Working Paper No. 15-42

Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements

Philip G. Berger
University of Chicago Booth School of Business

Michael Minnis
University of Chicago Booth School of Business

Andrew Sutherland
MIT Sloan School of Management

Published as "Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements." Berger, Phillip G., Michael Minnis, and Andrew Sutherland. *Journal of Accounting and Economics* Vol. 64, No. 2-3 (253-277). DOI: 10.1016/j.jacceco.2017.06.005

Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements*

Philip G. Berger
Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637

Michael Minnis
Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637

and

Andrew Sutherland[+]
MIT Sloan School of Management, 100 Main St., Cambridge, MA 02142

May 2017

The online appendix is available at the end of this manuscript.

* We appreciate helpful comments from Mary Barth, Anne Beatty (discussant), John Core, Joao Granja, Glen Jamarron (discussant), Anil Kashyap, Randall Kroszner, Alina Lerman, Petro Lisowsky, Maria Loumioti (discussant), William Minnis, Stephen Ryan, Nemit Shroff, Eric So, Philip Strahan (discussant), Rodrigo Verdi, Todd Vermilyea, Joe Weber, participants of workshops at Chicago, Cornell, Drexel, Miami, MIT, NYU, Stanford Accounting Summer Camp, UIC, Federal Reserve System/Conference of State Bank Supervisors Fourth Annual Community Banking Research and Policy Conference, UNC/Duke Fall Camp, the 2016 FARS Midyear Meeting, and the 2016 Journal of Accounting and Economics Conference, and especially an anonymous referee and Bob Holthausen (editor). Any remaining errors or omissions are ours. We thank the Risk Management Association for providing data. Minnis acknowledges support from the ARAMARK Faculty Research Fund at Booth, Sutherland thanks MIT for funding, and all authors appreciate financial support from the University of Chicago's Booth School of Business.
[+] Corresponding author: ags1@mit.edu; 617.324.3894

Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements

Abstract

Lending concentration features prominently in models of information acquisition by banks, but empirical evidence on its role is limited because banks rarely disclose details about their exposures or information collection. Using a dataset of bank-level commercial loan exposures, we find banks are less likely to collect audited financial statements from firms in industries and regions in which they have more exposure. These findings are stronger in settings in which adverse selection is acute and muted when the bank lacks experience with an exposure. Our results offer novel evidence on how bank characteristics are related to the type of financial information they use and support theoretical predictions suggesting portfolio concentration reveals a bank's relative expertise.

JEL Classification: G21, G38, M40, D82, L14.

Keywords: commercial lending; monitoring; information economics; lending concentration; financial statements; bank regulation; auditing; hard and soft information; theory of the firm.

1. Introduction

Theory suggests banks' demand for high quality financial reports varies with their lending strategies and internal information structures. For example, Stein (2002) predicts that larger intermediaries, with a need to transmit information to multiple agents within their hierarchical organizations, more frequently request verified reports from borrowers. Smaller financial intermediaries, which typically have fewer layers of internal information transfer, are conversely more likely to use soft information acquired through personal interactions with firms (Berger et al. 2005). In addition to bank size, theories suggest that exposure concentration may play an equally important role in shaping banks' interactions with firms (Dell'Ariccia et al. 1999; Winton 1999). These theories suggest a bank's concentration in a sector implies a degree of expertise: banks with more exposure to a sector have more interactions with borrowers and are thus more informed. Banks with less exposure, and thus fewer substitute sources of information, may demand more detailed and verified information when contracting with those borrowers, resulting in a negative relation between exposure concentration and high quality information demand.

A negative relation between the concentration of a bank's loan portfolio and the extent of information collection from borrowers is not obvious, however. Concentration in an industry or region not only increases information about that sector, but also amplifies portfolio risk. Collecting more reliable and precise information about concentrated exposures is a plausible strategy for managing the risk and related scrutiny from regulators, depositors, and the board of directors.[1] Thus, a positive association may emerge between a bank's lending concentration and its high quality information collection.

[1] For an example of both managing credit risks and mitigating regulatory scrutiny accompanying concentration, see the Office of the Comptroller of the Currency's discussion document concerning concentration risk in Commercial Real Estate lending (OCC 2006).

Understanding the relation between the concentration of a bank's portfolio and its information collection practices is of interest for several reasons. First, the banking market has consolidated considerably following more than two decades of mergers and bank failures. As a result, the loan market has become more concentrated and individual bank loan portfolios have become more diverse. Depending on the relation between portfolio concentration and information collection, these market shifts can affect the demand for auditing from the predominantly private commercial borrowers that choose whether to have audited financial statements, in part, as a function of lending relationships. Second, C&I loan exposures are rarely securitized, and banks typically retain their loan exposures after origination (Loutskina 2005, FDIC call reports). As a result, regulators expect banks to document more information collection from their larger positions to mitigate portfolio concentration risk (Basel 2000; OCC 2011). Requiring more collection of hard information about concentrated positions could, however, impede lending to opaque firms when a bank is willing to lend based on knowledge arising from its specialization in the borrower's sector. Despite a close link between exposure concentration and borrower monitoring techniques in theoretical models, empirical evidence establishing a link between concentration and information collection is sparse because banks disclose limited details about their exposures or information collection practices.

We use a bank-level dataset supplied by the Risk Management Association (RMA) to examine the relation between a bank's commercial lending concentration and its level of financial information quality requested of borrowers. The dataset includes the financial statement collection records and commercial loan exposures of banks representing a substantial portion of the U.S. commercial and industrial (C&I) loan market. RMA compiles financial reports collected by member banks and categorizes them according to report type as: unqualified audit, review, compilation,

tax return, or other. Unqualified audits provide the highest level of independent verification and the most financial information. Each category is tabulated by six-digit NAICS code, six borrower size groups, six regions, and the ten years 2002-2011. We use these data to measure the frequency of banks' audit collection from borrowers and their C&I exposure concentration within a bank-year. RMA confidentially provided these data at the bank level, allowing us to map lending exposures and statement collection records by bank to FDIC call reports.

Our central finding is that banks' audited statement collection is negatively related to portfolio concentration. We begin by using bank-year observations and find that banks with more concentrated commercial loan portfolios collect audited statements from borrowers less frequently, controlling for bank and borrower sizes and year fixed effects. Although these results reflect differences unrelated to bank or borrower size, they could be the result of omitted bank characteristics that are related to both concentration and financial report collection practices. Therefore, we use the panel structure of the data and bank-year and industry-region-year fixed effects to mitigate concerns about unobservables. We find banks collect audited statements at lower rates from borrowers in industry-regions in which they have more concentration. In our main specification, a one standard deviation increase in a bank's exposure to an industry reduces the rate of audited statement collection by approximately 2.4 percentage points, or about 19% of the unconditional mean audited statement collection rate in the sample.[2]

In additional tests using portfolio sorts of bank size, borrower size, and concentration, we compare the magnitude of our loan concentration finding to the more broadly studied characteristic

[2] Throughout the paper we use the term "exposure" to indicate the amount of activity that a bank engages in within a given industry-region. The dataset does not measure the dollar value of lending between banks and commercial borrowers, but instead measures the number of financial statements and total amount of sales of the firms from which the financial statements are collected. We use this information to calculate a bank's exposure to a given industry-region. In Section 3 we provide specific details of both the dataset and the variable construction and in Section 4, we demonstrate the robustness of our results to a number of alternative specifications.

of bank size. Prior work (e.g., Stein 2002, Berger et al. 2005; Liberti and Mian 2009) argues larger banks are more reliant on hard information. Consistent with this argument, we find larger banks are more likely to collect audited financial statements from borrowers compared to smaller banks, even after controlling for the average borrower size across banks. Moreover, our estimates suggest the difference in borrower audit rates across different concentration levels within bank is as large as the difference in borrower audit rates across banks of different sizes.

An explanation consistent with our evidence is that concentration fosters lending expertise. A bank with more exposure to an industry has better information about it and, thus, less need to obtain high quality (and costly) financial performance information from borrowers in the industry. By contrast, banks with less exposure to (and thus less information about) the same borrowers face an adverse selection concern, and thus demand high quality information to substitute for their relative lack of expertise. We conduct cross sectional tests to further examine whether exposure concentration builds expertise, which in turn reduces demand for audited statements. First, if concentration fosters bank expertise, our results should be stronger in settings with more severe adverse selection and information asymmetry concerns. We identify such settings using the extent of bank market competition within an industry-region and the performance dispersion of borrowers within an industry. Dell'Ariccia (2001) suggests that adverse selection concerns are most acute in imperfectly competitive banking markets, because potential entrant banks are particularly uninformed. In such situations, novice banks will demand high quality information from borrowers to offset their information disadvantage. Next, we argue performance dispersion makes it more difficult for banks with less exposure in those industries to interpret and utilize low quality information gathered from individual firms, which increases their demand for high quality reports. Thus, we expect to find stronger results in industries with more performance dispersion. We find

the negative relation between a bank's exposure to a given industry-region and its audited report collection from borrowers in that industry-region is stronger when either bank market competition is lower or borrower performance dispersion is higher.

Second, if concentration allows a bank to acquire information about a given sector over time, the negative relation should strengthen as the bank gains experience. To examine this, we perform cross sectional tests based on variation in the length of bank exposures. We use only the banks that are in the dataset for all ten years and count the years each bank has been lending to each industry-region as of the last year of data. We find a negative relation between experience and audit collection and that this negative relation increases in the bank's concentration in the given industry-region. Moreover, the negative relation between concentration and audit collection emerges only after the bank has been exposed to a sector for about four and half years, suggesting the beneficial information acquisition from concentration accumulates over time.

To further examine whether expertise through concentration is accumulated over time, and to ensure that the negative relation between concentration and audited report collection is not the result of banks lowering standards to attract new borrowers, we examine instances when banks first enter an industry-region. We find that when banks first lend to a sector, the audit rate for new exposures is 6.4% higher than for the bank's other contemporaneous exposures. We then show the incremental audit collection rate declines as the bank gains experience in the industry-region. This finding is consistent with banks facing an adverse selection problem when entering new markets and mitigating it by collecting audited statements; however, as banks better understand the new market the need for verified statements declines.

These results, however, do not address the concern that banks choose which new exposures they enter into. One example of how this concern could be a threat to our inferences is that banks

could first hire a sector expert and, only after having acquired sector expertise via the new employee, subsequently increase exposure to the sector while decreasing the audit collection rate. This explanation differs from ours, as our hypothesis is that a bank's increase in sector concentration leads to the expertise increase.

Although we do not have random assignment of banks to new exposures, we address the concern that banks choose exposures by using the housing boom of the 2000s as a setting where bank entry into a sector is plausibly driven less by endogenous bank strategy and more by demand-side factors. The housing boom increased loan demand from construction borrowers (Lisowsky et al. 2017) and we argue banks not exposed to construction borrowers in 2002 ("novice banks") gained exposure to construction firms by 2005, at least in part, because of the significant demand shock. We find novice banks indeed increase their relative exposure to construction lending substantially during the period. We next compare the information collection of novice banks to more established construction lenders in 2005 and find that novices are more likely to collect audited statements from construction firms. These tests help corroborate our finding that banks entering a new exposure are more likely to rely on audited financial statements.

Is either an audit verification strategy or an expertise strategy related to better performance? We find no indication either future loan charge-offs or future bank return on assets is related to audited statement collection, exposure concentration, or their interaction. This offers suggestive evidence that specialized banks are trading off alternative information sources rather than recklessly forgoing audit requests.[3] Given we observe performance at only the bank rather than the exposure level, we caution that our findings do not necessarily indicate banks are pursuing

[3] Somewhat similar trade-offs have been documented in prior literature with, for example, Demsetz and Strahan (1997) finding that more diversified banks do not have less risk than specialized banks, but instead offset the risk-reducing impact of their diversification by operating with greater leverage and larger commercial and industrial loan portfolios.

an optimal monitoring strategy.

Our study makes two contributions. First, it furthers our understanding of how organizational characteristics of a contracting party are related to the use of financial information by that party (e.g., Berger et al. 2005; Brickley et al. 2009). Our finding that bank concentration within an industry is negatively related to the use of higher quality financial reports is consistent with Paravisini et al.'s (2015) result that banks concentrating in particular Peruvian export markets possess expertise that shapes their lending decisions. We advance this line of work by showing how a bank's expertise relates to its screening and monitoring activities, and examining different channels through which banks use and develop expertise.

Second, we contribute to an emerging literature linking financial reporting to the characteristics of capital providers (e.g., Gormley et al. 2012; Lo 2014; Kalay 2015).[4] Our hypothesis is that banks have more substitute sources of knowledge about borrowers from industries in which they have more expertise and this reduces their demand for costly audits for such borrowers. Our results are consistent with this hypothesis, implying that one source of bank demand for verified financial reports is a lack of substitute knowledge from sources such as relationships or industry experience. Considering that the typical U.S. bank is much larger and less specialized following decades of consolidation, our findings are relevant to the literature examining factors affecting the use of financial reporting in capital allocation.

2. Prior literature and motivation

Seminal research by Diamond (1984) and Boyd and Prescott (1986) models banks as delegated monitors with an advantage in providing loans. A key force driving the net benefit of

[4] Our study also relates to a broader literature examining how intermediaries accumulate and employ expertise in the industries of their clients (e.g., Clement 1999; Cahan et al. 2008; Bills et al. 2015).

delegating private information collection about, and monitoring of, borrowers to a bank is the diversification of borrowers in the bank's portfolio. Diversification reduces the risks associated with idiosyncratic shocks and lowers the costs of monitoring the bank by depositors, thus increasing the bank's likelihood of meeting its obligations to depositors and allowing the bank to bear lending risks at a lower risk premium than individual investors. Moreover, such diversification need not arise exclusively from the mix of loans the bank originates, but can be bolstered when banks rebalance their loan portfolio exposures by buying and selling loans (Cebenoyan and Strahan 2004).

Traditional arguments based on these foundational papers suggest banks will avoid concentration risk. These arguments, however, presume banks are equally informed about all exposures in the economy. Winton (1999) argues that a bank's monitoring effectiveness is lower in new sectors. Dell'Ariccia et al. (1999) develop a model of bank competition in which banks accumulate information about borrowers over time. Incumbent banks become sector experts, creating information asymmetry between themselves and potential entrants. In both models, adverse selection makes it difficult for banks to perfectly diversify their lending portfolios. More broadly, a within-firm agency cost framework suggests specialized lenders are better able to collect information that is less verified and that such lenders also have more incentives to do so (Rajan 1992, Stein 2002, Berger et al. 2005, Liberti and Mian 2009).

Empirical research has only recently begun investigating the tensions between concentration risk and bank expertise. Acharya et al. (2006) use 105 Italian banks during 1993-1999 and find that, for high-risk (low-risk) banks, expansion by lending to new industries is associated with riskier loans and lower returns (marginally higher returns). Tabak et al. (2011) also examine banks' exposures by industry, finding increases in the scope of lending are associated with lower returns and higher bank default risk in a sample of Brazilian banks. Loutskina and Strahan (2011)

investigate geographic concentration in U.S. banks by studying mortgage lenders. They find that when such lenders specialize in a few markets they invest more in information collection than their more geographically diversified peers. As a result, concentrated mortgage lenders tilt their lending more toward the information-intensive non-conforming (jumbo) mortgages and toward high-risk borrowers. Liberti et al. (2017) show that lenders expand their geographic and sector exposures after sharing information with one another, and trace individual lenders' expansion decisions to their collateral expertise.[5] In addition to considerable theory and some empirical evidence on the tensions between bank concentration and information collection, these issues are of interest to policy makers and bank regulators. Regulators offer guidance cognizant of the tradeoffs between specialization and concentration risk:

> . . . due to a bank's trade area, geographic location or lack of access to economically diverse borrowers or counterparties, avoiding or reducing concentrations may be extremely difficult. In addition, banks may want to capitalize on their expertise in a particular industry or economic sector. [B]anks should not necessarily forego booking sound credits solely on the basis of concentration, [and] must be careful not to enter into transactions with borrowers or counterparties they do not know or engage in credit activities they do not fully understand simply for the sake of diversification. **(**Basel Committee on Banking Supervision 2000, Item 67).

The unsettled state of the policy debate surrounding how concentration risk should be managed (see Barth et al. 2004, Boyd and De Nicolo 2005, and Beck et al. 2006 for a review) aligns with the conflicting regulatory guidance. On one hand, regulators advise banks to diversify their exposures both geographically and by industry to avoid concentration risk. Conversely, regulatory documents detail various ways in which regulators expect banks to gather, store, and expertly assess information about borrowers and loan applicants to reduce adverse selection and moral hazard **(**OCC 2011, 2014; OIG 2012**).** These two aspects of regulatory advice for banks are in conflict if

[5] Banks accumulate borrower and industry expertise along several dimensions, including the ability to interpret financial information, assimilate soft information, and value and redeploy collateral.

adverse selection or monitoring costs tend to be higher when lending to a new region or sector. We thus contribute to the policy debate by measuring how banks trade off concentration with collection of high quality borrower information.

