Do "MEASURES" of Bank Diversification Measure Up? *

Priyank Gandhi
Rutgers Business School

Darius Palia
Rutgers Business School

Jasper Pan
Rutgers Business School

This version: September 10, 2024

Abstract

We analyze the effectiveness of several widely-used measures of bank business segment diversification in capturing the 'diversification effect', i.e., the ability of the measure to explain variation in idiosyncratic risk over time and across banks. Portfolio theory suggests that bank business segment diversification is negatively correlated with idiosyncratic risk (especially if income from these segments is imperfectly or negatively correlated with each other). We find that several commonly used measures of bank business segment diversification are either poorly or positively correlated with idiosyncratic risk, suggesting that they are inaccurate or misleading indicators of bank business segment diversification. We instead propose an 'Entropy' measure that accounts for both the number of businesses segments that a bank operates in as well as the proportion of banks' incomes from the business segments. In horse-races Entropy performs significantly better in capturing the diversification effect and thus measuring the degree of bank diversification.

JEL Codes: G20, G21, G24, G34.

Keywords: Bank diversification, Noninterest income, Entropy measure of diversification.

*Gandhi: (Corresponding author) 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

Traditionally banks have engaged in just two distinct activities – deposit-taking and lending. Modern banks, however, have diversified into a myriad set of business segments including trading, brokerage, investment banking, market-making, advisory, underwriting, insurance, and venture capital. Income from these diversified segments has increased significantly and now accounts for the majority of income of all U.S. banks.[1] While many academic papers explore how bank diversification into multiple business segments impacts its valuation, risk-taking, and equity returns, this literature has paid surprisingly little attention to the effectiveness of the measures of bank business segment diversification itself – a starting point for all these studies. None of the papers in this literature ask what is a reasonable measure of bank business segment diversification and how well do these existing popular measures fare in terms of capturing the extent of bank business segment diversification.[2] Our paper is the first to systematically analyze the efficacy of various bank diversification measures and make a practical recommendation for regulators and researchers in this area.

We evaluate the performance of various measures of bank diversification by simply examining how well they do in capturing the 'diversification effect', i.e., the ability of the measure to explain variation in idiosyncratic risk over time and across banks. Portfolio theory suggests that bank diversification is negatively correlated with measures of idiosyncratic risk, especially if such diversification produces income streams imperfectly or negatively correlated with each other. Compared to a noisy or inaccurate measure of bank diversification, a true or accurate measure of bank diversification should be better at capturing the diversification effect, i.e., the true measure must be negatively correlated with the bank's idiosyncratic or firm-specific risk. This basic insight from modern portfolio theory forms the basis for nearly all our empirical tests in the paper.

Existing measures of bank diversification can be inaccurate or misleading indicators of a bank's

[1]For the aggregate U.S. bank sector, noninterest income (i.e., income from activities other than deposit-taking and lending) accounted for only 18.11% of net income in 1988 but was substantially higher at nearly 54% of net income by 2016.

[2]The banking literature focuses on both geographical and business segment diversification. Our paper, however, focuses only on bank business segment diversification. Henceforth, we use bank diversification to refer to bank business segment diversification.

true diversification for one simple reason – all of the existing measures used in the literature are calculated using just two items on banks' income statements: interest and noninterest incomes and expenses. They do not account for the fact that bank interest and noninterest incomes are in turn derived from a variety of business segments. For instance, bank interest incomes and expenses can accrue from loans, deposits, trading in securities, participation in the Federal Funds market interbank markets. Similarly, bank noninterest income can also stem from a variety of business segments such as trading, insurance, securities underwriting, venture capital, etc.

A simple example illustrates why this can matter. Consider two banks 'A' and 'B': Bank A earns interest income from loans (\$50) and from trading debt securities (\$50). It also earns noninterest income from insurance (\$100) and venture capital activities (\$100). Bank B earns interest income from loans (\$100) and noninterest income from insurance (\$200). Since both banks earn \$100 in interest and \$200 in noninterest income, existing measures of bank diversification would deem these banks to be equally diversified, despite the fact that bank A operates across twice the number of business segments as compared to bank B. While the ratio of interest to noninterest income is the same for both banks, Bank A could have lower idiosyncratic risk, especially if incomes from loans, trading debt securities, insurance, and venture capital are imperfectly correlated. Thus, in empirical tests, existing diversification measures may perform poorly as compared to measures that use more granular data when attempting to capture the diversification effect.

We consider seven different measures of bank diversification, six of which are widely used in the literature: These are (i) one minus the Herfindahl-Hirschman index of total noninterest and total interest income (the Hhindex measure), (ii) one minus the absolute value of the difference between net interest and total noninterest income divided by the sum of net interest and total noninterest income (the Absdiff measure), (iii) the ratio of *net* noninterest income to the sum of *net* noninterest and interest income (the R-netnet measure), (iv) the ratio of *total* noninterest income to the sum of *total* noninterest and *net* interest income (the R-totnet measure), (v) the ratio of *total* noninterest income to *total* noninterest and interest income (the R-tottot measure), and (vi) finally, the *simple* ratio of total noninterest to interest income (the R-simple measure). These measures have been used widely in studies such as Stiroh (2004), Stiroh (2006), Stiroh and Rumble (2006), Baele, De Jonghe, and Vander Vennet (2007a), Laeven and Levine (2007), Lepetit, Nys, Rous, and Tarazi (2008a), Demirguc-Kunt and Huizinga (2010), Guerry and Wallmeier (2017), and Saunders, Schmid, and Walter (2020a), among many others.

Our final measure of bank diversification, the one we propose as the most informative measure based on the results of our empirical exercise, is the Entropy (Entropy) measure of bank diversification. It is computed as the weighted sum of income that a bank derives from various business segments, where the weights are the logarithm of the inverse of the income that the bank derives from that segment. This measure was first introduced by industrial economists (see, for example, Jacquemin and Berry (1979)) but has only recently been adopted in academic papers in finance to measure firm diversification (Khanna and Palepu (2000)). Entropy is computed using data for income that a bank derives from sixteen different categories of business segments – seven for interest income and nine for noninterest income – that are the most granular data for bank income that one can get from the publicly-available, quarterly call reports required to be filed by all bank holding companies in the U.S. The Entropy measures offers a clear conceptual advantage as it not only accounts for the number of distinct business segments in which a bank operates but also considers the distribution of a bank's total income across these business segments. For our hypothetical banks 'A' and 'B' above Entropy would equal 1.33 and 0.64, respectively, indicating (correctly) that Bank 'A' is more diversified than Bank 'B'.

We begin by documenting that at the aggregate bank sector level the six widely used measures of bank diversification mostly exhibit low correlation with each other, despite the fact that each of these measures claims to accurately assess the degree of diversification by banks into various business segments. These low correlations are all the more surprising given that the only difference among many of these measures is whether they use net or total interest and noninterest income for banks to measure bank diversification. The Entropy measure of bank diversification also exhibits very low correlation with all existing measures of bank diversification. However, this may be due to the fact that computation of the Entropy measure relies on much more granular and detailed data than the computation of these other measures. The correlation of Entropy with existing measures ranges from a minimum of -0.02 (with R-simple) to a maximum of 0.41 (with R-netnet). Overall, the low correlations among various measures of bank diversification provide yet another rationale for our study.

Time series plots for Entropy (for the aggregate U.S. bank sector) show that variation in Entropy coincides with the passage of major legislations related to banking, providing an important validation for this measure. For instance, Entropy increases significantly and remains at elevated levels for all banks post-1999 – when the Gramm-Leach-Bliley Act repealed the restrictions placed on banking activities by the Glass-Steagall Act of 1933. No other measure of diversification shows a significant increase post-1999. Similarly, Entropy drops significantly and remains at low levels for all banks post 2007-2009, and especially after the passage of the Dodd-Frank Act of 2010. While all other measures also fall during 2007-2009, indicating perhaps that bank noninterest income fell during the crisis, they soon revert to their pre-crisis levels. Thus, existing measures of bank diversification suggest that commercial banks in the U.S. were as diversified after Dodd-Frank as they were before, which perhaps may not actually be the case.

Next, we turn the bank-level analysis and estimate all seven measures of bank diversification for each bank for each quarter in our sample. We then systematically investigate the link between all measures of bank diversification and the diversification effect, i.e., the ability of the measure to explain variation in bank idiosyncratic risk over time and across banks. We do this by running predictive (panel) regressions and checking whether measures of bank diversification predict bank idiosyncratic risk one quarter ahead. We expect that an accurate measure of bank diversification should predict lower (i.e., be negatively correlated with) idiosyncratic risk. In all our tests idiosyncratic risk is measured using the idiosyncratic volatility of bank stock returns, i.e., by the standard deviation of residuals by regressing daily bank-level stock returns on the 3-factor Fama and French (1993) model.

In all empirical specifications, Entropy emerges as the strongest (negative) predictor of idiosyncratic volatility. In univariate tests, a one standard deviation increase in Entropy at the bank-level implies that bank idiosyncratic volatility next quarter will be lower by nearly 0.18%. Given that the average quarterly idiosyncratic volatility for the banks in our sample if 1.91%, this implies that higher Entropy is associated with a nearly 10% lower idiosyncratic volatility over the next quarter as compared to the sample mean. Controlling for bank-level characteristics, a one standard deviation increase in Entropy is now associated with nearly 0.08% reduction in bank idiosyncratic volatility over the next quarter (i.e., a nearly 5% reduction as compared to the sample mean).

Compared to Entropy, the ability of other measures of bank diversification to capture the diversification and predict one quarter ahead idiosyncratic volatility is either ambiguous or weak at best. After controlling for bank-level characteristics, none of the other measures predict one quarter ahead idiosyncratic volatility. In fact, an increase in some commonly used diversification measures (such as the R-netnet measure) is associated with an increase rather than a decrease in one-quarter ahead idiosyncratic volatility.

These results survive a battery of robustness tests and changes in the empirical specification. In horse races, i.e., when including multiple measures of bank diversification in the same regression specification, Entropy is the only measure that consistently (and negatively) predicts one quarter ahead idiosyncratic volatility across all specifications. The relation between Entropy and one quarter ahead idiosyncratic volatility remains statistically significant in all sub-samples, and during normal times, as well as during periods of recessions and financial crisis. Entropy emerges as the strongest predictor for bank idiosyncratic volatility regardless of the factor model used to estimate such volatility (CAPM or Fama-French five factor models).

