only depend on the bank's characteristics and supervisors' preference weight for the bank, but also on the shadow cost of supervisory resources, as measured by the Lagrange multiplier λ_{dt}. If the shadow cost only affects the probability of distress through supervisory hours after conditioning on bank characteristics — a point to which we return below — λ_{dt} is a valid instrument for log hours.[13]

To measure λ_{dt} variation, we first have to determine at what level of aggregation resources are set — that is, to what extent resources can be reallocated at a given point in time from one bank to another. For example, if resources were perfectly mobile across district lines and allocated in a fully centralized manner, only a single Lagrange multiplier would exist at each point in time, measuring the shadow cost for all banks across districts. As noted in Section 2.1, under delegated authority from the Board of Governors, each of the twelve Reserve Banks supervises banks located in its own district with a local supervisory staff, resulting in a decentralized allocation of supervisory resources. This approach could be because short-run budgetary processes are set independently at different Reserve Banks, or because geographically distant supervisory resources are not as effective as those that are close — for example, if local supervisors know more about local economic

[13]More generally, the shadow cost of supervisory resources would also increase under the less stringent assumption that total resources can be partially, but not fully, adjusted at a given point in time, for example due to budgetary processes or simply the time it takes to hire additional staff.

and banking sector conditions (Gopalan, Kalda, and Manela, 2017). In the data, we find that nearly all hours allocated to a bank (95 percent) are from staff at its district's Reserve Bank; we also show below that changes in λ_{dt} in other districts do not affect hours allotted to a bank in a given district. Based on this evidence, we consider a budget constraint set at the district and year level.

To instrument supervisory hours with variation in the Lagrange multiplier λ_{dt}, first note that the first stage (12) determines supervisory hours as a function of banks' characteristics, the preference shifter and the Lagrange multiplier. Because the Lagrange multiplier is common to all banks in a district and year, estimates of λ_{dt} can be constructed from district-year averages of supervisory hours, bank characteristics, and preference shifters up to the unknown coefficients. Estimates of the coefficient, in turn, also depend on the estimates of the Lagrange multipliers. We form a "plug-in" estimator of λ_{dt} by directly including district $\times$ year averages of supervisory hours and bank characteristics in the first stage and then jointly estimating all parameters. Following work on judicial outcomes identified through variation in judge leniency (Dahl et al., 2014; Dobbie et al., 2018), for each characteristic x we denoted by $\overline{x}_{-idt}$ the leave-out average in district d in year t excluding bank i. As noted by Dobbie et al. (2018), leave-out averages are equivalent to leave-out fixed effect estimators and can be interpreted as reduced-form jackknife IV estimators (Angrist et al., 1999). These leave-out averages are recommended in our setting because, without leaving out, measurement error in the dependent variable would also appears in the independent variable. The plug-in estimator is then obtained by replacing λ_{dt} in the first stage with the leave-out averages $\overline{\log H}_{-idt}$, $\overline{\log A}_{-idt}$ and of the five $\overline{\mathbb{I}[r]}_{-idt}$.[14]

Column (1) of Table 3 shows regression estimates of log hours at bank i on within-district leave-out averages of log hours, log assets, and each of the rating indicators. As noted at the bottom of the table, the regression also includes bank i's own log assets and supervisory rating indicators (coefficients omitted from the table). The estimated coefficients on the leave-out averages have the correct signs: log hours at bank i are increasing in average log hours (coef. $= 0.761$, p-val< 0.01) of other banks in the district, declining in average log assets (coef. $= -0.791$, p-val< 0.01) and declining in the rating indicators

[14] The judge leniency literature exploits quasi-random assignment of cases to judges who vary in their leniency. Variation in leniency is constructed from leave-out averages of the judges' other decisions after controlling for other covariates in a supplementary regression. Similarly here, the Lagrange multipliers are measured from supervisory hours allocations at other banks. But differently from the judge leniency literature, the variation in the Lagrange multiplier is determined from the leave-out average of the first stage itself, because the structural model fully characterizes the hours allocation. Intuitively, in our model, variation in the shadow cost of hours is measured by the gap between the average size and risk of other supervised banks relative to average hours assigned to them, which can be interpreted as a measure of "supervisory workload."

with larger effects for worse ratings. Because the elasticity of hours with respect to size is close to 1 in the data, the loadings on average log assets and average log hours have similar magnitudes. As shown at the bottom of the table, the F-statistic for the null that the coefficients on average log hours, log assets, and rating indicators are zero is 43.6.

We can characterize between- and within-district variation in the shadow cost, calculated as $-\bar{\delta}_H\overline{\log H}_{dt} + \bar{\delta}_A\overline{\log A}_{dt} + \sum_{r=2}^{5}\bar{\delta}_r\overline{\mathbb{I}[r]}_{dt}$ where $\overline{x}_{dt}$ denotes the average of variable x within district d and year t,[15] and where $\bar{\delta}_H$, $\bar{\delta}_A$ and $\bar{\delta}_r$ are the coefficients on average log hours, log assets, and rating indicators from Table 3, column (1). The overall variation of the total shadow cost is close to equally split into variation between and within districts (Appendix Table A3). To determine the source of variation, we decompose total shadow cost into two additive components, the "assets/hours component" $\bar{\delta}_A\overline{\log A}_{dt} - \bar{\delta}_H\overline{\log H}_{dt}$ (because assets and hours co-vary strongly) and the "ratings component" $\sum_{r=2}^{5}\bar{\delta}_r\overline{\mathbb{I}[r]}_{dt}$.[16] The majority of the overall variation in λ is due to variation in the assets/hours component, which varies both between and within districts. In contrast, the ratings component varies mostly within districts.

As noted above, we use leave-out averages in our baseline specification. Instead, the second column of Table 3 constructs averages when also including bank i. In this case, the coefficient on log hours is (by construction) equal to 1 and the remaining loadings are the same (with opposite signs) as those on bank i's own log assets and rating indicators when adding district $\times$ year fixed effects to Table 2, column (5).[17] In this specification, the F-statistic increases to 69, suggesting a "stronger instrument" than when we use leave-out averages. As noted above, however, because of the likely bias when using leave-in averages, we use leave-out averages as our preferred specification. In the third column of Table 3, we augment the regression with averages of log hours, log assets, and rating indicators for banks in all districts other than the one where bank i is located. We find insignificant coefficients, consistent with the assumption that the resource constraint that matters is the one at the level of a Federal Reserve district, as opposed to one set at the level of the entire Federal Reserve System. Thus far, we have only considered bank log hours and characteristics to measure resource scarcity. In column (4), we augment the specification with averages of the preference shifters discussed in Section 4.2 and find similar results.

[15] Even though we construct leave-out averages $\overline{x}_{-idt}$ to instrument for each individual bank's hours, the average across banks of leave-out averages equals the overall leave-in average used to illustrate the sources of variation, $\frac{1}{|\mathcal{I}_{dt}|}\sum_{i\in\mathcal{I}_{dt}}\overline{x}_{-idt} = \frac{1}{|\mathcal{I}_{dt}|}\sum_{i\in\mathcal{I}_{dt}}x_{idt} = \overline{x}_{dt}$.

[16] The squared overall standard deviations of the two components sum to about the squared overall standard deviation of the total shadow cost, indicating only limited covariance between the two components.

[17] Note that, $-\lambda_{dt} = \overline{\log H}_{dt} - \left(\sum_{r=2}^{5}\delta_r\overline{\mathbb{I}[r]}_{dt} + \delta_A\overline{\log A}_{dt}\right)$, or, in econometric terms, column (2) of Table 3 is a demeaned district $\times$ year fixed effect estimator.

As show in Appendix Table A5, we find very similar results when we include bank fixed effects in the regressions.

Correlated shocks within districts are a potential threat to the exclusion restriction. In this case a regional shock would not only lower supervisory hours through λ_{dt} but also imply that conditions at other banks may have worsened. Importantly, however, all specifications include banks' current ratings, which, as a supervisory summary of a bank's risk, should contain relevant information about regional shocks. In robustness checks (Section 5.4), we augment the specifications to also include additional risk controls and year fixed effects to account for aggregate time varying factors, such as the business cycle, and find similar results.

4.2 Supervisory preference shifters

We consider three supervisory preference shifters w_{idt}: an indicator for the largest five banks within each Federal Reserve district; an interaction indicator for banks with assets $\geq$ \$10 billion and the post-2008 period; and the number of supervisory examinations in the year prior for non-complex banks with assets $<$ \$10 billion and a satisfactory rating (equal to 1 or 2). Below we discuss the validity of each instrument and why these variables act as preference shifters. The instruments differ in the variation that they exploit. The "district top 5" instrument, which was proposed by Hirtle et al. (2018), relies on cross-sectional variation; the "post-2008 $\times$ large" interaction instrument uses cross-sectional and time-series variation over two groups and time intervals; finally, the lagged examination count for small non-complex well-rated banks exploits full panel variation in the data.

District top 5. Hirtle et al. (2018) posit that regional supervisors may be most concerned with the performance of the largest banks in their district, and show that the largest banks in each Federal Reserve district receive additional supervisory attention, after accounting for their size and risk. As a result, "district top 5" banks have lower risk — as measured by lower volatility of earnings, higher market returns, and better asset quality — relative to other banks, consistent with a risk-reduction effect of supervision. Column (1) of Table 4 corroborates these findings by showing that supervisory hours at "district top 5" banks are about 56 percent higher than at other banks (p-val$<$ 0.01, F-statistic of 13.6).[18] As noted at the bottom of the table, this is true after controlling for a bank's supervisory rating and, most importantly, the size of the bank as measured by log assets.

The exclusion restriction for this instrument is that "district top 5" banks differ from

[18] Recall that in regressions with log hours as dependent variables, the coefficient δ_d on a dummy variable d has to be transformed as $\exp(\delta_d) - 1$ to calculate the percentage change in hours for a dummy value of 1 vs. 0. Here, $\exp(0.442) - 1 \approx 0.56$.

other banks with the same size and rating only because the ordinal ranking within their district implies additional supervisory resources. Hirtle et al. (2018) include a large set of additional controls and also match "district top 5" banks with similar non-top 5 banks in other districts and continue to find evidence of lower risk. But, importantly, identification based on this instrument relies on cross-sectional variation, as bank size and ordinal size rankings within a district are relatively persistent. Indeed, after including bank fixed effects (Appendix Table A6, column 1) we find that the "district top 5" indicator is no longer statistically significant at conventional levels.

Large banks post 2008. The choice of supervisory hours is not purely under the discretion of supervisors at each point in time because supervisory hours are set in part to satisfy law, regulation and supervisory guidance that is set in advance. These requirements result in variation in supervisory hours that is independent of the distress shock η_{idt}. Regulation as well as supervisory guidance mandate enhanced supervision at larger banks, with intensities that differ before and after the financial crisis. The determination of what constitutes a "large" bank is typically based on asset thresholds, suggesting potential discontinuities in the allocation of supervisory hours.[19] As discussed in more detail below in the context of examination frequencies, according to the bank supervisory manual, exam requirements differ around a size threshold of \$10 billion. Furthermore, banks over \$50 billion are subject to the Federal Reserve's "Consolidated Supervision Framework for Large Financial Institutions" that implies enhanced supervision (and annual stress tests). SR letter 12-17 explicitly bases the supervisory framework for large banks on the crisis experience and explicitly incorporates macroprudential concerns, which are captured in our framework by the supervisory preference weight.

