Predicting Bank Failures per Quarter in the United States

Kay Duong, Angel Miranda-Acosta, Carrie Hashimoto, Annie Xu

2023-04-21

Statement of the Problem

In March 2023, the United States experienced the second-largest bank failure in the nation's history. The Silicon Valley Bank Failure disrupted many financial services such as lending and investment banking. Due to the increasing relevance of bank failures, our goal for this project is to create a multiple linear regression model that can be used to predict the number of American federally insured banks that will fail in a fiscal quarter based on various economic indicators.

1 Introduction

A bank failure occurs when a bank cannot meet its financial responsibilities to depositors and creditors and is influenced by various factors such as corporate governance and financial conditions. Banks have a particular risk for failure relative to other types of firms because they tend to be very highly leveraged, with ratios between debt and equity of 30:1 or higher (Mohan 1). Their high level of debt provides tax benefits, but if the value of loans given by a bank decreases a relatively small amount, it can trigger fear in depositors. If they begin to worry about bank stability, they will attempt to withdraw deposits, and the bank will need to sell loans to acquire cash; however, since deposits are more liquid than loans, getting this liquidity to meet depositor demands can be challenging, and the bank may have to sell loans at a markdown, leading to further losses. This precarious system can result in bankruptcy and even bring down other banks due to the interconnectedness of the banking system (Mohan 1).

Various views on bank failure's causes

Various views on the causes of banking crises exist, including a "sunspot" perspective, a view that places the blame on poor bank management and a flawed system, and a view that banking crises emerge out of the macroeconomic environment. Under the "sunspot" view, the precariousness of the system as detailed above implies that banks are never far from failure and a small change in circumstances can trigger a crisis (Mohan 8). Research on banking management has indicated that certain indicators correlate with better banking stability; for example, wealth concentration in a financial institution has been shown to positively affect risk management, and having a credit risk officer on a bank's executive board can also improve its outcome (Polyzos et al, 98). CEO ownership and institutional ownership tend to have better outcomes than public ownership, which is associated with lower risk and lower profitability.

Bank failure and the macroeconomy

Empirically, a troubled macroeconomy does not always result in banking crises. For example, although the United States and Canada both experienced the Great Depression, only one Canadian bank failed during that event, compared to 15,000 U.S. banks (Mohan 9). Furthermore, correlation between bank failures and firm failures from 1875 to 1933 was only 0.24, and a study of countries that experienced banking crises between 1970 and 1988 found that only 20 to 30 percent of them had abnormal GDP growth in the three years before the crisis (Mohan 9). Academic views differ on whether banking crises lead to economic contraction or the opposite.

A federally-endorsed system for determining commercial banks' soundness is the Uniform Financial Rating System introduced in 1979 (Cole 5). This system evaluates banks' capital, assets, management, earnings, liquidity, and sensitivity to market risk on a scale of 1 to 5 and aggregates the values to get a proxy for bank soundness. Such metrics provided this project with a starting point for potential regressors, such as net loans and leases and liquidity ratio.

The consequences of bank crises are varied and include problems like the direct cost to the federal government in providing assistance, a loss in output reflected in reduced GDP growth, and an increase in public debt (Mohan 14). On a more individual level, small business owners with good credit histories tend to fare better after bank failures, which can exacerbate inequalities, and local philanthropy tends to decline after small banks fail (GAO 3).

2 Data

2.1 Data Collection

Our primary source of data is the Federal Deposit Insurance Corporation's (FDIC) *Quarterly Banking Profile*. This data source provides a comprehensive summary of financial results for all FDIC-insured institutions by fiscal quarter. The data spans from 1984 to 2023, but due to missing values, we had to subset the data to only include 1987 to 2022. There are more than fifty variables in the dataset to select from. Using knowledge we gained from the background research, we selected four variables from the dataset: Liquidity Ratio, Net loans and Leases to Total Deposits, Number of Unprofitable Institutions, and Income Before Taxes and Extraordinary Items. Liquidity Ratio describes an institution's ability to pay off short term debt obligations for all insured institutions. The Net Loans and Leases variable is calculated by adding loans and lease-financing receivables net of unearned income and the allowance for possible loans and lease financing receivable losses divided by total assets. This variable is another measure of liquidity. Unprofitable Institutions is a count of unprofitable institutions in the fiscal quarter. Income before taxes is income before taxes and extraordinary items, the gains or losses from a disclosed unusual event.

