Combining expert expectations with machine learning: An analysis of bank going concern opinions before and after the financial crisis

Stuart Dearden
University of Nebraska-Lincoln
sdearden2@unl.edu

Emily Hunt
Joshua Hunt
Mississippi State University
ehunt@business.msstate.edu
jhunt@business.msstate.edu

Vern Richardson
University of Arkansas
vjricha@uark.edu

October 2021

Preliminary Draft. Please Do Not Circulate

Combining expert expectations with machine learning: An analysis of bank going concern opinions before and after the financial crisis

ABSTRACT:

We use anomaly detection machine learning techniques to predict bank failure during and after the financial crisis. We compare the accuracy of our model to the accuracy of auditor going concern opinions (GCOs) in predicting bank failure and find that auditors outperform our models. We observe an increase in auditors' propensity to issue GCOs after the financial crisis which increases their Type I GCO errors but allows them to capture almost all bank failures. Finally, we combine our machine learning predictions with auditor GCOs and capture the same number of bank failures as auditors while simultaneously reducing Type I GCO errors. Our study provides evidence that auditors have insight into the overall health of the bank client that is not reflected in readily available data and information provided in call reports. We also show that auditors responded to regulatory and market scrutiny by behaving more conservatively in the post-financial crisis period. Finally, we conclude that our machine learning method, when combined with GCOs, can improve auditor GCO accuracy. Our findings are important to investors interested in the informational value of GCOs for banks as well as auditors and regulators interested in the role that more accurate GCOs could play in preventing another financial crisis.

I. INTRODUCTION

The perceived failure of auditors to timely identify failing banks during the financial crisis has led to the question, 'Were auditors asleep at the wheel?' (Doogar, Rowe, and Sivadasan 2015). This failure to warn stakeholders arguably contributed to the financial crisis which resulted in updates to how auditors evaluate potential bank failures (Basel 2008; Kroeker 2011). In this study we provide evidence that, even though auditors failed to issue warnings that banks might fail, auditors were optimizing the information set they had at the time. We also examine how auditor behavior changed following the financial crisis. More specifically, we provide evidence that auditors became more conservative, i.e., auditors were more willing to make Type I GCO errors with their bank clients following the financial crisis.

Type I errors for banks are important for at least two reasons: First, the effects of bank failure are pervasive and extend to that bank's clients. Banks provide financing for public and private companies of all sizes in all industries. When a company's lender fails, the company is faced with reduced liquidity and increases in the cost of debt (Krishnamurthy 2010). Second, the receipt of a GCO can have a more severe "self-fulfilling prophecy" effect for bank clients compared to other industry clients. A GCO can affect a bank's regulatory capital and can reduce depositor confidence, resulting in liquidity problems that may further precipitate the bank's failure (Basel 2014). With this in mind, auditors might be more reluctant to issue a GCO to a bank client that can prematurely lead to failure.

We also argue that auditors' reluctance to issue a GCO to a bank client should have shifted following the onset of the financial crisis. As banks became riskier clients and their auditors faced greater regulatory and market scrutiny, auditors focused more resources and attention on bank clients (Cassell, Hunt, Narayanamoorthy, and Rowe 2019). As such, we expect to observe an increase in auditor conservatism following the financial crisis.

In this paper we provide evidence that auditors optimized their information set during and after the financial crisis when making going concern determinations for bank clients. We also provide evidence that auditors responded appropriately to regulatory and market scrutiny by increasing conservatism in issuing GCOs to bank clients following the financial crisis.

To evaluate auditors' ability to accurately predict bank failures, we compare the auditors' bank failure prediction, i.e., going concern opinions, and our own bank failure prediction models. We implement state-of-the-art bank failure prediction models using machine learning techniques specifically designed for anomaly detection. We incorporate publicly available data as well as information from call reports which are due just 30 days following the bank's fiscal year end and are used by regulators to identify troubled banks. To predict bank failure, we use two one-class learning models, Isolation Forests (IF) and One-Class Support Vector Machines (OCSVM). One-class learning allows the model to learn using data with only one outcome instead of two or more outcomes. In our setting, bank failures are anomalous, meaning their occurrence is extraordinarily rare. Our use of these algorithms that allow for the prediction of anomalous events is crucial to this study because more traditional prediction models are not designed to predict such rare events.

We find that auditors are better than our machine learning techniques at predicting bank failure during and after the financial crisis. Furthermore, we find that auditors increase their conservatism, i.e, increase their Type I GCO errors for bank following the financial crisis. This increase in conservatism translates to a significant increase in the number of bank failures captured by auditors in the post-financial crisis period. This indicates that auditors responded to the increased regulatory and market scrutiny. Finally, we evaluate whether our machine learning model can improve the accuracy of auditor GCOs. When we combine our bank failure prediction with auditor GCOs, we capture the same number of bank failures in the post-financial crisis period while simultaneously reducing Type 1 errors.

Our findings extend prior literature that evaluates the competence of auditors during the financial crisis. We focus on the unique regulatory environment of the banking industry and the unique challenges involved in auditing bank clients during and after the financial crisis, and we provide evidence that auditors became more conservative regarding the issuance of bank GCOs in the post-financial crisis period. We also contribute to the accounting literature by introducing powerful anomaly detection techniques that are ideal for examining extremely rare events. We contribute to the going concern opinion literature by developing a machine learning model that, when combined with GCOs, reduces Type 1 errors while maintaining the same number of correct GCOs, i.e., GCOs that preceded a bank failure.

The rest of the paper proceeds as follows. Section II provides background and develops our hypotheses. Section III presents our research methodology and sample selection. We tabulate and discuss our results in Section IV, perform additional analyses in Section V, and conclude in Section VI.

II. BACKGROUND AND DEVELOPMENT OF HYPOTHESES

Bank Failures and the Financial Crisis

While any corporate failure has negative economic consequences, bank failures have a particularly large negative effect (Kupiec and Ramirez 2013). Banks provide financing for many businesses. A bank failure reduces liquidity and increases the cost of debt for these businesses (Krishnamurthy 2010). Kandrac (2013) even finds that bank failures lead to lower income and compensation growth, higher poverty rates, and lower employment. Thus, widespread bank failure that took place during the financial crisis had pervasive effects. The problem began with failing banks but soon permeated to other industries through various ties to financial institutions.

