Bank Diversification and Tail Risk *

Priyank Gandhi
Rutgers Business School

Darius Palia
Rutgers Business School

Jasper Pan
Rutgers Business School

This version: March 17, 2025

Abstract

We examine the relationship between bank business line diversification and tail risk by assessing diversification across 16 business lines using a unique entropy-based measure. The results reveal that a one standard deviation increase in diversification is associated with a 2.5% reduction in tail risk in the subsequent quarter. This effect lasts up to four quarters ahead and is present in both good and bad times. Furthermore, diversified banks exhibit higher future stock returns in the next three quarters, higher profitability, lower default risk, higher change in loan supply, and lending resilience during the Great Recession. Lastly, diversification of core business lines (related diversification) is associated with lower tail risk and higher returns while diversification of noncore business lines (unrelated diversification) is not. This paper emphasizes the crucial role of diversification across business lines in mitigating tail risk and enhancing overall bank performance particularly during periods of financial instability and documents that it is related diversification that benefits banks.

JEL Codes: G20, G21, G24, G34.

Keywords: Bank Diversification, Related Diversification, Tail Risk, Stock Returns, Lending.

*Gandhi: Rutgers Business School, Rutgers University Newark and New Brunswick; Email: priyank.gandhi@business.rutgers.edu. Palia: Rutgers Business School, Rutgers University Newark and New Brunswick; Email: dpalia@business.rutgers.edu. Pan: Rutgers Business School, Rutgers University Newark and New Brunswick; Email: jasper.pan@rutgers.edu. We thank...for detailed and helpful comments.

Introduction

In the ever-evolving banking industry, the diversification of business lines has become a topic of significant academic interest and debate. Traditionally, banks have focused on a narrow set of activities, primarily generating interest income from core operations like loans. However, recent decades have seen modern banks increasingly diversify into a broader range of business lines, particularly those generating noninterest income, such as insurance, trading, and investment banking. This shift toward diversification has sparked discussions about its impact on banks' overall risk profiles, particularly their exposure to tail risk.

The debate around business line diversification is divided between two perspectives. Critics argue that excessive diversification increases risk by introducing a wider range of potentially volatile business activities. They contend that managing a diverse portfolio can heighten the accumulation of tail risks, where losses in one area can cascade into broader institutional instability. As a result, these critics often advocate for regulatory measures to limit diversification and prevent banks from overextending into unrelated business lines. On the other hand, proponents of diversification argue that expanding into multiple business lines allows banks to spread risk across different income streams, thereby reducing their overall risk exposure. They suggest that diversification acts as a hedge against downturns in any single area, enhancing banks' stability, profitability, and resilience in the face of economic shocks.

This paper contributes to the debate on business line diversification and tail risk in the banking sector by examining their relationship using a panel regression approach. An entropy-based diversification measure assesses income across 16 business lines—seven core lines related to interest income and nine noncore lines tied to noninterest income. The baseline measure of tail risk is defined as the negative of the average excess return during the worst 5% return days over a quarter. Various bank-level characteristics are controlled for, including size (log book value of assets), leverage or capitalization (capital-to-asset ratio), profitability (net income-to-asset ratio), operating efficiency (cost-to-income ratio), funding structure (deposits-to-liabilities ratio), growth opportunities (asset growth rate), and risk-taking behavior (loan loss provisions and Z-score). Given tail risk's persistence, the model includes lagged tail risk for each bank and quarter. All regressions incorporate bank and time fixed effects, and independent variables are standardized by subtracting their mean and dividing by their standard deviation. Statistical significance is computed using standard errors clustered at the bank level.

The baseline results suggest that a one standard deviation increase in diversification corresponds with a 0.112 percentage point reduction in tail risk in the following quarter, equivalent to a 2.45% quarterly decrease (9.8% annualized) relative to the sample mean of 4.57%. This reduction in tail risk is consistent across four alternative measures: *Tail Risk 10%*, defined as the negative of the average excess return on the worst 10% return days over a quarter; *CAPM Tail Risk*, defined as the negative of the average CAPM residual return on the worst 5% return days; *FF3 Tail Risk*, utilizing the Fama and French [1993] residual return; and *FF5 Tail Risk*, which applies the Fama and French [2015] residual return during the worst 5% return days over a quarter.

Our analysis further reveals that diversification is a strong predictor of reduced tail risk over a horizon extending up to four quarters. While the effect is initially observed in the subsequent quarter, it intensifies over time, reaching 0.188 percentage points (4.1% in economic magnitude) by the third quarter and remaining substantial at 0.173 percentage points (3.8%) in the fourth quarter. These findings are statistically significant at the 1% level, confirming that diversification consistently predicts lower tail risk across both short and extended horizons.

Further investigation examines the effect of diversification on tail risk under varying economic conditions and around the time of the global financial crisis. Diversification consistently predicts lower tail risk in bank stock returns across both favorable and adverse economic conditions, with a notably stronger effect during "bad times" (e.g., NBER-defined recessions), where the coefficient more than doubles compared to "good times." This implies that diversification's tail-risk-reducing impact becomes more pronounced during periods of financial stress. Moreover, while diversification's predictive power holds before the crisis, it significantly weakens afterward, pointing to a diminished association with tail risk post-Great Recession.

Our subsequent analysis decomposes overall diversification into related (core business lines generating interest income) and unrelated (noncore lines generating noninterest income) categories to assess which type more effectively predicts reduced tail risk in banks. The results indicate that only related diversification, involving interest-generating lines, significantly correlates with lower tail risk, whereas unrelated diversification does not. Specifically, a one-standard-deviation increase in related diversification results in a 0.075 percentage point reduction in tail risk, representing an economic magnitude of 1.64%. This novel finding suggests that banks' core, interest-generating diversification mitigates risk, while noncore diversification lacks this risk-reducing effect, underscoring the risk implications of banks diversifying into noncore activities.

To explore the mechanisms underlying the link between diversification and lower tail risk, we examine two primary channels: the hedging channel and the business model channel. The hedging channel suggests that banks diversify proactively to hedge against future risks, learning from past negative shocks and adjusting their diversification accordingly. Conversely, the business model channel attributes diversification and risk behaviors to a bank's intrinsic risk culture and strategic priorities, with minimal adjustment to external shocks. Analyzing banks' responses to the 1998 Russian Financial Crisis, the evidence supports the business model channel, as banks with higher tail risk did not increase diversification post-crisis, indicating that their diversification strategies reflect a stable risk culture rather than a reactive hedging approach.

In examining performance, our analysis reveals that diversification also predicts higher stock returns. A one standard deviation increase in diversification is associated with a 65% increase in quarterly excess returns and an approximately 27% increase in abnormal returns, according to models based on CAPM and Fama-French regressions. These results are statistically significant, with diversification's impact on excess returns being significant at the 5% level and on abnormal returns at the 1% level, highlighting a robust positive association between diversification and stock performance across various return measures.

The analysis also demonstrates that diversification strongly predicts stock returns over the subsequent three quarters. Initially associated with a 65% increase in quarterly excess returns in the next quarter, diversification's effect grows stronger in the second and third quarters, with returns increasing by 113% and 125%, respectively, both significant at the 1% level. However, by the fourth quarter, diversification's impact on returns diminishes and loses statistical significance. These findings suggest a substantial, lasting positive effect of diversification on returns, particularly within the first three quarters.

Another decomposition of diversification into related (core business lines generating interest income) and unrelated (noncore lines generating noninterest income) categories examines their impact on stock returns. Results show that only related diversification is positively associated with higher returns, while unrelated diversification lacks a significant effect and sometimes even shows negative coefficients. Specifically, a one-standard-deviation increase in related diversification correlates with a 52% increase in quarterly excess returns and an 18% increase in abnormal returns, with statistical significance across models. These findings align with earlier results linking related diversification to lower tail risk, emphasizing that related diversification enhances both return and risk outcomes, unlike unrelated diversification.

Additionally, our analysis addresses the nonlinear effects of diversification on the risk-return trade-off, suggesting that optimal diversification levels for minimizing tail risk differ from those maximizing returns. Including squared diversification terms in the regressions, the findings reveal that while diversification reduces tail risk, this effect plateaus beyond a certain level; the optimal level for tail risk reduction is 2.14 standard deviations above the mean. Conversely, diversification enhances returns but peaks at 1.28 standard deviations above the mean. This indicates a balance point for diversification between 1.28 and 2.14 standard deviations, where return maximization and risk minimization intersect, suggesting a hump-shaped relationship for returns and a U-shaped relationship for tail risk.

The subsequent analysis examines diversification's impact on bank lending, revealing that it significantly predicts loan supply growth, with more diversified banks expanding their lending activities. Using a comprehensive model with controls for economic conditions, monetary policy, and bank-specific factors, the results consistently show a strong link between diversification and loan growth across different model specifications. Even after adjusting for bank size, capital adequacy, profitability, and liquidity, diversification remains a robust predictor. Economically, a one-standard-deviation increase in diversification is associated with a 14% rise in loan supply in the following quarter, highlighting diversification's role in banks' lending decisions and supporting financial stability and credit access.

Moreover, our analysis finds that diversified banks not only predict higher loan growth but also display lending resilience during the 2007-2008 financial crisis. Fixing diversification and control variables at pre-crisis levels allows for isolating diversification's impact on loan supply under stress, scaled by 2006 total assets. The results reveal that a one-standard-deviation increase in pre-crisis diversification led to a 4.9 percentage point increase in total loans, with specific rises of 1.7 and 2 percentage points in commercial and industrial loans and real estate loans, respectively. These findings underscore diversification's role in bolstering banks' lending capacity during downturns.

Finally, the paper demonstrates that diversification substantially enhances banks' accounting performance, predicting higher profitability and lower default risk. A one-standard-deviation increase in diversification results in a 0.011 percentage point increase in return on assets (ROA), equating to a 5% rise relative to the mean, and a 0.151 percentage point increase in return on equity (ROE), or 6.5% relative to the mean. Additionally, diversification is associated with a 0.006 unit rise in the Z-score, indicating a 1.4% improvement in default risk resilience. These results underscore that diversification strengthens banks' profitability, equity efficiency, and stability, reinforcing their overall financial health.

This comprehensive analysis underscores the multifaceted impact of business line diversification on the banking sector's risk-return profile, lending capacity, and overall financial health. By decomposing diversification into related and unrelated categories, the study demonstrates the distinct benefits of core interest-generating activities in reducing tail risk and enhancing stock returns. The findings also emphasize that while diversification is associated with reduced tail risk and higher returns, the effect is non-linear, pointing to a potential optimal diversification level that balances risk reduction with return maximization. These insights highlight critical considerations for bank management and regulators in assessing the role of diversification in promoting financial stability and supporting sustained growth.

Our paper is linked to the vast literature on diversification by both financial and nonfinancial firms (see, for example, Lang and Stulz [1994], Berger and Ofek [1995], and Campa and Kedia [2002], Hann, Ogneva, and Ozbas [2013], and Kuppuswamy and Villalonga [2016], among many others). Specifically, we contribute to the literature on bank diversification and its impact on tail risk, return, and lending resilience. A comprehensive review of this vast literature is beyond the scope of this paper. This literature has explored various dimensions of bank diversification which includes both geographic diversification (Deng and Elyasiani [2008], Goetz, Laeven, and Levine [2013], Goetz, Laeven, and Levine [2016], Levine, Lin, and Xie [2021], and Gelman, Goldstein, and MacKinlay [2023]) as well as bank loan portfolio diversification (Acharya, Hasan, and Saunders [2006] and Shim [2019]). Our paper, instead, relates to the large literature on diversification by banks into multiple business lines. Important papers in this area include Demsetz and Strahan [1997], Stiroh [2004], Stiroh and Rumble [2006], Laeven and Levine [2007], and Saunders, Schmid, and Walter [2020], and many others. Our paper distinguishes itself from others in three key ways. First, it uniquely applies a detailed entropy measure of bank diversification, leveraging the most granular data from the quarterly call reports that all U.S. banks are required to file, covering 16 distinct business lines. Second, it is the first study to link bank business line diversification with reduced tail risk and enhanced returns, thereby highlighting a risk-return trade-off. Third, our paper is the first to establish a significant relationship between related diversification—across core business lines that generate interest income—and lower tail risk and higher returns, while showing that unrelated diversification—across lines that generate noninterest income—has no meaningful relation to risk or return.

The paper is structured as follows: Section 1 details the data sources and methodologies used to calculate the primary dependent and independent variables. Section 2 presents the core empirical findings. Finally, Section 3 provides a summary and concludes the study.

1 Data and summary statistics

In this section, we identify the set of banks used in our analysis, provide a thorough description of our bank diversification measure, and present detailed summary statistics and correlations for the main variables. Additionally, we examine the time-series and cross-sectional variations in diversification and conduct an in-depth analysis of the driving factors influencing bank diversification.

1.1 Sample selection

We collect balance sheet and income statement data from the Consolidated Financial Statements for Bank Holding Companies (FR Y-9C), filed quarterly by U.S. bank holding companies (henceforth referred to as banks). Banks with a total book value exceeding $500 million are required to file these reports, providing comprehensive data on assets, liabilities, and income. For our main sample, we focus specifically on publicly listed banks, as stock return data is necessary to construct the dependent variable, tail risk. This sample includes 560 unique banks, which represent the largest U.S. banks, collectively accounting for over 90% of total U.S. banking sector assets at any given time. Restricting the sample to publicly listed banks with a total book value above $500 million enables analysis at the highest available frequency and captures meaningful variations in financial data. Detailed income data by category in the FR Y-9C reports begins in September 1996, which establishes the start date of our sample.[1]

1.2 Measuring overall, related, and unrelated diversification

We begin by collecting the most granular data for income that is available across all business lines for all banks using the publicly-available, quarterly FR Y-9C reports required to be filed by all bank holding companies in the U.S. Specifically, we collect data for income across sixteen different categories of bank business lines - seven for interest income and nine for noninterest income. The seven sources of interest income include income that a bank accrues from: (i) loans in both domes-

[1] As of March 2024, there are 4,568 banks in the U.S.; however, most are small or privately owned, and stock return data is unavailable for these institutions. Consequently, measures of tail risk cannot be computed for banks excluded from this analysis.

tic and international branches, (ii) leases - including both direct and leveraged leases, (iii) balances at depository institutions, (iv) securities, including both U.S. Treasury and agency obligations as well as mortgage-backed securities, (v) trading assets, (vi) federal funds sold and repurchase agreements, and (vii) any other sources of interest income. The nine sources of noninterest income include income that a bank accrues from: (i) fiduciary activities, (ii) services on domestic deposit accounts, (iii) trading activities, (iv) activities related to securities and insurance, including brokerage services, investment banking, annuity sales, and insurance or reinsurance operations, (v) venture capital, (vi) servicing activities related to mortgages, credit cards, and other financial products, (vii) securitization, encompassing gains, losses, and fees associated with securitization and structured finance, (viii) sale of loans, leases, and real estate, and (ix) any other sources of non-interest income (for e.g., income from safe deposit box rentals and U.S. savings bond redemptions, etc.).