We used an additional data source to find S&P 500, an index that measures the stock values of the 500 largest corporations by market capitalization listed on the New York Stock Exchange or Nasdaq. Additionally, we included selected market interest rates published by the Federal Reserve System by fiscal quarter, known as the Interest Rates variable.

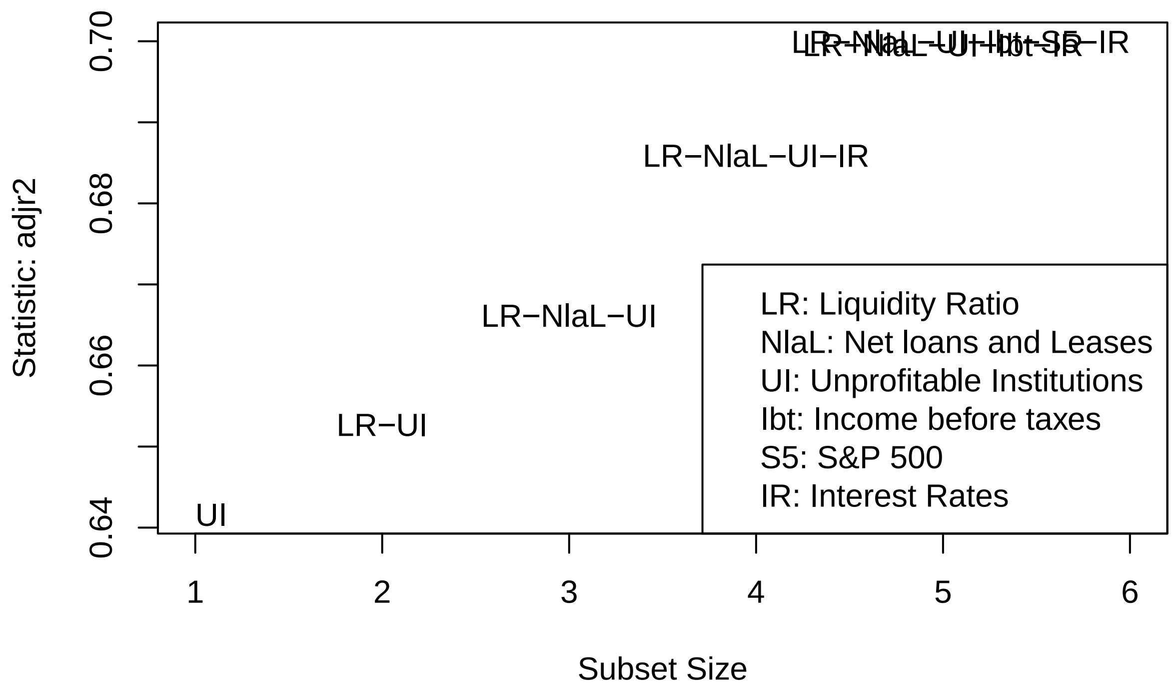
We collected and inputted the above variables onto a spreadsheet. We encountered an issue while obtaining the data for our response variable, number of failed institutions. The *Quarterly Bank Profile* does not have aggregated data on the number of failed institutions by fiscal quarter. Therefore, we obtained the data by searching through each *Quarterly Bank Profile* PDF report and subtracting the annual number of failed institutions from the previous quarter report to get the quarterly number of failed institutions.