Figure 2 shows the relation between log hours and log assets in the pre-crisis period (1995–2008) and the post-crisis period (2009–2014), controlling for rating and allowing for breaks in the log-linear relationship at the size thresholds of \$10 billion and \$50 billion. The log-linear relation fits the data quite well pre-crisis but there are clear breaks in the post-crisis period. Large banks ($\geq$ \$10 billion) and, to a lesser extent, the largest banks ($\geq$ \$50 billion) seem to receive discretely more attention than small banks in the post-2008 sample. This change in the treatment of large banks is a change in the supervisory weights on large banks resulting from changes in regulations and supervisory guidance. We capture this discontinuity in the first stage (12) as a component of the preference shifter w_{idt} equal to the interaction of the size thresholds and time samples.

In Table 4, columns (2) and (3), we augment the baseline regression specification with

[19] The continuity of the cost function specification $h(s_{idt}, A_{idt})$ in (4) with respect to size implies that any discontinuities in the allocation of hours are not attributed to technology.

dummy variables for the post-2008 period, banks larger than \$10 billion and \$50 billion, as well as their interactions.[20] Controlling for log assets and rating, banks with more than \$10 billion in assets received about 40 percent more hours than other banks (p-val$<$ 0.1, column 2). After controlling for log assets and for the greater-than-\$10-billion dummy, banks with more than \$50 billion in assets do not have unusually higher supervisory hours pre-2008. Turning to the interactions with the post-2008 dummy, which are the source of the identification in our difference-in-differences setting, hours at banks with assets greater than \$10 billion increased by 87 percent (p-val$<$ 0.01) after the financial crisis but, analogous to the pre-2008 period, controlling for the interacted break at \$10 billion, the one at \$50 billion is not statistically significant. Results are robust to including bank fixed effects (Appendix Table A6). To estimate the coefficients in the second stage in the next section, we only use as instrument for hours the interaction between the post-2008 dummy and the \$10 billion size dummy, which leads to very similar point estimates with an F-statistic of 32 (Table 4, column 3).

Examination frequency. We form a last preference shifter instrument based on minimum mandated examination frequencies for BHCs, which depend on size, supervisory rating, and complexity (as assessed by the supervisor) but are set in advance and therefore are uncorrelated with the date t realization of the shock η_{idt}. According to the BHC supervisory manual, BHCs with assets below \$10 billion can be examined only every other year (if non-complex and rated 1 or 2) while BHCs with assets above \$10 billion have to be examined at least once every year.[21] As we show next, at smaller banks, the temporal pattern of examinations in years prior is strongly predictive of current supervisory hours.

We classify banks in our sample into "low examination frequency" ($<$ \$10 billion in assets, not complex, and rated 1 or 2) and "high examination frequency" (all other banks). As expected, more than 95 percent of "low examination frequency" banks have either zero or one exam per year, fairly equally split with 41 percent and 55 percent, respectively. Consistent with minimum mandated examinations every other year, these banks either have one or zero exams in a given year. In contrast, over 80 percent of "high frequency" banks have at least one exam per year.

For banks examined at the lower biannual frequency, we expect a negative correlation

[20] A better approach to a simple difference-in-differences strategy is to use a regression discontinuity, comparing only banks with assets close to the \$10 billion and \$50 billion size thresholds. Unfortunately, the distribution of banks around these thresholds is rather sparse, and thus we only use a standard difference-in-differences approach.

[21] Appendix Table A4 lists in detail the exam frequency requirements. In addition, as noted in Eisenbach et al. (2015) the largest and most complex banking institutions (typically with assets greater than \$50 billion) are monitored continuously as opposed to by means of examinations.

between current hours and the lagged number of exams. That is because when an exam took place in the year prior, an exam in the current year is less likely. We use lagged exams and the interaction with a dummy for banks on a high-frequency exam schedule as shifters w_{idt} for log hours in equation (12). Column (4) of Table 4 shows that an additional exam in the year prior lowers current hours by over 60 percent (p-val< 0.01) at banks that are neither stressed, nor large or complex. The interacted coefficients of lagged exams and the large, complex,and non-stressed dummies, have very similar magnitudes ranging from 67 percent to 87 percent (p-val < 0.01). Adding the interacted and uninteracted lagged exam terms, these estimates imply that, for these "high frequency" banks, the lagged number of examinations is not negatively correlated with current hours (first-stage F-statistic of 16).

In column (5), we pool all interaction dummies in a "high exam frequency" dummy, which excludes banks that have assets below \$10 billion, are not complex and have a satisfactory supervisory rating of 1 or 2. We again find that banks on a biannual cycle have 64 percent (p-val< 0.01) lower hours for an additional exam in the year prior but using a single dummy for high exam frequency increases the F-statistic to 37.[22]

The exclusion restriction for this instrument is that the number of examinations in the year prior does not contain information about the soft information observed by the supervisor (but not the econometrician) in the current year, η_{idt}. This exclusion restriction is likely to hold as ratings include soft information as of the examination date and the specifications using exam count condition on both current and lagged supervisory ratings.[23]

The last column of Table 4 combines all preference shifters. Consistent with the fact that these preference shifters exploit different variation in the data, the point estimates on the shifters are essentially unchanged compared to when they are included one-by-one and each remains statistically significant (p-vals< 0.01) with a joint F-statistic of 22.7.

5 Estimation results

We estimate the coefficients of the IV probit in equations (10)–(12) and obtain estimates of the underlying structural parameters measuring the effect of supervision γ, the economies of scale of supervision α, and systematic supervisory preference loadings on size $\widetilde{\alpha}$ and risk

[22] In unreported results, we also find that a strong negative serial correlation in supervisory hours for "low-frequency" banks, which is again consistent with substitution of resources over time.

[23] We control for lagged supervisory rating to account for variation in lagged number of exams N_{idt-1} driven by bank characteristics instead of examination frequency requirements. More specifically, the instrumental variable exclusion restriction is that the unobserved distress shock η_{idt} affecting current hours is not correlated with the previous year's number of exams, $E[N_{idt-1}\eta_{idt} \mid r_{idt-1}, r_{idt}, A_{idt}] = 0$, where we are conditioning on the lagged rating r_{idt-1} which contains all information gathered in the exams of year $t-1$. Therefore $\eta_{idt} \mid r_{idt-1}, r_{idt}, A_{idt} \sim \mathcal{N}(0, \sigma_\eta^2)$ and independent of N_{idt-1}.

$\widetilde{\rho}$. We then inspect the mechanism by which supervision affects bank distress based on the response to supervision of banks' regulatory ratios, asset quality and major categories of income and expense. Finally, we present robustness tests of the main specification.

Tables 5 and 6 present second-stage estimates for the three bank distress outcomes: severe stress, failure, and low ROA. Outcomes are measured in the year after supervisory hours are recorded. Each table reports coefficient estimates, standard errors clustered by bank in brackets, and estimated average marginal effects in curly braces. In the previous section, we presented estimates of the first-stage regression, which are the same, up to sample differences, to first stages estimates in the IV probits. For brevity, we do not revisit those estimates and report results in Appendix Tables A7, A8, and A9. Each second-stage table in the main text reports effective F-statistics from the first stages, and critical values, for the weak instrument test of Olea and Pflueger (2013), which is robust to heteroskedasticity, autocorrelation, and clustering. Depending on the specification, the critical values range between 10 and 23.[24]

The different columns in the tables show parameter estimates using different instrumental variables. We first instrument log supervisory hours with variation in resource scarcity by including the leave-out district $\times$ year averages of hours and bank characteristics in the first stage. We then instrument log hours with preference shifters. The "district top 5" and post-2008 $\times$ large bank indicators measure the increase in supervisory hours for the largest banks within a Federal Reserve district and for large banks after the financial crisis, respectively. Next, we use the number of examinations in the year prior for banks with assets less than \$10 billion that are not stressed and have a satisfactory rating (1 or 2).[25] Finally, we combine all instruments in column (5) of Table 5 for outcome severe stress and columns (3) and (8) of Table 6 for outcomes failure and low ROA, respectively. As additional controls, the second stage includes log assets and rating indicators. For instruments based on interaction terms, uninteracted variables are also included; for example, the post-2008 and Assets$\geq$\$10b dummies when using the post-2008 $\times$ Assets$\geq$\$10b instrument (noted at the bottom of the table, coefficients omitted).

5.1 Effect of supervision

The effect of supervision on the probability of bank distress is measured by the parameter γ, which is the loading of the distress threshold D_{idt} on the intensity of supervision s_{idt}. The

[24]We use the critical value for the commonly used 5 percent significance level for the test that approximate asymptotic bias does not exceed 10 percent.

[25]When bank failure is the outcome variable (Table 6), we only consider the lagged number of examinations as preference shocks, because only one "district top 5" and zero large BHC post-2008 fail in the sample.

marginal effect of supervision on the probability of distress is then equal to the product of γ and the density function, $\gamma\phi(\cdot)$. We first discuss estimates of γ and then of the average marginal effect; finally, we present estimates of the marginal effects evaluated at different ratings that show non-linear effects of supervision on banks with different risk.

We find statistically and economically significant effects when using the probability of severe stress as the outcome (Table 5). All p-values are smaller than 1 percent with the exception of the post-2008 $\times$ Assets$\geq$\$10b instrument (column 3). In terms of economic significance, as shown by the estimates of the average marginal effect in curly braces, an increase in supervisory hours of 100 percent lowers the future probability of distress by 2.8 percentage points on average when including all instruments (column 5). Other columns show similar magnitudes, with the exception of "district top 5" (-13.3 percentage points) and post-2008 $\times$ Assets$\geq$\$10b (-.09 percentage points). These estimated average marginal effects suggest large economic effects of supervision. For example, doubling supervisory hours implies a reduction in the probability of severe stress by about half the unconditional probability of 5.3 percent (Table 1).

Table 6 presents estimates for the year-ahead probability of failure and the year-ahead probability of low ROA. Even including the financial crisis, bank failure only occurs 0.5 percent of the time (Table 1). Here, we only consider the lagged number of examinations as a preference shifter because "district top 5" and post-2008 $\times$ Assets$\geq$\$10b perfectly predict (lack of) failure. We find an insignificant effect of log hours on failure probability when using resource scarcity as the instrument (column 1) but a strong negative impact (p-val$<$ 0.01) from the lagged examination count (column 3) or when including both instruments (column 3). A doubling of supervisory hours therefore implies a reduction in probability of failure of about three-quarters of the unconditional probability. We find a significant effect of supervisory hours reducing the probability of a low ROA for the resource scarcity instrument (column 4) and the "district top 5" instrument (column 5). When combining all instruments, the estimated marginal effect of log hours is -2.1 percentage points (column 8).

