Federal Reserve Bank of New York
Staff Reports

Parsing the Content of Bank Supervision

Paul Goldsmith-Pinkham
Beverly Hirtle
David Lucca

Staff Report No. 770
March 2016

Parsing the Content of Bank Supervision

Paul Goldsmith-Pinkham, Beverly Hirtle, and David Lucca

Federal Reserve Bank of New York Staff Reports, no. 770
March 2016
JEL classification: G21, G28

Abstract

We measure bank supervision using the database of supervisory issues, known as matters requiring attention or immediate attention, raised by Federal Reserve examiners to banking organizations. The volume of supervisory issues increases with banks' asset size, especially for the largest and most complex banks, and decreases with profitability and the quality of the loan portfolio. Stressed banks are faster at resolving issues, but all else equal, resolving new issues takes longer the more issues a bank faces, which may suggest capacity constraints in addressing multiple supervisory issues. Using computational linguistic methods on the text of the issue description, we define five categorical issue topics. The subset of issues related to capital levels and loan portfolio are the most consequential in terms of supervisory rating downgrades and are directly related to changes in banks' balance sheet characteristics and profitability. Other issues appear to reflect soft information and are less correlated with bank observables. By categorizing questions asked by analysts at banks' quarterly earnings calls using the same linguistic approach, we find that market monitors raise issues similar to those of supervisors when the issues are related to hard information (such as loan quality or capital) and public supervisory assessment programs.

Key words: bank supervision, bank regulation, market monitoring, text classification, Latent Dirichlet Allocation

---

Goldsmith-Pinkham, Hirtle, Lucca: Federal Reserve Bank of New York (e-mails: paul.goldsmith-pinkham@ny.frb.org, beverly.hirtle@ny.frb.org, david.lucca@ny.frb.org). The authors thank John Goldsmith for useful discussions. Eilidh Geddes provided excellent research assistance.

1 Introduction

While important to economic growth, banks have also posed threats to economic prosperity in a number of severe banking crises over the past century. Banking authorities are tasked with ensuring financial stability to mitigate these threats.[1] The Basel Committee on Banking Supervision coordinates efforts of these authorities around three pillars: regulatory capital, market discipline, and supervisory review. Supervisory review involves examining institutions to assess their safety and soundness and using the resulting information to demand corrective actions if banks' conditions or practices are deemed unsafe or unsound, or if they are not in compliance with regulations. This paper extracts textual information from issues raised by Federal Reserve supervisors to banks to provide a unique measure of supervision and its interaction with the other Basel pillars.

The analysis is based on the timing, volume and content of confidential supervisory actions known as matters requiring attention (MRA) and *immediate* attention (MRIAs), which are supervisory demands to banks for corrective action. While these type of demands have long existed, MRAs and MRIAs were formalized in 2008 to improve the consistency and clarity of written communication of Federal Reserve (Fed) supervisors to banks (SR 08-1). We use an unsupervised machine learning method known as latent dirichlet allocation (LDA; see Blei, Ng, and Jordan, 2003) to classify issues into five categories, or topics. We use these topics to study what supervision is and how it varies, as well as to understand the extent to which market monitors complement bank supervisory efforts.

The data include about 38,000 supervisory actions issued between 2009 and 2014 either to bank holding companies (BHCs) or state member banks (SMBs), henceforth jointly referred to as banks. Our linguistic method categorizes these issues into five topics: compliance & regulation, internal controls, risk modeling, loan portfolio and capital & liquidity. The categories reflect the different information on which Fed supervisors focus their analysis: some categories relate primarily to what we classify as soft information (e.g. internal processes are related to information technology and audits), while for others, hard information appears to play a more important role (e.g. loan portfolio and capital & liquidity). When compared to soft information, we think of hard information as verifiable to outsiders and more easily quantified. Overall, only nine percent of issue openings are related to risk modeling, and almost half (47%) relate to soft information (internal processes and compliance & regulation).

We first examine what factors are associated with issue openings and closings – that is, when new MRAs and MRIAs are issued by examiners and when bank remediation has been

[1] For example, the Federal Reserve was established "[...] to furnish an elastic currency, to afford means of rediscounting commercial paper, to establish *a more effective supervision of banking in the United States*, and for other purposes," Federal Reserve Act (1913) (emphasis added by the authors).

sufficient for examiners to close the finding. The overall number of issue openings is associated with factors related to the extent of supervisory attention on a bank, such as a BHC being included in publicly disclosed supervisory assessments (e.g., the Comprehensive Capital Analysis and Review (CCAR) and Dodd-Frank Act (DFAST) stress testing programs), or during onsite examinations, as well as to the underlying bank conditions, as measured by supervisory ratings. Inclusion in the stress testing programs is associated with a significantly higher number of issues, above and beyond the difference associated with asset size, consistent with the increased attention of banking authorities to these institutions (section 165 of the 2010 Dodd-Frank Act).

When we examine the issue categories, we find that bank stress (as measured by supervisory ratings) is correlated with more issue openings in all categories except those relating to risk modeling (e.g. stress testing). Moreover, deteriorating bank health as measured by banks' observables is mostly associated with issues related to hard information (loan portfolio and capital & liquidity). Finally, while BHCs included in publicly disclosed supervisory assessments have more issue openings in all categories, over 70% these are concentrated in issues related to risk modeling and internal controls. These types of issues, however, do not tend to be associated with more severe supervisory actions. We also find that improvements in asset quality and capital ratios are associated with closures of loan portfolio and capital & liquidity issues, which are directly related to those bank characteristics.

In terms of more severe supervisory actions, we study two measures: changes in supervisory ratings, which provide a comprehensive measure of a firm's performance; and formal actions, which as opposed to MRAs and MRIAs, are publicly disclosed and have force of law. We find that increases in the number of outstanding issues are associated with rating downgrades, specifically within topics related to loan portfolio and capital and liquidity. Increase in the stock of issues outstanding are associated with more formal actions, mostly with respect to issues related to compliance & regulation and capital & liquidity. While all issues deal with a combination of soft and hard information, our estimates suggest that changes in supervisory ratings are mostly related to MRAs and MRIAs involving hard information, as defined above. Formal actions, instead, are associated with MRAs and MRIAs related to both.

Since a reduction in the stock of issues is associated with an improvement in supervisory ratings, we examine what correlates with the speed of issue closure. We find substantial variation across issue types, with risk modeling issues closing in over 500 days on average, and compliance & regulation issues closing in almost half that time. We also find the time to closure is shorter at stressed banks, but is slower at banks with a higher stock of outstanding issues, even when controlling for firm fixed effects. This result suggests there could be capacity constraints in banks' ability to close issues. But because we do not observe exogenous variation in the stock of issues, this evidence may instead result from banks being in worse financial

strength when they have more issues outstanding, although the inclusion of a large set of observables should help account for this alternative explanation.

Finally, turning to the last Basel pillar of market monitoring, we combine textual data from analyst questions on large BHCs' earnings calls with the LDA model parameters estimated on Fed supervisory issues. We first find that the share of questions asked by market analysts about risk modeling, loan portfolio and capital & liquidity (i.e. mostly related to hard information) are predicted by firm balance sheet information and stress test program participation. The share of questions raised by analysts on risk modeling, loan portfolio and capital & liquidity are also correlated with the number of issues raised by Fed supervisors in those topics.

Related literature In the theoretical literature, a few papers have explicitly focused on supervision, as opposed to regulation (e.g. Dewatripont and Tirole, 1994; Prescott, 2004; Rochet, 2007; Harris and Raviv, 2012). Motivated by the inclusion of supervision as a pillar in the Basel framework, a few papers study the interaction of supervision and regulation, such as Bhattacharya et al. (2002) and Decamps et al. (2004). In the empirical literature, papers focusing on the supervisory pillar that rely on supervisory ratings include Cole and Gunther (1995); Hirtle and Lopez (1999) and Agarwal, Lucca, Seru, and Trebbi (2014). Berger, Davies, and Flannery (1998), Cargill (1989) and Hirschhorn (1998) focus on the interaction between bank supervisory ratings and market pricing information, and Berger and Davies (1998) find evidence of interaction between regulatory discipline and supervision.

The text used in the earning call analysis has been analyzed in terms of its informativeness vis-a-vis announcement data using tone/sentiment analysis in the spirit of Tetlock (2007) and Loughran and McDonald (2011). Many of these papers study the content of the press release rather than of the earnings call. Frankel, Mayew, and Sun (2010) test whether investor relations costs are one possible incentive for managers to avoid small negative earnings surprises. This paper provides the first view of supervision using the content of supervisory issues raised to banks, as opposed to summary measures such as supervisory ratings or the frequency of public enforcement actions (Lucca, Seru, and Trebbi, 2014). This data provides a much broader and in-depth view of the content of supervision in the context of the remaining Basel pillars.

The rest of the paper is organized as follows. Section 2 discusses information on supervisory issues and the data used in the analysis. Section 3 discusses the linguistic methodology. Section 4 presents results on supervisory data, while Section 5 compares market monitors concerns to supervisory issues. Finally, Section 6 concludes.

2 Bank Supervision, supervisory issues and data

2.1 Bank supervision and supervisory issues

Federal Reserve bank supervision. The Federal Reserve is responsible for supervising all U.S. bank holding companies (BHCs), Financial Stability Oversight Council-designated systemically important financial institutions, state-chartered commercial banks that are members of the Federal Reserve System (SMBs) and the U.S. operations of foreign banking organizations. The analysis of this paper relies on information on domestic institutions only. Prudential supervisory responsibilities primarily revolve around the assessment of the financial health, or safety and soundness (S&S), of an institution. In addition, Fed supervisors are also tasked with assessing compliance with anti-money laundering, consumer protection legislation among other areas. In terms of S&S, the Fed coordinates with state supervisors, which also oversee SMBs as their chartering authority, through alternating examination schedules and other means (see Agarwal, Lucca, Seru, and Trebbi, 2014, for more detail). For BHCs, the main source of the Fed's supervisory authority is Section 5 of the Bank Holding Company Act, including subsequent amendments, such as those introduced with the 2010 DFA. BHC supervision is conducted on a consolidated basis and through coordination with other federal regulators that oversee depository institutions within bank holding companies.[2] We refer to Eisenbach, Haughwout, Hirtle, Kovner, Lucca, and Plosser (2015) for a detailed institutional description of supervisory activities at the Federal Reserve.

Banking supervision and supervisory actions. Prudential supervision is tightly linked with, but distinct from, regulation of banking organizations, which involves the development and promulgation of the rules under which BHCs and other banks operate. There are two main elements to bank supervision. The first is monitoring institutions to assess the safety and soundness (S&S) including the institutions' risk-management systems, financial condition, and compliance with applicable banking laws and regulations. A second, and equally important element, is the use of information uncovered through supervisory assessments to demand corrective actions to banks should their conditions or practices be deemed unsafe or unsound, or not to comply with the law. Compliance with law and regulations includes not only compliance to statutes and rules related to S&S but also to those related to anti-money laundering (BSA/AML) and consumer protection. While the focus of this paper lies in S&S, our data include all supervisory issues including those related to non S&S roles.[3]

[2]Other federal banking supervisors are the Federal Deposit Insurance Corporation (FDIC), the Office of the Comptroller of the Currency (OCC) and the National Credit Union Administration (NCUA).

[3]The Bank Secrecy Act (BSA) require federal banking agencies to review that banks (SMBs in the case of the Fed) comply with procedures mandated by the BSA to detect and prevent money laundering (12 USC 1818(s)). In terms of consumer protection, Fed supervisors are tasked to assess that SMBs comply with a number of federal

Corrective supervisory actions can be formal or informal. Formal supervisory actions are publicly disclosed by the Federal Reserve Board and include written agreements, cease and desist orders, and fines ("civil money penalties").[4] Formal supervisory actions have legal force, meaning that should the firm fail to meet the terms of the action, it can face fines and other actions, such as a requirement to restrict its growth or to divest certain assets. Informal actions are labeled as such because the Fed's authority to impose these actions is based on supervisory practice as described in various Supervision and Regulation (SR) Letters. In this paper we focus on the content and volume of informal actions known as matters requiring attention (MRA) and matters requiring immediate attention (MRIAs). While these actions are less severe than formal ones (or other informal actions, for example, memorandum of understandings, or MOUs), given their higher incidence, they provide as a whole a detailed description of the issues that are uncovered, and brought to banks' attention, by Fed examiners.

MRAs and MRIAs. In January 24, 2008, matters requiring attention (MRAs) and immediate attention (MRIAs) were introduced with Supervision and Regulation (SR) 08-1 (subsequently superseded by SR 13-13 on which we based this discussion) to improve the consistency and clarity of written communication of Federal Reserve supervisors to supervised entities about findings uncovered during supervisory activities to bring attention on the entities' deficiencies found during the supervisory process. Depending on the size of the organization, supervisory findings are communicated in writing in the "Matters Requiring Attention" section of S&S formal examination or inspection reports or during the annual "roll-up" of those reviews into a report for the largest banks. Reports are typically presented to the board of directors, which directs the organization's management to take corrective action.[5]

Safety-and-soundness (S&S), consumer compliance or BSA/AML examinations can give rise to MRA/MRIAs.[6] While our data include issues arising from each type of examinations, as we

consumer financial laws, including home mortgage disclosure (HMDA) and the Community Reinvestment Act (CRA). Following the 2010 DFA, these responsibilities are shared with the Consumer Financial Protection Bureau (CFPB) for SMBs with assets greater than $10 billion.

[4]Because of their public nature, Lucca, Seru, and Trebbi (2014) use the incidence of formal actions as a proxy for supervisory activity. Written agreements and certain cease and desist orders (referred to as "consent orders") are agreed between the Federal Reserve and an institution, and stipulate certain findings about the firm and its actions and specify a course of correction to address the findings. Cease and desist orders can also be imposed without the agreement of the firm. 4(m) agreements (or 4Ms, named after the corresponding section of the BHC act) are another type of formal action, which are not publicly disclosed and may be issued when a BHC is either engaged in non-permissible activities, or when the holding company or one of its depository institution subsidiaries is either inadequately capitalized or not well managed.

[5] While examiners may provide bank management feedback on matters that are informative or advisory in nature that address the range of acceptable practices, these do not constitute MRAs or MRIAs. Only when examiners expect the banking organization to take action to address practices in a particular area or business function are those matters treated as MRIAs or MRAs. MRAs/MRIAs include all significant supervisory issues.

[6]Consumer-related issues are presented to banks in the consumer compliance examination reports to the banks.

discussed below in Section 4, most of these and much of our focus, centers around S&S. MRAs and MRIAs specify the particular concern being raised as well as a timeframe by which the firm must remediate the deficiency. Firms receiving MRAs or MRIAs will typically develop a plan for remediating the shortcomings being raised; the supervisory team then reviews the plans and is responsible for following up to ensure that the firm has followed it. This follow-up can take the form of a subsequent examination or regular or enhanced continuous monitoring. In general terms supervisory guidance requires that communication of supervisory findings must be: written in clear and concise language; (2) prioritized based upon degree of importance; and (3) focused on any significant matters that require attention.

The key distinction between MRIAs and MRAs is the nature and severity of matters requiring corrective action, as well as the immediacy with which the banking organization must begin and complete corrective actions. MRIAs are matters of significant importance and urgency that the Fed requires banking organizations to address immediately. The expected timeframe to address MRIAs is generally short, and may be "immediate," when safety-and-soundness risk is heightened. MRAs are matters that are important and that the Fed is expecting a banking organization to address over a reasonable period of time. Issues giving rise to MRAs pose a threat to S&S of the organization that is less immediate than those giving rise to MRIAs.[7]

A matter requiring attention remains an open issue until resolution and Fed examiners confirm the banking organization's corrective actions. When follow-up indicates that the firm failed to address sufficiently the deficiency identified in the MRA or MRIA, MRAs can be escalated into MRIAs, and for both types of issues additional formal or informal investigation or formal enforcement action (issued by the Board of Governors) can be issued against the organization. As it will be clear from the descriptive statistics below, it is typical for a relatively large banking organization to have many outstanding MRAs and MRIAs at any given time, reflecting the outcomes of the range of supervisory activities undertaken by the firm-focused supervisory team and other Federal Reserve supervisory staff. We discuss the types and content of supervisory issues after outlining the computational linguistic methodology.

2.2 Data

We use three sources of regulatory and supervisory data: financial data from regulatory filings of domestic bank holding companies (BHCs) and state member banks (SMBs); confidential ex-

[7] Issues classified as MRIAs include "[...](1) matters that have the potential to pose significant risk to the safety and soundness of the banking organization; (2) matters that represent significant noncompliance with applicable laws or regulations; (3) repeat criticisms that have escalated in importance due to insufficient attention or inaction by the banking organization[...]" (SR 13-13). MRIAs are communicated to the board of directors with the standardized language: "The board of directors (or executive-level committee of the board), or banking organization is required to immediately [...]"; MRAs are, instead, introduced as: "The board of directors (or executive-level committee of the board), or banking organization is required to [...]"

amination information and regulatory ratings from the National Examination Database (NED); confidential information on all issues (MRAs and MRIAs) raised by Fed examiners to both types of banking organizations. In our analysis of market monitors (Section 5), we use transcripts of quarterly earnings call from Lexis Nexis. We review the earnings call data in that section and focus on detailing the supervisory data here.

We use quarterly public regulatory filings of banking organizations to measure bank size, regulatory capital and asset quality. Data for BHCs are from Y9-C forms, which report balance sheet information on a consolidated basis, while data for SMBs are from Call Reports. In our sample period (2009-14) the minimum asset requirement for Y-9C reporting is $500 million.

