

# Rotten Banks: Predicting Bank Failures After Great Recession with Binary Classification

Maxim Ananyev

November 2018

I investigate the determinants of bank failures after the financial crisis of the years 2007 - 2009 to build a predictive model of bank failures. I use two paradigms for prediction: accuracy maximization and Neyman-Pearson framework. Accuracy-maximization implies that Type I errors and Type II errors are equally costly, thus out-of-sample predictive accuracy is the most important parameter for evaluation. Neyman-Pearson paradigm implies setting an upper bound for Type I errors and minimizing Type II errors within that bound. In this case, the costs associated with Type I and Type II errors can be different. I find that, because the bank failures are rare events, many of the accuracy-maximizing classifiers tend to assign all the observations to the class of non-failing banks. This achieves out-of-sample predictive accuracy of 96 percent, but misses all the failures. Two algorithms, post-Lasso logit and random forest tend to have relatively low level of Type II errors. The classification with the Neyman-Pearson paradigm performs better in terms of minimizing Type II errors while containing Type I errors. All of the algorithms, in out-of-sample testing, were able to identify at least 50 percent of the failing banks, while having false positive rate below ten percent. The minimum share of Type II errors were displayed by Ada-Boost algorithm (24 percent), while GLM with LASSO penalty and sparse LDA did not perform much worse (the level of Type II errors was 27 percent).

## 1. INTRODUCTION

Financial institutions play many important roles in the economy: they manage savings, fund entrepreneurial projects, provide mortgages and consumer credit, support the system of payments, and diversify risk through financial innovation thus contributing to economic growth (King and Levine (1993), Rajan and Zingales (1998), Bekaert et al. (2005)). When the financial industry works well, it brings

a lot of benefits to the society. Unfortunately, when financial industry does not work well, every sector of the economy suffers. As was exemplified in the financial crisis of 2007-08, according to one of the common explanations, just a handful of banks overexposed to a default risk of the securities in their portfolio (as a result of a search for higher returns when monetary policy keeps interest rates low) can trigger a sudden dry up of liquidity leading to a chain reaction of a self-enforcing financial panic that inflicts lasting damage on the economy (see Campello et al. (2010), Helleiner (2011), Berger and Bouwman (2013), Bernanke (2013) and Gorton and Ordonez (2014) among many others on this topic).

This danger is exacerbated by the difficulties in the accurate valuation of the assets on balance sheets of banks. Two most common approaches to valuation – mark-to-market and historical cost can be misleading. Mark-to-market (or fair value) approach implies that a value of an asset equals its market price. This approach can be unreliable because the market price is volatile, so that a mark-to-market values can change rapidly thus providing little insight to the economic situation of the bank (Bernard et al. (1995), Allen and Carletti (2008), and Heaton et al. (2010) are among the authoritative discussions of this issue). Another issue is that for many securities originated on the thin markets, market price might not be available. Their valuation according to fair value principles requires strong assumptions.

Historical-cost valuations may not be as volatile, but they are often not relevant for the current economic conditions. For example, if a rating agency downgrades certain asset-backed securities (as happened during 2007-08 financial crisis), this change reflects on their market price, but not on their historical-cost valuations, and a balance sheet of an institution that carries those securities would appear healthier than it should be (Ellul et al. (2015) has a discussion of this and other issues related to historical-cost accounting). This problem is especially acute given that prudential regulation - for example, requiring banks to have a certain fraction of its assets in cash or highly liquid securities has not been successful in reducing the risks of bank failures (Calomiris and Jaremski (2016)).

Overexposure to default risks and incorrect valuation of assets do not exhaust a list of potential problems a financial institution might suffer from. Many banks, especially in the developing nations, often engage in activities that are different from those allowed by the regulators. Such activities often involve moving cash from their host jurisdiction to another, usually more secretive, jurisdiction. In the extreme form, this activity amounts to money laundering (Reuter (2004)). The banks who engage in such transactions are dangerous for the economy because they facilitate illegal transfer of assets, tax avoidance, and, potentially, corruption and organized crime. Detecting insolvent and outright fraudulent financial institutions is important for the regulators and investors because of a grave danger such institutions can pose to the society.

In this project, I attempt to build a predictive model of bank failure using financial statements from banks. I use a proprietary dataset to collect financial statements from the sample of US banks on their performance in the year 2008. I then collect information from the U.S. Federal Deposit Insurance Corporation about which banks failed *after* the year 2008. I use not only the items from the financial statements themselves, but also various performance ratios common in the economics and finance literature.

With those features, I deploy a set of classification algorithms, including post-LASSO logistic regression, Support Vector Machine Classification, and Random Forest to build a predictive model of bank failures. I evaluate predictive accuracy of those models and predictive accuracy of an algorithm based on stacking of the previous models. I find that Neyman-Pearson paradigm of classification that minimizes Type II error rate while enforcing an upper bound for Type I errors delivers promising results by being able to predict around 70 percent of actual failures while also being able to hold false alarms below 10 percent.

In this context, two aspects of this exercise might be useful. First, predictive accuracy of the models ( $\hat{y}$  problem) and other characteristics of the classifier (like Type 1 error, Type 2 error, false discovery rate, and others). This information

enables creating an early warning systems of bank failures that might be of practical importance. Second, the marginal contributions of each of the factors into the probability of failure (beta problem) is also of theoretical interest.

Predicting corporate bankruptcy is an important topic within the literature on management and expert systems (Zmijewski (1984), Lau (1987), Altman et al. (1994) are among the earlier studies that influenced the ever-growing field since then. Sun et al. (2014) provides a review) Unlike those studies, this project focuses solely on banks, since banks have different business model and different accounting standards. For example, the size of the cashflow have different economic interpretation for financial and non-financial firms. Financial firms also have an additional set of items they need to report that can improve the quality of the prediction.

Another important strand of literature is a literature on earnings management (Jones (1991), Dechow et al. (1995), Burgstahler and Dichev (1997), Durnev and Guriev (2007)). This literature develops a set of indicators that allow detecting if the management of a firm is systematically using accounting rules to change the amount or reported earnings. The simplest of such indicators is a difference between earnings and cashflow. Earnings should be recorded when a service is performed (or a merchandise is shipped), but cashflow should be recorded when the cash changes hands. Depending on the nature of the business, the difference between earnings and a cashflow can be innocuous, but it might also mean that the management is either over-reporting or under-reporting the profits of a firm.

In sum, this thesis attempts to build a system of early warning of bank failure, using advances predictive modeling and finance. One of the contributions of this work is that, unlike some of the previous work on financial expert systems, my main dependent variable is not a measure of financial troubles extracted from the financial reports (that can be manipulated by banks themselves) but independently recorded event of bank failure. I view my contributions as complementary to the earlier work on financial statistical expert systems.

My findings can potentially be used by regulators to detect problems in the

banking system, investors to decide where to allocate their funds, CEOs and directors of banks to assess the quality of the bank's management, auditors to use as one of the indicators for the soundness of financial reporting of firms, and social scientists to use as a dependent or independent variable in their studies.

This thesis proceeds as follows. Section 2 discusses the sources of data and presents some descriptive statistics. Section 3 presents and discusses additional features (performance ratios) that I am using as covariates. Section 4 presents and evaluates predictive models for bank failures that use flexible data-driven algorithms. Section 5 concludes.

## 2. DATA AND DESCRIPTIVE STATISTICS

The data come from two sources. I use harmonized financial statements of banks from database *Orbis* compiled by Bureau van Dijk. This is a firm-level proprietary database that contains a set of items extracted from the regulatory filings of firms and other sources. I download all the items from firms that have NACE code 6418 ("financial institutions – other financial intermediation") that have industry category "Banks" and that are registered in the United States. This search categories yield a representative sample of the US retail banks. The information available in the database includes balance sheet items (such as total assets, total liabilities, loans, deposits etc.) and the items from income statement (net income, costs, taxes etc.).

Unfortunately, not all banks have all the financial information – in fact, many banks do not have any. After removing the banks for which the majority of items contain missing data, I end up with 6614 banks in my dataset. Then I remove variables that have missing values for more than 10 percent of the remaining observations. This leads to exclusion of five variables. Then I remove all the observations with missing data, and I end up with 6471 observations. Alternative approach for dealing with missing data would be to impute the missing values that would preserve variance-covariance structure of the missing data. I leave the inference based

on multiple imputations for further research.