New exposures can present adverse selection and moral hazard costs. In part, these costs may arise due to banks having less expertise about borrowers from industries or regions in which they have less experience. If so, one way banks might lower adverse selection costs when expanding their industrial or geographic reach is by increasing their demand for audited financial reports, which are highly verified and contain more information than alternatives such as tax returns. Thus, especially for the generally non-public borrowers likely to constitute the overwhelming majority of loan applicants and borrowers, the bank's degree of expertise in a given exposure may have important implications for the demand for external audits.

Audits add direct costs (e.g., fees) and indirect costs (e.g., opportunity costs of time) relative to other report types, but create three benefits. First, audited information is more contractible because a third party has verified it (Watts and Zimmerman 1983). Second, audited statements have less reporting noise than financial statements that are merely reviewed or compiled by an external auditor, leading such statements to better reflect a firm's performance and health—i.e., audited statements are more informative (Minnis 2011). Third, relative to reports such as tax returns, audited statements contain all three financial statements and full footnote disclosure, and thus have more information. Despite these benefits, banks are also able to use other mechanisms such as collateral, past relationship history (or soft information), and credit scoring models in lieu of audited financial statements (Berger and Udell 2006; Cassar et al. 2015; Sutherland 2016). Collectively, we hypothesize that audited financial statements are more useful for a bank's loan monitoring and screening when it is less familiar with a borrower's industry or region, but because

audits are more costly, and substitute mechanisms are available, expert banks will use alternative report types more frequently.

3. Sample, data and variable construction

Our data come from the RMA's Annual Statement Studies. Each year, RMA compiles the financial statements gathered by member banks from commercial borrowers and loan applicants, and publishes summary statistics in its Annual Statement Studies. The Studies detail the number of statements collected from firms according to five mutually exclusive categories: unqualified audit, review, compilation, tax return, and other. Financial reports collected are further partitioned according to six-digit NAICS code, borrower size group (under $1M of revenue, $1M-$3M, $3M-$5M, $5M-$10M, $10M-$25M, and greater than $25M) and region (Northeast, Southeast, Central, South Central, North Central, and West).[6] While the publicly available studies report aggregate figures across all banks, RMA has confidentially provided us with the disaggregated data linked to the contributing banks. Appendix A provides additional details about the construction of the dataset.

These data allow us to observe the exposure and financial statement collection practices of a broad set of U.S. banks. In its annual Survey of Credit Underwriting Practices, the OCC notes financial reporting is a key element of underwriting, defining underwriting standards as "terms and conditions under which banks approve, extend, or renew credit such as *financial reporting*,

[6] Table A5 of the online appendix compares the distribution of financial reports by industry as reported in the RMA dataset to the overall economy, the firms in Compustat, and the firms in the Sageworks dataset (Minnis 2011). The distribution of firms in RMA is similar to that of the economy, except RMA has more manufacturing, wholesale trade, and real estate firms, and fewer services-related firms. Compared to Compustat, the RMA dataset has fewer firms in mining, utilities, manufacturing, and insurance (industries with firms that are typically large, regulated and have a higher propensity for being public) and more wholesale trade and real estate firms. The RMA distribution is most similar to the private firm dataset collected by Sageworks, with the exception that the Sageworks dataset (which is primarily generated from accounting firms) has a higher concentration of construction firms, which have a surety-driven demand for accounting services.

collateral requirements, repayment terms, pricing, and covenants" (OCC 2014, p. 2, emphasis added). Consistent with the overall population of firms, the vast majority of RMA sample firms are privately held, meaning that banks' requests for GAAP financial statements with unqualified audits are often costly, providing meaningful tension for our study.

Three features of the RMA data raise selection and measurement issues. First, while banks have to submit their firms' financial statements according to RMA's protocol if they participate, bank participation is voluntary. This is problematic if participants are not representative of the banking market or if their choice to participate is related to their monitoring and exposure strategies. Fortunately, we observe which banks participate each year and can compare them to those banks not participating. In Table A1 in the online appendix, we tabulate a variety of descriptive analyses related to bank participation. We do not find any significant differences in the amount of capital or financial performance (as measured by return on assets) between participants and non-participants, but do identify two differences. Not surprisingly, participating banks have more of their loan portfolios invested in commercial loans (relative to consumer or mortgage loans, for example) and the RMA dataset banks are larger than banks not in the dataset.

While ideally one would want randomly selected banks in the dataset to ensure generalizability, neither the commercial lending nor the size difference seems particularly problematic for our study. Banks with more commercial lending activity are the banks we are more interested in, so participation based on commercial lending activity seems innocuous. Consistent with the skewness in the population of bank assets, more than 40% of the total commercial and industrial loan balances during our sample period are held by the ten largest C&I lenders (FDIC call reports). In each year of our sample, at least eight of these top ten commercial lenders participate in the RMA dataset, except for two years in which seven (six) participate. Thus, because the U.S. banking

market is highly concentrated and large banks are overrepresented in our sample, our dataset includes a large portion of the commercial lending activity in the U.S. At the same time, the dataset includes a broad cross section of banks according to size, location, business model, and performance. Nevertheless, because participation is not random we interpret our results with caution as they may not be generalizable.[7]

Second, the financial statements collected by banks can be part of either the application or ongoing monitoring process, meaning some of the statements we use to measure exposures are from firms that were rejected or received loans elsewhere. Unfortunately, the RMA data do not detail whether the statements were provided by actual borrowers of the bank nor the dollar value of any loans made. That said, we are comforted by the strong correlation (0.74) between the cumulative borrower sales for the bank (which we use to calculate our exposure measure) and the size of the C&I loan portfolio reported by the bank to the FDIC. Moreover, banks develop industry expertise through both loans *and* information collection during screening, meaning statements collected during the application stage are informative about the bank's specialization. Ultimately, any noise that this data feature introduces will make it harder for us to find relations consistent with our predictions. A third and related issue is that we do not observe loan terms, which prevents us from exploring any link between exposure concentration and loan terms or loan-level outcomes

[7] Major lenders including Bank of America, JPMorgan Chase, SunTrust Banks, Wells Fargo, and Zions Corporation appear every year, and these five banks alone are responsible for nearly one-fifth of U.S. C&I lending by banks during the sample period according to call report data. The sample also contains other large banks that appear every year until they failed or were acquired (e.g., National City, Wachovia, and Washington Mutual). The top commercial lenders are also among the biggest small business lenders. Per call report data, the top ten overall commercial lenders hold one-third of the loans made below $1 million, suggesting sample selection toward bigger banks still covers a significant portion of small business lending. We report the top ten commercial lenders and their participation each year in Table A2 of the online appendix. While a significant portion of C&I lending is conducted by bigger banks, the sample also includes smaller banks. As we report in Table A1 of the online appendix, the 25th percentile of total assets for participating banks is $278 million.

(e.g., De Franco, et al. 2016). Related, we cannot identify syndicated loans in our sample or separate statements collected by lead arrangers from those collected by other syndicate participants.[8]

Table 1 provides a variety of statistics about the reports compiled by RMA and the banks that collect them. Panel A reports that RMA compiled almost 1.8 million financial reports collected by financial institutions between 2002 and 2011.[9] We eliminate 72,220 statements from twelve bank-years that have over 100 statements from firms with at least $25 million of revenue but zero unqualified audits, given these likely reflect data errors.[10] The total number of statements compiled by RMA generally increases over the sample period. While the portion of statements that are unqualified audits averages 21.3%, it is declining over our sample period.[11] In 2002, the 30,157 unqualified statements collected represent 23.4% of statements collected, whereas in 2011 the 40,130 unqualified statements constitute 19.5% of statements obtained. The South Central and North Central regions provide fewer statements than the remaining four regions. Finally, firms with sales in the $3-$5 million range provide fewer statements than other size groups, whereas those with sales in excess of $25 million contribute the most statements.

Table 1, Panel B summarizes bank-level reporting for the 728 financial institutions that provide at least one year of data during 2002-2011. The median bank contributes data to RMA for

[8] We conduct two robustness tests to assess whether the inclusion of statements from syndicated deals affects inferences. First, we eliminate every statement collected from firms with more than $25 million of revenue, a borrower segment that comprises the majority of syndicated transaction volume. Second, we exclude the 30 biggest banks in each year of our sample, which underwrite the majority of syndicated loans. In both cases, our main results maintain.
[9] We begin our sample in 2002 because this is the earliest year for which RMA has maintained the data electronically at the bank level, which allows us to merge the RMA data to bank-level data from call reports.
[10] Our results are unchanged if we retain these observations.
[11] Some readers may perceive the 21.3% average audit rate as unusually low for firms borrowing from banks; however, first note from Table 1 that 56% of the sample is derived from firms with less than $10 million in revenues and the vast majority of these firms do not receive an audit (Minnis 2011). For further assurance that the RMA data set is not unusual, we benchmark the RMA audit rate against the audit rates observed in two independent databases: Sageworks (firm-level data collected from accounting firms) and IRS tax returns (firm-level data for medium to large firms in which the IRS asks the firm if its financial statements are audited). In brief, the comparison reveals consistent audit rates across the three databases (if anything, the RMA audit rate is slightly higher than the audit rate of the Sageworks database, which may not be surprising since the RMA data set conditions on firms borrowing from banks). For more details of the benchmarking, please see Table A1 of the Supplemental Appendix to Lisowsky and Minnis (2016).

three years, in part because consolidation in the banking industry has eliminated banks during this period and in part because some banks do not provide data to the RMA for all possible consecutive years. The majority of statements come from large banks that participate regularly in the RMA Annual Statement Studies: 56.4% (89.1%) of sample statements come from banks that contribute every year (at least five years). Tables A3 and A4 in the online appendix tabulate additional details regarding bank participation conditional on the number of years in the dataset.

Table 2, Panel A provides statistics for our main variables at the bank-year level. Our main dependent variable of interest is the proportion of financial reports that a bank collects that are unqualified audits. We refer to this variable as *% Unqualified*. We focus on audited statements because they provide substantially more verification and information than the other report types. Though other RMA statement categories include reports with some level of verification, audits require positive assurance (in contrast to reviews which have negative assurance or tax returns with statistical or implicit monitoring). Audited statements also provide the most information (e.g., in contrast to tax returns, audited financial statements have a Statement of Cash Flows and complete footnotes). This additional verification and information is reflected in the significantly higher cost of audits relative to the other report types, providing revealed preference evidence of the incremental benefit of audited statements. However, our results are similar if we also consider reviewed statements as high quality financial information. An average of 12.8% of the financial statements collected have unqualified audits. This figure is well below the Table 1, Panel A statistic which reports that 21.3% of all statements in the database have unqualified audits. The difference occurs because bigger banks—which have bigger borrowers and higher audit rates—supply relatively more of the RMA data but the bank-year statistics in Table 2 are equally weighted.

Table 2 also reports that there is skewness in the data, consistent with the skewness in the

population of banks. The mean (median) bank-year has 527 (72) financial statements collected each year.[12] Cumulative borrower sales for the average bank is $161 billion and the average borrower firm size for the average bank-year is $279 million; however, both distributions are highly right-skewed, as the median bank's borrowers' cumulative sales volume is just over $2 billion and the average borrower size for the median bank is $21 million in sales.[13]

Our main analyses examine how the collection of financial statements with unqualified audits varies with the concentration of the bank's C&I loan exposures. The remaining Table 2, Panel A statistics summarize our measures of concentration. We first measure the overall level of a bank's C&I exposure concentration. The *# Unique Industry (Region) Exposures* measures, at the bank-year level, the number of unique OCC industries (regions) from which the bank collected financial statements.[14] The average bank-year has data from 36 industries and slightly more than 2 (out of a possible of 6) U.S. regions. Combining these two dimensions into the variable *# Unique Industry-Region Exposures* indicates that banks operate in approximately 52 industry-regions, on average. We use a standard measure of concentration, the Hirschman-Herfindahl index (*HHI*), to measure the overall extent of exposure concentration within a bank-year. *HHI* is equal to the sum of the squares of the relative exposures, where each relative exposure is the exposure as a fraction of the bank's total exposure. Deflating by bank rather than total industry exposure allows us to

[12] As the number of financial statements that the median bank collects is an interesting descriptive statistic, and potentially surprisingly small, it is worth putting this number into context. Table A1 in the online appendix shows the median bank has $62 million in C&I loans. Therefore, the 72 financial statements should be compared to the $62 million in loans, or about $860k per borrower, which is a reasonable C&I loan size for the median bank, given that the median bank has only $635 million in total assets and would limit its exposure to any particular borrower (Gup and Kolari 2005). This statistic reinforces the skewness in banking activity in the U.S., which is reflected in the data: big banks hold a very large portion of the C&I loans (to big *and* small borrowers) in the U.S.

[13] In the publicly available version of the RMA Annual Statement Studies, RMA truncates firms with more than $250 million in assets. These firms were not truncated in the data made available to us in order to provide more complete exposure detail.

[14] To examine industry exposure through a similar lens as bank regulators, we map three-digit NAICS industries into OCC industry definitions using Appendix A of OCC (2011). Roughly 80% of three-digit NAICS industries map directly into OCC industries. Also see Table A6 of the online appendix for a list of OCC industry groups and the frequency with which audited financial statements are collected across those groups.

disentangle the effects of bank size and concentration, and account for the fact that banks of all sizes accumulate more expertise in their larger exposures.[15] We calculate *HHI* using the industry-region level of data and measure exposure to an industry *i* in region *r* for bank *b* in year *t* by summing the sales of all firms providing financial statements to bank *b*:

$$HHI_{b,t} = \sum \left(\frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{i=1}^{I} \sum_{r=1}^{6} \sum_{f=1}^{F} Sales_{b,f,i,r,t}} \right)^2 . \quad (1)$$

The closer *HHI* is to one, the more the bank's commercial exposures are concentrated within fewer industry-regions. Panel B of Table 2 shows the time trend of both the *# of Unique Industry-Region Exposures* and *HHI* from 2002 to 2011. Overall, bank C&I exposure concentrations are decreasing, but the time trend differs considerably across bank sizes. The largest banks, which begin with very diverse portfolios, are becoming more concentrated over time. By contrast, the smaller banks (which begin with more concentrated C&I exposures) are becoming more diverse.

We next measure banks' exposures to each particular industry-region within a year. We sum the commercial loan exposure bank *b* has to all firms *f* in industry *i* in U.S. region *r* in year *t* and then divide this numerator by bank *b*'s total commercial loan exposure across all industries and regions in year *t*. This variable, which we refer to as *Share_bank*, measures bank *b*'s exposure to industry *i* in region *r* relative to all other commercial exposures in its *own portfolio* in year *t*:

$$Share\_bank_{b,r,i,t} = \frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{i=1}^{I} \sum_{r=1}^{6} \sum_{f=1}^{F} Sales_{b,f,i,r,t}} \quad (2)$$

[15] For robustness, we calculate a variety concentration measures, including a count of the number of industries required to accumulate 50% (66%, 75%) of a bank's C&I exposure, what portion of a bank's exposure the top 5 (10) industries constitute, and alternative specifications of the definition of industry. The results in our paper are not sensitive to the calculation of concentration and these specifications have been tabulated in Table A9 in the online appendix.

To ensure robustness of our results, we calculate a bank's exposure to an industry-region two alternative ways. First, we use the same numerator as *Share_bank*, but instead divide it by the total commercial loan exposure in the same industry, region, and year for all banks in the sample. This approach measures bank *b*'s exposure to industry *i* relative to *all other banks* in region *r*:

$$Share\_market_{b,r,i,t} = \frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{b=1}^{B}\sum_{f=1}^{F} Sales_{b,f,i,r,t}}. \quad (3)$$

Second, in the variable *Share_bank*, we calculate both the numerator and denominator using the sales of firms in each industry-region; however, our results are robust to using the number of financial statements collected instead. We refer to this alternative variable as *Share_statements* in Table 2.[16] Table 2 reports that the average bank has 6.1% of its C&I portfolio exposed to a given industry-region. Similar to the bank-level exposure concentration measures reported in Table 1, the exposure concentration statistics in Table 2 indicate that these industry-region-based measures also have right-tail skewness.

Collectively, these descriptive statistics highlight two important facts. First, there are extensive differences in financial statement collection practices across banks. Second, banks have significant heterogeneity in concentration levels across industry exposures. We now examine how these characteristics co-vary in our main tests.