The probability of future distress strongly depends on current supervisory risk assessments. Relative to a rating of 1, a rating of 3 increases probability of severe stress by 21 percentage points (Table 5, column 5). Due to the non-linearity of the probit specification, the marginal effect of supervision at a particular rating can differ from the average marginal effect. Figure 3 shows marginal effects conditional on the five rating categories compared to the unconditional marginal effect for each of the three outcome variables. The marginal effect of supervision varies considerably across banks of different riskiness. For severe stress, the effect increases (in absolute value) from about 0 for a bank currently rated 1 to about 8 percentage points for a bank rated 3 or worse.

Sub-categories of ROA. To gain a deeper insight into how increased supervision lowers bank distress, we estimate IV probits for the year-ahead realization of a high nonperforming loans ratio; low tier-1 capital ratio; and of low (high) realizations of the major income (expense) sub-categories of ROA, scaled by assets (Appendix Table A10). For each measure, a low (high) realization is defined as in the 10th (90th) percentile. We estimate the impact of log hours on each dependent variable using all instruments and including the current value of the dependent variable as a control. As shown in the table, increased supervision lowers the likelihood of a high nonperforming loan ratio (column 2, p-val<0.01) and of a low tier 1 capital ratio, although the latter effect is not significant at conventional levels (column 3). With respect to the ROA subcategories, the lower probability of a low ROA realization (column 1 replicates the last column of Table 6) is driven by lower probabilities of low non-interest income, high loan loss provisioning and low realized gains on securities.[26] These results are consistent with bank de-risking when supervision increases: Lower probabilities of high nonperforming loan ratio and high loan loss provisioning are consistent with safer loan portfolios; lower probability of low realized gains on securities indicates less-risky security holdings; lower probability of low non-interest income is consistent with more conservative positions in cash and derivative instruments. In unreported results, we also use averages, rather than tail events, for each outcome variable and find either smaller or insignificant effects. These results suggest that supervision has a greater effect on the tails of the distributions of bank performance and risk than its has on the averages.

5.2 Economies of scale

The hours cost function in (4) has a size elasticity of α, meaning that to achieve the same intensity of supervision at a bank with double the assets requires 2α hours, and economies of scale in supervision exist for $\alpha < 1$. To measure α, we can divide the coefficient β_A on log assets in the second stage (11), which is an estimate of $\alpha\gamma$, by the coefficient on log hours, which is an estimate of γ. This implies point estimates for the size elasticity of hours cost α between 0.5 and 0.7, indicative of large scale economies (significantly less than 1 for severe stress and low ROA, p-val<0.05). Such strong scale economies suggest that banks have not grown to be "too large to be supervise" and that breaking up large banks would require significant additional supervisory resources.

[26] The income-to-asset categories are: non-interest income, interest income and realized gains on securities not held to maturity. The expense-to-asset categories we consider are non-interest expense and loan loss provisioning.

Identification of α relies on the exclusion restriction that, controlling for risk, the failure threshold D_{idt} is not directly affected by size but only indirectly through the intensity of supervision s_{idt}, which is decreasing in size when hours are held fixed. Assessing this assumption requires testing for an effect of size on distress probability while controlling for the endogenous variation in hours. In unreported reduced-form IV probit estimates, we find no significant effect of log assets, consistent with the exclusion restriction.

5.3 Supervisory preferences

The preference weight W_{idt} given by the supervisor to a particular bank in the optimization problem (5) contains a preference shifter component, which we use to identify the model parameters, and a systematic component, with loadings $\widetilde{\alpha}$ on log assets and $\widetilde{\rho}_2, \ldots, \widetilde{\rho}_5$ on ratings $2, \ldots, 5$ (relative to rating 1). Proposition 1 shows that supervisory hours are increasing in both bank size and risk, even when the distress probability is weighted equally across banks (zero loadings in the preference weight) — because large banks require additional resources for the same intensity of supervision and because the marginal impact of supervision is higher at riskier banks. A positive loading on size ($\widetilde{\alpha} > 0$) implies a higher weight on the distress probability of a larger bank, for example, because a larger bank's failure may cause greater systemic spillovers; if such spillovers grew disproportionately with bank size, then so would the preference weight, $\widetilde{\alpha} > 1$. In contrast, there is no obvious prior whether the distress probability of a riskier bank (which already receives extra hours under equal weighting) should receive a greater or smaller preference weight.

Proposition 2 shows that the coefficients in the first stage (12) are estimates of convex combinations of the structural and preference parameters. For any variable x_{idt} entering the probability of distress with loading κ and the supervisory preference weight with loading $\widetilde{\kappa}$, while the second-stage coefficient β_x is an estimate of κ, the first-stage coefficient δ_x is an estimate of

$$\widehat{\delta}_x \mathrel{\widehat{=}} \overline{\pi} \frac{\kappa}{\gamma} + (1 - \overline{\pi}) \widetilde{\kappa}, \tag{14}$$

where the structural parameter κ appears normalized by γ corresponds to the second stage coefficient on log hours β_H. Without relying on estimates of $\overline{\pi}$, equation (14) implies that the unobserved preference parameter $\widetilde{\kappa}$ is *greater* than the first-stage coefficient $\widehat{\delta}_x$ if and only if the normalized second stage coefficient $\widehat{\beta}_x / \widehat{\beta}_H$ is *smaller* than the first-stage coefficient $\widehat{\delta}_x$: Figure 4 shows the normalized second-stage coefficients and the first-stage coefficients for log assets and the dummies for ratings 2 to 5 for each of the three distress outcomes. For log assets, the normalized second-stage coefficient is weakly smaller than the first-stage coefficient for all distress outcomes. This implies $\widetilde{\alpha} \geq 0$ — that is, supervisors give more weight to the distress probability of larger banks. In contrast, for ratings 2 to 5, the normalized second-stage coefficient is significantly larger than the first-stage coefficient for all distress outcomes (p-val<0.05). This implies $\widetilde{\rho}_r \ll \rho_r$, that is, the additional weight $\widetilde{\rho}_r$ that supervisors place on a bank rated r (compared to a bank rated 1) is considerably *lower* than the relative effect of rating r on the probability of distress.

To obtain point estimates of $\widetilde{\alpha}$ and $\widetilde{\rho}_2, \ldots, \widetilde{\rho}_2$, we construct the sample analog of the weight $\overline{\pi}$ from Proposition 2, $\overline{\pi} = E[\pi_{idt}]$, using fitted values $\widehat{\text{PD}}_{idt}$ and estimates of γ and σ_ε.[27] Using the estimated $\overline{\pi}$, the preference size elasticity $\widetilde{\alpha}$ is estimated to be 0.72 for severe stress, 0.71 for failure, and 0.65 for low ROA, which are all positive, consistent with the bounds above. These values of less than 1 would indicate that the increase in preference weight is less than proportional with the increase in bank size. However, the first stage includes several dummy variables that depend on size, which reduces the first-stage coefficient on log assets and therefore our estimate of $\widetilde{\alpha}$. If we use the overall unit size elasticity of hours, controlling only for ratings (Table 2, column 5), we obtain an estimate for $\widetilde{\alpha}$ of 1.24 for severe stress, 1.17 for failure, and 1.07 for low ROA. These values suggest that the preference weight increases at least proportionally with bank size as would be expected if larger banks had disproportionately larger distress costs.

Figure 5 shows the estimates for the preference weight loadings on rating, $\widetilde{\rho}_2, \ldots \widetilde{\rho}_5$, relative to rating 1. The negative values suggest that supervisors weight the distress probability of the higher-risk banks rated $2, \ldots, 5$ *less* than the distress probability of a low-risk bank rated 1. It may seem counter-intuitive that supervisors would weight less the banks that are already in relatively bad condition because, as shown above, they are much more likely to fail. Recall, however, that even under equal-weighting, riskier banks receive more attention since the marginal effect on their distress probability — and therefore the marginal benefit of hours — is higher (Proposition 1). Given that, empirically, riskier banks receive considerably more hours, the inverse preference weighting in Figure 5 only attenuates the cost-benefit effect. One interpretation of negative loadings on ratings is a difference between objective distress probabilities in the data and subjective distress probabilities used in the supervisory optimization. Such a difference could arise from probability weighting, where very small and very large probabilities receive disproportionate weight (Kahneman and Tversky, 1979) or stem from a supervisory assessment that distress probabilities are different than estimated from historical data (Gennaioli and Shleifer, 2018). A second interpretation is that an apparent preference tilt toward safer or riskier banks really is evidence of frictions in the allocation, such as adjustment costs or

[27] We obtain an estimate of $\sigma_\varepsilon = \sqrt{\text{var}(\varepsilon_{idt})}$ from the IV probit estimates of $\text{var}(v_{idt})$ and $\text{corr}(u_{idt}, v_{idt})$, using the fact that $v_{idt} = \delta_\eta \eta_{idt}$ and $u_{idt} = \eta_{idt} + \varepsilon_{idt}$, and the normalization $\text{var}(u_{idt}) = 1$.

minimum mandated allocations because of regulatory examination frequencies.

5.4 Robustness checks

The baseline results in Tables 5 and 6 present estimated effects of supervisory hours on bank outcomes using three outcome variables and four different instruments, both independently and jointly. In Appendix C, we consider two additional robustness exercises: (i) including an additional set of controls and (ii) running the estimation in a linear probability instrumental variable setting and comparing IV probit estimates with un-instrumented ones. We find that the IV probit effects are robust to including additional controls and to changing the specification to a linear probability IV model and show evidence of the endogeneity that requires our instrumental variables approach.

6 Counterfactual supervisory policy experiments

We study quantitative implications of three aspects of supervision documented above. First, supervisory resources at the Federal Reserve rose significantly post-2008 and were reallocated to larger banks. Second, the shadow cost of supervisory resources is not equalized across the twelve Federal Reserve districts and disproportionate attention is devoted to the largest banks in each district. Third, supervisory resources are reallocated to riskier banks but not as much as under a friction-less allocation resulting from neutral supervisory risk weights ($\widetilde{\rho} = 0$).

As counterfactual policy experiments, we alter the allocation of supervisory resources (or reduce them overall) and then trace out the implications on distress probabilities. We construct a counterfactual hours allocation $\widehat{\log H^*_{idt}}$ that differs from the one predicted by the first stage (12), then use the second stage (11) to predict a counterfactual probability of distress $\widehat{\mathrm{PD}}\left(\widehat{\log H^*_{idt}}, A_{idt}, r_{idt}\right)$, and obtain the predicted change in distress probability:[28]

$$\widehat{\Delta \mathrm{PD}}_{idt} = \widehat{\mathrm{PD}}\left(\widehat{\log H^*_{idt}}, A_{idt}, r_{idt}\right) - \widehat{\mathrm{PD}}\left(\widehat{\log H_{idt}}, A_{idt}, r_{idt}\right)$$

In all but one experiment, we impose constraints on total hours under the counterfactual allocation and construct lump-sum transfers such that overall resources are kept constant. In one experiment, we hold total resources fixed at pre-crisis levels. Equal weighted and asset-size weighted averages for the predicted changes in distress probability are shown in

[28]We use predicted hours $\widehat{\log H_{idt}}$ as a baseline — instead of actual hours $\log H_{idt}$ — to not conflate the effects of the prediction error with the effects of the counterfactual.

panels (b)–(g) of Table 7. Panel (a) shows baseline distress probabilities for comparison.

