Cyber Risk in Banking: Measuring and Predicting Vulnerability*

Steven D. Baker† Dimuthu Ratnadiwakara‡

August 4, 2026

Abstract

We construct a bank–quarter panel linking external cybersecurity ratings, realized cyber incidents, and regulatory financial data for U.S. banks. Using quarterly expanding-window random forest models, we predict whether a bank will experience a cyber incident within the next quarter, two quarters, or year. Both bank characteristics and cybersecurity posture provide independent and complementary predictive content, with the combined model achieving the highest out-of-sample accuracy. Predictive performance remains robust across bank sizes and forecast horizons, is not driven by simple persistence in incident history, and exceeds that of alternative textual metrics. Interpretation of the fitted models highlights the consistent importance of size, balance sheet composition, and specific security controls, along with interactions between these controls, in predicting cyber incidents.

Keywords: cyber risk, banking, predictive modeling

*The results and views are those of the authors and do not reflect those of the Federal Reserve Bank of Richmond, or the Federal Reserve System.

†The Federal Reserve Bank of Richmond (steven.baker@rich.frb.org)

‡The Federal Reserve Bank of Richmond (dimuthu.ratnadiwakara@rich.frb.org)

1. Introduction

Cyber risk has rapidly become one of the most pressing threats in the banking sector, commanding heightened attention from industry and regulators (Kashyap and Wetherilt, 2019; Duffie and Younger, 2019; The International Monetary Fund, 2024). Banks' heavy reliance on digital infrastructure and online services has vastly expanded their attack surface, exposing them to sophisticated cyber threats. A successful cyberattack can impair payment systems, disrupt credit provision, and undermine confidence in the broader financial system, creating channels for rapid contagion across institutions (Eisenbach, Kovner, and Lee, 2022, 2025; Kopp, Kaffenberger, and Wilson, 2017). Despite growing supervisory attention, there are no standardized measures of bank-level cyber risk that supervisors or industry can apply across the sector. Existing approaches—largely disclosure-based or market-implied (Florackis, Louca, Michaely, and Weber, 2023a; Jamilov, Rey, and Tahoun, 2023; Jiang, Khanna, Yang, and Zhou, 2024)—are designed for large publicly traded firms and cannot be readily extended to include small and mid-sized banks.

We assemble a bank-quarter panel that links externally observed cybersecurity posture, realized cyber incidents, and detailed balance sheet information representing the entire U.S. banking sector. The dataset combines BitSight technical risk assessments, Zywave incident reports, and FFIEC Call Report filings, allowing for consistent coverage of banks of all sizes from 2015 through 2024.

BitSight provides standardized indicators of network hygiene, configuration practices, and evidence of compromise, while Zywave classifies and dates cyber incidents using proprietary monitoring. Call Reports supply quarterly financial and structural characteristics for U.S. banks. Integrating these sources produces a unified framework for forward-looking analysis of cyber vulnerability, enabling sector-wide assessment of how security posture relates to future incidents and supporting the development of supervisory met-

rics applicable beyond large, publicly traded institutions.

We begin by presenting descriptive statistics that summarize the cybersecurity landscape of U.S. banks. These include the cross-sectional distribution of technical security indicators, the incidence and composition of cyber events over time, and systematic variation in these patterns by bank size and balance sheet characteristics. The key takeaways from this exercise are twofold: first, while most banks score highly on core security metrics, there is meaningful dispersion—particularly in certain risk vectors—that leaves subsets of institutions more exposed; and second, realized incidents are disproportionately concentrated among the largest banks, consistent with their greater digital complexity and visibility to attackers and with the key role of size in attacker incentives, as modeled in equilibrium by Ramírez (2025).

We then evaluate the out-of-sample predictive content of cybersecurity posture and bank characteristics for future incidents, where the dependent variable is a binary indicator equal to one if the bank experiences at least one incident within the next quarter. Using a quarterly random forest framework, we find that both information sets contribute independent forecasting power. Models based solely on BitSight signals achieve an average out-of-sample AUC of 87.1%, while bank-characteristics-only models reach 86.8%.[1] Combining the two yields the highest average AUC of 88.9%. This gain underscores the complementarity between externally observed security posture and financial characteristics in predicting cyber events.