At the beginning of the financial crisis, bank failures were unexpected, even among private banks. Before the financial crisis, there were multiple years with no bank failures, followed by sharp increases in failures. In 2005 and 2006 there were zero bank failures, while in 2007 and 2008 there were 3 and 25 respectively. Then, during the height of the crisis in 2009 and 2010, there were 140 and 157 bank failures, respectively.[1] When considering only public banks, we see a similar trend (see Table 1). Auditors did not issue GCOs for many of these bank failures and thus received bad press for their role in the financial crisis (IAG 2011) and regulatory agencies called for auditors to change how banks are audited (Casey 2009; FDIC 2016). Auditors responded by making changes to how they audit (Basel 2008; Kroeker 2011), and regulators made changes to how certain risks are measured. For example, the Dodd-Frank Act of 2010 requires banks to perform stress tests. The FDIC changed how it adjusts capital, asset quality, management, earnings, liquidity and market sensitivity (CAMELS) ratings in 2011, and clarified definitions for higher risk assets in 2012.

Bank Going Concern Opinions

As part of issuing the audit opinion, auditors are required to opine on whether there is substantial doubt that an entity will continue as a going concern during the twelve months after the 'as of' date on the financial statements. If the auditor believes there is substantial doubt that the client will continue as a going concern, auditors add explanatory language to the audit opinion, which is a going concern opinion. Bank going concern opinions have more severe

[1] These bank failure counts include public and private commercial banks. See https://www.fdic.gov/bank/historical/bank/index.html. Our sample consists of public commercial banks with available data.

implications than those in other industries. This is because a GCO can affect a bank's regulatory capital and can reduce depositor confidence, resulting in liquidity problems that further precipitate the bank's failure (Basel 2014). For example, issuing a GCO means a bank must reduce asset values to reflect the collectability of the assets (e.g., writing-off deferred tax assets or impairing goodwill and other intangible assets) - which is much lower when bank failure is likely – which can affect Tier 1 capital ratios. Bank regulators require minimum thresholds for the capital ratios (typically 4% of equity capital divided by the total risk-weighted assets). Also, the FDIC must pay out on the insured portion of depositor balances (Basel 2014). Consistent with the notion that bank going concern opinions have more severe implications, Albrecht, Glendening, Kim, and Pereira (2020) find that auditors are less likely to issue a going concern opinion to systemically risky banks which reduces Type I errors, a behavior that is more pronounced during the financial crisis.

In summary, it is unclear whether, or to what extent, auditors will modify their likelihood of issuing GCOs after the financial crisis. On one hand, auditors are more likely to issue GCOs in order to avoid the negative attention they received surrounding the financial crisis. On the other hand, auditors are unlikely to issue a GCO that can lead to premature bank failure. Thus, we phrase our first hypothesis in null, as follows:

H1: *Regulatory and market scrutiny had no effect on the auditors' propensity to issue a going concern opinion during and after the financial crisis.*

Predicting Bank Failure

Studies that provide bank failure prediction models generally measure their success in terms of accuracy in predicting failed banks both during and outside of financial crises and most use variables that proxy for capital, asset quality, management, earnings, liquidity and market sensitivity (CAMELS). Meyer and Pifer (1970) use a set of financial ratios to predict bank failure. Thomson (1991) uses a single-equation model with many CAMELS proxies to predict bank failures. Wheelock and Wilson (2005) also associate bank failures before financial crisis with a number of CAMELS rating categories. Estrella, Park, and Perstiani (2000) use capital ratios to forecast U.S. bank failures occurring from 1988 to 1993. More recently, Cole and White (2012) find a significant association between equity to assets, non-performing assets to assets, cash to assets, brokered deposits to assets, investment securities to assets, and ROA and bank failure. Ng and Roychowdhury (2014) associate loan loss reserves with the probability of bank failure. Kosmidou and Zopounidis (2008) also predict bank failures from 1993-2003 using CAMELS components and a multicriteria approach.

Still others contribute to this literature by examining additional variables and how they affect the predictability of bank failure. Jin, Kanagaretnam, and Lobo (2011) examine the predictability of bank failure using accounting and audit quality variables. Aubuchon and Wheelock (2010) find a correlation between regional economic factors and bank failure in their 2007-2010 sample period. Kanagaretnam, Lim, and Lobo (2014) find that individualism and cultures that encourage higher risk-taking are associated with bank failure and troubled banks during the financial crisis. Barth and Landsman (2010) find a significant association between loan loss provisions and bank health. Akins, Li, Ng, and Rusticus (2016) find that less competition among banks is associated with engagement in risky activities, regulatory intervention, and ultimately bank failure. Clearly and Hebb (2016) predict bank distress using variables from prior literature known to be correlated with bankruptcy and bank failure including Altman (1968) and Collier, Forbush, Nuxoll, and O’Keefe (2005).

Several papers also examine various methods used to model bank failure. Bell (1997) compares logistic regression and neural networks in their respective effectiveness at predicting bank failures and find that each model performs similarly but neural networks are better at predicting bank failure when the bank is only in marginal distress compared to logistic regression. Jin, Kanagaretnam, and Lobo (2011) use logistic regression to examine whether accounting and audit variables are useful in predicting bank failure. Aubuchon and Wheelock (2010) examine geographic patterns in bank failure and find that many bank failures, especially small banks, are reflecting local economic conditions.

We extend prior literature that examines bank failure by using a state-of-the-art machine learning model that specializes in anomaly detection to predict bank failure. We want to benchmark auditors' GCOs (bank failure predictions) against our anomaly predictions. Auditors possess not only expertise in bank audits but also have "soft" information about their bank clients that is not disclosed in financial statements. This includes an understanding of the bank's processes, strategic decisions, management integrity, and so forth, which is required of auditors by the PCAOB when considering an entity's ability to continue as a going concern.[2] Auditors may also gain contextual knowledge through the process of obtaining, evaluating, and documenting management's plans to mitigate any substantial doubt about the company's ability to continue as a going concern. This contextual knowledge improves auditors' bank failure predictions in immeasurable ways.

However, machine learning might outperform auditor bank failure prediction because these models allow for non-linear relationships that a human could not anticipate. This non-linearity may allow for greater accuracy in prediction. Machine learning methods require fewer

[2] PCAOB Audit Standard 2415 – *Consideration of an entity's ability to continue as a going concern*.

assumptions about the data generating process (Mullainathan and Spiess, 2017). Consequently, they can anticipate unidentified relationships (Gu et al. 2020).

This leads to our second hypothesis, stated in the null:

H2: *Bank failure prediction accuracy is similar between auditors and machine learning methods.*

Combining the strengths of auditor's experience and access to "soft" data with the ability to detect non-linear relationships through machine learning should result in more accurate predictions. Thus, our third hypothesis, stated in the alternate is:

H3: *Combining auditor and machine learning bank failure predictions improves going concern opinion accuracy*.