Summary statistics of the key measures in the paper are reported in Table 3. ROA is expressed as the annualized percent ratio of net income to assets; Tier 1 capital is the ratio of Tier 1 capital to risk weighted assets; Log Assets is the logarithm of total assets expressed in $ thousands; NPL ratio is the percentage ratio of non-performing loans (30+ delinquent and non-accruals) to total loans; Charge-offs are the annualized percentage fraction of charge-offs net of recoveries to total loans. Summary statistics are split by whether banking organizations are SMBs, smaller BHCs or the 32 BHCs that have ever been part of a Fed supervisory stress test programs, the 2009 Supervisory Capital Assessment Program (SCAP) or any of the yearly Comprehensive Capital Analysis and Reviews (CCAR) starting in 2011. We refer to these banks as "CCAR" banks. We split these BHCs because in our sample period they have been subject to enhanced supervisory programs, such as stress testing (see SR 12-17 for additional detail).[8]

For supervisory rating, we use the so-called RFI/C(D) for BHCs and CAMELS for SMBs.[9] Both rating systems assign a 1-to-5 score with lower numbers indicating fewer issues, i.e. a better rating. Banks with a rating of 1 or 2 are considered in satisfactory condition and present few significant supervisory concerns. Banks with a 3, 4, or 5 rating present moderate to extreme levels of supervisory concerns. In the analysis we construct a dummy variable "stressed rating", which is equal to one when the supervisory rating is equal to 3, 4 or 5, and also control supervisory ratings both in levels and by dummying out each level. A banking organization

[8]The number of institutions included in the CCAR program has changed over time. In 2011, CCAR included the same 19 BHCs that participated in the 2009 SCAP. In the Fall of 2011 the set of institutions was expanded to all U.S. BHCs with total consolidated assets of $50 billion or more. Over a phasing-in period, these additional institutions were assessed under a different capital plan review program. In 2014, 30 BHCs with total consolidated assets greater than $50 billion were part of CCAR.

[9]For BHCs, 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 (i.e. the overall rating considering and weighing the ratings on "R", "F" and "I"), and "D" is the rating assigned to the depositories (e.g. commercial banks or thrifts) owned by the holding company. Similarly for SMBs, the letters in CAMELS stand for the subcomponents that are evaluated by examiners: capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to market risk.

rating is assigned yearly (18-months if an SMB has assets less than $500 million and a non-stressed rating) following a full-scope examination, meaning that each rating subcomponent is being evaluated, or during a "roll-up" period for the largest BHCs, which are monitored under "continuous monitoring," meaning with examination staff continuously assigned to the banking organization. Because issues are more likely to be raised in the context of examinations, we construct an "ongoing examination" dummy variable that indicates the time between the quarter of the start of a full-scope examination (or roll-up) and its termination. In addition to studying stress we also look at FDIC resolutions for SMBs, although only briefly as very few such events take place in our sample period. Resolutions are measured with a dummy variable that indicates if such event takes place in the next four quarters.

While rating and financial information is available from the early 1990s, data on supervisory issues (MRAs and MRIAs) starts in 2009. As noted above SR 08-01 introduced MRAs and MRIAs as formal categories but data in 2008 are not as complete. We access two separate databases containing supervisory issue information, the first containing information for so-called LISCC institutions, and the second for all remaining BHCs as well as all SMBs. LISCC stands for the Fed's Large Institution Supervision Coordinating Committee, which oversees the supervision of the largest, most systemically important financial institutions. There are currently sixteen such institutions including large bank holding companies, U.S. operations of certain foreign banking organizations, and nonbank financial companies that are designated by the Financial Stability Oversight Council. Our sample includes the subset of domestic holding companies within the 16 LISCC firms (see Eisenbach et al., 2015, for more detail). The issue database contains detailed data on all MRAs and MRIAs that have been issued by Fed examiners including a description of the issue (average of about 150 words), the organization to which the issue was raised, the date when issue was communicated to the banking organization and the when and if the issue was resolved, meaning that the bank conduced satisfactory remedial action. The next section discusses a methodology to classify issues based on the co-occurrence of words in the description of the issue.

3 Linguistic Methodology

Supervisory issues vary in their context and scope, reflecting the breadth of supervisory responsibilities in the context of S&S and other goals, such as BSA/AML and consumer protection. A key feature of the data on supervisory issues is in their textual description. While such information is valuable, it poses obvious quantification challenges in an econometric analysis. In this section we discuss our method for categorizing the textual information into a set of topics.

A researcher attempting to characterize banking supervision through MRAs and MRIAs faces two main obstacles. First, these texts are confidential supervisory information making

them either unaccessible to researchers, or, as for this paper, not describable with any detail to safeguard confidentiality. But even abstracting from these confidentiality issues, thousands of MRAs and MRIAs have been issued to banking organization in our sample period requiring a more formal statistical approach.

Our goal is to use the text of an issue description to label each MRA or MRIA with a salient and meaningful topic, or category, and do so in a structured and disciplined fashion. We use an unsupervised computational linguistic method known as Latent Dirichlet Allocation (LDA) to identify topic weightings for each issue. LDA is a machine learning algorithm first described by Blei, Ng, and Jordan (2003) that uses hierarchical Bayesian modeling to identify underlying latent structures over large datasets containing text. Crucially, this approach places limited *ex ante* restrictions on the topic labels and instead uses the distribution of words to identify different topics. At a high level, LDA can be understood as a structural model that uses both parametric assumptions and large amounts of data to infer latent characteristics of the data generating process. LDA requires some basic assumptions on the data generating process in order to estimate the underlying topics. Our first assumption is a common assumption known as "bag-of-words," meaning that within a document, the order of the words is irrelevant, and so can be represented as a vector of word counts.[10] Second, we will impose standard text preprocessing on our dataset.[11] Next, a document is assumed to be characterized by two sets of distributions: first, the unobserved topic weight for a document is assumed to be drawn from a Dirichlet distribution; then, each topic has an unobserved multinomial distribution over the vocabulary of words. These two sets of distributions give the data generating process its hierarchical nature. Formally, given a document d, the following data generating process for a given issue is assumed:

I. Draw length of document $N \sim \text{Poisson}(\xi)$.

II. Choose the topic weights $\theta_d \sim \text{Dirichlet}(\alpha)$.

III. For each of the N words w_n,

   A. Draw a topic $z_n \sim \text{Multinomial}(\theta_d)$.

   B. Draw a word from $p(w_n|\beta, z_n)$, a multinomial distribution conditioned on topic z_n.

The model contains several structural elements. First, β is a $k \times V$ matrix, where k is the (fixed) number of underlying topics, and V is the size of the vocabulary (the number of distinct

[10] This is a common assumption in computational linguistics, and methods still typically do quite well despite discarding relevant information related to the ordering of words.

[11] This includes removing words that occur only once across all documents, removing "stop words" which reflect very common words like "and" and "the," and lemmatizing words. Lemmatization involves normalizing words such that "bank" and "banks" are viewed as identical.

words). Each row in β defines the probability of drawing a word given a topic. The overall β matrix is assumed to be fixed across all documents. Next, θ_d is a $k \times 1$ vector defined for each document and gives the topic weighting for a given document. A prior α is prespecified over the distribution of θ_d, and ξ is a scalar that defines separately the average length of an issue. Each document is assumed to be independent, and within a document, words are assumed to be exchangeable.

When applied a full set of documents of number D, the parameters of interest will be β (a $k \times V$ matrix) and θ (a $k \times D$ matrix). Practically speaking, estimation of these parameters can be implemented using straightforward Python packages. However, there are two important caveats in implementation: first, the topics are not given semantically meaningful labels by the algorithm; and second, the number of topics is prespecified by the researcher.

Labeling topics. The naive approach to topic labeling would be to examine words in a given topic that have large values of β. However, this method is flawed due to the fact that words like "risk" are prone to occur frequently across all supervisory issues addressed to banks. Instead, we expand on a method focusing on "salient" words in a given topic, words which are particularly informative about a given topic's probability. This method builds on Chuang, Manning, and Heer (2012) and focuses on saliency as measured by:

$$\text{saliency}(w|z) = p(w) \times \text{ distinctiveness}(w|z) \tag{3.1}$$

where distinctiveness$(w|z)$ is the pointwise contribution to the Kullback-Leibler divergence:

$$\text{distinctiveness}(w|z) = p(z|w) \log \frac{p(z|w)}{p(z)}. \tag{3.2}$$

Note that $p(z|w)$ is not a direct output of the model, and instead requires a transformation of the given model parameters.[12] Distinctiveness$(w|z)$ weights words that, when observed, substantially increase the probability of a given topic, when compared the marginal distribution of the topic. In other words, given a word like "risk", how much can someone easily identify one topic over another? Saliency$(w|z)$ weights distinctiveness by word frequency in order to avoid overemphasizing infrequent words that may by chance be extremely distinctive.

Choosing the number of topics. Next, to select the number of topics we use a novel methodology described in Goldsmith-Pinkham and Lucca (2016) and outlined in Appendix A. Briefly, this methodology trades off identifying topics that have sufficiently salient words and ensures that each topic is "important" in explaining content. This ensures that we identify meaningful topics that label a substantial fraction of the overall issues. In our application, we find five

[12] Note that by Bayes' Rule, $p(z|w) = p(w|z)p(z)/p(w)$. $p(z)$ is a straightforward summation over θ and $p(w|z)$ is exactly β. $p(w)$ is the empirical marginal distribution of each word.

topics to be optimal in trading off between these two criterion.

Model output. Our model generates two sets of estimated outputs: a set of parameter values identifying the distribution over the vocabulary for each topic, and the topic weights for each issue. For each issue, a 5×1 vector of topic weights π is defined such that $\sum_i \pi_i = 1$.

Bank characteristics are observed at the quarterly level and we run most our analysis at this level of aggregation. We use issue-open dates and regulatory entity identifiers associated to each issue j to group them by bank i and quarter t. The total number of issues, irrespective or topic, opened in a given quarter is then:

$$\text{NumIssuesOpen}_{it} = \sum_{j \in J_{it}} 1(\text{Issue}_j), \tag{3.3}$$

where with J_{it} is the set of all issues j opened in quarter t for bank i. For each topic k obtained from LDA, we construct

$$\text{NumIssuesOpenTopic}_{ikt} = \sum_{j \in J_{it}} 1(\text{Issue}_j)\, \hat{\theta}_{jk} \tag{3.4}$$

where $\hat{\theta}_{j,k}$ is the estimated LDA weight for topic k and issue j. The number of opened issues, or for a given topic, is set to zero if no issues were raised in that quarter for that bank.[13] We define the total number of issue closings for bank, $\text{NumIssuesClose}_{ijt}$ accordingly using information on the dates in which the issue was closed.

4 Supervisory issues

We first present results of the linguistic analysis and overview descriptive statistics. We then study association of openings and resolutions of issues in terms of bank characteristics, as well as the implications of issue openings and closings for supervisory rating and issuance of formal supervisory actions. This analysis is run at the bank level and use aggregated issue level data. We finally study time-to-close at the issue level.

4.1 Topic description and summary statistics

Topics. We start discussing the estimation results of the topic selection criteria discussed in Section 3. Figure 1 presents a graphical summary of the parameter estimates for the eight most salient words in each topic. We focus on presenting the most salient words, rather than the most probable words in a given topic, because some words are highly probable across all topics, and are not particularly informative about the semantic content in a topic. This fact is

[13] In the analysis below we will also make use of shares that are constructed as $\text{ShareOpenTopic}_{ikt} = \text{NumIssuesOpenTopic}_{ikt} / \sum_l \text{NumIssuesOpenTopic}_{ilt}$, where $\sum_k \text{ShareOpenTopic}_{ikt} = 1$.

quite apparent in Figure 1. The eight most salient words for each topic are shown in the left panel, and the width of each circle reflects $p(w|z)$, the probability of each word given the topic. The topic numberings are shown on the top axis. Words such as "bank" and "risk" are highly probable across many topics, while words like "regulation" and "audit" are highly concentrated, or distinct, within a particular topic. This fact is evident from the right panel, which shows the marginal probability of each word. Reflecting the high loading of each topic on words like "risk," their marginals are high, which, along with a word's distinctiveness, determines the saliency of a word (equation 3.1).

As we examine the columns in Figure 1, we start to identify semantic labels for each topic. Topic 1 has words such as regulation, compliance and violation, and so we label it "Compliance & Regulations." Topic 2 appears to relate to internal processes – audit, internal, IT and control – so we label this "Internal Controls." Topic 3 focuses on risk models and hence we label it "Risk Modeling." Topic 4 is clearly related to loan portfolios, focused specifically on credit risk, and we label it "Loan Portfolio." Finally, Topic 5 highlights capital and liquidity funding risks, and so we label it "Capital & Liquidity."

Table 1 provides more detail on the most common words in each topic. Since single words (unigrams, which appear in Figure 1 and form the basis of the LDA estimation) can have many meanings, it is useful to examine frequent pairs of words (bigrams) within topics. To do this, we identify bigram counts within each document and weight these counts by the topic weights produced by the model. We then sum up these weighted counts for each bigram and present the top five most frequent bigrams per topic, ignoring bigrams which are common across all topics.[14] In order to avoid issues of ordering, all bigrams are counted in alphabetical order – in presentation of the bigrams, we reorder the bigrams in order to improve readability. For Compliance & Regulation, we note that "due diligence" and "compliance regulation" are common bigrams and consistent with our labeling. For Internal Controls, we see that these are related to "internal audits", "information security" and "operational risk" – seemingly non-credit and liquidity risk-based issues. For Risk Modeling, we notice that the most important bigram is related to "stress testing", as well as "risk model" and "model validation". This suggests that a big component of this topic is related to CCAR and its predecessor SCAP. Loan Portfolio is highly related to "credit risk", "real estate" (an especially important source of credit risk over our sample period) and "ALLL methodology". ALLL is the "Allowance for Loan and Lease Losses", and its purpose is to reflect the estimated credit losses within a bank's portfolio of loans and leases. This is an important element of prudential bank behavior, and hence a

[14] We ignore "management must", "management [is] required", "management [is] expected", "risk assessment", "board [of] director[s]" and "risk management", as they each show up in the top five for three out of the five topics, and hence are not informative for topic labeling.

component of supervision is to ensure proper ALLL behavior. Finally, Capital & Liquidity relates to "capital plan," "liquidity risk" and "contingency funding."

These topics are intuitive given the highly salient words and bigrams, but we next also perform a "sanity" check on the unsupervised categorization by comparing them to those applied by data users.[15] As seen in Table 2, approximately 60% of Compliance & Regulations are issues categorized by users as Bank Secrecy Act and Anti-Money Laundering (BSA / AML) and Consumer Compliance / Community Reinvestment Act (CRA) issues. In column 2, Internal Controls, we note that the majority of these issues are categorized by users as Information Technology (IT), Operational Risks, and Audit. Risk Modeling corresponds to several topics, but most importantly Market / Liquidity Risk and Credit Risk. Loan Portfolio is almost exclusively concentrated in Credit Risk issues, while Capital & Liquidity are concentrated in Market / Liquidity Risk and Risk Management. It is interesting that for several topics (such as Loan Portfolio and Compliance & Regulations), the topics line up quite tightly with the internal categories, while for other topics they split between categories set by users. For example, issues categorized as Market/Liquidity are split in the unsupervised method between capital & liquidity and risk modeling. While these results are comforting, unsupervised methods are based on statistical techniques, which will be associated with estimation error.

Descriptive statistics. In Table 3, we present basic summary statistics on supervisory issues. The first pair of columns present the pooled summary statistics for all banks in our sample, while second, third and fourth pairs of columns present the summary statistics for State Member Banks, Non-CCAR BHCs and CCAR BHCs, respectively.[16] Since a unit of observation here is a bank-quarter, we see that the average bank receives roughly an issue a quarter, with non-CCAR BHCs receiving 0.16 a quarter, SMBs receiving 1.42 a quarter, and CCAR BHCs receiving 6 issues a quarter. This appears to suggest a strong correlation with respect to size and institution type, and we will control for these characteristics in our regressions. We also see that almost all Risk Modeling issues are opened for CCAR BHCs, with less than 0.1 per quarter for SMBs

[15] Aside from being explicit about the classification criterion, there a number of reasons why one would perform an unsupervised classification rather than using pre-determined ones. First, the human classification is based on business needs, which may not be as relevant for economic analysis. Second, the categorization we employ here is for only a subset of all issues and exclude those for LISCC banks, which are classified under a different methodology. Third, the distribution of issues across the internal categories is highly skewed, with large chunks concentrated in one label (25 percent in Credit Risk) and lots of categories containing a small number of issues. As a result, some topics would be measured well, while others would be dispersed and poorly measured, creating difficulty when we look at bank characteristics related to openings. Fourth, while could recombine categories, the process would be arbitrary, and violate the spirit of our attempt to provide a data-driven categorization of the issues. Finally, the semantic content within the topics may be useful for labels whose meaning may not be as transparent ("Management / Risk Management").

[16] Recall that we define as CCAR the 32 BHCs that have ever been part of a Fed supervisory stress test programs, the 2009 Supervisory Capital Assessment Program (SCAP) or any of the yearly Comprehensive Capital Analysis and Reviews (CCAR) starting in 2011. See Section 2.2 for more details.

and 0.01 per quarter for Non-CCAR BHCs. Finally, the stock of outstanding issues is 4.85 for the average bank, while CCAR BHCs have 34.54 issues outstanding.

We can also examine the opening and stock of issues across time. In Figure 2, we present the total number of issues in each topic raised per quarter for SMB and BHCs. There are several notable features of these graphs. First, the occurrence of risk modeling is very low for SMBs, and spikes during the stress testing periods for BHCs. Second, Loan Portfolio issues dominate during the crisis and immediately afterwards for SMBs, while Capital and Liquidity issues are largest in 2009 for BHCs. Finally, excluding Risk Modeling, the number of issues opened for both types of banks peaked in early 2010, and has declined since then. In Figure 3, we plot the stock of outstanding issues of each topic, across the two bank types. Interestingly, while the number of openings over time has declined since 2010, the stock of issues has stayed relatively constant since then. The stock of Risk Modeling issues is by far the largest set of outstanding issues for BHCs.

4.2 Issue openings, closings and bank characteristics

Empirical specification. We study the association between bank characteristics and the total number of MRAs and MRIAs issued by Fed supervisors to bank organization i in quarter t, as well as the number of issues in a given topic. The first regression model specifications that we consider are

$$\text{NumIssuesOpen}_{it} = \alpha + \beta X_{it} + \gamma W_{it} + \epsilon_{it} \tag{4.1}$$

$$\text{NumIssuesOpenTopic}_{ikt} = \alpha_k + \beta_k X_{it} + \gamma_k W_{it} + \epsilon_{it}. \tag{4.2}$$

Openings at date t include those that start at any time during the quarter, while the set of characteristics X_{it} and W_{it} are measured as of the beginning of the quarter, so that they are known to bank and supervisors at the time in which an issue is open.