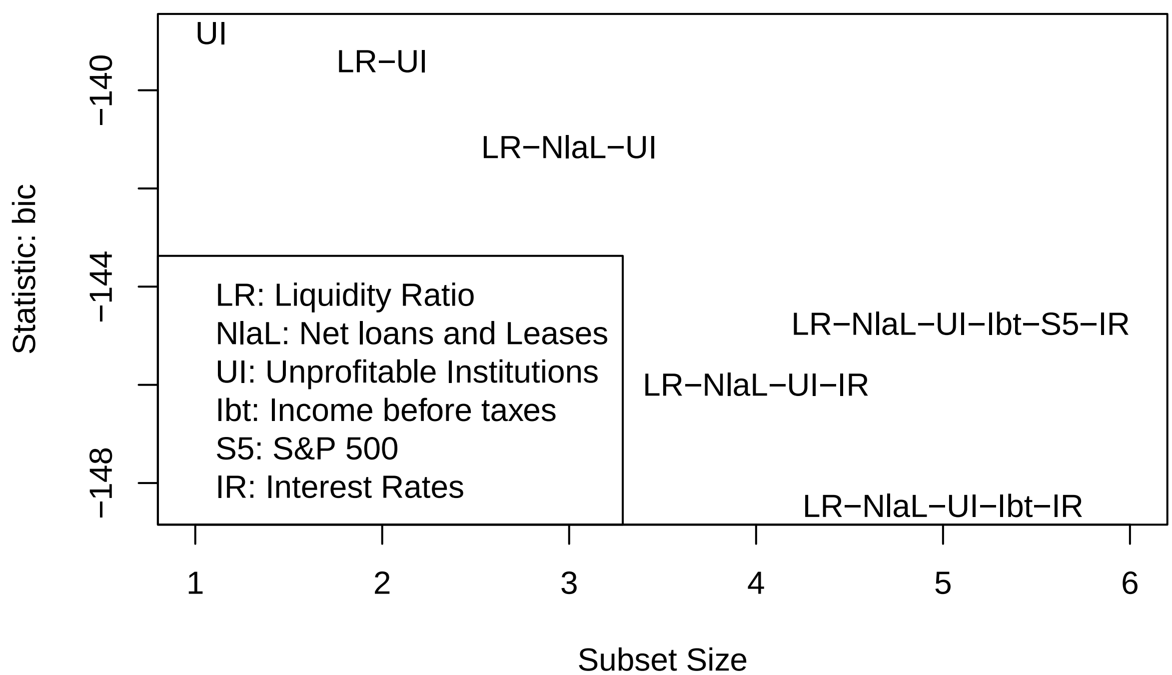
2.2 Data Cleaning

Once we loaded the spreadsheet into R, we subsetted the data to include the fiscal quarters of 1987 through 2022. Some data cleaning was necessary: R identified the variables as a character type due to the commas, so we used a function to take out any commas and convert them into a numeric type. Additionally, we adjusted three of the variables in the data. We converted the S&P 500 index to the change in percentage of the S&P 500 index by fiscal quarter because we felt that the response might be more sensitive to changes in the feature rather than its value. We transformed the liquidity ratio into a quadratic to better fit the data, having observed its relationship to the response using the pairs function. Finally, we also offset the number of unprofitable institutions by one so the previous number predicts the next quarter's unprofitable institutions.

3 Data Analysis and Modeling Results

3.1 Data Analysis





We used best subset selection to reach our final model. While forward or backward selection or other methods may have performed well on data with many possible regressors, we were working with only 6 possible regressors which yields a relatively manageable number of calculations using the best subset method.

We chose a model of size 5 based on plots of BIC and adjusted R-squared against Subset Size. Whereas the 5 and 6 regressor models have similar adjusted R-squared of about 0.7, the BIC plot shows that the 5 regressor model has a lower BIC of less than -1.48 compared to a 6-regressor model with BIC around -1.44. The 1, 2, 3, and 4-regressor models also perform worse than the 5 regressor model, with lower adjusted R-squared and higher BIC. Therefore, we decided to use the model of size 5 to explain our data.

By conducting t-tests on all 6 regressors and examining the R-squared value for our model, we further confirmed that our 5 regressor model might be a good fit for our data. t-tests yielded significant p-values of far less than 0.05 for each of our 5 regressors while yielding a not-significant p-value of 0.27 for S&P 500, the regressor that we exclude from our model. The adjusted R-squared value of 0.6695 and multiple R-squared of 0.7101 for our model are both high enough that we are willing to conclude sufficient fit by our model for our data.

Additionally, the confidence intervals for our 5 and 6-regressor models do not vary widely for difference choices of variables included in the model. This suggests that the general structure of our model works well for our data. If we had observed major differences in the confidence intervals - say, a difference between the two models in order of magnitude for a regressor or a difference in sign - we may have questioned whether our linear regression model had the best form to describe our data. Instead, we see that both the 5 and 6-regressor models produce comparable confidence intervals for each regressor, helping us to conclude that our model is in an acceptable form.