III. METHODOLOGY AND SAMPLE

Methodology

We have two data sets, training and testing. The training data uses 3-year rolling windows starting in 2003 to 2018 and the testing data set begins in 2006 to 2019.[3] Our outcome variable is *FAILURE,* an indicator variable equal to one if the bank fails within 15 months after the calendar year-end (i.e., within about 12 months of the report date) as measured by the FDIC, and zero otherwise.

We test H1 by comparing the proportion of issued going concern opinions during and after the financial crisis. We test the overall proportion as well as the occurrence of Type I errors (false positives) using a chi-squared test.

[3] Results are qualitatively similar if we use chaining training widows.

For H2 we compare two machine learning models to auditors' expectations. We compare auditors' bank failure predictions to what bank fundamentals would predict. Typical multivariate regression techniques are ill-suited for this setting due to the low occurrence of bank failures. We compare the two machine learning models' best predictions to the auditors' going concern opinions. We limit the number of predictions to the equivalent number of going concern opinions. For example, for the full sample period, we examine the models' top 127 best predictions because there are approximately 127 going concern opinions. For the financial crisis period, we examine the models' top 47 best predictions. For the period after the financial crisis, we examine the models' top 80 best predictions.

For H3, we combine the machine learning models predictions with auditors going concerns. We do this by keeping only the going concerns and then within in the going concern expectations we optimize the model's trade-off between sensitivity and specificity.

Bank Failure Predictors

We borrow from prior literature to identify predictors of bank failure (Jin, Kanagaretnman, and Lobo 2011; Cleary and Hebb 2016; Masli, Porter, and Scholz 2018). We include predictors for bank size including the log of total assets (*SIZE*) and then the raw amounts of total assets (*TOTAL_ASSETS*), loans (*TOTAL_LOANS)* and the net amount of loans *(NET_LOANS).* We include predictors for overall bank performance, such as the risk-weighted Tier 1 capital ratio (*CAP*), the ratio of securitized loans divided by total loans (*PSLOANS*), and the ratio of non-performing loans divided by total assets (*NPL*). We also include other predictors for bank health such as the ratio of off-balance sheet amounts divided by total assets (*OFF_BS*), the amount of unearned income (*UNEARN_NI*), total interest income (*TOT_INT_INC*), total non-interest income (*TOT_NONINT_INC*), net income (*NET_INCOME*), total charge offs (*TOT_CHRG*), total loan loss provision (*LLP*), net income before discontinued operations

(*NI_PREOPS*), and other real estate owned (*OREO*). Other predictors of bank health include the ratio of cash to total assets (*CA_TA*), the ratio of retained earnings to total assets (*RE_TA*), the ratio of total equity to total assets (*EQ_TA*), net income divided by total assets (*ROA*), net loans divided by total assets (*LOAN1*), net loans divided by total deposits and short-term funding (*LOAN2*), net loans divided by total deposits and borrowings (*LOAN3*). We also include predictors of loan quality such as the loan loss reserve divided by total loans (*LOANQUAL1*), net charge-offs divided by total loans (*LOANQUAL2*), and the total risk-weighted capital ratio (*TOTAL_CAP*).

We also include predictors for the mix of bank loans including the total amount of commercial and industrial loans (*CILOANS*), ratio of commercial and industrial loans to lagged total assets (*GCOMM*), the total amount of all other real estate assets (*AO_RE*), growth in all other real estate divided by lagged total assets (*GRESTATE*), growth in total loans divided by lagged total assets (*GLOANS*), and the sum of commercial and industrial loans, direct lease financing, all other real estate loans, agriculture loans, and foreign loans, divided by total loans (*LOAN_MIX*). We include a measure for stock market participant expectations with the 12-month stock return for the calendar year (*STOCK_RETURN*).

We also include a set of indicator variables for the words used in non-confidential narrative disclosures from any call report in the last three calendar quarters and for which quarter the words were disclosed (see full list in Table 2).[4] One advantage of using narrative disclosures in call reports is that such information allows us to capture, to some extent, the context around the call report information. Also, the information from the call reports are required to be submitted within 30 days of each quarter end, while financial statements have later deadlines.

[4] We only include sparse words that occur in approximately 0.005 percent of the sample, which is approximately the occurrence rate of bank failures for all call reports before we reduce the sample due to variable construction.

Finally, we include percentage changes in all continuous variables. This allows us to capture how changes in the variables can alter predictions. Many of our predictors capture different yet related aspects of bank health.

Unsupervised Models

We form predictions using Isolation Forest and One-Class Support Vector Machine algorithms. Isolation forest (IF) is an unsupervised model. Unsupervised models are not trained with targets. The IF model is built with decision trees. The IF model was designed around the idea that anomalies represent data that are different and rare. IF randomly samples the data based on randomly selected inputs. Observations that move deeper into the trees are less likely to be an anomaly because they require more branches to isolate them. Observations that end on short branches are more likely to be anomalies because they are easier to separate.

Isolation forests are an ensemble of decision trees. Each tree is an isolation tree. The algorithm begins by building isolation trees. Given a dataset, the model generates a random sub-sample of the data for each isolation tree. Branching of the tree starts by selecting a random input and splitting on a random threshold. This process continues until each observation is completely isolated. The isolation trees are combined into an Isolation forest (Liu et al. 2008). During prediction, the data must travel down each isolation tree. An anomaly score is assigned to each of the observations based on how deep it goes into the tree. The score is aggregated among all of the trees.

One class support vector machine (OCSVM) is essentially a standard two-class support vector machine (SVM) model with one class representing the training data (not containing any anomalous events) and the other class representing the origin of a feature space. See Krupa and Minutti-Meza (2021) for more details on SVM. OCSVM model maps inputs into a feature space and maximizes the margin between the inputs and the origin. This hyperplane separates the two classes. OCSVM uses a function (kernel) to allow for a non-linear separation between classes. In practice there is a tradeoff between maximizing the distance between class and the observations that fall within that separation. Ideally, we would like for the two classes to be completely separated with nothing falling in the hyperplane, but practically speaking this is not likely. Weighted points or support vectors solely determine the maximum separation between origin and data points. The model allows for a degree of error. (Schölkopf 1999, Heller 2003).