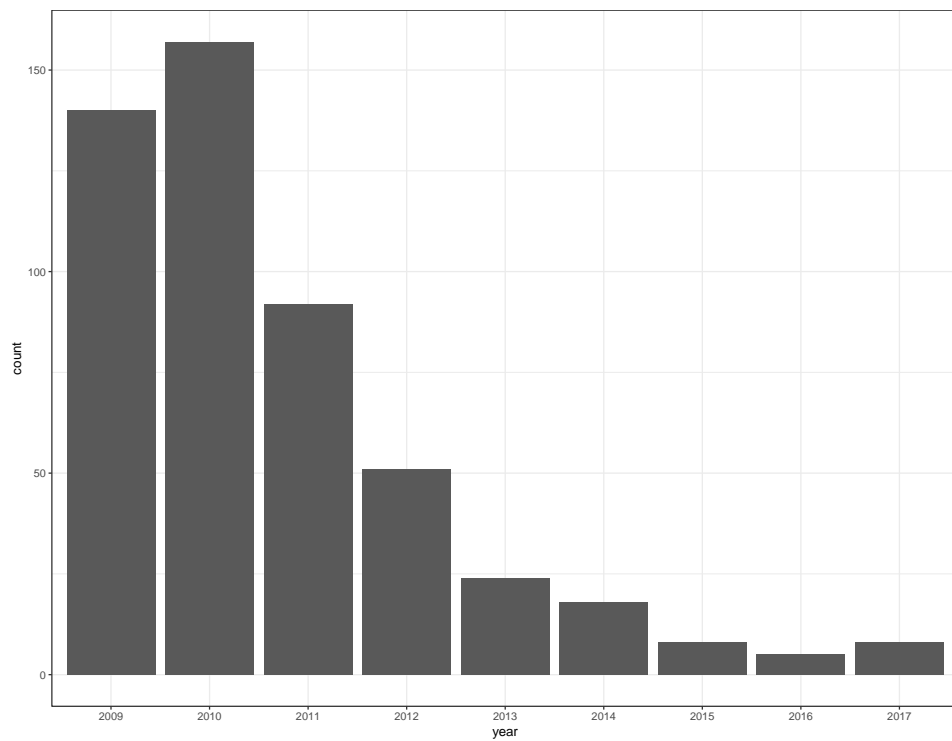
The second source of data is Federal Deposit Insurance Corporation (FDIC). FDIC is a governmental agency that oversees retail banks and processes the events of bank failures. Once management of a distressed bank realizes that the bank cannot meet its liabilities, a request should be filed within FDIC. The banks very rarely undergo a formal bankruptcy procedure which is required for other firms (filing Chapter 10 or Chapter 11). Usually, FDIC steps in to guarantee the deposits and facilitate finding another bank, which, with the help of the government, assumes the liabilities of a distressed bank. This event is categorized as “bank failure” by FDIC. FDIC maintains a list of all bank failures. For these events, they record a name of the distressed bank, a name of the acquiring bank and the date. For years 2009 - 2018, FDIC records 503 failure events. Figure 1 shows the number of bank failures per year. The purpose my predictive models is to predict and provide some qualitative insights into those failures using only financial data from the year 2008.

To construct my data set, I merge financial data from Orbis with the failure events data from FDIC. I end up with 6473 observations. Out of those observations, 203 observations are the banks that eventually failed. The goal of this analysis is to estimate an array of flexible models that allow predicting the eventual bank failure from the 2008 financial data.

Table 1 shows the descriptive statistics of the financial data of the full sample for a set of substantively important covariates. I show net income, total assets, return on assets (a measure of profitability), share of equity to assets (a standardized size of equity). It also shows to important characteristics of the vulnerability to liquidity crisis: impaired loans to equity ratio, and liquid assets to borrowed funds ratio.

Impaired loans to equity is a commonly used measure of asset quality. According to the US regulations, a loan becomes impaired when a borrower fails to make a payment. Impaired Loans to Equity ratio is an intuitive measure of the size of “bad” loans in the bank’s portfolio. Ratio of liquid assets to borrowed funds is a measure of the bank’s resilience to possible liquidity crises. Liquid assets are the assets that

Figure 1: Bank Failures per Year



**Note:** Bank failures per year. The data come from FDIC.

Table 1: Descriptive Statistics

	Variable	Mean	Median	SD
1	Net Income, th USD	2579.85	767.00	205207.80
2	Total Assets, th USD	1710338.27	142804.00	33797991.92
3	ROA	0.45	0.86	2.02
4	Equity/Assets	11.14	9.88	5.03
5	Impaired Loans/Equity	14.80	7.46	27.23
6	Liquid/Borr	9.70	6.59	10.64

**Note:** Source: Orbis Bureaus Van Dijk database and author's calculations. *ROA* is return on assets. *Liquid/ Borr* is a ratio of liquid assets to borrowed funds. All statistics are calculated using the full sample. *Impaired Loans/Equity* is a ratio of impaired loans to equity. See Section 3 for detailed discussion various performance ratios used in the analysis.

could be sold immediately at a market price (usually, those assets are stocks, bonds, derivatives and other financial instruments). Assets that are *not* liquid are usually land, real estate, inventory, art, rare wines etc.

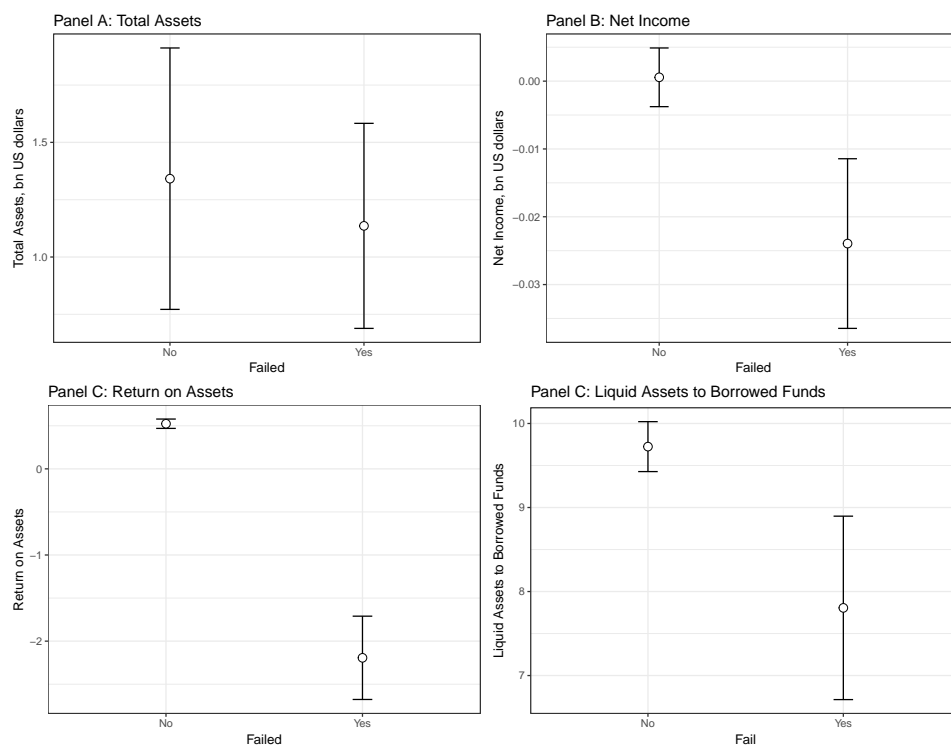
As one can see from the table, there is a significant variation in the values of those measures. For example, the standard deviation of liquid assets to borrowed funds ratio is around the same size as the mean. For impaired loans to equity ratio, the standard deviation is almost twice as large as the mean. This suggest a huge variation in the vulnerability to liquidity crises among the banks in my sample. Also, for all the measures except equity to assets ratio the median is smaller than the mean, so the data exhibit large right tails. This presents a problem for the feature engineering, especially for the linear model, since the least-squares loss function is vulnerable to the outliers.

After preparing the data set, I randomly divide it into test set and training set. Test set constitutes the 20 percent of the original data set: 1294 observations with 49 of them being banks that eventually failed. The rest of the data set is in the training set. Just by looking at the training data, we can see that the there are important differences in the distribution of the values of the ratios between the banks that failed and the banks that have not yet failed.

Figure 2 shows the some of those features plotted for the banks that have failed



Figure 2: Some Potentially Predictive Features



**Note:** Panels A to D show point estimates and 95-percent confidence intervals for the various characteristics of banks that have failed afterwards and not failed. Only training data were used for these estimates.

afterwards and banks that have not failed. We see that the banks that are to fail have, on average, less assets, smaller income, smaller return on assets, and smaller ratio of liquid assets to borrowed funds. I consider all these covariates (as well as many others) in developing predictive models for bank failures.

### 3. ADDITIONAL FEATURES

As has been noted previously, the values of the items in the banks' financial reports might be insufficient for assessing the vulnerability of the bank and its financial conditions. It is the ratios between those variables that are often viewed as important determinants of survival probabilities of banks. This section briefly reviews the ratios that are calculated from the various items of the banks' balance sheets and income statements.

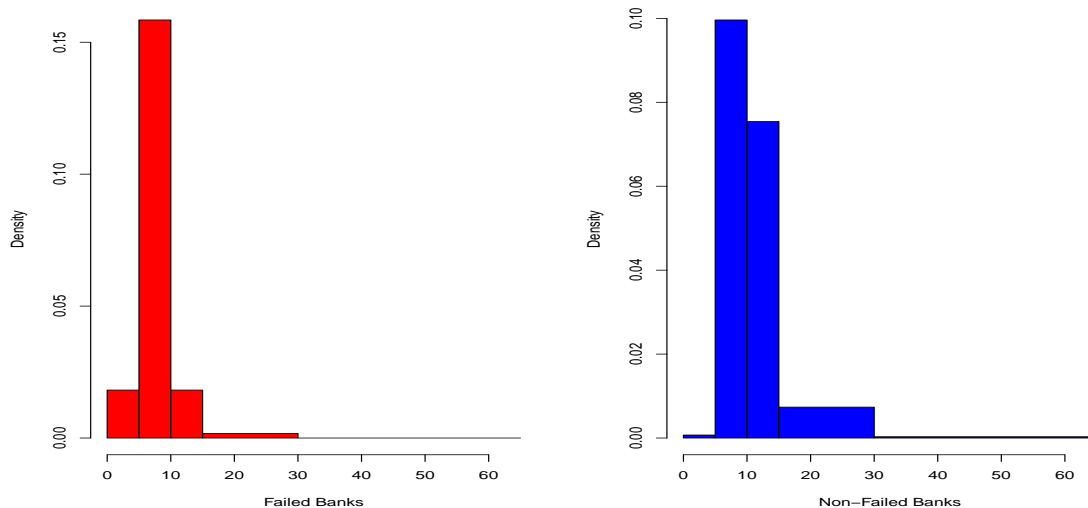
1. **Solvency Ratio.** Solvency Ratio is an after-tax net income divided by the total liabilities. Also known as "acid-test ratio", the solvency ratio quantifies the ability of a bank to pay its debts. The higher is the ratio the better. Usually, the ratio less than 20 percent can be a cause for concern<sup>1</sup>. The solvency ratio is a crude way to assess the solvency that has two main drawbacks. The first disadvantage is that the denominator does not distinguish between long-term liabilities and short-term liabilities. Of course, the higher is the share of the short-term liabilities the more problematic is the situation of the bank. The second reason is equilibrium effects: other financial institution may see that the bank is in trouble and refuse to provide funding for the bank. The depositors might also sense trouble and start withdrawing their deposits. These processes that happen because the bank is perceived to be in trouble decrease the denominator thus increasing the solvency ratio. In this case, paradoxically, the bank would appear more "healthy" than it is. Despite these problems, sol-