The set of baseline characteristics W_{it} includes the logarithm of bank assets, a dummy variable for the supervisory rating indicating stress (rating of 3,4 or 5) as well as either bank-type or bank-fixed effects. In addition, W_{it} also includes dummy variables for whether a full-scope examination is occuring in a given quarter interacted with the CCAR indicator. These interactions are meant to control for the variability of issues assigned to banks in the course of a year for CCAR and other banks. As discussed in Section 2.1, issues are often typically raised in the context of an examination, which occur about every 12 months for large BHCs and SMBs. The largest banking organizations are, instead, typically monitored on a continuous basis and so we separate out the effect for these banks.

We also include a set of bank observables X_{it}, which include measures of profitability (ROA),

asset quality (nonperforming loans and charge-offs) as well as regulatory capital levels (Tier 1 ratio). These controls enter equation 4.1 in levels. Because the opening and, in particular, the closing of an issue may depend on improvements or worsening in these characteristics we also consider the following specification for issue openings:

$$\begin{aligned}\text{NumIssuesOpen}_{it} &= \alpha + \beta X_{i,t-4} + \beta \Delta_{t-4,t} X_i + \gamma W_{i,t-4} + \epsilon_{it} \\ \text{NumIssuesOpenTopic}_{ikt} &= \alpha_k + \beta X_{i,t-4} + \beta \Delta_{t-4,t} X_i + \gamma_k W_{i,t-4} + \epsilon_{it},\end{aligned}$$

as well the analogous specification for closings, where $\Delta_{t-4,t} X_i$ reflects changes in banks observables in the four quarters to the opening of the issues. We use a four quarter change, which is the time between two full-scope examinations.

Before turning to the results, it is important to note that because the number of issues by topic always sum up to the total number (equation 3.4), issue loadings on the independent variables always add up to the corresponding coefficients in the total issue regression. For example, in terms of X_{it}, $\beta = \sum_k \beta_k$. This means that the β_ks not only provide information on the importance of a bank characteristics for an opening or closing of that particular issue topic, but also help us decompose each characteristic's effect across the different topics.

Results for issue openings. We first study the determinants of issue openings on basic bank characteristics when excluding bank fixed effects (Table 4, panel A). A bank with a stressed rating at the beginning of the quarter, receives 0.35 more issues (column 1) than banks with a non-stressed rating. This effect is substantial as banking organizations in our sample receive about .9 issues in a given quarter (unconditional means are reported at the bottom of the table and in Table 3). Looking at opening by issue type, we see that much of these openings are related to loan portfolio and capital & liquidity issues, which as a whole are about 0.25 higher for a stressed bank. That said banks under stress receive more issues for all topics with the exception of risk modeling, which as discussed above is related to stress testing.

The table also includes a dummy variable measuring whether an FDIC resolution takes place in the next four quarters. While this event is particularly meaningful for an analysis of banking supervision, it is important to remember that it only occurs with an annualized probability of 22 basis points in our sample (Table 3). Banks leading to failure tend to receive a lower number of new issues even accounting for the fact that these banks have a stress rating, as measured by netting the effect of the two dummies. Looking at topics, these institutions tend to have fewer issues related to internal controls and compliance & regulation. Instead, issue levels for capital & liquidity are as high as for stressed banks.

Larger banks tend to receive more issues. This effect is order of magnitudes smaller than the effect of stress when thinking of a given bank. For example, doubling a bank's size increases

quarterly issue openings by about 0.08, or 10% of the unconditional mean. That said there will be significant differences driven by size given the width of the support in the asset distribution. In terms of issue topic, while capital & liquidity issues were the most important for stressed banks, larger banks aren't more likely to receive these types of issues.

Our sample includes SMBs and BHCs, including those that underwent a public supervisory assessment at some point in the sample (either CCAR or the 2009 SCAP). The regressions includes SMB and CCAR bank-effects so that the left-out category is non-CCAR BHCs. The coefficient on CCAR BHCs, indicates that, controlling for the logarithm of assets, these banks receive about 5.5 additional issues per-quarter, with about 2.5 (1.4) additional risk modeling (internal control) issues. The additional issues to CCAR banks may come about because the relation between issue counts and size is non-linear in log assets, or because CCAR banks do in fact receive more issues associated with increased attention from supervisory assessments or the post-DFA enhanced focus on these institutions. The fact that much of the increase in issues is related to risk modeling would point to increased attention.

To tell these two alternatives apart more clearly, we consider a more flexible size specification by bucketing average issues by size categories around $50 billion, which as discussed above is, to a first-order, the threshold for BHCs being included in supervisory assessments in the sample period. As shown Figure 6, there is a discrete jump in the quarterly volume of new issues from about 1 to about 4 when moving from the [25,50) to the [50,75] asset categories. Instead, we see limited difference in size between banks in the (10,25] and [25,50) buckets, and the [50,75] and >=75 ones. This evidence suggests that the CCAR dummy in the regression is picking up an effect related to a BHC being included in the supervisory assessment rather a globally non-linear relation between issues and assets. In fact, risk modeling issues, which are related to stress testing, account for about 50% of all issues for CCAR banks as opposed to less than ten percent for other BHCs, and less than five percent for SMBs (Figure 5).

In terms of SMB versus non-CCAR BHC differential, SMBs receive about one additional issue per-quarter, which may be related to differences between consolidated holding company supervision with the effect not concentrated in any of the different issues, as evidenced by the similarity of the coefficients on the SMB dummy across rows. Finally, in terms of the ongoing examination dummy, it is indeed the case that the number of supervisory issues spike during an examination for banks that are not being continuously monitored (as proxied by non-CCAR), with the number of issues increasing by about 1.5 for banks that are under an exam. Taking the CCAR and ongoing full scope examination together suggest that, in addition to bank risk, key determinants of new issue opening include the extent of supervisory attention. While these results are intuitive, the point estimates help assess the relative importance of these factors.

Panel B of Table 4 reports estimates when dropping institution-type effect and instead

including bank fixed effects. We include bank fixed effects in all specifications discussed below. The inclusion of these controls does not significantly affect the remaining coefficients, with estimated effects of stressed rating and ongoing examination as discussed above. The within-bank asset variation in size is smaller than the between, and locally the effect of size is now estimated to be larger. Results on failures are also not significant in the within-specification, and since that they only affect a handful of institutions, we will not include them in the analysis below.

Table 5 extends the baseline specification of issue opening by including a number of bank characteristics. In including these additional measures, we attempt to understand to what extent opening of new issues are related to bank observables. To control for possible interactions between the stock of current outstanding issues and newly opened ones, each regression also includes the stock of issues in each topic (all five coefficients are omitted from the table). As shown in column 1, controlling for rating and size, total number of issues increase with a decline in Tier 1 capital and ROA, as well as a worsening of the loan book, as measured by increasing NPLs and charge-offs. All of these measures are expressed in percent, and looking at the (pooled) standard deviations of these measures (Table 3), the economic effects are comparable in magnitude. A standard deviation increase in NPLs, and a decline in Tier 1 ratio, are both associated with an increase in issue openings of about 0.1, or 15 percent of the unconditional mean. Looking at issue openings by topic, all bank observables are associated with issues related to loan portfolio and capital & liquidity, while bank observables are not associated with openings of other issues (with a handful of exceptions).

Beginning-of-quarter characteristics are the result of past quarter levels as well as recent changes. We next study how much openings are driven by the historical levels of characteristics versus recent changes. We address this question in Table 6, which shows separate loadings of issue openings on $t-4$ levels, X_{it-4}, and four quarter changes $\Delta_{t-4,t}X_i$. Estimated loadings are similar on past levels and four-quarter changes for Tier 1 capital, NPL and charge-off ratios, meaning that only the most recent levels matter. In terms of ROA, we see that the effect on changes is larger and more precisely estimated. In sum, the association of openings with bank observables is generally driven only by recent levels, with the exception of profitability for recent dynamics appear to matter more.

Results for issue closings. We discuss the association between issue closings and bank characteristics. As discussed in Section 2.1, issues remain open until resolution, and when follow-up indicates failure to address issues, supervisory actions are escalated. How banks respond to supervisory issues is therefore a key component of the supervisory process. Table 7 presents regression results for the total volume of closings over four quarters on $t-4$ bank characteristics and the change in these characteristic over the period. We focus on a year, because this

is the time frame between two examinations, and leave a detailed analysis to time-to-close to Section 4.4. As for the regression on openings, each regression includes the stock of issues in each topic (all five coefficients are omitted from the table). Among others, these variables control for the mechanical effect of stock outstanding on closing volumes. For similar reasons, we also control for the volume of issue opening over the period, as an increased opening flow may be associated to more closures. In terms of levels, large and stressed banks, which are more likely to be assigned the more-urgent MRIAs, are more likely to close issues. In addition, better capitalized banks (as measured by Tier 1 capital ratios) are also more likely to resolve issues. Looking at issue type, stressed banks are more likely to close all issues, although only the effect on capital & liquidity issues is statistically significant. The effect on size is mostly driven by loan portfolio and capital & liquidity issues, while better capitalized banks are more likely to close all type of issues, with the exception of risk modeling concerns.

In terms of changes, increases in bank assets are associated with increased closings, especially for loan portfolio and capital & liquidity issues. Intensified supervisory attention, as measured by more issues, can place constraints in the ability to expand assets. One interpretation for the association of bank size and closures is that as banks address their issues, those constraints may be lifted. While our ability to identify causality in this contemporaneous change is limited due to omitted variables, it is useful to keep in mind that the regressions control for contemporaneous improvements in bank profitability (and asset quality), which should not drive the contemporaneous association between size and closures. In terms of asset quality, the effect of non-performing loans is mainly in closures of loan portfolio issues, which, up to the aforementioned identification issues, may indicate that as banks improve the quality of their loan portfolio, the issues are subsequently closed. Finally, we see than an improvement in capital levels is associated with closures of issues related to capital & liquidity and internal controls.

Summary. Summing up the results, both supervisory attention (as measured by inclusion in supervisory assessment programs or undergoing a yearly full-scope examination) as well as bank stress (as measured by a bank's supervisory rating) drive the dynamic of issues. Bank observables are mostly associated with issues related to loan portfolio and capital & liquidity.

4.3 Ratings and formal supervisory actions

As discussed in Section 2.1, when follow-up on MRAs and MRIAs indicates that banks have not resolved the issues that were raised, Fed supervisors can issue so-called formal actions as well as downgrade the bank's supervisory rating. Institutions with worse supervisory ratings (i.e. higher) are subject to increased levels of supervisory attention, and moreover, these ratings are the key input for a number of policy outcomes ranging from the cost of the Federal Re-

serve's discount window to approvals on merger and branching decisions. Formal enforcement actions are posted publicly and are based on statutory authority as opposed to being based on supervisory practice as it is the case for MRAs and MRIAs.

As a result, one would expect both rating changes and the volume of supervisory issues to be related to the the stock and volume of new issues. We investigate here to what extent this may the case and for what issue topic in particular. We first construct a set of measures $\text{StockNumIssues}_{it}$ and $\{\text{StockNumIssues}_{ikt}\}_{k=1}^{5}$, which are the stock of outstanding open issues raised by the supervisors. We then estimate

$$\Delta_{t,t+4}\text{Rating}_i = \alpha + \beta\Delta_{t,t+4}\text{StockNumIssues}_i + \gamma X_{i,t} + \tau W_{i,t} + \epsilon_t \tag{4.3}$$

$$\text{Formal Action}_{t,t+4} = \alpha + \beta\Delta_{t,t+4}\text{StockNumIssues}_i + \gamma X_{i,t} + \tau W_{i,t} + \epsilon_t \tag{4.4}$$

where $\Delta_{t,t+4}\text{Rating}_i = \text{Rating}_{t+4} - \text{Rating}_t$ is change in supervisory rating over the next four quarters and $\text{Formal Action}_{t,t+4} = 1(\sum_{t=0}^{t+4}\text{Formal} > 0)$, which is an indicator for any formal actions being issued in the next year. The variable $\Delta_{t,t+4}\text{StockNumIssues}_i = \text{StockNumIssues}_{i,t+4} - \text{StockNumIssues}_{it}$ is the change in outstanding issues over the same annual period, and $X_{i,t}$ and $W_{i,t}$ include bank characteristics and regulatory controls such as the time t stock of issues, log(assets) and date/firm fixed effects. Importantly, we also include controls for each of the date-t rating levels to account for possible non-linear effect of changes on the starting rating, for example, because a bank with a starting rating of one cannot be upgraded.

In Table 8, we present estimates of Equation 4.3 and 4.4. An increase in the stock of issues corresponds to a future increase (downgrade) of about 0.01, or approximately 20% of the unconditional mean (column 1). Similarly, an increase in the stock of issues raises the probability of a formal action in the next year by 0.3 percentage points, approximately 10% of the unconditional mean. When we examine the specific topics related to changes in the rating, we see that issues related to loan portfolio and capital & liquidity are strongly correlated with rating, while the effect of changes in the stock of issues related to risk modeling, internal controls and compliance & regulations is both small and insignificant. This is both notable and perhaps to be expected – these ratings are most concerned with the factors at the bank related to capital adequacy and asset quality. Interestingly, we found above that issue openings for loan portfolio and capital & liquidity were the only ones that could be consistently matched to changes in bank observables. In terms of formal actions, changes in issues related to capital & liquidity as well as compliance & regulations have the largest effect on the probability of receiving a formal action. The effect of issues related to compliance & regulations may be driven by consumer compliance and BSA/AML related ones, as shown in Table 2. According to 12 USC 1818(s), for example, when violations of BSA/AML are not addressed supervisors are

required to issue formal actions (cease and desist orders). Regulatory capital is the main pillar to S&S, and consistently we find that formal actions are more likely to be issued when the stock of these issues increases. Notably, issues related to risk modeling and internal controls are not correlated with either type of additional supervisory action. This suggests one of two potential stories: these issues are addressed promptly and directly, such that additional regulatory responses are unnecessary; or these issues are viewed by supervisors as less time sensitive or important to the supervision mandate.

In Table 9, we rerun the same specification, but focus on whether it is the openings of new issues or closing of old issues that drives the rating changes and formal actions. Note that since $\Delta_{t,t+4}\text{StockNumIssues}_i = \text{StockNumIssues}_{t+4} - \text{StockNumIssues}_t = \text{Openings}_{t,t+4} + \text{Closings}_{t,t+4}$, a natural specification to run is the analog to Equation 4.3:

$$\Delta_{t,t+4}\text{Rating}_i = \alpha + \beta_1 \text{Openings}_{t,t+4} + \beta_2 \text{Closings}_{t,t+4} + \gamma X_{i,t} + \tau W_{i,t} + \epsilon_t \tag{4.5}$$

$$\text{Formal Action}_{t,t+4} = \alpha + \beta_1 \text{Openings}_{t,t+4} + \beta_2 \text{Closings}_{t,t+4} + \gamma X_{i,t} + \tau W_{i,t} + \epsilon_t. \tag{4.6}$$

As before we focus on a year-horizon, which the full-scope inspection frequency, but the time to close for a given issue can vary dramatically, as we will discuss in Section 4.4. As a result, we may expect to find less of an effect in closures as many closures take over a year.

We note that in Column 1 of Table 9, opening of issues is positively associated with a supervisory rating downgrade (increase) while closing of issues is associated with an improvement in rating. When we examine the specific topics, we again see that issue openings related to loan portfolio and capital & liquidity are strongly related with a deterioration in rating, while closure of these issues is not correlated with rating changes. However, closure of issues related to internal controls appears to be related with improvements in ratings.

For formal actions in Columns 3 of Table 9, we see that closures of issues are not correlated with the probability of formal actions, while openings have a strong positive effect – an issue opening is associated with a 1.3 percentage point increase in the probability of a formal action, which is over 25 percent of the unconditional mean. Moreover, we see that openings in compliance & regulations and capital & liquidity issues lead to more formal actions, while closures are not.

4.4 Time to Issue Closure

As discussed in the previous subsection, closing issues is associated with an improvement in a bank's supervisory rating. This suggests that banks have an incentive to resolve outstanding issues. But there are many potential challenges in closing issues. For example, supervisory

demands may be difficult or complex, or the bank may not be in a sufficiently strong financial position to address even less complex issues. Figure 8 presents the unconditional average time it takes to close an issue by topic type. We see that the average closure time for issues related to risk modeling is about 500 days, while issues related to compliance & regulation take under 270 days. These differences, however, may be driven by bank, rather than issue, heterogeneity, and so we reexamine this association within a regression framework that includes key controls identified in the analysis of issue openings.

We estimate regressions at the issue level, with the log of an issue's time-to-closure as the left-hand-side. For issues that have not yet been closed, we are faced with a right-censoring issue that we address in a few ways. Traditional methods for dealing with this issue are to run hazard models that incorporate the right-censoring issue. We report standard OLS results due to the fact that we are interested in including bank fixed effects within our specification, and non-linear hazard models cannot consistently estimate these effects (and other parameters) because of the incidental parameter problem. In unreported results, we use a semi-parametric Cox hazard model to estimate the effect of issue and bank characteristics on the time to closure, and find quantitatively very similar results to those reported below. [17]

Table 10 reports parameter estimates of the regression of the opening of an issue j at bank i in time period t:

$$\log \text{TimeToClose}_{ijt} = \alpha + \beta X_{ijt} + \tau W_{it} + \gamma \log \text{StockIssues}_{it} + \epsilon_{ijt}, \tag{4.7}$$

on issue characteristics, X_{ijt}, bank and regulatory characteristics, W_{it} and the log of the number of currently open issues at the bank at time of opening, log StockIssues. In Columns 1 and 2, we report the regressions with only year-quarter fixed effects, while in Columns 3 and 4, we include firm fixed effects. When examining the speed of closure by issue topic we label a topic as such if the issue-weight on that topic is above 0.5, and exclude the small fraction of issues (6 percent) that cannot be labeled in such way. Note that this restriction is only applied to the single-issue analysis. We also set the risk modeling topic as the omitted category, so that are estimated issue dummies are measured as a contrast to that category.

Stressed banks close issues between 10 percent to 16 percent faster than other banks across specifications (Table 10), which as discussed above may be because of increased urgency with which they may need to resolve issues. Across banks, larger institutions do not close issues faster than small; but within a bank, as banks grow in size, the length of time to closure increases. Recall that issues related risk modeling is the omitted dummy variable, and so we see that all other topics close their issues faster than this baseline. In the extreme case, issues

[17] As an additional robustness check, we impute the closing time for unclosed issues to be January 1st, 2015, and rerun our results. We find qualitatively similar results in this case as well.

related to compliance & regulation close 70% faster, while issues related to capital & liquidity close only nine percent faster. These results are consistent with Figure 8. We also see that issues that are defined as "Matters Requiring *Immediate* Attention" (MRIA) are closed slightly more slowly than MRAs. This suggests that while supervisors would like faster responses on MRIAs, they are also more difficult to close. We also see that issues at CCAR BHCs and State Member Banks are not resolved more quickly than at non-CCAR BHCs.