[16] A mechanism underlying the hypothesis that concentration leads to expertise is that a bank gathers more information as it concentrates in an industry-region. It is not clear whether using the amount of borrower sales or the total number of financial statements collected is a better approach to measure a bank's lending exposure. Measuring exposure based on borrower sales has the advantage that the financial statements collected from larger firms may be more informative and therefore should receive more weight in the analysis. On the other hand, collecting financial statements from more firms could provide the bank with more information, regardless of firm size (i.e., each firm should get equal weight). All of our results are robust to either approach. One additional approach could be to measure the level of lending that the bank has in each exposure. We cannot use this measure because the dataset does not provide this information, but it is not clear this would be a superior metric. Banks receive information from firms that apply for loans but do not become bank borrowers, which our sales and financial statement measures identify, but a lending-based measure would not.

4. Tests and results

4.1 Concentration and audited statement collection across banks

We begin by examining the association between audited statement collection (*% Unqualified*) and portfolio concentration across banks. Table 3 presents our bank-level results using regressions of the following form:

$$\begin{aligned} \%Unqualified_{b,t} = {} & \beta_1 HHI_{b,t} + \beta_2 LogCumulativeBorrowerSales_{b,t} \\ & + \beta_3 LogAverageBorrowerSize_{b,t} + \gamma^t + \varepsilon_{b,t} \end{aligned} , \quad (4)$$

where the unit of observation is a bank-year. Each column includes a control for the bank's average borrower size that is calculated as the natural logarithm of the total sales for all of the bank's exposures divided by the number of statements collected for these exposures. Each regression also includes year fixed effects (γ^t) so that any secular trends will not affect our inferences (e.g., Lisowsky et al. 2017). To address concerns about serial correlation in our loan concentration measures, standard errors are clustered by bank.

In column (1) of Table 3, the coefficient on *HHI* is significantly negative, indicating that banks with more concentrated exposures collect audited financial statements less frequently. In column (2) we regress *% Unqualified* on *Log Bank Size* (equal to the log of total borrowers sales that year) and find a significantly positive coefficient estimate. This result is consistent with the prediction of Stein (2002) and findings of Berger et al. (2005) in which bigger banks are more likely to collect hard information (in this case, audited financial statements) even after controlling for the average borrower size. In column (3) we include both *HHI* and *Log Bank Size* and find that, while the magnitudes of both variables attenuate, both remain significant. Our results also remain after measuring bank size using bank size tercile indicators (column 4), and interacting these indicators with year fixed effects (column 5)

In summary, Table 3 provides evidence that loan exposure concentration has a strong negative association with audited statement collection even after controlling for the positive association that bank size has with audited statement collection.

4.2 Specialization and audited statement collection within-bank

Our findings in Table 3 that concentrated lenders collect a lower portion of audited financial statements is consistent with more specialization facilitating more substitute information acquisition by the lender. These results could, however, be driven by unobservable bank characteristics, such as bank-wide policies about what information loan officers must collect, or differences in audit rates across industries that are spuriously correlated with bank-level exposures. Therefore, we now examine the relation between exposure concentration and financial statement collection both across and within bank.

In Table 4, we first partition each bank's portfolio within bank-year into terciles based on the bank's exposure to a given industry-region. That is, each bank's portfolio is allocated evenly across the concentration terciles (i.e., one-third of the industry-regions of a given bank are in each tercile based on the variable *Share_statements* within a bank-year).[17] This is the "within bank" portion of the analysis. We then partition the banks based on size, according to the total borrower sales for each bank-year as reported by RMA. This is the "across bank" portion of the test. One-third of the banks are in each bank-size tercile. We tabulate the results of the two way sort (concentration tercile by bank size tercile) by borrower size group. This partition allows us to compare the rate of unqualified audit collection within borrower size group and bank-year, across industries in which the bank has a small, medium, or large portfolio share. We calculate the percentage of

[17] In untabulated robustness tests, we also form the concentration terciles based on sales weighting the observations, and the results are broadly consistent. The only difference is that there are few observations in the low concentration portfolios for the largest borrower size group, affecting the results in these portfolios.

financial statements that are unqualified audits in each portfolio.

The main messages from Table 4 are: (1) concentration is negatively related to audited financial statement collection, and, (2) the importance of concentration is of similar magnitude as bank size. This point can be seen by examining the bottom two sets of figures in Table 4. The section captioned "Bank size tercile differences" shows, within each of the six borrower size groupings, the difference in the percentage of collected statements with unqualified audits across the top versus bottom tercile of bank size. Consistent with theory (e.g., Stein 2002), audited statement collection is higher for larger banks for all six borrower size groupings.

The bottom section captioned "Concentration tercile differences within bank size tercile" summarizes how audited statement collection differs across exposure concentration terciles *within* a given bank size tercile and for a given borrower size, and reports the statistical significance for these differences. The difference is significantly negative in 15 cells (and insignificantly negative in the remaining three cells) in this bottom section, demonstrating that across all bank and borrower sizes the collection of audited financial statements is lower when the bank's exposure to the borrower's industry-region is higher.[18] Note that the absolute values of the figures in these bottom 18 cells match quite closely the magnitudes of the "Bank size tercile differences" immediately above. For example, for borrowers with less than $1 million of sales, the "bank size tercile difference" is 1.3% and the absolute values of the concentration tercile differences are 1.0%, 1.4%, and 1.7% for the three bank size terciles. Thus, the negative relation between exposure concentration and audited statement collection is of similar magnitude to the positive relation between bank size and

[18] Why, then, do any firms bother approaching banks less concentrated in their industry, when more concentrated banks are less likely to request a costly audit? Prior research points to several explanations. First, foregoing an audit makes the firm more opaque, which can expose it to hold-up problems (Rajan 1992). Second, providing an audit can reduce the interest cost of the loan and facilitate contracting on the firm's financial statements (Blackwell, Noland, and Winters 1998). Third, from the bank's perspective, at some exposure level the costs of concentration risk begin to outweigh the benefits of relying on economies of scale in screening and monitoring within the same industry (Diamond 1984; Boyd and Prescott 1986).

audited statement collection.

We now turn to Table 5. Our identifying variation in these tests is the exposure to different industry-regions within a given bank-year, thus the unit of observation is bank-industry-region-year. Each Table 5 regression includes a control for the average borrower size and clusters standard errors at the bank level.[19] The specification in Panel A, column (1) uses *Share_bank* as the concentration variable of interest and the regression includes separate indicators for each bank, region, industry, and year:

$$\begin{aligned} \%Unqualified_{b,r,i,t} = {} & \beta_1 Share\_bank_{b,r,i,t} + \beta_2 LogCumulativeBorrowerSales_{b,r,i,t} \\ & + \beta_3 LogAverageBorrowerSize_{b,r,i,t} + \alpha^b + \eta^r + \delta^i + \gamma^t + \varepsilon_{b,r,i,t} \end{aligned} . \quad (5)$$

Column (2) also uses *Share_bank* as the concentration variable, but uses a multidimensional fixed effect specification of bank-year and industry-region-year fixed effects:

$$\begin{aligned} \%Unqualified_{b,r,i,t} = {} & \beta_1 Share\_bank_{b,r,i,t} + \beta_2 LogCumulativeBorrowerSales_{b,r,i,t} \\ & + \beta_3 LogAverageBorrowerSize_{b,r,i,t} + \pi^{b,t} + \lambda^{i,r,t} + \varepsilon_{b,r,i,t} \end{aligned} . \quad (6)$$

Our goal in using these fixed effects is to isolate the relation between concentration and audited statement collection separately from other factors, such as bank-wide policies, differences in collateral across industries, or industry trends and regional shocks that affect demand for credit. Columns (3) and (4) have the same fixed effects specification as equation (6) but replace *Share_bank* with *Share_market* and *Share_statements* as the independent variables of interest. Column 5 examines *Share_bank* but uses bank-industry-region and year fixed effects.[20]

[19] We have too few years in our dataset to cluster by time. However, in untabulated robustness analyses of our main results, we find that two-way clustering by bank and industry or including only one randomly selected bank-year per bank has no effect on our inferences. Also, our main tests are repeated for one year (2011) only in Table 8, column 1.

[20] The variation in this specification comes from changes in a bank's exposure to an industry-region. Our hypothesis development is based on the premise that banks are strategic in the exposures they have to various industry-regions and in the nature of the information they collect from these exposures. If banks generally select their exposures strategically, the exposure of each bank to any given industry-region may tend to remain quite stable over time. Indeed, our main independent variable of interest (*Share_bank*) does not have much variation year-to-year, and the

The inference across all specifications is the same, namely that greater exposure within a bank to a given industry-region is associated with a significantly lower percentage of audited statements. The coefficient estimate on *Share_bank* is reduced modestly in column 5 relative to columns 1 and 2, which is unsurprising given the inclusion of bank-industry-region fixed effects in this final specification. Moreover, the estimates from Table 5, Panel A are economically significant. For example, the coefficient estimate of -0.271 in column (2) indicates that a one standard deviation increase in a bank-years' exposure to the industry-region (9.0%, see Table 2, Panel A) reduces the rate of audited statement collection in that industry by 2.4 percentage points. This represents a meaningful one-fifth reduction relative to the unconditional mean audit rate of 12.8%.[21]

Next, we perform two specification checks to validate our Panel A findings. First, we measure how audited statement collection for a given industry is associated with the bank's exposure to *related industries*. To do so, we construct an exposure variable *Share_related* that is based on our *Share_bank* measure, but differs in two respects. Specifically, we aggregate borrower sales at the OCC Group level rather than the OCC Industry level, and exclude borrowers in that particular industry. To illustrate, we calculate *Share_related* for the Fuel Distributors industry using the bank's exposures to Oil, Gas, & Coal Extraction and Petroleum & Coal Products Manufacturing—the non-Fuel Distributor industry members of the Oil, Gas, and Coal group as defined by the OCC.

variation that does exist is in the extreme tails (perhaps indicative of shocks or large strategy shifts). We thus caution against generalizing our results under this approach.

[21] As another way to consider economic magnitudes, we provide an illustrative example. In the western region in 2011 there were 60 banks with exposure to the industry "Merchant wholesalers, durable goods" (NAICS 423). The top five banks in the region in terms of their portfolio exposure had approximately 26% of their C&I portfolios in this industry-region, on average. The bottom five banks had a negligible portion (<0.1%) of their portfolios allocated to this industry-region. Our results suggest, ceteris paribus, that if a merchant wholesale borrower matched with an "expert" bank, its likelihood of providing an audited financial statement would be 7 percentage points lower (-0.271 x .26), or 56% lower than the unconditional mean audit rate of 12.5% for this industry-region in 2011. As the ceteris paribus conditions are not likely to hold (and other factors, such as bank size, borrower size, etc. could be further considered), this is only an illustration.

Our tests retain the firm size controls and fixed effects used in Panel A, column 2. Panel B, column 1 shows a significantly negative coefficient for *Share_related* that is smaller in magnitude than the Panel A, column 2 coefficient for *Share_bank*. This indicates that bank expertise can transfer across related industries, although a bank's industry expertise is more important than their related exposure expertise in explaining the reliance on non-audited statements.

Second, we consider how *Share_bank* is related to the collection of other statement types. If expertise reduces a bank's demand for high quality information, and audited statements are the highest quality report, then we should find *Share_bank* to have the most negative relation with audited statements. Columns 2-4 model the proportion of financial reports that a bank collects that are reviews, compilations, and tax returns, respectively. We find a significantly *positive* coefficient for *Share_bank* in all three tests. Moreover, the coefficient is decreasing in the degree of ex-ante verification provided—the larger the bank's exposure, the greater their collection of the least verified statements (tax returns). Together, our Panel A and B findings reinforce that bank expertise reduces the demand for audited financial statements.

Our theoretical framework suggests two cross sectional tests to reinforce our main inference. First, Dell'Ariccia et al. (1999) suggest that imperfect competition in banking markets arises in part from adverse selection problems. Their argument is that when more of a lending market is served by relatively few banks, those with little exposure (or potential entrant banks) face an adverse selection concern when assessing borrowers. Therefore, in less competitive banking markets the high exposure banks (which know more about the borrowers) will have less demand for verified information from borrowers, whereas low exposure banks will have greater demand for verified information to combat the adverse selection problem. In Table 6, panel A, we split each

industry-region based on its degree of banking competition.[22] We find the greatest sensitivity of verification standards to concentration in those markets with the lowest level of banking competition. Moreover, as we increase the threshold defining "low competition," the difference in the coefficient estimates of interest across the partitions becomes larger, although the differences are only significant at conventional levels for the most extreme partitions.

Second, if expertise is particularly important in industries where borrower performance is dispersed because information about any given firm is less useful for evaluating the creditworthiness of other firms in the same sector, then our results will be stronger where the performance dispersion of firms is higher. When firms within an industry perform significantly differently than one another, the financial performance of one firm is not particularly informative about the creditworthiness of another and, as a result, industry expertise is more crucial to screening and monitoring.[23] In Table 6, Panel B we partition the Table 5, Panel A results after characterizing industry-years as having low or high firm performance dispersion, where firm performance dispersion is based on the interquartile range (IQR) of return on assets (ROA).[24] In column 1 of the table, we split the data based on the median industry-year ROA IQR and find that the relation between a bank's concentration and audited financial statement collection is higher in the high dispersion industries. In columns 2 and 3 we partition the sample on increasingly higher thresholds of ROA

[22] Specifically, we use the RMA data and calculate a Herfindahl-Hirschman index for each industry-region-year. Note that in contrast to the *Share_bank* variable which calculates the relative portfolio holdings of an industry-region *within a bank*-year, the industry-region index we calculate here is the degree of lending concentration among banks within a given industry-region-year, similar to a standard industrial organization view of market concentration.

[23] The more intense relation between audit requests and concentration in industries with low dispersion is not obvious, however. For example, if firms within a given industry are very similar, then (as an extreme example) one audited financial statement may be sufficient for a bank to learn everything needed about all firms in the industry. As the bank's concentration then increases in the industry, the audit rate mechanically reduces because the marginal financial statements requested would not need to be audited financial statements.

[24] We use data from Compustat and calculate each firm-year's return on assets (defined as *NI* divided by *AT*). We then calculate the interquartile range of ROA for each industry-year. Industry-years with larger (smaller) IQRs are called "high (low) dispersion" industries.

IQR. We again find that the disparity in our coefficient estimates of interest becomes progressively larger, though the differences are not always statistically significant. These results suggest that the concentration-verification tradeoff is more acute in industries in which borrower performance may be difficult to discern for a bank with little exposure (i.e., expertise). In sum, these cross sectional tests reveal that when information asymmetry or adverse selection concerns are severe, the concentration-verification tradeoff is more apparent, consistent with the inference that concentration reveals a bank's expertise.[25]

4.3 Specialization and syndicated lending

In this section, we investigate whether bank expertise is relevant to lending decisions in the syndicated loan setting. Our goal is not to further explore how a bank's demand for audited financial statements is related to its concentration, but is instead to provide preliminary evidence on how bank concentration is related to lending behavior beyond that captured in our RMA data. The syndicated loan setting offers two key advantages for the purpose of further studying the link between expertise and lending decisions. First, the lead arrangers of syndicated loans perform a delegated monitoring role in collecting and using information about borrowers (Sufi 2007; Wittenberg-Moerman 2008; Ball et al. 2010). Similar to the banks in our main setting, lead arrangers develop expertise through repeated interactions with a similar group of firms, and this expertise can shape their contracting decisions. Second, because firm-level data on syndicated loans are available, we are able to explore contracting decisions that are relevant to furthering our understanding of the role of expertise, but that are not observable in the RMA data.

[25] To ensure the robustness of our results, we tabulate three additional cross sectional tests. In Tables A10, A11, and A12 of the online appendix we partition the sample based on bank size, portion of lending that is to C&I borrowers, and amount of regulatory capital, respectively. We find a significantly negative association between portfolio concentration and unqualified audit collection across all of the partitions, except banks within the smallest size tercile, which have small C&I portfolios and contribute few statements.

To investigate how banks use their expertise in the syndicated loan setting, we collect data from DealScan on syndicated loans originated between 2002 and 2011.[26] We construct our sample using the intersection of the RMA and DealScan datasets. We identify 17,745 loans made by lead arrangers participating in RMA in the year of loan origination. We access data on firm size and ownership type (age and R&D spending; rating coverage) from DealScan (Compustat; Capital IQ).

We then explore two ways in which banks might use their expertise in the syndicated loan setting. First, we examine whether opaque firms are more likely to borrow from a lead arranger with a concentration in their industry. Our proxies for opacity include the firm's sales, age, R&D spending, ownership type, and whether the firm is rated. Our prediction is that banks with greater concentration in an industry possess the expertise necessary to monitor more opaque borrowers, and have sufficient reputation to attract other lenders to participate in the loan.