In addition, we follow Jacquemin and Berry [1979] and Khanna and Palepu [2000] to define the entropy measure of bank diversification. To compute diversification, we first compute the share of income derived by bank i in quarter t from source j – i.e. $(S_{j,i,t})$. That is, $S_{j,i,t}$ is simply the ratio of the income derived by bank i, in quarter t from source j to the total income derived from all sixteen interest and noninterest income sources listed above. In each quarter t, for each bank i, diversification is then defined as the weighted sum of the income shares $S_{j,i,t}$, where the weight equal the natural logarithm of the reciprocal of the income share. Thus, the diversification measure for bank i in quarter t equals:

$$Diversification_{i,t} = \sum_{16}^{1} S_{j,i,t} ln \left(\frac{1}{S_{j,i,t}} \right) \tag{1}$$

Diversification values can range from a maximum of 2.77 to a minimum of 0 for any bank. Diversification for a bank that gets \$1B from each of the 16 sources would equal 2.77. Such a bank, with income uniformly distributed across all 16 business lines, is maximally diversified. Diversification for a bank that reports \$15B for one income item, \$1B for another, and zero for all remaining source would equal 0.23. Such a bank has almost the lowest level of diversification.

Furthermore, we define related diversification as the diversification across the seven core business lines that generate interest income, while unrelated diversification is the diversification calculated across the nine non-core business lines generating noninterest income. The rationale behind this distinction is that core business lines may have operational synergies, potentially enhancing financial stability and efficiency, while non-core lines may lack these synergies. We use a horserace regression between these two measures to assess which diversification approach is more beneficial for banks in terms of risk reduction and return enhancement.

Panel A of Table 1 presents summary statistics for the cross-section of banks, highlighting notable variation in diversification measures. The average bank has an overall diversification level of 1.18, with a standard deviation of 0.36, ranging from 0.95 at the 25th percentile to 1.43 at the 75th percentile, indicating a substantial spread. The maximum observed diversification value is 2.47, close to the theoretical limit of 2.77. Related diversification has a mean of 0.46, a standard deviation of 0.27, and a maximum of 1.61, while unrelated diversification shows a mean of 1.06, a standard deviation of 0.37, and a maximum of 2.02. Panel B of Table 1 presents correlations among these diversification types: related and unrelated diversification are correlated at 0.21, related and overall diversification at 0.55, and unrelated and overall diversification at 0.57, reflecting both unique and overlapping diversification dimensions across banks.

1.3 Measuring bank tail risk, return, and control variables

We obtain data on banks' stock prices, holding period returns (including dividends), and total shares outstanding from the Center for Research on Security Prices (CRSP). To identify banks within CRSP, we select firms with a two-digit primary standard industrial classification (SIC) code of 60 or a four-digit SIC code of 6712. Additionally, some studies classify banks using SIC codes in the range 6000–6199. Each bank in CRSP is then matched to its FR Y-9C data via the "CRSP-FRB Link" provided by the Federal Reserve Bank of New York. This tool matches each bank's unique RSSD ID (assigned by the Federal Reserve for regulatory reporting) with its CRSP-assigned PERMCO, updating regularly to reflect mergers, acquisitions, failures, and delistings.

The baseline measure of a bank's tail risk at the quarterly level is defined as the negative of the average excess return during the worst 5% of return days within a quarter. To ensure robustness, we also construct four alternative measures. Tail Risk 10% is defined as the negative of the average excess return during the worst 10% of return days in a quarter. CAPM Tail Risk represents the negative of the average CAPM residual return during the worst 5% of return days within a quarter. Similarly, FF3 Tail Risk is the negative of the average residual return based on the Fama and French [1993] model, during the worst 5% of return days in a quarter, and FF5 Tail Risk is the negative of the average residual return based on the Fama and French [2015] model, during the worst 5% of return days in a quarter.

The baseline quarterly return measure, Qr Return, is the natural logarithm of the buy-and-hold stock excess return over a quarter. For robustness, we construct alternative measures as well: CAPM Abnormal Return, defined as the natural logarithm of the quarterly residual return based on the CAPM; FF3 Abnormal Return, the natural logarithm of the quarterly residual return based on the Fama and French [1993] model; and FF5 Abnormal Return, the natural logarithm of the quarterly residual return based on the Fama and French [2015] model.

Panel A of Table 1 presents the summary statistics for tail risk for the cross-section of banks. Mean tail risk equals nearly 4.57%. However, there is considerable variation over time and in the cross-section, as the standard deviation of tail risk itself is about 3.23, which is similar to the mean. The inter-quartile range (difference in the tail risk between the 25^{th}- and the 75^{th}-percentile) of 2.61% is also indicative of the considerable cross-section and time-series variation in tail risk across banks.

In all our analyses, we include a variety of control variables that can affect a bank's tail risk. Panel A of Table 1 presents the summary statistics for these additional control variables across the cross-section of banks. Specifically, for each bank in the sample, we collect data on the log of book value of assets as a control for bank size, the ratio of total capital to total book value of assets as a control for bank leverage or capitalization, the ratio of net income to total book value of assets as a control for bank profitability, the cost-to-income ratio (computed by dividing total noninterest and interest expenses by total noninterest and interest income) as a control for operational efficiency, the ratio of total deposits to total liabilities as a control for funding structure, the ratio of total loan loss provisions to total loans as a control for risk-taking, the growth rate of total book value of assets (computed over the last three years) as a control for growth opportunities, and, finally, the bank's Z-Score as a control for default risk.

The extant literature suggests that it is important to control for the variables listed above as they can influence bank risk taking and hence its tail risk. For instance, Laeven and Levine [2007] argue that a bank with greater capitalization (or lower leverage) may not indulge in excessive risk-taking, lowering tail risk. Elsas, Hackethal, and Holzhäuser [2010] suggests that we should control for operational efficiency in our analysis as this too can influence bank risk. Further, Laeven and Levine [2007] show that a bank with a higher proportion of deposits to liabilities can easily tap an inexpensive source of funding that benefits from government-subsidized deposit insurance, which can lower bank-specific (tail) risk. In Baele, De Jonghe, and Vander Vennet [2007], loan loss provisions are an important indicator of the amount of bank-specific credit risk. Finally, Saunders, Schmid, and Walter [2020] document that a bank's Z-score serves as an indicator of bank risk-taking behavior and is inversely correlated with the likelihood of bank insolvency. Table A2 in the Appendix provides a summary of the definition and data sources for all control variable listed in Table 1.

1.4 What drives bank diversification?

In this section, we explore the factors driving variation in bank diversification both across the cross-section and over time. We start by plotting the time-series data for Entropy, representing diversification levels in the aggregate U.S. banking sector, in Figure 1. The plot shows that changes in diversification align with key legislative events in banking, reflecting shifts in regulatory adjustments and banks' responses. For example, diversification increases significantly and stabilizes at higher levels across all banks after 1999, following the Gramm-Leach-Bliley Act, which removed the activity restrictions imposed by the Glass-Steagall Act of 1933.

Similarly, diversification declines sharply and remains low post-2007-2009, especially after the Dodd-Frank Act of 2010 was enacted. These trends in Figure 1 strongly suggest that the Entropy measure effectively captures shifts in banks' business line diversification, reacting to regulatory changes that either restrict or expand banks' permissible activities. This responsiveness underscores Entropy's value as a policy-sensitive indicator, highlighting its relevance for tracking the effects of legislative impacts on diversification.

The plot in Figure 2 indicates that the Dodd-Frank Act, passed in 2010, had a pronounced impact predominantly on large banks, while the effects on small and medium-sized banks appear minimal. Specifically, the time series data for bank-sector diversification demonstrate that large banks underwent noticeable adjustments following the enactment of the Act. In contrast, small and medium-sized banks show trends in diversification that remain relatively consistent, suggesting that these banks were less sensitive or responsive to the regulatory changes introduced by Dodd-Frank. Notably, the 2014 implementation of the Volcker Rule, a key provision within the Act, did not seem to prompt a significant reaction in diversification patterns among small and medium-sized banks. This lack of response underscores that regulatory impacts were more heavily concentrated among larger financial institutions, while smaller banks continued their operations with minimal observable shifts in diversification.

The plot in Figure 3 demonstrates substantial cross-sectional variation in diversification across banks. We calculate the average diversification within each quarter for each bank and present it in a box-and-whisker plot. Each year, we observe notable variations in diversification levels, with the median entropy distribution for all banks shifting annually. This median rises significantly following the Gramm-Leach-Bliley Act (GLB) and stabilizes at a lower level after the Great Recession.

The entropy measure of diversification reveals not only time-series variation, consistent with changes in the financial environment, but also considerable cross-sectional differences. Alternative diversification measures do not adequately capture these variations over time and across banks; hence, they are not shown here for clarity. In a separate study, Gandhi, Palia, and Pan [2025] demonstrate that the entropy measure is superior in capturing the diversification effect, as evidenced by reduced idiosyncratic volatility in stock returns. They find that, after controlling for entropy, existing measures of bank business line diversification fail to explain variations in idiosyncratic volatility. Based on the analysis of Gandhi, Palia, and Pan [2025], we adopt the entropy measure for the analysis in this paper.

We utilize predictive panel regression models to thoroughly investigate the various factors that impact a bank's degree of diversification across multiple business lines. Specifically, we perform regressions where the entropy measure of diversification at quarter t+1 is regressed on an array of balance sheet and income statement variables that are hypothesized to influence a bank's choice to diversify its business lines. The predictor variables include a comprehensive set of proxies that capture aspects such as bank size, leverage, profitability, interest margin, funding structure, exposure to credit risk, growth in assets, and probability of default. By examining these variables, we aim to identify and understand the underlying drivers that motivate banks to diversify their activities and income sources.

Table 2 displays the regression results, where a positive (negative) coefficient indicates that higher values of the variable correspond to greater (lower) entropy, signifying higher (lower) business line diversification. For example, in column (1), the coefficient for the log of book value of assets is positive and statistically significant at the 1% level (with a t-statistic of 9.94), suggesting that, as expected, larger banks or those with greater book asset values tend to be more diversified, reflected in higher Entropy. This well-known result is widely recognized and frequently discussed in the literature, underscoring the relationship between size and business diversification in banking institutions.

In column (2), a negative coefficient of -0.044 on the ratio of bank capital to total book value of assets (statistically significant at the 1% level, with a t-statistic of -7.71) suggests that an increase in bank capitalization (i.e., higher capital relative to assets or lower leverage) in a given quarter is associated with a reduction in diversification, in the following quarter. This effect highlights a trade-off between stability and diversification, where highly capitalized banks may prioritize financial stability over expanding into new business lines.

In Table 2, several patterns emerge regarding factors influencing diversification. As anticipated, banks with higher profitability in one quarter tend to exhibit higher diversification in the subsequent quarter. This is intuitive, as profitable banks likely have the resources to expand across multiple business lines and invest in new opportunities, thereby enhancing their diversification. Additionally, banks with higher cost-to-income ratios appear to pursue greater diversification, potentially indicating that these institutions rely on diversification as a strategy to offset operational inefficiencies. Conversely, banks with elevated credit risk, reflected by higher loan loss provisions, tend to be less diversified, implying that banks facing higher credit risk may focus on core lending activities rather than expanding into other areas. Banks experiencing rapid asset growth are also less diversified, possibly indicating a preference for growth within specific segments rather than a broadening of their business scope.

The coefficient on the proxy for a bank's bankruptcy risk (Z-score) is positive at 0.039, statistically significant at the 1% level, with a t-statistic of 8.17. This positive coefficient indicates that banks with lower bankruptcy risk in one quarter tend to exhibit greater diversification, as measured by Entropy, in the subsequent quarter. This finding reinforces the link between stability and diversification, suggesting that banks with lower financial distress risk are better positioned to expand their operations across diverse business lines.

2 Results

2.1 Diversification predicts lower tail risk in the next quarter.

In this section, we explore how bank business line diversification relates to future tail risk for the cross-section of U.S. banks. We do so by relating each bank's measure of business line diversification measured in quarter t to the tail risk of its stock returns at time $t+1$ using standard panel regressions. The exact specification of our panel regression is as follows:

$$Tail\ Risk_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Tail\ Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t} \quad (2)$$

Here, $Tail\ Risk_{i,t+1}$ includes the following five tail risk variables. *Tail Risk 5%* is the negative of the average excess return during the worst 5% return days over a quarter. *Tail Risk 10%* is the negative of the average excess return during the worst 10% return days over a quarter. *CAPM Tail Risk* is the negative of the average CAPM residual return during the worst 5% return days over a quarter. *FF3 Tail Risk* is the negative of the average Fama and French [1993] residual return during the worst 5% return days over a quarter. *FF5 Tail Risk* is the negative of the average Fama and French [2015] residual return during the worst 5% return days over a quarter. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. We control for several bank-level characteristics that influence the relation between bank diversification and bank tail risk. Specifically, we control for size (log book value of assets), leverage or capitalization (ratio of total capital to total book value of assets), profitability (net income to total assets), operating efficiency (cost to income ratio), funding structure (total deposits to total liabilities), growth opportunities (asset growth rate), and bank risk taking (loan loss provisions and Z-score). In addition, since tail risk can be highly persistent, we also control for lagged tail risk for bank i measured in quarter t. All regressions include bank fixed effects and time fixed effects. All right hand side variables are standardized by subtracting the mean and dividing by the standard deviation of the variables. Statistical significance is computed using standard errors clustered at the bank level. The main coefficients of interest is β_i, i.e., the coefficient on $Diversification_{i,t}$. We expect the sign on β_i to be negative, indicating that higher bank diversification is associated with lower tail risk.

Table 3 presents the estimates for regression (2) and shows that the diversification is correlated with lower tail risk. Each column of this table shows the estimates for a separate regression specification – one for each of the five different measures of tail risk. In column (1) where we use the baseline measure of tail risk, we notice that the coefficient on diversification is negative and statistically significant at the 1% level. The negative coefficient of -0.112 on diversification indicates that a one-standard deviation increase in diversification for a particular bank in a particular quarter is associated with a nearly 0.112% lower tail risk for this bank over the next quarter.

The negative relation between diversification and future bank tail risk is not only statistically but economically significant as well. Given that the average quarterly tail risk for the banks in our sample is 4.57%, this implies that higher Diversification is associated with nearly 2.45% lower tail risk over the next quarter as compared to the sample mean.

Notice also that in Table 3, the signs and significance of the coefficients on the control variables are as expected. For example, the coefficient on lagged tail risk is positive and statistically significant, indicating that tail risk is highly persistent. Similarly, the coefficient on leverage (i.e, the ratio of total capital to total book value of assets) is negative, indicating that banks with low leverage or higher capitalization have lower future tail risk. Banks that are less efficient (i.e., have higher cost to income ratios) or take on more risk (as indicated by higher loan loss provisions) indeed have more future tail risk, as indicated by the positive, statistically significant coefficients on these variables in all specifications in Table 3.

2.2 Diversification predicts lower tail risk four quarters ahead.

To assess whether the ability of diversification to predict lower tail risk extends beyond a 1-quarter horizon, we analyze the relationship between diversification and tail risk over longer periods. Table 4 presents the results of this analysis. Column (1) replicates the findings from the first column of Table 3, showing that a one standard deviation increase in diversification is associated with a 0.112 percentage point reduction in tail risk in the next quarter. This effect is economically significant, amounting to a 2.5% decrease in tail risk relative to the sample mean of 4.57%. The adjusted R-squared for this model is 0.697.

In column (2), where the dependent variable is tail risk in quarter $t + 2$, the diversification effect rises to 0.171 percentage points, with an economic magnitude of 3.7 percentage points. The adjusted R-squared declines to 0.65. In Column (3), examining tail risk at $t + 3$, the effect of diversification further increases to 0.188 percentage points, or an economic magnitude of 4.1 percentage points, with an adjusted R-squared of 0.624. Column (4) extends the horizon to quarter $t + 4$, where diversification continues to predict a reduction in tail risk of 0.173 percentage points (3.8 percent economically), with an adjusted R-squared of 0.605.