Also, Entropy emerges as the best predictor when we use alternate market-based measure of bank diversification such as the R^2 from a regression of bank stock returns on systematic asset pricing factors (Demsetz and Strahan (1997)) as the dependent variable. Additionally, Entropy is not only the best predictor of idiosyncratic volatility just 1-quarter ahead but remains so up to 4-quarters ahead. None of the other measures of bank diversification used in the literature have any statistically significant ability to predict idiosyncratic volatility more than 1-quarter ahead.

We also estimate to what extent a market participant could have predicted one quarter ahead idiosyncratic volatility of banks in real time, using the data available to that point in time using the seven different measures of bank diversification. That is, we conduct an out-of-sample predictability test. We estimate the ability of a measure to predict out-of-sample bank idiosyncratic volatility by computing the root mean squared error defined using the actual (i.e., realized) and predicted values of bank idiosyncratic volatility. Our results indicate that when forecasting one quarter ahead bank idiosyncratic volatility out-of-sample, the Entropy measure generally outperforms all other measures of bank diversification.

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 bank valuations. 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 segments and activities. Important papers in this area include Demsetz and Strahan (1997), Stiroh (2004), Stiroh and Rumble (2006), Baele, De Jonghe, and Vander Vennet (2007a), Laeven and Levine (2007), and Saunders, Schmid, and Walter (2020b), and many others. Our paper is different from these papers as it focuses on systematically analyzing which diversification measure for banks used in the literature is the most effective. It is also the first to construct a detailed measure of bank diversification using the most granular data for the 16 different business segments available from the quarterly call reports required to be filed by all banks in the U.S. Thus, our study can help reconcile some of the divergent results in the literature regarding bank diversification – such as Demirguc-Kunt and Huizinga (2010) and Saunders, Schmid, and Walter (2020b) who find bank diversification leads to higher insolvency risk and Lepetit, Nys, Rous, and

Tarazi (2008b) who show that diversified banks have lower Z-scores.

Our paper is closest to Demsetz and Strahan (1997), who compute a market-based measure of bank diversification (the R^2 from a regression of bank stock returns on systematic asset pricing factors) and relate it to bank size and risk-taking. In this paper, we use a similar approach to ask which measure of bank diversification derived from balance sheet data is the most effective at capturing the market-based diversification effect. The advantage of identifying the best measure of bank diversification derived from balance sheet data is that once identified, it can be computed for both publicly-listed as well as private banks.

Our analysis only focuses on identifying which of the popular measure of bank diversification used in the extant literature are best at capturing the diversification effect. Thus, our study has nothing to say about why banks choose to diversify into a wide range of business segments, and why such diversification varies over time and across banks. The degree of business segment diversification that a particular bank chooses to undertake is of course an endogenous choice, but is outside the scope of our analysis.

The rest of the paper is organized as follows: Section 1 discusses our research design. In section 2 we describe our data sources and the methodology used to compute key dependent and independent variables. Section 3 presents our key empirical results and analyzes the efficacy of various bank diversification measures in capturing the diversification effect. Finally, section 4 summarizes and concludes.

1 Research design

We begin by establishing an empirical benchmark to evaluate which of the measures of diversification proposed in the literature is the 'best' measure. We do so by relying on three straightforward economic insights. The first economic insight comes from the standard leverage and capital structure invariance effect of Modigliani and Miller (1958) which allows us to relate the return on real assets for any bank to the weighted average of the return on financial assets, the weights in all cases being determined by the relative market value of each of the financial assets.

To illustrate, consider a bank i, with total assets A_i, funded by total debt of D_i and total equity of E_i. In this example, i can stand for an individual bank or the entire aggregate banking sector. Denoting the bank's return on equity and debt by $R_{i,E}$ and $R_{i,D}$, respectively, we can compute the return on bank's assets as:

$$R_{i,A} = \frac{E_i}{A_i} R_{i,E} + \frac{D_i}{A_i} R_{i,D} \tag{1}$$

Hanson, Shleifer, Stein, and Vishny (2015) show that the average bank finances nearly 80% of its assets with deposits. That is, deposits raised from customers comprises nearly all of debt financing for a typical bank. Typically, the volatility of debt is much smaller than the volatility of equity, and this is likely to be even more true for deposit financing. Further, if we assume, as is reasonable, that the correlation between debt returns (i.e., deposit rates) and equity returns is small, we can use equation (1) to relate the asset variance to the equity variance of bank i as:

$$\begin{aligned} \sigma^2_{i,A} &= \frac{E_i^2}{A_i^2} \sigma^2_{i,E} \\ \sigma^2_{i,E} &= \frac{A_i^2}{E_i^2} \sigma^2_{i,A} \end{aligned} \tag{2}$$

The above description is fairly simplistic, but it captures our core idea very well. More sophisticated models provide a more accurate description of this relationship. For e.g., in the Merton (1976) model, equity variance is related to asset variance in a non-linear manner. Nagel and Purnanandam (2020) show that the application of the Merton (1976) model can be problematic for banks given the special nature of bank assets. However, in all classes of models, the broad relation between equity variance and asset variance remains positive.

Our second economic insight comes from the arbitrage pricing theory of Roll and Ross (1984) which suggests that returns on any class of assets can be expressed as a linear combination of factors, i.e., a linear factor model. If bank equity returns follow a factor structure or a factor model, then bank equity returns and bank equity return variance can be further decomposed as in the equation below. In this equation, $\beta_{i,F}$ captures the sensitivity of stock returns for bank i to the selected factor (F), σ_F^2 is the variance of factor F, and $sigma_{i,\epsilon}^2$ represents idiosyncratic variance:

$$\beta_{i,F}^2 \sigma_F^2 + \sigma_{i,\epsilon}^2 = \frac{A_i^2}{E_i^2} \sigma_{i,A}^2 \tag{3}$$

We use a one-factor model for both equity returns and equity variance in equation (3) to show only a parsimonious representation. Advanced factor models such as, the five-factor Fama and French (2015) model or the q-factor model of Hou, Xue, and Zhang (2015), that relate equity returns to multiple factors could easily be used, and will provide a more accurate explanation of the time-series and cross-sectional variation in equity returns and variance.

Our final economic insight comes from standard portfolio theory that allows us to decompose the asset variance of a bank that invests in a portfolio of multiple business segments using both the asset variance of each segment as well as its correlation structure with all other business segments. In other words, consider a bank that distributes its total assets A among N business segments. Let the share of bank assets invested in each segment be given by x_j with $j = 1, ..., N$. Then, we can decompose the asset variance of the bank as:

$$\begin{aligned} \sigma_A^2 &= \sum_j x_j^2 \sigma_j^2 + \sum_j \sum_k x_j x_k \sigma_{j,k} \qquad (4) \\ \sigma_\epsilon^2 &= \frac{A^2}{E^2} \left(\sum_j x_j^2 \sigma_j^2 + \sum_j \sum_k x_j x_k \sigma_{j,k} \right) - \beta_F^2 \sigma_F^2 \end{aligned}$$

Equation (4) relates bank idiosyncratic variance directly to the extent of bank balance sheet diversification across multiple business segments. That is, controlling for bank leverage, factor volatility, and factor exposures, more diversified banks should have lower idiosyncratic variance. This relation, which we refer to as the diversification effect throughout the rest of the paper, forms the basis for all our empirical tests. In our empirical section, we compute different measures of diversification proposed by the literature to measure bank asset diversification and test how well they correlate with or predict equity idiosyncratic variance or volatility.

Note that Equation (4) suggests that the extent of diversification be measured using the market value of dollars that a bank invests in different assets i.e., the market value of assets the bank has devoted to multiple business segments. For banks, data on market value of assets devoted to multiple business segments is not readily available. For a typical bank in our sample, commercial, real-estate, and personal loans account for nearly 90% of all assets and most of these are recorded at historical book values, with no adjustment for current market value of these assets. The market value of bank loans is also extremely hard to compute given data accessible to researchers (for e.g., see Gorton and Pennacchi (1995)). Further, book values of bank asset can also be very noisy due to the accounting treatment of assets such as goodwill and investment in subsidiaries as well as variation over time in rules proposed by the Federal Reserve in how to classify certain bank assets.[3]

For all of the reasons listed above, we follow the extant banking literature and use data for income that a bank derives from multiple business segments (rather than book value of assets) to measure the extent of its diversification across multiple business segments. Measures of diversification based on income depend on the flow of earnings that a bank generates from multiple business segments, and in our view are better proxies for the market value of assets that a bank devotes to multiple business segments than historical book value of assets.

2 Data and summary statistics

In this section, we identify the set of banks used in our analysis, describes various measure of bank diversification used in the literature, and presents summary statistics for the cross-section as well as the aggregate U.S. bank sector. We also describe data sources and present summary statistics for all our dependent, explanatory, and control variables.

2.1 Sample selection

We collect balance sheet data from the 'Report for Condition and Income' (henceforth, the Call Report) required to be filed by all FDIC-insured bank holding companies (henceforth, banks). In

[3] See for e.g., Beatty and Liao (2014) for a comprehensive analysis of how accounting rules and regulatory regimes can impact the book value of bank assets and liabilities as well as bank behavior.

the U.S., banks with total book value above $500 million file this report quarterly whereas other banks file this report semi-annually. We restrict our sample to banks which file the Call Report quarterly and are publicly listed (i.e., data for their stock returns and market capitalization is available). This restriction implies that our sample includes 560 unique banks. Our sample includes the largest banks in the U.S. that collectively account for more than 90% of total U.S. banking sector assets at any point in time. Focusing on banks with total book value above $500 million that are publicly listed is that it allows us to analyze data at the highest frequency possible. Call Report data with details for income that a bank derives from different categories starts in September 1996, and this determines the start date of our sample.[4]

A typical bank owns multiple subsidiaries that provide commercial banking or other financial services. Banks can also have stakes in non-financial firms although such ownership cannot exceed 5% of the non-financial firm's outstanding equity. For Call Reports, a bank is required to aggregate data only for subsidiaries that provide commercial banking or other financial services. Thus, by definition our data excludes non-financial subsidiaries owned by a bank, if any.

A drawback of our aggregated data is that we are unable to say how diversification within an individual commercial banking subsidiary impacts its operations. However, since most banks with several subsidiaries manage capital centrally (Avraham, Selvaggi, and Vickery (2012)), our aggregated data provides the ideal empirical setting for our analysis. In addition, for all banks in our sample, traded equity prices reference the entire firm, and not individual subsidiaries. Were we to use data only for individual commercial banking subsidiaries, we would be unable to carry the analysis that relies on traded equity returns.