6.1 Post-2008 supervisory resource allocation

Following the 2007–09 financial crisis, Federal Reserve supervisory staff increased roughly 50 percent while the share of resources devoted to smaller banks nearly halved (Figure 1). We use the parameter estimates to trace out changes in the shadow cost of resources to compare the scale of the expansion to changes in the size and riskiness of supervised banks following the financial crisis. We then study implications for the probability of distress in a counterfactual in which resources are kept constant at their 2008 level.

During and after the financial crisis, new banks came under Fed supervision and the risk profile of the supervised banks changed. Figure 6 (left panel) shows total supervisory hours starting in 2006, which is the first year for which information on supervisory hours for all Fed districts is available, split by whether resources are spent at existing or new banks. Total supervisory hours increased about 50 percent following the financial crisis, matching the increase in headcounts in Figure 1, but new banks accounted for a significant fraction of the increase in total hours (blue area). The right panel of Figure 6 shows that the estimated shadow cost of supervisory resources (averaged across districts) rose starting in 2006, increasing by 30 percent during and following the crisis before declining to 10 percent below the 2006 level by 2014. This means that Fed supervisors faced tight resource constraints from 2007 through 2011, and only by 2013 had the expansion of resources reduced the shadow cost to the level of 2006.

No resource expansion. As a counterfactual, we consider an allocation in which resources are held constant at the level of 2008. In this counterfactual, resources at bank i in district d and year t are

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} - \hat{\tau}^{\text{expan}}_{dt}, \tag{15}$$

where $\hat{\tau}^{\text{expan}}_{dt}$ is a lump-sum transfer to keep total hours constant at the level of 2008 for each district (see Appendix A.4 for details on the calculation of τ). Under this counterfactual, the shadow cost (orange line in Figure 6, right panel) would have nearly doubled from its 2006 level. Table 7, panel (a) shows that average distress probability increases considerably compared to its baseline level. The probability of severe stress, for example, increases about 16 percent from the baseline (-88 basis points versus 534 for the baseline). The increase in distress probability is larger at risky banks (rating ≥ 3) and small banks ($<$ \$10 billion), the latter reflecting the post-2008 reallocation of resources from small to large banks.

No reallocation to large banks. Controlling for bank size and risk, point estimates in column (3) of Table 4 imply that hours at small banks ($<$ \$10 billion) dropped 24 percent post-2008 while hours at large banks increased 93 percent. To assess the effect of this reallocation, we consider counterfactual distress probabilities in the post-2008 period when resources are not reallocated to large banks,

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} - \widehat{\delta}_{\text{post-large}} \mathbb{I}[t > 2008]\, \mathbb{I}[A_{idt} > \$10\text{b}] + \widehat{\tau}^{\text{reall}}_{dt},$$

where $\widehat{\tau}^{\text{reall}}_{dt}$ is a transfer to keep total hours constant at the district-year level. Compared to the actual allocation, this counterfactual first removes the additional hours large banks receive post-2008. This creates slack in the budget constraint, implying a drop in the shadow cost λ_{dt}. To make use of the slack, the counterfactual then increases all banks' hours by the change in the shadow cost, represented by $\widehat{\tau}^{\text{reall}}_{dt}$.

Table 7, panel (c), and Figure 7a show that undoing this reallocation has the expected effect of increasing the distress probability at large banks while decreasing it at small banks. The decrease at small banks outweighs the increase at large banks, resulting in an overall reduction of distress probability even under bank-size-weighted averages (row "All banks," even columns). This means that the post-2008 reallocation made the banking system riskier from a microprudential perspective and suggests the role of macroprudential objectives such as disproportionate spillovers from distress at large banks. In fact, the estimated supervisory preference weight is increasing more than proportionally with bank size ($\widetilde{\alpha} > 1$).

6.2 Decentralization of resource allocation

We next study implications of the decentralized allocation of resources across Federal Reserve districts.

Resources perfectly mobile across district boundaries. The variation in the shadow cost of resources λ_{dt} across districts is as large as the time series variation (Section 4.1). We assess the size of this apparent inefficiency with a counterfactual allocation that equalizes the shadow cost across districts within each year. We construct the counterfactual allocation by offsetting the estimated effect $\widehat{\lambda}_{dt}$ of the district-specific Lagrange multiplier (which enters with a negative sign) and then applying a lump-sum transfer $\widehat{\tau}^{\text{mobile}}_{t}$ to keep total hours (across all districts) constant each year:

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} + \widehat{\lambda}_{dt} + \widehat{\tau}^{\text{mobile}}_{t} \tag{16}$$

Compared to the realized allocation, this counterfactual reallocates hours from districts with low shadow costs to districts with high shadow costs.

Table 7, panel (d), shows larger effects for size-weighted averages (even columns) than for equal-weighted averages (odd columns). When weighted by bank size, the probability of severe stress decreases 32 basis points compared to only 6 basis points when weighted equally, suggesting that districts with larger banks have systematically fewer resources. As shown in Figure 7b, compared to the average change in distress probability across districts (orange dashed line), the effects are several orders of magnitude larger at the districts receiving and losing the most resources. From Table 7, the effect is largest for risky banks (rating ≥ 3), implying that the benefits of flexible resource allocation across districts would mostly accrue at riskier banks.

No disproportionate supervision of district top 5 banks. The largest BHCs in each district receive additional attention; we consider a counterfactual by removing the top-5 effect to construct the allocation

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} - \widehat{\delta}_{\text{top5}} \mathbb{I}[i \in \text{top5}_{dt}] + \widehat{\tau}^{\text{top5}}_{dt},$$

where $\widehat{\tau}^{\text{top5}}_{dt}$ is a lump-sum transfer to keeps total hours constant at the district-year level. While Table 7, panel (e) shows that this reallocation away from the largest banks in each district leaves the size-weighted average distress probability almost unchanged, even among large banks ($\geq$ \$10 billion) sufficiently many receive additional hours to decrease their equal-weighted average distress probability.

6.3 Alternative response to bank risk

As noted in Section 3, supervision responds strongly to bank risk. For example, column (5) of Table 2 shows that relative to hours at a bank with the best possible rating of 1, hours at a bank rated 4 or 5 more than triple.[29] However, we estimate larger supervisory risk weights $\widetilde{\rho}$ for safer banks meaning that resources are not reallocated in proportion to the actual risk. We consider two counterfactuals. In the first, resources cannot respond to risk, while in the second, they respond proportionately to risk.

No response to risk. When resources cannot respond to risk, counterfactual hours are

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} - \textstyle\sum_{r=2}^{5} \widehat{\rho}_{Hr} \mathbb{I}[r_{idt} = r] + \widehat{\tau}^{\text{risk}}_{dt}, \tag{17}$$

[29] The coefficients of 1.49 and 1.64 imply increases of 340 percent and 420 percent, respectively ($\exp(1.49) - 1 \approx 3.4$).

where $\widehat{\tau}_{dt}^{\text{risk}}$ is a transfer to keep total hours constant at the district-year level. As shown in panel (f) of Table 7, the average probability of severe stress increases 51 basis points, or about 10 percent of the baseline (534 basis points, column 1). However, this overall average masks significant underlying heterogeneity. As shown in Figure 7c, the effect of the reallocation is monotonic in rating and, for example, the probability of failure increases by over 2 percentage points for 5-rated banks (middle panel).

Response proportional to risk. We next consider the effects of an allocation that fully responds to risk — that is, where the supervisory preference weights all ratings equally, $\widetilde{\rho}_r = 0$, such that $\delta_r = \overline{\pi}\frac{\rho_r}{\gamma} + (1-\overline{\pi})\,\widetilde{\rho}_r = \overline{\pi}\frac{\rho_r}{\gamma}$. Counterfactual hours are then given by

$$\widehat{\log H^*_{idt}} = \widehat{\log H_{idt}} - \textstyle\sum_{r=2}^{5}\widehat{\rho}_{Hr}\mathbb{I}[r_{idt}=r] + \sum_{r=2}^{5}\overline{\pi}\frac{\widehat{\rho}_r}{\widehat{\gamma}}\mathbb{I}[r_{idt}=r] + \widehat{\tau}_{dt}^{\text{prop}}, \tag{18}$$

where $\widehat{\tau}_{dt}^{\text{prop}}$ is a transfer to keep total hours constant at the district-year level. This counterfactual reallocates hours towards riskier banks and results in large decreases in their distress probabilities (Figure 7d). Table 7, panel (g), shows that the decrease at riskier banks more than offsets any increase at safer banks. For example, for probability of severe stress, the average decrease at risky banks (rating $\geq$ 3) is -5.7 percentage points while the average increase at safe banks is only 78 basis points, resulting in an overall average reduction of -28 basis points (column 1).

The effects of these two extreme scenarios for responding to risk are roughly equal in absolute magnitude (but with opposite sign). This suggests that the actual allocation is roughly equidistant from the two extreme scenarios. In sum, while the empirical sensitivity of supervisory resources with respect to bank risk is large, it is not as large as implied by a frictionless benchmark.

7 Conclusion and overall supervisory resources

We find that supervision has large economic effects on bank distress. Doubling hours at a supervised bank lowers the probability of future severe stress by 2.8 percentage points against a baseline probability of 5.3 percent. The marginal effect is stronger at riskier banks, with a marginal effect of about 8 percentage points for banks rated 3 or worse. We also find significant frictions in the allocation of resources. Across banks, the safest institutions seem to receive excess resources. Across Federal Reserve districts, shadow costs of resources are not equalized, even in the longer run. Counterfactual policy experiments reveal large benefits from the ability to reallocate resources across banks, especially for the riskiest ones. We finally find evidence of economies of scale in supervision, suggesting that the largest institutions have not grown to be "too large too supervise."

Our analysis takes the overall level of resources as given, but an important question is how the overall Federal Reserve budget for supervision should be determined. Although answering this question is beyond the scope of this analysis, we consider a simple back-of-the-envelope calculation of increasing the budget by 1 percent and comparing the marginal expected benefit to the marginal cost. According to the Federal Reserve's 2017 annual report, total operating expenses for supervision and regulation by the Federal Reserve System were $1.6 billion. Thus, increasing the budget by 1 percent would cost $16 million.

Assuming that total supervisory hours grow at the same rate as budgeted costs, and abstracting from estimation uncertainty, our estimates indicate that the resulting 1 percent increase in supervisory hours would lower the probability of failure by 0.004 percentage points on average. With bankruptcy cost estimates in the literature of about 12 percent and total bank holding company assets under Federal Reserve supervision of $19 trillion in 2017, this calculation implies a reduction in expected bankruptcy costs of roughly $90 million.