Sample Selection and Descriptive Statistics

We obtain commercial bank data from call reports, bank holding company data from FR Y-9C reports, and auditor information from Audit Analytics. We tie each commercial bank to its highest-level bank holding company, if any, and then to the audit opinion in Audit Analytics. We obtain other financial statement and stock return data from Compustat. We obtain bank failure information from the FDIC failed bank list and data related to the Capital Purchase Program (CPP) and the Community Development Capital Initiative (CPCI) from the US Treasury website. Our sample period is from 2003 through 2019. We keep observations that have all required predictor variables. Our final sample consists of 8,504 commercial bank-year level observations.

In Table 1, we present the frequency of bank failures and going concern opinions, which on average throughout the samples period are quite rare. During the first training period (2003-2005), there are no bank failures, followed by 52 bank failures over a span of 6 years. Typical prediction methods, such as logistic regression, would not be able to form a prediction with such a rare event. Consistent with the notion that auditors became more conservative following the financial crisis, the univariate data show that, before the financial crisis in 2008, auditors seemed to be reluctant to issue GCOs to bank clients. Starting in 2009, GCOs became more frequent for bank clients.

(INSERT TABLE 1 HERE)

We present descriptive statistics for the predictor variables in our sample in Table 2. Some of the mean values for our descriptive statistics are well outside the interquartile range, suggesting outliers (e.g., *TOTAL_ASSETS*, *TOTAL_LOANS*, *CAP*, etc.). However, one of the benefits of our machine learning models is that outliers are less likely to affect the predictions.

(INSERT TABLE 2 HERE)

In Table 3 we present differences in means grouped by failure years and non-failure years. Statistical differences between these means suggests that our variables have the capacity to discriminate between failure years and non-failure years. We find that many of our variables are significantly different between failure years and non-failure years. In fact, all financial statement variables are significantly different between the two groups with a p-value of at least $p < 0.10$ except for the loan loss provision (*LLP*) and other real estate owned (*OREO*). That is, we find that banks that fail have significantly lower stock returns (*STOCK_RETURN*), are smaller (*TOTAL_ASSETS*, *TOTAL_LOANS*, *NET_LOANS*, *CILOANS*, *AO_RE*, *DIRECT_LEASEFIN*, *SIZE*), have greater proportion of riskier assets (*GCOMM*, *LOANMIX*) and performed poorly (*GLOANS*, *GRESTATE*, *RE_TA*, *ROA*, *LOANQUAL1*, *LOANQUAL2*, *NET_INCOME_PRE_OPS*). The narrative disclosure elements suggest that earlier and more disclosure is consistent with non-failure years. That is, if a bank adds narrative disclosure on the call report in Q2 of the fiscal year, the bank is less likely to fail. If a company includes more rare words in the disclosure, that bank is less likely to fail in the next year. Altogether, this suggests that the variables in our models should have ability to discriminate between failure years and non failures.

(INSERT TABLE 3 HERE)

IV. RESULTS

Table 4 presents results related to H1. We find that although auditors did not significantly increase their issuance of going concern opinions, they did significantly increase their Type 1 errors. These findings suggest that auditors became more conservative in the post financial crisis period.

(INSERT TABLE 4 HERE)

Table 5 presents results related to H2. We form predictions using IF and OCSVM and compare those predictions to auditors' predictions over the entire sample period. We specifically examine the top 127 best guesses for each machine learning model. We find that IF and OCSVM perform fairly poorly. IF predicts 9.6% and OCSVM predicts 0% of actual bank failures. Auditors correctly predict 71.2% of bank failures (37 of 52 total failures). Auditors only missed 15 actual bank failures.

We acknowledge that our models' probability cut-off to classify a prediction as a bank failure may be somewhat arbitrary. Thus, we also examine the best predictions our models can form when we choose the predicted probability that correctly predicts the most bank failures and correctly predicts the most non-bank failures. Here, we see that IF and OCSVM predict 98.1% of bank failures (51 of 52 failures), missing one. These predictions correctly classify nearly all bank failure years, but this comes at a cost. IF has a 26.4% (1,408 out of 5,331 bank-years) false positive rate, and OCSVM has a 27.3% (1,460 out of 5,331 bank-years) false positive rate.

Altogether, this suggests that auditors outperform machine learning methods when predicting bank failures over the sample period. This is surprising and suggests that auditors' contextual knowledge and soft information is very important in predicting bank failures. Call report information alone performs rather poorly in the highly regulated bank setting.

(INSERT TABLE 5 HERE)

Next, we examine our models and auditors' top 47 best guesses through the financial crisis (from 2007 through 2009) in Table 6 panel A. We find that IF correctly predicts 4 of 41 failures, missing 37 bank failures. OCSVM has 1 correct bank failure prediction and misses 40 bank failures. During this period, auditors correctly predict 65.8% of bank failures (27 out of 41 failures), missing 14 bank failures. Auditors have 20 false positives compared to IF's 43 and SVM's 46. The best model optimizing sensitivity and specificity for IF and OCSVM predicts 95.0% and 97.5% of bank failures, respectively. These models establish our baseline predictions through the financial crisis. It suggests that auditors performed better than prediction models that incorporate information from call reports and financial statements.

In Table 6 panel B, we present results on how auditors perform after the financial crisis (2010 through 2019). We examine the top 80 best guesses for each model because auditors issued 80 going concern opinions during this period. Here, the prediction models perform even more poorly, and auditors improve their performance. That is, IF and OCSVM correctly predicted 0 and 1 of the 11 bank failures and predicted 80 and 79 false positives. Auditors perform much better, accurately predicting 90.9% of bank failures (10 of 11 failures), missing one. Similar to the results for the financial crisis period, the best models optimize sensitivity and specificity. IF and OCSVM capture 10 and 11 of the bank failures, respectively, but yield 553 and 897 false positives (16.1% and 26.1%), respectively.

In summary, auditors accurately predicted 65.8% (i.e., 27 out of 41) of bank failures during the financial crisis and accurately predicted 90.9% (10 out of 11) of bank failures after the crisis. IF correctly predicts 4 bank failures, and SVM correctly predicts 0 bank failures during the financial crisis. After the crisis, they predicted 0 and 1 bank failure. Overall, this provides further support that auditors perform better than prediction models that incorporate information from call reports and financial statements.

(INSERT TABLE 6 HERE)

Table 7 presents the results for H3. Within the going concern predictions we optimize the sensitivity and specificity of the models and examine their accuracy. We first show the full sample. IF captures 32 of the 37 bank failures and reduces the false positives from 90 to 52. OCSVM captures 31 of the 37 bank failures and reduces the false positives from 90 to 46. During the financial crisis the IF captures 20 of the 27 bank failures and reduces the false positives from 20 to 6, while OCSVM captures 26 of the 27 bank failures and reduces the false positives from 20 to 12. Post crisis each algorithm captures all of the bank failures and reduces the false positives from 70 to 43 and 36 for IF and OCSVM. Taken together these results show that combining machine learning with auditor expertise yields consistent true positives and reduces false positives yielding higher accuracy. We believe that auditors can combine these models with their expertise and inside information to reduce their overall false positive rate while still correctly predicting bank failures.