In thinking about the number issues assigned to a bank, an important issue is the extent to which these may impose frictions. For example, increased attention will result in a higher number of issues, but banks may be unable to address issues as their number increases because of capacity constraints. We are not able to address this issue causally as we do not have an exogenous source of variation that drives the stock of issues outstanding. However, we can examine the correlation of the stock of outstanding issues on the time to completion for a new issue. A one percent increase in the current stock of issues is correlated with a 0.28 percent or 0.07 percent increase in the time of completion for a new issue, respectively (column 1-3 in the table). Figure 9 presents a graphical representation of this estimated relation using the specification in Column 1. The tight relation between the stock of outstanding issues and the time to close could come about for several reasons: banks that have a large stock of open issues may be in worse conditions, and thus unable to close subsequent issues; or banks may have finite resources to address issues and hence previous supervisory attention delays their response time to new issues. To address the first concern we control for the stressed rating of the bank and firm fixed effects, however, a more exogenous source of variation would have been certainly preferable.

When we examine the subcomponents of the stock of outstanding issues, we note that the stock of issues related to capital & liquidity and internal controls predict a slower closure time, while the stock of issues related to compliance & regulations actually predict a faster response time.

5 Market monitor issues

We first describe the textual data used in this analysis. Then we turn to a discussion of how we apply the topic labeling from supervisory issues to the earnings call text. We finally discuss the results.

5.1 Earnings analysts as market monitors

Earnings calls are made by publicly traded companies shortly after their quarterly earnings announcements. While calls are not required by regulations, large publicly traded companies hold regular earnings calls as part of their relation with investors, while smaller companies

may not because of lower interest from investors and analysts. Most earnings calls take place on the same day, or the day after, press earnings announcements, which in turn are released in the month following the end of the quarter during "earnings season."

Earnings calls are presented by a company's CEO and/or CFO and start with a legal statement about safe harbor for forward looking statements. They include a discussion of quarterly results by senior management (the presentation) and are then followed by a question and answer session with analysts the content of which we will focus on. While the presentation discusses the content of the press release, the question and answer session may cover new material and provide additional "color" for analysts. The analysts on the call can phone in questions. Each analyst usually asks a short series of questions. In few cases, the earnings call is pre-recorded in which case there will not be a question and answer session.

Earnings calls were opened to the public by the SEC's rule Regulation Fair Disclosure (FD) in the year 2000. Regulation FD mandates that companies must release material information to all investors at the same time. If the call is more than 48 hours away from the earnings press release (which will be provided to the SEC on Form 8-K), the call must be separately reported. However, this does not mean that the transcripts are freely available. Typically, press releases, presentations that were used in the call, and a recording of the call are placed in the Investor Relations section of the company's website. Transcripts may be made by the firm or by a third party, which in our case is Fair Disclosure Wire via Lexis Nexis.

Our sample includes all BHCs that file FR-Y9C and that existed in the CRSP-FRB crosswalk.[18]

Our sample runs from 2002 to 2014, and our analysis uses 200 firms over this period with 3,770 BHC-quarter observations. Each document is split into three components: preamble, analyst questions and bank answers. In our following analysis, we focus on the text in analyst questions as a measure of market monitors, since the questions asked by analysts will reflect concerns of market participants that are independent of the information released by the firm itself.

5.2 Supervisory topics on earnings calls and statistical model

Linguistic methods With estimated values of β and θ from the supervisory dataset, it is straightforward to take a new document and infer $p(z|d)$, where d is the vector of words in the new document. The crucial assumption that we are forced to make is to ignore the words in the new document that were not found in the supervisory issues. Obviously, there is no estimated value of β for that word, and hence it provides no meaningful information in the context of the model. In the context of earnings calls, this is a meaningful assumption – the vocabulary

[18] This is available on the Federal Reserve Bank of New York website: Federal Reserve Bank of New York. 2014. CRSP-FRB Link.

overlap between the two datasets (following standard preprocessing) is 9,825 words, when compared to a vocabulary of 19,789 words and 18,491 in the supervisory and earnings calls, respectively.

With our dataset of earnings calls, we read in the analyst questions and estimate the topic weights of each question, given the model parameters. This is done using Bayes' rule – since $p(w|z)$ and $p(z)$ are known from the original dataset, it is straightforward to calculate $p(z|w)$. Given that words are assumed to be exchangeable, this leads to an estimate of the topics given a document of words.

Statistical model We are first interested in identifying whether the labeling procedure is informative, and what bank characteristic are associated with the resulting issue categories. We examine the characteristics of the analyst calls in a manner similar to Equation 4.1. However, since the questions of analysts are not signed, an increase in the number of questions of a given type is not as meaningful as a new supervisory issue. For example, follow-up questions regarding capital may be positive, rather than negative, while supervisory issues are exclusively negative. As a result, we focus on the percentage share of questions devoted to a particular topic:

$$\text{ShareTopicAnalyst}_{ikt} = \alpha_k + \beta_k X_{it} + \gamma_k W_{it} + \epsilon_{it}, \tag{5.1}$$

where $\text{ShareTopicAnalyst}_{ikt} = \text{NumTopicAnalyst}_{ikt} / \sum_{l=1}^{5} \text{NumTopicAnalyst}_{ilt}$, where NumTopic Analyst_{ikt} is the sum of the topic k weight for bank i in period t. We control for firm characteristics in X_{it}, which include Tier 1 capital ratios, ROA, NPLs, charge-offs and log assets as of the beginning of the quarter. W_{it} controls for bank regulatory characteristics such as whether the BHC is a CCAR bank and whether the bank is stressed (has a supervisory rating greater than 2). Note that in these regressions, since the outcome variables sum to 1, the coefficients on the firm characteristics sum to zero. Since our sample is substantially more limited over this period (with only 200 firms), we run a fully pooled regression.

It is first valuable to examine the unconditional means of the different outcome variables (reported at the bottom of Table 11). Shares of analyst questions are highly concentrated in topics related to risk modeling, loan portfolio and capital & liquidity. A very small fraction of questions are related to internal controls as compared to about 20% for Fed examiners as shown in Figure 4. Turning to the firm characteristics, larger banks have a substantially higher fractions of questions related to risk modeling, both in terms of log assets, as well as through the effect of the dummy for CCAR BHCs. This is perhaps unsurprising given that stress tests are supervisory assessments that are publicly disclosed, and may therefore elicit questions from analysts. BHCs with worse asset quality, i.e. higher NPL and charge-offs have a higher share of questions devoted to loan portfolio, which suggest that analysts pay more attention to these issues as these the measured quality of the loan book worsens. Lastly, firms with higher Tier 1

capital ratios have more questions devoted to capital and liquidity, which is likely a result of the fact that our method does not distinguish sentiment of the questions. A BHC may raise capital in the previous quarter, and an analyst may ask why the firm did so, for example. This was of course not a concern when measuring supervisory issues, which are only raised, for example, when capital levels are low, rather than high. Broadly speaking, firm characteristics appear to be quite predictive for the share of questions related to risk modeling and loan portfolio (R^2 values of 0.33 and 0.20, respectively), but explain much less variation for the other three topics (R^2 values less than 0.05).

We next study the association between these analyst share measures and the corresponding ones from supervisory issues:

$$\text{ShareTopicAnalyst}_{ikt} = \alpha + \tau_k \text{NumIssuesOpenTopic}_{ikt} + \tau \text{NumOtherIssues}_{ikt} + \beta X_{it} + \gamma W_{it} + \epsilon_{it}, \tag{5.2}$$

where $\text{NumIssuesOpenTopic}_{ikt}$ is the number of new issues in topic k raised by the Federal Reserve supervisors for BHC i in period t, and $\text{NumOtherIssues}_{ikt}$ is the number of issues raised in the other topics. Since we only observe issues from 2009 to 2014, we restrict our analyst sample to this period, leaving us with 2133 observations. Both X_{it} and W_{it} contains the controls as in the previous regression, along with indicators for whether the BHC has an ongoing exam and date-by-CCAR fixed effects.

We report regression estimates in Table 12, where we omit all controls except for the estimates of τ_k. Three topics stand out: higher number of issues related to risk modeling, loan portfolio and capital & liquidity predict a higher share of analyst questions devoted to these same topics. Note that this effect is conditional on bank characteristics, as well as date fixed effects, which suggests that the same supervisory concerns of the Federal Reserve are being raised by market monitors during earnings calls.

Issues related to compliance and regulations and internal controls do not appear to correlate with the share of questions discussed by analysts. One reason for this may be that these issues are not tightly linked to short term profitability, which may be the primary concern of analysts. Indeed as shown in Table 2, compliance & regulation are often related to BSA/AML and consumer protection rules, while issues related to internal controls are related to operational risk and IT.

But the results in Table 12 could also be explained in terms of a distinction of soft versus hard information. Federal Reserve supervisors present issues to banks that are based both on verifiable information as well as information that cannot be observed by markets (see Eisenbach et al., 2016, for a formalization). While all issues deal with a combination of both types of information, capital & liquidity and loan portfolio issues are those for which verifiable public

information is more easily accessible to market monitors thanks to financial or regulatory filings (FR-Y9Cs or SEC). Although non-verifiable information of supervisors will enter those issues as well, we find that consistent with a hard/soft information distinction, openings for these issues are related to observables such capital levels and loan portfolio measures. Similarly, given that results of supervisory assessments are made public, they can translate into hard information and as such we see them discussed in analyst questions. In contrast, compliance & regulations and internal controls are based on information that is non-verifiable, and as such they will not be reflected in market monitors' concerns.

6 Conclusion

Bank supervisors collect information through their monitoring activities, and they mandate corrective actions to banks when this information points to risks. This paper uses information on the timing, volume and description of confidential supervisory actions known as matters requiring attention (MRA) and immediate attention (MRIAs) to characterize Fed supervisory activities.

The volume of supervisory actions is indeed associated with higher bank risk, as proxied by supervisory ratings, but it also varies in terms of the extent of supervisory attention. New issues at smaller institutions, which are not monitored on a continuous basis, are mostly associated with yearly on-site examination activity. Larger banks generally receive more issues, but the data also suggest a discontinuity in the number of new and stock issues outstanding for banks with assets crossing $50 billion, a threshold that is associated to the inclusion in supervisory assessment programs as well as enhanced prudential standards under the 2010 DFA. While the jump in the volume of issues is mostly associated with risk modeling and internal controls, we find that an increase in these issues are generally not associated with more severe supervisory actions such as formal enforcement ones and supervisory rating downgrades.

Issues related to loan portfolio and capital & liquidity appear to be the most closely associated with rating downgrades. These same issues are also correlated with changes in bank characteristics, in particular, Tier 1 capital ratios and NPLs. Instead, others appear to be mostly associated with soft information.

Bank supervision is one of the three Basel pillars, along with minimum capital requirements and market discipline. These pillars may be considered both complements and substitutes to one another (Eisenbach et al., 2016). Stricter regulations can help achieve financial stability objectives, but rules require oversight – in fact, we find that 19% of issues raised are related to compliance with rules and 17% to capital & liquidity. This implies that regulation relies on supervision for its enforcement. In contrast, market discipline works through investors rewarding banks that are prudent in their risk management, and penalizing those that are not

(Lopez, 2003). We find a correlation between issues raised by Fed supervisors and the topics of market monitors' questions at quarterly earning calls for issues that are mostly involved with hard and verifiable information. If one assumes that the correlation of measured topics that we find maps into a correlation of content, and that banking authorities and market monitors share common preferences, one could interpret our findings as suggesting that market monitors can substitute for supervisors when monitoring hard information. To make this argument, however, one also needs to assume that market disclosure is not affected by the absence of bank supervisors, which is in contrast to results in the literature on the auditing effects of bank supervisors (Costello et al., 2015; Gunther and Moore, 2003). Similarly, while it is the case that market monitors are more likely to discuss risk modeling when Fed supervisors raise those issues, this likely revolves around the fact that results of supervisory assessments have been made public. Exploring further the interaction between bank supervision and the other Basel pillars remains an open and important question in banking policy.

References

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

Berger, A. N. and S. M. Davies (1998). The information content of bank examinations. *Journal of Financial Services Research 14*(2), 117–144.

Berger, A. N., S. M. Davies, and M. J. Flannery (1998). Comparing market and supervisory assessments of bank performance: who knows what when? *FEDS Paper* (98-32).

Bhattacharya, S., M. Plank, G. Strobl, and J. Zechner (2002). Bank capital regulation with random audits. *Journal of Economic Dynamics and Control 26*(7), 1301–1321.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *the Journal of machine Learning research 3*, 993–1022.

Cargill, T. F. (1989). Camel ratings and the cd market. *Journal of Financial Services Research 3*(4), 347–358.

Chuang, J., C. D. Manning, and J. Heer (2012). Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 74–77. ACM.

Cole, R. A. and J. W. Gunther (1995). A camel rating's shelf life.

Costello, A., J. Granja, and J. Weber (2015). Do effective regulators increase the transparency of the banking system. Technical report, Working paper, MIT.

Decamps, J.-P., J.-C. Rochet, and B. Roger (2004). The three pillars of basel ii: optimizing the mix. *Journal of Financial Intermediation 13*(2), 132 – 155. Bank Capital Adequacy Regulation under the New Basel Accord.

Dewatripont, M. and J. Tirole (1994). *The prudential regulation of banks*. MIT Press.

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).

Eisenbach, T. M., D. O. Lucca, and R. Townsend (2016). The economics of bank supervision. *FRB of New York Staff Report*.

Frankel, R., W. J. Mayew, and Y. Sun (2010). Do pennies matter? investor relations consequences of small negative earnings surprises. *Review of Accounting Studies 15*, 220–242.

Goldsmith-Pinkham, P. and D. O. Lucca (2016). An intuitive criterion for selecting topics in latent dirichlet allocation. Technical report.

Gunther, J. W. and R. R. Moore (2003). Loss underreporting and the auditing role of bank exams. *Journal of Financial Intermediation 12*(2), 153–177.

Harris, M. and A. Raviv (2012). How to get banks to take less risk and disclose bad news. *Chicago Booth Research Paper* (12-55).

Hirschhorn, E. (1998). The information content tof bank examination ratings. *FDIC Banking and Economic Review*.

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

Lopez, J. A. (2003). Disclosure as a supervisory tool: Pillar 3 of basel ii. *FRBSF Economic Letter 22*, 47–63.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance 61*, 35–65.

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.

Prescott, E. S. (2004). Auditing and bank capital regulation. *Federal Reserve Bank of Richmond Economic Quarterly 90*(4), 47–63.

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

Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*, 1139–1168.

Table 1: **Topic labels, unigrams and bigrams.** This table presents the topic output from the methodology described in Section 3. Label gives our proposed label for the topic, Unigrams presents the eight most salient words per topic, and Bigrams give the five most frequent bigrams per topic. We ignore "management must", "management [is] required", "management [is] expected", "risk assessment", "board [of] director[s]" and "risk management", as they each show up in the top five for three out of the five topics, and hence are not informative for topic labeling. Bigrams are calculated ignoring ordering of the words, however we attempt to present them here in the most sensible order. In places where natural stop words have been removed, such as "of", we reintroduce them for clarity. Bigrams are reported in order of frequency. Unigrams are reported in order of saliency.

Topic	Label	Unigrams	Bigrams
1	**Compliance & Regulations**	regulation, account, bank, compliance, customer, violation, procedure, transaction	policy procedure, due diligence, suspicious activity, reserve bank, compliance regulation
2	**Internal Controls**	audit, internal, program, assessment, committee, IT, control, function	internal audit, policy procedure, audit committee, information security, operational risk
3	**Risk Modeling**	model, risk, firm, process, data, validation, loss, framework	stress testing, policy procedure, risk rating, risk model, model validation
4	**Loan Portfolio**	loan, credit, ALLL, appraisal, analysis, impairment, collateral, portfolio	credit risk, ALLL methodology, loan policy, real estate, loan review
5	**Capital & Liquidity**	capital, plan, liquidity, funding, limit, board, stress, contingency	capital plan, liquidity risk, interest rate, contingency funding, funding plan

Table 2: **Cross tabulation of topics.** This table compares Fed categories with those obtained from applying LDA to the issue descriptions as detailed in Section 3.

	Compl. & Reg.	Internal Ctrl.	Risk Model.	Loan Port.	Cap. & Liquid.	Total
Asset & Wealth Mgmt	6.3	1.3	1.1	0.9	0.9	2.2
Audit Issues	0.7	16	4.9	0.3	0.2	4.9
BSA / AML	18.9	6.4	13.6	0.1	0.2	6.4
Consumer Compliance / CRA	40.2	8.4	3.7	5.2	1.6	13.1
Credit Risk	4.1	3.3	15.5	77.9	6.1	24.8
IT-Related	1.2	30.1	8.8	0.2	0.3	9
Market/Liquidity Risk	0.5	1.5	28.6	1.3	49.9	11.6
Mgmt / Risk Mgmt	9.2	8.1	11.3	7.2	22	11
Operational Risk	8.6	20.8	4.9	1.1	3.3	8.8
Other	10.3	4	7.6	5.7	15.4	8.3
Total	100	100	100	100	100	100

Table 3: **Summary statistics.** This table presents summary statistics for the variables included in the regressions specifications.