---

<sup>1</sup>Often, in solvency ratio the numerator includes not the net income, but a sum of net income and depreciation, but because the depreciation values are not available (and are usually trivially small for the financial firms anyway), the depreciation is not included in the definition

vency ratio is often considered to be an important indicator for assessing the bank's financial situation.

Figure 3: Solvency Ratio Histogram



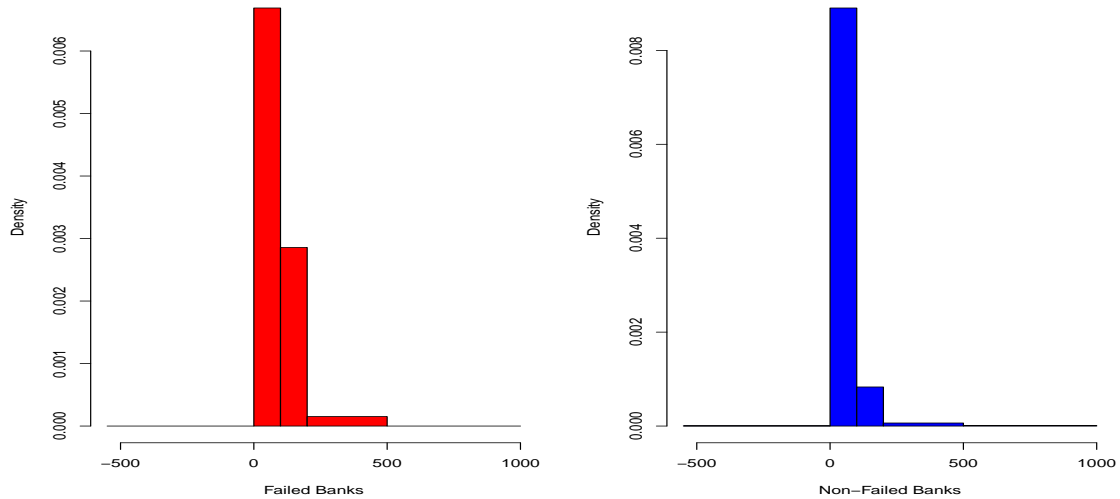
**Note:** Solvency ratio of the the banks that have failed (shown in red) and the banks that have not failed (in blue).

Figure 3 shows the the distribution of the solvency ratios for the banks that have not failed (in red) and the banks that have failed (in green). One could see the the failed banks occupy the very left tail of the distribution, suggesting that solvency ratio can indeed be predictive of bank failure. One can also notice that for the same level of solvency ratio, some banks do fail and some do not, suggesting that a significant proportion of the variation in probability of failure can not be explained by solvency ratio.

2. **Cost to Income Ratio.** Cost to income ratio is a popular measure of efficiency of a bank. It is calculated as operating income divided by operating expenses. Usually, the smaller value of this ratio indicates a more efficient bank. Importantly, neither assets nor liabilities influence this ratio directly. Assets only influence the cost to income ratio through the income from interest payments on a loans that the bank gave out, and liabilities influence the ratio only through

the payments that the bank has to make to meet its obligations. This is a useful ratio, but because it does not consider any balance sheet variables, it should only be used with the other ratios.

Figure 4: Cost to Income Ratio Histogram



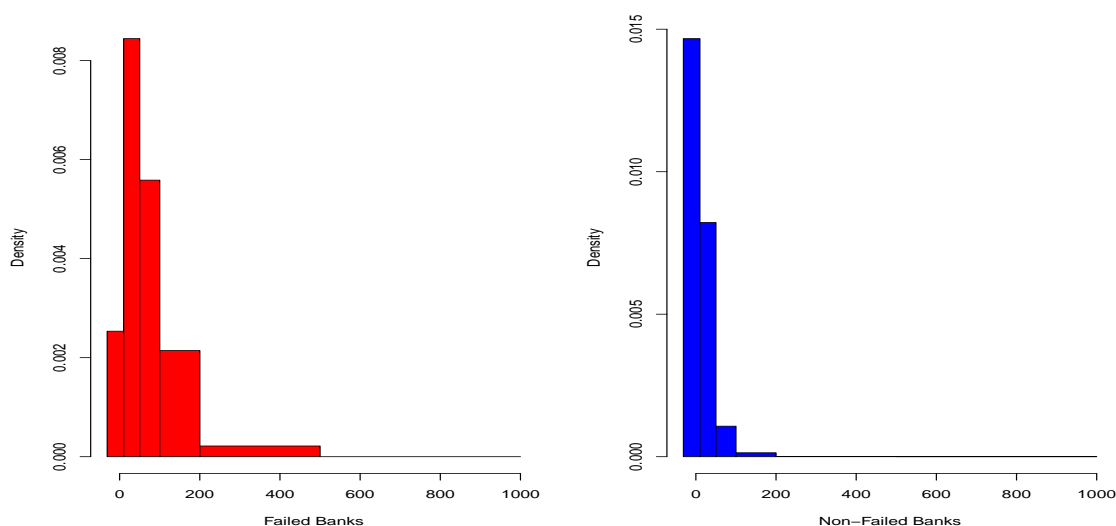
**Note:** Cost to income ratio of the banks that have failed (shown in red) and the banks that have not failed (in blue).

Table 4 shows the distributions of the cost to income ratio among the banks in my training set. First, we see a huge variation: cost to income ratio can be as low as -500 and as large 500. Secondly, unlike with the solvency ratio, there is no discernible pattern in where the failed banks are located.

3. **Loan Loss Provisions to Net Income.** Loan Loss Provisions is money that the bank reserves to cover for bad loans. If the person or an organization that borrowed money from the bank defaults on its loan, then in many cases the bank has to cover for that loan (if for example, this loan has been used as a collateral in the bank's own borrowing). This ratio can be interpreted in different ways. One interpretation is that if this ratio is higher, then the bank is more conservative in its projections, and thus is less vulnerable to the liquidity shocks. Another interpretation is that the bank with the higher loss provisions

to net income ratio has given out more bad loans and is likely to experience more defaults of its borrowers in the future. Like the previous performance ratios, this ratio has to be interpreted with caution.

Figure 5: Loss Provision to Net Income Histogram



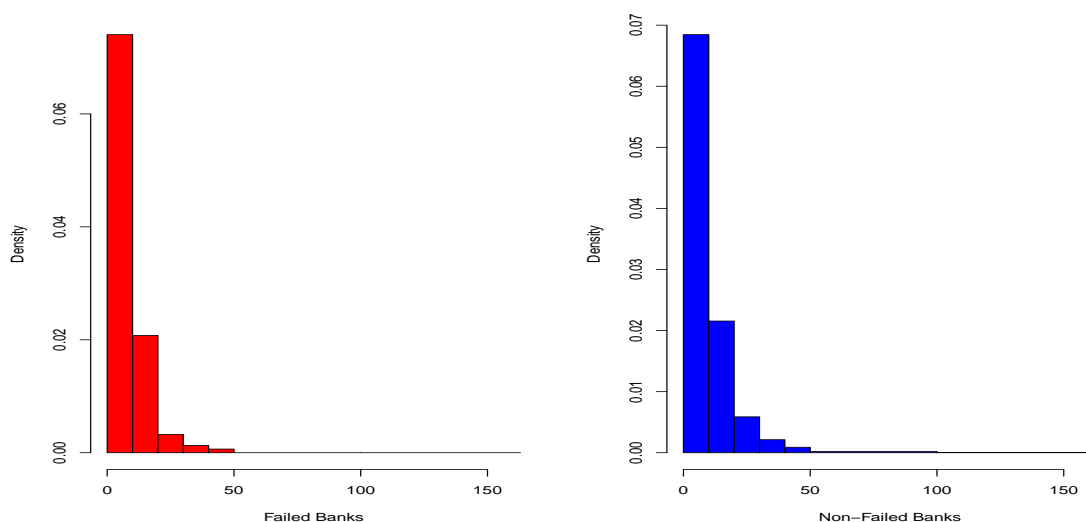
**Note:** Loan loss provision to net income ratio of the the banks that have failed (shown in red) and the banks that have not failed (in blue).