Across all four horizons, the coefficients on diversification are statistically significant at the one percent level, indicating robust predictability. In conclusion, diversification significantly predicts lower tail risk up to four quarters ahead, with economic magnitudes of 2.5 percent, 3.7 percent, 4.1 percent, and 3.8 percent for quarters $t+1$ through $t+4$, respectively.

2.3 Diversification predicts lower tail risk in subsamples.

In Table 5, we examine whether the effect of diversification on tail risk varies over time, specifically in differing economic conditions and across periods surrounding the global financial crisis. We define "bad economic times" as quarters marked by NBER-defined recessions or notable financial crises (e.g., the failure of Long-Term Capital Management, the Russian sovereign debt crisis). Recessionary quarters are identified by the NBER Business Cycle Dating Committee.

Table 5 demonstrates that both in good and bad economic conditions, as well as before and after the crisis, diversification remains significantly negatively correlated with future tail risk in bank stock returns. The diversification coefficient is -0.089 in good economic times and more than doubles to -0.210 in bad economic times, suggesting that diversification's predictive power for reduced tail risk is enhanced during periods of financial distress. This coefficient in good periods is statistically significant at the 1% level (t-statistic = -2.73), while in bad periods, it is significant at the 5% level (t-statistic = -1.96).

Additionally, the third and fourth columns of Table 5 reveal that diversification's predictive capacity for lower tail risk diminishes post-Great Recession. Although potential explanations are explored, results are not included here. The pre-crisis period spans from 1996Q3 to 2007Q3, while the post-crisis period is defined from 2009Q3 to 2020Q4. In the pre-crisis period, the diversification coefficient is -0.134, significant at the 1% level (t-statistic = -2.66). In contrast, during the post-crisis years, the coefficient is -0.08 and statistically insignificant (t-statistic = -1.1).

2.4 Related diversification predicts lower tail risk.

In Table 6, we decompose overall diversification into related and unrelated diversification. Related diversification pertains to activities across core business lines—specifically, the seven lines that generate interest income. In contrast, unrelated diversification refers to diversification across noncore business lines, encompassing nine lines that generate noninterest income. The literature has yet to address which type of diversification is associated with reduced tail risk. To the best of our knowledge, this paper is the first to empirically decompose and test these types of diversification in relation to tail risk prediction. Columns (1) to (5) present five alternative measures of tail risk. For each measure, we conduct a horserace regression to determine whether related or unrelated diversification significantly impacts tail risk reduction. Results indicate that across all columns, only the coefficients for related diversification are statistically significant, with each reaching significance at the 1% level, while those for unrelated diversification are not.

In column (1), for one baseline measure of tail risk, a one-standard-deviation increase in related diversification corresponds to a 0.075 percentage point reduction in tail risk in the following quarter. This reduction translates to an economic magnitude of 1.64% relative to the sample mean of 4.57%, which is smaller than the economic impact of overall diversification (2.5%) as reported in column (1) of Table 3. This finding is novel as it distinguishes between related and unrelated diversification, showing that only related diversification across banks' core, interest-generating lines is linked to reduced tail risk. Conversely, diversification into noncore, noninterest-generating lines does not show this association. As banks increasingly venture into noncore business lines, these results provide important insights into the potential risk implications of such diversification behavior.

2.5 Channels through which diversification predicts lower tail risk.

In Table 7, we explore the mechanisms driving the primary result that diversification is associated with lower tail risk, examining two main channels that could explain this relationship: the hedging channel and the business model channel. According to the hedging channel, banks pursue diversification as a strategy to protect themselves against potential tail risks, adapting their practices by learning from unexpected adverse events. In this view, a bank that experiences a significant negative shock would diversify further, hedging against the possibility of similar future risks. This channel implies that diversification is proactive and responsive, guided by the institution's ability to learn from past experiences and mitigate potential future shocks.

In contrast, the business model channel suggests that a bank's inherent risk culture, shaped by its operational structure and strategic priorities, influences both its approach to risk and its diversification strategies. Here, diversification and risk are outcomes of a bank's broader, more rigid approach to risk management rather than responses to particular market events. This rigid risk culture implies that banks are slow to adjust to unexpected adverse experiences, such as financial crises, as their strategic orientation and established practices prioritize stability over adaptability. Consequently, institutions may continue their diversification strategies without modifying them in response to external shocks, even if such shocks reveal potential vulnerabilities.

Using a sample period spanning from 1998 to 2009, we analyze banks' responses to the 1998 Russian Financial Crisis, a significant event that presented unforeseen tail risks. We find that banks that exhibited higher tail risk during this period subsequently showed lower or unchanged levels of diversification. This empirical evidence supports the business model channel over the hedging channel, as banks did not increase diversification in response to the crisis. Instead, their diversification practices appear more aligned with an inflexible risk culture that limits adaptation to new risk insights. These findings suggest that rather than adjusting in response to market shocks, banks' diversification choices reflect deeper organizational traits and strategic priorities embedded within their business models.

2.6 Diversification predicts higher returns.

In Table 8, we demonstrate that diversification not only predicts lower tail risk but is also associated with higher stock returns. In column (1), we use diversification to predict quarterly buy-and-hold excess returns. In column (2), diversification is used to predict quarterly returns calculated with residual returns from the quarterly CAPM regression. In column (3), we use diversification to predict quarterly returns based on residual returns from the quarterly Fama and French [1993] regression, and in column (4), diversification predicts quarterly returns based on residual returns from the quarterly Fama and French [2015] regression. Across all columns, return variables are transformed using the natural logarithm.

In column (1), a one standard deviation increase in diversification is associated with a 0.502 increase in quarterly excess return, statistically significant at the 5 percent level with a t-statistic of 2.10, indicating an economic magnitude of 65%. In column (3), a one standard deviation increase in diversification is associated with a 0.243 increase in quarterly abnormal return based on the Fama and French [1993] model, with a similar effect size observed in the CAPM and Fama and French [2015] models. This result is statistically significant at the 1 percent level, with a t-statistic of 2.88 and an economic magnitude of 27%. Overall, a one standard deviation increase in diversification corresponds to a 65% increase in quarterly excess returns and a 27% increase in abnormal quarterly returns.

2.7 Diversification predicts higher returns three quarters ahead.

In Table 9, we provide detailed evidence that diversification serves as a robust predictor of stock returns not only for the immediate subsequent quarter but also across the following three quarters. The dependent variable, used to measure returns, is defined as the natural logarithm of the buy-and-hold excess stock return over each quarter.

For the first quarter following the measurement period ($t + 1$), the coefficient on diversification is estimated at 0.502, and this value is statistically significant at the 5% level, with a t-statistic of 2.12. This suggests that diversification positively impacts returns in the immediate quarter, likely reflecting an initial market response to diversified risk profiles. Moving to the second quarter ($t+2$), the coefficient on diversification increases substantially to 0.755, reaching significance at the 1% level (t-statistic = 3.08). This notable rise in both the coefficient size and statistical significance implies that the positive effects of diversification continue to strengthen as time progresses.

In the third quarter ($t + 3$), the coefficient on diversification climbs further to 0.812, remaining significant at the 1% level with a t-statistic of 3.03. By the fourth quarter ($t + 4$), however, the coefficient on diversification no longer reaches statistical significance, indicating that the predictive power of diversification on returns may begin to diminish as the time horizon extends beyond three quarters.

From an economic perspective, the results reveal substantial magnitudes associated with diversification. Specifically, a one-standard-deviation increase in diversification correlates with a 65% increase in returns in the next quarter, a 113% increase in returns two quarters ahead, and a 125% increase in returns three quarters ahead. These findings highlight not only the immediate but also the enduring impact of diversification on stock returns, particularly within the first three quarters following its measurement.

2.8 Related diversification predicts higher returns.

In Table 10, we further decompose overall diversification into related and unrelated diversification. Related diversification refers to diversification across the seven core business lines that generate interest income, while unrelated diversification refers to diversification across the nine noncore business lines that generate noninterest income. Ex ante, it is unclear which type of diversification may yield higher stock returns. To investigate, we conduct a horserace regression between related and unrelated diversification. The results indicate that unrelated diversification is not associated with higher returns; instead, it is related diversification that is positively associated with returns. This finding can be connected to the results in 2.4, which show that related diversification is associated with lower tail risk, whereas unrelated diversification is not.

In Table 10, across all four columns, the coefficients of unrelated diversification are consistently statistically insignificant. Moreover, for abnormal returns in columns (2), (3), and (4), the coefficient for unrelated diversification even becomes negative. In contrast, the coefficients for related diversification are statistically significant across all columns. In column (1), where the dependent variable is the natural logarithm of the buy-and-hold stock excess return over a quarter, the coefficient for related diversification is 0.415, statistically significant at the 5% level (t-statistic = 2.3). This coefficient implies an economic magnitude of 52% higher quarterly returns for a one-standard-deviation increase in related diversification. In column (3), where the dependent variable is the natural logarithm of the quarterly residual return based on the Fama and French [1993] model, the coefficient for related diversification is 0.171, statistically significant at the 1% level (t-statistic = 2.64). This translates to an economic magnitude of 18% higher quarterly returns for a one-standard-deviation increase in related diversification.

In conclusion, alongside the findings in 2.4, these results demonstrate that related diversification is associated with both lower tail risk and higher returns, whereas unrelated diversification exhibits no statistically significant relationship with either tail risk or returns. This finding is novel and has not been previously documented in the literature.

2.9 How does diversification affect the risk-return trade-off?

In Tables 11 and 12, we examine how diversification impacts the risk-return trade-off. The hypothesis is that the level of diversification optimal for reducing tail risk may not coincide with that optimal for maximizing return. This section tests that idea by analyzing these nonlinear effects.

It is reasonable to propose that maximum diversification does not necessarily result in the greatest tail risk reduction. After a certain point, diversification's effect on mitigating tail risk may diminish. To identify this point, we include squared diversification in the regression to capture a potential nonlinear relationship between diversification and tail risk. In Table 11, for columns (1) through (3), where tail risk is measured using excess returns and residual returns from the CAPM, the coefficient of diversification is negative and statistically significant at the 1% level, while squared diversification is not statistically significant. However, in columns (4) and (5), where tail risk is measured on residual returns based on the Fama and French [1993] and Fama and French [2015] models, the squared diversification coefficient is positive while the diversification coefficient remains negative. For example, a simple calculation for column (4) suggests that the level of diversification associated with maximum tail risk reduction is 2.14 standard deviations above the sample mean. Recall that diversification levels range between 0 and 2.77.

Similarly, maximum diversification may not yield the greatest return. Beyond a certain point, the positive effect of diversification on returns may start to decline. To assess this, we again introduce squared diversification in the regression to explore a potential nonlinear relationship between diversification and returns. In Table 12, across all columns, the diversification coefficient is positive, while the squared diversification coefficient is negative. In column (4), for instance, where the dependent variable is the natural logarithm of quarterly residual return based on the Fama-French [1993] model, a basic calculation indicates that the level of diversification associated with maximum return enhancement is 1.28 standard deviations above the sample mean.

Summarizing these findings, the level of diversification associated with maximum tail risk reduction is 2.14 standard deviations above the sample mean, while that associated with maximum return enhancement is 1.28 standard deviations above the sample mean. This suggests that a diversification level between 1.28 and 2.14 standard deviations above the mean achieves a balance between risk reduction and return enhancement. For example, when the diversification level rises above 1.28 standard deviations, return is no longer maximized but approaches the level where tail risk is minimized (at 2.14), resembling a hump-shaped relationship between diversification and return with a peak at 1.28. Conversely, if the diversification level falls from 2.14 standard deviations, tail risk is no longer minimized but trends toward the level where return is maximized (at 1.28), suggesting a U-shaped relationship between diversification and tail risk with a minimum at 2.14.

2.10 Diversification predicts greater changes in loan supply.

In Table 13, we present evidence that diversification is a robust and significant predictor of increases in loan supply, supporting the hypothesis that diversified banks are more likely to expand lending activities. Following well-established models in the banking literature, we employ a comprehensive set of controls to account for various economic and bank-specific factors that may influence loan supply changes. Specifically, we control for four lags of loan supply changes to capture persistent trends in lending, as well as four lags of changes in the federal funds rate to account for the effects of monetary policy. Additionally, we include four lags of GDP growth in columns (1) and (2) to account for macroeconomic conditions, while in columns (3) and (4), we employ year-quarter fixed effects as an alternative to capture time-specific effects and broader economic shifts. Across all specifications, we include bank fixed effects to control for unobserved heterogeneity among banks, ensuring that the observed relationship between diversification and loan supply is not driven by bank-specific characteristics that remain constant over time.

In columns (1) and (3), which do not include additional bank-level controls, the results indicate a strong and statistically significant association between diversification and loan supply changes, suggesting that banks with higher diversification tend to increase their lending activity. This relationship remains significant in columns (2) and (4), where we introduce a set of eight additional bank-level control variables, including measures of bank size, capital adequacy, profitability, and liquidity, to address potential confounding factors. Even with these additional controls, diversification continues to be a significant predictor of increased loan supply.

Economically, the magnitude of this effect is notable. In column (4), a one standard deviation increase in diversification is associated with a 14% increase in loan supply changes in the subsequent quarter, indicating that diversification has a meaningful impact on banks' lending behavior. This finding underscores the role of diversification as an important factor in banks' lending decisions, with implications for how diversification may support financial stability and credit availability in the economy.

2.11 Diversified banks show lending resilience in the 2007-2008 crisis.

In Table 14, we demonstrate that diversified banks not only predict greater increases in loan supply but also exhibit lending resilience during the 2007-2008 financial crisis. To assess this resilience, we fix diversification and eight bank-level control variables at their values in the fourth quarter of 2006, prior to the onset of the Great Recession. By maintaining these pre-crisis levels, we can isolate the impact of diversification on lending behavior during the crisis, providing a clearer picture of how diversification affects loan supply under financial stress.

In this analysis, we scale loan supply by total assets as measured in the fourth quarter of 2006 to standardize across banks and focus on relative changes in lending. Furthermore, we examine not only total loans but also specific loan categories, including commercial and industrial loans and real estate loans, to understand if diversification influences various lending activities differently. This approach allows for a comprehensive view of how diversified banks maintained their lending capacity across sectors during a period of economic turmoil.

The results are striking. A one standard deviation increase in pre-crisis diversification is associated with a 4.9 percentage point increase in total loans, scaled by pre-crisis assets. For commercial and industrial loans, this diversification effect corresponds to a 1.7 percentage point increase, and for real estate loans, a 2 percentage point increase, both scaled by pre-crisis assets. These findings underscore the role of diversification in promoting lending resilience, suggesting that banks with broader diversification were better positioned to sustain their loan portfolios during the financial crisis.

2.12 Diversification predicts better accounting performance.

In Table 15, we present evidence that diversification predicts improved accounting performance. Specifically, in column (1), we examine whether diversification is associated with a higher return on assets (ROA) in the subsequent quarter. Return on assets is defined as net income divided by total assets. We find that a one standard deviation increase in diversification corresponds to a 0.011 percentage point increase in return on assets, with the coefficient statistically significant at the one percent level (t-statistic = 4.54). In terms of economic significance, this increase translates to a five percent rise in future ROA relative to the sample mean of 0.22 percent. This result indicates that diversification has a meaningful impact on ROA, suggesting that banks that diversify more extensively can achieve higher profitability relative to their asset base.