2.2 Measuring the extent of bank diversification

We begin by collecting the most granular data for income and expenses that is available across all business segments for all banks using the publicly-available, quarterly call reports required to be

[4]As of March 2024, the total number of banks in the U.S. is 4,568. However, most of these banks are small or are privately-owned and stock return data is not available for them. Thus, we are unable to compute measures of idiosyncratic risk for banks that are excluded from our analysis.

filed by all bank holding companies in the U.S. Specifically, we collect data for income and expenses across sixteen different categories of bank business segments - seven for interest income and nine for noninterest income. The seven sources of interest income include income and expenses that a bank accrues or incurs from: (i) loans in both domestic 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 fixed income. The nine sources of noninterest income include income and expenses that a bank accrues or incurs 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 noninterest income (for e.g., revenue from safe deposit box rentals and U.S. savings bond redemptions, etc.).

We separately aggregate the data for the seven sources of interest income and the nine sources of noninterest income to compute the total and net interest and noninterest incomes for each bank for each quarter. Using the aggregated values of total and net interest and noninterest incomes, we construct the six measures of bank business segment diversification widely used by practitioners, regulators, and academics to measure bank diversification: These are (i) one minus the Herfindahl-Hirschman index of total noninterest and total interest income (the HHindex measure), (ii) one minus the absolute value of the difference between net interest and total noninterest income divided by the sum of net interest and total noninterest income (the Absdiff measure), (iii) the ratio of *net* noninterest income to the sum of *net* noninterest and interest income (the R-netnet measure), (iv) the ratio of *total* noninterest income to the sum of *total* noninterest and *net* interest income (the R-totnet measure), (v) the ratio of *total* noninterest income to *total* noninterest and interest income (the R-tottot measure), and (vi) finally, the *simple* ratio of total noninterest to interest income (the R-simple measure). These measures have been used widely in studies such as Stiroh (2004), Stiroh (2006), Stiroh and Rumble (2006), Baele, De Jonghe, and Vander Vennet (2007a), Laeven and Levine (2007), Lepetit, Nys, Rous, and Tarazi (2008a), Demirguc-Kunt and Huizinga (2010), Guerry and Wallmeier (2017), and Saunders, Schmid, and Walter (2020a), among many others. The detailed definition of these six measures of bank diversification is listed in Table 1

In addition, we follow Jacquemin and Berry (1979) and Khanna and Palepu (2000) to define the Entropy measure of bank diversification. To compute Entropy, 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, Entropy 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 Entropy measure for bank i in quarter t equals (Detailed definition in the last row of Table 2.2):

$$Entropy_{i,t} = \sum_{16}^{1} S_{j,i,t} ln \left(\frac{1}{S_{j,i,t}} \right) \tag{5}$$

Entropy values can range from a maximum of 2.77 to a minimum of 0 for any bank. Entropy 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 as measured by Entropy. Entropy for a bank that reports \$15B for one income item, \$1B for another, and zero for all remaining sources would equal 0.23. Such a bank has almost the lowest level of diversification per Entropy measure.

Panel A of Table 2 presents summary statistics for the cross-section of banks. The average bank in our sample has an Entropy of 1.18 with a standard deviation of 0.36. The dispersion in Entropy is large as it varies from 0.95 to nearly 1.43 for banks at the 25^{th} and the 75^{th} percentile, respectively. The maximum value of Entropy for any bank over our sample period is 2.47, compared to the maximum theoretical possible value of 2.77.

Panel A also presents summary statistics for the six other measures of diversification commonly used in the literature. We observe that the means of many measures of diversification differ significantly even though there are only minute differences in their definitions. For instance, the means of R-netnet, R-totnet, and R-tottot are -0.66, 0.25 and 0.19, respectively, even though these measures only differ in whether they use total or net interest and noninterest incomes to compute the diversification measures.

Panel B of Table 2 presents summary statistics for the aggregate bank sector. To compute aggregate time-series, 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 for each quarter Call Report data for a particular bank is available for the previous and current quarters. This requirement ensures that our series are not affected by entry or exit of banks.[5] We then aggregate the income and expenses data for across each of the sixteen different categories of bank business segment listed above to obtain time series data for the aggregate bank sector. We repeat this process for all sixteen categories, and then use the aggregated data to compute the seven measures of bank diversification listed in Table 1.

Panel B shows that there is substantial time variation in the measures of diversification for the aggregate bank sector across time. For example, Entropy for the aggregate bank sector has a mean of 2.04 with a standard deviation of 0.27. Other measures of diversification commonly used in the literature also vary over time. As was the case for the cross-section of banks, the mean values of commonly used measures differ substantially from each other.

Since in Table 2 bank diversification measures have substantially different means and standard deviations, in all our empirical tests (that compare the ability of these measures to capture the diversification effect), we use standardized variables. Thus, coefficients on all diversification measures in all our regressions are directly comparable.

Table 3 documents how various measures of bank diversifications correlate with each other. Panels A and B present correlations for the cross-section and the aggregate bank sector, respectively. A number of interesting facts emerge from this analysis. First, with few exceptions, most

[5] This requirement also means that the actual number of banks used in any quarter varies over time.

measures of bank diversification are positively correlated with each other, and that these correlations are statistically significant at conventional levels (1% level or better). For instance, for the cross-section of banks, as well as for the aggregate U.S. bank sector, the R-tottot and R-simple measures have a correlation of 90% or above.

Second, we note that the at the aggregate bank sector level some widely used measures of bank diversification also exhibit low (positive) correlation with each other, despite the fact that each of these measures claims assess the degree of bank business segment diversification. These low correlations are surprising given that the only difference among some of these measures is whether they use net or total interest and noninterest income. For e.g., the R-tottot measure has a correlation of just 0.05 (not statistically significant) with the R-netnet measure.

Finally, we note the low correlation of Entropy with all existing measures of bank diversification, which ranges from a minimum of -0.02 (with R-simple) to a maximum of 0.41 (with R-netnet). These low correlations for Entropy may be due to the fact that computation of this measure relies on much more granular and detailed data than the computation of other measures. Overall, the low correlations among various measures of bank diversification in Table 3 provide yet another rationale for our study.

2.3 Measuring bank idiosyncratic risk and data for control variables

We collect data for banks' stock prices, holding period returns including dividends, and total shares outstanding from the Center for Research on Security Prices (CRSP). For identifying banks in CRSP, we follow Gandhi and Lustig (2015) andGandhi (2018) and select all firms with the two-digit header standard industrial classification (SIC) code of 60 or a four-digit SIC code of 6712. Several studies also define banks using four-digit SIC codes ranging from 6000–6199. Gandhi and Lustig (2015) show this selection misses bank holding companies (listed under SIC code 6712). We match each bank in CRSP to its Call Report data (i.e., FRY9-C data) using the 'CRSP-FRB Link' provided by the Federal Reserve Bank of New York. This tool uniquely matches each RSSD IDs (a unique bank identifier allocated by the Federal Reserve for banks' regulatory reporting) with

PERMCO (a unique bank identifier allocated by CRSP) and is updated frequently to account for bank mergers, acquisitions, failures, and delistings.

Next, we follow Ang, Hodrick, Xing, and Zhang (2006) and Ang, Hodrick, Xing, and Zhang (2009) and estimate the idiosyncratic volatility for each bank in our sample by regressing bank stock returns on the three Fama and French (1993) stock factors, namely the market (mkt), small minus big (smb), and high minus low (hml). While previous empirical studies suggest that there are many other cross-sectional factors that have explanatory power for the cross-section of returns, we do not directly control for all such factors. Rather, we follow Ang, Hodrick, Xing, and Zhang (2006), who argue that controlling for these additional factors only adds noise. Specifically, we estimate the following regression using daily data:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,mkt} mkt_t + \beta_{i,smb} smb_t + \beta_{i,hml} hml_t + \epsilon_{i,t} \tag{6}$$

We use daily data to estimate equation 6 for each bank for each quarter over our sample period. Idiosyncratic volatility is simply the standard deviation of the residuals (i.e., $\sigma(\epsilon_{i,t})$). Thus, at the end of this exercise, we have a time-series of quarterly idiosyncratic volatility for each bank in our sample. Note that while our primary analyses utilizes the idiosyncratic volatility derived from the Fama and French (1993) three-factor model, in robustness tests, we use idiosyncratic volatility computed using CAPM and the Fama and French (2015) five-factor model. As Section 3 shows, our results are not sensitive to the choice of a factor model.

Table 4 presents the summary statistics for idiosyncratic volatility for the cross-section of banks. Mean idiosyncratic volatility equals nearly 1.91%. However, there is considerable variation over time and in the cross-section, as the standard deviation of idiosyncratic volatility itself is about 1.49, which is similar to the mean. The inter-quartile range (difference in the idiosyncratic volatility between the 25^{th}- and the 75^{th}-percentile) of 1.15% is also indicative of the considerable cross-section and time-series variation in idiosyncratic volatility across banks.

In all our analysis, we include data for a variety of control variables that can affect bank's idiosyncratic risk. Table 4 also presents the summary statistics for these additional control variables for the cross-section of banks. Specifically, for each bank in our sample, we collect data for log 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 the total noninterest and interest expenses by the total noninterest and interest income as a control for a bank's operational efficiency, the ratio of total deposits to total liabilities as a control for bank funding structure, the ratio of total loan loss provisions to total loans as a control for bank risk taking, the growth rate of total book value of assets (computed over the last three years) as a control for the growth opportunities available to a bank, and finally the bank's Z-Score as a control for bank risk-taking and idiosyncratic 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 idiosyncratic 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 idiosyncratic risk. Elsas, Hackethal, and Holzhäuser (2010) suggests that we should control for operational efficiency in our analysis as this too can influence idiosyncratic 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 (idiosyncratic) risk. In Baele, De Jonghe, and Vander Vennet (2007b), loan loss provisions are an important indicator of the amount of bank-specific credit risk. Finally, Saunders, Schmid, and Walter (2020b) 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 4.

3 Results

In this section, we present our main empirical results. We evaluate how well various measures of bank diversification perform in capturing the diversification effect. We begin by studying the ability of various diversification measures to predict bank idiosyncratic risk. We also directly compare the ability of various bank diversification measure to predict idiosyncratic volatility (horse races). We check if our results survive a battery of robustness tests. This section also investigates what factors or characteristics drive Entropy in the cross-section and over time. After documenting which of the seven measures of bank diversification is best at capturing the diversification effect, we revisit the question of how bank diversification relates to bank valuations.