Figure 5 shows the distribution of this ratio in my training sample. We find, first, that there is a lot of extreme variation in this ratio. Secondly, non-failing banks tend to have smaller loss provisions to income ratio, suggesting the the second interpretation might be correct: higher amount of loss provisions indicates higher amount of bad loans and not necessarily more conservative projections regarding the repayment.

4. **Ratio of Liquid Assets to Borrowed Funds** Liquid assets are cash and financial instruments that can be sold for cash fast and with small transactions costs (usually, stocks, bonds, and derivatives). Borrowed funds are deposits and loans. This ratio is a way to measure banks' resilience to liquidity crises. For example, if a proportion of the bank's clients decide to withdraw the deposits, higher amount of liquid assets would help the bank to cover for those deposits

and continue to operate.

Figure 6: Liquid Assets to Borrowed Funds



**Note:** Ratio of liquid assets to borrowed Funds of the the banks that have failed (shown in red) and the banks that have not failed (in blue).

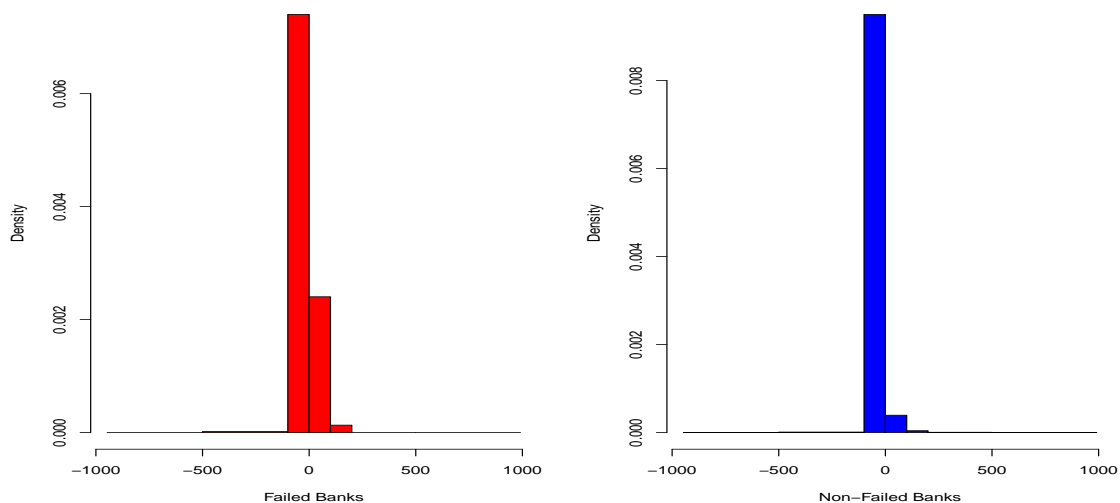
Figure 6 shows the distribution of this measure in my training sample. We observe that the banks that eventually failed had had smaller ratio of liquid assets to borrowed funds. This is consistent with the view that this measure might be predictive of bank failures.

5. **Non-operating Income to Net Income.** Banks are supposed to take deposits from the public and provide loans to businesses and households. This means that for a well-run retail bank, the major source of income should be interest payments on the loans it gives. All other sources of income – dividends, liquidation of collateral, income from buying and selling securities – should constitute a minor proportion of total income. If it constitutes a large proportion of net income, then it might indicate that the bank is engaging in businesses which not typical for a bank. This is often considered a bad sign.

Figure 7 shows the distributions of this ratio. We do see that non-failed banks alike the most common value is close to zero, and for non-failed banks devia-

tions to the right from zero are more likely.

Figure 7: Non-Operating Income to Net Income



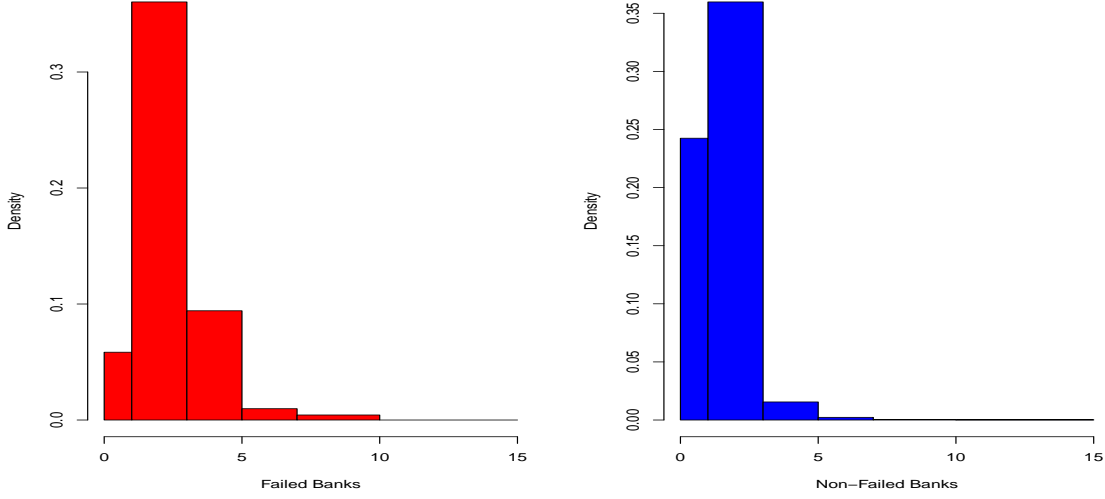
**Note:** Ratio of non-operating income to net income of the the banks that have failed (shown in red) and the banks that have not failed (in blue).

6. **Gross Reserves to Loans Ratio** This ratio, as the name suggests, shows the amount of reserves that the bank has as proportion to loans the bank has given. Usually, the bank is considered more resilient to shocks if this number is high. But, alternatively, it may indicate that the management is afraid that the loans in the bank's portfolio would become impaired.

Figure 8 shows the distribution of this ratio. Failed banks do indeed have more observations in the right side of the distribution, suggesting that higher level of gross reserves as proportion of loans might be positively related to the probability of future failure. However, non-failed banks have more extreme values.

I also use several different versions of these ratios, ending up with 58 predictors. In the next section, I use several flexible data-driven specifications to formulate the predictive models and evaluate their performance out-of-sample.

Figure 8: Ratio of Gross Reserves to Loans



**Note:** Ratio of gross reserves loans of the the banks that have failed (shown in red) and the banks that have not failed (in blue).

## 4. PREDICTIVE MODELS

This section presents a series of predictive models I use to predict the episodes of bank failures. It should be noted that my goal here it twofold: first, I aim at predictive accuracy so that those models could be used by regulatory agencies, policy makers, and the banks themselves to evaluate the risk of failure. I am also interested in building interpretable models.

### 4.1. Logistic Regression with Stepwise Forward Selection

The first model I consider is a logistic regression where the covariates are selected using stepwise forward procedure. The procedure is the following:

1. Start with null model that contains only an intercept and does not contain any predictors.
2. Fit  $K$  logistic regressions (where  $K$  is a total number of predictors in the data) adding to the previous model just one of the predictors. Select into the model



a predictor that results in a model with the largest AIC value.

3. Repeat Step 2 until the maximum value of AIC in all the models considered under this rule is reached.

I estimate this procedure using the package MASS in R. As a result, out of 58 initial predictors, 41 are selected into the resulting model (the list of those is reproduced in Appendix).

Table ?? shows the confusion matrix from the application of the classifier to the test data. The accuracy is rather high – 96 percent. However, this high accuracy is achieved by classifying almost all of the observations into the most prevalent class, non-failing banks. In the test data, 49 banks belong to the class of failing banks, but only three cases has been classified as such in such by the classifier.

Table 2: Confusion Matrix for Stepwise Logistic Classifier

	Actual Failure	Actual Non-Failure
Predicted. Failure	0.002	0
Predicted Non-Failure	0.035	0.96

**Note:** Confusion matrix is calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 4.2 presents separately Type I and Type II error rate on the test data for this classifier. We see that the rate of false alarms is zero, but the rate of missed failures is very high. Because bank failures are rare events, just classifying almost all of the observations into a non-rare class can produce high out-of-sample accuracy. In the contest on bank failures, however, the cost of Type II error is high, since when the regulators miss bank failures, this can lead to cascading adverse effects in the economy. Thus, a classifier that miss most of the adverse events can hardly be adequate.

#### 4.2. Logistic Regression after Selection via LASSO

The second model I consider belongs to a class of post-selection models described by Efron and Hastie (2016, pp. 394-417). The basic idea of such algorithms is to

Table 3: Error Rates for Stepwise Logistic Classifier

	Rate
Type I (False Alarm)	0
Type II (Missed Failure)	0.939

**Note:** Out-of sample error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

break down the selection into two steps: the first step uses some principled way to select a model from a certain class of models (in this section, I consider a class of linear sparse models), and the second step estimates the model and generates predictions.

Here, I apply the following algorithm:

1. Select the variable for logistic regression via LASSO.
2. Estimate logistic regression with an indicator for bank failure the outcome and the variables selected by LASSO on the previous step as predictors.

This procedure has been argued to have an advantage over just estimating logit with  $l^1$  penalty, since, in general the coefficient estimates from  $l^1$ -penalized regression will be biased (Belloni et al. (2012); Belloni, Chernozhukov, and Hansen (2014)). Breaking up the estimation procedure into two two steps can have a debiasing effect.