In column (2), we extend this analysis to test whether diversification is associated with a higher return on equity (ROE) in the subsequent quarter. Return on equity is calculated as net income divided by total equity. The results indicate that a one standard deviation increase in diversification is associated with a 0.151 percentage point rise in ROE. This coefficient is also statistically significant at the one percent level, with a t-statistic of 4.6, underscoring the robustness of the association between diversification and enhanced equity returns. Economically, this finding translates to a 6.5 percent increase in future ROE relative to the sample mean of 2.32 percent. The positive relationship between diversification and ROE implies that banks benefit from greater equity efficiency, as diversification allows them to generate higher returns on their equity investments.

Finally, in column (3), we assess whether diversification is associated with an improved Z-score in the subsequent quarter, where the Z-score is calculated as the natural logarithm of the ratio of the sum of return on assets and the capital-to-assets ratio to the standard deviation of return on assets over a rolling 12-quarter window. This metric provides an indication of default risk, with higher Z-scores corresponding to lower default risk. We find that a one standard deviation increase in diversification is associated with a 0.006 unit increase in the Z-score, with the coefficient statistically significant at the one percent level (t-statistic = 4.13). This result represents a 1.4 percent increase in the future Z-score, signifying that higher diversification levels are associated with lower default risk for banks. The economic significance of this result indicates that, by diversifying their activities, banks can enhance their stability and resilience against default.

Overall, the findings in Table 15 demonstrate that diversification is positively associated with better accounting performance across multiple metrics. Diversification leads to higher return on assets, higher return on equity, and a reduction in default risk, as reflected by an increase in the Z-score. These results have substantial economic implications, underscoring the value of diversification in enhancing bank performance and stability. By diversifying their operations, banks can achieve greater profitability, equity efficiency, and risk resilience, thereby improving their overall financial health.

3 Conclusion

This paper provides substantial evidence on the significant role of business line diversification in shaping banks' risk and return profiles, offering insights into both the benefits and limits of diversification strategies. By employing a detailed entropy-based measure, the study finds that diversification across multiple business lines—especially within core, interest-generating activities—has a considerable effect on reducing tail risk and enhancing overall bank performance. Specifically, a one-standard-deviation increase in business line diversification leads to a 2.5% decrease in tail risk in the subsequent quarter, an effect that persists for up to four quarters, thereby demonstrating the stability that a diversified business model can bring to banks' risk management.

The benefits of diversification extend beyond risk reduction. This analysis also shows that banks with diversified operations enjoy higher future stock returns, improved profitability, lower default risk, and increased lending capacity. For instance, diversification correlates with a significant rise in excess stock returns—up to 125% over three quarters—while also enhancing banks' ability to lend, particularly during times of financial stress such as the Great Recession. This finding suggests that diversification not only aids banks in stabilizing their income streams but also enables them to sustain essential lending functions, which are critical for broader economic stability during downturns.

A notable contribution of this study is its differentiation between related and unrelated diversification. By decomposing diversification into related (core, interest-generating) and unrelated (noncore, noninterest-generating) categories, the study reveals that only related diversification significantly reduces tail risk and enhances returns. Specifically, related diversification within core business lines is associated with a 1.64% reduction in tail risk and a substantial increase in returns, underscoring its effectiveness in strengthening a bank's financial stability and resilience. Conversely, unrelated diversification into noncore activities fails to provide similar risk or return benefits, and in some cases, can even detract from stability. This finding challenges the conventional notion that broader diversification across various activities is inherently beneficial for banks, highlighting the importance of focusing on core activities when developing diversification strategies.

The study also reveals a non-linear relationship between diversification and performance, suggesting an optimal level where diversification's benefits are maximized. The analysis indicates that while increasing diversification reduces tail risk up to a certain point (2.14 standard deviations above the mean), its effect diminishes beyond this level. Similarly, diversification's positive impact on returns peaks at a moderate level (1.28 standard deviations above the mean), after which it declines. This non-linear relationship underscores the need for a balanced approach to diversification, where banks maximize risk reduction without diluting returns. These insights imply that management should be cautious in pursuing excessive diversification, as it may lead to diminishing returns on risk mitigation and potentially reduce overall financial performance.

The findings of this study carry valuable implications for bank management, investors, and policymakers. For bank managers, focusing on related diversification within core business lines offers a path to mitigate risk and achieve sustainable growth, enhancing the bank's resilience in both stable and volatile market conditions. For policymakers, understanding the specific benefits of related diversification may inform regulatory policies that encourage banks to diversify strategically without overextending into riskier, unrelated activities. These results suggest that regulation should consider promoting core-focused diversification rather than imposing blanket limitations on diversification, thus supporting banks' ability to absorb economic shocks while maintaining lending capabilities that are vital for financial stability.

In addition to these practical implications, this study contributes to the existing literature on diversification by addressing gaps in the understanding of how different types of diversification impact bank risk and return profiles. Unlike prior studies, this paper is the first to use a highly detailed measure of diversification across 16 distinct business lines, allowing for a nuanced view of diversification's effects across varied revenue sources. It also enhances the understanding of how related diversification serves as a more effective risk management tool than unrelated diversification. By capturing the differential impacts of diversification across economic cycles and before and after the global financial crisis, this study provides a comprehensive picture of how diversification strategies can adapt to changing market conditions.

Future research could expand on these findings by exploring how the effects of related and unrelated diversification vary across different regulatory environments, bank sizes, and economic contexts. Additionally, examining the impact of diversification on smaller banks, which may face different challenges in achieving optimal diversification, could provide further insights into tailoring diversification strategies. Extending the analysis to international markets with varying regulatory structures could also reveal how diversification impacts banks operating under different economic conditions, potentially offering a broader understanding of diversification's role in global banking stability and performance.

In conclusion, this study sheds light on the importance of business line diversification in reducing tail risk, boosting returns, and enhancing bank stability. It underscores that related diversification within core activities provides banks with a robust framework for managing risk and capitalizing on growth opportunities, while unrelated diversification may compromise financial stability. These findings have meaningful implications for bank managers, regulators, and stakeholders, highlighting the critical balance between diversification and specialization to ensure resilience and sustained growth in a dynamic financial environment.

References

Viral V Acharya, Iftekhar Hasan, and Anthony Saunders. Should banks be diversified? evidence from individual bank loan portfolios. *Journal of Business*, 79(3):1355–1412, 2006.

Lieven Baele, Olivier De Jonghe, and Rudi Vander Vennet. Does the stock market value bank diversification? *Journal of Banking and Finance*, 31(7):1999–2023, 2007.

Philip G Berger and Eli Ofek. Diversification's effect on firm value. *Journal of Financial Economics*, 37(1):39–65, 1995.

Jose Manuel Campa and Simi Kedia. Explaining the diversification discount. *Journal of Finance*, 57(4):1731–1762, 2002.

Rebecca S Demsetz and Philip E Strahan. Diversification, size, and risk at bank holding companies. *Journal of Money, Credit, and Banking*, pages 300–313, 1997.

Saiying Deng and Elyas Elyasiani. Geographic diversification, bank holding company value, and risk. *Journal of Money, Credit and Banking*, 40(6):1217–1238, 2008.

Ralf Elsas, Andreas Hackethal, and Markus Holzhäuser. The anatomy of bank diversification. *Journal of Banking and Finance*, 34(6):1274–1287, 2010.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

Priyank Gandhi, Darius Palia, and Jasper Pan. Do "measures" of bank diversification measure up? *Available at SSRN 4952789*, 2025.

Michael Gelman, Itay Goldstein, and Andrew MacKinlay. Bank diversification and lending resiliency. *Available at SSRN 4147790*, 2023.

Martin R Goetz, Luc Laeven, and Ross Levine. Identifying the valuation effects and agency costs of corporate diversification: Evidence from the geographic diversification of us banks. *Review of Financial Studies*, 26(7):1787–1823, 2013.

Martin R Goetz, Luc Laeven, and Ross Levine. Does the geographic expansion of banks reduce risk? *Journal of Financial Economics*, 120(2):346–362, 2016.

Rebecca N Hann, Maria Ogneva, and Oguzhan Ozbas. Corporate diversification and the cost of capital. *Journal of Finance*, 68(5):1961–1999, 2013.

Alexis P Jacquemin and Charles H Berry. Entropy measure of diversification and corporate growth. *Journal of Industrial Economics*, 27(4):359–369, 1979.

Tarun Khanna and Krishna Palepu. Is group affiliation profitable in emerging markets? an analysis of diversified indian business groups. *Journal of Finance*, 55(2):867–891, 2000.

Venkat Kuppuswamy and Belén Villalonga. Does diversification create value in the presence of external financing constraints? evidence from the 2007–2009 financial crisis. *Management Science*, 62(4):905–923, 2016.

Luc Laeven and Ross Levine. Is there a diversification discount in financial conglomerates? *Journal of Financial Economics*, 85(2):331–367, 2007.

Larry HP Lang and Rene M Stulz. Tobin's q, corporate diversification, and firm performance. *Journal of Political Economy*, 102(6):1248–1280, 1994.

Ross Levine, Chen Lin, and Wensi Xie. Geographic diversification and banks' funding costs. *Management Science*, 67(5):2657–2678, 2021.

Anthony Saunders, Markus Schmid, and Ingo Walter. Strategic scope and bank performance. *Journal of Financial Stability*, 46:100715, 2020.

Jeungbo Shim. Loan portfolio diversification, market structure and bank stability. *Journal of Banking and Finance*, 104:103–115, 2019.

Kevin J Stiroh. Diversification in banking: Is noninterest income the answer? *Journal of Money, Credit and Banking*, pages 853–882, 2004.

Kevin J Stiroh and Adrienne Rumble. The dark side of diversification: The case of us financial holding companies. *Journal of Banking and Finance*, 30(8):2131–2161, 2006.

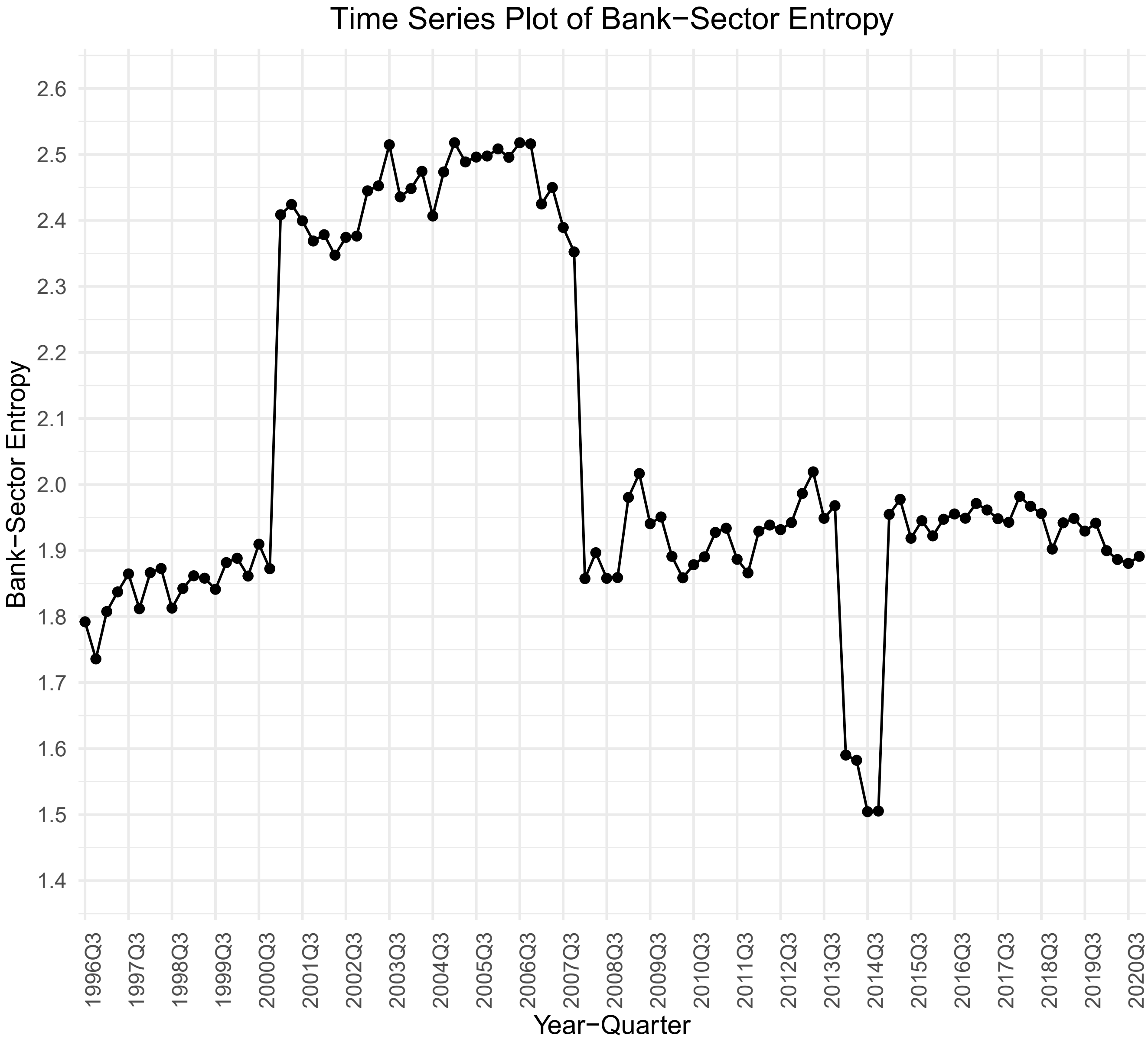

Figure 1: Time series plot of bank-sector entropy.

Notes: This figure shows the time series of bank-sector entropy. Banks must exist in both the current and previous quarters. For each quarter and each of the 16 income items, we aggregate income across all eligible banks, then apply the entropy formula to compute the bank-sector entropy for that specific quarter.