(INSERT TABLE 7 HERE)

V. ADDITIONAL ANALYSES

In additional analysis we add the going concern to the models. In table 8 we find that the models perform similarly with respect to out-of-sample accuracy. In table 9 panel A we also drop all bailout banks from our sample and find similar results. In table 9 Panel B we run our models on distressed only banks and find that they perform qualitatively similar but have a much lower false positive rate. In table 10 we show that the model fit between the models with and without GCO added are qualitatively similar.

VI. CONCLUSION

In this paper we evaluate GCOs during and after the financial crisis. We develop a bank failure prediction model using machine learning techniques designed specifically for anomaly detection. Our machine learning bank failure prediction model allows us to 1) predict bank failures with greater accuracy than those in prior literature, 2) benchmark auditor performance during and following the financial crisis, 3) provide a tool that we believe would be beneficial to auditors in more accurately capturing bank failures with their GCOs.

We use IF and SVM machine learning techniques to predict anomalous bank failures. We compare our machine learning-based bank failure predictions to auditors' GCOs during and after the financial crisis and find that auditors consistently outperform our model. This provides evidence that auditors private information and expertise give them a competitive advantage in predicting bank failure. This is also indicative of auditors optimizing their information set to form the most accurate predictions possible regarding their bank clients' likelihood of failure.

Next, we evaluate auditor conservatism during and after the financial crisis. Given the increased scrutiny from the market and regulators during the financial crisis, we expect auditors to respond with increased conservatism in the post-financial crisis period. To test this, we examine auditors' propensity to correctly predict bank failure as well as their tolerance for Type I GCO errors in the post-financial crisis period. We find that auditors capture ten out of eleven bank failures in the post-financial crisis period, a sharp increase from the financial crisis period where very few bank failures were preceded by a GCO. We also see a sharp increase in Type I

GCO errors in the post-financial crisis period, which is indicative of an increase in auditor conservatism.

Finally, we explore whether adding our machine learning techniques to auditor GCOs that inherently incorporate auditor private information and expertise can produce an even better bank failure prediction than GCOs alone. Once combined, we find that our machine learning models enhance the accuracy of auditor GCOs by reducing Type I errors by about 50% while maintaining the same number of correct bank failure predictions (i.e. GCO precedes a bank failure) in the post-financial crisis.

This study provides three primary contributions. First, we introduce anomaly detection machine learning techniques that more accurately predict bank failure . These techniques allow us to benchmark auditor performance and introduce new machine learning techniques to the accounting literature that are useful in overcoming the inherent limitations related to testing anomalous economic events. Second, we provide evidence suggesting that auditors optimized their information set in their decision to issue GCOs, i.e. bank failure predictions, during and after the financial crisis while increasing their conservativeness in response to increasing investor and regulator scrutiny. Third, we provide a bank failure prediction model tool that should be useful to auditors and regulators interested in more accurate GCOs for bank clients.

REFERENCES

Akins, B., L. Li, J. Ng, and T. O. Rusticus. 2016. Bank Competition and Financial Stability: Evidence from the Financial Crisis. *Journal of Financial and Quantitative Analysis* 51(1): 1 – 28.

Albrecht, A., M. Glendening, K. Kim, and R. Pereira. 2020. Auditor Going Concern Opinions and Bank Systemic Risk: Evidence from the 2007-2009 Financial Crisis. *Auditing: A Journal of Practice and Theory* 39(3): 1 – 28.

Altman, E. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance* 23(4): 589 – 609.

Aubuchon, C. P., and D. C. Wheelock. 2010. The Geographic Distribution and Characteristics of U.S. Bank Failures, 2007-2010: Do Bank Failures Still Reflect Local Economic Conditions? *Federal Reserve Bank of St. Louis Review* 92(5): 395 – 415.

Barth, M. E., and W. R. Landsman. 2010. How did Financial Reporting Contribute to the Financial Crisis? *European Accounting Review* 19(3): 399 – 423.

Basel Committee on Banking Supervision (BCBS). 2008. *External audit quality and banking supervision*. Available at https://www.bis.org/publ/bcbs146.pdf.

Basel Committee on Banking Supervision (BCBS). 2014. *External audits of banks*. Available at: https://www.bis.org/publ/bcbs280.pdf.

Beams, J., and Y. Yan. 2015. The Effect of Financial Crisis on Auditor Conservatism: U.S. Evidence. *Accounting Research Journal* 28(2): 160–171.

Bell, T. 1997. Neural Nets or the Logit Model? A Comparison of Each Model's Ability to Predict Commercial Bank Failures. *Intelligent Systems in Accounting Finance and Management* 6: 249 – 264.

Casey, K. 2009. Speech by SEC Commissioner: Lessons from the Financial Crisis for Financial Reporting, Standard Setting and Rule Making. Accessed at https://www.sec.gov/news/speech/2009/spch111709klc.htm, retrieved July 27, 2021.

Clearly, S., and G. Hebb. 2016. An Efficient and Functional Model for Predicting Bank Distress: In and Out of Sample Evidence. *Journal of Banking and Finance* 64: 101 – 111.

Cole, R. A., and L. J. White. 2012. Déjà Vu All Over Again: The Causes of U.S. Commercial Bank Failures *This* Time Around. *Journal of Financial Services Research* 42: 5–29.

Collier, C., S. Forbush, D. A. Nuxoll, and J. O'Keefe. 2005. The SCOR System of Off-Site Monitoring: Its Objectives, Functioning, and Performance. *FDIC Banking Review* 15(3): 17 – 32.

Deloitte. 2015. Cognitive Technologies: The Real Opportunities for Business. Accessed at https://www2.deloitte.com/us/en/insights/deloitte-review/issue-16/cognitive-technologies-business-applications.html.

Dickey, G., S. Blanke, and L. Seaton. 2019. Machine Learning in Auditing. *The CPA Journal* Accessed at https://www.cpajournal.com/category/magazine/june-2019-issue/.

Doogar, R., S. Rowe, and P. Sivadasan. 2015. Asleep at the Wheel (Again)? Bank Audits During the Lead-Up to the Financial Crisis. Contemporary Accounting Research 32(1): 358-391.