While they are on the same order of magnitude, the marginal benefit of $90 million seems to outweigh the marginal cost of $16 million. This calculation omits the required increase in resources at other agencies that Federal Reserve supervisors rely on. For example, the FDIC had 2017 operating expenditures of $1.9 billion, and total expenses in 2017 at the OCC were $1.2 billion. Including 1 percent increases in the budgets of both agencies increases the marginal cost to $47 million, still only half the marginal expected benefit. This would suggest that there would be significant benefits to increasing supervisory budgets from current levels. However, the simple back-of-the-envelope calculation omits important but hard to quantify effects of supervision, such as the benefit of avoiding spillover costs of bank failure and the cost of lost intermediation activity. We leave assessing the relative magnitude of such counteracting effects to future research.

The analysis in this paper also takes other pillars of banking policy, such as regulation, as given. It remains an open question to what extent resource-intensive supervision could be substituted with stricter regulation, such as requiring banks to hold more capital, as advocated, for example, by Admati and Hellwig (2013) or Cochrane (2011). We leave this question for future research.

References

Admati, A. and M. Hellwig (2013). *Bankers' New Clothes: What's Wrong with Banking and What to Do About It*. Princeton University Press.

Agarwal, S., D. Lucca, A. Seru, and F. Trebbi (2014). Inconsistent regulators: Evidence from banking. *The Quarterly Journal of Economics 129*(2), 889–938.

Aguiar, M. and E. Hurst (2007). Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics 122*(3), 969–1006.

Aikman, D., J. Bridges, A. Kashyap, and C. Siegert (2019). Would macroprudential regulation have prevented the last crisis? *Journal of Economic Perspectives 33*(1), 107–30.

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics 14*(1), 57–67.

Barth, J., G. Caprio, and R. Levine (2012). *Guardians of Finance: Making Regulators Work for Us*. MIT Press.

Basel Committee on Banking Supervision (2010). *Calibrating Regulatory Minimum Capital Requirements and Capital Buffers: A Topdown Approach*.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal 75*(299), 493–517.

Bisetti, E. (2018). The value of regulators as monitors: Evidence from banking. Working Paper.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2018). Children, time allocation, and consumption insurance. *Journal of Political Economy 126*(S1), S73–S115.

Board of Governors of the Federal Reserve System (2015). *Calibrating the GSIB Surcharge*.

Borio, C. (2011). Implementing the macro-prudential approach to financial regulation and supervision. In C. J. Green, E. J. Pentecost, and T. G. Weyman-Jones (Eds.), *The Financial Crisis and the Regulation of Finance*, pp. 101–117. Edward Elgar Cheltenham.

Carletti, E., G. Dell'Ariccia, and R. Marquez (2017). Supervisory incentives in a banking union. Working Paper.

Carpenter, D. and D. A. Moss (2013). *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press.

Cochrane, J. H. (2011). The more bank capital, the safer the bank. *Wall Street Journal 15*, a15.

Corbae, D., P. D'Erasmo, S. Galaasen, A. Irarrazabal, and T. Siemsen (2018). Structural stress tests. Working Paper.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014). Family welfare cultures. *The Quarterly Journal of Economics 129*(4), 1711–1752.

Dewatripont, M. and J. Tirole (1994). *The Prudential Regulation of Banks*. MIT Press.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–40.

Duffie, D. (2019). Prone to fail: The pre-crisis financial system. *Journal of Economic Perspectives 33*(1), 81–106.

Egan, M., A. Hortacsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review 107*(1), 169–216.

Eisenbach, T. M., A. Haughwout, B. Hirtle, A. Kovner, D. O. Lucca, and M. C. Plosser (2015). Supervising large, complex financial institutions: What do supervisors do? *FRB of New York Staff Report* (729).

FDIC (1997). *History of the Eighties – Lessons for the Future*.

Federal Reserve (2018). *Federal Reserve Supervision and Regulation Report*.

Flannery, M. J. (1998). Using market information in prudential bank supervision: A review of the u.s. empirical evidence. *Journal of Money, Credit and Banking 30*(3), 273–305.

Gennaioli, N. and A. Shleifer (2018). *A Crisis of Beliefs: Investor Psychology and Financial Fragility*. Princeton University Press.

Gopalan, Y., A. Kalda, and A. Manela (2017). Hub-and-spoke regulation and bank leverage. Working Paper.

Granja, J. and C. Leuz (2017). The death of a regulator: Strict supervision, bank lending and business activity. Working Paper.

Greenwood, R., J. C. Stein, S. G. Hanson, and A. Sunderam (2017). Strengthening and streamlining bank capital regulation. *Brookings Papers on Economic Activity 2017*(2), 479–565.

Harris, M. and A. Raviv (2014). How to get banks to take less risk and disclose bad news. *Journal of Financial Intermediation 23*(4), 437 – 470.

Hirtle, B., A. Kovner, and M. Plosser (2018). The impact of supervision on bank performance. *Journal of Finance* (forthcoming).

Hirtle, B. J. and J. A. Lopez (1999). Supervisory information and the frequency of bank examinations. Working Paper.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305–360.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Kisin, R. and A. Manela (2014). Funding and incentives of regulators: Evidence from banking. Working Paper.

Kroszner, R. S. and P. E. Strahan (1999, 11). What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics 114*(4), 1437–1467.

Lucca, D., A. Seru, and F. Trebbi (2014). The revolving door and worker flows in banking regulation. *Journal of Monetary Economics 65*(0), 17 – 32.

Marshall, D. A. and E. S. Prescott (2001). Bank capital regulation with and without state-contingent penalties. *Carnegie-Rochester Conference Series on Public Policy 54*(1), 139–184.

Olea, J. L. M. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Peek, J., E. S. Rosengren, and G. M. Tootell (1999). Is bank supervision central to central banking? *The Quarterly Journal of Economics 114*(2), 629–653.

Radner, R. and M. Rothschild (1975). On the allocation of effort. *Journal of Economic Theory 10*(3), 358–376.

Repullo, R. and J. Suarez (2012). The procyclical effects of bank capital regulation. *The Review of financial studies 26*(2), 452–490.

Rochet, J.-C. (2004). Macroeconomic shocks and banking supervision. *Journal of Financial Stability 1*(1), 93–110.

Rochet, J.-C. (2007). *Why Are there So Many Banking Crises? The Politics and Policy of Bank Regulation*. Princeton University Press.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science 2*(1), 3–21.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

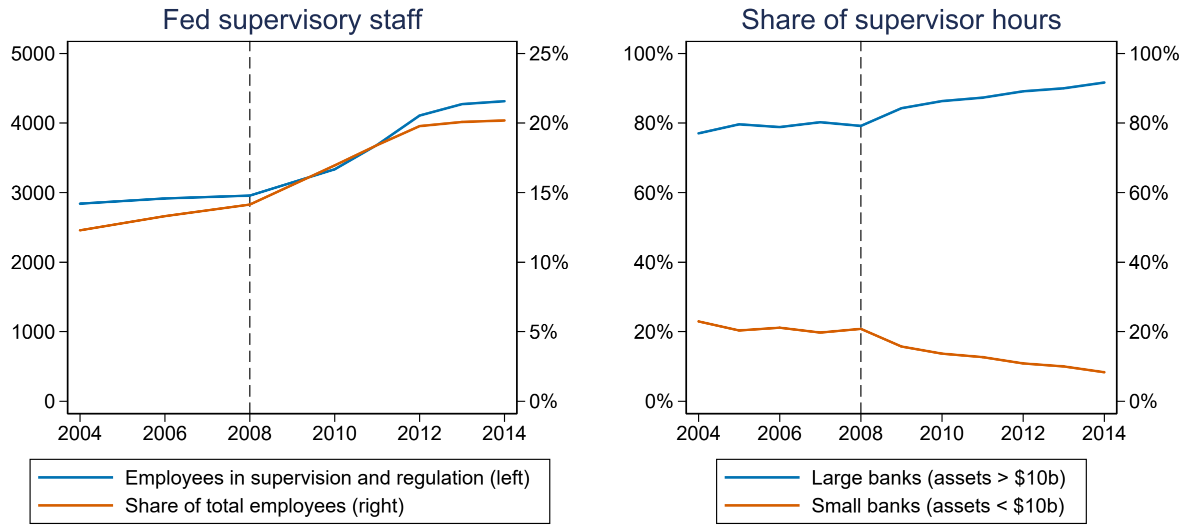


Figure 1: Federal Reserve supervisory staff and allocation of resources. The data on employees is from Federal Reserve Annual reports. The data on resources is from internal hours data for supervisory examiners at the Federal Reserve. The hours data in the right panel exclude resources allocated to institutions that were not under Federal Reserve supervision pre-2008.

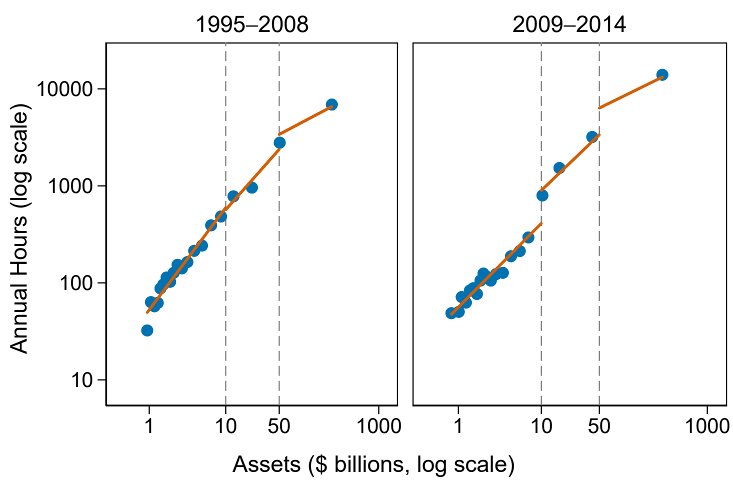


Figure 2: Increased attention to large banks. This figure presents binned scatter plots and the fitted lines of regressing log supervisory hours on log assets and supervisory ratings for different bank size categories ($10 billion and $50 billion asset thresholds) before and after 2008.