3.1 Bank business segment diversification and idiosyncratic volatility

In this section, we explore how various measures of bank business segment diversification relate to future idiosyncratic volatility for the cross-section of U.S. banks. In particular, we test which diversification measure is best at capturing the diversification effect. We do so by relating each bank's measure of business segment diversification measured in quarter t to the idiosyncratic volatility of its stock returns at time $t+1$ using standard panel regressions. The exact specification of our panel regression is as follows:

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,DIV} DIV_{i,t} + Controls + \eta_i + \gamma_t + \epsilon_{i,t} \tag{7}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of stock returns for bank i measured in quarter $t+1$, $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. We control for several bank-level characteristics that influence the relation between bank diversification and bank idiosyncratic volatility. 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 idiosyncratic volatility can be highly persistent (Ang, Hodrick, Xing, and Zhang (2006)), we also control for lagged idiosyncratic volatility 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 $\beta_{i,DIV}$, i.e., the coefficient on $DIV_{i,t}$. We expect the sign on $\beta_{i,DIV}$ to be negative for each of the seven measures of bank business segment diversification, indicating that higher bank diversification is associated with lower idiosyncratic risk (i.e., the diversification effect) as predicted by the framework in Section 1.

Table 5 presents the estimates for regression (7) and shows that the Entropy measure is best at capturing the diversification effect. Each column of this Table shows the estimates for a separate regression specification – one for each of the seven different measures of bank diversification defined in Table 1. We notice that the coefficient on Entropy is negative and statistically significant at the 1% level. The negative coefficient of -0.08 on Entropy indicates that a one-standard deviation increase in Entropy for a particular bank in a particular quarter is associated with a nearly 0.08% lower idiosyncratic volatility of daily returns for this bank over the next quarter.

The negative relation between Entropy and future bank idiosyncratic is not only statistically but economically significant as well. A one-standard deviation increase in Entropy at the bank level implies that idiosyncratic volatility will be lower by nearly 0.08% over the next quarter. Given that the average quarterly idiosyncratic volatility for the banks in our sample is 1.91%, this implies that higher Entropy is associated with nearly 4.19% lower idiosyncratic volatility over the next quarter as compared to the sample mean.

In addition, to Entropy, the coefficients on R-totnet, R-tottot, and R-simple (Columns (4), (6), and (7)) are also negative but not statistically significant, indicating that higher values of these measures are not associated with lower future idiosyncratic volatility for banks. Since all right hand side variables are standardized, the magnitude of the coefficients in columns (1) - (7) of Table 5 are directly comparable, and indicate that the Entropy measure is the best at capturing the diversification effect. While a one-standard deviation increase in R-totnet, R-tottot and R-simple is each associated with lower idiosyncratic volatility by -0.03%, the magnitude of these coefficients are at least 50% lower as compared to the coefficient on Entropy in Column (1) of Table 5.

Conversely, there is no statistically significant relation between the Hhindex and Absdiff measures of bank diversification and future idiosyncratic volatility of bank stock returns. The coefficient on R-netnet is positive, indicating that higher values of diversification (as measured by R-netnet) are associated with higher rather than lower idiosyncratic volatility.

Notice also that in Table 5, the signs and significance of the coefficients on the control variables are as expected. For example, the coefficient on lagged idiosyncratic volatility is positive and statistically significant, indicating that idiosyncratic volatility 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 idiosyncratic risk. For banks, government-guaranteed deposits are considered as a source of low risk, low cost, capital, it is not surprising that higher values of total deposits to total liabilities is also associated with lower idiosyncratic 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 idiosyncratic risk, as indicated by the positive, statistically significant coefficients on these variables in all specifications in Table 5.

Table 6 directly compares the ability of the Entropy measure to forecast idiosyncratic volatility with that of the six other measures of bank diversification used in the literature. Since all diversification measures are positively correlated with each other and since some of these correlations are quite high, to avoid any issues related to multi-collinearity, in each regression specification we include only Entropy and only one of the six other measures of bank diversification. Specifically, we run a panel regression of the form:

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t} \tag{8}$$

Here, $Entropy_{i,t}$ is the Entropy measure for bank i measured in quarter t and $Div_{i,t}$ is one of the six other measures of bank diversification i.e., Hhindex, Absdiff, R-totnet, R-netnet, R-tottot, or R-simple. Each column in Table 6 shows the results for a different specification of the regression in equation (8). As above, in all cases we control for several bank characteristics, the lagged bank idiosyncratic volatility, and all regressions include bank fixed effects.

Table 6 shows that in all cases, Entropy still emerges as the best at capturing the diversification effect. The coefficient on Entropy is always negative and statistically significant at the 1% level. This coefficient ranges from -0.08 (when including R-simple measure of bank diversification) to -0.09 (when including the Hhindex measure of bank diversification) and the *t*-statistics are all above 4.

Once we control for Entropy, the coefficient on most other measures of bank diversification is either positive or not statistically significant. For instance, the coefficient on Hhindex, Absdiff, R-totnet, and R-netnet are all positive and statistically significant indicating that an increase in these measures of bank diversification is associated with an increase rather than a decrease in future idiosyncratic volatility. The coefficient on R-tottot and R-simple is negative but not statistically significant, indicating that once we include the Entropy measure of bank diversification, it renders the relation between these latter measures and future bank idiosyncratic volatility meaningless. The magnitude of the coefficients on these other measures of diversification is also at best about one-third of the magnitude of the coefficient on Entropy, indicating that the Entropy measure is the best at capturing the expected negative relation between bank diversification and idiosyncratic volatility of bank stock returns.

In Table 7 we test if the ability of Entropy to capture the diversification effect varies over time, i.e., do our results differ in good or bad economic times or before and after the global financial crisis. We define bad economic times as quarters of NBER recessions as well as quarters with financial crisis (failure of Long-Term Capital Management, Russian sovereign debt crisis, etc.). Quarters with recessions are identified and published by the NBER business cycle dating committee.

Table 7 shows that in both good and bad economic times as well as in the pre- and post-crisis years, Entropy is significantly negatively correlated with future idiosyncratic volatility of bank stock returns. The coefficient on Entropy in this Table is -0.06 in good economic times but is nearly twice as high at -0.15 during bad economic times, indicating that if anything the ability of Entropy to capture the diversification effect is even better during periods of financial distress and crisis. Both these coefficients are statistically significant at the 1% level with t-statistics of -3.18 and -3.14, respectively.

Similarly, the third and fourth column of Table 7 shows that the ability of Entropy to predict lower idiosyncratic risk does not change after the global financial crisis. In both the pre- and post-crisis years, the coefficient is nearly the same (-0.08 and -0.09, respectively) and both coefficients are statistically significant with t-statistics of -3.46 and -2.18, respectively. Note that all coefficients in Table 7 are economically significant as well. For instance, in the second column, a coefficient of -0.15 on Entropy implies that a one-standard deviation increase in Entropy is associated with a -0.15% decline in idiosyncratic volatility of bank stock returns, which is nearly 7.85% of the sample mean idiosyncratic volatility of 1.91%.

3.2 Robustness tests

We carry out a series of robustness tests to ensure that the ability of Entropy to capture the diversification effect withstands any change in the design of our empirical tests. We begin by checking if our results are robust to the empirical asset pricing factor model employed to compute bank idiosyncratic risk. The results of this test are in Table 8.

Table 8 reruns the regression in Equation (8). Panels A and B of this Table uses volatility of idiosyncratic returns derived from the CAPM and the Fama-French Fama and French (2015) 5-factor models as the dependent variable, respectively. As above, all regressions control for all bank-level characteristics above (coefficients on these control variables are not reported for brevity). In addition, we control for overall market volatility (VIX) or a bank's exposure to market risk (β) as these variables can also determine the degree of bank idiosyncratic risk.

Regardless of the empirical asset pricing factor model employed to estimate bank idiosyncratic risk, Entropy emerges as the best at capturing the diversification effect. The coefficient on Entropy is always negative and statistically significant at the 1% level. This coefficient ranges from -0.08 to -0.09 and the t-statistics in all cases are close to -4. In fact, the results in Table 8 are even stronger than those in Table 6. In Table 8 coefficients on most other measures of bank diversification are either positive or not statistically significant (the only exception being R-simple for which the coefficient is only marginally statistically significant). The results in Table 8 also show that controlling for market volatility (VIX) or a bank's exposure to market risk factor (β) does not impact our results and conclusion.

In an influential and highly cited paper, Demsetz and Strahan (1997), argue that the problem with using volatility of idiosyncratic returns (as we have in all our tests above) is that this variable is not only influenced by the degree of bank business segment diversification but also by individual components of its balance sheet (i.e., its assets, liabilities, off-balance sheet positions, and leverage). Demsetz and Strahan (1997) go on to suggest that one should use the R^2 from a factor model instead of the volatility of idiosyncratic returns as an alternative market-based measure of the degree of diversification by a bank. Their suggestion is based on earlier papers such as, Barnea and Logue (1973) and Roll (1988), who advocate for the use of R^2 from a simple factor model to measure the degree of conglomerate diversification.

Given these arguments, Table 9 repeats the analysis in Table 6, but now uses the R^2 from a factor model to measure firm-specific risk or as the dependent variable. Following the arguments in Barnea and Logue (1973), Roll (1988), and Demsetz and Strahan (1997), systematic risk factors should explain a greater proportion of variation in stock market returns (i.e., R^2 should be higher) or firm-specific risk or variation (i.e., 1 - R^2) should be lower for banks with higher business segment diversification. In other words, in this next test, we expect the coefficients on all measures of bank business-segment diversification to be positive and statistically significant.

Table 9 presents the results and again Entropy emerges as the best at capturing the diversification effect. In all cases, the coefficient on Entropy is positive and statistically significant. The coefficient on almost all other measures either has the wrong sign or these coefficients are not statistically significant at conventional levels. For instance, the coefficient on R-simple is negative and statistically significant at the 10% level.

Next, we check if the ability of Entropy to predict lower idiosyncratic risk and capture the diversification effect is limited to just 1-quarter ahead or whether the relation between Entropy and idiosyncratic returns is statistically significant at longer horizons as well. Table 10 presents the results of this analysis. Panels A, B, and C present the estimates for horse-races that compare the ability of various diversification measures to predict bank idiosyncratic risk 2-, 3-, and 4-quarters ahead. These results indicate that of all the bank business segment diversification measures used in the literature, Entropy is the only one that has a statistically significant (negative) correlation with idiosyncratic risk beyond a horizon of 1-quarter. In all cases, when predicting idiosyncratic risk at a horizon of 2-, 3-, or 4-quarters ahead, the coefficient on Entropy is negative and statistically significant at the 1% level with t-statistics close to or above 4. The coefficients on the six remaining diversification measures are small, statistically insignificant, and often switch sign indicating these measures are sometimes associated with higher and at other times lower idiosyncratic volatility at longer horizons. Table 10 provides yet another piece of evidence that supports Entropy as the best measure at capturing the diversification effect for banks.