Estrella, A., S. Park, and S. Peristiani. 2000. Capital Ratios as Predictors of Bank Failure. FRBNY Economic Policy Review July 2000.

Federal Deposit Insurance Corporation (FDIC). 2016. FDIC Law, Regulations, Related Acts. Statements of Policy, Auditing Programs. Service Location 5000-5370.

Gutierrez, Krupa, Minutti-Mezza, and Vulcheva. 2020. Do Going Concern Opinions Provide Incremental Information to Predict Corporate Defaults? *Review of Accounting Studies* 25: 1344–1381.

Investor Advisory Group (IAG). 2011. "Report from the working group on: Lessons learned from the financial crisis." PCAOB. Available online at https://pcaobus.org/News/Events/Documents/03162011_IAGMeeting/ The_Watchdog_That_Didnt_Bark.pdf, retrieved July 27, 2021.

Jin, J. Y., K. Kanagaretnam, and G. J. Lobo. 2011. Ability of accounting and audit quality variables to predict bank failure during the financial crisis. *Journal of Banking and Finance* 35: 2811 – 2819.

Kanagaretnam, K., C. Y. Lim, and G. J. Lobo. 2014. Influence of National Culture on Accounting Conservatism and Risk-Taking in the Banking Industry. *The Accounting Review* 89(3): 1115 – 1149.

Kandrac. 2013. Bank Failure Relationship Lending, and Local Economic Performance. SSRN.

Kepes, B. 2016. Big Four Accounting Firms Delve into Artificial Intelligence. *Computerworld* Accessed at https://www.computerworld.com/article/3042536/big-four-accounting-firms-delve-into-artificial-intelligence.html.

Kosmidou, K., and C. Zopounidis. Predicting US commercial bank failures via a multicriteria approach. *International Journal of Risk Assessment and Management* 9(1): 26 – 43.

Krishnamurthy, A. 2010. How Debt Markets Have Malfunctioned in the Crisis. *Journal of Economic Perspectives* 24(1): 3–28.

Kroeker, J. 2011. Testimony Concerning the Role of the Accounting Profession in Preventing another Financial Crisis. Accessed at https://www.sec.gov/news/testimony/2011/ ts040611jlk.htm on August 26, 2021.

Kupiec, P., and C. Ramirez. 2013. Bank Failures and the Cost of Systemic Risk: Evidence from 1900-1930. *Journal of Financial Intermediation* 22(3): 285-307.

Lennox, C. S. 1999. The Accuracy and Incremental Information Content of Audit Reports in Predicting Bankruptcy. *Journal of Business Finance and Accounting* 26(5): 757 – 778.

Masli, A., C. Porter, and S. Scholz. 2018. Determinants of Auditor Going Concern Reporting in the Banking Industry. *Auditing: A Journal of Practice and Theory*. 37(4): 187 – 205.

Meyer, P. A., and H. W. Pifer. 1970. Prediction of Bank Failures. *The Journal of Finance* 25(4): 853–868.

Ng, J., and S. Roychowdhury. 2014. Do Loan Loss Reserves Behave Like Capital? Evidence from Recent Bank Failures. *Review of Accounting Studies* 19: 1234–1279.

Persico, F., and H. Sidhu. 2017. How AI will Turn Auditors into Analysts. *Accounting Today* Accessed at https://www.accountingtoday.com/opinion/how-ai-will-turn-auditors-into-analysts.

Thomson, J. 1991. Predicting Bank Failures in the 1980s. *Economic Review of the Federal Reserve Bank of Cleveland* 27(1): 9–20.

Wheelock, D. C., and P. Wilson. 2005. The Contribution of On-site Examinations Ratings to an Empirical Model of Bank Failures. *Review of Accounting and Finance* 4(4): 110–133.

Xu, Q. and R. Kalelkar. 2020. Consequences of Going-Concern Opinion Inaccuracy at the Audit Office Level. *Auditing: A Journal of Practice & Theory* 39(3): 185–208.

Table 1: Frequency

This table presents the frequency of bank failures and auditor bank going concern opinions from 2003 through 2019.

Year	*Failure*	*Going_Concern*
2003	0	1
2004	0	0
2005	0	0
2006	0	0
2007	3	0
2008	22	16
2009	16	31
2010	5	35
2011	1	19
2012	5	17
2013	0	3
2014	0	1
2015	0	4
2016	0	1
2017	0	0
2018	0	0
2019	0	0
Total	52	128

Table 2: Descriptive Statistics

This table presents descriptive statistics from 2006 through 2019 for variables in our testing dataset. We also include the percentage change in our non-text variables, but which are omitted from the table.

Statistic	Mean	St. Dev.	P(25)	Median	P(75)
STOCK_RETURN	0.829	63.195	-0.176	0.027	0.204
TOTAL_ASSETS	20,069,685.000	136,964,530.000	491,459.5	1,044,273	3,189,023.0
TOTAL_LOANS	10,476,509.000	65,454,854.000	329,214.5	728,520	2,227,703.0
NET_LOANS	10,109,906.000	63,129,015.000	320,735	706,828	2,151,054
CAP	0.135	0.658	0.102	0.118	0.139
PSLOANS	0.702	0.179	0.596	0.737	0.840
NPL	0.016	0.023	0.004	0.01	0.02
CLOANS	2,262,245.000	14,202,847.000	30,952	90,733	330,729
GCOMM	0.118	0.093	0.057	0.094	0.152
AO_RE	5,426,444.000	32,289,814.000	255,882.5	548,346	1,524,129
GRESTATE	0.051	0.166	-0.006	0.025	0.070
GLOANS	0.067	0.210	-0.004	0.038	0.090
DIRECT_LEASEFIN	205,021.000	1,442,204.000	0	0	1,190
LOAN_MIX	0.913	0.118	0.893	0.954	0.980
SIZE	14.185	1.738	13.105	13.859	14.975
CA_TA	0.051	0.057	0.021	0.032	0.059
RE_TA	0.031	0.050	0.016	0.036	0.057
EQ_TA	0.106	0.033	0.087	0.101	0.120
ROA	0.005	0.017	0.004	0.008	0.011
LOAN1	0.663	0.132	0.600	0.680	0.751
LOAN2	1.519	22.648	0.747	0.861	0.958
LOAN3	0.920	4.021	0.707	0.803	0.883
LOANQUAL1	0.016	0.010	0.010	0.013	0.018
LOANQUAL2	0.007	0.015	0.001	0.002	0.007
TOTAL_CAP	0.148	0.658	0.115	0.131	0.151
OFF_BS	0.158	0.350	0.053	0.095	0.165
UNEARN_NI	1,206.176	10,074.420	0	0	98
TOT_INT_INC	651,057.200	4,142,538.000	22,563	47,307	138,778.5
TOT_NONINT_INC	352,852.800	2,618,257.000	2,688.5	7,478	30,042