Confidence Intervals for the 5-Regressor Model

```
##                                    2.5 %        97.5 %
## (Intercept)                  61.424450724  1.489806e+02
## ‘Liquidity Ratio‘             0.014665650  5.151701e-02
## ‘Net loans and Leases‘       -2.701403573 -1.172852e+00
## ‘Unprofitable Institutions‘   0.006864067  1.276025e-02
## ‘Income before taxes‘        -0.000235599 -3.762584e-05
## ‘Interest Rates‘              0.813314702  2.906175e+00
```

Confidence Intervals for the 6-Regressor Model

```
##                                     2.5 %        97.5 %
## (Intercept)                  61.6939402480  1.491915e+02
## ‘Liquidity Ratio‘             0.0159780554  5.318846e-02
## ‘Net loans and Leases‘       -2.7043352028 -1.176831e+00
## ‘Unprofitable Institutions‘   0.0066729350  1.259864e-02
## ‘Income before taxes‘        -0.0002430884 -4.375939e-05
## ‘S&P 500‘                    -9.9268618869  3.504497e+01
## ‘Interest Rates‘              0.7812361495  2.875600e+00
```

One question that we considered in the course of our data analysis was whether to convert the Number of Unprofitable Institutions into a percentage (Number of Unprofitable Institutions / Total Number of Reporting Institutions * 100%) and similarly convert Number of Bank Failures into a percentage out of the total number of reporting institutions. The question arose when we considered that the Number of Bank Failures might correlate with the total number of federally insured institutions. We observed that the total number of banks in the first decade of our data (Q1 of 1986 to Q4 of 1995) exceeded 11,000 banks every quarter, whereas in the last decade (Q1 of 2013 to Q4 of 2022) fewer than 8,000 banks reported every quarter.

The number of bank failures also seemed to be much larger in the first decade of our data than in the last, with a mean of over 29 bank failures per quarter for 1986-1995 and a mean of only 2 bank failures per quarter for 2013-2022. Based on these numbers, we were motivated to investigate whether percentages would be better regressors than absolute numbers.

However, once we adjusted our regression so that our regressors included percentage of unprofitable institutions instead of absolute value of unprofitable institutions and, in turn, adjusting our response variable to be a percentage of bank failures rather than an absolute number of failures, we found that our R-squared decreased for our 5-regressor model and that the p-value for the unprofitable institutions regressor also decreased. These results suggest that, in fact, the absolute number of unprofitable institutions may be a better regressor than the percentage of unprofitable institutions. Based on this analysis, we decided to keep our original model using the absolute number of unprofitable institutions.

3.2 Modeling Results

The final model we decided to use to predict the number of bank failures in a given financial quarter is a Multiple Linear Regression model with 5 regressors and an intercept. The 5 regressors are Liquidity Ratio, Net Loans and Leases, Number of Unprofitable Institutions, Income before Taxes, and Interest Rates. The regressor with the best p-value and perhaps the tightest correlation with the number of bank failures was the Number of Unprofitable Institutions, with a p-value of 0.00000000121.

$$\begin{aligned} Number\ of\ Bank\ Failures = {} & 105.2 + 0.03309 * (40 - Liquidity\ Ratio)^2 \\ & - 1.937 * (Net\ Loans\ and\ Leases) \\ & + 0.009812 * (Number\ of\ Unprofitable\ Institutions) \\ & - 1.366 * 10^{-4} * (Income\ before\ taxes) \\ & + 1.86 * (Interest\ Rate) \end{aligned}$$

The model explains our data with a 71% multiple R-squared.

4 Summary & Discussion

4.1 Regression Analysis and Interpretation

We selected 5 out of the 6 original regressors in our model, with the excluded regressor being percent change in the S&P 500.

With liquidity ratio, for each squared distance from 40% liquidity ratio, we expect an additional 0.033 bank failures. The squared distance was visually chosen from the pairs call to the data. A financial explanation to this could be that if banks did not have enough liquidity, they would not be able to match deposits, but if they had too much, they were not operating well as a business. The cause of this relationship would need to be investigated further.

Net loans and leases had a coefficient of -1.937. For each additional million dollars of net loans and leases, we expect 1.9 fewer bank failures. At a first glance of the pairs call, this regressor seems to have a positive relation with failures, but the failures with high loans are likely to have been caught and represented more with other regressors, such as liquidity ratio and income.

The calculated coefficient for number of unprofitable institutions is 0.0098. Every 100 more unprofitable institutions results in an expected increase in the number of bank failures by 1. If all failed banks came from last quarter's unprofitable banks, then around 1% of unprofitable banks become failed; however, we know this is not the case, with the most notable example being SVB.

Income before taxes has a $-1.366 * 10^{-4}$ coefficient, meaning every additional million in income reduces the number of failed banks by $-1.366 * 10^{-4}$. Higher income means banks are less likely to fail.

The last regressor, interest rate, has a coefficient of 1.86. For each 1 point increase in interest rate, there is an expected increase in bank failures by around 2. This effect could result from a more competitive environment or because of an overall economic downturn.