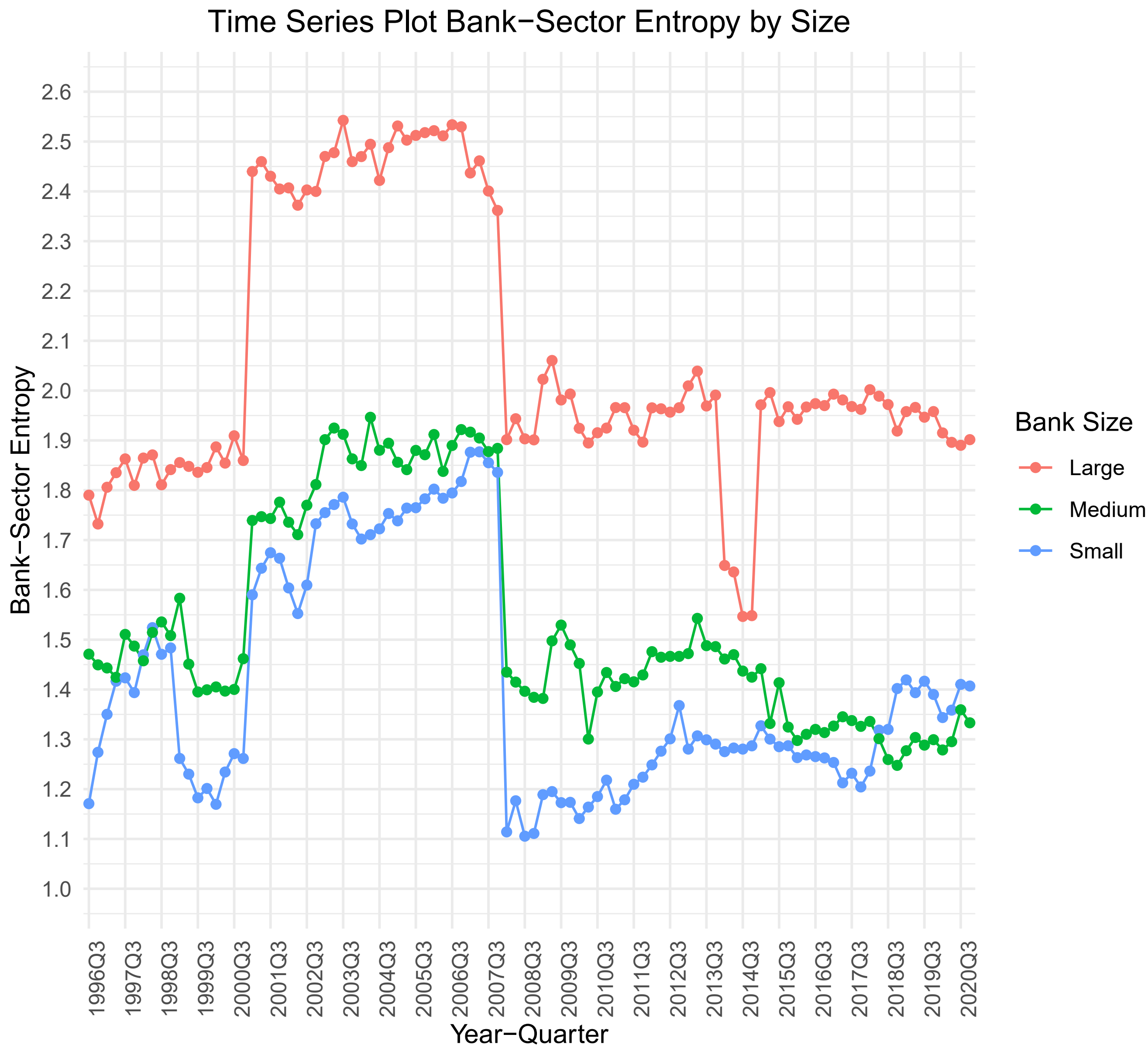

Figure 2: Bank-sector entropy by size.

Notes: This figure illustrates the time series of bank-sector entropy categorized into different size groups. As defined by the Federal Deposit Insurance Corporation, small banks are those with assets less than 1.503 billion, large banks have assets exceeding 10 billion, and medium banks fall in between these thresholds. Each quarter, banks are categorized into size groups based on their total assets. To calculate entropy, banks must exist in both the current and previous quarters. For each quarter and each of the 16 income items, we aggregate income across all eligible banks, then apply the entropy formula to compute the bank-sector entropy for that specific quarter. This process is repeated for each size group.

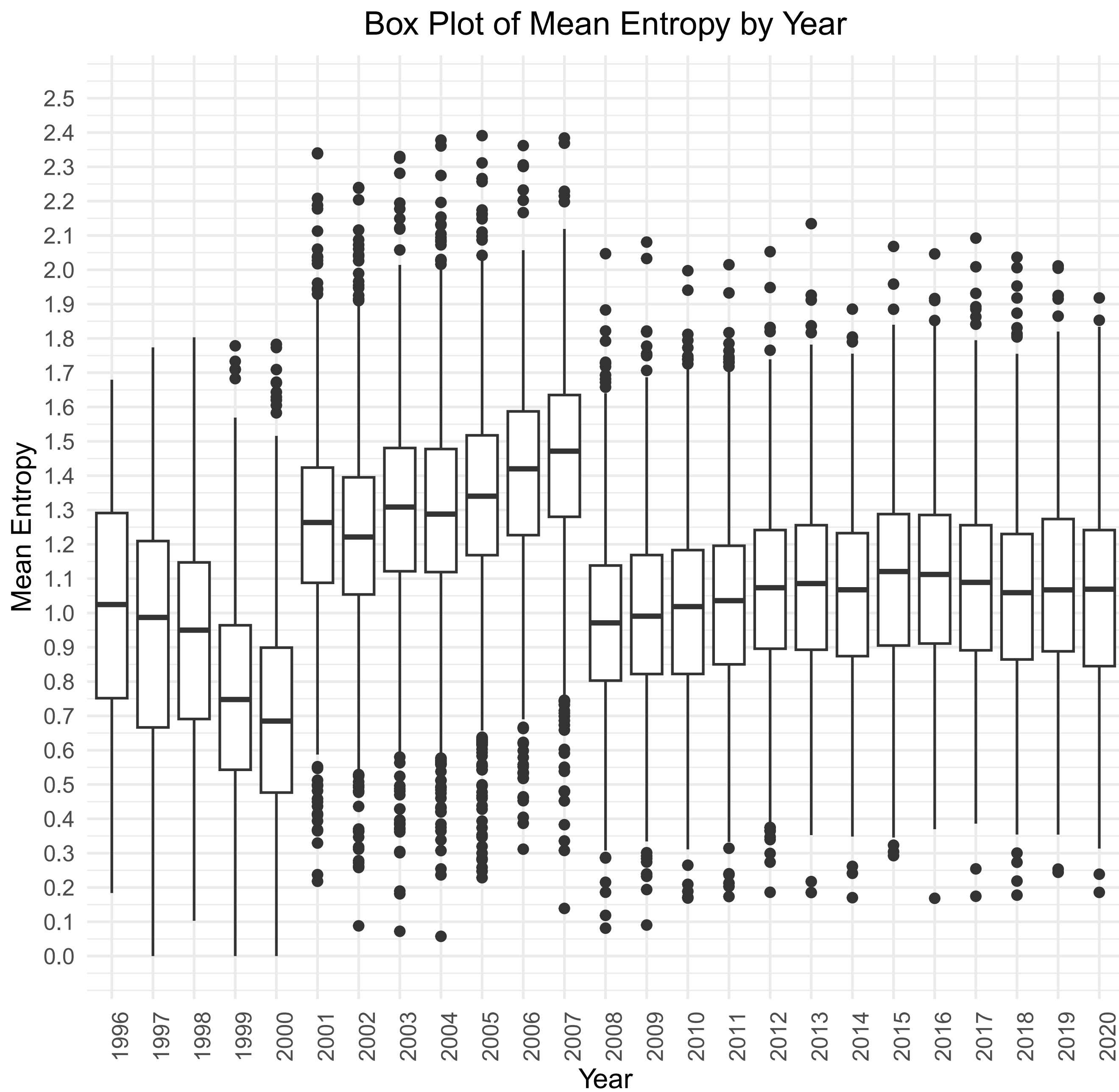

Figure 3: Entropy's cross-sectional distribution over time

Notes: This figure presents the time-evolving cross-sectional distribution of entropy. We compute the mean entropy for every bank annually. Subsequently, for each year, a box plot is utilized to depict the entropy's cross-sectional spread.

Table 1: Summary statistics and correlations.

Notes: The Panel A of this table shows the summary statistics for all variables in the main sample. The columns report mean, standard deviation, minimum, 25^{th}-percentile, median, 75^{th}-percentile, and maximum values, and the number of observations. The Panel B of this table shows the correlations of key variables. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Related Diversification* is the Entropy Index of seven interest income items. *Unrelated Diversification* is the Entropy Index of nine noninterest income items. *Tail Risk 5%* is the negative of the average excess return during the worst 5% return days over a quarter. *Tail Risk 10%* is the negative of the average excess return during the worst 10% return days over a quarter. *CAPM Tail Risk* is the negative of the average CAPM residual return during the worst 5% return days over a quarter. *FF3 Tail Risk* is the negative of the average Fama and French [1993] residual return during the worst 5% return days over a quarter. *FF5 Tail Risk* is the negative of the average Fama and French [2015] residual return during the worst 5% return days over a quarter. *Qr Return* is the natural logarithm of the buy-and-hold stock excess return over a quarter. *CAPM Abnormal Return* is the natural logarithm of the quarterly residual return based on the CAPM. *FF3 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [1993] model. *FF5 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [2015] model. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interest income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. *Δlog(Total Loans)* is the quarter-on-quarter change of the logarithm of total loans. *ROA* is the ratio of net income to total assets. *ROE* is the ratio of net income to total equity in percentage. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels. The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

Panel A: Summary statistics

	Mean	SD	Min	p25	Median	p75	Max	N
Diversification	1.18	0.36	0.01	0.95	1.19	1.43	2.47	23809
Related Diversification	0.46	0.27	0.00	0.27	0.43	0.60	1.61	23809
Unrelated Diversification	1.06	0.37	0.00	0.82	1.10	1.34	2.02	23809
Tail Risk 5%	4.57	3.23	0.20	2.64	3.62	5.25	42.25	23809
Tail Risk 10%	3.83	2.65	0.16	2.24	3.04	4.41	34.14	23809
CAPM Tail Risk	4.02	2.92	0.19	2.26	3.14	4.74	40.60	23809
FF3 Tail Risk	3.73	2.83	0.18	2.03	2.90	4.40	39.74	23809
FF5 Tail Risk	3.63	2.78	0.00	1.96	2.82	4.27	39.31	23809
Qr Return	0.17	18.72	-203.80	-6.48	1.32	9.21	104.86	23809
CAPM Abnormal Return	-2.05	4.99	-185.50	-1.78	-0.80	-0.43	-0.00	23809
FF3 Abnormal Return	-1.82	4.74	-183.10	-1.51	-0.69	-0.34	-0.00	23809
FF5 Abnormal Return	-1.74	4.59	-185.04	-1.44	-0.65	-0.32	0.00	23809
Log Assets	14.97	1.71	11.97	13.72	14.57	15.84	21.94	23809
Capital to Assets	9.00	1.98	4.91	7.76	8.77	9.95	16.75	23809
Operating Profits	0.41	0.22	-0.43	0.30	0.40	0.51	1.16	23809
Cost to Income	74.89	12.64	48.61	67.30	73.93	80.48	135.69	23809
Deposits to Liabilities	83.83	11.59	33.22	78.78	86.59	91.92	98.97	23809
Loan Loss Provisions	0.15	0.26	-0.11	0.03	0.07	0.15	1.60	23809
Assets Growth	50.99	66.71	-26.46	13.15	32.11	64.11	396.18	23809
Z-Score	4.10	0.46	2.66	3.86	4.18	4.43	4.92	23809
VIX	20.57	8.07	10.31	14.23	19.32	25.09	58.59	23809
Market Beta	0.68	0.49	-0.11	0.25	0.66	1.05	1.94	23809
Δlog(Total Loans) $\times$ 100	1.97	6.59	-210.65	-0.51	1.41	3.46	192.22	62096
ROA	0.22	0.24	-1.11	0.16	0.25	0.33	0.81	70683
ROE	2.32	3.37	-20.21	1.72	2.72	3.68	8.37	70683

Panel B: Correlations

	Diversification	Related Diver	Unrelated Diver	Log Assets	Tail Risk 5%	Qr Return
Diversification	1.00					
Related Diver	0.55***	1.00				
Unrelated Diver	0.57***	0.21***	1.00			
Log Assets	0.30***	0.39***	0.29***	1.00		
Tail Risk 5%	-0.24***	-0.04***	-0.09***	-0.08***	1.00	
Qr Return	0.06***	0.01	0.01**	-0.00	-0.46***	1.00

Table 2: Determinants of diversification.

Notes: This table shows the estimated coefficients for the forecasting regression:

$$Diver_{i,t+1} = \alpha_i + \beta_i Determinants_{i,t} + \epsilon_{i,t}$$

$Diver_{i,t+1}$ is the Entropy Index of seven interest income and nine noninterest income items, calculated for bank i at time $t+1$. $Determinants_{i,t}$ includes the following eight bank-level variables. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels. The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Diver(t+1)	(2) Diver(t+1)	(3) Diver(t+1)	(4) Diver(t+1)	(5) Diver(t+1)	(6) Diver(t+1)	(7) Diver(t+1)	(8) Diver(t+1)	(9) Diver(t+1)
Log Assets	0.081***								0.088***
	(9.94)								(11.11)
Capital to Assets		-0.044***							-0.054***
		(-7.71)							(-9.96)
Operating Profits			0.023***						0.074***
			(3.64)						(5.21)
Cost to Income				-0.015***					0.063***
				(-2.94)					(4.69)
Deposits to Liabilities					-0.045***				-0.003
					(-5.06)				(-0.42)
Loan Loss Provisions						-0.045***			-0.047***
						(-11.20)			(-12.13)
Assets Growth							-0.019***		-0.051***
							(-4.58)		(-12.47)
Z-Score								0.036***	0.039***
								(7.75)	(8.17)
Observations	68938	68938	68938	68938	68938	68938	68938	68938	68938
Adjusted R^2	0.054	0.016	0.004	0.002	0.017	0.016	0.003	0.011	0.124

Table 3: The impact of diversification on tail risk.

Notes: This table shows the estimated coefficients for the following regressions:

$$Tail\,Risk_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Tail\,Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Tail\,Risk_{i,t+1}$ includes the following five tail risk variables. *Tail Risk 5%* is the negative of the average excess return during the worst 5% return days over a quarter. *Tail Risk 10%* is the negative of the average excess return during the worst 10% return days over a quarter. *CAPM Tail Risk* is the negative of the average CAPM residual return during the worst 5% return days over a quarter. *FF3 Tail Risk* is the negative of the average Fama and French [1993] residual return during the worst 5% return days over a quarter. *FF5 Tail Risk* is the negative of the average Fama and French [2015] residual return during the worst 5% return days over a quarter. We control for Tail Risk$_{i,t}$ in all five specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the *t*-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk 5%(t+1)	(2) Tail Risk 10%(t+1)	(3) CAPM Tail Risk(t+1)	(4) FF3 Tail Risk(t+1)	(5) FF5 Tail Risk(t+1)
Diversification	-0.112***	-0.087***	-0.133***	-0.129***	-0.129***
	(-3.23)	(-3.29)	(-3.88)	(-3.85)	(-3.91)
Tail Risk 5%(t)	1.180***				
	(24.72)				
Tail Risk 10%(t)		1.146***			
		(30.77)			
CAPM Tail Risk(t)			1.142***		
			(24.43)		
FF3 Tail Risk(t)				1.141***	
				(24.71)	
FF5 Tail Risk(t)					1.122***
					(24.92)
Log Assets	-0.152	-0.084	-0.372***	-0.457***	-0.461***
	(-1.45)	(-1.08)	(-3.49)	(-4.30)	(-4.42)
Capital to Assets	-0.292***	-0.222***	-0.329***	-0.327***	-0.319***
	(-9.44)	(-9.64)	(-10.89)	(-10.95)	(-10.94)
Operating Profits	-0.092*	-0.068*	-0.089*	-0.066	-0.060
	(-1.77)	(-1.73)	(-1.74)	(-1.34)	(-1.24)
Cost to Income	0.205***	0.159***	0.189***	0.214***	0.227***
	(3.54)	(3.56)	(3.39)	(3.93)	(4.24)
Deposits to Liabilities	0.057	0.040	0.060	0.063	0.058
	(1.49)	(1.38)	(1.54)	(1.64)	(1.51)
Loan Loss Provisions	0.465***	0.350***	0.420***	0.408***	0.398***
	(14.38)	(14.26)	(13.13)	(12.91)	(12.77)
Assets Growth	-0.020	-0.014	-0.034*	-0.039**	-0.038**
	(-1.01)	(-0.97)	(-1.77)	(-2.10)	(-2.10)
Z-Score	-0.119***	-0.090***	-0.123***	-0.109***	-0.107***
	(-5.08)	(-5.12)	(-5.47)	(-4.96)	(-4.91)
VIX	0.955***	0.739***	0.734***	0.700***	0.693***
	(15.67)	(15.23)	(12.65)	(11.95)	(12.18)
Market Beta	0.163***	0.110***	0.069**	0.041	0.038
	(5.06)	(4.49)	(2.28)	(1.38)	(1.33)
Observations	23654	23654	23646	23644	23642
Adjusted R^2	0.697	0.737	0.642	0.621	0.620
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes

Table 4: The impact of diversification on tail risk four quarters ahead.