Post-selection inference in general has been discussed by Berk et al.(2013) with more specific approaches for high-dimensional models being presented by Van de Geer et al.(2014), Zhang and Zhang (2014), Javanmard and Montanari (2014). Belloni et al. (2012) derive rate of convergence for post-Lasso inference for linear models and apply post-Lasso inference to the study of the effect of eminent domain on land prices. Bryzgalova (2016) develops a version of post-Lasso inference to study risk factors in a cross-section of stock returns.

I assume the following data-generating process:

$$fail_i \sim \text{Bernoulli}(p_i) \quad (1)$$

Here,  $fail_i$  equals 1 if bank  $i$  eventually failed, and 0 otherwise;  $p_i$  is an (unobserved) probability of failure. This probability is given by an inverse logit link function:

$$p_i = \text{logit}^{-1}(\beta_0 + x_i^T \beta) \quad (2)$$

Here,  $x_i$  is a vector of covariates of bank  $i$ . I also make the assumption of sparsity:

$$\sum_{j=1}^K 1(\beta_j \neq 0) \ll n \quad (3)$$

Here,  $K$  is a number of possible predictors (I have 58 in my dataset, not counting higher-order polynomial terms and possible interactions).  $n$  is the number of observations. Intuitively, as in LASSO framework, this assumption postulates that the number of non-zero coefficients in a true data-generating process is very small.

This assumption in general is useful in situations when  $K \gg n$ . In my setup,  $K$  is smaller than number of observations in the training data. I still find this assumption useful because, by selecting only a subset of observations, I might get better qualitative insights into the process of bank failure and develop an interpretable model.

Equations 1 - 3 define the data-generating process. I estimate the vector of  $\beta$  using the two-step procedure I have outlined above: selecting a vector of non-zero  $\beta$ 's using LASSO and then fitting a non-penalized GLM.

LASSO, as described by Tibshirani (1996), estimate  $\beta$ 's by minimizing the least-squares cost function with  $l^1$  penalty:

$$L(\beta) = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \|\beta\|_1 \quad (4)$$

Here,  $\lambda$  is a hyperparameter that describes the magnitude of the penalty.

This setup has an inconsistency: equation 4 implicitly treats the variable  $y_i$ , an outcome in the model as continuous, however in my case it is an indicator for bank failure (a binary variable). One might argue that Equations 1 - 3 and Equation 4 imply different data-generating processes. This concern is valid. However, because, as described above, I have unbalanced data (5023 observations in the training set, 154 of them failures), the logit with  $l^1$  penalty (as implemented in package `glmnet` in R), fails to converge in the cross-validation stage – most likely, because the constant  $\beta_0$  can be estimated is  $-\infty$ <sup>2</sup> For this reason, I use ordinary LASSO (with normal family) as a device for variable selection. As I discuss later, this algorithm would have some desirable properties.

For LASSO, I select the penalty parameter ( $\lambda$ ) via cross-validation and choose the value of  $\lambda$  which is one standard deviation larger than the value of  $\lambda$  that minimizes cross-validation MSE.

Also, as has been shown previously, as almost all values of in my data-set have long right tails, a Box-Cox transformation might be appropriate. I apply cubic-root transformation to all the variables.

After these transformation, I estimate LASSO with `glmnet` package in R (Friedman, Hastie, and Tibshirani (2010)).

Out of 58 variables<sup>3</sup>, the following ones are selected by LASSO: net income, return on equity, impaired loans, gross reserves to loans, loss provisions to net income, impaired loans to equity.

Substantively, it turns out that, intuitively, what seems to matter for the survival of a bank is its profitability (measured in several ways) and amount of “bad” loans on its balance sheet (also measured in several ways). Some factors that are expected to matter turn out to have no predictive power. Those factors are amount of assets

---

<sup>2</sup>Asymptotically, with unbalanced data, the estimates of constant in a logistical regression approach infinity (Owen (2007)).

<sup>3</sup>In a separate estimation, I also included polynomial expansion of all the variable, up to 5th degree. None of the higher-degree variables ended up being selected.

Table 4: Logistical Regression after LASSO Selection

	Model 1
Net Income	−0.59* (0.24)
Return on Equity	−0.09 (0.26)
Impaired Loans	0.40** (0.13)
Gross Reserves to Loans	−0.11 (0.12)
Loss Provisions to Net Income	0.61** (0.20)
Impaired Loans to Equity	1.04*** (0.21)
Num. obs.	5177

**Note:** Logistic regression with the bank failure is dependent variable. Intercept is not shown. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

on the balance sheets, structure of those assets, and structure of income (interest income, dividends, etc.). The amount of assets only matters in relation to the amount of impaired loans.

As a next step, I estimate logistical regression with these covariates. Table 4.3 presents the result of the estimation. All the variables are standardized, so the coefficients can be compared. In terms of magnitude and the size of the p-value, the most important variable is the ratio of impaired loans to equity. Given the base rate of failure of 5 percent, a logistic coefficient of 1 corresponds (in the training data) to the change in predicted probability of 4.8 percentage points for the one standard deviations of impaired loans to equity ratio.

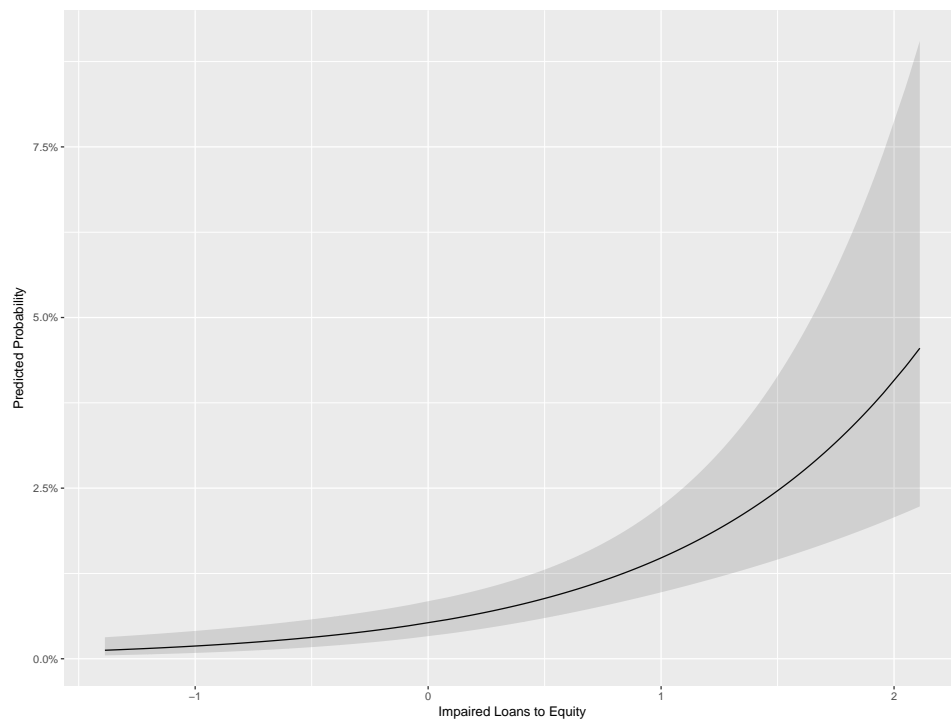
Figure 9 shows the change in predicted probabilities for training data, when the (standardized) impaired loans to equity ratio goes up.

These results indicate that, at least in the training data, the post-LASSO logit approach was able to identify a set of variable that are strongly related.

In the test data, the classifier based on post-LASSO logistical fit is also able to distinguish future failures from non-failures.

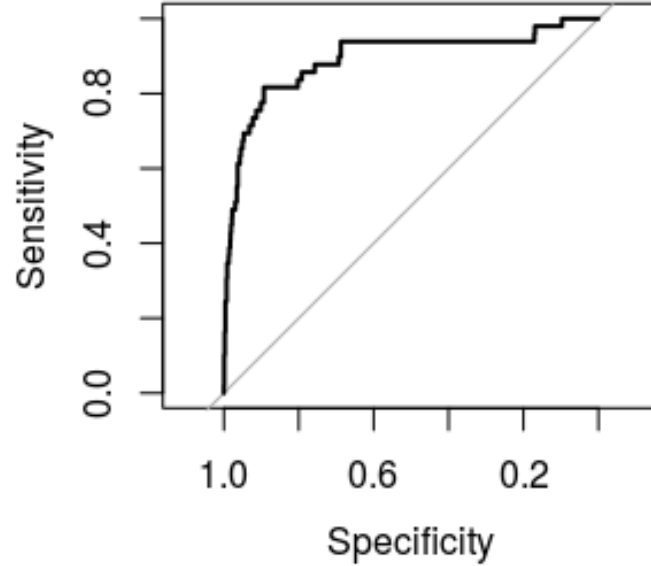
If we classify observations, where the predicted probability is larger than or

Figure 9: Effects of Impaired Loans to Equity Ratio



**Note:** Marginal effects of impaired loans to equity ratio on predicted probability of bank failure. Results from the logistic regression model.

Figure 10: ROC curve for Post-LASSO Logistic Classifier



**Note:** ROC curve for test data from the post-LASSO logistic classifier. Area under the curve: 0.89

equal to 0.5 as failures, and all other observations as non-failures, then we gen the following confusion matrix (Table 4.4 ).

Table 5: Confusion Matrix for Post-Lasso Logistic Classifier

	Actual Failure	Actual Non-Failure
Predicted. Failure	0.027	0.073
Predicted Non-Failure	0.01	0.89

**Note:** Confusion matrix is calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 4.5 separately presents the Type 1 and Type 2 error rates for this classifier calculated on the test data.