Table 7, Panel A regresses our opacity proxies on *Share_statements*, measured at the bank-industry-year level, and bank, industry, and year fixed effects.[27] We find a negative but insignificant relation between log sales for the firm and the bank's exposure to the firm's industry. We also do not find an association between firm age and bank expertise. On the other hand, we find industry expert banks are more likely to lend to firms with R&D spending and to privately held firms. Last, we find the bank's industry expertise is negatively related to the probability of arranging a loan to an unrated firm.

Our second set of tests in the syndicated loan setting studies whether banks more frequently lead deals in areas of their concentration. Our analysis models the share of the bank's syndicated

[26] We thank Maria Loumioti for sharing her syndicated loan data and providing assistance. Because many of the syndicated loans are made to private firms, data limitations prevent us from presenting a constant sample.
[27] Using a statement-based rather than a dollar-based share exposure measure avoids generating a mechanical relation between loan size (Dealscan loans are very large) and the exposure measure. Note, we also omit industry fixed effects from the R&D spending tests, given much of the variation in spending is across, rather than within, industry.

loans as a lead arranger that were allocated to a given industry as a function of *Share_statements*, both measured over 2002-2011. The unit of observation is bank-industry. Our initial sample contains 998 observations involving 53 unique banks with RMA data acting as lead arrangers between 2002 and 2011.[28] Panel B shows a significantly positive correlation between syndicated loan allocations and *Share_statements*. Economically, a one standard deviation increase in *Share_statements* is associated with a 1.4% increase in syndicated loan exposure to that industry, representing almost half of the mean syndicated loan exposure. Column 2 restricts the sample to the 43 banks leading at least ten syndicated loans across all industries between 2002 and 2011, and finds slightly weaker and marginally insignificant results.

In sum, we find mixed support for banks using their industry expertise, as measured in our RMA setting, in making their syndicated loan decisions. On one hand, we find significant overlap between a bank's small commercial loan activity in an industry and its arrangement of syndicated loans to firms in that industry. We also find private firms and those with R&D expenditures are more likely to borrow from a lead arranger with more exposure to their industry. On the other hand, we find exposure is not related to syndicated borrower size or age, and is *negatively* related to syndicated lending to unrated firms. Our exploration of the role of bank expertise in the syndicated loan market suggests several opportunities for future research. First, our tests center only on whether banks act as lead arrangers to certain types of firms. Because of data limitations, we do not consider how industry expertise influences the structure of the syndicate, the share held by the lead bank, or the extent of securitization. Second, because most syndicated lending is conducted by a limited number of banks, the sample size is small for our tests in this setting. As coverage in

[28] Our sample of banks is limited by two factors. First, large banks dominate the syndicated market during our sample period. Second, participation in RMA is voluntary; several large banks do not participate every year, and a number of others (e.g., ABN AMRO, Deutsche Bank, Royal Bank of Scotland) are based outside the U.S. and are therefore not covered by RMA.

these datasets improves, or as other expertise measures are developed (e.g., De Franco et al. 2016), additional dimensions and consequences of bank expertise can be investigated.

4.4 Audited statement collection and bank experience

In the previous section, we examined variation in the rate of unqualified audit collection for a given industry as a function of the bank's exposure to that industry. We argue these findings support the view that concentration reflects a bank's opportunity to accumulate expertise, which reduces the benefit of audited financial statements. If this argument has merit, we should find the relation between exposure size and audit rates varies as banks accumulate experience in the exposure. In this section we consider how a bank's lending experience to a sector over time is related to expertise as reflected in the audit requirements the bank imposes. We are unaware of prior evidence about how experience combines with the number of firms falling within a shared grouping to shape a lender's expertise about firms within the grouping. A contemporaneous paper that touches on a similar issue is Shroff et al. (2016), which finds that the number of peers a private firm has affects its initial interest rate for newly issued public debt.

We begin by measuring a bank's experience by counting the number of years it has contributed data to RMA from each industry-region. To mitigate noise in the measure, we use only banks that contribute data across all 10 years and examine only the final year of the sample.[29] A bank's average industry-region exposure has been in its portfolio for seven years. In column 1 of Table 8, we replicate our main result from Table 5, Panel A, column 2 but with our more restrictive sample from 2011 and continue to find a negative coefficient estimate on *Share_bank*. In column 2 we then include the time series variable, *Years Experience*, and find that the coefficient estimate

[29] The results are not sensitive to the choice of years. In untabulated results, we ease the 2011 restriction and include banks from 2009 to 2011. We also ease the restriction on the minimum number of years in the RMA dataset to six or eight. In either case, the results remain similar.

on *Share_bank* slightly attenuates from column 1 and a negative coefficient on *Years Experience*, suggesting that as banks spend more time in a given industry-region, audited report demand decreases. To provide some context on the economic magnitudes, the coefficient estimate on *Bank_share* from this table suggests that a one standard deviation change in *Bank_share* is associated with a 1.8 percentage point change (-0.2 x 9%) in the audit rate. This compares to about 2.6 additional years of experience in the sector (divide 0.018 by the coefficient estimate on *Years Experience* of 0.007). In column 3, we then interact *Share_bank* with *Years Experience* and find that the coefficient estimate on *Share_bank* is no longer significant (and in fact becomes weakly positive), but that the interaction term is significantly negative. Dividing the *Share_bank* main coefficient estimate by the interaction coefficient estimate indicates that after 4.5 years the relation between concentration and unqualified audits becomes negative, suggesting that only over time does concentration lead to expertise. Moreover, these results suggest that each year of experience provides more expertise when banks have more concentration in the industry-region.

To further investigate how quickly banks accumulate expertise within an exposure concentration, we examine banks' behavior in, and after, the year of entering new industry-regions. To implement this analysis, we create an indicator variable *New Exposure This Year* for industries the bank was not exposed to in the previous year *and* which are one of the top ten exposures for the bank-region in the current year. We restrict our attention to new exposures above this threshold to ensure we are measuring cases where the bank is making a meaningful effort to enter the industry. We omit observations that are in the first year that a bank appears in the RMA dataset because we have no way of knowing what exposures are new versus preexisting. Using this approach, the indicator for *New Exposure This Year* equals one for 4.9% of our observations. Our Table 9 specifications control for firm size, cluster standard errors at the bank level, and include bank-year and

industry-region-year fixed effects. The results indicate that when a bank takes on a meaningful exposure for the first time, collection of unqualified financial statements from firms in that new exposure is significantly higher than otherwise. Controlling for bank-year and industry-region-year effects, the average audit rate for new exposures is 6.4% higher than the same bank's other contemporaneous exposures.

Next, in columns 2-4, we examine how the audit rates for new exposures evolve over time. We find that after two years, the incremental audit rate for new exposures is only 5.0%, and after four (eight) years is only 4.4% (-0.1% and not significant). These results are thus consistent with Table 8 which indicates a significantly negative relation between lending experience and audited financial statement collection. As important, these results do not provide support for the hypothesis that banks lower audit standards to increase exposure, which would produce a spurious negative correlation between exposure and audit collection.[30]

5. Addressing Concerns that Banks Choose Exposures – the Construction Boom

The tests to this point control for bank-year and industry-region-year fixed effects, so our findings cannot be a simple artifact of bank policy, bank condition, differences in audit rates across sectors, or industry trends. However, the tests do not address concerns that banks choose which exposures to specialize in and that borrowers may know whether a bank specializes in their industry. This could be a concern, for example, if banks hire a sector expert and, as a result, concentration increases and audit rates subsequently decline.[31] Although we do not have random assignment

[30] Because we do not have loan level data, one hypothesis we cannot disentangle in this test is whether the negative relation is caused by banks learning about specific borrowers within an industry or about the industry more generally. However, our cross sectional tests using data at the OCC Group level in Table 5, Panel B suggest banks are learning about the industry more generally.

[31] Our hypothesis is that concentration leads to expertise, not that banks hire experts who then increase concentration and reduce audit collection. Therefore, although we expect banks to have experts in sectors where they have their portfolios concentrated, we are trying to rule out that our results are primarily driven by experts arriving first and then producing concentration.

of banks to new exposures, we address the concern that banks choose exposures by using the housing market boom of the early 2000s as a setting where exposure changes are plausibly due to a demand shock rather than supply-side shifts by banks. A key strength of this approach, which is widely used in labor economics (see Bartik 1991; Blanchard and Katz 1992), is that it exploits pre-existing cross-sectional differences in industry exposure by interacting them with national or regional shocks or trends.[32] We argue that a shift in the supply of financing for consumers to purchase houses led to a demand shift in financing from construction firms to build them. This construction demand shift then led banks with little or no initial construction exposure to enter construction lending without necessarily planning to do so. We thus compare the audit collection rates of "novice" banks that entered construction lending as a result of the boom to more seasoned banks that had been exposed to construction all along.

While our approach does not address all concerns related to the endogenous strategies of banks, there are two reasons it helps. First, our argument is that at least some portion of the housing boom change in construction loan exposure was a result of construction firms applying for credit from new lenders, as opposed to ex-ante strategizing by bank management to enter construction lending. Second, the regional pattern of the housing boom allows us to confirm that any exposure increases are happening where we expect, and to conduct within-bank tests that rule out bank-wide policies driving our results. While we sacrifice generalizability, the above features help to address endogeneity concerns.

We first establish that the housing shock shifted the exposure of novice banks. Table 10, Panel A examines whether banks with little construction exposure in 2002 increase their exposure

[32] Bartik (1991) develops a method of isolating local labor demand changes. The "Bartik Instrument" averages national employment growth across industries using local industry employment shares as weights to produce a measure of local labor demand that is orthogonal to changes in local labor supply.

by 2005 differentially by region.[33] We choose 2002 because it is our first sample year and when investment in housing began increasing above historical levels (US Bureau of Economic Analysis), and 2005 to allow time for novice banks to enter construction lending but not so much time that they become experts. Consistent with our prior tests, bank exposures are measured at the regional level to account for variation in industry concentration throughout the country and differences in banks' expertise across regions. We use the indicator *Hot Region* to separate areas where the boom was more dramatic, by setting this variable equal to one for the Southeast and West regions and to zero otherwise. We restrict the Panel A sample to observations from 2002 and 2005. Because we are interested in whether banks with little or no construction exposure enter during the housing boom, we restrict Panel A observations to those where construction was not a top 10 exposure for the bank-region in 2002 and label such observations as construction *Novices* in the Panel B tests. The remaining sample restrictions in Panel A are described below.

In the first three columns of Panel A, the dependent variable is the share of each industry in the bank-year-region's portfolio. The main effects are the indicator for whether or not the year is 2005 and the *Hot Region* indicator. The remaining independent variables are the two-way interactions *Construction * Hot Region, Hot Region * Year 2005*, and *Construction * Year 2005*, and the variable of main interest, the three-way interaction *Year 2005 * Hot Region * Construction*, predicted to be positive. In columns (1) and (2) we impose no additional sample restrictions, with the column (1) regression including industry fixed effects and column (2) adding bank fixed effects. Column (3) also uses industry and bank fixed effects, but imposes the additional sample restriction that included banks must have exposure to both hot regions and not hot regions. All three columns produce positive coefficient estimates on both the interaction term *Construction **

[33] Construction exposures involve firms in the following OCC Industries: Homebuilding, Nonresidential Building Contractors, Heavy and Civil Engineering Construction, and Specialty Trade Contractors.

Year 2005 and the triple interaction term. Thus, not only is the housing boom associated with novice banks increasing their exposure to construction in 2005 over 2002, but the extent to which this occurs is at least twice as great in the hotter housing boom regions of the Southeast and West.

In columns (4) and (5) of Panel A, we further restrict the sample to include only construction industry observations (and correspondingly change the dependent variable to be the share of the construction industry in the bank-year region's loan portfolio), cluster standard errors at the bank level, and include bank fixed effects. In column (4), the coefficient estimate on *Hot Region * Year 2005* is positive, consistent with banks increasing their exposure to construction by more in 2005 for loan portfolios located in the two hot regions (however, the estimate is not significant at the .10 level). In column (5), when the bank-region-year observations are further restricted to those for only banks with loan portfolios in both hot and not hot regions, the coefficient estimate on *Hot Region * Year 2005* increases by 30% and becomes marginally statistically significant at the .10 level. In sum, banks increase their exposure to construction lending in response to the housing boom precisely where the growth is greatest, consistent with our prediction.

Next, in Panel B, we test the impact of demand-driven shocks to loan exposures on banks' collection of audited financials. The two regressions in this panel compare unqualified audit collection in 2005 for bank-region-industries based on whether or not the industry is construction, whether or not the region is one of the two hot regions for the housing boom, and whether or not the bank-region is a novice to construction lending (based on its 2002 exposure). As in prior regressions using *% Unqualified* as the dependent variable, the regressions also control for firm size. Both regressions include industry and bank fixed effects and cluster standard errors by industry. The column (2) regression differs from that in column (1) by further restricting the sample to include only banks with any industry exposure in both hot and not hot regions.

The coefficient estimate of main interest is that on the triple interaction term *Hot Region * Construction * Novice*. This estimate is significantly positive in both regressions, consistent with audited financial statement collection being higher among banks that entered construction lending in a hot region if the bank is a novice to construction lending in that region.[34] Although the positive estimate is more statistically significant in column (1), the estimates do not materially differ across the column (1) and (2) regressions.[35]

Our results in Panels A and B show that "novice" banks responded to the housing boom by increasing their construction exposure, and entered this exposure using more highly verified financial statements than experienced construction lenders. This corroborates our main finding that banks with greater exposure to an industry are less likely to collect audited statements from firms in that industry.

6. Is Future Performance Related to Specialization and Audited Statement Collection?

Having established that a bank's portfolio concentration and information collection strategies are related, a natural question is whether these strategies are associated with performance. Table 11 provides evidence on how two key dimensions of bank performance (loan charge-offs and ROA) vary with loan concentration, percentage of statement collection that uses audited financials, and the interaction between these two variables.[36] We measure the dependent variables

[34] Because our goal in these tests is to compare audit rates for novice and experienced banks in hot regions for construction industry firms, we focus on the triple interaction term. Note, however, that the results for the noninteracted and double interacted terms in Table 10, Panel B are consistent with previous findings. Specifically, the coefficient on the *Novice* variable is significantly positive, suggesting that inexperienced banks collect audits at higher rates, unconditional on the region or industry, consistent with our results in Table 9. Moreover, the interaction term *Hot Region * Construction* is significantly negative, consistent with the findings of Lisowsky et al. (2017) that banks had lower audit rates for construction firms in fast growing construction regions relative to other regions within the U.S., on average, unconditional on bank experience.

[35] Moreover, we note that by restricting the sample to banks with exposure in hot and not hot regions, the column 2 tests exclude many novice banks which operate in a single region.

[36] We follow Loutskina and Strahan (2011) by including controls for log Total Assets, and the ratios of Securities, Deposit Interest, Capital, Net Income, Deposits, Real Estate Loans, C&I Loans, Unused Loan Commitments, and

one year after the measurement of the independent variables. Also, because we include an interaction between *HHI* and *%Unqualified*, we demean both variables to facilitate interpretation of the coefficient estimates (i.e., the marginal effect of one variable is conditional on the mean of the other variable). The loan charge-off dependent variable has the merit of being tied solely to loan performance, but omits other key aspects of performance related to the lending business, such as fee-based revenues. The return on assets (*ROA*) dependent variable includes all aspects of performance related to the bank-year, including any portion of performance unrelated to the commercial lending business.

Although our focus is on the interaction term, the main effects are also of interest. The extent of concentration of the bank-year's loans is not significantly associated with either next year's charge-offs or next year's ROA. These results contrast somewhat with Loutskina and Strahan's (2011) finding that more geographically concentrated lenders have higher profitability. They argue their finding is consistent with geographically concentrated lenders facing decreasing returns to scale due to their need to use soft knowledge in screening and monitoring (although their paper does not allow direct observation of the information their sample banks collect). We caution that, like Loutskina and Strahan, we observe charge-offs and profitability at the bank-year level and not at the loan exposure level. We also find that the bank-year's percentage of statements collected that have unqualified audits has no association with next year's loan charge-offs or ROA. Thus, after holding constant the effect of our other variables, collection of audited financials from borrowers and loan applicants is not associated with future performance of the loan portfolio or the bank overall.

The coefficient estimate for the interaction term is insignificantly different from zero in

Letters of Credit to Total Assets, but do not tabulate the coefficient estimates for brevity. Due to requiring consecutive years of RMA and FDIC call report data for a given bank, the sample size is 2,489 bank-years.

both regressions. Thus, the performance of concentrated lenders is, on average, no more or less sensitive to collection of audited financials. We arrive at the same conclusion if we conduct our tests separately for pre and post-crisis years. This is consistent with variation in exposure concentration across banks (and variation in audited statement collection) representing equilibrium choices rather than variation in how close to optimal some banks' choices are. Nevertheless, we cannot draw strong conclusions from our results for two reasons. First, because our dependent variables are measured at the bank-year level, we cannot directly tie loan losses to the degree of concentration and audit rates for specific exposures. Second, uncovering a non-relation between losses, exposure size, and audit rates *on average* does not preclude the possibility that losses are highly sensitive to lending standards in specific situations. For example, research surrounding the recent crisis indicates large exposures to the housing market, coupled with low screening standards, resulted in large losses (Keys et al. 2010; Dell'Ariccia et al. 2012; Lisowsky et al. 2017).