The excluded regressor, the percent change in the S&P 500, could have been explained better by other variables, or more likely that there is simply no correlation between the two.

4.2 Model Analysis

The regressors explain the number of bank failures fairly well. The R-squared value for our model is 0.71, meaning that these 5 regressors explain 71% of the variation in the number of bank failures. All five of the regressors are significant and were chosen with best subset selection. The percent change in the S&P 500 was not a significant regressor and additionally was excluded from the best subset.

The relationship between the regressors and the number of failed banks does not mean that there is causation between them, nor that one affects the other. The financial system is a highly complicated and interconnected system that does not always have a 'this changes that'. It is hard to tell from this model what factors drive change and which are simply indicators of the state of the financial system. For example, income and bank failures could both be symptoms of a wider economic downturn rather than income dictating failures. Our model does not account for that. Additionally, interest rates are set by policy. This could be an indicator of the current health of the economy, the future health, or the direction policymakers want the economy to be in. This might have an effect on bank failures directly, or it might affect the economy which affects bank failures.

4.3 Different Models

After looking at how effective shifting the number of unprofitable banks was at predicting the number of bank failures, one change we wanted to look at was shifting all of the data back one quarter. In effect, the previous quarter's numbers would predict the current quarter's bank failures. For this model, all regressors were significant. The best subset using adjusted R-squared and BIC was the complete model as well. This is quite different from the model that used the current quarter's data, which had only 4 regressors be significant. The BIC and adjusted R-squared for the new, shifted model was slightly better and has an R-squared of 0.74, 0.04 more than the model that uses current numbers. This model is also more interpretable and usable, since it creates a prediction for the next quarter, as opposed to an explanation of the current quarter. The regressor values are likely to be released in the same report at the same time as the number of bank failures, so getting a prediction beforehand could prove useful.

The resulting model was

$$\begin{aligned} Number\ of\ Bank\ Failures = {} & 76.57 + 0.03932 * (40 - Liquidity\ Ratio)^2 \\ & - 1.801 * (Net\ Loans\ and\ Leases) \\ & + 0.01037 * (Number\ of\ Unprofitable\ Institutions) \\ & - 1.319 * 10^{-4} * (Income\ before\ taxes) \\ & + 39.17 * (S\&P\ 500) \\ & + 1.456 * (Interest\ Rate) \end{aligned}$$

Most of the regressors retained values similar to what they previously had, except for the S&P 500, which is now significant and selected. What is interesting is the positive relationship between percent change in S&P 500 and number of bank failures. This could be because of an actual interplay between a booming economy and bank failure; it could be due to a lag between S&P 500 and bank failures; or it could be due to heavy hitting bank failures.

4.4 Model Improvements and Expansions

Our model does not explain or account for every possible regressor nor the relationships between other factors. One possible improvement to this model would be to have more regressors that attempt to explain the number of bank failures to try to account for the complexity of the financial system. We did find data on several different regressors but were limited in obtaining some data, since they were available in a PDF and not in csv form. Additionally, our data was restricted to quarters between 1987 and 2022. We would have liked for our data to go further back in time, and also be more granular to create a better model.

Another way of accounting for the complexity of the financial system would be to understand the system itself more. If we better understood the intricacies of the system, then we could attempt to fit a better model with more predictive power. Of course, we do not know much about the system, but if we did we could create a better model.

One change that we did to the Number of Unprofitable Institutions was to shift that data back a quarter, such that the previous quarter's unprofitable institutions would try to explain this quarter's failures. Intuitively, this makes sense because of the time relations of each quarter. We found that our model improved because of the shift. An additional change to the model could be to create more regressors that shift quarters back different amounts. We could also attempt to create a time series to try and explain the relationship of the data with time.

Our model focuses on the overall picture of the number of bank failures, but does not look at which banks are likely to fail. Nor does the model assess the severity of any given failure, which is what caused this exploration to begin in the first place with the SVB failure. An expansion to the project could be to get data on individual bank health to try and determine what might cause a bank to fail, or if it failed how severe the failure would be.