The misclassification rate in the test data is 0.083. Since in the test data only 3.7 percent of cases are actual failures this classifier performs worse than a benchmark that classifies all the cases as non-failures without looking at covariates. However, arguably, the benchmark classifier is useless since missing a possible failure in the context of financial regulation is much more costly than the false alarm. Since

Table 6: Error Rates for Post-Lasso Logistic Classifier

	Rate
Type I (False Alarm)	0.076
Type II (Missed Failure)	0.27

**Note:** Out-of sample error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

the non-failure benchmark always misses all the failures, and the post-LASSO logit classifier correctly predicts 73 percent of the failure and has a reasonable rate of false alarms, the predictive model developed in this section is arguably more useful for the practical purposes than the non-failure benchmark.

#### 4.3. *Random Forest*

Tree-based methods allow predictors to have a complicated structure of interactions. Because it is reasonable to expect that in the case of predicting bank failures the covariates might interact with one another in a way that is hard to evaluate ex-ante, I use tree-based methods for prediction.

A classification and regression tree (CART) for predicting an outcome  $y$  is an algorithm that uses recursive partitioning of the training data to achieve prediction (Efron and Hastie (2016, p. 124)). At each step of the partitioning, a variables  $X$  (or a set of variables) is selected for which the split is performed and, for every variable, a value ( $m$ ) is selected such that all the observations in current partition are split further into two groups, such that for one group a value of  $X$  is weakly larger than  $m$  and for another group the value of  $X$  is smaller than  $m$ . At each stem,  $X$  and  $m$  is selected to maximize the difference between the two groups regarding the outcome  $y$  (see Breiman (1996) and Efron and Hastie (2016, pp 126-127) for an overview of partitioning rules).

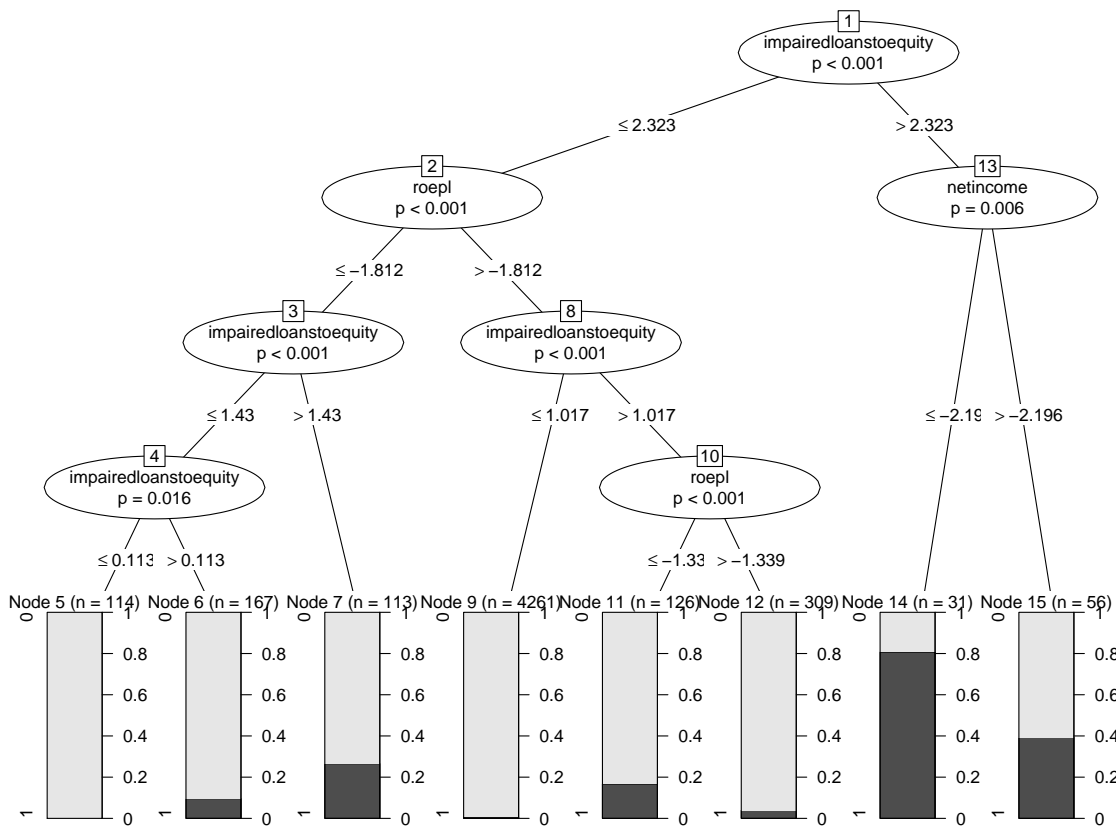
Random Forest builds on the CART approach by building many random trees using bootstrapped samples of the observations and the variables and the aggregating the resulting partitions. In this section, I attempt to build a predictive model



of bank failures using a Random Forest approach (as implemented in R package `randomForest` (Liaw and Wiener (2002))).

To substantively evaluate the possible structure of interaction, it might be instructive to start with a simple conditional inference tree (Hothorn, Konik, and Zeileis (2006)). This method uses recursive partitioning to build a decision tree for classification.

Figure 11: Tree-Based Classifier for Predicting Bank Failures



**Note:** Decision tree for classifying the banks into failures and non-failures on training data. Definition of the variables: *impairedloanstoequity* is a ratio of impaired loans to equity, *roepl* is return on equity, *netincome* is net income. The tree is produced using R package `party`

Figure 11 shows a decision tree constructed with the variables selected using LASSO<sup>4</sup>. One can see that even with the small number of variables selected by LASSO,

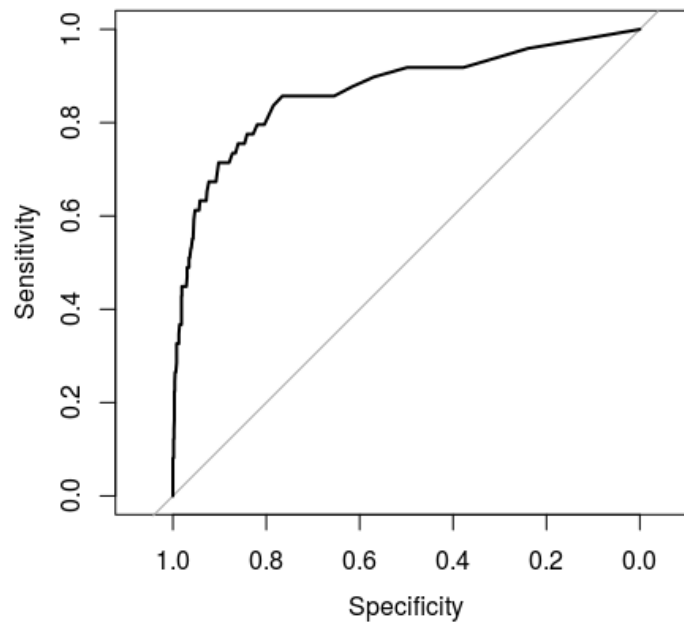
<sup>4</sup>In a Random Forest procedure, I will be using all the variables

the process might indeed depend on the interactions between variable. For example, the largest share of bank failures is observed among the banks with high ratio of impaired loans to equity *and* relatively low levels of net income. To take advantage of such interactions for the predictions, one might find tree-based methods useful.

I fit a Random Forest algorithm using all the covariates in my data set as predictors and choosing the hyperparameters with 10-fold cross-validation <sup>5</sup>.

I use the fitted model to calculate predicted probabilities of failure on the test data. The resulting ROC curve is shown in Figure 12.

Figure 12: ROC Curve for the Random Forest



**Note:** ROC curve calculated on the test data for the Random Forest predictions. Area under curve is 0.87

The confusion matrix is shown in Table 4.6. This classifier is worse than the post-Lasso logistic classifier, since it has the similar amount of correctly predicted failures but higher amount of false alarms<sup>6</sup>.

<sup>5</sup>As a result, the value of *mtry* parameter (number of variables used after a node is split) is set to 2

<sup>6</sup>This also evident from the comparison of AUC for post-LASSO logit (0.89) to the AUC of Random Forest (0.87)

Table 7: Confusion Matrix for Random Forest Classifier

	Actual Failure	Actual Non-Failure
Predicted. Failure	0.027	0.0714
Predicted Non-Failure	0.01	0.855

**Note:** Confusion matrix is calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 4.7 presents Type I and Type II error rates of the classifier calculated with the test data. One can see that the Type II error rate is the same as in the post-LASSO logit estimation, and Type I error rate is slightly larger.

Table 8: Error Rates for Random Forest Classifier

	Rate
Type I (False Alarm)	0.077
Type II (Missed Failure)	0.27

**Note:** Out-of sample error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

As in the previous approach, I attempt to formulate which variables are the most important for predicting bank failures. To assess the importance of the variables, I use Gini Impurity Criteria, which is a standard procedure for evaluating importance of features after estimating a Random Forest algorithm (Breiman (2002) and Louppe et al. (2013)). For this measure, for every variable, a weighted Gini impurity decrease is calculated once the variable is removed. The higher is the decrease, the more important that variable is, since its removal contributes to the decrease in the total difference between the nodes.

Table 4.8 presents top 10 variables with the largest mean decreases in Gini.