7. Conclusion

Our paper examines the relation between commercial lending portfolio concentration and audited financial statement collection at banks. Our analysis builds on well-developed theory relating bank characteristics to external information collection. There has not previously been strong empirical evidence based on this well-developed theoretical work due to lack of available data. We find the propensity across banks to collect audited financial statements from borrowers is lower for banks with more concentrated commercial loan portfolios. Within bank, we find more frequent collection of audited financial statements from borrowers in industries to which the bank has less exposure. We find the economic magnitude of this relation is comparable to that between the more frequently studied bank size variable and audited statement collection, suggesting that concentration is an important characteristic that affects information collection practices of banks. Finally,

we find banks collect audited financial statements from their borrowers significantly more frequently when they enter new exposures. Collectively, our results support the joint hypothesis that the concentration of bank exposures is related to the expertise of the bank and that this expertise substitutes for high quality information, such as audited financial statements.

Our study offers novel evidence on how organizational features interact with the use of financial information. Although we examine banks, other intermediaries including institutional investors and analysts acquire expertise through repeated interactions with firms and our results are potentially relevant to understanding the demand for high quality financial information in such settings. In addition, while we find some evidence of the role of expertise in the syndicated lending market, future research with more detailed data can further explore how expertise affects bank participation and loan performance in this setting. We also advance the literature investigating how banks use financial accounting in contracting. While prior literature typically focuses on characteristics of *borrowers* with respect to firm financial reporting decisions (e.g., Allee and Yohn 2009, Lisowsky and Minnis 2016), we find financial reporting variation is also related to characteristics of *lenders*. Moreover, recent disclosure regulation research links a decline in the use of financial information in debt contracts to shifts in accounting standards (Demerjian 2011). Our evidence suggests that consolidation in the banking sector, which has left the U.S. with fewer, more diverse banks, could generate its own trend in the propensity of firms to provide audited reports to their banks. While various implications of the consolidation of the financial services industry have been discussed and studied (e.g., Berger et al. 1999), the potential impact of such consolidation on the auditing and financial reporting of private firms remains unexplored.

References

Acharya, V., Hasan, I., & Saunders, A. (2006). Should banks be diversified? Evidence from individual bank portfolios? Journal of Business, 79, 1355–1412.

Allee, K. D., & Yohn, T. L. (2009). The demand for financial statements in an unregulated environment: An examination of the production and use of financial statements by privately held small businesses. The Accounting Review, 84(1), 1-25.Ball, R., Bushman, R. M., & Vasvari, F. P. (2008). The debt-contracting value of accounting information and loan syndicate structure. *Journal of Accounting Research, 46*(2), 247-287.

Barth, J. R., Caprio, G. & Levine, R. (2004). Bank supervision and regulation: What works best? Journal of Financial Intermediation, 13, 205-48.

Bartik, T. J. (1991). Who benefits from state and local economic development policies? Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Basel Committee on Banking Supervision. (2000). Principles for the management of credit risk. Basel, Switzerland.

Basel Committee on Banking Supervision. (2006). International convergence of capital measurement and capital standards. Basel, Switzerland.

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2006). Bank concentration, competition, and crises: First results. Journal of Banking & Finance, 30(5), 1581-1603.

Berger, A. N., Demsetz, R. S., & P. E. Strahan. (1999). The consolidation of the financial services industry: Causes, consequences, and implications for the future. Journal of Banking & Finance 23, 135-194.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., & Stein, J. C. (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. Journal of Financial Economics 76, 237-269.

Berger, A. N. & Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organizational structure. The Economic Journal 112, F32-F53.

Berger, A. N., & Udell, G. F. (2006). A more complete conceptual framework for SME finance. Journal of Banking & Finance, 30(11), 2945-2966.

Bills, K. L., Jeter, D. C., & Stein, S. E. (2015). Auditor industry specialization and evidence of cost efficiencies in homogenous industries. The Accounting Review 90(5): 1721-1754.

Blackwell, D., T. Noland, & D. Winters. (1998). The value of auditor assurance: Evidence from loan pricing. Journal of Accounting Research 36(1): 57-70.

Blanchard, O. J., & Katz, L. F. (1992). Regional evolutions. Brookings Papers on Economic Activity, 1, 1-61.

Boyd, J. H., & De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. Journal of Finance 60(3), 1329-43.

Boyd, J. H., & Prescott, E. C. (1986). Financial intermediary-coalitions. Journal of Economic Theory, 38(2), 211-232.

Brickley, J., C. Smith, and J. Zimmerman. (2009). Managerial Economics and Organizational Architecture. 5th ed. McGraw-Hill, New York, NY.

Cahan, S. F., Godfrey, J. M., Hamilton, J., & Jeter, D. C. (2008). Auditor specialization, auditor dominance, and audit fees: The role of investment opportunities. The Accounting Review, 83(6), 1393-1423.

Cassar, G., Ittner C., and Cavalluzzo K. (2015). Alternative information sources and information asymmetry reduction: Evidence from small business debt. Journal of Accounting and Economics, 59(2), 242-263.

Cebenoyan, A. S. & Strahan, P. E. (2004). Risk management, capital structure and lending at banks. Journal of Banking & Finance, 28, 19-43.

Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?. Journal of Accounting and Economics, 27(3), 285-303.

De Franco, G., Edwards, A., & Liao, S. (2016). Product market peers in lending: Information processing efficiencies and proprietary costs. Working paper, Tulane University and University of Chicago.

Dell'Ariccia, G. (2001). Asymmetric information and the structure of the banking industry. European Economic Review, 45(10), 1957-1980.

Dell'Ariccia, G., Friedman, E., & Marquez, R. (1999). Adverse selection as a barrier to entry in the banking industry. The RAND Journal of Economics, 515-534.

Dell'Ariccia, G., Igan, D., & Laeven, L. U. (2012). Credit booms and lending standards: Evidence from the subprime mortgage market. Journal of Money, Credit and Banking, *44*(2-3), 367-384.

Demerjian, P. (2011). Accounting standards and debt covenants: Has the "balance sheet approach" led to a decline in the use of balance sheet covenants? Journal of Accounting and Economics, 52(2), 178-202.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. The Review of Economic Studies, 51(3), 393-414.

Gormley, T. A., Kim, B. H., & Martin, X. (2012). Do firms adjust their timely loss recognition in response to changes in the Banking Industry? Journal of Accounting Research, 50(1), 159-196.

Gup, B. & Kolari, J. (2005). Commercial Banking: The Management of Risk. 3rd Ed. John Wiley & Sons, Hoboken, NJ.

Kalay, A. (2015). Investor sophistication and disclosure clienteles. Review of Accounting Studies, 20(2), 976-1011.

Keys, B., Mukherjee, T., Seru, A., & Vig, V. Did securitization lead to lax screening? Evidence from subprime loans. Quarterly Journal of Economics, 125 (1): 307-362.

Liberti, J. & Mian, A. (2009). Estimating the effects of hierarchies on information use. Review of Financial Studies 22, 4057-90.

Liberti, J., Sturgess, J., & Sutherland, A. G. (2017). Information sharing and lender specialization: Evidence from the U.S. commercial lending market. Working paper, DePaul University, Northwestern University, and MIT.

Lisowsky, P. and Minnis, M. (2017). Accounting choices and capital allocation: Evidence from large private U.S. firms. Working paper, University of Chicago and University of Illinois.

Lisowsky, P., Minnis, M., & Sutherland, A. G. (2017). Economic growth and financial statement verification. Journal of Accounting Research, forthcoming.

Lo, A. K. (2014). Do declines in bank health affect borrowers' voluntary disclosures? Evidence from international propagation of banking shocks. Journal of Accounting Research, *52*(2), 541-581.

Loutskina, E. (2005). Does securitization affect bank lending? Evidence from Bank Responses to Funding Shocks. Working paper, Boston College.

Loutskina, E., & Strahan, P. E. (2011). Informed and uninformed investment in housing: The downside of diversification. Review of Financial Studies, 24(5), 1447-1480.

Minnis, M. (2011). The value of financial statement verification in debt financing: Evidence from private U.S. firms. Journal of Accounting Research, 49(2), 457-506.

Minnis, M., & Sutherland, A. G. (2016). Financial statements as monitoring mechanisms: Evidence from small commercial loans. Journal of Accounting Research.

Office of the Comptroller of the Currency (OCC). (2006). Concentrations in Commercial Real Estate Lending, Sound Risk Management Practices. Washington, D.C.

Office of the Comptroller of the Currency (OCC). (2011). Comptroller's Handbook: Concentrations of Credit. Washington, D.C.

Office of the Comptroller of the Currency (OCC). (2014). Survey of Credit Underwriting Practices. Washington, D.C., December.

Office of Inspector General (OIG). (2012). Acquisition, development, and construction loan concentration study. Office of Audits and Evaluations. Report No. EVAL-13-001. October.

Paravisini, D., Rappoport, V., & Schnabl, P. (2015). Specialization in bank lending: Evidence from exporting firms. No. w21800. National Bureau of Economic Research.

Rajan, R.G. (1992). The choice between informed and arm's-length debt. Journal of Finance 47(4), 1367-1400.

Shroff, N., Verdi. R.S. & Yost, B.P. (2017). When does peer-firm information matter? MIT working paper.

Stein, J. (2002). Information production and capital allocation: Decentralized vs. hierarchical firms. Journal of Finance 57, 1891-1921.Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance*, *62*(2), 629-668.

Sutherland, A. (2017). The economic consequences of borrower information sharing: Relationship dynamics and investment.

Tabak, B. M., Fazio, D. M., & Cajueiro, D. O. (2011). The effects of loan portfolio concentration on Brazilian banks' return and risk. Journal of Banking & Finance, 35(11), 3065-3076.

US Bureau of Economic Analysis. Industry Economic Accounts. http://www.bea.gov/industry/gdpbyind_data.htm (accessed July 25, 2015).

U.S. House of Representatives. (1984). Inquiry into Continental Illinois Corp. and Continental Illinois National Bank. Serial No. 98-111.

Watts, R. & Zimmerman, J. (1983). Agency problems, auditing and the theory of the firm: Some evidence. Journal of Law and Economics, 26, 613-34.Wittenberg-Moerman, R. (2008). The

role of information asymmetry and financial reporting quality in debt trading: Evidence from the secondary loan market. *Journal of Accounting and Economics*, *46*(2), 240-260.

Winton, A. (1999). Don't put all your eggs in one basket? Diversification and specialization in lending. (September 27, 1999).

Appendix A: Risk Management Association Data Description[37]

Overview
The Risk Management Association (RMA) "is a not-for-profit, member-driven professional association serving the financial services industry. Its sole purpose is to advance the use of sound risk management principles in the financial services industry." It has approximately 2,500 institutional members and 18,000 individual members, including "relationship managers, credit officers, risk managers, and other financial services professionals."

The RMA has been publishing the Annual Statement Studies© for nearly a century (first published in 1919). The purpose of these studies is to provide financial institutions (hereafter, banks) with benchmarking data to better understand the financial performance of commercial borrowers and prospects. Data for these studies are collected annually. Each year, the RMA begins its campaign to encourage members to participate. Participating banks typically have a deadline of June or July of each year to provide annual financial statements that the bank has collected from a borrower or prospect over the time period of April 1 of the previous year to March 31 of the current year. Historically banks have submitted their collection of financial statements manually (e.g., via mail and fax); however, the overwhelmingly predominant form of submission more recently is via electronic submission (for example, in 2014, 95% of the financial statements submitted by banks were provided electronically). Several software packages that banks use to analyze commercial loans have a compatible export feature, allowing banks to simply push the "submit" button to create an RMA submission file.

RMA member banks collect financial reports from borrowers in all industries, sizes, and loan grades or risk ratings. However, as quoted from the RMA Handbook, observations will be rejected if any one of the following is not true:

- The fiscal year must fall within the current period—only 12-month fiscal statements falling within 4/1 to 3/31 are acceptable.
- The balance sheet must balance.
- The legal form of the entity must be noted.
- The type of financial statement must be noted.
- A valid NAICS or SIC code must be present. RMA accepts either an SIC code (four-digit) or a NAICS code (six-digit). Please note: RMA strongly encourages submission via 2012 NAICS.
- The income statement must be complete.

Importantly, reports are rejected if a valid industry and report type are not included. This mitigates concerns that industries or report types classified as "other" are simply picking up "missing" observations. RMA indicates that their credo is, "Contribute every statement you have," so they make a concerted effort to have each bank submit their entire portfolio of reports. The RMA also has controls in place to identify duplicate reports from the same borrower so the system will not allow more than one report from the same borrower within a bank. For the publicly available Annual

[37] This section quotes frequently from RMA's homepage (www.rmahq.org) and the RMA 2015-2016 Annual Statement Studies© Submission Campaign handbook (available at: http://www.rmahq.org/FileLibrary/Toolsand Publications/StatementStudies/Submission-Campaign-Brochure-2015.pdf) as accessed on August 25, 2015.

Statement Studies, the RMA truncates firms with assets above $250 million. For purposes of our study, however, the RMA did not eliminate either duplicate observations or those observations with more than $250 million in assets to provide the best proxy for a bank's portfolio.

The dataset that the RMA provided to us is aggregated at the bank-year-industry-region-borrower size level. For each unit of observation, the RMA tabulated for us the number of financial reports into one of five mutually exclusive, collectively exhaustive categories—unqualified audit, review, compilation, tax return, and other (see below for additional detail about these report types)—and the total sales ($) for all borrowers within the unit of observation.

Several important points and caveats regarding this dataset are worth noting:

- The data are not collected from a random sample of banks. Banks volunteer to participate. To the extent that this creates omitted variable selection bias in the data, we cannot control for this bias; however, the results reported in the paper are robust to including only those banks that participate in each year. Moreover, those banks that choose to participate in the RMA sample are typically more important commercial lenders than those banks that do not participate—i.e., these are the more important banks for our study.
- There is no guarantee that the data represent the entire bank portfolios. The RMA only "encourages" banks to submit all financial reports. Moreover, banks do not collect any financial reports for a minority of their smallest borrowers (Minnis and Sutherland 2016). However, given the simple electronic submission process and the very high correlation between magnitude of RMA participation and bank commercial lending portfolios as tabulated in Call Reports, we believe that the RMA dataset is a very reasonable proxy for the banks' commercial lending portfolios.
- See the online appendix for additional analysis investigating the banks participating in the RMA dataset.

Report Types

As noted above, RMA tabulates the number of financial reports collected by members into one of five different financial report types: unqualified audit, review, compilation, tax returns, and other. In this section, we will describe each of these report types.

Unqualified Audit

A financial statement audit provides positive assurance that the financial statements are reported in accordance with Generally Accepted Accounting Principles. An unqualified audit opinion indicates that the auditor believes that the financial statements are materially in accordance with GAAP. Unqualified audited financial statements are accompanied by complete footnote disclosure, providing the most complete set of information of all of the reports along with the highest level of assurance and no detected material deviations from GAAP.

Review

Financial statement reviews provide negative assurance. An independent accountant performs analytical procedures (e.g., ratio analyses) and interviews management to assess whether the financial statements are misstated; however, the accountant does not perform substantive procedures to obtain positive evidence of an account balance. Reviews are generally accompanied

by complete footnote disclosure; therefore, reviewed financial statements provide a similar information set to unqualified audits, but the information has a significantly lower level of assurance, reporting quality, and cost.

Compilation

A compilation provides no assurance about the financial statement balances reported in the financial statements. An accountant puts the firm's financial information in the form of financial statements but performs no procedures and provides no assurance as to the reporting quality. Compilations include all three standard financial statements, but are not required to report (and generally omit) footnote disclosures. Therefore, compilations provide substantially less assurance and information than either audits or reviews.

Tax Return

All firms are required to file a return with the Internal Revenue Service (IRS) annually. The nature of these returns differs by entity type (e.g., C Corporation, S Corporation, or Limited Liability Company) and entity size (e.g., firms with less than $250,000 in assets are not required to complete Schedule L which is a balance sheet). While all firms follow "tax basis" accounting to complete the form, the tax basis may differ based on firm size and various options that firms are able to elect (e.g., accrual versus cash basis; differing depreciation options, etc.). Therefore, even within the tax basis of accounting, the differing forms and various options result in heterogeneity. The focus of tax returns is the income statement, but firms exceeding $250,000 in assets also must provide a balance sheet. Important omissions from tax returns include both the statement of cash flows and financial footnotes. Moreover, while independent accountants are frequently involved in the production of these reports, they generally do not provide assurance about them. However, the IRS serves an implicit monitoring role, though the vast majority returns are not audited on an annual basis by the IRS. Collectively, tax returns provide useful but limited financial information and have some, but weaker (and implicit) verification.