	All Banks		SMBs		Non-CCAR BHCs		CCAR BHCs	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: Supervisory Issues								
Number of Issues Opened	0.89	3.00	1.42	3.71	0.16	1.06	5.99	6.42
Compliance and Regulations Opened	0.19	0.76	0.33	1.00	0.03	0.24	0.66	1.18
Internal Ctrls. Opened	0.23	0.98	0.38	1.23	0.05	0.44	1.28	1.71
Risk Model. Opened	0.09	0.63	0.07	0.28	0.01	0.13	2.82	3.73
Loan Portfolio Opened	0.21	0.89	0.38	1.19	0.02	0.24	0.46	0.86
Capital and Liquidity Opened	0.17	0.70	0.26	0.88	0.05	0.35	0.77	1.27
Number of Issues Closed	0.67	2.27	1.07	2.79	0.12	0.79	4.22	5.56
Compliance and Regulations Closed	0.15	0.64	0.27	0.85	0.02	0.17	0.48	1.01
Internal Ctrls. Closed	0.18	0.80	0.29	1.02	0.03	0.32	0.94	1.64
Risk Model. Closed	0.06	0.53	0.05	0.24	0.01	0.10	1.99	3.46
Loan Portfolio Closed	0.16	0.69	0.28	0.91	0.02	0.21	0.40	0.91
Capital and Liquidity Closed	0.12	0.55	0.19	0.68	0.04	0.26	0.62	1.19
Number of Issues Stock	4.85	10.26	7.74	10.45	0.80	3.02	34.54	34.96
Compliance and Regulations Stock	0.78	1.82	1.31	2.15	0.13	0.54	3.81	4.62
Internal Ctrls. Stock	1.26	2.90	2.03	3.30	0.24	1.23	7.33	7.14
Risk Model. Stock	0.49	3.38	0.37	0.67	0.06	0.36	16.95	20.55
Loan Portfolio Stock	1.28	2.93	2.36	3.73	0.11	0.58	2.25	2.57
Capital and Liquidity Stock	1.04	2.35	1.67	2.80	0.26	1.11	4.21	4.22
Panel B: Firm Characteristics								
ROA	0.63	1.03	0.73	1.01	0.53	1.04	0.58	0.89
Tier 1 Cap.	14.43	5.45	15.37	6.11	13.52	4.54	12.18	2.08
Log Assets	13.32	1.68	12.37	1.43	14.13	0.98	19.15	1.16
NPL	2.41	2.39	2.02	2.20	2.79	2.53	3.15	1.92
Charge-Offs	2.63	4.12	2.04	3.82	3.17	4.28	5.45	5.07
Stressed Rating	0.25	0.43						
Failure within year (x100)	0.23	4.78	0.43	6.55	0.02	1.52	0.00	0.00
Observations	36246		18344		17325		577	

Table 4: **Issue openings and basic firm characteristics.** This table reports estimates of the number of issues opened in a quarter on basic bank attributes:

$$\text{Open}_{it} = \alpha + \gamma W_{it} + \epsilon_{it}$$

for the overall number of issues (column 1) or for the number of issues in a given topic (columns 2-6). For each row, coefficients in columns 2-6 sum the coefficient in column 1. Topics are defined in Section 3. Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Panel A: Pooled Variation

	(1) Num. Issues	(2) Comp. & Regs	(3) Internal Ctrls.	(4) Risk Model.	(5) Loan Port.	(6) Cap. & Liq.
Stressed Rating	0.356***	0.036***	0.050***	0.011	0.140***	0.119***
	(0.052)	(0.012)	(0.017)	(0.022)	(0.016)	(0.012)
Log Assets	0.085***	0.017***	0.032***	0.030***	0.001	0.005
	(0.027)	(0.006)	(0.008)	(0.009)	(0.007)	(0.005)
State Member Bank	1.194***	0.249***	0.354***	0.136***	0.250***	0.205***
	(0.134)	(0.035)	(0.050)	(0.028)	(0.036)	(0.033)
CCAR BHC	5.671***	0.699***	1.374***	2.576***	0.503***	0.520***
	(0.988)	(0.242)	(0.232)	(0.631)	(0.151)	(0.122)
Ongoing, CCAR	0.044	-0.238	-0.184	0.373	-0.076	0.169
	(0.780)	(0.204)	(0.206)	(0.485)	(0.113)	(0.124)
Ongoing, Non-CCAR	1.470***	0.234***	0.449***	0.024***	0.438***	0.324***
	(0.049)	(0.012)	(0.018)	(0.007)	(0.017)	(0.011)
Failure within year	-0.690***	-0.195***	-0.293***	0.007	-0.163	-0.046
	(0.237)	(0.067)	(0.077)	(0.053)	(0.109)	(0.085)
R^2	0.143	0.060	0.079	0.270	0.094	0.082
Mean of Outcome	0.883	0.185	0.239	0.065	0.224	0.169
Firm FE?	No	No	No	No	No	No
Year-Quarter-BHC FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38000	38000	38000	38000	38000	38000

Panel B: Within variation (fixed effects)

	(1) Num. Issues	(2) Comp. & Regs	(3) Internal Ctrls.	(4) Risk Model.	(5) Loan Port.	(6) Cap. & Liq.
Stressed Rating	0.241***	0.034*	0.061**	-0.008	0.081***	0.072***
	(0.069)	(0.019)	(0.024)	(0.017)	(0.025)	(0.020)
Log Assets	0.633***	0.093***	0.144***	0.009	0.197***	0.190***
	(0.101)	(0.035)	(0.041)	(0.020)	(0.038)	(0.034)
Ongoing, CCAR	0.489	-0.228	-0.101	0.678***	-0.068	0.208
	(0.492)	(0.185)	(0.218)	(0.219)	(0.097)	(0.137)
Ongoing, Non-CCAR	1.463***	0.236***	0.453***	0.033***	0.426***	0.316***
	(0.051)	(0.012)	(0.019)	(0.003)	(0.017)	(0.012)
Failure within year	-0.851	-0.180	-0.279	-0.027	-0.109	-0.255
	(0.521)	(0.124)	(0.200)	(0.018)	(0.172)	(0.205)
R^2	0.219	0.123	0.143	0.483	0.154	0.144
Mean of Outcome	0.883	0.185	0.239	0.065	0.224	0.169
Firm FE?	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter-BHC FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38000	38000	38000	38000	38000	38000

Table 5: **Issue openings and bank characteristics.** This table reports estimates of the number of issues opened in a quarter on bank characteristics:

$$\text{Open}_{it} = \alpha + \beta X_{it} + \gamma W_{it} + \epsilon_{it},$$

for the overall number of issues (column 1) or for the number of issues in a given topic (columns 2-6). For each row, coefficients in columns 2-6 sum to the coefficient in column 1. Topics are defined in Section 3. Unreported controls include the stock of total issues outstanding at the beginning of the quarter (column 1) and issues by topic (columns 2-6). Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) Num. Issues	(2) Comp. & Regs	(3) Internal Ctrls.	(4) Risk Model.	(5) Loan Port.	(6) Cap. & Liq.
Log Assets	0.265*	0.064	0.074	0.009	0.046	0.073**
	(0.146)	(0.046)	(0.052)	(0.019)	(0.037)	(0.033)
Tier 1 Cap.	-0.017**	-0.003	-0.001	-0.002**	-0.006**	-0.006***
	(0.007)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
ROA	-0.057*	-0.001	-0.005	0.000	-0.033***	-0.018**
	(0.030)	(0.007)	(0.009)	(0.003)	(0.010)	(0.007)
NPL	0.051***	0.001	0.009*	0.001	0.027***	0.013***
	(0.014)	(0.003)	(0.004)	(0.001)	(0.005)	(0.004)
Charge-Offs	0.016***	0.002	0.004**	0.000	0.006***	0.005***
	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Stressed Rating	0.453***	0.065***	0.113***	0.007	0.136***	0.132***
	(0.075)	(0.020)	(0.025)	(0.012)	(0.024)	(0.021)
R^2	0.269	0.156	0.191	0.537	0.205	0.204
Mean of Outcome	0.892	0.191	0.234	0.086	0.212	0.169
Firm FE?	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE?	Yes	Yes	Yes	Yes	Yes	Yes
Ongoing-Exam FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36246	36246	36246	36246	36246	36246

Table 6: **Issue openings and changes in bank characteristics.** This table reports estimates of the number of issues opened in a quarter on the one-year-lagged and the year-over-year change in bank characteristic:

$$\text{Open}_{it} = \alpha + \beta X_{it-4} + \gamma \underbrace{(X_{it} - X_{it-4})}_{\text{Bank Changes}} + W_{it-4} + \epsilon_{it}$$

for the overall number of issues (column 1) or for the number of issues in a given topic (columns 2-6). For each row, coefficients in columns 2-6 sum to the coefficient in column 1. Topics are defined in Section 3. Unreported controls include the stock of total issues outstanding at the beginning of the quarter (column 1) and issues by topic (columns 2-6). Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) Num. Issues	(2) Comp. & Regs	(3) Internal Ctrls.	(4) Risk Model.	(5) Loan Port.	(6) Cap. & Liq.
Log Assets$_{t-4}$	0.283	0.076	0.052	0.014	0.064	0.077*
	(0.173)	(0.056)	(0.064)	(0.023)	(0.045)	(0.040)
Stressed Rating$_{t-4}$	0.210***	0.064***	0.060***	0.002	0.062***	0.022
	(0.069)	(0.018)	(0.023)	(0.009)	(0.022)	(0.018)
ROA$_{t-4}$	-0.059	0.004	-0.001	0.001	-0.033**	-0.030***
	(0.046)	(0.011)	(0.014)	(0.004)	(0.015)	(0.011)
Tier 1 Cap.$_{t-4}$	-0.030***	-0.005*	-0.005	-0.002	-0.009***	-0.010***
	(0.010)	(0.003)	(0.004)	(0.001)	(0.003)	(0.003)
NPL$_{t-4}$	0.095***	0.009*	0.019***	0.004*	0.038***	0.025***
	(0.018)	(0.005)	(0.006)	(0.002)	(0.007)	(0.005)
Charge-Offs$_{t-4}$	0.020**	0.003	0.006*	0.000	0.008***	0.004
	(0.009)	(0.002)	(0.003)	(0.001)	(0.003)	(0.002)
$\Delta_{t-4,t}$ Log Assets	0.328	0.129*	0.131	0.005	-0.014	0.077
	(0.245)	(0.066)	(0.081)	(0.034)	(0.063)	(0.065)
$\Delta_{t-4,t}$ ROA	-0.077**	-0.003	-0.008	-0.001	-0.040***	-0.025***
	(0.033)	(0.007)	(0.010)	(0.003)	(0.012)	(0.008)
$\Delta_{t-4,t}$ Tier 1 Cap.	-0.025**	-0.004	-0.004	-0.002	-0.009***	-0.006**
	(0.010)	(0.003)	(0.004)	(0.001)	(0.003)	(0.003)
$\Delta_{t-4,t}$ NPL	0.064***	0.003	0.013**	0.002	0.032***	0.015***
	(0.016)	(0.004)	(0.005)	(0.001)	(0.006)	(0.005)
$\Delta_{t-4,t}$ Charge-Offs	0.020***	0.002	0.004**	0.000	0.007***	0.006***
	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
R^2	0.284	0.163	0.202	0.548	0.220	0.220
Mean of Outcome	0.903	0.193	0.238	0.088	0.213	0.171
Firm FE?	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE?	Yes	Yes	Yes	Yes	Yes	Yes
Ongoing-Exam FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33325	33325	33325	33325	33325	33325

Table 7: **Issue closings and changes in bank characteristics.** This table reports estimates of the total number of issues closed in the next four quarters on current bank characteristics and the change in bank characteristics over the next four quarters:

$$\text{Closings}_{it,t+4} = \beta X_{it} + \beta \underbrace{(X_{it+4} - X_{it})}_{\text{Bank Changes}} + W_{it} + \epsilon_{it}$$

for the overall number of issues (column 1) or for the number of issues in a given topic (columns 2-6). For each row, coefficients in columns 2-6 sum to the coefficient in column 1. Topics are defined in Section 3. Unreported controls include the stock of total issues outstanding at the beginning of the quarter (column 1) and issues by topic (columns 2-6). Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}\,p < 0.1$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$

	(1) Issue Closings	(2) Comp. & Regs	(3) Internal Ctrls.	(4) Risk Model.	(5) Loan Port.	(6) Cap. & Liq.
Stressed Rating	0.115***	0.020*	0.024	0.009	0.021	0.030***
	(0.041)	(0.011)	(0.015)	(0.011)	(0.015)	(0.011)
Log Assets	0.404***	0.062*	0.046	0.006	0.121***	0.165***
	(0.112)	(0.033)	(0.045)	(0.015)	(0.032)	(0.034)
ROA	0.008	0.003	0.002	0.001	0.000	-0.002
	(0.015)	(0.004)	(0.005)	(0.002)	(0.005)	(0.005)
Tier 1 Cap.	0.023***	0.004*	0.008***	-0.000	0.005**	0.007***
	(0.007)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
NPL	-0.011	0.002	-0.001	-0.002	-0.009**	-0.004
	(0.012)	(0.003)	(0.004)	(0.001)	(0.004)	(0.004)
Charge-Offs	0.001	0.000	0.000	-0.000	-0.000	-0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\Delta_{t,t+4}$ Log Assets	0.266***	0.052*	0.060	0.004	0.106***	0.083***
	(0.101)	(0.030)	(0.044)	(0.015)	(0.032)	(0.029)
$\Delta_{t,t+4}$ ROA	0.007	0.004	-0.001	0.001	0.002	-0.002
	(0.010)	(0.003)	(0.004)	(0.001)	(0.003)	(0.003)
$\Delta_{t,t+4}$ Tier 1 Cap.	0.010*	0.002	0.005**	-0.002**	0.001	0.003*
	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
$\Delta_{t,t+4}$ NPL	-0.016*	0.002	-0.002	-0.001	-0.011***	-0.003
	(0.009)	(0.002)	(0.003)	(0.001)	(0.003)	(0.003)
$\Delta_{t,t+4}$ Charge-Offs	-0.000	0.001	0.001	0.000	-0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
R^2	0.810	0.771	0.733	0.924	0.745	0.704
Mean of Outcome	0.903	0.209	0.240	0.086	0.219	0.168
Firm FE?	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter-BHC FE?	Yes	Yes	Yes	Yes	Yes	Yes
Ongoing-Exam FE?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26988	26988	26988	26988	26988	26988

Table 8: **Supervisory actions and changes in the stock of outstanding issues.** This table reports estimates of the probability of a formal action and the rating change in the next four quarters on current bank characteristics and the net change in the stock of outstanding issues over the next four quarters:

$$Y_{it} = \alpha + \beta \Delta_{t,t+4} \text{Stock}_i + \epsilon_{it},$$

where $Y_{it} = \Delta_{t,t+4}\text{Rating}_i$ or $1(\sum_{t=0}^{t+4} \text{Formal} > 0)$, for the overall number of issues (column 1 and 3) or for the number of issues in within topics (columns 2 and 4). Topics are defined in Section 3. Unreported controls include the stock of total issues outstanding at the beginning of the quarter (column 1 and 3) and issues by topic (columns 2 and 4). Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) $\Delta_{t,t+4}$ Rating	(2) $\Delta_{t,t+4}$ Rating	(3) Formal Action$_{t,t+4}$	(4) Formal Action$_{t,t+4}$
$\Delta_{t,t+4}$ Stock of Issues	0.013***		0.003***	
	(0.001)		(0.001)	
$\Delta_{t,t+4}$ Comp. & Regs		-0.003		0.006**
		(0.006)		(0.003)
$\Delta_{t,t+4}$ Internal Ctrls.		-0.004		0.001
		(0.004)		(0.002)
$\Delta_{t,t+4}$ Risk Model.		-0.005		0.003
		(0.005)		(0.005)
$\Delta_{t,t+4}$ Loan Port.		0.035***		-0.001
		(0.005)		(0.002)
$\Delta_{t,t+4}$ Cap. & Liq.		0.037***		0.008***
		(0.007)		(0.003)
R^2	0.639	0.644	0.406	0.408
Mean of Outcome	0.057	0.057	0.046	0.046
Firm FE?	Yes	Yes	Yes	Yes
Year-Quarter-BHC FE?	Yes	Yes	Yes	Yes
Ongoing-Exam FE?	Yes	Yes	Yes	Yes
Observations	30568	30568	30568	30568

Table 9: **Changes in rating, issue openings and closings.** This table reports estimates of the probability of a formal action and the rating change in the next four quarters on current bank characteristics and opening and closings of issues over the next four quarters:

$$Y_{it} = \alpha + \beta \underbrace{\text{Openings}_{it,t+4} + \gamma \text{Closings}_{it,t+4}}_{\Delta_{t,t+4}\text{Stock}_i} + \epsilon_{it}$$

where $Y_{it} = \Delta_{t,t+4}\text{Rating}_i$ or $1(\sum_{t=0}^{t+4}\text{Formal} > 0)$, for the overall number of issues (column 1 and 3) or for the number of issues within topics (columns 2 and 4). Topics are defined in Section 3. Unreported controls include the stock of total issues outstanding at the beginning of the quarter (column 1 and 3) and issues by topic (columns 2 and 4). Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) $\Delta_{t,t+4}$ Rating	(2) $\Delta_{t,t+4}$ Rating	(3) Formal Action$_{t,t+4}$	(4) Formal Action$_{t,t+4}$
Issue Closings	-0.019***		-0.003	
	(0.007)		(0.003)	
Comp. & Regs Close		0.005		-0.006
		(0.024)		(0.012)
Internal Ctrls. Close		-0.042**		-0.006
		(0.019)		(0.009)
Risk Model. Close		0.018		-0.026
		(0.030)		(0.038)
Loan Port. Close		-0.004		0.019**
		(0.023)		(0.009)
Cap. & Liq. Close		-0.057		-0.016
		(0.034)		(0.014)
Issue Openings	0.016***		0.004***	
	(0.002)		(0.001)	
Comp. & Regs Open		-0.009		0.027**
		(0.023)		(0.012)
Internal Ctrls. Open		-0.029		0.001
		(0.018)		(0.009)
Risk Model. Open		-0.014		0.011
		(0.025)		(0.024)
Loan Port. Open		0.182***		-0.002
		(0.022)		(0.010)
Cap. & Liq. Open		0.179***		0.042***
		(0.032)		(0.014)
R^2	0.643	0.652	0.408	0.411
Mean of Outcome	0.057	0.057	0.046	0.046
Firm FE?	Yes	Yes	Yes	Yes
Year-Quarter-BHC FE?	Yes	Yes	Yes	Yes
Ongoing-Exam FE?	Yes	Yes	Yes	Yes
Observations	30568	30568	30568	30568