5 Appendix

5.1 R Code

```
#Load the data into R.
data <- read.csv("~/Desktop/STAT 410/bank_failure_data.csv")

#Take out the comma and turn the variable S&P 500 into a numeric type.
sp <-  as.numeric(gsub(",", "", data$S.P.500))[13:156]
new_sp <- rep(0, 13) #Add zeroes for missing data rows.
# Convert the  S&P 500 index to the change in percentage of the S&P 500 index
# by fiscal quarter.
for(i in seq.int(1, length(sp) - 1)){
  new_sp <- c(new_sp, (sp[i + 1] - sp[i])/sp[i])
}

#Get design matrix.
#Change the liquidity ratio into a quadratic.
x <- cbind((40 - as.numeric(
  data$Liquidity.ratio...All.Insured.Institutions.))^2,
        as.numeric(gsub(",", "",
        data$Net.loans...Leases.to.Total.Deposits..All.Insured.Institutions.)),
        c(as.numeric(gsub(",", "", data$Number.of.unprofitable.institutions))[1],
        as.numeric(gsub(",", "",
        data$Number.of.unprofitable.institutions))[1:155]),
        as.numeric(gsub(",", "",
```

```
        data$Income..loss..before.income.taxes.and.extraordinary.items)),
        new_sp,
        as.numeric(gsub(",", "", data$Interest.rates)))

#Subset the data to the years with no missing data and including the response variable.
new_data <- data.frame(cbind(
            as.numeric(data$Number.of.failed.assisted.banks),x))[13:156,]
#Add the variable names for readability.
names(new_data) <- c("Banks failed",
                    "Liquidity Ratio",
                    "Net loans and Leases",
                    "Unprofitable Institutions",
                    "Income before taxes",
                    "S&P 500",
                    "Interest Rates")
#Get the model.
model <- lm(new_data)
#summary(model)

#Perform best subset selection.
x <- data.frame(x)
names(x) <- c("Liquidity Ratio",
            "Net loans and Leases",
            "Unprofitable Institutions",
            "Income before taxes",
            "S&P 500",
            "Interest Rates")
library(car)
library(leaps)
b <- regsubsets(as.matrix(x[13:156,]),
     as.numeric(data$Number.of.failed.assisted.banks[13:156]))

#Search for the optimal subset size.
#Adjusted R Squared
#subsets(b, statistic = c("adjr2"), legend = "bottomright")
#BIC
#subsets(b, statistic = c("bic"), legend = "bottomleft")

#Selected Subset
newx <- cbind((40 - as.numeric(
        data$Liquidity.ratio...All.Insured.Institutions.))^2,
        as.numeric(gsub(",", "",
        data$Net.loans...Leases.to.Total.Deposits..All.Insured.Institutions.)),
        c(as.numeric(gsub(",", "",
        data$Number.of.unprofitable.institutions))[1],
        as.numeric(gsub(",", "",
        data$Number.of.unprofitable.institutions))[1:155]),
        as.numeric(gsub(",", "",
        data$Income..loss..before.income.taxes.and.extraordinary.items)),
        as.numeric(gsub(",", "", data$Interest.rates)))

#Get the new model.
newnew_data <- data.frame(cbind(
```

```r
  as.numeric(data$Number.of.failed.assisted.banks), newx))[13:156,]
names(newnew_data) <- c("Banks failed",
                        "Liquidity Ratio",
                        "Net loans and Leases",
                        "Unprofitable Institutions",
                        "Income before taxes",
                        "Interest Rates")
newmodel <- lm(newnew_data)
#summary(newmodel)

#Confidence Intervals
fiveregressorconf <- confint.default(newmodel)
sixregressorconf <- confint.default(model)
```

5.2 Acknowledgements

Board of governors of the Federal Reserve System. Federal Reserve Board - H15 - Data Download Program - Build. (n.d.). https://www.federalreserve.gov/datadownload/Build.aspx?rel=H15

Cole, R. A., & White, L. J. (2012). Deja vu all over again: The causes of U.S. commercial bank failures this time around. Journal of Financial Services Research, 42(1-2), 5-29. doi:https://doi.org/10.1007/s10693-011-0116-9

Mohan, T. T. R. (2017). Towards a safer world of banking : Bank regulation after the subprime crisis. Business Expert Press.

Stathis Polyzos, Khadija Abdulrahman and Apostolos Christopoulos (2018). Good management or good finances? An agent-based study on the causes of bank failure. Banks and Bank Systems, 13(3), 95-105.

S&P 500 historical rates (SPX). Investing.com. (n.d.). https://www.investing.com/indices/us-spx-500-historical-data

Government Accountability Office. Financial Institutions: Causes and Consequences of Recent Bank Failures (2013). 1-3.

Quarterly banking profile. FDIC. (n.d.). https://www.fdic.gov/analysis/quarterly-banking-profile/