To ensure that the observed predictability is not simply capturing persistence in which institutions report or experience incidents, we augment the combined model with lagged incident history—defined as an indicator for whether the bank experienced a qualifying event in the previous four quarters. A model using only this variable achieves AUC of 78.8%, indicating that recent history alone leaves much variation in incident rates un-

[1]Predictive accuracy is assessed via the Area Under the Receiver Operating Characteristic Curve, or AUC. Models with AUC above 50% have discriminatory power, with AUC of 100% indicating perfect prediction.

explained. When lagged incidents are added to the combined specification, the AUC remains about the same as when they are omitted, suggesting that forecasting power comes from contemporaneous security posture and bank characteristics rather than mechanical recurrence of past events.

Next, we assess whether the model's predictive power is concentrated in specific parts of the banking sector or tied to a particular forecast window. Splitting the sample by bank size shows high accuracy across small, mid-sized, and large institutions, underscoring that the combination of technical security signals and balance sheet characteristics generalizes beyond the largest, most visible banks. Likewise, performance remains strong for prediction horizons ranging from one quarter to one year, indicating that the same predictors support both near-term monitoring and longer-term supervisory assessments.

Finally, we compare our model performance to textual metrics developed in Florackis et al. (2023a) and Jamilov et al. (2023) for public banks. Our model substantially outperforms the textual metrics, achieving out-of-sample AUC averaging above 90% versus around 60% for each of the textual metrics. Adding the textual metrics to our model results in a small performance improvement of around 0.5% AUC on average.

Beyond demonstrating high predictive accuracy, our analysis provides new insights into the factors underlying cyber risk in the banking sector. Variable importance results reveal that bank size and balance sheet composition—particularly total assets, deposits-to-assets, and loans-to-assets—are among the most consistent predictors of incidents, rivaling or exceeding the importance of individual security indicators. On the cybersecurity side, specific features such as web application security, patching cadence, TLS/SSL configuration, TLS/SSL certificates, and DKIM records emerge as the most informative BitSight measures, while several commonly tracked indicators, including botnet infections and spam propagation, show limited relevance in recent years.

Interaction analysis further reveals that cyber vulnerability often reflects combinations of risk factors rather than isolated weaknesses. For bank characteristics, size (total assets)

interacts strongly with other characteristics, especially asset growth, suggesting IT challenges when large banks rapidly expand or shrink. Among security indicators, patching cadence consistently interacts with multiple controls—most notably TLS/SSL configuration and TLS/SSL certificates—suggesting that the co-occurrence of unpatched systems and insecure communication protocols materially elevates incident risk. These patterns underscore the value of combining operational and technical perspectives when assessing cyber resilience and point to areas where supervisory attention could yield the greatest risk reduction.

This paper contributes to three strands of literature: the study of cyber incidents as a source of operational and systemic risk in banking, the measurement of firm-level cyber risk exposures, and the policy discussion on supervisory tools for mitigating cyber vulnerabilities.

First, a growing body of research documents the substantial and persistent financial repercussions of cyber incidents. Between 2012 and 2017, major banks incurred roughly $200 billion in operational risk losses, with cyber events consistently among the most significant drivers (The International Monetary Fund, 2024). Empirical evidence shows that breaches can erode franchise value and impair operations well beyond the immediate remediation period. Kamiya, Kang, Kim, Milidonis, and Stulz (2021) find that firms suffering cyberattacks with customer data losses experience shareholder wealth declines far exceeding direct costs, as well as significantly lower sales growth for up to three years. Such shocks can propagate beyond the affected institution, causing payment system disruptions, reputational contagion, and heightened sector-wide risk perceptions. Chernobai, Ozdagli, and Wang (2021) show that greater business complexity increases the frequency of operational losses, while Berger, Curti, Mihov, and Sedunov (2022) demonstrate that large operational losses can raise systemic risk through both direct solvency effects and correlated losses across banks. Kopp et al. (2017) highlight how cyber risk can generate market failures that threaten financial stability, motivating supervisory standards and col-