Notes: This table shows the estimated coefficients for the following regressions:

$$Tail\,Risk_{i,t+1/2/3/4} = \alpha_i + \beta_i Diversification_{i,t} + Tail\,Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, *Tail Risk* is the negative of the average excess return during the worst 5% return days over a quarter. We control for *Tail Risk* at quarter t in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects ($\sum \eta_i$) and year-quarter fixed effects ($\sum \gamma_t$). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk (t+1)	(2) Tail Risk (t+2)	(3) Tail Risk (t+3)	(4) Tail Risk (t+4)
Diversification	-0.112***	-0.171***	-0.188***	-0.173***
	(-3.27)	(-4.21)	(-4.13)	(-3.44)
Tail Risk (t)	1.180***	0.941***	0.751***	0.525***
	(25.02)	(17.86)	(11.94)	(10.14)
Log Assets	-0.152	0.180	0.409***	0.622***
	(-1.46)	(1.46)	(2.84)	(3.73)
Capital to Assets	-0.292***	-0.337***	-0.349***	-0.325***
	(-9.56)	(-9.28)	(-8.07)	(-6.95)
Operating Profits	-0.092*	-0.004	-0.084	-0.119
	(-1.79)	(-0.07)	(-1.24)	(-1.57)
Cost to Income	0.205***	0.253***	0.155**	0.113
	(3.58)	(4.07)	(2.27)	(1.57)
Deposits to Liabilities	0.057	0.087*	0.101*	0.113*
	(1.51)	(1.86)	(1.76)	(1.79)
Loan Loss Provisions	0.465***	0.409***	0.392***	0.421***
	(14.55)	(10.50)	(9.57)	(10.16)
Assets Growth	-0.020	-0.031	-0.029	-0.029
	(-1.02)	(-1.17)	(-0.89)	(-0.83)
Z-Score	-0.119***	-0.131***	-0.130***	-0.102***
	(-5.14)	(-4.68)	(-3.82)	(-2.86)
VIX	0.438**	-0.376	0.116	0.233***
	(2.11)	(-1.51)	(1.62)	(3.41)
Market Beta	0.163***	0.151***	0.100**	0.027
	(5.12)	(3.80)	(2.12)	(0.52)
Observations	23654	23394	23130	22856
Adjusted R^2	0.697	0.650	0.624	0.605
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table 5: The impact of diversification on tail risk in subsamples.

Notes: This table shows the estimated coefficients for the following regression using subsamples:

$$Tail\ Risk_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Tail\ Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Tail\ Risk_{i,t+1}$ is the negative of the average stock return during the worst 5% returns days for bank i in quarter $t+1$. We control for $Tail\ Risk_{i,t}$ in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. *Bad times* are defined as the failure of Long-Term Capital Management (LTCM) and the Russian Crisis in the first and second quarters of 1999 and the recessions dated by National Bureau of Economic Research (NBER). *Good times* are periods not defined as *Bad Times*. *Pre-Crisis* is defined as the period between 1996Q3 and 2007Q3. *Post-Crisis* is defined as the period between 2009Q3 and 2020Q4. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects ($\sum \eta_i$) and year-quarter fixed effects ($\sum \gamma_t$). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	Good Times Tail Risk (t+1)	Bad Times Tail Risk (t+1)	Pre-Crisis Tail Risk (t+1)	Post-Crisis Tail Risk (t+1)
Diversification	-0.089***	-0.210**	-0.134***	-0.080
	(-2.73)	(-1.96)	(-2.66)	(-1.10)
Tail Risk (t)	0.846***	0.982***	1.033***	0.838***
	(24.42)	(6.63)	(17.44)	(13.45)
Log Assets	-0.281**	0.146	0.135	0.504**
	(-2.51)	(0.41)	(0.69)	(2.15)
Capital to Assets	-0.299***	-0.217**	-0.056	-0.343***
	(-9.71)	(-2.39)	(-1.19)	(-5.66)
Operating Profits	-0.033	-0.211	0.007	-0.228***
	(-0.66)	(-1.09)	(0.07)	(-2.76)
Cost to Income	0.182***	0.265	0.282***	-0.044
	(3.21)	(1.20)	(2.69)	(-0.49)
Deposits to Liabilities	0.043	0.152	0.055	0.084
	(1.17)	(1.23)	(0.82)	(1.25)
Loan Loss Provisions	0.353***	0.741***	0.467***	0.410***
	(10.32)	(7.54)	(10.19)	(7.38)
Assets Growth	-0.025	0.017	0.032	-0.068*
	(-1.26)	(0.28)	(1.02)	(-1.91)
Z-Score	-0.084***	-0.250**	-0.069**	-0.076**
	(-3.89)	(-2.35)	(-2.21)	(-2.07)
VIX(t)	0.366***	1.679***	0.974***	0.518***
	(4.31)	(12.26)	(12.50)	(5.54)
Market Beta(t)	0.151***	0.170	0.243***	0.026
	(4.73)	(1.28)	(6.37)	(0.44)
Observations	19453	4201	12985	9720
Adjusted R^2	0.598	0.654	0.721	0.630
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fiexed Effects	Yes	Yes	Yes	Yes

Table 6: The impact of related and unrelated diversification on tail risk.

Notes: This table shows the estimated coefficients for the following regressions:

$$Tail\,Risk_{i,t+1} = \alpha_i + \beta_i Related\,Diversification_{i,t} + Unrelated\,Diversification_{i,t} + Tail\,Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Tail\,Risk_{i,t+1}$ includes the following five tail risk variables. *Tail Risk 5%* is the negative of the average excess return during the worst 5% return days over a quarter. *Tail Risk 10%* is the negative of the average excess return during the worst 10% return days over a quarter. *CAPM Tail Risk* is the negative of the average CAPM residual return during the worst 5% return days over a quarter. *FF3 Tail Risk* is the negative of the average Fama and French [1993] residual return during the worst 5% return days over a quarter. *FF5 Tail Risk* is the negative of the average Fama and French [2015] residual return during the worst 5% return days over a quarter. We control for $Tail\,Risk_{i,t}$ in all five specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. The numbers in parenthesis are the *t*-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk 5%(t+1)	(2) Tail Risk 10%(t+1)	(3) CAPM Tail Risk(t+1)	(4) FF3 Tail Risk(t+1)	(5) FF5 Tail Risk(t+1)
Related Diversification	-0.075***	-0.057***	-0.076***	-0.075***	-0.075***
	(-2.92)	(-2.96)	(-3.05)	(-3.05)	(-3.09)
Nonrelated Diversification	-0.046	-0.038	-0.043	-0.044	-0.043
	(-1.48)	(-1.55)	(-1.42)	(-1.47)	(-1.47)
Tail Risk 5%(t)	1.180***				
	(24.64)				
Tail Risk 10%(t)		1.146***			
		(30.65)			
CAPM Tail Risk(t)			1.144***		
			(24.47)		
FF3 Tail Risk(t)				1.143***	
				(24.75)	
FF5 Tail Risk(t)					1.124***
					(24.99)
Log Assets	-0.143	-0.076	-0.359***	-0.444***	-0.449***
	(-1.35)	(-0.98)	(-3.38)	(-4.20)	(-4.32)
Capital to Assets	-0.292***	-0.222***	-0.327***	-0.326***	-0.318***
	(-9.43)	(-9.60)	(-10.81)	(-10.88)	(-10.86)
Operating Profits	-0.100*	-0.075*	-0.097*	-0.075	-0.068
	(-1.88)	(-1.85)	(-1.85)	(-1.47)	(-1.37)
Cost to Income	0.201***	0.155***	0.184***	0.209***	0.222***
	(3.41)	(3.42)	(3.26)	(3.78)	(4.08)
Deposits to Liabilities	0.041	0.028	0.043	0.047	0.041
	(1.11)	(0.99)	(1.12)	(1.23)	(1.09)
Loan Loss Provisions	0.465***	0.350***	0.421***	0.408***	0.398***
	(14.34)	(14.21)	(13.10)	(12.87)	(12.73)
Assets Growth	-0.017	-0.012	-0.031	-0.036*	-0.035*
	(-0.86)	(-0.83)	(-1.63)	(-1.96)	(-1.96)
Z-Score	-0.121***	-0.092***	-0.125***	-0.112***	-0.109***
	(-5.17)	(-5.21)	(-5.57)	(-5.07)	(-5.01)
VIX	0.965***	0.746***	0.738***	0.704***	0.698***
	(15.67)	(15.29)	(12.70)	(12.05)	(12.28)
Market Beta	0.162***	0.109***	0.068**	0.039	0.037
	(4.97)	(4.41)	(2.20)	(1.32)	(1.27)
Observations	23654	23654	23646	23644	23642
Adjusted R^2	0.697	0.737	0.641	0.621	0.620
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes

Table 7: Performance during the 1998 crisis and diversification in 1999-2009.

Notes: This table reports the results of panel regressions relating banks' stock performance in the 1998 Russian crisis to their diversification in subsequent years. These regressions are limited to the 1999 to 2009 period. *High Tail Risk 1998* is a dummy variable that equals 1 if a bank's *Tail Risk* in 1998 is higher than the medium value across all banks during that quarter. *Tail Risk* is the negative of the average excess return during the worst 5% return days over a quarter. All regressions include size-decile fixed effects and the eight bank-level control variables from Table 2, but we do not report the coefficients in order to conserve space. *Diver* is the Entropy Index of seven interest income and nine noninterest income items. In column (1), we estimate the effect of *High Tail Risk 1998* on *Diver* between 1999 and 2009. In column (2), we estimate a first-difference specification. The dependent variable in column (3) is the increase in *Diver* over 1998 to 2000, and we control for all the bank characteristics for 1998. Similarly, the dependent variable in column (4) (column (5)) is the increase in *Diver* over 2000 to 2003 (2003 to 2006), and we control for all the bank characteristics for 2000 (2003). *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster.

	(1)	(2)	(3)	(4)	(5)
	Diver	ΔDiver	ΔDiver_1998-00	ΔDiver_2000-03	ΔDiver_2003-06
High Tail Risk 1998	-0.034	-0.003**	-0.084	0.031	-0.038
	(-1.51)	(-2.28)	(-1.32)	(0.61)	(-1.20)
Constant	1.124***	0.024	0.040	0.682***	0.114*
	(23.72)	(1.59)	(0.28)	(2.81)	(1.75)
Observations	11228	10880	95	196	221
Adjusted R^2	0.521	0.347	-0.041	0.038	0.030
Year-Quarter Fixed Effects	Yes	Yes	No	No	No
Size-Decile Fixed Effects	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes

Table 8: The impact of diversification on return.

Notes: This table shows the estimated coefficients for the following regressions:

$$Return_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Return_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Return_{i,t+1}$ includes the following four return variables. *Qr Return* is the natural logarithm of the buy-and-hold stock excess return over a quarter. *CAPM Abnormal Return* is the natural logarithm of the quarterly residual return based on the CAPM. *FF3 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [1993] model. *FF5 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [2015] model. We control for $Return_{i,t}$ in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects ($\sum \eta_i$) and year-quarter fixed effects ($\sum \gamma_t$). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Qr Return(t+1)	(2) CAPM Abnormal Return(t+1)	(3) FF3 Abnormal Return(t+1)	(4) FF5 Abnormal Return(t+1)
Diversification	0.502**	0.249***	0.243***	0.242***
	(2.10)	(2.90)	(2.88)	(2.97)
Qr Return(t)	-2.078***			
	(-8.85)			
CAPM Abnormal Return(t)		1.812***		
		(9.77)		
FF3 Abnormal Return(t)			1.736***	
			(9.62)	
FF5 Abnormal Return(t)				1.657***
				(9.12)
Log Assets	-6.680***	0.482**	0.560**	0.539**
	(-9.06)	(2.16)	(2.55)	(2.54)
Capital to Assets	0.091	0.612***	0.610***	0.591***
	(0.40)	(6.96)	(6.96)	(6.85)
Operating Profits	1.986***	0.198*	0.165	0.149
	(3.99)	(1.87)	(1.57)	(1.47)
Cost to Income	-0.669	-0.396***	-0.395***	-0.406***
	(-1.31)	(-2.97)	(-3.01)	(-3.15)
Deposits to Liabilities	-0.244	-0.182*	-0.181*	-0.178*
	(-0.96)	(-1.83)	(-1.86)	(-1.86)
Loan Loss Provisions	-3.975***	-0.805***	-0.752***	-0.718***
	(-12.81)	(-7.83)	(-7.44)	(-7.34)
Assets Growth	-0.145	0.132***	0.129***	0.125***
	(-0.94)	(2.75)	(2.78)	(2.88)
Z-Score	-0.130	0.242***	0.225***	0.217***
	(-0.74)	(4.62)	(4.51)	(4.56)
VIX	-6.438***	-1.025***	-0.947***	-0.936***
	(-12.77)	(-7.55)	(-7.36)	(-7.79)
Market Beta	-0.160	-0.180***	-0.126**	-0.119**
	(-0.63)	(-2.97)	(-2.16)	(-2.09)
Observations	23654	23654	23654	23654
Adjusted R^2	0.399	0.420	0.392	0.388
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table 9: The impact of diversification on return four quarters ahead.

Notes: This table shows the estimated coefficients for the following regressions:

$$Qr\ Return_{i,t+1/2/3/4} = \alpha_i + \beta_i Diversification_{i,t} + Qr\ Return_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, *Qr Return* is the natural logarithm of the buy-and-hold stock excess return over a quarter. We control for *Qr Return* at quarter t in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Qr Return (t+1)	(2) Qr Return (t+2)	(3) Qr Return (t+3)	(4) Qr Return (t+4)
Diversification	0.502**	0.755***	0.812***	0.427
	(2.12)	(3.08)	(3.03)	(1.55)
Qr Return (t)	-2.078***	-0.727***	0.129	0.976***
	(-8.95)	(-2.81)	(0.51)	(4.37)
Log Assets	-6.680***	-7.722***	-7.687***	-7.413***
	(-9.17)	(-8.82)	(-8.49)	(-7.93)
Capital to Assets	0.091	0.186	0.080	-0.132
	(0.40)	(0.84)	(0.34)	(-0.56)
Operating Profits	1.986***	-0.172	0.530	-0.420
	(4.04)	(-0.38)	(1.20)	(-0.97)
Cost to Income	-0.669	-1.019*	-0.033	-0.833*
	(-1.33)	(-1.94)	(-0.07)	(-1.75)
Deposits to Liabilities	-0.244	-0.294	-0.275	-0.137
	(-0.97)	(-1.10)	(-0.99)	(-0.50)
Loan Loss Provisions	-3.975***	-1.289***	-1.050***	-1.141***
	(-12.96)	(-4.79)	(-3.74)	(-4.20)
Assets Growth	-0.145	-0.087	-0.199	-0.180
	(-0.95)	(-0.46)	(-0.95)	(-0.93)
Z-Score	-0.130	0.202	0.074	-0.110
	(-0.75)	(1.08)	(0.40)	(-0.67)
VIX	21.907***	3.268*	-1.200***	-3.409***
	(11.48)	(1.66)	(-2.71)	(-6.59)
Market Beta	-0.160	-0.328	-0.009	0.294
	(-0.64)	(-1.24)	(-0.03)	(1.06)
Observations	23654	23394	23130	22856
Adjusted R^2	0.399	0.362	0.360	0.356
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table 10: The impact of related and unrelated diversification on return.