Table 10: **Time-to-close an issue and bank characteristics.** This table reports estimates of time it takes (measured in years) for an issue to close on characteristics of the bank and issue at the time of the issue opening. The unit of observation is an issue. Columns 1 and 2 include year-quarter fixed effects as controls, while columns 3 and 4 add bank level fixed effects. Topics are defined in Section 3. Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) Log Time to Close	(2) Log Time to Close	(3) Log Time to Close	(4) Log Time to Close
Log Assets	-0.01	-0.00	0.35***	0.34***
	(0.01)	(0.01)	(0.09)	(0.09)
MRIA	0.03	0.04	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
Stressed Rating	-0.12***	-0.09**	-0.15***	-0.16***
	(0.04)	(0.04)	(0.06)	(0.06)
CCAR BHC	-0.21**	-0.07		
	(0.10)	(0.10)		
SMB	-0.28***	-0.26***		
	(0.06)	(0.06)		
Comp. & Regs	-0.65***	-0.63***	-0.70***	-0.61***
	(0.06)	(0.06)	(0.05)	(0.05)
Internal Ctrls.	-0.16***	-0.21***	-0.17***	-0.19***
	(0.04)	(0.04)	(0.04)	(0.03)
Loan Port.	-0.13***	-0.19***	-0.19***	-0.18***
	(0.05)	(0.05)	(0.04)	(0.04)
Cap. & Liq.	-0.05	-0.12**	-0.09**	-0.09**
	(0.05)	(0.05)	(0.04)	(0.04)
Log Current Stock	0.28***		0.07**	
	(0.02)		(0.03)	
Log Comp. & Regs Stock		-0.02		-0.08***
		(0.01)		(0.01)
Log Internal Ctrls. Stock		0.08***		0.09***
		(0.01)		(0.02)
Log Risk Model Stock		0.00		0.01
		(0.01)		(0.02)
Log Loan Port. Stock		0.06***		-0.02
		(0.01)		(0.01)
Log Cap. & Liq. Stock		0.05***		0.04**
		(0.01)		(0.02)
R^2	0.17	0.17	0.35	0.35
Mean of Outcome	13.31	13.31	13.31	13.31
Firm FE?	No	No	Yes	Yes
Year-Quarter FE?	Yes	Yes	Yes	Yes
Observations	35494	35494	35494	35494

Table 11: **Analyst questions and firm characteristics.** This table reports estimates of the percentage of each topic in the analyst questions asked during earnings call on basic bank attributes. Coefficients sum to zero across columns, as topic fraction sum to 1. Topics are defined in Section 3 and information on its application to earnings analyst calls is provided in Section 5.2. Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) Comp. & Regs	(2) Internal Ctrls.	(3) Risk Modeling	(4) Loan Port.	(5) Cap. & Liq.
Log Assets	-0.41***	-0.05	1.76***	-1.24***	-0.06
	(0.12)	(0.06)	(0.19)	(0.24)	(0.13)
Tier 1 Cap.	-0.15***	-0.00	0.23***	-0.29***	0.21***
	(0.04)	(0.02)	(0.07)	(0.07)	(0.05)
ROA	0.27*	0.14	0.46**	-1.00***	0.13
	(0.15)	(0.09)	(0.19)	(0.27)	(0.16)
NPL	-0.11	-0.08**	-0.32***	0.57***	-0.06
	(0.08)	(0.04)	(0.12)	(0.14)	(0.09)
Charge-Offs	0.02	-0.02	-0.12**	0.29***	-0.18***
	(0.04)	(0.02)	(0.05)	(0.07)	(0.04)
CCAR BHC	-0.83	-0.09	2.38***	-1.54	0.09
	(0.57)	(0.26)	(0.71)	(1.05)	(0.51)
Stressed Rating	-0.38	0.01	0.08	-1.14**	1.43***
	(0.42)	(0.16)	(0.58)	(0.57)	(0.45)
R^2	0.05	0.01	0.33	0.20	0.03
Mean of Outcome	18.35	5.65	32.21	21.96	21.83
Firm FE?	No	No	No	No	No
Year-Quarter FE?	No	No	No	No	No
Observations	3770	3770	3770	3770	3770

Table 12: **Market monitors and bank supervision.** This table reports estimates of the percentage of each topic in the analyst questions asked during earnings call on the number of Fed issues raised in each topic. Coefficients sum to zero across columns, as topic fraction sum to 1. Topics are defined in Section 3 and information on its application to earnings analyst calls is provided in Section 5.2. Bank characteristics are measured as of the beginning of the quarter. Standard errors clustered by bank reported in parenthesis. Significance: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

	(1) Comp. & Regs	(2) Internal Ctrls.	(3) Risk Modeling	(4) Loan Port.	(5) Cap. & Liq.
Fed Issues - Same	-0.16	0.06	0.30***	1.11***	0.52***
	(0.11)	(0.07)	(0.11)	(0.28)	(0.13)
Fed Issues - All Other	0.02	0.02	0.01	-0.16***	-0.21***
	(0.04)	(0.02)	(0.07)	(0.06)	(0.05)
R^2	0.07	0.05	0.43	0.33	0.10
Mean of Outcome	17.55	5.50	32.36	22.32	22.28
Firm FE?	No	No	No	No	No
Year-Quarter FE?	Yes	Yes	Yes	Yes	Yes
Ongoing Exam FE?	Yes	Yes	Yes	Yes	Yes
Observations	1978	1978	1978	1978	1978

Figure 1: **Topic labels.** This graph plots in the left panel the probability, given a topic, of the top eight most salient words per topic. The right panel plots the marginal probability of each word across the issues. The size of the circles in the figure reflects the size of the probability given topic. The first eight words (starting from the top) are the most salient words, in decreasing order from top to bottom, for Topic 1. The next eight words are the most salient words for Topic 2, etc. The methodology for how topics are defined and labeled and the definition for saliency is outlined in Section 3.

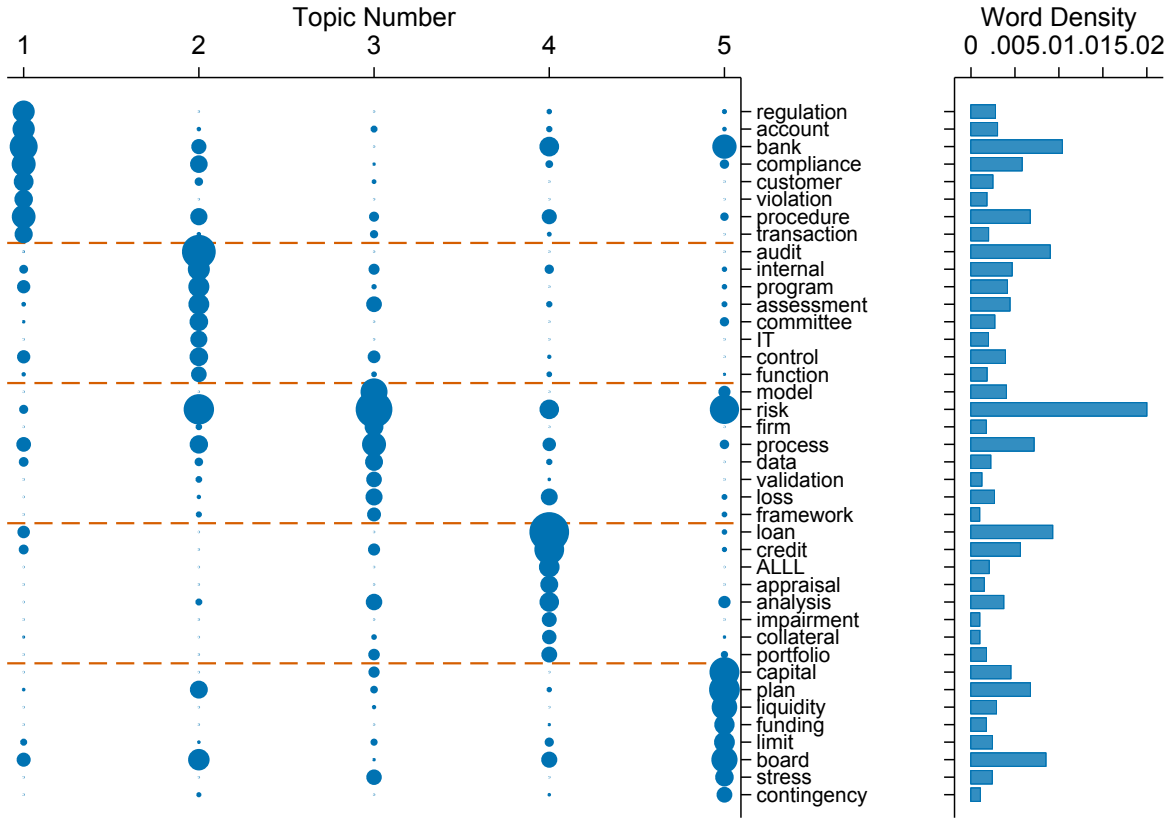


Figure 2: **Quarterly total topic openings by bank type.** This graph plots the total number issues of each topic raised per quarter by the Federal Reserve, across each type of bank supervised by the Fed. SMB denotes State Member Bank, BHC denotes Bank Holding Companies. The methodology for how topics are defined and labeled is outlined in Section 3.

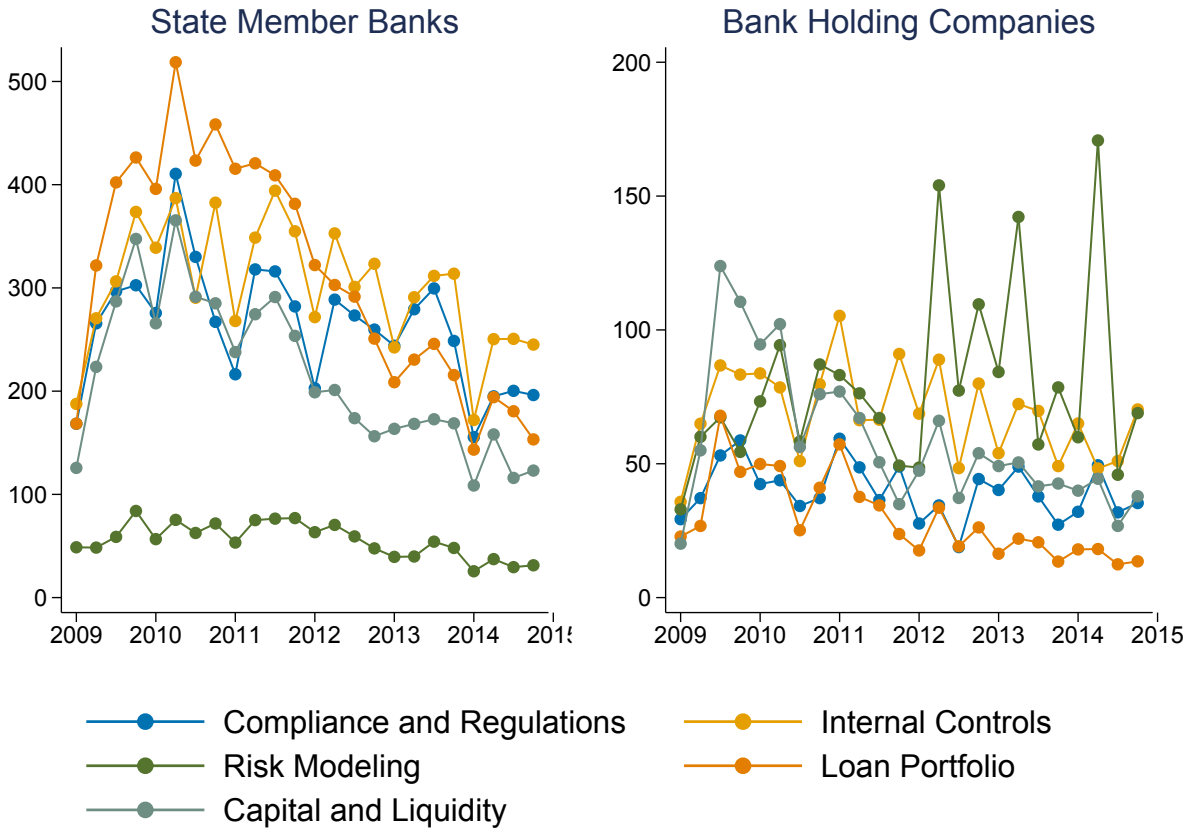


Figure 3: **Quarterly total topic stock by bank type.** This graph plots the total number issues of each topic outstanding per quarter, across each type of bank supervised by the Fed. SMB denotes State Member Bank, BHC denotes Bank Holding Companies. The methodology for how topics are defined and labeled is outlined in Section 3.

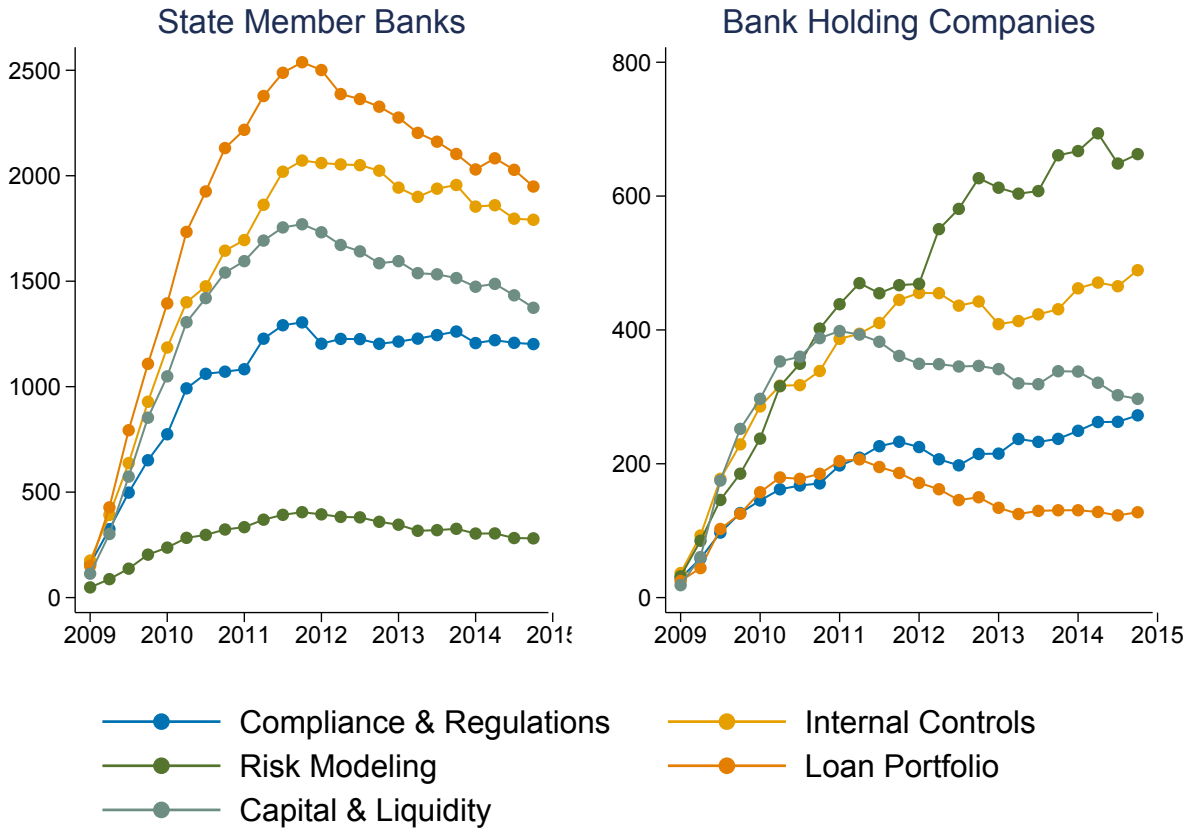


Figure 4: **Average quarterly total topics by bank type.** This graph plots the average number issues of each topic raised per quarter by the Federal Reserve, across each type of bank supervised by the Fed. If no issues are raised in that quarter, all 5 topics are set to missing for a given bank. SMB denotes State Member Bank, CCAR BHC and Non-CCAR BHC denotes Bank Holding Companies that do and do not fall under the provision of CCAR regulations. The methodology for how topics are defined and labeled is outlined in Section 3.

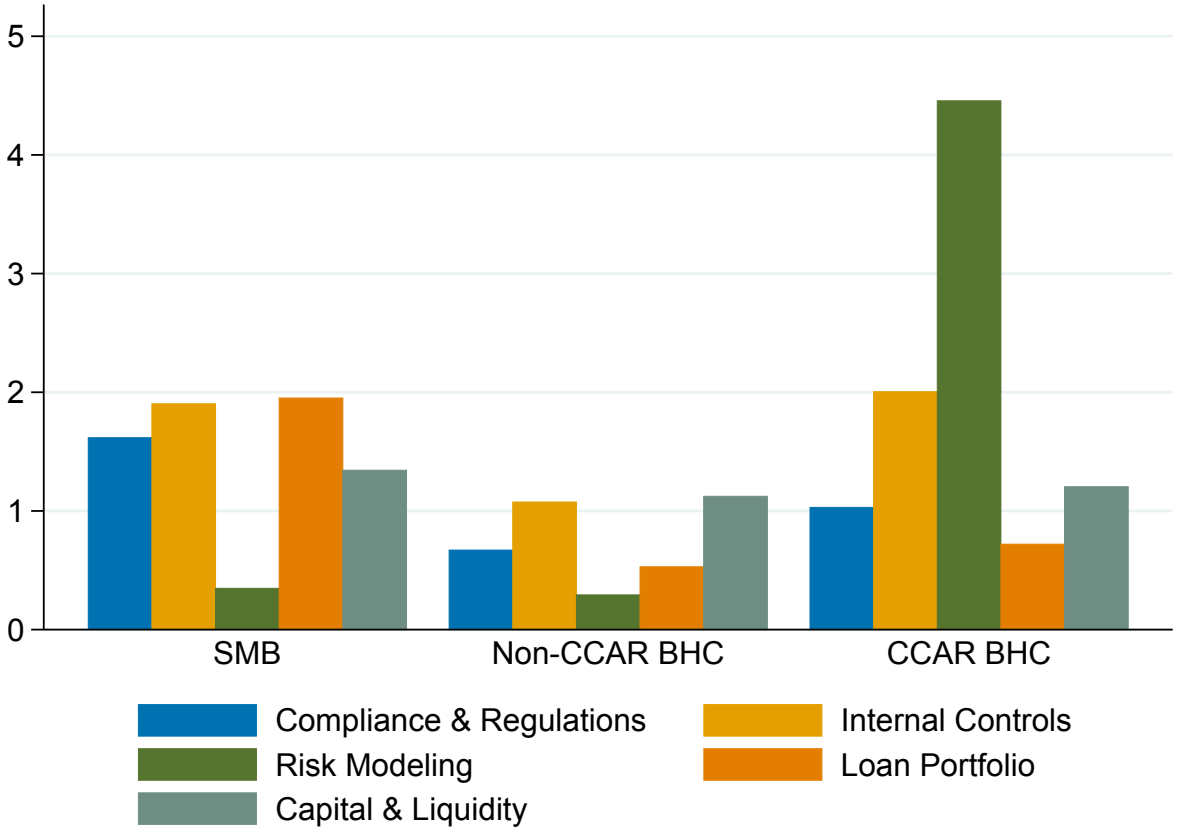


Figure 5: **Average quarterly topic fractions by bank type.** This graph plots the average topic fraction of issues raised per quarter by the Federal Reserve across each type of bank supervised by the Fed. If no issues are raised in that quarter, topic fractions are set to missing for a given bank. SMB denotes State Member Bank, CCAR BHC and Non-CCAR BHC denotes Bank Holding Companies that do and do not fall under the provision of CCAR regulations. The methodology for how topics are defined and labeled is outlined in Section 3.