Other

The "other" category captures all reports that are not one of the above, and per our discussions with RMA, mostly consists of two report types: company prepared financial statements (the vast majority of this category) and qualified audit reports. Company prepared financial reports are those prepared internally by management and provided to the bank without the involvement of an external accountant. Qualified audit reports are audits similar to "unqualified" audit reports described above but a qualification was made regarding some aspect of the financial statements. For example, the company prefers not to follow a particular accounting rule, so the independent accounting firm provides an "except for" opinion which states that the financial statements follow GAAP except for this aspect. Historically, the RMA reported qualified reports as a separate category, but because this category was infrequently used, RMA consolidated it with "other." Unfortunately, we are unable to disentangle qualified audits from other statements in this category, which is one caveat of our study, but given the assurance and information provided by qualified opinions, this omission works against our findings.

Appendix B: Variable Definitions

This appendix provides definitions for all variables used throughout the paper. The subscripts *b, i, r, t, f* denote bank, industry, region, year, and firm, respectively.

Variable	Description
Exposure	The bank is classified as having an exposure to a given industry if it collects any financial statements from firms belonging to the respective industry. See OCC (2011) for industry definitions.
# Unique Exposures	The number of unique OCC Industries included in the bank's exposures. The unit of observation is bank-year.
HHI	A Hirschman-Herfindahl index (HHI) measure, equal to the sum of the squares of the Industry Shares. The HHI is calculated using the total sales of borrowers for each bank within a region-industry. The unit of observation is bank-year. The formula is as follows: $$HHI_{b,t} = \sum \left(\frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{i=1}^{I} \sum_{r=1}^{6} \sum_{f=1}^{F} Sales_{b,f,i,r,t}} \right)^2$$
Share_bank	The ratio of total firm sales for a given Industry-Region within a bank to total sales for all of the bank's commercial customers in a given year. The unit of observation is bank-year-industry-region. The formula is as follows: $$Share\_bank_{b,r,i,t} = \frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{i=1}^{I} \sum_{r=1}^{6} \sum_{f=1}^{F} Sales_{b,f,i,r,t}}$$
Share_market	The ratio of total firm sales for a given Industry-Region within a bank to total sales for all bank commercial customers for all banks for the same industry-region in a given year. The unit of observation is bank-year-industry-region. The formula is as follows: $$Share\_market_{b,r,i,t} = \frac{\sum_{f=1}^{F} Sales_{b,f,i,r,t}}{\sum_{b=1}^{B} \sum_{f=1}^{F} Sales_{b,f,i,r,t}}$$
Share_statements	The ratio of the number of financial statements collected for a given industry-region within a bank to the total financial statements collected from all of the bank's commercial customers in a given year.

	The unit of observation is bank-year-industry-region. The formula is as follows: $$Share\_statements_{b,r,i,t} = \frac{\sum_{f=1}^{F} \#Statements_{b,f,i,r,t}}{\sum_{i=1}^{I}\sum_{r=1}^{6}\sum_{f=1}^{F} \#Statements_{b,f,i,r,t}}$$
Share_related	The ratio of total firm sales for a given OCC Group-Region, *excluding those of that particular industry*, within a bank to total sales for all of the bank's commercial customers in a given year. The unit of observation is bank-year-industry-region. The formula is as follows: $$Share\_related_{b,r,g,t} = \frac{\sum_{f=1}^{F} Sales_{b,f,g-i,r,t}}{\sum_{g=1}^{G}\sum_{r=1}^{6}\sum_{f=1}^{F} Sales_{b,f,g,r,t}}$$
New Exposure This Year	An indicator equal to one for industries that the bank is exposed to this year that it was not exposed to in the prior year, and zero otherwise. For observations in the bank's first year of RMA reporting, the variable is recorded as missing. The unit of observation is bank-year-industry-region.
Number of Statements	The total number of financial statements collected by the bank, including Unqualified Audits, Reviews, Compilations, Tax Returns, and Other Statements.
Log Bank Size	The total firm sales for all of the bank's exposures. We use the natural log of this variable.
Average Borrower Size	The ratio of total firm sales for all of the bank's exposures to the number of statements. We use the natural log of this variable.
% Unqualified	The percent of financial statements collected that are Unqualified Audits. The unit of observation can be either bank-year or bank-year-industry-region.
Hot Region	An indicator equal to one if the financial statements were collected from the Southeast or West, and zero otherwise.
Construction	An indicator equal to one for industries related to construction activity, including the following Industries: Homebuilding, Nonresidential Building Contractors, Heavy and Civil Engineering Construction, and Specialty Trade Contractors; and zero otherwise. We

	omit industries related to real estate (Industries: Real Estate Developer/Owner and Rental and Leasing Services) because audited financials are not typically used to monitor real estate loans.
Novice	An indicator equal to one for bank-regions where construction was not a top 10 exposure in 2002, and zero otherwise.

Table 1, Panel A: Number of Financial Statements by Type, Region, and Firm Size
This table describes the sample of financial statements submitted to RMA by year, region, and industry.

Financial documents submitted to RMA	1,755,576
Exclude Bank-Years missing Unqualified data	(72,220)
Final Sample	1,683,356

Year	Unqualified	Reviews & Compilations	Tax & Other	All Statements
2002	30,157	44,223	54,518	128,898
2003	31,442	55,000	64,884	151,326
2004	33,961	47,075	69,583	150,619
2005	33,077	43,228	68,843	145,148
2006	38,169	47,056	86,930	172,155
2007	36,046	42,387	89,666	168,099
2008	38,323	44,224	100,321	182,868
2009	38,756	43,917	103,959	186,632
2010	38,669	43,606	109,516	191,791
2011	40,130	44,763	120,927	205,820
Total	358,730	455,479	869,147	1,683,356
	21.3%	27.1%	51.6%	100.0%

Region	# Statements	%
Northeast	314,633	18.7%
Southeast	377,996	22.5%
Central	338,359	20.1%
South Central	124,740	7.4%
North Central	209,650	12.5%
West	317,978	18.9%
Total	1,683,356	100.0%

Firm size	# Statements	%
<$1M	272,201	16.2%
$1M-$3M	279,334	16.6%
$3M-$5M	162,464	9.7%
$5M-$10M	233,583	13.9%
$10M-$25M	275,374	16.4%
>$25M	460,400	27.4%
Total	1,683,356	100.0%

Table 1, Panel B: Bank-level Descriptive Statistics

This table provides descriptive statistics for banks submitted data to the RMA over our sample period 2002-2011. The top half of the panel provides descriptive statistics for the number of years that the banks report data to the RMA and when the typical bank begins and ends reporting data to the RMA. The bottom half of the table reports the number of financial statements in the dataset distributed according to the number of years that the bank which collected the financial statements has been reporting data to RMA.

Bank-level reporting	Mean	Std Dev	Min	25%	50%	75%	Max	N
# Years of RMA data	4.4	3.2	1.0	2.0	3.0	7.0	10.0	728
Earliest Reporting Year	2003	2	2002	2002	2002	2004	2011	728
Latest Reporting Year	2007	3	2002	2004	2007	2011	2011	728
Longest Streak of Consecutive Reporting (Years)	3.7	3.0	1.0	1.0	2.0	5.0	10.0	728

# Reporting Years	Statements	# Banks	% Statements	% Cumulative
10	948,878	87	56.4%	56.4%
9	108,377	51	6.4%	62.8%
8	89,782	31	5.3%	68.1%
7	153,015	38	9.1%	77.2%
6	97,375	34	5.8%	83.0%
5	103,250	51	6.1%	89.1%
4	50,603	61	3.0%	92.2%
3	46,816	77	2.8%	94.9%
2	65,160	118	3.9%	98.8%
1	20,100	180	1.2%	100.0%
	1,683,356	728	100.0%	

Table 2, Panel A: Bank-year level Summary Statistics

This table describes the number and type of statements banks are collecting, and concentration measures of the banks' portfolios at the bank-year unit of observation. See Appendix B for variables definitions.

	Mean	Std Dev	25%	50%	75%	N
# F/S Collected	527	1,746	27	72	202	3,193
% Unqualified	12.8%	12.6%	4.2%	10.3%	18.2%	3,193
Cumulative Borrower Sales ($ millions)	160,742	745,408	214	2,111	30,623	3,193
Average Borrower Size ($ millions)	278.7	1,042.4	6.7	21.1	165.0	3,193
# Unique Industry Exposures	36.1	24.8	17.0	30.0	50.0	3,193
# Unique Region Exposures	2.2	1.7	1.0	1.0	3.0	3,193
# Unique Industry-Region Exposures	51.8	68.6	17.0	31.0	55.0	3,193
HHI	26.5%	22.2%	10.7%	19.0%	34.7%	3,193
Share_bank	6.1%	9.0%	2.0%	3.4%	6.9%	3,193
Share_market	1.4%	3.1%	0.1%	0.3%	1.3%	3,193
Share_statements	6.1%	3.2%	4.4%	4.8%	6.8%	3,193

Table 2, Panel B: Trends in Bank Portfolio Composition
This table summarizes bank-level HHI (the Hirschman-Herfindahl index) of the banks' commercial loan exposures and the number of unique industry-region exposures for each bank-year. The sample size is 3,193 bank-years.

	All Banks		Banks >=$100B Assets		Banks <$100B Assets	
Year	HHI	# Unique Industry-Regions	HHI	# Unique Industry-Regions	HHI	# Unique Industry-Regions
2002	0.280	40.0	0.041	376.0	0.281	39.3
2003	0.281	44.5	0.037	380.0	0.281	43.7
2004	0.276	47.3	0.031	300.0	0.277	45.9
2005	0.275	49.0	0.041	250.0	0.276	47.7
2006	0.270	52.4	0.041	261.5	0.273	49.7
2007	0.243	56.2	0.027	263.5	0.247	53.1
2008	0.262	59.0	0.036	267.7	0.268	53.4
2009	0.254	61.6	0.039	252.0	0.258	57.9
2010	0.221	62.2	0.068	251.4	0.224	58.4
2011	0.253	62.8	0.057	228.0	0.257	59.4

Table 3: Commercial Portfolio Concentration and Unqualified Audit Collection across Banks

This table examines the across bank association between the proportion of unqualified audits collected at the bank-year level and the bank's overall degree of concentration (*HHI*), bank size (*Log Bank Size* or *Bank Size Tercile #*), and average borrower size (*Log Average Firm Size*). *HHI* is increasing in exposure concentration. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified
HHI	-0.108***		-0.073***	-0.040***	-0.038**
	[-6.50]		[-3.52]	[-2.62]	[-2.47]
Log Bank Size		0.014***	0.009**		
		[4.49]	[2.19]		
Bank Size Tercile 2				0.016**	
				[2.09]	
Bank Size Tercile 3				0.062***	
				[6.27]	
Log Average Borrower Size	0.026***	0.005	0.014**	0.015***	0.015***
	[9.71]	[0.95]	[2.00]	[6.96]	[6.91]
Adj R2	0.184	0.181	0.191	0.191	0.193
N	3,193	3,193	3,193	2,547	2,547
Year Fixed Effects	Yes	Yes	Yes	Yes	No
Bank Size Tercile x Year Fixed Effects	No	No	No	No	Yes
Clustering	Bank	Bank	Bank	Bank	Bank
Obs Level	Bank-Year	Bank-Year	Bank-Year	Bank-Year	Bank-Year

Table 4: Unqualified Audit Collection within Bank—Portfolio Sort

This table examines differences in unqualified audit collection across bank size and exposure concentration within bank. Bank size terciles are formed according to cumulative borrower sales in the bank-year. Concentration terciles are formed by sorting each industry-region within a bank-year into a tercile based on the variable *Share_statements*. We report differences in *% Unqualified* across the first and third Bank size and Concentration terciles at the bottom of the table. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively. See Appendix B for variable definitions.

Bank Size Tercile #1

	%Unqualified	<$1M	$1M-$3M	$3M-$5M	$5M-$10M	$10M-$25M	>$25M	All
Concentration Tercile	1	3.5%	5.6%	7.1%	11.1%	20.4%	40.2%	9.1%
	2	3.0%	5.5%	7.8%	10.3%	16.3%	29.8%	7.1%
	3	2.5%	3.8%	4.1%	9.0%	13.4%	23.6%	5.2%
	All	2.9%	4.8%	5.8%	10.0%	16.8%	32.7%	7.0%

Bank Size Tercile #2

	%Unqualified	<$1M	$1M-$3M	$3M-$5M	$5M-$10M	$10M-$25M	>$25M	All
Concentration Tercile	1	4.2%	7.0%	9.7%	14.7%	24.2%	45.9%	15.1%
	2	4.3%	6.6%	10.9%	15.4%	24.2%	47.0%	14.6%
	3	2.9%	4.4%	7.0%	10.9%	17.6%	36.4%	9.6%
	All	3.3%	5.3%	8.3%	12.4%	20.0%	40.5%	11.4%

Bank Size Tercile #3

	%Unqualified	<$1M	$1M-$3M	$3M-$5M	$5M-$10M	$10M-$25M	>$25M	All
Concentration Tercile	1	5.7%	8.0%	11.3%	16.8%	28.5%	56.1%	31.8%
	2	5.5%	7.7%	11.2%	15.9%	25.7%	51.3%	26.8%
	3	4.0%	7.5%	11.0%	15.0%	23.1%	44.8%	21.6%
	All	4.2%	7.6%	11.0%	15.2%	23.5%	46.2%	22.5%

Bank size tercile differences (tercile 3 minus tercile 1)

	1.3%**	2.8%***	5.2%***	5.1%***	6.7%***	13.5%***

Concentration tercile differences (tercile 3 minus tercile 1) within bank size tercile

Bank size Tercile	1	-1.0%*	-1.7%**	-3.1%***	-2.1%	-7.0%***	-16.5%***
	2	-1.4%***	-2.5%***	-2.6%***	-3.8%***	-6.6%***	-9.6%***
	3	-1.7%***	-0.4%	-0.3%	-1.8%**	-5.5%***	-11.3%***

Table 5: Unqualified Audit Collection within Bank

This table models the proportion of unqualified audits collected by a bank (*% Unqualified*) for a given industry-region as a function of the relative degree of exposure for that industry-region. In Panel A, relative exposure is measured as *Share_bank*, *Share_market*, or *Share_statements* in models (1)-(2) and (5), (3), and (4), respectively. We use various combinations of bank, time, industry, and region fixed effects across the specifications as labeled at the bottom of the table. In Panel B, we measure relative exposure as *Share_related* in model 1. In models 2, 3, and 4 the dependent variable is the proportion of reviews, compilations, and tax returns collected by a bank (*% Reviewed*, *% Compilation*, and *% Tax Return*). See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

Panel A: Main results

	(1)	(2)	(3)	(4)	(5)
	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified
Share_bank	-0.235***	-0.271***			-0.195***
	[-7.14]	[-8.39]			[-5.48]
Share_market			-0.301***		
			[-11.08]		
Share_statements				-0.681***	
				[-9.53]	
Log Average Borrower Size	0.054***	0.056***	0.059***	0.053***	0.040***
	[21.17]	[23.01]	[24.68]	[24.35]	[13.48]
Adj R2	0.386	0.409	0.412	0.408	0.615
N	165,374	165,374	165,374	165,374	165,374
Bank FE?	Yes	No	No	No	No
Industry FE?	Yes	No	No	No	No
Year FE?	Yes	No	No	No	Yes
Region FE?	Yes	No	No	No	No
Bank-Year FE?	No	Yes	Yes	Yes	No
Industry-Region-Year FE?	No	Yes	Yes	Yes	No
Bank-Industry-Region FE?	No	No	No	No	Yes
Clustering	Bank	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year

Panel B: Specification checks

	(1)	(2)	(3)	(4)
	% Unqualified	% Reviewed	% Compilation	% Tax Return
Share_related	-0.082***			
	[-5.61]			
Share_bank		0.039*	0.047***	0.101***
		[1.80]	[2.91]	[4.45]
Log Average Borrower Size	0.052***	-0.001	-0.015***	-0.033***
	[24.09]	[-0.67]	[-20.00]	[-22.27]
Adj R2	0.407	0.149	0.148	0.335
N	165,374	165,374	165,374	165,374
Bank-Year FE?	Yes	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year