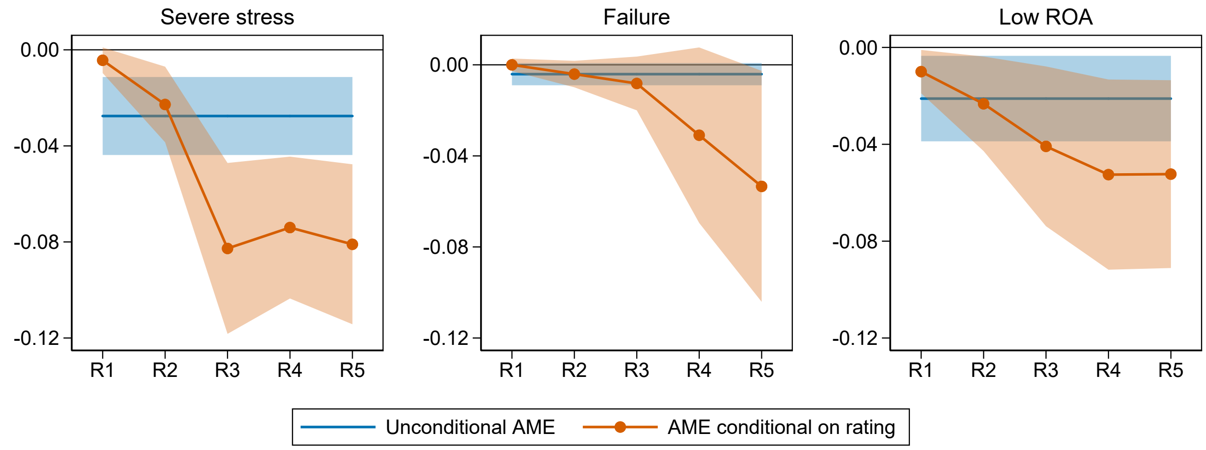


Figure 3: Marginal effect of log hours on probability of distress conditional on different ratings. The figure shows the (average) marginal effect of log hours on a bank's next-year probability of distress (left panel: severe stress; middle panel: failure; right panel: low ROA) unconditionally (blue) and evaluated at each of the five supervisory ratings (orange). Shaded areas represent 95% confidence intervals computed based on standard errors clustered at the bank level via the delta method.

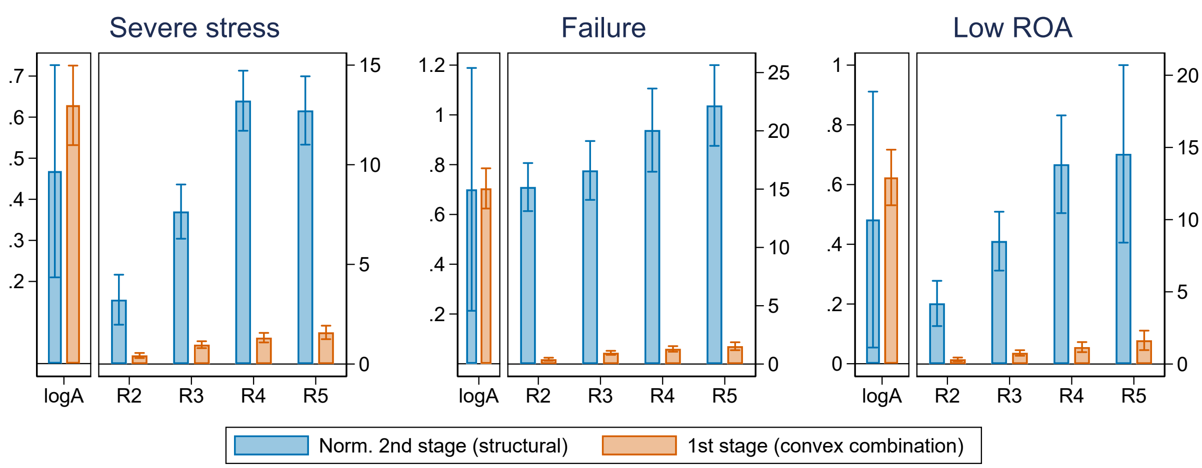


Figure 4: Comparison of coefficients on size and ratings from first and second stage. The figure shows the coefficients on log assets and the dummies for rating 2 to rating 5 from the second-stage equation (11) and the first-stage equation (12) for the three outcome variables probability of severe stress, probability of failure, and probability of low ROA. The second stage coefficients are normalized by the second stage coefficient on log hours. Whiskers represent 95% confidence intervals.

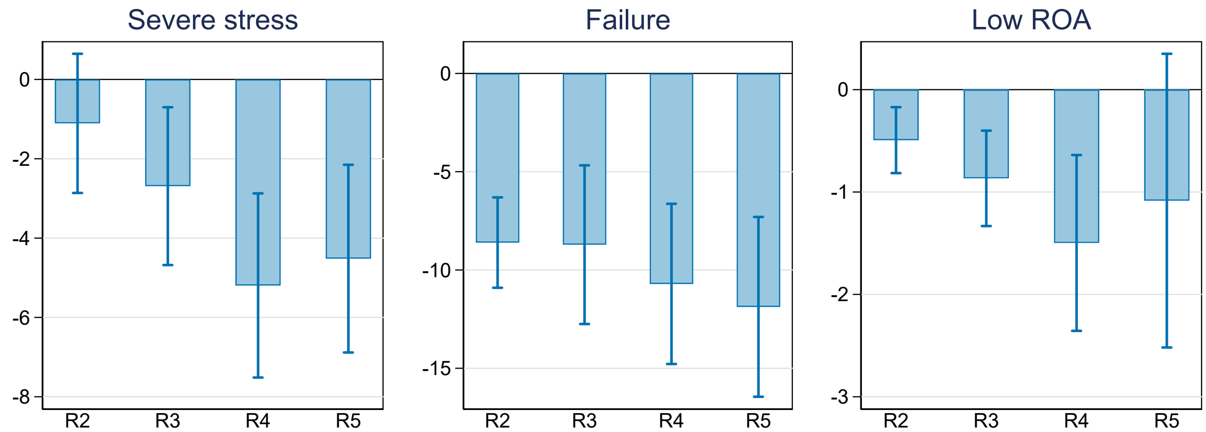


Figure 5: Implied supervisory preference weight relative to rating 1. The figure shows the implied loading of the supervisory preference weight W_{idt} for rating 2 to rating 5 (left panel: probability of failure or rating 4 or 5; right panel: probability of failure only). Whiskers represent 95% confidence intervals based on bootstrapped standard errors clustered at the bank level (1,000 replications).

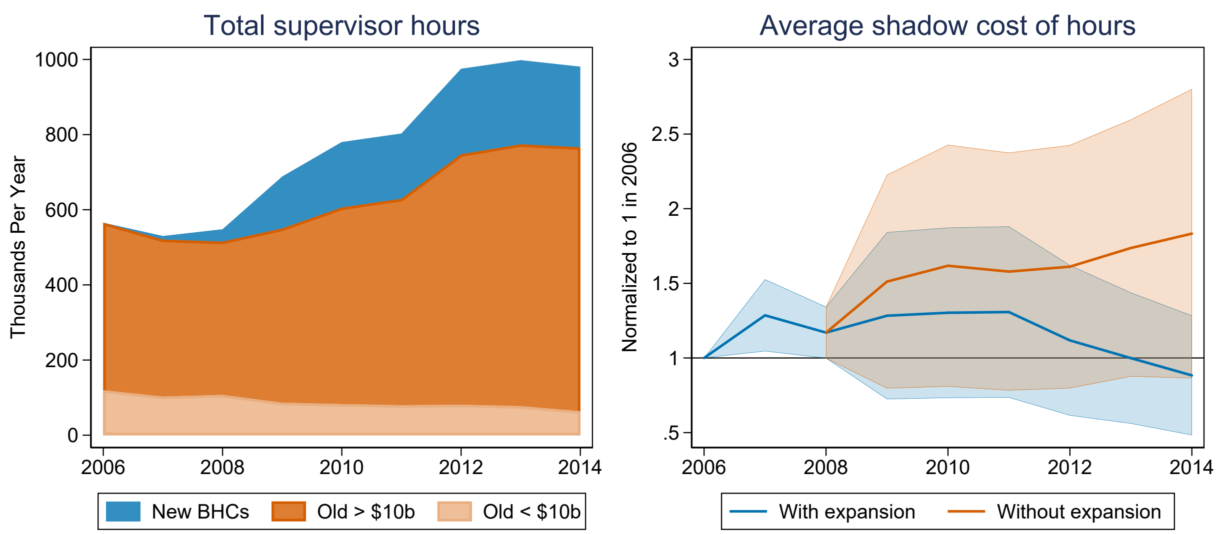
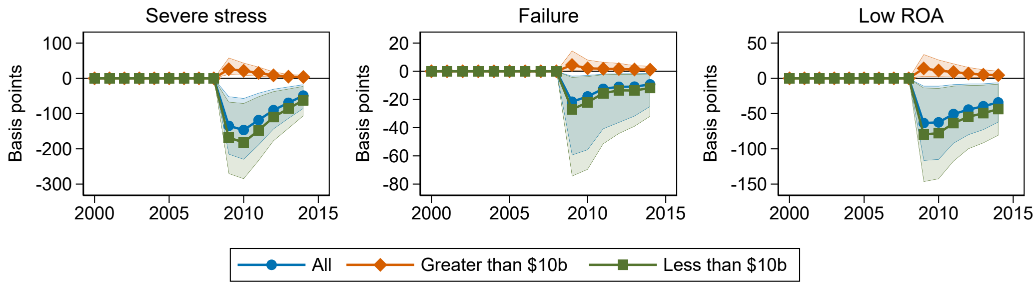
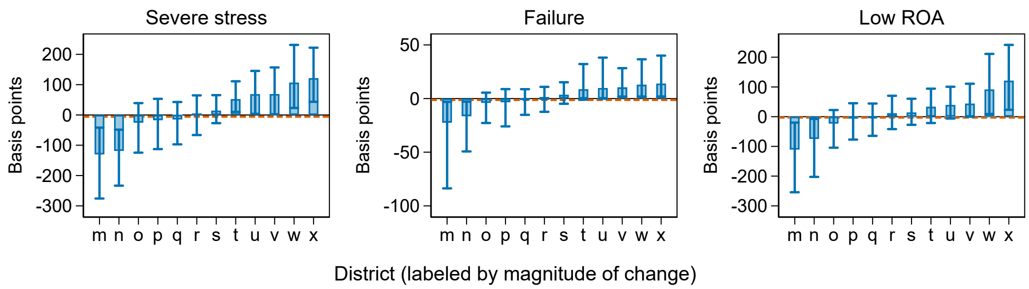


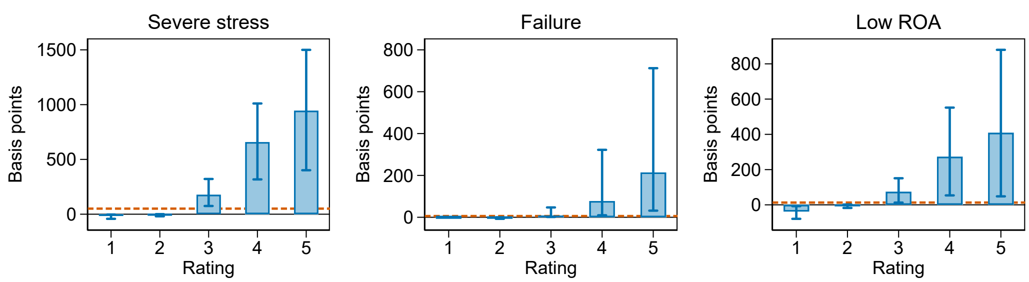
Figure 6: Allocation of supervisor hours and average shadow cost of hours. The left panel shows the allocation of total supervisor hours across BHCs under Fed supervision as of 2006, split between large BHCs ($\geq$\$10b) and small BHCs (<\$10b), as well as new BHCs coming under Fed supervision in the following years. The right panel shows the average across districts of the estimated shadow cost of supervisor hours, under the actual allocation and under the counterfactual allocation if resources were kept fixed at the level of 2008. Shaded areas represent 95% confidence intervals based on bootstrapped standard errors clustered at the bank level (1,000 replications).