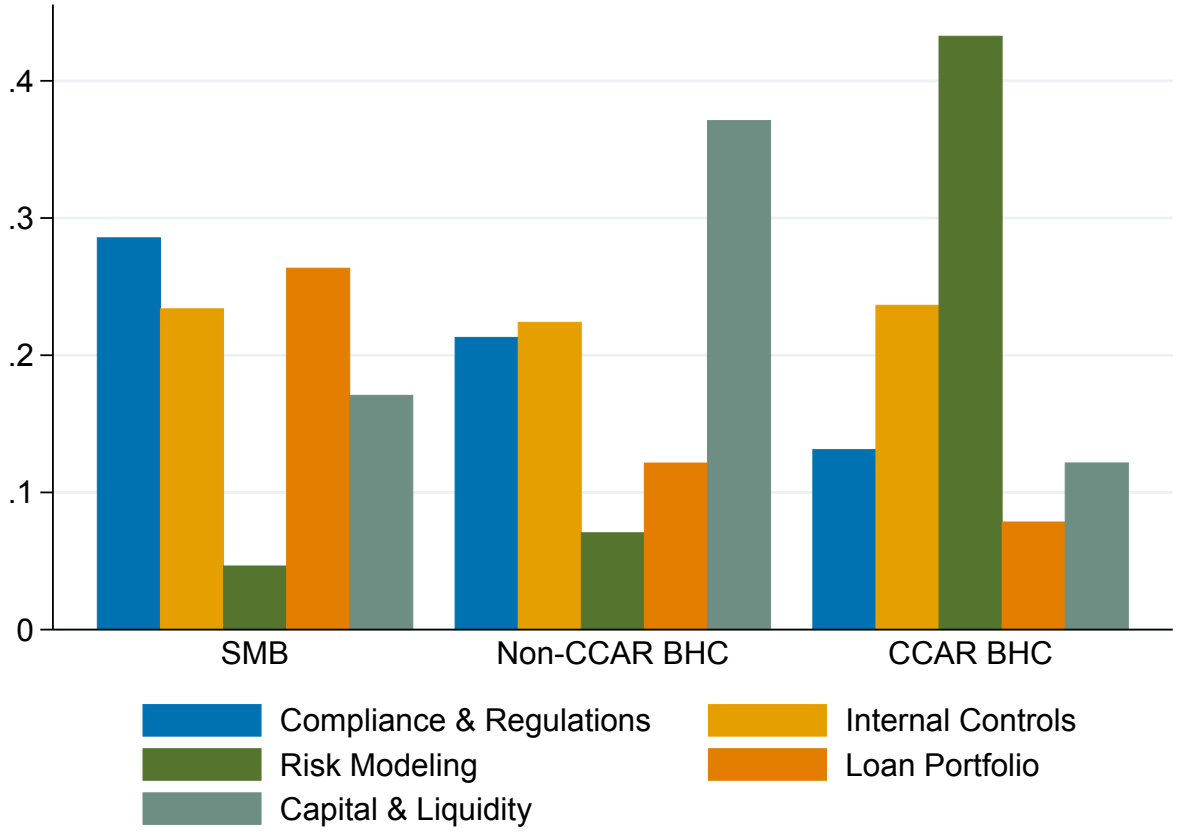


Figure 6: **Extent of of supervisory attention across asset size.** This graph plots the average number of issues raised in a quarter for Bank Holding Companies supervised by the Federal Reserve against groups of BHCs broken into billions of consolidated asset holdings. The methodology for how topics are defined and labeled is outlined in Section 3. This figure excludes State Member Banks.

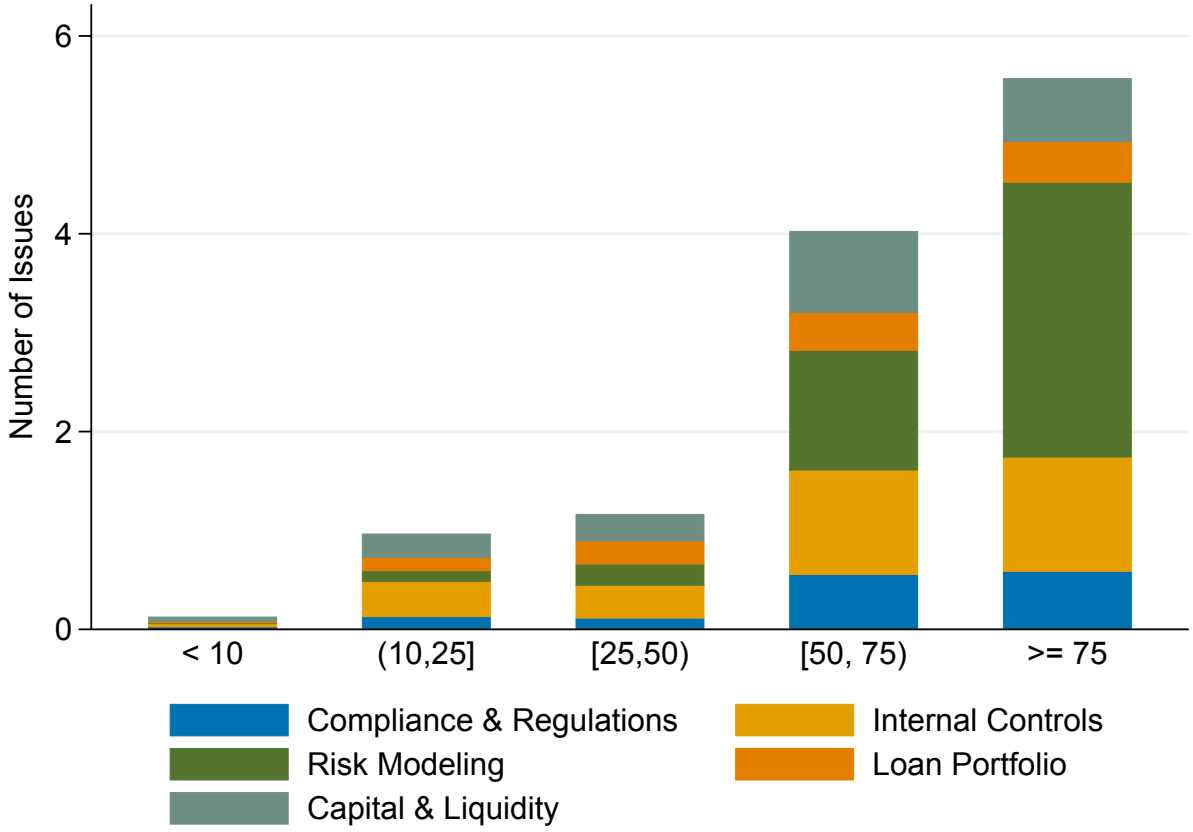


Figure 7: **Extent of of supervisory attention across asset size.** This graph plots the topic shares of issues raised in a quarter for Bank Holding Companies supervised by the Federal Reserve against groups of BHCs broken into billions of consolidated asset holdings. The methodology for how topics are defined and labeled is outlined in Section 3. This figure excludes State Member Banks.

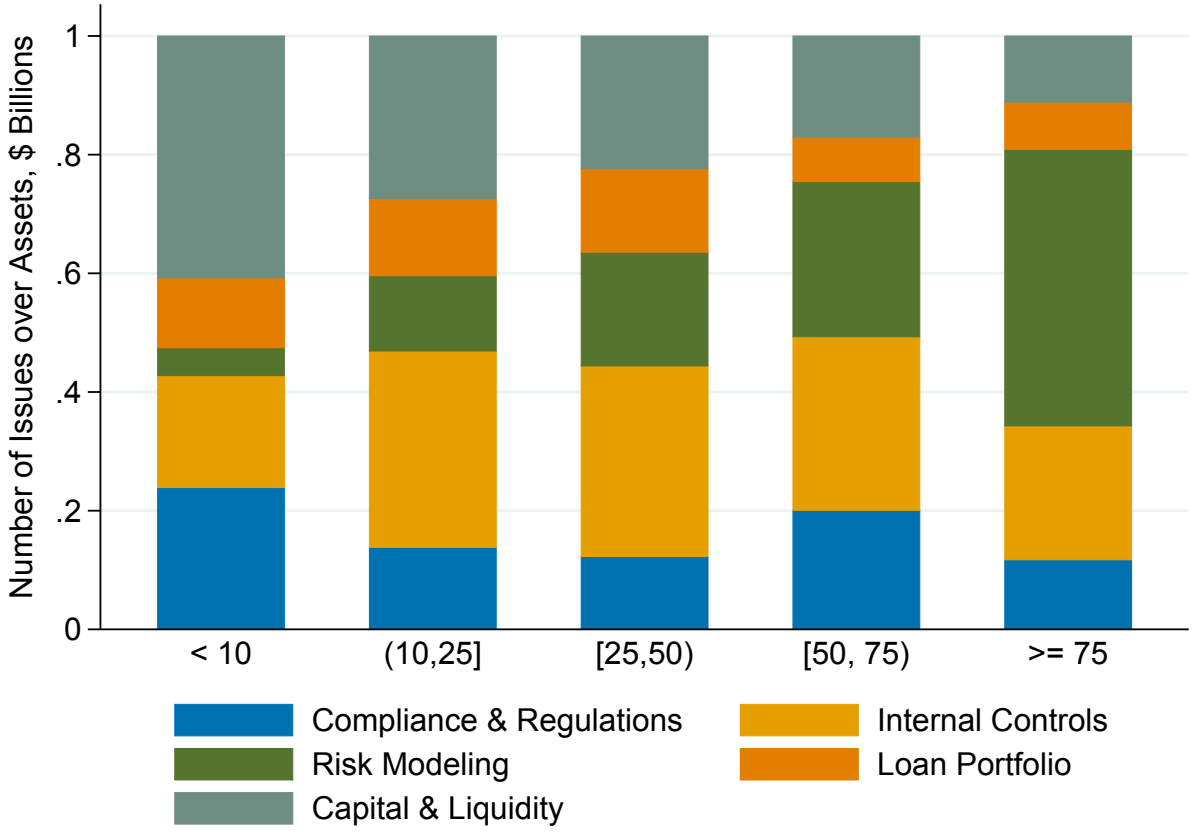


Figure 8: **Length of issue closures by topic.** This graph plots the average number of days for an issue to close by topic. Issues are marked as a topic if an issue has over fifty percent weight in a particular topic. Issues with no topic assigned are excluded, and account for 6.82 percent of the sample of issues. The methodology for how topics are defined and labeled is outlined in Section 3. Issues not yet closed are marked as missing.

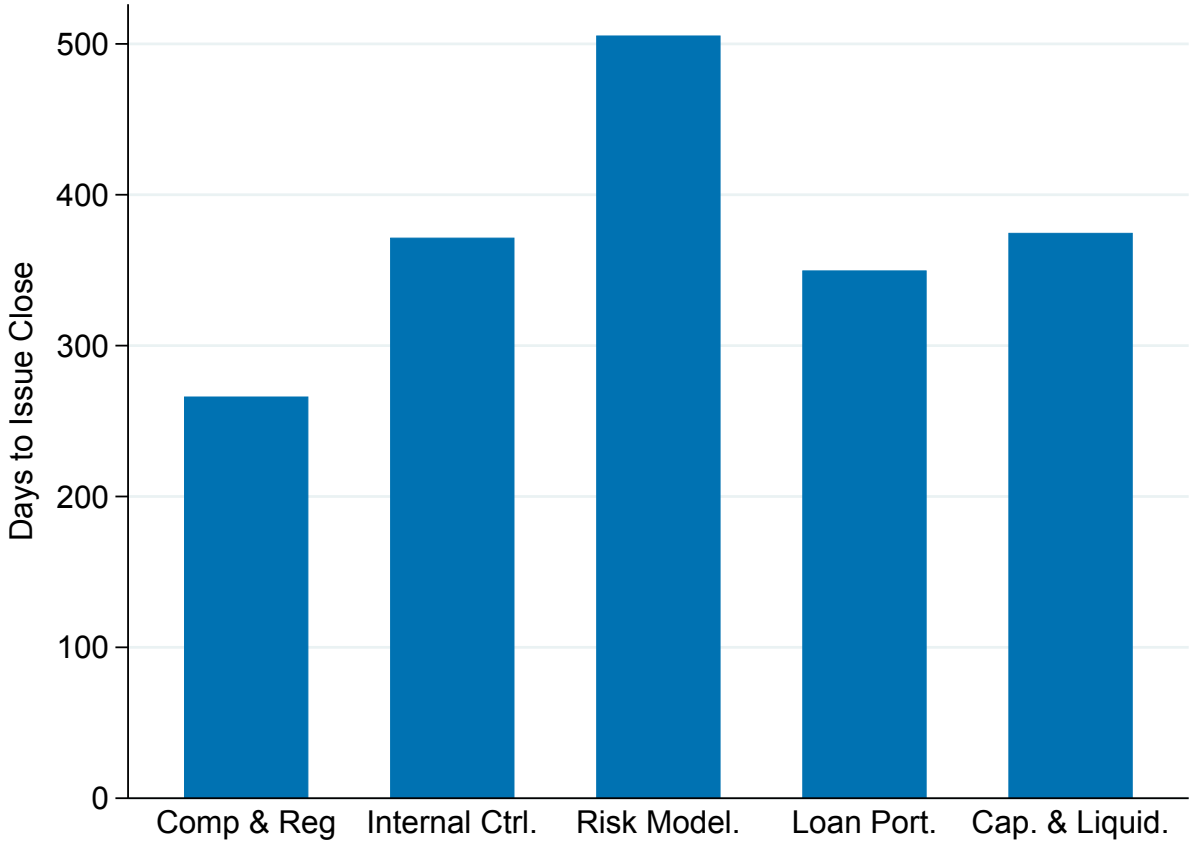
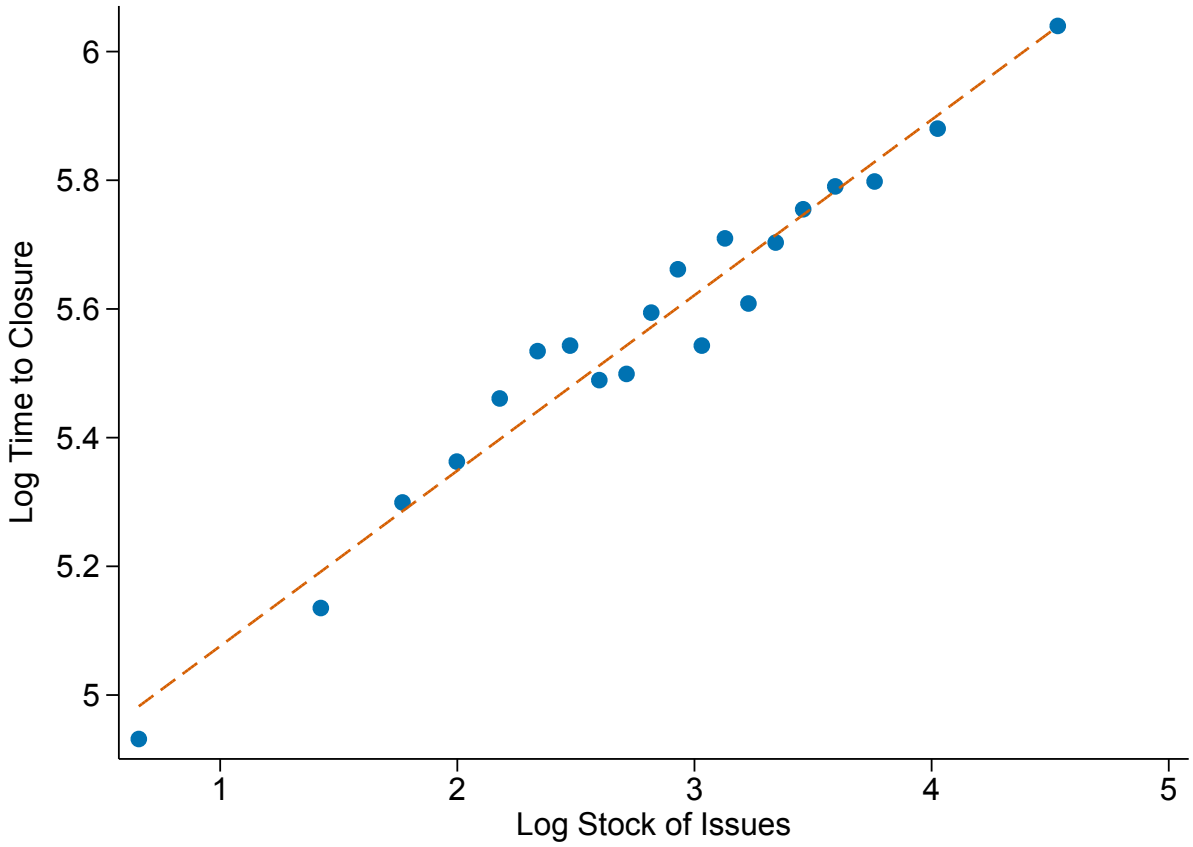


Figure 9: **Time to closure and stock of issues.** This graph plots the average time to completion for an issue against the stock of outstanding issues for the bank at issue opening. Each point represents the conditional mean for five perecent of the sample, and controls for year-month fixed effects.