Table 6: Unqualified Audit Collection within Bank—Cross Sectional Tests

This table reports the Table 5, Panel A, column 2 regression results after partitioning the sample. In Panel A, industry-regions are partitioned based on the competition of the banking market. Banking market competition is measured with the Herfindahl-Hirschman index using borrower sales data from the RMA dataset. In Panel B, industry-regions are partitioned based on dispersion of ROA of firms within the industry. ROA dispersion is measured as the interquartile range (IQR) for each industry-year using data from Compustat. See Appendix B for variables definitions. We report two-tailed test statistics for differences in *Share_bank* coefficients across the high and low samples below each panel. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

Panel A: Partitioning on bank market concentration

	(1)	(2)	(3)	(4)	(5)	(6)
	Split at 50th %ile		Split at 75th %ile		Split at 90th %ile	
	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified
	Low Comp	High Comp	Low Comp	High Comp	Low Comp	High Comp
Share_bank	-0.313***	-0.261***	-0.352***	-0.265***	-0.534***	-0.269***
	[-7.25]	[-7.96]	[-5.66]	[-8.32]	[-4.69]	[-8.54]
Log Average Borrower Size	0.056***	0.057***	0.058***	0.056***	0.064***	0.056***
	[21.53]	[23.48]	[20.88]	[23.30]	[18.98]	[23.26]
Adj R2	0.385	0.424	0.383	0.414	0.350	0.411
N	63,096	102,278	26,349	139,025	7,719	157,655
Bank-Year FE?	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
Difference: (High=1-High=0)	-0.052		-0.087		-0.265	
Chi-square	1.950		2.310		5.890	
Difference: p-value	0.1628		0.1282		0.0152	

Panel B: Partitioning on firm performance dispersion

	(1)	(2)	(3)	(4)	(5)	(6)
	Split at 50th %ile		Split at 75th %ile		Split at 90th %ile	
	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified
	High Disper	Low Disper	High Disper	Low Disper	High Disper	Low Disper
Share_bank	-0.310***	-0.272***	-0.359***	-0.267***	-0.444**	-0.273***
	[-7.81]	[-7.75]	[-6.04]	[-8.38]	[-2.61]	[-8.65]
Log Average Borrower Size	0.059***	0.055***	0.065***	0.054***	0.064***	0.056***
	[23.09]	[21.59]	[21.44]	[22.16]	[16.08]	[23.04]
Adj R2	0.394	0.406	0.382	0.401	0.331	0.410
N	77,434	83,719	33,165	127,988	11,753	149,400
Bank-Year FE?	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
Difference: (High=1-High=0)	-0.038		-0.092		-0.171	
Chi-square	1.140		3.240		1.100	
Difference: p-value	0.2860		0.0721		0.2939	

Table 7: Bank Expertise and Syndicated Lending

This table examines how bank expertise is employed in syndicated lending. In Panel A, the dependent variable takes on one of five measures of firm opacity: sales (1), age (2), and indicators for R&D spending (3), being privately held (4), and being unrated (5). In Panel B, the dependent variable is the proportion of the bank's syndicated loans between 2002 and 2011 allocated to a given industry. In both panels, the independent variable is *Share_statements*, and the sample is restricted to U.S. loans the bank completed as a lead arranger. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

Panel A: Firm opacity and lead arranger loans

	(1) Log Sales	(2) Log Age	(3) Has R&D	(4) Private Firm	(5) Unrated
Share_statements	-1.457	0.843	3.497***	0.438**	-0.605**
	[-1.31]	[0.65]	[4.66]	[2.30]	[-2.11]
Adj R2	0.255	0.167	0.044	0.165	0.144
N	10,367	7,903	7,903	14,809	17,745
Bank FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank	Bank
Obs Level	Loan	Loan	Loan	Loan	Loan

Panel B: Lead arranger loans and bank expertise

	(1) Share Dealscan Loans	(2) Share Dealscan Loans
Share_statements	0.564**	0.382
	[2.02]	[1.59]
Adj R2	0.139	0.094
N	998	969
Bank FE	No	No
Industry FE	Yes	Yes
Clustering	Bank	Bank
Obs Level	Bank-Industry	Bank-Industry
At least 10 Dealscan Loans?	No	Yes

Table 8: Unqualified Audit Collection, Concentration, and Experience

This table reports the Table 5, Panel A column 2 regression results after including a variable which measures the number of years of experience that a bank has lending to each industry-region. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)
	% Unqualified	% Unqualified	% Unqualified
Share_bank	-0.203**	-0.202**	0.344
	[-2.00]	[-2.02]	[1.57]
Years Experience		-0.007***	-0.006***
		[-3.70]	[-3.16]
Share * Years Experience			-0.077***
			[-3.45]
Log Avg Borrower Size	0.056***	0.057***	0.057***
	[10.32]	[10.44]	[10.39]
Adj R2	0.404	0.408	0.409
N	7,068	7,068	7,068
Bank FE?	Yes	Yes	Yes
Industry-Region FE?	Yes	Yes	Yes
Clustering	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
Sample	2011 only	2011 only	2011 only
Required Reporting Yrs	10	10	10

Table 9: Unqualified Audit Collection and New Exposures

This table examines unqualified audit collection (*% Unqualified*) for exposures that are both new to the bank and comprise a significant portion of their portfolio. The explanatory variable of interest in column 1 (*New Exposure This Year*) is an indicator variable equal to one if the bank collects no financial statements from a given industry within a region in year *t-1* (i.e., has no exposure to a given industry within a particular region) but that the industry is a top 10 exposure for the bank-region in year *t*. In columns 2-4, we modify this variable by looking at exposures that were new and top 10 exposures two, four, or eight years ago. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	% Unqualified	% Unqualified	% Unqualified	% Unqualified
New Exposure This Year	0.064***			
	[8.37]			
New Exposure Two Years Ago		0.050***		
		[5.67]		
New Exposure Four Years Ago			0.044***	
			[4.25]	
New Exposure Eight Years Ago				-0.013
				[-0.74]
Log Average Borrower Size	0.051***	0.052***	0.053***	0.053***
	[22.07]	[22.16]	[22.14]	[22.14]
Adj R2	0.407	0.406	0.406	0.406
N	136,911	136,911	136,911	136,911
Bank-Year FE?	Yes	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year

Table 10, Panel A: Construction Entry Following the Housing Boom

This table analyzes whether the housing boom led banks with little construction exposure in 2002 to increase their exposure by 2005 relative to banks which were lending to the construction industry in 2002, and whether the exposure change differs by regional variation in construction activity. *Hot Region* is an indicator variable equal to one for the Southeast and West regions, and zero for other regions. We restrict the sample to observations where construction was not a top 10 exposure for the bank-region in 2002, and classify such observations as construction "Novices" for subsequent tests. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at either the industry (columns 1, 2, and 3) or bank (columns 4 and 5) level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	Share_bank	Share_bank	Share_bank	Share_bank for Construction	Share_bank for Construction
Year 2005	-0.020***	-0.011***	-0.007	0.042***	0.030**
	[-6.49]	[-3.02]	[-1.48]	[3.22]	[2.31]
Hot Region	0.031***	0.024***	0.026**	0.003	0.000
	[5.27]	[2.65]	[2.52]	[0.48]	[0.02]
Construction * Hot Region	-0.031***	-0.014***	-0.014**		
	[-5.25]	[-2.83]	[-2.44]		
Hot Region * Year 2005	-0.029***	-0.010*	-0.013*	0.038	0.050*
	[-4.57]	[-1.69]	[-1.70]	[1.36]	[1.69]
Construction * Year 2005	0.061***	0.027***	0.020		
	[4.85]	[3.51]	[1.46]		
Year 2005 * Hot Region * Construction	0.062**	0.042***	0.044***		
	[2.43]	[3.41]	[3.32]		
Fixed Effects	Industry	Industry, Bank	Industry, Bank	Bank	Bank
Clustering	Industry	Industry	Industry	Bank	Bank
adj. R-sq	0.048	0.289	0.287	0.022	0.024
N	14,288	14,288	10,926	887	654
Obs Level	Bank-Region-Industry-Year	Bank-Region-Industry-Year	Bank-Region-Industry-Year	Bank-Region-Year	Bank-Region-Year
2002 and 2005 only?	Yes	Yes	Yes	Yes	Yes
Construction not in bank's top 10 in 2002?	Yes	Yes	Yes	Yes	Yes
Only Banks with both Hot and Not Hot Regions?	No	No	Yes	No	Yes
Only Construction Industry?	No	No	No	Yes	Yes

Table 10, Panel B: Unqualified Audit Collection during the Housing Boom

This table compares unqualified audit collection in 2005 for construction-related industries across various regions and banks with different levels of construction expertise. Novice bank-regions are those where construction was not a top 10 exposure in 2002. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the industry level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)
	% Unqualified	% Unqualified
Novice	0.076***	0.072***
	[6.11]	[5.92]
Hot Region	-0.047***	-0.050***
	[-3.98]	[-4.21]
Hot Region * Novice	0.029*	0.034**
	[1.78]	[2.07]
Novice * Construction	-0.021	0.013
	[-1.27]	[0.60]
Hot Region * Construction	-0.042**	-0.043***
	[-2.15]	[-3.05]
Hot Region * Construction * Novice	0.089***	0.075*
	[2.89]	[1.77]
Log Average Borrower Size	0.054***	0.060***
	[26.17]	[27.76]
Fixed Effects	Industry, Bank	Industry, Bank
Clustering	Industry	Industry
adj. R-sq	0.387	0.376
N	12,260	8,828
Obs Level	Bank-Region-Industry-Year	Bank-Region-Industry-Year
Sample	2005 only	2005 only
Only Banks with both Hot and Not Hot Regions	No	Yes

Table 11: Unqualified Audit Collection and Bank Outcomes

This table tests for a link between unqualified audit collection (*% Unqualified*), concentration (*HHI*), and subsequent bank performance. We examine two bank performance measures: (1) the ratio of C&I chargeoffs to total C&I loans at the beginning of the year; and (2) the ratio of net income to total assets at the beginning of the year. We include the same bank controls as Loutskina and Strahan (2011): Securities/Assets, Interest on Deposits/Deposits, Log Assets, Capital/Assets, Deposits/Assets, Net Income/Assets, Real Estate Loans/Assets, C&I Loans/Assets, Unused Loan Commitments/Assets. Both % Unqualified and HHI have been demeaned to facilitate interpretation of the coefficients. See Appendix B for variables definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1) Scaled C&I Chargeoffs *t+1*	(2) ROA *t+1*
% Unqualified	-0.003	-0.006
	[-0.36]	[-1.36]
HHI	0.000	0.001
	[0.08]	[0.82]
% Unqualified * HHI	0.028	0.004
	[0.81]	[0.35]
Adj R2	0.078	0.397
N	2,489	2,489
Fixed Effects	Year	Year
Bank Controls	Yes	Yes
Clustering	Bank	Bank
Obs Level	Bank-Year	Bank-Year

Online Appendix to:

Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements

May 2017

This online appendix provides descriptive detail of the data provided by the Risk Management Association and tabulates additional analyses.

Table A1: RMA Bank Characteristics versus the Population of US Commercial Banks

This table compares bank-years included in the RMA dataset with the population of bank-years reported in FDIC call reports. Total assets are the total assets of the bank (RCFD 2170). ROA is the return on assets defined as net income (RIAD 4340) divided by total assets. Regulatory capital is defined as Tier 1 (RCFD 8274) plus Tier 2 (RCFD 8275) Capital divided by Risk-Weighted Assets (RCFDA 223). C&I Loans/Total Loans is the dollar value of commercial and industrial loans (RCFD 1766) divided by the total loans and leases (RCFD 1400). The number of bank-years in this table is less than the number of bank-years used in the paper because not all RMA participant banks can be traced to the FDIC dataset (e.g., credit union or thrift). Panel A presents descriptive statistics for all variables. Panel B presents the results of a linear probability model where the dependent variable equals 1 if the bank-year is included in the RMA dataset and 0 if not. Standard errors are clustered at the bank level. *, **, *** represent significance at the 10%, 5%, and 1% level, respectively, using two-tailed p-values.

Panel A: Descriptive Statistics

RMA Bank-Years	Mean	Std Dev	25%	50%	75%	N
Total Assets ($ millions)	8,580.6	63,900.0	278.2	634.8	1,886.1	2,546
ROA	0.9%	1.0%	0.6%	1.0%	1.3%	2,546
Regulatory Capital	13.9%	4.6%	11.3%	12.8%	14.9%	2,546
C&I Loans/Total Loans	18.0%	10.4%	10.8%	16.0%	22.7%	2,546
C&I Loans ($ millions)	916.6	5,250.7	26.3	62.2	204.8	2,546

Non-RMA Bank-Years	Mean	Std Dev	25%	50%	75%	N
Total Assets ($ millions)	1,135.2	23,600.0	58.5	119.6	265.7	73,656
ROA	0.7%	1.7%	0.4%	0.9%	1.3%	73,656
Regulatory Capital	18.5%	28.1%	12.2%	14.7%	19.2%	73,656
C&I Loans/Total Loans	14.6%	10.8%	7.5%	12.5%	19.1%	73,656
C&I Loans ($ millions)	100.8	1,774.9	3.3	8.9	23.3	73,656

Panel B: OLS Regression of RMA Participation on Bank Characteristics

	In the RMA dataset?
LN(Assets)	0.034***
	[15.66]
ROA	0.009
	[0.22]
Regulatory Capital	0.003
	[1.36]
C&I Loans/Total Loans	0.078***
	[5.34]
AdjR2	0.066
N	76,202
Fixed effects	Year
Clustering	Bank
Obs level	Bank-year

Table A2: Participation Frequency of the Largest US C&I Lending Banks

This table reports how frequently the 10 largest US-based commercial and industrial lending banks participate in the RMA dataset. The 10 largest banks were determined each year by using the balance of the C&I loan portfolio as of the December FDIC call report (variable RCFD 1766).