As in the post-LASSO algorithm, the ratio of impaired loans to equity is by far the most important predictor of bank failure. Intuitively, according to this measure those banks are more likely which have larger proportion of “bad” loans. The inclusion of the measure of loss provisions to net income and the size of impaired loans points to this mechanism also. Unlike in post-LASSO logit estimation, the solvency ratio seems to be an important predictor. In general, there is a significant overlap

Table 9: Variable Importance for Random Forest

Name	Mean Decrease Gini
Impaired Loans to Equity	14.9
Loss Provisions to Net Income	8.4
Return on Equity	8.4
Pre-tax Net Income	8.4
Impaired Loans	8.3
Equity to Assets	8
Solvency Ratio	7.8
Net Income	7.7
Equity to Liability	7.5

**Note:** Mean Gini Decrease is reported for all the variable. Only training data have been used in the calculation. See Breiman (2002) on the details of Gini Impurity Criteria

in the features selected by post-Lasso logit algorithm and variables selected by a Random Forest.

#### 4.4. Support Vector Machine Classifier

This section uses a Support Vector Machine (SVM) classifier for predicting bank failures using all available covariates. This method has a disadvantage of producing results that might not be interpretable, but it might achieve higher accuracy than the other methods.

A linear SVM classifier minimizes hinge loss plus a penalty (Efron and Hastie (2016, p. 379)):

$$L(\beta) = \sum_{i=1}^n [1 - y_i(\beta_0 + x_i^T \beta)]_+ + \lambda \|\beta\|_2^2 \quad (5)$$

Here,  $y_i$  is an outcome, and  $x_i$  is a vector of covariates.

As in the previous examples, I transform all the predictors using cubic roots. I fit the linear SVM model with the hyperparameters tuned using three-fold cross-validation resampled 10 time<sup>7</sup>.

The confusion matrix for the test data is presented in Table 4.9. This classifier

<sup>7</sup>Variations in the number of folds and number of resamples turn out to be inconsequential

has lower misclassification rate than the post-LASSO logistic classifier, however from the practical perspective it might be worse even though it is more accurate on average. The reason for this is, as I mentioned before, different costs associated with Type 1 and Type 2 errors. Because missing a looming bank failure can be extremely costly for the financial system and for the economy, while implementing preventive measures in the case of alarm might not be as costly, it might be better for the classifier to catch as many real bank failures as possible (while also containing the level of false alarms at a reasonable level).

Table 10: Confusion Matrix for SVM Classifier

	Actual Failure	Actual Non-Failure
Predicted Failure	0.002	0
Predicted Non-Failure	0.036	0.962

**Note:** Confusion matrix is calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 4.10 shows the Type I and Type II errors for linear SVM classifier. The Type I error rate is zero, but Type II error rate is higher than 99 percent.

Table 11: Error Rates for Linear SVM Classifier

	Rate
Type I (False Alarm)	0
Type II (Missed Failure)	0.994

**Note:** Out-of sample error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

The SVM classifier has zero false alarms in the test data, but the level of correctly predicted failures is also very low. In sum, this predictor is too conservative and almost never predicts failures. As a result, it is only able to identify correctly 0.6 percent of true failures, which is inadequate for all practical purposes.

#### 4.5. Model Stacking

Finally, I use an ensemble method, model stacking, to combine information from all the models. The resulting confusion matrix is presented in Table 4.11.

Table 12: Confusion Matrix for Model Stacking

	Actual Failure	Actual Non-Failure
Predicted Failure	0.0007	0.012
Predicted Non-Failure	0.037	0.95

**Note:** Confusion matrix is calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 4.12 presents the results for model stacking.

Table 13: Error Rates for Model Stacking

	Rate
Type I (False Alarm)	0.01
Type II (Missed Failure)	0.98

**Note:** Out-of sample error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

From these table, the benefits of stacking are not immediately obvious. Some characteristics are slightly improved in comparison to each of the individual methods, but some are made worse. The out-of-sample misclassification rate is smaller than the misclassification rate of post-Lasso logit and Random Forest, but slightly larger than the misclassification rate of SVM classifier. Most importantly, this classifier is still too conservative misclassifying most of the actual bank failures.

In sum, in terms of minimization of misclassification rate, and SVM classifier turns out to be the best, but I would not recommend adopting it since it achieves its high accuracy by missing a lot of actual bank failures. Because the price of these kinds of misclassifications (Type 2 errors) is high in the context of financial regulation, a classifier that allows for some manageable false positive rate but also correctly predicts most of the failures seems more preferable. For these reasons, I would recommend post-Lasso logistic classifier for the practitioners as a useful device to augment information from other forms of analysis.

#### 4.6. Other Accuracy-Maximizing Algorithms

In this section, I present results of a set of additional algorithms I used to for the prediction. It turns out that all the cases those algorithms produce results that are inferior to the results presented earlier. Table 4.13 presents the out-of-sample performance metrics of those approaches.

Table 14: Additional Predictive Algorithms

Algorithm	Accuracy	Type I Error	Type II Error
Logistic Regression	0.96	0	1
SVM with Radial Basis Functions	0.96	0	1
SVM with Polynomial Kernel	0.96	0	1
Weighted Subspace Random Forest	0.96	0	1
Genetic Tree Models	0.54	0.47	0.28
Sparse LDA	0.58	0.44	0.06
Single C5.0 Tree	0.29	0.73	0.06
Neural Network	0.96	0	1
Naive Bayes	0.96	0	1
Maximum Uncertainty LDA	0.003	0.04	0.999

**Note:** Error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Many of these approaches render a non-informative prediction, when all the observations are assigned to the most prevalent class. Other approaches render classifiers that can be outperformed by a random classification. The three classifiers I have presented early turn out to be the best out of larger sample of classifiers.

#### 4.7. Neyman-Pearson Classifiers

Classifiers I used in the previous section optimized overall predictive accuracy in the training data. While this approach is valid, an alternative would be to hold Type I error fixed and minimize Type II errors. In the context of predicting bank failures this approach might be useful because, as I mentioned previously, the failure of a bank is a costly event that might adversely impact the economic system. This is why some level of false alarms might be tolerable, if the early warning system is able to catch most of the banks that are about to fail.

To capture this difference in costs of Type I and Type II errors, I apply an approach developed by Tong, Feng, and Li (2018) who propose an umbrella algorithm for the classifiers that enforce a pre-selected upper bound for Type I errors, while minimizing Type II errors. Using R package `nproc`, I set up an upper bound for the rate Type I errors to 0.1, and perform the classification using ten classification algorithms. The out-of sample diagnostics are presented in Table 4.14.

Table 15: Classification with NP paradigm

Algorithm	Accuracy	Type I Error	Type II Error
Logistic Regression	0.91	0.07	0.37
SVM	0.92	0.07	0.35
Penalized GLM	0.91	0.09	0.27
Random Forest	0.92	0.07	0.35
LDA	0.91	0.08	0.29
Sparse LDA	0.91	0.09	0.27
Naive Bayes	0.90	0.09	0.49
Nonparam. Naive Bayes	0.90	0.09	0.49
Ada-Boost.	0.91	0.08	0.24

**Note:** Error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

Table 16: Classification with NP paradigm: Type II Error Limited

Algorithm	Accuracy	Type I Error	Type II Error
SVM	0.15	0.88	0.18
Sparse LDA	0.89	0.11	0.22
Penalized GLM	0.90	0.10	0.20
Ada-Boost.	0.88	0.08	0.24

**Note:** Error rates are calculated by classifying predicted probabilities larger of equal to 0.5 as failures, other as non-failures.

In the Neyman-Pearson paradigm, the rate of Type II errors becomes tolerable. Naive Bayes and Nonparametric Naive Bayes perform worse than others – their rate of Type II errors is 49 percent. Penalized GLM performs close to the the post-Lasso logit presented earlier. The best algorithm, in terms of minimizing Type II error rate in the test data is Ada-Boost (Friedman, Hastie, and Tibshirani (2000)). Penalized GLM and sparse LDA perform worse, but the difference is relatively small. I expect



on different samples all three methods, Penalized GLM, Sparse LDA, and Ada-Boost, when Neyman-Pearson paradigm is applied perform reasonably well.

Table 4.15 shows the results when I run `nproc` when limiting Type II errors to 30 percent and minimizing Type I errors. For Sparse LDA and Penalized GLM this results in smaller out of sample Type II error rates and Type I error rates. For Ada-Boost, the Type I error rates became larger, but Type II error rates remained the same. For other algorithms, like SVM, this resulted in high Type I error rates.

## 5. CONCLUSION

The goal of this paper is to investigate the determinants of bank failures after the financial crisis of the years 2007 - 2009 and to build a predictive model of bank failures.

In the existing work, the most common dependent variable is “financial distress” which is measured using financial statements of the banks. Instead, I use an independently recorded event – bank failure reported by FDIC – as my main outcome. I use literature on the economics and corporate behavior of banks to come up with a set of performance ratios that might be plausibly connected to bank failures. With those covariates and all the other items from the banks’ income statements and balance sheet, I evaluate a set of predictive models.

I use two paradigms for prediction: accuracy maximization, and Neyman-Pearson paradigm. Accuracy maximization implies that Type I errors and Type II errors are equally costly, thus out-of-sample predictive accuracy is the most important parameter for evaluation. Neyman-Pearson paradigm implies setting an upper bound for Type I errors and minimizing Type II errors within that bound. In this case, the costs associated with Type I and Type II errors can be different.

I find that, because the bank failures are rare events, many of the accuracy-maximizing classifiers tend to assign all the observations to the class of non-failing banks. This achieves out-of-sample predictive accuracy of 96 percent, but, of course,

misses all the failures. This is inadequate for practical purposes. Two algorithms, post-Lasso logit and random forest tend to have relatively low level of Type II errors.