A LDA methodology example

To make the issues in estimating the Latent Dirichlet Allocation (LDA) model concrete, we provide a stylized example. Consider the following vocabulary:

- {apple, apricot, abolish, accent, affliction, anchor, albino, archer}
- {beach, banana, bear, book, bias, blonde, bind, beer}
- {cat, center, chant, cycle, corn, cyclone, colony, cornbread}
- {deer, donut, dimple, dog, dynamite, demon, dwell, devious}
- {ear, egg, earnest, elephant, enzyme, elf, elderweiss, email}
- {fall, freedom, fondue, filler, fundamental, france, fix, forced}
- {gross, gab, glee, glitter, gloss, glib, gowanus, goose}
- {zoo, zing, zip, zulu}.

Assume that there are seven topics, with each topic being made up of the words starting with the letters a-g, respectively. The words beginning with z are noisy words that are included in all topics. These words can be considered filler, or "stop words", that may not reflect thematically interesting information in the categorization. Recall that the two main unobserved parameter vectors we are interested in are β, the distribution over the words given topics, and θ, the distribution of the topics over documents. In our stylized example, we define

$$\tilde{\beta} = \begin{pmatrix} \beta_a & 0_8 & \dots & 0_8 \\ 0_8 & \beta_b & \dots & 0_8 \\ & & \vdots & \\ 0_8 & 0_8 & \dots & \beta_g \\ \beta_z & \beta_z & \dots & \beta_z \end{pmatrix} + \Sigma, \tag{A.1}$$

where $\beta_a, \dots, \beta_g$ are 8×1 vectors of weights on each of the subset of vocabulary words beginning with a-g, β_z is a 4×1 vector of weights on the z words. Σ is a noise component added to each word. Then, the final β vector is a rescaled $\tilde{\beta}$ such that the column sums of β are equal to one. For the remaining components of the data-generating process, innocuous assumptions are made: α is considered to be symmetric and equal to the inverse of the number of topics. ξ is set so that the average number of words per document is 100. We define β and Σ such that within initial letter, the probability of a given word declines. The β_z is defined such that they are frequent and prominent, but equally so across topics, relative to the most frequent words in $\beta_a - \beta_g$.

We generate a sample dataset of documents with 10,000 documents. We then estimate the LDA model on this sample of documents. However, we are faced with the same considerations that we faced with the supervisory issues. First, given the set of documents with these words, the number of topics (here, seven) is not obvious *ex ante*. Second, given a set of topic labels for words, it is not immediately obvious what semantic label to give a topic. First consider

the second issue, and assume that we know there are seven topics: Figure A.1 displays all the words with their estimated model parameter values, similar to Figure 1. In this scenario, however, we have organized the words into their categories using our pre-existing knowledge, and not saliency. By noting the size of the circles, we see that if we attempted to label the topics by sorting the size of the parameter values (the circles), we would get many erroneous "z" words. It is possible that we might remove these values prior to estimation (as we do with stop words), but that requires knowledge beforehand. More usefully, we can instead sort the words based on saliency, as in Figure 1, and present the same table for the top 5 most salient words in each topic (sorted by saliency). In Figure A.2, we see that the "z" words are completely excluded, and the words sort perfectly into topics.[19]

To address the issue of *ex ante* knowledge of the number of topics, consider the same set of data, but now we estimate the model with four topics instead of seven. Our saliency output in this scenario is described in Figure A.3, and we can see that the topics are not correctly labeled nor organized correctly in this case. Clearly, choosing the right number of topics is important to have semantically meaningful categorizations. To formalize this selection process, we define a criterion function that trades off between saliency, defined in the text, and topic load:

$$\text{Load}_{ik} = \sum_{d=1}^{N} \underbrace{\theta_{dk}}_{\text{Doc } d \text{ Topic } k \text{ Weight}} \tag{A.2}$$

This load variable effectively measures how much weight a topic gets over the entire corpus. We are interested in finding a set of topics where there is a small set of salient words that are mostly assigned to a particular topic. As we increase the number of topics, two things occur. First, more words become salient, as they are able to stand out in a particular topic. In the limit, each word gets its own topic, and becomes perfectly salient. Second, as the number of topics increases, the load on a given topic will decline, as the loads are distributed across topics. We are interested in still providing "meaningful" topics.

Our proposal is to trade-off these two factors. First, we sum the saliency values for the top 5 (this can be varied) most salient words in each topic. We then weigh each topic saliency by the average load for that topic. This generates our weighted saliency measure:

$$\text{Weighted Saliency}_K = \sum_{i=1}^{K} \left(\text{Top Saliency Sum}_i\right) \times \text{Avg Load}_i \tag{A.3}$$

We can then plot this versus K, the number of topics for a given model. For our simulated data, this value is maximized at 7 in Figure A.4, the true number of topics. In our supervisory issue data, the value is maximized at 5 in Figure A.5.

[19] This perfect fit obviously decreases as we increase the amount of noise in the data generating process, and we pursue this exercise in (Goldsmith-Pinkham and Lucca, 2016)

Figure A.1: **Simulated data parameter values.** This figure presents the parameter values of the LDA model for the words in the simulated data described in Appendix A. The size of the circles in the left panel represents the estimated probability of a word, given the topic. The right panel presents the marginal distribution of each word. The words are sorted into the prespecified topics by the authors.

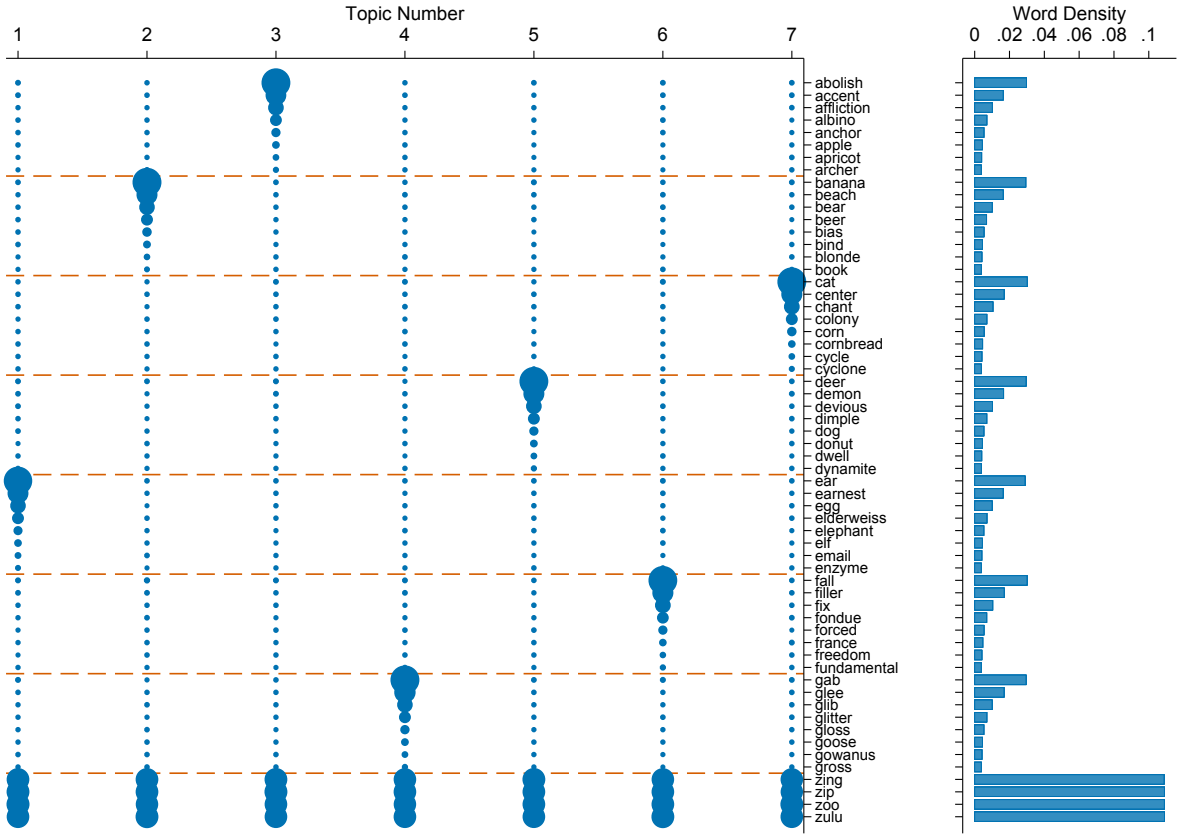


Figure A.2: **Labeling simulated data, correct number of topics.** This figure presents the parameter values of the LDA model for the five most salient words in each topic, estimated on the simulated data described in Appendix A. We estimated the model with seven topics, which is the true value in the data generating process. The size of the circles in the left panel represents the estimated probability of a word, given the topic. The right panel presents the marginal distribution of each word. The top five words are sorted by saliency within topic, then stacked.

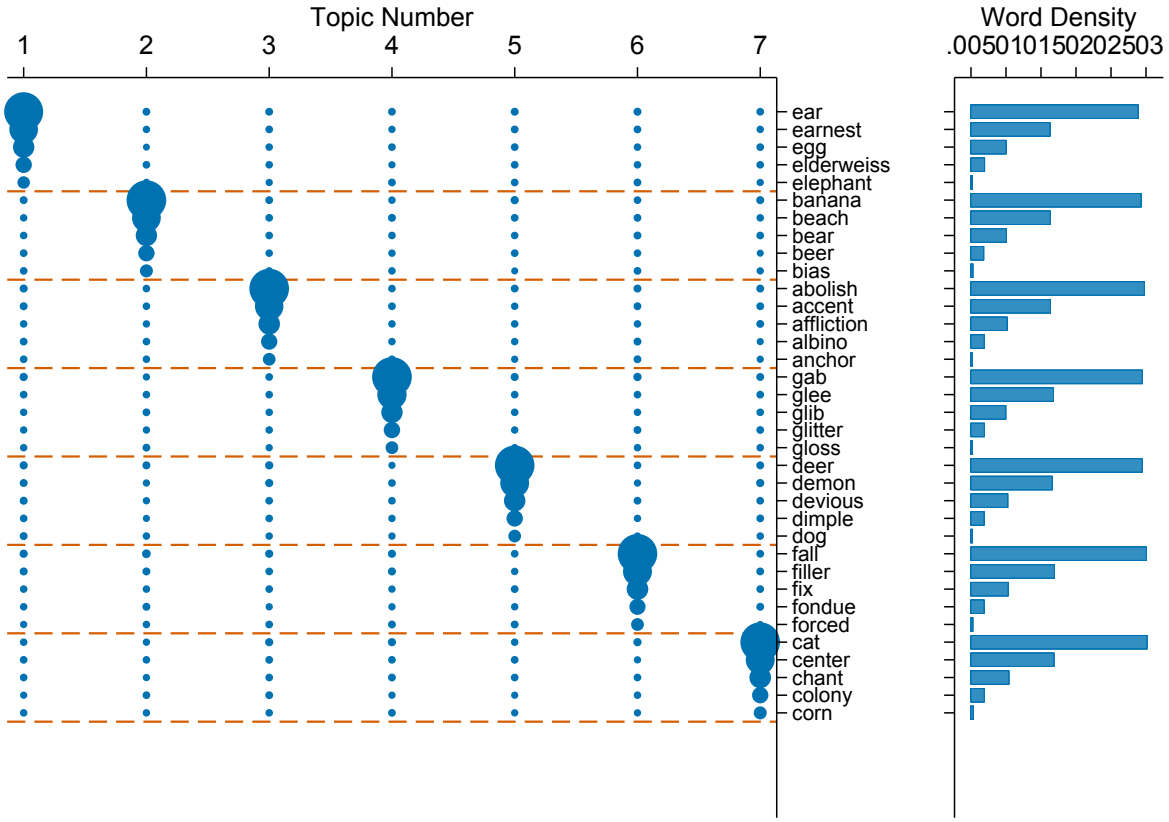


Figure A.3: **Labeling simulated data, incorrect number of topics.** This figure presents the parameter values of the LDA model for the five most salient words in each topic, estimated on the simulated data described in Appendix A. We estimated the model with four topics, which is different from the value in the data generating process. The size of the circles in the left panel represents the estimated probability of a word, given the topic. The right panel presents the marginal distribution of each word. The top five words are sorted by saliency within topic, then stacked.

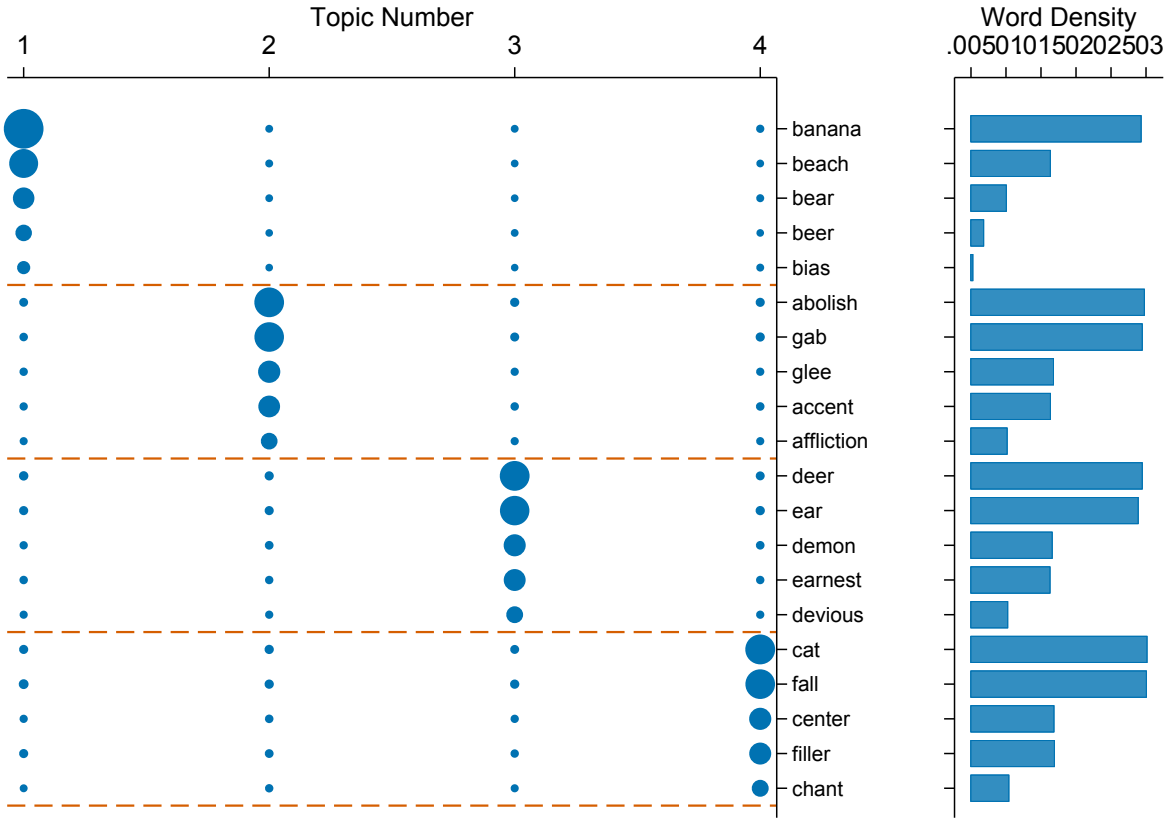


Figure A.4: **Criterion for choosing topics in simulated data.** This figure presents the criterion function estimated on the simulated data, and calculated using the top five most salient words. See Appendix A for more detail.

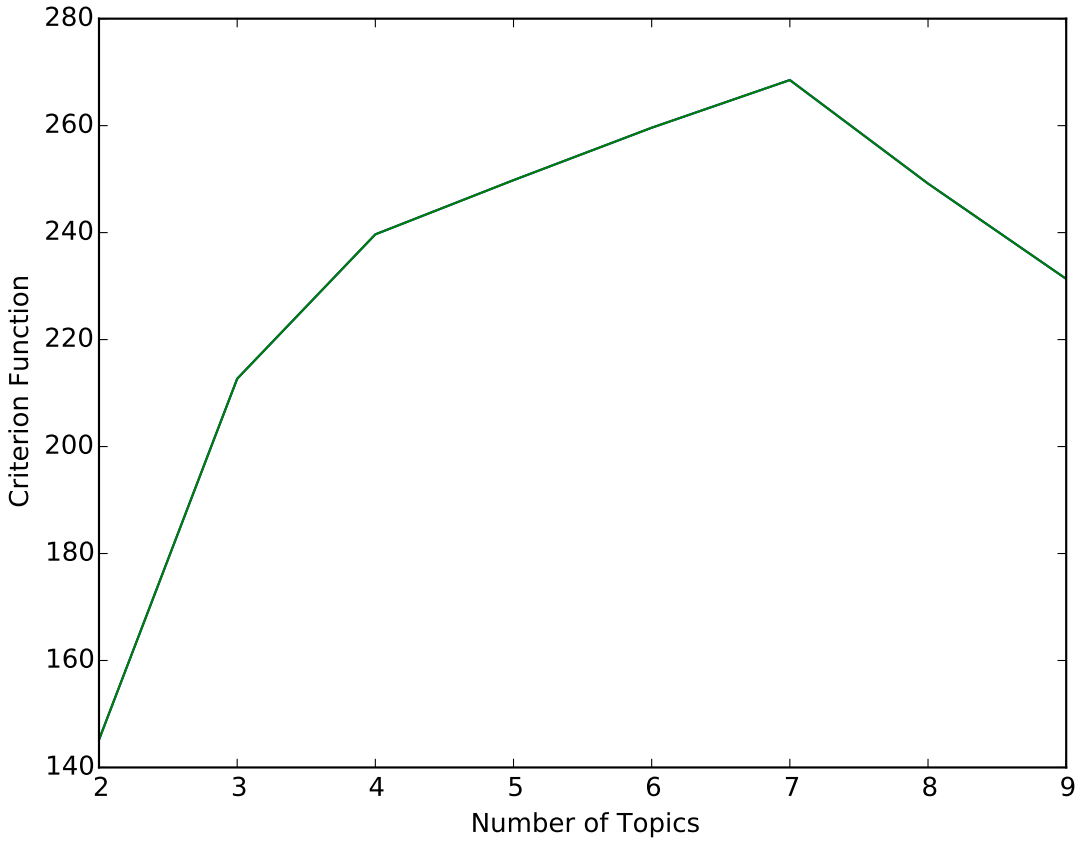


Figure A.5: **Criterion for choosing topics in supervisory issue data.** This figure presents the criterion function estimated on the issue data used in the paper, and calculated using the top five most salient words. See Appendix A for more detail.