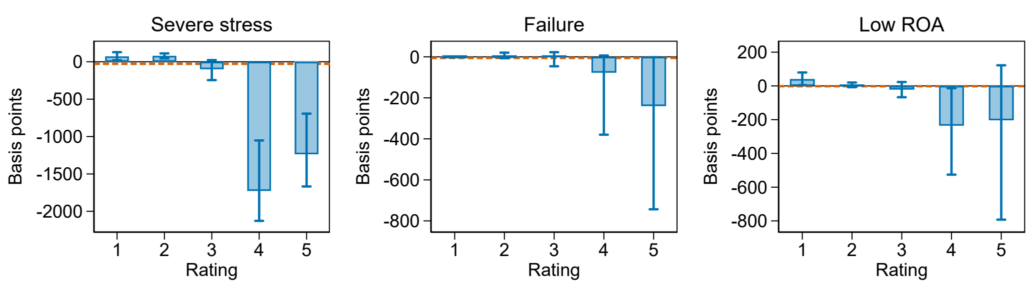
(a) No post-2008 reallocation to large banks.



(b) Resources perfectly mobile across district boundaries.



(c) No response to risk.



(d) Response proportional to risk.

Figure 7: Change in distress probability under counterfactual policy experiments. The figure shows equal-weighted averages of the change in bank-level distress probability implied by the counterfactuals. See Section 6 for details on the counterfactuals. Shaded areas and whiskers represent 95% confidence intervals based on bootstrapped standard errors clustered at the bank level (1,000 replications).

Table 1: Summary statistics. This table presents summary statistics for the variables included in the regression specifications. For detailed variable definitions see Section 3 and Appendix E. Sample is 1998–2014.

(a) Outcome variables (percent).

	All banks			Assets<\$10b		Assets≥\$10b	
	Mean	StDev	Obs	Mean	StDev	Mean	StDev
Failure	0.53	7.28	5445	0.59	7.65	0.29	5.41
Severe stress	5.34	22.49	5445	5.88	23.53	3.03	17.15
Low ROA	10.34	30.45	5405	10.35	30.47	10.30	30.42
High NPL ratio	10.21	30.28	5465	10.09	30.12	10.75	30.98
Low tier-1 capital ratio	10.29	30.39	5470	8.08	27.25	19.73	39.82
Low noninterest income	9.52	29.36	5502	10.32	30.42	5.92	23.62
High noninterest expense	10.31	30.42	5498	9.65	29.53	13.27	33.95
High loan-loss provisions	10.23	30.31	5452	9.46	29.26	13.61	34.31
Low net interest income	9.95	29.94	5528	8.05	27.21	18.74	39.04
Low real. gains on securities	10.20	30.26	5404	9.74	29.65	12.22	32.77

(b) Explanatory variables.

	All banks			Assets<\$10b		Assets≥\$10b	
	Mean	StDev	Obs	Mean	StDev	Mean	StDev
Hours (thousands)	1.54	4.94	5900	0.31	0.50	6.87	9.70
Assets (real, \$ bil.)	32.97	174.62	5900	2.96	2.26	162.75	376.44
Log Hours	5.34	2.08	5900	4.74	1.70	7.91	1.57
Log Assets (real)	8.35	1.44	5900	7.77	0.63	10.83	1.33
Rating	1.98	0.78	5900	1.97	0.81	2.02	0.64
Return on assets	0.01	0.01	5451	0.01	0.01	0.01	0.01
Non-perf. loans ratio	0.02	0.02	5542	0.02	0.02	0.02	0.02
Tier-1 capital ratio	0.12	0.03	5565	0.12	0.03	0.11	0.03
District top 5	0.15	0.35	5900	0.03	0.17	0.65	0.48
Post-2008	0.46	0.50	5900	0.47	0.50	0.42	0.49
Assets≥\$10b	0.19	0.39	5900	0.00	0.00	1.00	0.00
Assets≥\$50b	0.07	0.26	5900	0.00	0.00	0.40	0.49
Complex	0.28	0.45	5900	0.18	0.39	0.67	0.47
Stressed (rating≥3)	0.15	0.36	5900	0.15	0.35	0.16	0.37
High exam frequency	0.51	0.50	5900	0.39	0.49	1.00	0.00
Exam count	1.47	2.39	5900	0.84	0.78	4.18	4.32
Noninterest income	1.62	3.61	5900	1.42	3.47	2.49	4.05
Noninterest expense	3.33	2.82	5900	3.27	2.84	3.55	2.73
Loan-loss provisions	0.57	1.17	5900	0.57	1.23	0.57	0.86
Net interest income	3.23	1.06	5900	3.32	1.05	2.86	0.99
Realized gains on securities	0.01	0.27	5898	0.01	0.27	0.00	0.30

Table 2: Basic determinants of supervisory hours. The table shows estimates from linear regressions of log supervisory hours on the listed controls. For detailed variable definitions, see Section 3 and Appendix E. Standard errors clustered by bank reported in brackets; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sample is 1998–2014.

	Log(Hours)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Assets (real)	1.003***	0.673***	0.994***	0.689***	0.966***	0.750***
	[0.027]	[0.095]	[0.027]	[0.097]	[0.030]	[0.094]
Return on assets	-0.159***	-0.224***	-0.018	-0.110**		
	[0.048]	[0.048]	[0.047]	[0.048]		
Non-perf. loans ratio	0.101***	0.076***	0.011	0.013		
	[0.021]	[0.024]	[0.022]	[0.026]		
Tier-1 capital ratio	-0.024**	-0.006	-0.012	-0.005		
	[0.010]	[0.012]	[0.010]	[0.012]		
Rating = 2			0.376***	0.218**	0.419***	0.238***
			[0.069]	[0.093]	[0.069]	[0.089]
Rating = 3			1.032***	0.740***	1.153***	0.870***
			[0.109]	[0.127]	[0.093]	[0.107]
Rating = 4			1.429***	1.041***	1.489***	1.176***
			[0.193]	[0.188]	[0.115]	[0.134]
Rating = 5			1.192***	1.152***	1.642***	1.356***
			[0.383]	[0.324]	[0.169]	[0.184]
Bank FEs	No	Yes	No	Yes	No	Yes
Adj. R^2	0.51	0.62	0.52	0.62	0.49	0.61
Observations	5243	5243	5243	5243	5900	5900
Distinct BHCs	739	739	739	739	769	769

Table 3: Instruments for supervisory hours: resource scarcity. The table shows estimates from linear regressions of log supervisory hours on the listed controls. Other controls are noted at the bottom. District averages either leave out or leave in bank i, as noted at the bottom. National averages leave out bank i's district. For detailed variable definitions, see Section 3 and Appendix E. F-statistics are for the test that the coefficients on the instruments are zero. Standard errors clustered by bank reported in brackets; significance: $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample is 1998–2014.

	Log(Hours)			
	(1)	(2)	(3)	(4)
District avg. Log Hours	0.761***	1.000***	0.738***	0.726***
	[0.053]	[0.053]	[0.052]	[0.062]
District avg. Log Assets	-0.791***	-0.985***	-0.792***	-0.740***
	[0.087]	[0.088]	[0.088]	[0.103]
District avg. (Rating = 2)	-0.033	-0.347*	0.028	0.218
	[0.182]	[0.183]	[0.213]	[0.244]
District avg. (Rating = 3)	-0.783***	-1.205***	-0.528	-0.423
	[0.297]	[0.297]	[0.360]	[0.380]
District avg. (Rating = 4)	-1.400**	-1.635**	-1.453**	-1.246*
	[0.647]	[0.644]	[0.733]	[0.700]
District avg. (Rating = 5)	-2.483***	-1.913**	-2.206**	-2.329***
	[0.856]	[0.852]	[0.910]	[0.846]
National avg. Log Hours			-0.138	
			[0.153]	
National avg. Log Assets			0.099	
			[0.272]	
National avg. (Rating = 2)			-0.505	
			[0.561]	
National avg. (Rating = 3)			0.390	
			[1.097]	
National avg. (Rating = 4)			-3.391	
			[3.687]	
National avg. (Rating = 5)			1.006	
			[3.526]	
Dist. avg. Post-2008×(Assets≥$10b)				-0.438
				[0.338]
Dist. avg. District top 5				0.738
				[0.543]
Dist. avg. Lagged exam count				0.314
				[0.232]
Dist. avg. Lag exam ct.×(Hi. exam freq.)				-0.345
				[0.248]
Dist. avg. High exam frequency				0.260
				[0.209]
Log Assets, Ratings	Yes	Yes	Yes	Yes
Leave-out average	Yes	No	Yes	Yes
F-statistic	43.6	69.0	22.5	24.2
Adj. R^2	0.53	0.56	0.53	0.54
Observations	5900	5900	5900	5188
Distinct BHCs	769	769	769	722

Table 4: Instruments for supervisory hours: preference shocks. The table shows estimates from linear regressions of log supervisory hours on the listed controls. Other controls are noted at the bottom. For detailed variable definitions, see Section 3 and Appendix E. F-statistics are for the test that the coefficients on the instruments are zero. Standard errors clustered by bank reported in brackets; significance: $^{*}\,p < 0.1$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. Sample is 1998–2014.