Notes: This table shows the estimated coefficients for the following regressions:

$$Return_{i,t+1} = \alpha_i + \beta_i Related\,Diversification_{i,t} + Unrelated\,Diversification_{i,t} + Return_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Return_{i,t+1}$ includes the following four return variables. *Qr Return* is the natural logarithm of the buy-and-hold stock excess return over a quarter. *CAPM Abnormal Return* is the natural logarithm of the quarterly residual return based on the CAPM. *FF3 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [1993] model. *FF5 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [2015] model. We control for $Return_{i,t}$ in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Qr Return(t+1)	(2) CAPM Abnormal Return(t+1)	(3) FF3 Abnormal Return(t+1)	(4) FF5 Abnormal Return(t+1)
Related Diversification	0.419** (2.30)	0.178*** (2.70)	0.171*** (2.64)	0.169*** (2.67)
Unrelated Diversification	0.076 (0.30)	-0.004 (-0.06)	-0.006 (-0.08)	-0.002 (-0.03)
Qr Return(t)	-2.080*** (-8.85)			
CAPM Abnormal Return(t)		1.813*** (9.76)		
FF3 Abnormal Return(t)			1.737*** (9.62)	
FF5 Abnormal Return(t)				1.658*** (9.12)
Log Assets	-6.680*** (-9.02)	0.482** (2.15)	0.559** (2.54)	0.538** (2.53)
Capital to Assets	0.085 (0.37)	0.608*** (6.93)	0.606*** (6.93)	0.587*** (6.82)
Operating Profits	1.987*** (3.96)	0.193* (1.72)	0.160 (1.44)	0.146 (1.36)
Cost to Income	-0.678 (-1.32)	-0.404*** (-2.90)	-0.403*** (-2.94)	-0.413*** (-3.07)
Deposits to Liabilities	-0.171 (-0.67)	-0.148 (-1.55)	-0.148 (-1.59)	-0.145 (-1.58)
Loan Loss Provisions	-3.976*** (-12.82)	-0.807*** (-7.84)	-0.755*** (-7.44)	-0.721*** (-7.35)
Assets Growth	-0.159 (-1.03)	0.127*** (2.64)	0.124*** (2.67)	0.120*** (2.76)
Z-Score	-0.120 (-0.69)	0.248*** (4.70)	0.231*** (4.59)	0.223*** (4.64)
VIX	-6.474*** (-12.80)	-1.022*** (-7.48)	-0.943*** (-7.27)	-0.933*** (-7.70)
Market Beta	-0.148 (-0.58)	-0.174*** (-2.88)	-0.121** (-2.06)	-0.113** (-2.00)
Observations	23654	23654	23654	23654
Adjusted R^2	0.399	0.420	0.392	0.388
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table 11: The impact of diversification on tail risk (nonlinear).

Notes: This table shows the estimated coefficients for the following regressions:

$$Tail\,Risk_{i,t+1} = \alpha_i + \beta_{1,i}Diversification_{i,t} + \beta_{2,i}Diversification^2_{i,t} + Tail\,Risk_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Tail\,Risk_{i,t+1}$ includes the following five tail risk variables. *Tail Risk 5%* is the negative of the average excess return during the worst 5% return days over a quarter. *Tail Risk 10%* is the negative of the average excess return during the worst 10% return days over a quarter. *CAPM Tail Risk* is the negative of the average CAPM residual return during the worst 5% return days over a quarter. *FF3 Tail Risk* is the negative of the average Fama and French [1993] residual return during the worst 5% return days over a quarter. *FF5 Tail Risk* is the negative of the average Fama and French [2015] residual return during the worst 5% return days over a quarter. We control for $Tail\,Risk_{i,t}$ in all five specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Diversification*^{2} is the squared Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk 5%(t+1)	(2) Tail Risk 10%(t+1)	(3) CAPM Tail Risk(t+1)	(4) FF3 Tail Risk(t+1)	(5) FF5 Tail Risk(t+1)
Diversification	-0.108*** (-3.17)	-0.084*** (-3.23)	-0.127*** (-3.80)	-0.120*** (-3.66)	-0.120*** (-3.74)
Diversification Squared	0.012 (0.93)	0.009 (0.89)	0.017 (1.37)	0.028** (2.24)	0.025** (2.10)
Tail Risk 5%(t)	1.180*** (24.70)				
Tail Risk 10%(t)		1.146*** (30.77)			
CAPM Tail Risk(t)			1.142*** (24.36)		
FF3 Tail Risk(t)				1.139*** (24.60)	
FF5 Tail Risk(t)					1.121*** (24.82)
Log Assets	-0.159 (-1.50)	-0.089 (-1.14)	-0.381*** (-3.55)	-0.471*** (-4.42)	-0.474*** (-4.52)
Capital to Assets	-0.292*** (-9.43)	-0.222*** (-9.63)	-0.328*** (-10.88)	-0.326*** (-10.94)	-0.319*** (-10.93)
Operating Profits	-0.093* (-1.78)	-0.069* (-1.73)	-0.089* (-1.75)	-0.066 (-1.35)	-0.060 (-1.25)
Cost to Income	0.205*** (3.54)	0.159*** (3.56)	0.189*** (3.40)	0.215*** (3.95)	0.228*** (4.25)
Deposits to Liabilities	0.056 (1.48)	0.039 (1.37)	0.059 (1.53)	0.062 (1.62)	0.057 (1.50)
Loan Loss Provisions	0.465*** (14.37)	0.350*** (14.26)	0.420*** (13.13)	0.407*** (12.90)	0.397*** (12.76)
Assets Growth	-0.019 (-0.96)	-0.014 (-0.93)	-0.032* (-1.71)	-0.037** (-2.00)	-0.036** (-2.00)
Z-Score	-0.118*** (-5.07)	-0.090*** (-5.11)	-0.122*** (-5.44)	-0.108*** (-4.91)	-0.106*** (-4.86)
VIX	0.954*** (15.64)	0.738*** (15.18)	0.733*** (12.61)	0.697*** (11.90)	0.691*** (12.13)
Market Beta	0.164*** (5.09)	0.111*** (4.52)	0.071** (2.33)	0.044 (1.48)	0.041 (1.42)
Observations	23654	23654	23646	23644	23642
Adjusted R^2	0.697	0.737	0.642	0.621	0.620
Bank Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes

Table 12: The impact of diversification on return (nonlinear).

Notes: This table shows the estimated coefficients for the following regressions:

$$Return_{i,t+1} = \alpha_i + \beta_{1,i}Diversification_{i,t} + \beta_{2,i}Diversification^2_{i,t} + Return_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Return_{i,t+1}$ includes the following four return variables. *Qr Return* is the natural logarithm of the buy-and-hold stock excess return over a quarter. *CAPM Abnormal Return* is the natural logarithm of the quarterly residual return based on the CAPM. *FF3 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [1993] model. *FF5 Abnormal Return* is the natural logarithm of the quarterly residual return based on the Fama and French [2015] model. We control for $Return_{i,t}$ in all four specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. $Diversification^2$ is the squared Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all independent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Qr Return(t+1)	(2) CAPM Abnormal Return(t+1)	(3) FF3 Abnormal Return(t+1)	(4) FF5 Abnormal Return(t+1)
Diversification	0.425* (1.82)	0.223*** (2.78)	0.215*** (2.72)	0.214*** (2.80)
Diversification Squared	-0.228** (-2.30)	-0.075** (-2.26)	-0.084*** (-2.59)	-0.083*** (-2.66)
Qr Return(t)	-2.086*** (-8.87)			
CAPM Abnormal Return(t)		1.809*** (9.73)		
FF3 Abnormal Return(t)			1.732*** (9.58)	
FF5 Abnormal Return(t)				1.653*** (9.09)
Log Assets	-6.567*** (-8.95)	0.521** (2.31)	0.603*** (2.72)	0.582*** (2.70)
Capital to Assets	0.083 (0.36)	0.610*** (6.97)	0.608*** (6.97)	0.589*** (6.86)
Operating Profits	1.989*** (4.00)	0.198* (1.89)	0.165 (1.60)	0.150 (1.50)
Cost to Income	-0.670 (-1.32)	-0.397*** (-2.99)	-0.396*** (-3.03)	-0.407*** (-3.17)
Deposits to Liabilities	-0.235 (-0.92)	-0.179* (-1.81)	-0.178* (-1.83)	-0.175* (-1.83)
Loan Loss Provisions	-3.970*** (-12.79)	-0.803*** (-7.83)	-0.750*** (-7.43)	-0.717*** (-7.34)
Assets Growth	-0.163 (-1.07)	0.126*** (2.65)	0.122*** (2.67)	0.119*** (2.76)
Z-Score	-0.142 (-0.81)	0.238*** (4.58)	0.221*** (4.45)	0.213*** (4.50)
VIX	-6.414*** (-12.73)	-1.018*** (-7.54)	-0.939*** (-7.34)	-0.928*** (-7.77)
Market Beta	-0.185 (-0.73)	-0.189*** (-3.08)	-0.136** (-2.30)	-0.128** (-2.23)
Observations	23654	23654	23654	23654
Adjusted R^2	0.399	0.420	0.393	0.389
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table 13: The impact of diversification on lending.

Notes: This table shows the estimated coefficients for the following regressions:

$$\Delta log(Total\,Loans)_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $\Delta log(Total\,Loans)_{i,t+1}$ is the change in the logarithm of total loans. We control for four lags of the dependent variable and four lags of the change in federal funds in all four specifications. In columns (1) and (2), we control for year-quarter fixed effects (γ_t) while in columns (3) and (4) we control for four lags of the change in GDP. All regressions include bank fixed effects (η_i). *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the *t*-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Δlog(Total Loans)(t+1)	(2) Δlog(Total Loans)(t+1)	(3) Δlog(Total Loans)(t+1)	(4) Δlog(Total Loans)(t+1)
Diversification	0.329*** (3.91)	0.211*** (2.59)	0.349*** (6.78)	0.133*** (2.60)
Log Assets		-3.805*** (-10.65)		-3.018*** (-12.35)
Capital to Assets		0.634*** (6.93)		0.702*** (7.44)
Operating Profits		-0.185 (-1.19)		-0.540*** (-4.50)
Cost to Income		-0.302* (-1.90)		-0.612*** (-4.79)
Deposits to Liabilities		-0.184* (-1.77)		-0.094 (-1.01)
Loan Loss Provisions		-0.397*** (-9.59)		-0.511*** (-12.89)
Assets Growth		0.210*** (3.35)		0.141** (2.52)
Z-Score		0.639*** (9.72)		0.742*** (12.03)
Observations	62096	62096	62096	62096
Adjusted R^2	0.064	0.086	0.029	0.060
Four Lags of Dependent Variable	Yes	Yes	Yes	Yes
Four Lags of Rate Change	Yes	Yes	Yes	Yes
Four Lags of GDP Change	No	No	Yes	Yes
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	No	No

Table 14: Lending resilience during the 2007-2008 crisis.

Notes: This table reports the results of panel regressions relating banks' pre-crisis entropy to their lending during the 2007-2008 crisis. These regressions are limited to the 2007 to 2008 period. All regressions include year-quarter fixed effects and the eight bank-level control variables from column (1) of Table 2 and *Diversification*, all of which are fixed at their values in the fourth quarter of 2006. In column (1), we estimate the effect of pre-crisis entropy on total loan supply scaled by pre-crisis assets without bank-level controls. In column (2), we estimate the effect of pre-crisis entropy on total loan supply scaled by pre-crisis assets with bank-level controls fixed in the fourth quarter of 2006. In column (3), we estimate the effect of pre-crisis entropy on commericial and industrial loan supply scaled by pre-crisis assets without bank-level controls. In column (4), we estimate the effect of pre-crisis entropy on commericial and industrial loan supply scaled by pre-crisis assets with bank-level controls fixed in the fourth quarter of 2006. In column (5), we estimate the effect of pre-crisis entropy on real estate loan supply scaled by pre-crisis assets without bank-level controls. In column (6), we estimate the effect of pre-crisis entropy on real estate loan supply scaled by pre-crisis assets with bank-level controls fixed in the fourth quarter of 2006. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Loans/PC Assets	(2) Loans/PC Assets	(3) CI/PC Assets	(4) CI/PC Assets	(5) RE/PC Assets	(6) RE/PC Assets
Pre-Crisis Diversification	4.349***	4.931***	2.055***	1.741***	0.801	2.032***
	(7.24)	(8.73)	(8.79)	(7.03)	(1.24)	(3.53)
Log Assets_2006		-2.948***		1.058***		-4.645***
		(-4.71)		(3.24)		(-7.12)
Capital to Assets_2006		-0.062		-0.013		0.491
		(-0.09)		(-0.04)		(0.69)
Operating Profits_2006		6.496***		1.143		3.804*
		(2.77)		(1.31)		(1.83)
Cost to Income_2006		4.891**		0.705		3.038
		(2.20)		(0.81)		(1.52)
Deposits to Liabilities_2006		1.834**		0.848***		1.029
		(2.36)		(2.96)		(1.32)
Loan Loss Provisions_2006		0.058		-0.173		-0.792
		(0.10)		(-0.67)		(-1.33)
Assets Growth_2006		5.925***		0.374		6.721***
		(10.16)		(1.48)		(10.61)
Z-Score_2006		2.193***		-0.143		1.984***
		(3.87)		(-0.53)		(3.36)
Observations	6761	6761	6761	6761	6761	6761
Adjusted R^2	0.094	0.240	0.073	0.098	0.031	0.220
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Table 15: The impact of diversification on accounting performance.

Notes: This table shows the estimated coefficients for the following regressions:

$$Accounting\,Variable_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Accounting\,Variable_{i,t} + Controls_{i,t} + \sum \eta_i + \sum \gamma_t + \epsilon_{i,t}$$

Here, $Accounting\,Variable_{i,t+1}$ includes the following three accounting variables. *ROA* is the ratio of net income to total assets in percentage. *ROE* is the ratio of net income to total equity in percentage. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. We control for $Accounting\,Variable_{i,t}$ in all three specifications. *Diversification* is the Entropy Index of seven interest income and nine noninterest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets in percentage. *VIX* is the average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. *Market Beta* is the sum of the two beta coefficients after we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. We drop missing values and winsorize all ratio-based control variables at the top and bottom 1%. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) ROA(t+1)	(2) ROE(t+1)	(3) Z-Score(t+1)
Diversification	0.011***	0.151***	0.006***
	(4.54)	(4.60)	(4.13)
ROA(t)	0.040***		
	(9.02)		
ROE(t)		0.784***	
		(10.60)	
Z-Score			0.398***
			(239.03)
Log Assets	-0.041***	-0.669***	-0.013**
	(-4.60)	(-4.86)	(-2.40)
Capital to Assets	0.016***	0.009	0.002
	(6.50)	(0.23)	(1.16)
Cost to Income	-0.046***	-0.523***	-0.016***
	(-14.75)	(-10.75)	(-13.63)
Deposits to Liabilities	-0.003	-0.088**	0.002
	(-1.06)	(-2.35)	(1.10)
Loan Loss Provisions	-0.035***	-0.417***	-0.015***
	(-11.25)	(-9.12)	(-14.44)
Assets Growth	0.010***	0.233***	0.004***
	(6.22)	(8.54)	(3.67)
Observations	68938	68938	68938
Adjusted R^2	0.335	0.317	0.840
Bank Fixed Effects	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes

Appendix
Bank Diversification and Tail Risk

A Definitions and construction of variables

We collect balance sheet and income statement data for banks from the Consolidated Financial Statements for Holding Companies (henceforth FR Y-9C) required to be filed by all U.S. bank holding companies. Definitions for the variables are available at https://www.federalreserve.gov/apps/reportingforms/Report/Index/FR_Y-9C. Banks with total book value of assets above $500 million file this report quarterly. We restrict our sample to banks which file the Call Reports quarterly and report a positive book value of assets. Between June 1986 and December 2020, this yields 182,038 observations. The actual number of observations in our analysis is less for several reasons. First, we eliminate data for all banks whose total capital is missing, zero, or negative. This yields a dataset with just 132,937 observations. Second, we eliminate observations if any of the control variables is missing. This leaves me with 70,683 bank-quarter observations. Third, after merging with variables constructed from CRSP, we require that the banks in our sample have at least three consecutive years (12 quarters) of data available. This leaves us with 23,809 bank-quarter observations between September 1996 and December 2020.