In our final robustness test in Table 11 we compare the out-of-sample performance of various diversification measures to predict bank idiosyncratic volatility. Specifically, we compare the out-of-sample performance of models that (separately) use the seven diversification measures – Entropy, Hhindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple to predict bank idiosyncratic volatility. We also compare the performance of models that use these seven diversification variables by themselves (i.e., univariate regressions or a model without any control variables) as well as a model that contains each of these seven diversification variables along with all of the control variables from our baseline regression in Table 5 (i.e., multivariate regressions or a model with all control variables listed above).

For each specification and model, we measure to what extent a market participant could have predicted bank idiosyncratic volatility in real time, using the most recent data that was available to her up to that point in time. That is, we first run forecasting regressions with the selected model using data from the most recent 3-year or 5-year window. We then use the estimated parameters of this model to predict idiosyncratic volatility for one quarter following the window. For the 3-year window, our out-of-sample forecasts start in December 1999, when we first have 3 years of data to estimate the parameters of the model. Similarly, for the 5-year window, our out-of-sample forecasts start in December 2001, when we first have 5 years of data to estimate the parameters of the model. In all cases, we estimate the root mean squared error (i.e., the RMSE), which is defined as the square root of the squared differences between the actual (i.e., realized) and predicted values of idiosyncratic volatility.

Table 11 presents the results for out-of-sample forecasts for the models without (Panel A) and with (Panel B) the inclusion of all the control variables. In each panel, each row presents the RMSEs for a model that either uses a 3- or 5-year window to estimate the model parameters. The columns report the RMSEs for separate models that use one of the seven diversification measure to forecast bank idiosyncratic volatility 1-quarter ahead. For instance the number in the first row and column of panel A of Table 11 reports the RMSE when data for Entropy is used by itself over 3-year rolling windows to predict idiosyncratic volatility 1-quarter ahead out-of-sample. In all cases, we multiply the RMSEs by 100 and express these in percentages.

Panel A of Table 11 indicates that when predicting bank idiosyncratic volatility without the use of any control variables, the model with Entropy has the lowest RMSE at both the 3- and 5-year horizons (1.4011% and 1.3160%, respectively) as compared to all other models. Models that use other diversification measures never outperform the model that uses Entropy by itself as measured by out-of-sample RMSE.

This is clear even in Panel B of this Table that shows the results for similar out-of-sample tests but now uses each of the seven diversification measures along with all control variables listed above. Again, the model that uses Entropy along with all control variables has significant predictive ability for future bank idiosyncratic volatility (lowest RMSE at 1.6795% and 1.7627% for 3- and 5-year windows, respectively).

Note that comparing the results in Panels A and B of Table 11 indicates that across all our models, the one that uses Entropy by itself (1st column of Panel A) produces the lowest RMSE, thus indicating that predicting with multiple variables (diversification measure and control variables) appears to only add noise. This result is consistent with the fact that in most out-of-sample predictive tests, it is often the most parsimonious model (such as the one with single variables) that performs well

3.3 What drives bank Entropy?

In this section we investigate what drives or explains variation in Entropy in the cross-section or over time. We begin by plotting time-series data for Entropy for the aggregate U.S. bank sector in Figure 1. In this figure, the gray-shaded bars represent periods of NBER recessions or financial crisis. The time series plot shows that variation in Entropy coincides with the passage of major legislations related to banking. For instance, Entropy increases significantly and remains at elevated levels for all banks post-1999 – when the Gramm-Leach-Bliley Act repealed the restrictions placed on banking activities by the Glass-Steagall Act of 1933. In similar plots (not shown for clarity) no other measure of diversification shows a significant increase post-1999.

Similarly, Entropy drops significantly and remains at low levels for all banks post 2007-2009, and especially after the passage of the Dodd-Frank Act of 2010. While all other measures also fall during 2007-2009, indicating perhaps that bank noninterest income fell during the crisis, they soon revert to their pre-crisis levels. Thus, existing measures of bank diversification suggest that commercial banks in the U.S. were as diversified after Dodd-Frank as they were before, which perhaps may not actually be the case. Along with the regression results in Section 3.1 above, the plot in Figure 1 suggests once again that the Entropy measure is best at capturing the degree of business segment diversification by banks, as it is the only measure that seemingly reacts to changes in regulations that place or remove restriction on banks' activities.

We use standard contemporaneous panel regressions to investigate what factors cause Entropy for a particular bank to be higher or lower. In other words, we regress Entropy on a number of balance sheet and income statement variables that could drive banks' decision to diversify into various business segments. These variables include proxies for bank size, leverage, profitability, asset growth, and bank risk.

Table 12 presents the results of this regression, where a positive (negative) coefficient implies that higher values of that variable are associated with higher (lower) Entropy, and thus higher (lower) business segment diversification. For instance, in column (1), the coefficient on log book value of assets is positive (statistically significant at the 1% level with a t-statistic of 10.55), implying that, as expected, larger banks or banks with more book value of assets are more diversified or have higher Entropy. This result is well-known in the literature.

In column 2, a negative coefficient of -0.045 on the ratio of bank capital to total book value of assets (again statistically significant at 1% level with a t-statistic of -8.83) indicates that an increase in the amount of capital held by banks (i.e., higher capitalization or lower leverage) in a particular quarter is associated with a decrease in its Entropy or business segment diversification over the following quarter.

In Table 12, the only other variables that are significantly related with Entropy are the bank's profitability and Z-score. As expected banks with higher profitability in one quarter tend to have higher Entropy in the subsequent quarter. This makes intuitive sense as banks with higher profitability are the ones that are more likely to have the resources to expand into multiple business segments.

Similarly, in the Table, the coefficient on a proxy for a bank's banruptcy risk (Z-score) is positive at 0.033. This coefficient is statistically significant at the 1% level with a t-statistic of 7.24. The positive coefficient indicates that banks with higher bankruptcy risk in one quarter tend to have higher diversification as measured by Entropy in the subsequent quarter. This result is consistent with prior literature that suggests a relation between diversification and the probability and cost of financial distress. For e.g., results in Lewellen (1971), Mansi and Reeb (2002), and Leland (2007) suggest that nonfinancial firms with higher risk of distress tend to diversify into multiple business segments to seek 'coinsurance' of income benefits. In other words, business segment diversification tends to be higher among firms with higher distress risk as imperfect correlation among cash flows from different business segments can help reduce such bankruptcy risk. The results in Table 12 suggests that this argument also holds for banks.

The results in Column (6) of Table 12 for a specification including all independent variables mirror those in Columns (1) through (5). Comparing coefficients across columns, we note that all our conclusions survive, and the statistical significance as indicated by the t-statistics in column (6) either remains comparable to the regressions in Columns (1) – (5) or becomes slightly stronger. While a bank's asset growth was not a statistically-significant determinant of its Entropy in only marginally statistically significant with a t-statistic of just -1.87.

4 Conclusion

Our study introduces an innovative Entropy-based measure for bank business segment diversification incorporating granular information from both the number of distinct business lines and the proportion of income that a bank derives from these business lines. The Entropy measure is uncorrelated with the six popular measures of bank business segment diversification used in the extant literature and is best at capturing the diversification effect both in-sample as well as in out-of-sample testing. Our results survive extensive robustness checks and hold regardless of how idiosyncratic volatility is measured, and are evident across different economic scenarios and sample periods.

Our study has several practical application which can be of interest to both practitioners and regulators. For instance, the existing literature documents conflicting results regarding bank diversification and valuation with some studies indicating a positive, while others documenting a negative relation between bank business segment diversification and valuation. Given that we document that the Entropy measure is the 'best' measure of bank business segment diversification, and that other popular measures commonly used in the literature are unable to capture the diversification effect, our study can be used revisit studies that examine the link between bank diversification and valuation.

Since the Entropy measure can be estimated quite easily at a fairly high frequency (quarterly level) for all banks (public or private), future studies can use the Entropy measure to re-examine many of the existing studies on the impact of bank business segment diversification on bank risk-taking, executive compensation, and can serve as a practical tool for regulators and researchers in this area.

References

Acharya, V. V., I. Hasan, and A. Saunders. 2006. Should banks be diversified? Evidence from individual bank loan portfolios. *Journal of Business* 79:1355–1412.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2009. High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics* 91:1–23.

Avraham, D., P. Selvaggi, and J. Vickery. 2012. A Structural View of U.S. Bank Holding Companies. *Federal Reserve Bank of New York Economic Policy Review* 18:65–81.

Baele, L., O. De Jonghe, and R. Vander Vennet. 2007a. Does the stock market value bank diversification? *Journal of Banking and Finance* 31:1999–2023.

Baele, L., O. De Jonghe, and R. Vander Vennet. 2007b. Does the stock market value bank diversification? *Journal of Banking and Finance* 31:1999–2023.

Barnea, A., and D. E. Logue. 1973. Stock-market based measures of corporate diversification. *The Journal of Industrial Economics* pp. 51–60.

Beatty, A., and S. Liao. 2014. Financial accounting in the banking industry: A review of the empirical literature. *Journal of accounting and Economics* 58:339–383.

Berger, P. G., and E. Ofek. 1995. Diversification's effect on firm value. *Journal of financial economics* 37:39–65.

Campa, J. M., and S. Kedia. 2002. Explaining the diversification discount. *The journal of finance* 57:1731–1762.

Demirguc-Kunt, A., and H. Huizinga. 2010. Bank activity and funding strategies: The impact on risk and returns. *Journal of Financial Economics* 98:626–650.

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

Deng, S., and E. Elyasiani. 2008. Geographic diversification, bank holding company value, and risk. *Journal of Money, Credit and Banking* 40:1217–1238.

Elsas, R., A. Hackethal, and M. Holzhäuser. 2010. The anatomy of bank diversification. *Journal of Banking and Finance* 34:1274–1287.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.

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

Gandhi, P. 2018. The Relationship between Credit Growth and Expected Returns of Bank Stocks. *European Financial Management, Special Issue, Corporate Policies and Asset Prices* 24.

Gandhi, P., and H. Lustig. 2015. Size Anomalies in Bank Stock Returns. *Journal of Finance* 70:733–768.

Gelman, M., I. Goldstein, and A. MacKinlay. 2023. Bank diversification and lending resiliency. *Available at SSRN 4147790* .

Goetz, M. R., L. Laeven, and R. Levine. 2013. Identifying the valuation effects and agency costs of corporate diversification: Evidence from the geographic diversification of US banks. *Review of Financial Studies* 26:1787–1823.