	Log(Hours)					
	(1)	(2)	(3)	(4)	(5)	(6)
District top 5	0.442***					0.407***
	[0.120]					[0.116]
Post-2008		-0.269***	-0.270***			-0.262***
		[0.065]	[0.065]			[0.063]
Assets≥\$10b		0.335**	0.325**	0.727***		
		[0.146]	[0.134]	[0.136]		
Post-2008×(Assets≥\$10b)		0.625***	0.659***			0.534***
		[0.140]	[0.117]			[0.109]
Assets≥\$50b		-0.106				
		[0.222]				
Post-2008×(Assets≥\$50b)		0.089				
		[0.216]				
Small (assets<\$10b), complex				0.529***		
				[0.105]		
Small (assets<\$10b), stressed				0.082		
				[0.151]		
High exam frequency					0.527***	0.504***
					[0.088]	[0.087]
Lagged exam count				-0.627***	-0.636***	-0.620***
				[0.119]	[0.119]	[0.121]
Lag exam ct.×(Assets≥\$10b)				0.727***		
				[0.122]		
Lag exam ct.×(Small, complex)				0.871***		
				[0.128]		
Lag exam ct.×(Small, stressed)				0.668***		
				[0.156]		
Lag exam ct.×(Hi. exam freq.)					0.740***	0.718***
					[0.122]	[0.123]
Log Assets, Ratings	Yes	Yes	Yes	Yes	Yes	Yes
F-statistic	13.6	15.6	31.9	16.2	36.7	22.7
Adj. R^2	0.49	0.50	0.50	0.57	0.56	0.57
Observations	5900	5900	5900	5188	5188	5188
Distinct BHCs	769	769	769	722	722	722

Table 5: Second stage of IV probit with outcome variable 1y-ahead probability of severe stress. The table shows estimates from the second stages of IV probit regressions of distress probability on the listed controls where log hours are instrumented for (corresponding first stages in Table A7). The instruments used and other controls are noted at the bottom. Instrument abbreviations: "Lambda" is shadow cost, "Top5" is district top 5, "P08G10" is post-2008 × assets≥\$10b, "LExCt" is lagged exam count, "All" is all instruments. Other controls abbreviations: "P08" is post-2008, "G10" is assets≥\$10b, "HF" is high exam frequency, "LRT" is lagged ratings. For detailed variable definitions, see Section 3 and Appendix E. The effective F-statistic and critical value are for the weak-instrument test of Olea and Pflueger (2013), robust to heteroskedasticity, autocorrelation, and clustering, from the respective first stage. Average marginal effects reported in curly braces. Standard errors clustered by bank reported in brackets; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sample is 1998–2014.

	Severe stress				
	(1)	(2)	(3)	(4)	(5)
Log Hours	-0.238***	-0.612***	-0.122	-0.216***	-0.287***
	[0.069]	[0.054]	[0.309]	[0.076]	[0.060]
	{-0.022}	{-0.133}	{-0.009}	{-0.019}	{-0.028}
Log Assets (real)	0.151**	0.546***	0.012	0.081	0.134**
	[0.071]	[0.072]	[0.274]	[0.069]	[0.057]
	{0.014}	{0.119}	{0.001}	{0.007}	{0.013}
Rating = 2	0.604***	0.497***	0.636***	0.898***	0.926***
	[0.151]	[0.097]	[0.185]	[0.188]	[0.184]
	{0.055}	{0.108}	{0.048}	{0.079}	{0.089}
Rating = 3	1.624***	1.358***	1.757***	2.062***	2.194***
	[0.166]	[0.187]	[0.337]	[0.199]	[0.199]
	{0.148}	{0.295}	{0.134}	{0.181}	{0.211}
Rating = 4	2.942***	2.136***	3.150***	3.661***	3.790***
	[0.158]	[0.361]	[0.353]	[0.221]	[0.221]
	{0.268}	{0.464}	{0.240}	{0.322}	{0.364}
Rating = 5	2.754***	2.115***	2.962***	3.502***	3.650***
	[0.208]	[0.338]	[0.438]	[0.254]	[0.251]
	{0.251}	{0.460}	{0.226}	{0.308}	{0.351}
Instrument	Lambda	Top5	P08G10	LExCt	All
Other controls			P08 G10	HF LRT	P08 HF LRT
F-statistic	53.0	10.1	29.5	33.9	37.0
Critical value	15.1	23.1	23.1	10.5	16.3
Observations	5445	5445	5445	4764	4764
Distinct BHCs	744	744	744	704	704

Table 6: Second stage of IV probit with outcome variables 1y-ahead probability of failure and 1y-ahead probability of low ROA. The table shows estimates from the second stages of IV probit regressions of distress probability on the listed controls where log hours are instrumented for (corresponding first stages in Table A8). The outcome variable is noted at the top. The instruments used and other controls are noted at the bottom. Instrument abbreviations: "Lambda" is shadow cost, "Top5" is district top 5, "P08G10" is post-2008 × assets≥\$10b, "LExCt" is lagged exam count, "All" is all instruments. Other controls abbreviations: "P08" is post-2008, "G10" is assets≥\$10b, "HF" is high exam frequency, "LRT" is lagged ratings. For detailed variable definitions, see Section 3 and Appendix E. The effective F-statistic and critical value are for the weak-instrument test of Olea and Pflueger (2013), robust to heteroskedasticity, autocorrelation, and clustering, from the respective first stage. Average marginal effects reported in curly braces. Standard errors clustered by bank reported in brackets; significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sample is 1998–2014.

	Failure			Low ROA				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Hours	0.060	-0.347***	-0.235***	-0.122**	-0.391***	-0.167	0.056	-0.149**
	[0.134]	[0.045]	[0.081]	[0.056]	[0.151]	[0.180]	[0.079]	[0.059]
	{0.001}	{-0.010}	{-0.004}	{-0.018}	{-0.074}	{-0.024}	{0.008}	{-0.021}
Log Assets (real)	-0.123	0.244***	0.165**	0.101*	0.364**	0.072	-0.082	0.072
	[0.126]	[0.065]	[0.076]	[0.058]	[0.152]	[0.149]	[0.065]	[0.053]
	{-0.002}	{0.007}	{0.003}	{0.015}	{0.069}	{0.010}	{-0.012}	{0.010}
Rating = 2	0.251	3.585***	3.566***	0.415***	0.466***	0.545***	0.500***	0.627***
	[0.222]	[0.272]	[0.256]	[0.079]	[0.071]	[0.092]	[0.122]	[0.119]
	{0.004}	{0.099}	{0.062}	{0.060}	{0.089}	{0.078}	{0.071}	{0.089}
Rating = 3	0.114	3.974***	3.902***	0.653***	0.882***	1.070***	0.872***	1.271***
	[0.329]	[0.329]	[0.310]	[0.116]	[0.141]	[0.192]	[0.165]	[0.155]
	{0.002}	{0.110}	{0.068}	{0.095}	{0.168}	{0.153}	{0.124}	{0.180}
Rating = 4	0.622*	4.766***	4.714***	1.087***	1.368***	1.606***	1.608***	2.066***
	[0.353]	[0.425]	[0.437]	[0.183]	[0.193]	[0.259]	[0.269]	[0.258]
	{0.010}	{0.132}	{0.082}	{0.158}	{0.261}	{0.229}	{0.229}	{0.292}
Rating = 5	0.979**	5.269***	5.211***	1.368***	1.764***	1.891***	1.664***	2.173***
	[0.435]	[0.430]	[0.419]	[0.355]	[0.382]	[0.458]	[0.442]	[0.468]
	{0.016}	{0.146}	{0.091}	{0.199}	{0.336}	{0.270}	{0.237}	{0.307}
Instrument	Lambda	LExCt	All	Lambda	Top5	P08G10	LExCt	All
Other controls		HF LRT	HF LRT			P08 G10	HF LRT	P08 HF LRT
F-statistic	53.0	33.9	41.8	53.7	14.9	33.4	27.6	34.8
Critical value	15.1	10.9	16.5	15.2	23.1	23.1	14.3	17.2
Observations	5445	4764	4764	5274	5274	5274	4594	4594
Distinct BHCs	744	704	704	745	745	745	675	675

Table 7: Change in distress probability under counterfactual policy experiments. The table shows averages of the bank-year level distress probabilities in panel (a) and changes in distress probability under counterfactual policy experiments in panels (b)–(g), both in basis points. See Section 6 for details on the counterfactuals. Distress outcome variables are noted at the top. "EW" denotes equal-weighted and "SW" size-weighted averages. Panels (b) and (c) include only the post-2008 period. Significance based on bootstrapped standard errors clustered at the bank level (1,000 replications): * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	Severe stress		Failure		Low ROA	
	(1) EW	(2) SW	(3) EW	(4) SW	(5) EW	(6) SW
(a) Baseline probability						
All banks	534.4***	337.5*	53.3***	8.0	1034.2***	1125.9***
Assets<\$10b	588.0***	540.6***	58.8***	55.2***	1035.1***	913.7***
Assets≥\$10b	303.0***	321.5	29.3	4.2	1030.4***	1141.6***
Rating<3	162.5***	40.0**	43.3***	6.3	750.1***	735.8***
Rating≥3	2602.4***	1188.3*	108.4***	12.7	3035.7***	2061.2***
(b) No post-2008 expansion						
All banks	88.3	54.1	15.0	7.9	45.5	32.0
Assets<\$10b	93.0	89.6	16.1	16.5	46.5	45.6
Assets≥\$10b	66.7	51.7	9.9	7.3	40.7	31.2
Rating<3	51.6*	17.1	14.7	9.7	38.0	24.5
Rating≥3	184.4	88.7	15.7	6.2	65.1	39.1
(c) No post-2008 reallocation						
All banks	-101.2***	-3.1	-13.8	-0.5	-49.0**	-1.0
Assets<\$10b	-125.9***	-113.6***	-17.2	-17.3	-61.4**	-58.4**
Assets≥\$10b	12.9**	4.2**	2.0	0.7	8.3*	2.8*
Rating<3	-38.9***	-3.3*	-10.0	-0.4	-32.1**	-1.4
Rating≥3	-263.9***	-2.9	-23.7	-0.5	-93.2**	-0.6
(d) Resources perfectly mobile						
All banks	-5.5	-32.2*	-1.2	-3.9	-2.6	-25.0*
Assets<\$10b	-3.3	-6.6	-1.0	-1.2	-1.0	-3.3
Assets≥\$10b	-15.0	-34.2*	-2.1	-4.1	-9.5	-26.7*
Rating<3	-5.2	-11.8	-0.4	-3.9	-2.6	-22.0
Rating≥3	-7.5	-80.7*	-5.5	-4.0	-2.7	-32.1
(e) No top-5 effect						
All banks	-17.5	-0.5	-2.3	-0.0	-12.9	0.3
Assets<\$10b	-20.8	-18.7	-2.7	-2.6	-15.3	-14.4
Assets≥\$10b	-4.3	0.8	-0.6	0.2	-3.4	1.4
Rating<3	-10.1	-0.4	-1.7	0.1	-11.2	0.5
Rating≥3	-55.3	-0.8	-5.0	-0.2	-21.5	-0.2
(f) No response to risk						
All banks	51.0***	27.2***	5.8	0.6	13.0*	6.7**
Assets<\$10b	56.7***	45.4***	6.9	4.8	15.0*	10.0
Assets≥\$10b	27.8***	25.9***	1.6	0.3	4.9	6.4**
Rating<3	-10.6*	-3.2	-2.1	-1.9	-14.6**	-7.4**
Rating≥3	365.4***	96.6***	46.5	6.3	153.8**	38.7**
(g) Response proportional to risk						
All banks	-27.8*	-10.7*	-5.4	0.2	-1.8	-0.2
Assets<\$10b	-32.8*	-14.3	-6.5	-4.6	-2.8	0.1
Assets≥\$10b	-7.7	-10.5	-1.1	0.5	2.6	-0.2
Rating<3	78.0***	32.4***	1.3	1.7	14.9**	7.4*
Rating≥3	-567.8***	-109.2***	-39.9	-3.3	-86.8**	-17.5**