NET_INCOME	175,314.600	1,474,825.000	2,001.5	7,078	24,797
TOT_CHRG	99,529.950	798,159.500	650	2,635	12,122
TOT_INT_EXP	144,917.900	1,199,282.000	4,132	10,758	30,708.5
TOT_NONINT_EXP	529,467.800	3,529,193.000	12,937	28,833	84,559.5
LLP	85,444.440	798,674.200	504	2,360	10,595.5
NET_INCOME_PRE_OPS	175,481.200	1,475,578.000	2,003	7,082	24,821.5
OREO	25,432.730	183,214.800	298	1,962	8,101
DISTRESS	0.138	0.345	0	0	0

Text Variables

Statistic	**Mean**	**St. Dev.**	**P(25)**	**Median**	**P(75)**
Q2DISCLOSURE	0.006	0.080	0	0	0
Q3DISCLOSURE	0.004	0.065	0	0	0
Q4DISCLOSURE	0.032	0.176	0	0	0
NONCONFIDENTIAL	0.054	0.227	0	0	0
CALL	0.022	0.223	0	0	0
DEPOSITS	0.014	0.155	0	0	0
EFFECTIVE	0.009	0.097	0	0	0
FDIC	0.011	0.144	0	0	0
REPORT	0.022	0.211	0	0	0
LOAN	0.012	0.163	0	0	0
LOANS	0.018	0.186	0	0	0
QUARTER	0.012	0.123	0	0	0
ASSETS	0.011	0.146	0	0	0
BANK	0.047	0.420	0	0	0
BANKS	0.016	0.199	0	0	0
DUE	0.012	0.131	0	0	0
MILLION	0.017	0.176	0	0	0
CAPITAL	0.017	0.247	0	0	0
SCHEDULE	0.011	0.116	0	0	0
LINE	0.010	0.134	0	0	0
REPORTED	0.008	0.097	0	0	0
DECEMBER	0.010	0.128	0	0	0

Table 3: Difference in Means

This table presents the difference in means for variables in our testing dataset. The data are separated by non-failure and failure bank-years.

Variables	*Avg. for No Failures*	*Avg. for Failures*	*Diff*	*P-value*	*Sig.*	*t-stat*
STOCK_RETURN	0.841	-0.614	>	0.0768	.	1.77
TOTAL_ASSETS	20226432.634	1962266.231	>	<1e-06	***	9.814
TOTAL_LOANS	10555389.963	1364199.269	>	<1e-06	***	9.985
NET_LOANS	10186660.073	1243330.865	>	<1e-06	***	10.206
CAP	0.135	0.06	>	<1e-06	***	8.086
PSLOANS	0.701	0.827	<	1.97e-06	***	-5.354
NPL	0.015	0.083	<	<1e-06	***	-9.974
CILOANS	2280596.62	142303.596	>	<1e-06	***	11.386
GCOMM	0.119	0.08	>	0.00023	***	3.955
AO_RE	5463384.824	1159036.135	>	<1e-06	***	8.305
GRESTATE	0.052	-0.048	>	<1e-06	***	6.116
GLOANS	0.068	-0.063	>	<1e-06	***	7.052
DIRECT_LEASEFIN	206766.756	3354.192	>	<1e-06	***	10.807
LOAN_MIX	0.912	0.977	<	<1e-06	***	-17.563
SIZE	14.191	13.537	>	0.00157	**	3.336
CA_TA	0.051	0.095	<	0.000297	***	-3.881
RE_TA	0.032	-0.078	>	7.86e-06	***	4.972
EQ_TA	0.106	0.05	>	<1e-06	***	16.121
ROA	0.005	-0.049	>	<1e-06	***	10.192
LOAN1	0.663	0.7	<	0.0104	*	-2.659
LOAN2	1.525	0.838	>	0.0195	*	2.336

LOAN3	0.921	0.757	>	0.00258	**	3.016
LOANQUAL1	0.016	0.04	<	<1e-06	***	-8.925
LOANQUAL2	0.007	0.041	<	<1e-06	***	-9.278
TOTAL_CAP	0.148	0.075	>	<1e-06	***	7.882
OFF_BS	0.159	0.041	>	<1e-06	***	20.097
UNEARN_NI	1211.014	647.192	>	0.0186	*	2.39
TOT_INT_INC	655761.16	107654.404	>	<1e-06	***	8.892
TOT_NONINT_INC	355821.287	9931.673	>	<1e-06	***	10.133
NET_INCOME	177444.143	-70691.596	>	<1e-06	***	9.08
TOT_CHRG	99865.181	60803.692	>	0.0494	*	1.99
TOT_INT_EXP	145700.359	54531.596	>	6.05e-05	***	4.101
TOT_NONINT_EXP	533482.749	65668.096	>	<1e-06	***	9.1
LLP	85511.303	77720.115	>	0.732		0.344
NET_INCOME_PRE_OPS	177612.243	-70691.385	>	<1e-06	***	9.084
OREO	25402.316	28945.865	<	0.69		-0.401
DISTRESS	0.131	0.962	<	<1e-06	***	-30.46
PUBLIC	0.019	0	>	<1e-06	***	10.827
Q2DISCLOSURE	0.006	0	>	<1e-06	***	6.265
Q3DISCLOSURE	0.004	0.038	<	0.207		-1.279
Q4DISCLOSURE	0.032	0.019	>	0.508		0.666
NONCONFIDENTIAL	0.054	0.096	<	0.312		-1.02
CALL	0.022	0.019	>	0.888		0.141
DEPOSITS	0.013	0.038	<	0.517		-0.653
EFFECTIVE	0.009	0	>	<1e-06	***	7.386
FDIC	0.011	0.019	<	0.671		-0.427
REPORT	0.022	0.019	>	0.888		0.141
LOAN	0.012	0	>	<1e-06	***	5.513

LOANS	0.018	0	>	<1e-06	***	7.59
QUARTER	0.012	0	>	<1e-06	***	7.395
ASSETS	0.011	0.019	<	0.671		-0.427
BANK	0.047	0.019	>	0.168		1.395
BANKS	0.016	0.019	<	0.881		-0.15
DUE	0.011	0.038	<	0.322		-1
MILLION	0.016	0.038	<	0.571		-0.571
CAPITAL	0.017	0.038	<	0.58		-0.557
SCHEDULE	0.01	0.019	<	0.652		-0.453
LINE	0.01	0	>	<1e-06	***	6.052
REPORTED	0.008	0.019	<	0.568		-0.575
DECEMBER	0.01	0	>	<1e-06	***	6.018

Table 4: Model Restricted to most likely-to-fail guesses

This table presents results for our tests for our first hypothesis. The auditors' most likely-to-fail guess is their predicted failure and is measured by the issuance of a GCO to a bank client. Actual failure is a bank failure listed on the FDIC failed bank list.