Goetz, M. R., L. Laeven, and R. Levine. 2016. Does the geographic expansion of banks reduce risk? *Journal of Financial Economics* 120:346–362.

Gorton, G. B., and G. G. Pennacchi. 1995. Banks and loan sales marketing nonmarketable assets. *Journal of monetary Economics* 35:389–411.

Guerry, N., and M. Wallmeier. 2017. Valuation of diversified banks: New evidence. *Journal of Banking and Finance* 80:203–214.

Hann, R. N., M. Ogneva, and O. Ozbas. 2013. Corporate diversification and the cost of capital. *Journal of Finance* 68:1961–1999.

Hanson, S. G., A. Shleifer, J. C. Stein, and R. W. Vishny. 2015. Banks as patient fixed-income investors. *Journal of Financial Economics* 117:449–469.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* 28:650–705.

Jacquemin, A. P., and C. H. Berry. 1979. Entropy measure of diversification and corporate growth. *Journal of Industrial Economics* 27:359–369.

Khanna, T., and K. Palepu. 2000. Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups. *The Journal of Finance* 55:867–891.

Kuppuswamy, V., and B. Villalonga. 2016. Does diversification create value in the presence of external financing constraints? Evidence from the 2007–2009 financial crisis. *Management Science* 62:905–923.

Laeven, L., and R. Levine. 2007. Is there a diversification discount in financial conglomerates? *Journal of Financial Economics* 85:331–367.

Lang, L. H., and R. M. Stulz. 1994. Tobin's q, corporate diversification, and firm performance. *Journal of political economy* 102:1248–1280.

Leland, H. E. 2007. Financial synergies and the optimal scope of the firm: Implications for mergers, spinoffs, and structured finance. *The Journal of finance* 62:765–807.

Lepetit, L., E. Nys, P. Rous, and A. Tarazi. 2008a. Bank income structure and risk: An empirical analysis of European banks. *Journal of Banking and Finance* 32:1452–1467.

Lepetit, L., E. Nys, P. Rous, and A. Tarazi. 2008b. Bank income structure and risk: An empirical analysis of European banks. *Journal of Banking and Finance* 32:1452–1467.

Levine, R., C. Lin, and W. Xie. 2021. Geographic diversification and banks funding costs. *Management Science* 67:2657–2678.

Lewellen, W. G. 1971. A pure financial rationale for the conglomerate merger. *The journal of Finance* 26:521–537.

Mansi, S. A., and D. M. Reeb. 2002. Corporate diversification: what gets discounted? *The journal of Finance* 57:2167–2183.

Merton, R. C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics* 3:125–144.

Modigliani, F., and M. H. Miller. 1958. The cost of capital, corporation finance and the theory of investment. *The American economic review* 48:261–297.

Nagel, S., and A. Purnanandam. 2020. Banks risk dynamics and distance to default. *The Review of Financial Studies* 33:2421–2467.

Roll, R. 1988. Rˆ2. *The Journal of Finance* 43:541–566.

Roll, R., and S. A. Ross. 1984. The arbitrage pricing theory approach to strategic portfolio planning. *Financial analysts journal* 40:14–26.

Saunders, A., M. Schmid, and I. Walter. 2020a. Strategic scope and bank performance. *Journal of Financial Stability* 46:100715.

Saunders, A., M. Schmid, and I. Walter. 2020b. Strategic scope and bank performance. *Journal of Financial Stability* 46:100715.

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

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

Stiroh, K. J. 2006. A portfolio view of banking with interest and noninterest activities. *Journal of Money, Credit and Banking* pp. 1351–1361.

Stiroh, K. J., and A. Rumble. 2006. The dark side of diversification: The case of US financial holding companies. *Journal of Banking and Finance* 30:2131–2161.

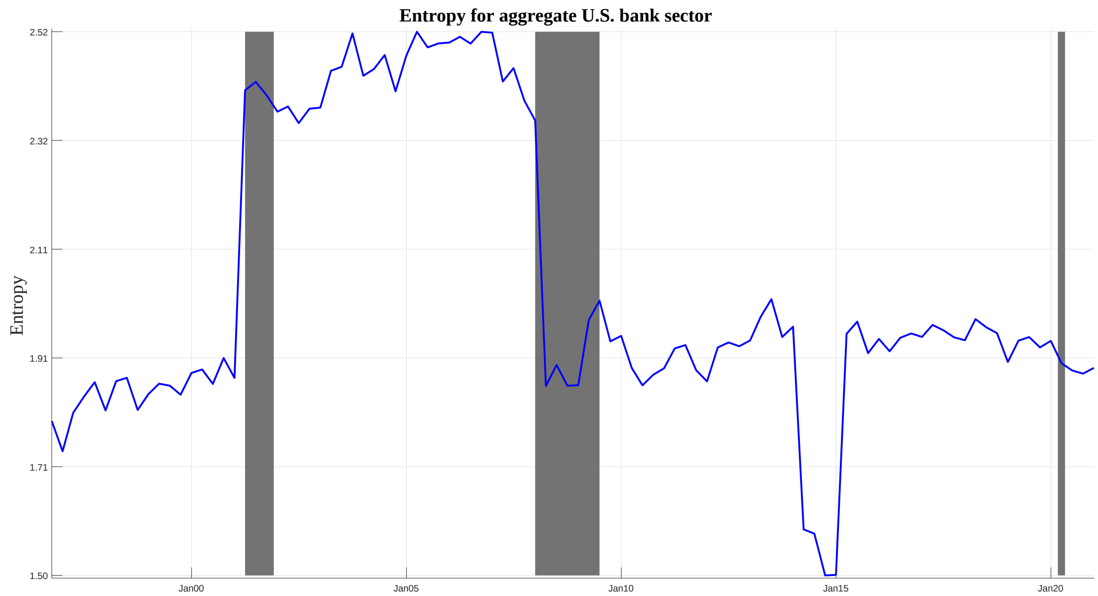

Figure 1. Time series plot for Entropy for the aggregate U.S. bank sector.

Notes: This figure plots Entropy for all domestic banks in the U.S. The blue solid line plots Entropy and the grey shaded regions represent National Bureau of Economic Research (NBER) recessions as well as periods of financial crisis. Years and months are on indicated on the X-axis. The NBER recession dates are published by the NBER Business Cycle Dating Committee. Quarterly data, 1990 – 2023.

Table 1. Business line diversification measures.

Notes: This table provides details regarding the construction of measures of bank diversification used in our empirical analysis. To construct these measures, we collect balance sheet data for banks from the 'Report of Condition and Income (Call Report)' 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 measure of bank diversification in our empirical analysis. Column titled 'Definition of measure' provides a brief description of how the measure is constructed.

Measure	Business line diversification measure	
Entropy	$\sum_{i=1}^{16} S_i \ln \frac{1}{S_i}$ S_i is the share of each income item	This paper
HHindex	$1 - [(\frac{\text{Noninterest income}}{\text{Sum}})^2 + (\frac{\text{Net interest income}}{\text{Sum}})^2]$ $Sum = \text{Noninterest} + \text{Net interest income}$	Stiroh and Rumble (2006)
Absdiff	$1 - \lvert\frac{\text{Net interest income - Noninterest income}}{\text{Net interest income + Noninterest income}}\rvert$	Laeven and Levine (2007)
R-netnet	$\frac{\text{Net noninterest income}}{\text{Net noninterest income + Net Interest income}}$	Stiroh (2004)
R-totnet	$\frac{\text{Noninterest income}}{\text{Noninterest income + Net Interest income}}$	Lepetit, Nys, Rous, and Tarazi (2008b)
R-tottot	$\frac{\text{Noninterest income}}{\text{Noninterest income + Interest income}}$	Baele, De Jonghe, and Vander Vennet (2007b)
R-simple	$\frac{\text{Noninterest income}}{\text{Interest income}}$	Saunders, Schmid, and Walter (2020a)

Table 2. Summary statistics: Diversification measures.

Notes: This table presents summary statistics for the key variables for the cross-section of banks. Column 1 indicates the variable for which the summary statistics is computed. Columns 2 - 6 report the mean, the standard deviation, the minimum, the 25^{th}-percentile, the median, the 75^{th}-percentile, and the maximum values for each variable. Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Panel A shows the summary statistics for the cross-section of banks and Panel B shows the summary statistics for the aggregate U.S. bank sector. Quarterly data, 1996Q3 – 2020Q4.

Variable	Mean	σ	Min	25^{th}	Median	75^{th}	Max
Panel A: Cross-section of banks							
Entropy	1.18	0.36	0.01	0.95	1.19	1.43	2.47
HHindex	0.33	0.11	-0.03	0.26	0.35	0.42	0.50
Absdiff	0.46	0.21	-0.03	0.31	0.45	0.60	0.96
R-netnet	-0.66	0.62	-3.92	-0.81	-0.52	-0.35	1.13
R-totnet	0.25	0.15	-0.01	0.16	0.23	0.31	0.84
R-tottot	0.19	0.13	-0.01	0.11	0.17	0.24	0.75
R-simple	0.29	0.38	-0.01	0.12	0.20	0.32	2.93
Panel B: Aggregate bank sector							
Entropy	2.04	0.27	1.49	1.87	1.94	2.37	2.52
HHindex	0.49	0.02	0.37	0.49	0.50	0.50	0.50
Absdiff	0.91	0.09	0.48	0.88	0.93	0.97	1.00
R-netnet	-0.32	0.24	-1.83	-0.37	-0.25	-0.19	-0.07
R-totnet	0.46	0.05	0.24	0.44	0.47	0.50	0.54
R-tottot	0.36	0.07	0.16	0.30	0.37	0.40	0.48
R-simple	0.58	0.17	0.20	0.43	0.58	0.66	0.93

Table 3. Correlations.

Notes: This table presents the correlation among the seven measures of bank diversification, i.e., Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple computed as described in Table 2.2. Panel A shows the correlations for the cross-section of banks and Panel B shows the correlations for the aggregate U.S. bank sector. Statistical significance at the 10%, 5% and 1% levels is denoted by *, **, and *** respectively. Quarterly data, 1996Q3 – 2020Q4.