	Bank	C&I Loans	Rank	In RMA this year?	Total Years in RMA
2002	CITIBANK NA	$94,559,000	1	0	4
	BANK OF AMER NA	$80,724,000	2	1	10
	JPMORGAN CHASE BK	$52,446,000	3	1	10
	WACHOVIA BK NA	$44,036,000	4	1	8
	FLEET NA BK	$42,586,000	5	1	1
	BANK ONE NA	$35,978,000	6	1	2
	U S BK NA	$28,188,264	7	0	6
	WELLS FARGO BK NA	$21,996,000	8	1	10
	SUNTRUST BK	$21,266,051	9	1	10
	COMERICA BK	$16,766,439	10	1	10
				8	
2003	CITIBANK NA	$84,106,000	1	0	4
	BANK OF AMER NA	$64,467,689	2	1	10
	JPMORGAN CHASE BK	$45,551,000	3	1	10
	WACHOVIA BK NA	$41,279,000	4	1	8
	FLEET NA BK	$33,921,000	5	0	1
	BANK ONE NA	$30,562,000	6	1	2
	U S BK NA	$26,435,646	7	0	6
	WELLS FARGO BK NA	$24,650,000	8	1	10
	COMERICA BK	$22,399,882	9	1	10
	SUNTRUST BK	$22,265,600	10	1	10
				7	
2004	CITIBANK NA	$96,894,000	1	0	4
	JPMORGAN CHASE BK NA	$67,492,000	2	1	10
	BANK OF AMER NA	$66,192,087	3	1	10
	WACHOVIA BK NA	$44,454,000	4	0	8
	WELLS FARGO BK NA	$35,678,000	5	1	10
	U S BK NA	$27,960,042	6	0	6
	FLEET NA BK	$25,368,065	7	0	1
	COMERICA BK	$22,096,056	8	1	10
	SUNTRUST BK	$19,872,418	9	1	10
	KEYBANK NA	$17,581,190	10	1	10
				6	
2005	BANK OF AMER NA	$107,228,401	1	1	10
	CITIBANK NA	$102,841,000	2	0	4
	JPMORGAN CHASE BK NA	$92,184,000	3	1	10
	WACHOVIA BK NA	$59,210,000	4	1	8
	WELLS FARGO BK NA	$39,799,000	5	1	10
	U S BK NA	$29,893,923	6	0	6
	SUNTRUST BK	$23,176,105	7	1	10
	COMERICA BK	$21,018,959	8	1	10
	KEYBANK NA	$20,005,334	9	1	10
	LASALLE BK NA	$19,698,680	10	1	5
				8	
2006	CITIBANK NA	$134,009,000	1	0	4
	BANK OF AMER NA	$115,861,798	2	1	10
	JPMORGAN CHASE BK NA	$98,956,000	3	1	10
	WACHOVIA BK NA	$67,064,000	4	1	8
	WELLS FARGO BK NA	$47,095,000	5	1	10
	U S BK NA	$32,024,668	6	1	6
	NATIONAL CITY BK	$25,083,199	7	1	7
	SUNTRUST BK	$24,612,132	8	1	10
	COMERICA BK	$23,483,137	9	1	10
	LASALLE BK NA	$21,238,763	10	1	5
				9	
2007	CITIBANK NA	$175,308,000	1	0	4
	BANK OF AMER NA	$143,655,570	2	1	10
	JPMORGAN CHASE BK NA	$133,106,000	3	1	10
	WACHOVIA BK NA	$81,813,000	4	1	8
	WELLS FARGO BK NA	$61,570,000	5	1	10
	U S BK NA	$35,625,587	6	1	6
	NATIONAL CITY BK	$29,284,568	7	1	7
	COMERICA BK	$24,784,783	8	1	10
	SUNTRUST BK	$23,958,854	9	1	10
	LASALLE BK NA	$21,131,668	10	0	5
				8	
2008	BANK OF AMER NA	$177,567,660	1	1	10
	JPMORGAN CHASE BK NA	$140,768,000	2	1	10
	CITIBANK NA	$131,518,000	3	1	4
	WACHOVIA BK NA	$89,023,000	4	1	8
	WELLS FARGO BK NA	$82,256,000	5	1	10
	U S BK NA	$40,428,261	6	1	6
	SUNTRUST BK	$28,419,017	7	1	10
	NATIONAL CITY BK	$28,208,252	8	1	7
	PNC BK NA	$24,599,963	9	1	10
	COMERICA BK	$24,451,203	10	1	10
				10	
2009	BANK OF AMER NA	$145,715,980	1	1	10
	CITIBANK NA	$96,538,000	2	1	4
	JPMORGAN CHASE BK NA	$95,801,000	3	1	10
	WELLS FARGO BK NA	$69,566,000	4	1	10
	WACHOVIA BK NA	$61,386,000	5	1	8
	PNC BK NA	$41,567,804	6	1	10
	U S BK NA	$34,231,916	7	1	6
	SUNTRUST BK	$22,394,217	8	1	10
	FIFTH THIRD BK	$22,049,506	9	1	9
	COMERICA BK	$20,559,136	10	1	10
				10	
2010	BANK OF AMER NA	$149,553,874	1	1	10
	WELLS FARGO BK NA	$118,116,000	2	1	10
	CITIBANK NA	$96,357,000	3	1	4
	JPMORGAN CHASE BK NA	$90,552,000	4	1	10
	PNC BK NA	$41,095,999	5	1	10
	U S BK NA	$33,552,969	6	1	6
	SUNTRUST BK	$25,069,556	7	1	10
	FIFTH THIRD BK	$22,340,117	8	0	9
	COMERICA BK	$21,251,795	9	1	10
	ALLY BK	$16,904,717	10	0	0
				8	
2011	BANK OF AMER NA	$173,399,570	1	1	10
	WELLS FARGO BK NA	$140,035,000	2	1	10
	CITIBANK NA	$117,547,000	3	1	4
	JPMORGAN CHASE BK NA	$115,243,000	4	1	10
	PNC BK NA	$46,826,231	5	1	10
	U S BK NA	$39,436,806	6	1	6
	SUNTRUST BK	$29,895,126	7	1	10
	FIFTH THIRD BK	$25,587,644	8	1	9
	ALLY BK	$25,158,101	9	0	0
	COMERICA BK	$23,400,526	10	1	10
				9	

Table A3: Number of Commercial Banks by Year

This table reports the number of US commercial banks by year over the years 2002 to 2011. Column 1 reports the number of banks in the population, calculated as the number of call reports in the FDIC database. Column 2 reports the number of banks which have at least 10% of their loan portfolio consisting of commercial and industrial (C&I) loans. Column 3 reports the minimum number of banks needed each year to cumulate 75% of the total C&I market for loans. Column 4 presents the number of banks participating in the RMA dataset.

	(1)	(2)	(3)	(4)
Year	Population of banks with FDIC call reports	With at least 10% C&I lending	Consisting of 75% of C&I market	Number of banks contributing RMA data
2002	8,751	5,722	117	465
2003	8,609	5,540	129	426
2004	8,432	5,375	126	372
2005	8,302	5,233	110	317
2006	8,238	5,166	97	319
2007	8,097	5,097	82	271
2008	7,873	4,855	78	266
2009	7,613	4,495	92	256
2010	7,284	4,221	91	252
2011	7,072	4,089	77	249
% change 2002 to 2011	-19%	-29%	-34%	-46%

Table A4: Distribution of Bank Years and Financial Statements by Frequency of Bank Participation in RMA Dataset

This table presents the number of bank years and financial statements contributed by banks, conditional on how frequently the bank participates in the dataset. Panel A includes only banks that participate in consecutive years. Panel B includes banks with only one year of missing data between consecutive years of participation. Panel C includes banks with two separate missing years (in one year increments only) among consecutive years. Panel D includes banks with at least two consecutive missing years.

	(1)	(2)	(3)	(4)	(5)
	Number of years in RMA	Bank years		Financial Statements	
		Number	Cumulative %	Number	Cumulative %
Panel A:Banks with no intermediate missing years	10	870	27%	948,878	56%
	9	108	31%	41,855	59%
	8	72	33%	17,714	60%
	7	126	37%	71,961	64%
	6	126	41%	88,807	69%
	5	175	46%	40,956	72%
	4	128	50%	39,919	74%
	3	168	56%	43,442	77%
	2	204	62%	60,427	80%
	1	180	68%	20,100	82%
		2,157	68%	1,374,059	82%
Panel B: Banks with only one missing year	9	351	11%	66,522	4%
	8	88	14%	56,852	7%
	7	35	15%	71,234	12%
	6	30	16%	3,371	12%
	5	45	17%	32,397	14%
	4	68	19%	7,694	14%
	3	39	21%	2,850	14%
	2	16	21%	893	14%
		672	21%	241,813	14%
Panel C: Banks with two different episodes of missing data of only 1 year	8	72	2%	13,153	1%
	7	49	4%	4,738	1%
	6	36	5%	3,634	1%
	5	25	6%	29,625	3%
	4	40	7%	2,547	3%
	3	15	7%	339	3%
	2	16	8%	3,840	3%
		253	8%	57,876	3%
Panel D: Banks with more than two consecutive missing years	8	16	1%	2,063	0%
	7	56	2%	5,082	0%
	6	12	3%	1,563	1%
	5	10	3%	272	1%
	4	8	3%	443	1%
	3	9	3%	185	1%
		111	3%	9,608	1%

Table A5: Distribution of Financial Statements by Industry

This table reports the distribution of firms by industry across several different datasets. Column 1 reports the distribution of firms from the RMA dataset in the year 2011. Column 2 reports the distribution of private firms from the Sageworks dataset across the years 2002 to 2008 as reported in Minnis (2011). Column 3 reports the distribution of US-based firms in Compustat with a non-missing value for the total assets (AT) variable in 2011. Column 4 reports the distribution of firms in the entire US economy, determined as those firms with an assigned Dun & Bradstreet (DUNS) number.

		RMA	Sageworks	Compustat	Economy
Code	**Industry Title**	**%**	**%**	**%**	**%**
11	Agriculture, Forestry, Fishing and Hunting	1.9%	1.7%	0.3%	2.9%
21	Mining	1.0%	0.5%	6.4%	0.2%
22	Utilities	1.0%	0.7%	4.1%	1.1%
23	Construction	7.4%	22.3%	1.2%	9.3%
31-33	Manufacturing	14.9%	16.2%	34.8%	2.2%
42	Wholesale Trade	10.7%	10.4%	2.7%	4.8%
44-45	Retail Trade	8.4%	10.0%	3.8%	9.8%
48-49	Transportation and Warehousing	3.6%	2.8%	2.3%	2.6%
51	Information	1.5%	1.4%	9.6%	2.1%
52	Finance and Insurance	2.8%	6.6%	18.1%	4.5%
53	Real Estate Rental and Leasing	11.9%	9.0%	5.3%	4.7%
54	Professional, Scientific, and Technical Services	7.5%	5.4%	3.6%	13.1%
55	Management of Companies and Enterprises	0.4%	0.2%	0.0%	0.2%
56	Admin/Support and Waste Mgt/Remed Svs	2.9%	2.3%	1.8%	12.0%
61	Educational Services	2.1%	0.9%	0.4%	2.1%
62	Health Care and Social Assistance	8.9%	2.7%	1.8%	8.4%
71	Arts, Entertainment, and Recreation	2.0%	1.3%	0.7%	1.9%
72	Accommodation and Food Services	4.7%	3.1%	1.6%	5.1%
81	Other Services (except Public Administration)	5.3%	2.5%	0.3%	11.5%
92	Public Administration	1.0%	0.0%	1.3%	1.5%
Total	Total	205,820	13,614	5,940	14,772,275

Table A6: Unqualified Audit Collection and Portfolio Share by OCC Group

This table reports descriptive statistics of industry concentration within banks and the rate of unqualified audit collection. *% Unqualified* is the percentage of financial reports collected from that industry that are unqualified audits. Top 5? is the frequency with which a particular industry is one of the bank-year's top 5 exposures within its portfolio.

OCC Group	% Unqualified	Top 5?
Agribusiness	16.6%	4.2%
Apparel & Textiles Manufacturing	24.4%	2.3%
Auto-Related	13.0%	24.1%
Banks	33.3%	9.5%
Commercial Services	15.9%	6.9%
Consumer Services	24.6%	4.5%
Durables Manufacturing Exc. Auto	21.5%	7.7%
Entertainment & Recreation	31.8%	4.1%
Finance & Insurance	33.3%	6.7%
Food & Beverage Manufacturing	29.9%	9.5%
Food & Drug Stores	12.0%	11.5%
Government & Education	63.5%	9.5%
Health Care & Pharmaceuticals	31.7%	9.7%
Loans to Individuals on Commercial Systems	14.2%	0.9%
Materials & Commodities Exc. Energy	25.4%	5.6%
Media & Telecom	34.2%	4.7%
Oil & Gas & Coal	25.2%	13.1%
Professional Services	16.8%	17.7%
Real Estate & Construction	16.0%	14.5%
Restaurant & Hotel	10.8%	9.1%
Retail Stores Exc. Food & Drug	11.3%	6.3%
Transportation Services	19.5%	5.4%
Utilities	58.0%	11.2%
Wholesale Distribution	17.3%	29.0%

Table A7: Sector and Firm Size Coverage

This table reports the number of financial reports collected by borrower size and OCC Sector (industry sector as defined by the OCC) over the full sample period.

OCC Sector	<$1M	$1M-$3M	$3M-$5M	$5M-$10M	$10M-$25M	>$25M	Total
Commodities	9,584	18,824	14,431	22,039	27,476	61,392	153,746
Distribution	18,396	38,323	28,066	45,920	60,567	100,768	292,040
Financial	13,860	12,232	6,184	7,922	8,778	21,310	70,286
Government	3,778	6,907	4,348	6,634	8,841	16,187	46,695
Manufacturers	10,907	29,213	23,599	38,566	51,773	88,917	242,975
Real Estate	112,237	46,672	23,049	30,438	31,518	39,238	283,152
Services	103,159	126,888	62,670	81,917	86,247	132,392	593,273
Not Elsewhere Classified	280	275	117	147	174	196	1,189
Total Statements	272,201	279,334	162,464	233,583	275,374	460,400	1,683,356

Table A8: Correlation Matrix

This table provides a Pearson correlation matrix for bank-level concentration variables used in the study and supplemental analyses. The unit of observation is bank-year and there are 3,193 observations. See Appendix B in the paper for variable definitions.

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Number of Statements Collected	1.00					
(2) % Unqualified	0.20	1.00				
(3) # Unique Industry-Region Exposures	0.81	0.32	1.00			
(4) HHI	-0.25	-0.13	-0.35	1.00		
(5) Share of Portfolio in top 5 Industry exposures	-0.50	-0.23	-0.65	0.75	1.00	
(6) # Industries in top 50% of Bank's Portfolio	0.61	0.19	0.67	-0.66	-0.90	1.00

Table A9: Unqualified Audit Collection and Bank Specialization across Banks

This table tabulates robustness tests to Table 3 in the paper, which examines the across bank association between the proportion of unqualified audits collected at the bank-year level and the bank's overall degree of concentration during the year. The HHI and Share measures are increasing in concentration, whereas the # Industries in Bank's Top portion of portfolio measure is decreasing in concentration. Column 1 below reproduces column 1 of Table 3 in the paper. See Appendix B in the paper for variable definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified	% Unqualified
HHI	-0.108***							
	[-6.50]							
HHI - 6 Digit NAICS		-0.107***						
		[-6.38]						
HHI - Group			-0.118***					
			[-6.97]					
Share Top 5				-0.144***				
				[-6.96]				
Share Top 10					-0.210***			
					[-7.78]			
# Industries in Bank's Top 50%						0.010***		
						[5.92]		
# Industries in Bank's Top 66%							0.006***	
							[6.27]	
# Industries in Bank's Top 75%								0.005***
								[6.68]
Log Average Borrower Size	0.026***	0.026***	0.025***	0.024***	0.023***	0.024***	0.024***	0.024***
	[9.71]	[9.70]	[9.85]	[9.95]	[9.65]	[9.49]	[9.45]	[9.47]
Adj R2	0.184	0.184	0.195	0.200	0.208	0.178	0.182	0.185
N	3,193	3,193	3,193	3,193	3,193	3,193	3,193	3,193
Fixed Effects	Year	Year	Year	Year	Year	Year	Year	Year
Clustering	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
Obs Level	Bank-Year	Bank-Year	Bank-Year	Bank-Year	Bank-Year	Bank-Year	Bank-Year	Bank-Year

Table A10: Unqualified Audit Collection within Bank—Conditioning on Bank Size

This table reports the paper's Table 5, column 2 regression results after partitioning on bank size. Banks are sorted into terciles according to Total Assets in a given bank-year per Call Report data. Column 4 restricts the sample to observations in bank-years which exceed $100B of Total Assets. See Appendix B in the paper for variable definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	% Unqualified	% Unqualified	% Unqualified	% Unqualified
	Small Bank	Medium Bank	Large Bank	TBTF Bank
Share_bank	0.020	-0.105**	-0.450***	-2.557***
	[0.48]	[-2.09]	[-6.68]	[-6.86]
Log Average Borrower Size	0.029***	0.040***	0.058***	0.053***
	[7.19]	[8.50]	[14.24]	[6.96]
Adj R2	0.350	0.345	0.374	0.254
N	18,552	23,493	69,245	9,487
Bank-Year FE?	Yes	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes	Yes
Clustering	Bank	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
Bank Size (Total Assets)	<$363M	$363M-$1.218B	>$1.218B	>$100B

Table A11: Unqualified Audit Collection within Bank—Conditioning on Amount of C&I Lending

This table partitions the Table 5, column 2 result from the paper after conditioning on the ratio of C&I Loans to Total Loans in a bank's portfolio per Call Report data. Banks are sorted into terciles according to the C&I Loans-to-Total Loans ratio in given bank-year. See Appendix B in the paper for variable definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)
	% Unqualified	% Unqualified	% Unqualified
	Low C&I	Medium C&I	High C&I
Share_bank	-0.070*	-0.252***	-0.309***
	[-1.68]	[-4.78]	[-5.12]
Log Average Firm Size	0.034***	0.052***	0.059***
	[8.96]	[13.21]	[13.42]
Adj R2	0.357	0.397	0.378
N	24,045	42,207	45,038
Bank-Year FE?	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes
Clustering	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
C&I Loans/Total Loans	<12.8%	12.8%-20.3%	>20.3%

Table A12: Unqualified Audit Collection within Bank—Conditioning on Regulatory Capital

This table partitions the Table 5, column 2 result from the paper after conditioning on the level of Regulatory Capital per Call Report data. Banks are sorted into terciles according to the ratio of Tier 1 plus Tier 2 Capital to Risk-Weighted Assets in a given bank-year. See Appendix B in the paper for variable definitions. Reported below the coefficients are t-statistics calculated with standard errors clustered at the bank level. *, **, *** indicate significance at the two-tailed 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)
	% Unqualified	% Unqualified	% Unqualified
	Low Capital	Medium Capital	High Capital
Share_bank	-0.341***	-0.205***	-0.161***
	[-6.31]	[-4.28]	[-3.30]
Log Average Firm Size	0.054***	0.051***	0.047***
	[13.17]	[14.03]	[11.06]
Adj R2	0.376	0.388	0.385
N	44,495	36,624	30,171
Bank-Year FE?	Yes	Yes	Yes
Industry-Region-Year FE?	Yes	Yes	Yes
Clustering	Bank	Bank	Bank
Obs Level	Bank-Industry-Region-Year	Bank-Industry-Region-Year	Bank-Industry-Region-Year
Regulatory Capital	<11.8%	11.8%-14.0%	>14.0%