Auditor Predictions 2007-2009			Auditor Predictions 2010-2019		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	1,830	No	No	3,359
Yes	No	14	Yes	No	1
No	Yes	20	No	Yes	70
Yes	Yes	27	Yes	Yes	10

Test for Hypothesis 1

Chi-squared test of total number of GC during vs after the financial crisis: p-value = 0.7853

Chi-squared test of total number of false positives during vs after the financial crisis: p-value = 0.01113

Table 5: Model Restricted to 127 most likely-to-fail guesses full time period

This table presents how well Isolation Forest and OCSVM predict bank failures for the top 127 most likely-to-fail predictions and compares those predictions to auditors' predictions. Auditor predictions are measured by the issuance of a GCO.

Isolation Forest Predictions			OCSVM Predictions			Auditor Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*	*Actual Failure*	*Going Concern*	*N*
No	No	5,157	No	No	5,152	No	No	5,189
Yes	No	47	Yes	No	52	Yes	No	15
No	Yes	122	No	Yes	127	No	Yes	90
Yes	Yes	5	Yes	Yes	0	Yes	Yes	37

Best Model Optimizing Sensitivity and Specificity

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	3,871	No	No	3,819
Yes	No	1	Yes	No	1
No	Yes	1,408	No	Yes	1,460
Yes	Yes	51	Yes	Yes	51

Table 6 Panel A: Model Restricted to 47 most likely-to-fail guesses Before Financial Crisis

This table presents how well Isolation Forest and OCSVM predict bank failures for the top 47 most likely-to-fail predictions before the financial crisis and compares those predictions to auditors' predictions. Auditor predictions are measured by the issuance of a GCO.

Isolation Forest Predictions			OCSVM Predictions			Auditor Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*	*Actual Failure*	*Going Concern*	*N*
No	No	1,807	No	No	1,804	No	No	1,830
Yes	No	37	Yes	No	40	Yes	No	14
No	Yes	43	No	Yes	46	No	Yes	20
Yes	Yes	4	Yes	Yes	1	Yes	Yes	27

Best Model Optimizing Sensitivity and Specificity

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	1,500	No	No	1298
Yes	No	2	Yes	No	1
No	Yes	350	No	Yes	552
Yes	Yes	39	Yes	Yes	40

Table 6 Panel B: Model Restricted to 80 most likely-to-fail guesses After Financial Crisis

This table presents how well Isolation Forest and OCSVM predict bank failures for the top 80 most likely-to-fail predictions after the financial crisis and compares those predictions to auditors' predictions. Auditor predictions are measured by the issuance of a GCO.

Isolation Forest Predictions			OCSVM Predictions			Auditor Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*	*Actual Failure*	*Going Concern*	*N*
No	No	3,349	No	No	3,350	No	No	3,359
Yes	No	11	Yes	No	10	Yes	No	1
No	Yes	80	No	Yes	79	No	Yes	70
Yes	Yes	0	Yes	Yes	1	Yes	Yes	10

Best Model Optimizing Sensitivity and Specificity

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	2,876	No	No	2,532
Yes	No	1	Yes	No	0
No	Yes	553	No	Yes	897
Yes	Yes	10	Yes	Yes	11

Table 7

Within GCO using the best cutoff with ranges from approximately the top 35 – 25 percent of predictions

Full Sample 2007-2019

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	38	No	No	44
Yes	No	5	Yes	No	6
No	Yes	52	No	Yes	46
Yes	Yes	32	Yes	Yes	31

During Crisis 2007-2009

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	14	No	No	8
Yes	No	7	Yes	No	1
No	Yes	6	No	Yes	12
Yes	Yes	20	Yes	Yes	26

Post Crisis 2010-2019

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	27	No	No	34
Yes	No	0	Yes	No	0
No	Yes	43	No	Yes	36
Yes	Yes	10	Yes	Yes	10

Table 8: Model Restricted to 127 most likely-to-fail guesses full time period with GCO

This table presents how well Isolation Forest and OCSVM predict bank failures for the top 127 most likely-to-fail predictions with GCOs added and compares those predictions to auditors' predictions. Auditor predictions are measured by the issuance of a GCO.

Isolation Forest Predictions			OCSVM Predictions			Auditor Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*	*Actual Failure*	*Going Concern*	*N*
No	No	5,157	No	No	5,158	No	No	5,189
Yes	No	47	Yes	No	46	Yes	No	15
No	Yes	122	No	Yes	121	No	Yes	90
Yes	Yes	5	Yes	Yes	6	Yes	Yes	37

Best Model Optimizing Sensitivity and Specificity

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	4,026	No	No	3,817
Yes	No	1	Yes	No	1
No	Yes	1,253	No	Yes	1,462
Yes	Yes	51	Yes	Yes	51

Table 9

Predicting Bank Failure whole sample best model with NO bailouts

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	3,368	No	No	3,970
Yes	No	1	Yes	No	2
No	Yes	1,562	No	Yes	960
Yes	Yes	51	Yes	Yes	50

Predicting Bank Failure whole sample best model Distress only

Isolation Forest Predictions			OCSVM Predictions		
Actual Failure	*Predicted Failure*	*N*	*Actual Failure*	*Predicted Failure*	*N*
No	No	306	No	No	417
Yes	No	2	Yes	No	4
No	Yes	462	No	Yes	351
Yes	Yes	48	Yes	Yes	46

Table 10 Panel A: Without GC

Metric	*Isolation_Forest*	*Support_Vector_Machine*
LogLoss	0.982	0.978
AUC	0.898	0.863
Gini	0.797	0.727
PRAUC	0.048	0.037
LiftAUC	2.542	2.300
GainAUC	0.894	0.860
KS_Stat	71.405	70.420

Table 10 Panel B: With GC

Metric	*Isolation_Forest_GC*	*Support_Vector_Machine_GC*
LogLoss	0.981	0.978
AUC	0.906	0.865
Gini	0.812	0.730
PRAUC	0.052	0.037
LiftAUC	2.617	2.309
GainAUC	0.902	0.861
KS_Stat	74.341	70.382