Variable	Entropy	HHindex	Absdiff	R-netnet	R-totnet	R-tottot	R-simple
Panel A: Cross-section of banks							
Entropy	1.00						
HHindex	0.54***						
Absdiff	0.54***	0.97***	1.00				
R-netnet	0.14***	0.14***	0.15***	1.00			
R-totnet	0.39***	0.66***	0.70***	0.28***	1.00		
R-tottot	0.35***	0.60***	0.64***	0.19***	0.95***	1.00	
R-simple	0.17***	0.27***	0.31***	0.22***	0.84***	0.90***	1.00
Panel B: Aggregate bank sector							
Entropy	1.00						
HHindex	0.28***						
Absdiff	0.29***	0.91***	1.00				
R-netnet	0.41***	0.76***	0.65***	1.00			
R-totnet	0.33***	0.85***	0.95***	0.63***	1.00		
R-tottot	0.02	0.51***	0.64***	0.05	0.66***	1.00	
R-simple	-0.02	0.45***	0.60***	0.01	0.64***	0.99***	1.00

Table 4. Summary statistics: Control variables.

Notes: This table presents summary statistics for the key variables for the cross-section of banks. Column 1 indicates the variable for which the summary statistics is computed. Columns 2 - 6 report the mean, the standard deviation, the minimum, the 25^{th}-percentile, the median, the 75^{th}-percentile, and the maximum values for each variable. Quarterly data, 1996Q3 – 2020Q4.

Variable	Mean	σ	Min	25^{th}	Median	75^{th}	Max
Idio volatility	1.91	1.49	0.01	1.06	1.49	2.21	33.10
Log Assets	14.97	1.72	11.97	13.71	14.57	15.85	21.94
Capital/Assets	9.01	1.97	5.10	7.76	8.77	9.96	16.75
Operating Profits	0.41	0.24	-0.11	0.30	0.41	0.51	1.16
Cost/Income	74.78	12.22	48.63	67.33	73.96	80.47	131.40
Deposits/Liabilities	83.81	11.59	33.22	78.76	86.57	91.90	98.97
Loan loss provisions	0.15	0.24	-0.10	0.03	0.07	0.15	1.50
Assets growth	51.08	66.43	-26.31	13.28	32.23	64.28	393.92
Z-score	2.12	0.46	0.64	1.88	2.20	2.44	2.93
Market beta	0.68	0.49	-0.11	0.25	0.66	1.05	1.93
VIX	20.58	8.02	10.31	14.23	19.32	25.09	58.59

Table 5. Predicting idiosyncratic volatility.

Notes: This table shows the estimated coefficients for the univariate forecasting regressions:

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Idiosyncratic volatility is defined as the standard deviation of the residuals obtained by regressing daily bank stock returns on the three Fama and French (1993) stock return factors. We control for the lagged dependent variable (i.e., $\sigma_{i,t}$). $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Entropy	-0.08***						
	(-4.23)						
Hhindex		0.01					
		(0.14)					
Absdiff			0.01				
			(0.04)				
R-totnet				-0.02			
				(-0.72)			
R-netnet					0.01		
					(0.22)		
R-tottot						-0.03	
						(-1.53)	
R-simple							-0.03
							(-1.57)
$\sigma_{i,t}$	6.10***	6.20***	6.20***	6.20***	6.20***	6.20***	6.20***
	(18.97)	(28.80)	(28.78)	(28.75)	(28.76)	(28.74)	(28.75)
Assets	-0.25***	-0.23***	-0.23***	-0.23***	-0.23***	-0.24***	-0.23***
	(-4.33)	(-4.15)	(-4.17)	(-4.21)	(-4.26)	(-4.28)	(-4.23)
Capital/Assets	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***
	(-10.29)	(-10.14)	(-10.14)	(-10.14)	(-10.09)	(-10.13)	(-10.12)
Operating profits	-0.04*	-0.05*	-0.05*	-0.04	-0.05*	-0.03	-0.03
	(-1.69)	(-1.80)	(-1.77)	(-1.36)	(-1.92)	(-0.94)	(-1.12)
Cost/Income	0.11***	0.10***	0.10***	0.11***	0.10***	0.11***	0.11***
	(3.88)	(3.57)	(3.57)	(3.68)	(3.60)	(3.90)	(3.84)
Deposits/Liabilities	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	(1.43)	(1.13)	(1.13)	(1.08)	(1.13)	(1.09)	(1.09)
Loan loss provisions	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***
	(10.21)	(10.23)	(10.26)	(10.26)	(9.98)	(10.23)	(10.26)
Asset growth	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*
	(-1.74)	(-1.73)	(-1.73)	(-1.73)	(-1.73)	(-1.70)	(-1.71)
Z-Score	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***
	(-5.43)	(-5.65)	(-5.64)	(-5.69)	(-5.69)	(-5.72)	(-5.73)
VIX	0.36***	0.34***	0.34***	0.34***	0.34***	0.35***	0.34***
	(12.29)	(11.71)	(11.71)	(11.73)	(11.80)	(11.74)	(11.75)
β	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**
	(2.03)	(2.05)	(2.05)	(2.07)	(2.04)	(2.08)	(2.05)
Adjusted R^2	0.60	0.60	0.60	0.60	0.60	0.60	0.60
Bank fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*

Table 6. Predicting idiosyncratic volatility: Horseraces.

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

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Idiosyncratic volatility is defined as the standard deviation of the residuals obtained by regressing daily bank stock returns on the three Fama and French (1993) stock return factors. We control for the lagged dependent variable (i.e., $\sigma_{i,t}$). $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)
Entropy	-0.09***	-0.09***	-0.08***	-0.08***	-0.08***	-0.08***
	(-4.27)	(-4.28)	(-4.10)	(-4.23)	(-3.97)	(-4.14)
Hhindex	0.03					
	(1.31)					
Absdiff		0.02				
		(1.28)				
R-totnet			0.01			
			(0.12)			
R-netnet				0.01		
				(0.32)		
R-tottot					-0.01	
					(-0.60)	
R-simple						-0.02
						(-1.11)
$\sigma_{i,t}$	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***
	(18.98)	(18.97)	(18.96)	(18.95)	(18.97)	(18.96)
Assets	-0.24***	-0.24***	-0.25***	-0.25***	-0.25***	-0.25***
	(-4.18)	(-4.22)	(-4.30)	(-4.39)	(-4.33)	(-4.34)
Capital/Assets	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***
	(-10.25)	(-10.27)	(-10.29)	(-10.26)	(-10.28)	(-10.29)
Operating profits	-0.05*	-0.05*	-0.04	-0.04*	-0.03	-0.03
	(-1.87)	(-1.88)	(-1.54)	(-1.77)	(-1.17)	(-1.09)
Cost/Income	0.10***	0.10***	0.10***	0.11***	0.11***	0.11***
	(3.75)	(3.68)	(3.65)	(3.89)	(3.84)	(3.95)
Deposits/Liabilities	0.03	0.03	0.03	0.03	0.03	0.03
	(1.43)	(1.43)	(1.46)	(1.44)	(1.42)	(1.41)
Loan loss provisions	0.19***	0.19***	0.19***	0.19***	0.19***	0.19***
	(10.20)	(10.22)	(10.20)	(9.89)	(10.18)	(10.19)
Assets growth	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*	-0.02*
	(-1.75)	(-1.73)	(-1.74)	(-1.74)	(-1.73)	(-1.72)
Z-Score	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***
	(-5.40)	(-5.40)	(-5.43)	(-5.49)	(-5.47)	(-5.50)
VIX	0.36***	0.36***	0.36***	0.36***	0.36***	0.36***
	(12.22)	(12.19)	(12.18)	(12.30)	(12.12)	(12.18)
β	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***
	(1.98)	(1.99)	(2.03)	(2.03)	(2.05)	(2.04)
Adjusted R^2	0.60	0.60	0.60	0.60	0.60	0.60
Bank fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Table 7. Predicting idiosyncratic volatility: Sub-samples.

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

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Idiosyncratic volatility is defined as the standard deviation of the residuals obtained by regressing daily bank stock returns on the three Fama and French (1993) stock return factors. We control for the lagged dependent variable (i.e., $\sigma_{i,t}$). $Entropy_{i,t}$ is the Entropy measure of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. Bad Times are periods of economic or financial sector distress and include quarters with the failure of Long-Term Capital Management (LTCM) and the Russian Crisis between the first and second quarters of 1999 and the recessions dated by National Bureau of Economic Research (NBER). Good Times are defined as all quarters not included as Bad Times. Pre-crisis is defined as the period between the third quarter of 1996 and the third quarter of 2007, and post-crisis is defined as the period between the third quarter of 2009 and the fourth quarter 2020. All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	Good times	*Bad times*	*Pre − crisis*	*Post − crisis*
Entropy	-0.06***	-0.15***	-0.08***	-0.09**
	(-3.18)	(-3.14)	(-3.46)	(-2.18)
$\sigma_{i,t}$	0.48***	0.57***	0.48***	0.44***
	(19.88)	(6.00)	(12.88)	(9.31)
Assets	-0.31***	-0.21	-0.07	-0.07
	(-5.35)	(-1.23)	(-0.76)	(-0.61)
Capital/Assets	-0.17***	-0.22***	-0.10***	-0.19***
	(-10.29)	(-4.83)	(-4.76)	(-6.28)
Operating profits	-0.02	0.09	0.05	-0.13***
	(-0.96)	(0.96)	(1.11)	(-3.11)
Cost/Income	0.08***	0.35***	0.19***	-0.04
	(3.33)	(3.07)	(3.68)	(-0.89)
Deposits/Liabilities	0.01	0.12*	0.03	0.02
	(0.35)	(1.93)	(0.83)	(0.53)
Loan loss provisions	0.14***	0.30***	0.18***	0.17***
	(8.43)	(4.91)	(7.56)	(6.30)
Assets growth	-0.02*	-0.04	0.01	-0.03
	(-1.90)	(-1.26)	(0.06)	(-1.47)
Z-Score	-0.06***	-0.13**	-0.03*	-0.05***
	(-4.92)	(-2.37)	(-1.77)	(-2.75)
VIX	0.02	0.56***	0.38***	0.32***
	(0.91)	(8.56)	(10.74)	(7.31)
β	0.02	0.02	0.04**	0.01
	(1.38)	(0.22)	(2.05)	(0.22)
Adjusted R^2	0.50	0.51	0.63	0.38
Bank fixed effects	*Yes*	*Yes*	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*	*Yes*	*Yes*

Table 8. Predicting idiosyncratic volatility: Horseraces: Alternative factor models.