The data present a number of challenges in terms of creating a consistent time-series. Due to changing reporting requirements, some of the data items in the FR Y-9C used for the construction of key variables in our analysis are not comparable across quarters. The Chicago Federal Reserve Bank provides instructions for the construction of consistent time-series for the data in the FR Y-9C. These instructions are available at http://www.chicagofed.org/webpages/banking/financial_institution_reports/bhc_data.cfm and are summarized in table A1.

Once we define time-series for individual banks, we also compute data for all U.S. banks (i.e. the aggregate U.S. bank sector) to report summary statistics in section 1. To compute the time-series for all U.S. banks, we start with data for individual banks. We filter the top and bottom 1-percentile of banks based on the quarterly growth rate in total book value of assets. This filter removes observations for those bank-quarters in which banks are involved in significant mergers. For aggregation, we require that in each quarter, banks included in our sample have FR Y-9C data available for at least 12 previous quarters (3 years). We also require that for each quarter FR Y-9C data for a particular bank is available for the previous and current quarters. This requirement ensures that the time-series of core and non-core income are not affected by entry or exit of banks. This requirement also means that the actual number of banks used in any quarter to compute the time-series for all U.S. banks varies over time.

Table A2 presents the definition for all key variables used in the paper.

Table A1: Computation of consistent time-series

Notes: This table provides details regarding the construction of key variables used in our empirical analysis. We collect balance sheet and income statement data for banks from the Consolidated Financial Statements for Holding Companies (FR Y-9C) required to be filed by all FDIC-insured bank holding companies in the U.S. The first column lists the mnemonic used to identify each variable in our empirical analysis. Column titled 'Name' provides a brief description. Column titled 'Call Report Data Item' lists the exact Federal Reserve item codes used to construct each variable. Finally, column titled 'Adjustment Rules' details adjustments made to the definition of each variable to render them time-consistent.

Mnemonic	Name	Call Report Data Items	Adjustment Rules
Total assets	Book value of assets	BHCK2170	
Capital	Tier 1 Capital	BHCA8274	After 2014 use BHCA8274. Between 1996 and 2014, use BHCK8274. Before 1996 use the sum of BHCK3210, BHCK3247, BHCK3455, and BHCK3456.
	TIER 1 CAPITAL	BHCK8274	
	Total Equity Capital	BHCK3210	
	Undivided profits and capital reserves	BHCK3247	
	Unsecured long-term debt	BHCK3247	
	Mandatory convertible securities	BHCK3247	
Deposits	Domestic noninterest bearing deposits	BHDM6631	Sum of items.
	Domestic interest bearing deposits	BHDM6636	
	Foregin noninterest bearing deposits	BHFN6631	
	Foregin interest bearing deposits	BHFN6636	
Noninterest income	Total noninterest income	BHCK4079	
	Total interest income	BHCK4107	
	Net income	BHCK4340	
	Total liabilities	BHCK2948	
	Preferred stocks	BHCK3283	
	Net interest income	BHCK4074	
	Total interest expense	BHCK4073	
	Noninterest expense	BHCK4093	
	Loan loss provisions	BHCK4230	
	Income from fiduciary activities	BHCK4070	
	Service charges on deposits accounts in domestic offices	BHCK4483	
	Trading revenue	BHCKA220	
	Fees and commissions from securities brokerage	BHCKC886	
	Investment banking, advisory, and underwriting fees and commisions	BHCKC888	
	Fees and commissions from annuity sales	BHCKC887	
	Underwriting income from insurance and reinsurance activities	BHCKC386	
	Income from other insurance activities	BHCKC387	
	Venture capital revenue	BHCKB491	
	Net servicing fees	BHCKB492	
	Net securitization income	BHCKB493	
	Net gains (losses) on sales of loans and lease	BHCK8560	
	Net gains (losses) on sales of other real estate owned	BHCK8561	
	Net gains (losses) on sales of other assets	BHCKB496	
	Other noninterest income	BHCKB497	
Interest income	Loans secured by 1-4 family residential properties (domestic)	BHCK4435	
	All other loans secured by real estate (domestic)	BHCK4436	
	All other loans (domestic)	BHCKF821	
	In foreign offices, Edge and Agreement subsidiaries, and IBFs	BHCK4059	
	Income from lease financing receivables	BHCK4065	
	Interest income on balances due from depository institutions	BHCK4115	
	U.S. Treasury securities and U.S. government agency obligations (excluding mortgage-backed securities)	BHCKB488	
	Mortgage-backed securities	BHCKB489	
	All other securities	BHCK4060	
	Interest income from trading assets	BHCK4069	
	Interest income on federal funds sold and securities purchased under agreements to resell	BHCK4020	
	Other interest income	BHCK4518	

Table A2: Variables and data sources.

Notes: This table shows the definition for all key variables used in the paper and the data sources used to collect the data for the construction of these variables.

Variable	Description and sources
Diversification	Entropy Index of seven ninterest income and nine interest income items. FR Y-9C
Related Diversification	Entropy Index of seven interest income items (core business). FR Y-9C
Unrelated Diversification	Entropy Index of nine noninterest income items (noncore business). FR Y-9C
Tail Risk 5%	The negative of the average excess return during the 5% worst returns days over the quarter. CRSP
Tail Risk 10 %	The negative of the average excess return during the 10% worst returns days over the quarter. CRSP
CAPM Tail Risk	The negative of the average CAPM residual return during the 5% worst returns days over the quarter. CRSP
FF3 Tail Risk	The negative of the average Fama and French [1993] residual return during the worst 5% return days over the quarter. CRSP
FF5 Tail Risk	The negative of the average Fama and French [2015] residual return during the worst 5% return days over the quarter. CRSP
Qr Return	The natural logarithm of the buy-and-hold stock excess return over the quarter. CRSP
CAPM Abnormal Return	The natural logarithm of the quarterly residual return based on the CAPM over the quarter. CRSP
FF3 Abnormal Return	The natural logarithm of the quarterly residual return based on the Fama and French [1993] model over the quarter. CRSP
FF5 Abnormal Return	The natural logarithm of the quarterly residual return based on the Fama and French [2015] model over the quarter. CRSP
log(Total Loans)	The natural logarithm of the change in total loan supply. FR Y-9C
Change in Federal Funds Rate	The change in effective federal funds rate. FRED
Change in GDP	The natural logarithm of the change in nominal GDP. FRED
ROA	Ratio of net income to total assets.
ROE	Ratio of net income to equity capital. FR Y-9C
Log Assets	Natural logarithm of total assets. FR Y-9C
Capital to Assets	Ratio of equity capital to total assets. FR Y-9C
Operating Profits	Ratio of the sum of noninterest and interest income to total assets. FR Y-9C
Cost to Income	Ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income. FR Y-9C
Deposits to Liabilities	Ratio of total deposits to total liabilities. FR Y-9C
Loan Loss Provisions	Ratio of loan loss provisions to total loans. FR Y-9C
Assets Growth	Three-year growth in total assets. FR Y-9C
Z-Score	The common logarithm of Z-score which is the sum of return on assets and capital to assets ratio divided by the standard deviation of return on assets over a rolling window of 12 quarters. FR Y-9C
VIX	The average daily closing price of the CBOE S&P 500 Volatility Index within a quarter. CBOE Indexes
Market Beta	Regress daily excess returns on excess market returns and lagged excess market returns in a rolling window of eight quarters. Market beta is equal to the sum of the two beta coefficients. CRSP, Kenneth French's Data Library

Table A3: The impact of diversification on tail risk (10%) four quarters ahead.

Notes: This table shows the estimated coefficients for the following regressions:

$$\text{Tail Risk } 10\%_{i,t+1/2/3/4} = \alpha_i + \beta_i Diversification_{i,t} + \text{Tail Risk } 10\%_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, *Tail Risk 10%* is the negative of the average stock return during the 10% worst returns days over the quarter. We control for *Tail Risk 10%* at time *t*. *Diversification* is the Entropy Index of nine noninterest income and seven interest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE SP 500 Volatility Index within a quarter. For *Market Beta*, we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. Market beta is equal to the sum of the two beta coefficients. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all dependent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. All regressions include bank fixed effects (η_i) and year-quarter fixed effects (γ_t). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk 10%(t+1)	(2) Tail Risk 10%(t+2)	(3) Tail Risk 10%(t+3)	(4) Tail Risk 10%(t+4)
Entropy	-0.080***	-0.119***	-0.138***	-0.138***
	(-3.00)	(-3.69)	(-3.91)	(-3.47)
Tail Risk 10%(t)	1.117***	0.840***	0.667***	0.487***
	(31.44)	(20.94)	(15.00)	(11.56)
Log Assets	-0.039	0.200**	0.366***	0.545***
	(-0.52)	(2.08)	(3.17)	(4.02)
Capital to Assets	-0.202***	-0.249***	-0.253***	-0.252***
	(-9.06)	(-9.29)	(-8.10)	(-7.01)
Operating Profits	-0.066*	0.012	-0.059	-0.071
	(-1.76)	(0.27)	(-1.13)	(-1.37)
Cost to Income	0.150***	0.197***	0.112**	0.107**
	(3.61)	(4.21)	(2.16)	(2.06)
Deposits to Liabilities	0.037	0.057	0.061	0.078
	(1.35)	(1.58)	(1.38)	(1.55)
Loan Loss Provisions	0.320***	0.305***	0.295***	0.332***
	(14.51)	(10.93)	(9.56)	(10.38)
Assets Growth	-0.005	-0.018	-0.012	-0.020
	(-0.34)	(-0.92)	(-0.47)	(-0.75)
Z-Score	-0.088***	-0.106***	-0.107***	-0.083***
	(-5.09)	(-4.80)	(-3.92)	(-2.88)
VIX	0.314**	-0.372*	0.045	0.146***
	(2.01)	(-1.89)	(0.84)	(2.70)
Market Beta	0.103***	0.108***	0.066*	0.010
	(4.26)	(3.41)	(1.76)	(0.23)
Observations	23795	23545	23289	23021
Adjusted R^2	0.735	0.688	0.662	0.644
Bank Fixed Effects	Yes	Yes	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes	Yes

Table A4: The impact of diversification on tail risk (using different fixed effects).

Notes: This table shows the estimated coefficients for the following regressions:

$$Tail\,Risk_{i,t+1} = \alpha_i + \beta_i Diversification_{i,t} + Tail\,Risk_{i,t} + Controls_{i,t} + \epsilon_{i,t} + (\eta_i + \gamma_t)$$

Here, *Tail Risk* is the negative of the average stock return during the 5% worst returns days over the quarter. We control for lagged *Tail Risk*. *Entropy* is the Entropy Index of nine noninterest income and seven interest income items. *Log Assets* is the natural logarithm of total assets. *Capital to Assets* is the ratio of equity capital to total assets in percentage. *Operating Profits* is the ratio of the sum of noninterest and interest income to total assets in percentage. *Cost to Income* is the ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income in percentage. *Deposits to Liabilities* is the ratio of total deposits to total liabilities in percentage. *Loan Loss Provisions* is the ratio of loan loss provisions to total loans in percentage. *Assets Growth* is the three-year growth in total assets. *Z-Score* is equal to the common logarithm of Z-score, where the Z-score is the ratio of the sum of return on assets and capital to assets ratio to the standard deviation of return on assets over a rolling window of 12 quarters. *VIX* is the average daily closing price of the CBOE SP 500 Volatility Index within a quarter. For *Market Beta*, we regress daily excess returns on market and lagged market factor in a rolling window of eight quarters. Market beta is equal to the sum of the two beta coefficients. We drop missing values, winsorize all ratio-based control variables at the top and bottom 1%, and standardize all dependent variables. The numbers in parenthesis are the t-statistics. Statistical significance is indicated by *, **, and *** at the 10%, 5% and 1% levels respectively using cluster-robust standard errors with each bank as a cluster. Column (1) does not include bank fixed effects (η_i) and year-quarter fixed effects (γ_i). Column (2) does not include bank fixed effects (η_i) but includes year-quarter fixed effects (γ_i). Column (3) includes bank fixed effects (η_i) but does not include year-quarter fixed effects (γ_i). Column (4) includes both bank fixed effects (η_i) and year-quarter fixed effects (γ_i). The panel data is on the bank-quarter level between 1996Q3 and 2020Q4.

	(1) Tail Risk(t+1)	(2) Tail Risk(t+1)	(3) Tail Risk(t+1)	(4) Tail Risk(t+1)
Diversification	-0.223***	-0.085***	-0.360***	-0.102***
	(-12.03)	(-4.25)	(-12.92)	(-2.95)
Tail Risk	1.724***	1.484***	1.447***	1.156***
	(39.47)	(31.85)	(31.27)	(24.48)
Log Assets	-0.122***	-0.220***	0.774***	-0.091
	(-4.35)	(-7.58)	(8.93)	(-0.88)
Capital to Assets	-0.127***	-0.116***	-0.247***	-0.267***
	(-6.89)	(-5.60)	(-8.09)	(-8.81)
Operating Profits	0.001	-0.048	0.267***	-0.088*
	(0.03)	(-1.37)	(4.73)	(-1.76)
Cost to Income	0.301***	0.192***	0.595***	0.196***
	(8.70)	(5.11)	(9.79)	(3.60)
Deposits to Liabilities	0.003	0.026	-0.189***	0.053
	(0.14)	(1.13)	(-4.04)	(1.45)
Loan Loss Provisions	0.426***	0.425***	0.412***	0.427***
	(14.22)	(15.34)	(12.84)	(14.39)
Assets Growth	-0.002	0.036**	-0.053***	-0.008
	(-0.13)	(2.45)	(-2.70)	(-0.43)
Z-Score	-0.085***	-0.151***	-0.069**	-0.117***
	(-4.26)	(-7.45)	(-2.57)	(-4.99)
VIX	0.105***	0.282**	0.230***	0.955***
	(3.88)	(1.99)	(7.52)	(15.98)
Market Beta	0.164***	0.140***	0.125***	0.154***
	(6.65)	(5.39)	(3.85)	(4.81)
Observations	23795	23795	23795	23795
Adjusted R^2	0.508	0.715	0.459	0.695
Bank Fixed Effects	No	No	Yes	Yes
Year-Quarter Fixed Effects	No	Yes	No	Yes