I find that the highest out-of-sample accuracy, 96.22 percent, is achieved by Linear SVM classifier. It is slightly larger than a no-information benchmark (just predicting no failure for any observation which yields an accuracy of 96.06 percent). However, given that the bank failure is a rare event, this high accuracy is achieved at a cost of misclassifying most of the actual failures. Because of this reason, I cannot recommend SVM classifier for adoption by finance practitioners.

The classification with the Neyman-Pearson paradigm performs better in terms of minimizing Type II errors while containing Type I errors. All of the algorithms, in out-of-sample testing, were able to identify at least 50 percent of the failing banks, while having false positive rate below ten percent. The minimum share of Type II errors were displayed by Ada-Boost algorithm (24 percent), while GLM with LASSO penalty and sparse LDA did not perform much worse (the level of Type II errors were 27 percent). When Type II error is set to be bound by 30 percent, and Type I error is minimized, then penalized GLM and Sparse LDA perform slightly better, but other algorithms perform worse. More research is needed on whether it would be more efficient to bound Type I errors and minimize Type II errors, or to bound Type II errors and minimize Type I errors.

My analysis produces additional substantive insights. Some of the methods I am using are interpretable and allow tentative conclusions about which factors are more likely to be associated with bank failure. I find that low profitability and high proportion of impaired loans are the most important factors. Interestingly, other factors that are sometimes mentioned in the context of bank failures – size of assets, structure of assets, balance between short-term and long-term loans, leverage – do not seem to be selected as important predictors, conditional on the measures of profitability and bad loans. These findings might contribute to the literature on the factors of solvency of retail banks, but more research is needed.

One of biggest limitations of my analysis, which is also a limitation of my data is that the events that I trying to predict are rare. In the whole sample (in the training set and in the test set combined), I only have 203 failure events out of 6471 observations. This makes the prediction problem especially hard since many algorithms, trying to optimize for accuracy will treat the observations of failure as noise and converge on predicting nonfailure all the time (this is what happens with SVM with Radial Basis Functions, for example), or will only predict failure in one or two cases and will miss all the other failure events (this is what happens with Linear SVM). A high true positive rate can only be achieved at a cost of substantially increasing false positive rate. More research is needed on how this issue can be remedied.

One promising way to remedy the limitations of these data is to use the Neyman-Pearson framework by setting the tolerable upper bound for the level of Type I errors and minimize Type II errors. This framework can also be used for the risk analysis in a variety of situations: predicting bankruptcies, extreme losses in a portfolio of assets, and rapid changes in commodity prices. I leave those applications to further research.

## 6. APPENDIX

### 6.1. *Variables Selected by Stepwise Procedure*

Turnover

Net Income Before Tax

Solvency Ratio

Return on Assets

Solvency Ratio

Employees

Net Interest Margin

Loans

Gross Loans

Loans to Banks  
Other Securities  
Total Earning Assets  
Fixed Assets  
Non-earning Assets  
Deposits from Customers  
Bank Deposits  
Other Liabilities  
Trading Liabilities  
Long-term Funding  
Other Non-Interest Bearing Assets  
Reserves  
Equity  
Impaired Loans  
Liquid Assets  
Intangible Assets  
Net Interest Revenue  
Other Operating Income  
Net Fees  
Overheads  
Loans to Provisions  
Taxes  
Dividends  
Loss Provisions to Net Income  
Impaired Loans to Equity  
Equity to Assets  
Equity to Liabilities  
Capital Funds to Assets  
Net Loans to Assets

Net Loans to State Funds

Net Loans to Borrowed Funds

## REFERENCES

Allen, Franklin and Elena Carletti. Mark-to-market accounting and liquidity pricing. *Journal of accounting and economics*, 45(2):358378, 2008.

Altman, Edward, Giancarlo Marco, and Franco Varetto. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the italian experience). *Journal of banking and Finance*, 18(3):505529, 1994.

Bekaert, Geert, Campbell R Harvey, and Christian Lundblad. Does financial liberalization spur growth? *Journal of Financial economics*, 77(1):355, 2005.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. "Sparse models and methods for optimal instruments with an application to eminent domain." *Econometrica* 80, no. 6 (2012): 2369-2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. "High-dimensional methods and inference on structural and treatment effects." *Journal of Economic Perspectives* 28.2 (2014): 29-50.

Berger, Allen N and Christa HS Bouwman. How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109(1):146176, 2013.

Berk, Richard, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. "Valid post-selection inference." *The Annals of Statistics* 41, no. 2 (2013): 802-837.

Bernanke, Ben. The Federal Reserve and the financial crisis. Princeton University Press, 2013.

Bernard, Victor L , Robert C Merton, and Krishna G Palepu. Mark-to-market accounting for banks and thrifts: Lessons from the danish experience. Journal of Accounting Research, pages 132, 1995.

Breiman, Leo. "Some properties of splitting criteria." Machine Learning 24.1 (1996): 41-47.

Bryzgalova, Svetlana. "Spurious factors in linear asset pricing models." LSE manuscript (2015).

Burgstahler, David and Ilia Dichev. Earnings management to avoid earnings decreases and losses. Journal of accounting and economics, 24(1):99126, 1997.

Calomiris, Charles W, and Matthew Jaremski. Deposit insurance: Theories and facts. Annual Review of Financial Economics, 8:97120, 2016.

Campello, Murillo, John R Graham, and Campbell R Harvey. The real effects of financial constraints: Evidence from a financial crisis. Journal of financial Economics, 97(3):470487, 2010.

Dechow, Patricia M, Richard G Sloan, and Amy P Sweeney. Detecting earnings management. Accounting review, pages 193225, 1995.

Durnev, Art and Sergei M Guriev. The resource curse: A corporate transparency channel. 2007.

Durtschi, Cindy , William Hillison, and Carl Pacini. The effective use of benfords law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, 5(1):1734, 2004.

Efron, Bradley, and Trevor Hastie. *Computer age statistical inference*. Vol. 5. Cambridge University Press, 2016.

Ellul,Andrew. Chotibhak Jotikasthira, Christian T Lundblad, and Yihui Wang. Is historical cost accounting a panacea? market stress, incentive distortions, and gains trading. *The Journal of Finance*

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)." *The annals of statistics* 28.2 (2000): 337-407.

Friedman, Jerome H., Trevor Hastie, and Rob Tibshirani. "glmnet: lasso and elastic-net regularized generalized linear models, 2010b." URL <http://CRAN.R-project.org/package=glmnet>. R package version: 1-1.

Gorton, Gary and Guillermo Ordo nez. Collateral crises. *The American Economic Review*, 104(2):343-378,2014.

Heaton, John. Deborah Lucas, and Robert L McDonald. Is mark-to-market accounting destabilizing? analysis and implications for policy. *Journal of Monetary Economics*, 57(1):64-75, 2010.

Helleiner, Eric. Understanding the 2007-2008 global financial crisis: Lessons for scholars of international political economy. *Annual Review of Political Science*,

14:67-87, 2011.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. "Unbiased recursive partitioning: A conditional inference framework." *Journal of Computational and Graphical statistics* 15.3 (2006): 651-674.

Javanmard, Adel and Andrea Montanari. "Confidence intervals and hypothesis testing for high-dimensional regression." *The Journal of Machine Learning Research* 15.1 (2014): 2869-2909.

Jones, Jennifer. Earnings management during import relief investigations. *Journal of accounting research*, pages 193-228, 1991.

Kay, John. *Other people's money: the real business of finance*. Public Affairs, 2015.

King, Robert and Ross Levine. Finance and growth: Schumpeter might be right. *The quarterly journal of economics*, 108(3):717-737, 1993.

Liaw, A and M. Wiener (2002). Classification and Regression by randomForest. *R News* 2(3), 18–22.

Lau, Amy Hing-Ling A five-state financial distress prediction model. *Journal of accounting research*, pages 127-138, 1987.

Louppe, Gilles, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. "Understanding variable importances in forests of randomized trees." In *Advances in neural information processing systems*, pp. 431-439. 2013.

Nigrini, Mark. A taxpayer compliance application of benford's law. *The Journal*



of the American Taxatio Association, 18(1):72, 1996.

Owen, Art B. "Infinitely imbalanced logistic regression." *Journal of Machine Learning Research* 8.Apr (2007): 761-773.

Raghuram G Rajan and Luigi Zingales. Financial dependence and growth. *American Economic Review*, pages 559586, 1998.

Reuter, Peter, Chasing dirty money. Institute for International Economics, 2004.

Shiller, Robert Finance and the good society. Princeton University Press, 2013.

Sun, Jie, Hui Li, Qing-Hua Huang, and Kai-Yu He. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, 57:41-56, 2014.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267-288.

Tong, X., Feng, Y., and Li, J. J. (2018). Neyman-Pearson classification algorithms and NP receiver operating characteristics. *Science advances*, 4(2), eaao1659.

Van de Geer, Sara, Peter Bhlmann, Yaacov Ritov, and Ruben Dezeure. "On asymptotically optimal confidence regions and tests for high-dimensional models." *The Annals of Statistics* 42, no. 3 (2014): 1166-1202.

Zhang, Cun Hui, and Stephanie S. Zhang. "Confidence intervals for low dimensional parameters in high dimensional linear models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1 (2014): 217-242.

Zmijewski, Mark Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, pages 5982, 1984.