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

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Idiosyncratic volatility is defined as the standard deviation of the residuals obtained either by regressing daily bank stock returns on the market risk factor (Panel A) or the five Fama-French Fama and French (2015) stock return factors (Panel B). We control for the lagged dependent variable (i.e., $\sigma_{i,t}$). $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. In addition, we control for overall market volatility (VIX) or a bank's exposure to market risk (β). All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: CAPM.			
Entropy	-0.09***	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***
	(-4.19)	(-4.18)	(-3.96)	(-4.17)	(-3.81)	(-4.03)
Hhindex	0.02					
	(1.23)					
Absdiff		0.02				
		(1.15)				
R-totnet			-0.01			
			(-0.25)			
R-netnet				0.01		
				(0.32)		
R-tottot					-0.03	
					(-1.11)	
R-simple						-0.03*
						(-1.82)
VIX	0.38***	0.38***	0.38***	0.38***	0.38***	0.38***
	(12.69)	(12.67)	(12.67)	(12.77)	(12.65)	(12.71)
β	0.05***	0.05***	0.05***	0.05***	0.05***	0.05***
	(3.23)	(3.24)	(3.29)	(3.27)	(3.31)	(3.29)
Adjusted R^2	0.62	0.62	0.62	0.62	0.62	0.62
Bank fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Other controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
			Panel B: Fama-French 5 factors.			
Entropy	-0.09***	-0.08***	-0.08***	-0.08***	-0.08***	-0.08***
	(-4.30)	(-4.32)	(-4.14)	(-4.27)	(-4.02)	(-4.19)
Hhindex	0.02					
	(1.30)					
Absdiff		0.02				
		(1.18)				
R-totnet			0.01			
			(0.15)			
R-netnet				0.01		
				(0.31)		
R-tottot					-0.01	
					(-0.56)	
R-simple						-0.02
						(-1.03)
VIX	0.36***	0.35***	0.36***	0.36***	0.36***	0.36***
	(12.54)	(12.52)	(12.50)	(12.61)	(12.44)	(12.50)
β	0.03*	0.03*	0.03*	0.03*	0.03**	0.03*
	(1.90)	(1.90)	(1.95)	(1.95)	(1.97)	(1.96)
Adjusted R^2	0.60	0.60	0.60	0.60	0.60	0.60
Bank fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Other controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*

Table 9. Predicting regression r-squared values: Horseraces.

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

$$R^2_{i,t} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Regression R^2 is defined as the R-squared value obtained from a return generating model as in Demsetz and Strahan (1997). We control for the lagged dependent variable (i.e., $R^2_{i,t}$). $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)
Entropy	0.44**	0.45**	0.47**	0.45**	0.48**	0.49**
	(2.14)	(2.22)	(2.35)	(2.26)	(2.41)	(2.43)
Hhindex	0.06					
	(0.34)					
Absdiff		0.02				
		(0.08)				
R-totnet			-0.11			
			(-0.47)			
R-netnet				0.14		
				(1.14)		
R-tottot					-0.17	
					(-0.69)	
R-simple						-0.55**
						(-2.17)
VIX	2.11***	2.11***	2.12***	2.11***	2.14***	2.21***
	(4.17)	(4.18)	(4.23)	(4.18)	(4.30)	(4.45)
β	2.47***	2.47***	2.48***	2.48***	2.48***	2.48***
	(10.42)	(10.44)	(10.42)	(10.43)	(10.41)	(10.39)
Adjusted R^2	0.57	0.57	0.57	0.57	0.57	0.57
Bank fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Time fixed effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Other controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*

Table 10. Predicting idiosyncratic volatility: 2- to 4-quarters ahead.

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

$$\sigma_{i,t+1} = \alpha_i + \beta_{i,Entropy} Entropy_{i,t} + \beta_{i,DIV} DIV_{i,t} + Controls_{i,t} + \eta_i + \gamma_t + \epsilon_{i,t}$$

Here, $\sigma_{i,t+1}$ is the idiosyncratic volatility of bank i measured at time $t+1$. Idiosyncratic volatility is defined as the standard deviation of the residuals obtained either by regressing daily bank stock returns on the market risk factor (Panel A) or the five Fama-French stock return factors (Panel B). We control for the lagged dependent variable (i.e., $\sigma_{i,t}$). $DIV_{i,t}$ is one of the seven diversification measure for bank i at time t – Entropy, HHindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple are the seven measures of diversification computed as described in Table 2.2. Controls include log assets, ratio of total capital to total book value of assets, ratio of operating profits to total assets, ratio of cost to income, ratio of deposits to total liabilities, ratio of loan loss provisions to total loans, the three-year growth in total book value of assets, and the bank's Z-Score. All right-hand-side variables are standardized by subtracting the mean and normalizing by the standard deviation of the variable. 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 and time fixed effects. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 2 quarters ahead.						
Entropy	-0.09***	-0.10***	-0.10***	-0.09***	-0.09***	-0.09***
	(-3.78)	(-3.87)	(-4.06)	(-3.99)	(-3.88)	(-3.97)
Hhindex	0.01					
	(0.01)					
Absdiff		0.01				
		(0.24)				
R-totnet			0.03			
			(0.83)			
R-netnet				-0.02		
				(-1.09)		
R-tottot					-0.01	
					(-0.27)	
R-simple						-0.01
						(-0.60)
Adjusted R^2	0.56	0.56	0.56	0.56	0.56	0.56
Panel B: 3-quarters ahead.						
Entropy	-0.11***	-0.11***	-0.11***	-0.11***	-0.11***	-0.11***
	(-3.93)	(-3.95)	(-4.17)	(-4.24)	(-4.09)	(-4.29)
Hhindex	-0.01					
	(-0.37)					
Absdiff		-0.01				
		(-0.44)				
R-totnet			-0.01			
			(-0.06)			
R-netnet				-0.01		
				(-0.57)		
R-tottot					-0.02	
					(-0.61)	
R-simple						0.01
						(0.17)
Adjusted R^2	0.51	0.51	0.51	0.51	0.51	0.51
Panel C: 4-quarters ahead.						
Entropy	-0.10***	-0.10***	-0.09***	-0.09***	-0.09***	-0.09***
	(-3.51)	(-3.38)	(-3.45)	(-3.46)	(-3.37)	(-3.48)
Hhindex	0.02					
	(0.93)					
Absdiff		0.01				
		(0.48)				
R-totnet			0.01			
			(0.27)			
R-netnet				0.01		
				(0.51)		
R-tottot					-0.01	
					(-0.16)	
R-simple						0.01
						(0.38)
Adjusted R^2	0.49	0.49	0.49	0.49	0.49	0.49

Table 11. Predicting idiosyncratic volatility: Out of sample forecasts.

Notes: This Table reports the out of sample root-mean-squared errors (RMSE) for predicting bank idiosyncratic volatility. The row headers indicate the horizon (either 3- or 5-year windows) used to estimate the model. The column headers – Entropy, Hhindex, Absdiff, R-totnet, R-netnet, R-tottot, and R-simple indicate the variable used to predict idiosyncratic volatility. Panel A reports the results for the model that only uses the diversification measure without control variables and Panel B reports the results for the model that uses diversification measure along with other control variables to predict idiosyncratic volatility. The predicted values from the model for idiosyncratic volatility are compared to the realized values to compute the RMSE. Values in the table are RMSEs that are multiplied by 100 and expressed in percentages.

$H =$	Entropy	Hhindex	Absdiff	R-totnet	R-netnet	R-tottot	R-simple
Panel A: Without control variables							
3-year window	1.4011	1.4069	1.4023	1.4058	1.4069	1.4057	1.4069
5-year window	1.3160	1.3352	1.3355	1.3340	1.3277	1.3342	1.3349
Panel B: With control variables							
3-year window	1.6795	1.7707	1.7736	1.7723	1.7678	1.7616	1.7665
5-year window	1.7627	1.8542	1.8583	1.8572	1.8478	1.8357	1.8467

Table 12. Determinants of Entropy.

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

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

$Entropy_{i,t}$ is the Entropy measure of diversification computed as described in Table 2.2. $Determinants_{i,t}$ includes bank-level characteristics – Assets, Capital to Assets, Operating profits, Asset growth, and Z-Score. 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. Quarterly data, 1996Q3 – 2020Q4.

	(1)	(2)	(3)	(4)	(5)	(6)
Assets	0.084***					0.084***
	(10.55)					(10.75)
Capital/Assets		-0.045***				-0.055***
		(−8.83)				(-9.75)
Return on assets			0.016			0.027***
			(1.60)			(5.41)
Assets growth				-0.008		-0.026*
				(-0.98)		(-1.87)
Z-Score					0.033***	0.038***
					(7.24)	(8.22)
Adjusted R^2	0.058	0.017	0.002	0.001	0.009	0.094

Appendix
Which "MEASURE" of Bank Diversification Measure Up?

A Definitions and construction of variables

We collect balance sheet data for banks from the 'Report for Condition and Income' (henceforth Call Reports) required to be filed by all FDIC-insured bank holding companies in the U.S. This data is available at https://www.chicagofed.org/applications/bhc_data/bhcdata_index.cfm. Definitions for the variables are available at http://www.federalreserve.gov/apps/mdrm/. Banks with total book value of assets above $500 million file this report quarterly. Other banks file this report only semi-annually. 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 us with 75,020 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 23,785 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 Call Reports 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 Call Report. 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 2. 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 call report data available for at least 12 previous quarters (3 years). We also require that for each quarter Call Report data for a particular bank is available for the previous and current quarters. This requirement ensures that the time-series of core, non-core, and trading incomes 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 data for banks from the 'Report of Condition and Income (Call Report)' 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	
	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	
Noninterest income	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. Appendix - 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
Entropy	Entropy Index of nine noninterest income and seven interest income items. FR Y-9C
Hhindex	1 - Herfindahl-Hirschman Index of noninterest income and net interest income. FR Y-9C
Absdiff	1 - the absolute value of the ratio of net interest income minus noninterest income to the sum of net interest income and noninterest income. FR Y-9C
R-totnet	Ratio of noninterest income to the sum of noninterest income and net interest income. FR Y-9C
R-netnet	Ratio of net noninterest income to the sum of net noninterest income and net interest income. FR Y-9C
R-tottot	Ratio of noninterest income to the sum of noninterest income and interest income. FR Y-9C
R-simple	Ratio of noninterest income to interest income. FR Y-9C
Idiosyncratic volatility	Standard deviation of the residuals obtained by regressing daily bank stock returns on Fama-French three factors. CRSP, Kenneth French's Data Library
Assets	Natural logarithm of total assets. FR Y-9C
Capital/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/Income	Ratio of the sum of noninterest and interest expense to the sum of noninterest and interst income. FR Y-9C
Deposits/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, 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. FR Y-9C
Tobin's Q	Ratio of the sum of the market value of common equity, the book value of total liabilities, and the book value of preferred stocks to the book value of total assets. CRSP, FR Y-9C