lective action to enhance resilience. The International Monetary Fund (2024) emphasizes that cyber risk has become a macrofinancial stability concern with systemic dimensions, noting that severe incidents can impair critical financial infrastructure, trigger liquidity strains, and generate broad confidence shocks. They stress that rising interconnectedness and reliance on digital technologies amplify the potential for contagion across institutions and borders, underscoring the need for coordinated supervisory responses and robust resilience frameworks. Our study extends this literature by documenting the incidence and drivers of cyber events across the entire U.S. banking sector, providing the first system-wide evidence that includes small and mid-sized banks absent from prior analyses.

Second, the analysis relates to research developing measures of firm-level cyber risk. Florackis et al. (2023a) construct a disclosure-based index from cyber-related language in 10-K filings, showing that it predicts future incidents and is priced in equity markets. Similarly, Jamilov et al. (2023) derive a market-based "cyber risk exposure" factor from earnings call transcripts, linking heightened cyber discussion to lower valuations, higher volatility, and sizable aggregate costs even without realized attacks. Jiang et al. (2024) employ machine learning to create a predictive cyber risk index that outperforms individual indicators and is associated with a positive risk premium. Ottonello and Rizzo (2024) focus on the software supply chain as a source of cyber risk, showing that new software vulnerabilities are a significant source of risk to which market participants react, but often slowly. While these studies advance measurement, they focus almost exclusively on large publicly traded firms with rich market or disclosure data.

Extending the scope to the banking sector, Murphy, Tindall, Klemme, Suek, and Dunbar (2025) combine commercial cybersecurity loss estimates with financial structure information to estimate average annual cyber loss rates for U.S. banks. Complementing these approaches, Eling and Wirfs (2019) use operational risk data to distinguish between routine and extreme cyber event costs through actuarial and statistical techniques, identifying key drivers—such as human error—that disproportionately affect financial institu-

tions. Jin, Li, Liu, and Nainar (2023) link banks' discretionary loan loss provisions — an indicator of internal control and risk management quality — to the likelihood of future cyber attacks, identifying an accounting-based early warning signal for cyber vulnerability. Heo (2024) applies the textual metric of Florackis et al. (2023a) banks, finding that cyber risk increases the probability of bank default. Gogolin, Lim, and Vallascas (2021) find that cyber incidents decrease branch deposit growth rates at small banks, which they attribute to reputational damage. Our work contributes along different dimensions, focusing on ex-ante risk indicators, especially those with specific technological underpinnings, that are applicable to nearly the entire US banking sector.

A distinct but related line of literature studies how banks react to the cyber risks of their customers in setting loan terms. Huang and Wang (2021) finds that firms with reported data breaches face higher loan spreads and increased likelihood of collateral requirements and loan covenants. Sheneman (2025) reaches similar conclusions but focuses on ex-ante risk, more in line with our approach. Whereas these papers study bank reactions to customers' cyber risks, we study the risks to banks themselves.

Third, the study contributes to the policy literature on cyber risk to the banking sector. On the one hand, our work supports a partial market-based solution. Evidence that commercially available risk-indicators are informative across the banking sector strengthens incentives to invest in cybersecurity, as such investments are at least partially observable to clients and counterparties who acquire the risk-indicators. Ahnert, Brolley, Cimon, and Riordan (2024) investigate such costly signals in a theoretical model, finding that they increase cybersecurity investment and welfare in equilibrium. However, they also show that regulation offers avenues to further improve the equilibrium. Kashyap and Wetherilt (2019) emphasize that cyber risk differs from other operational risks in its intentionality, potential for stealth, and rapid propagation, and argue for supervisory frameworks tailored to these features. Bouveret (2018) discuss the limitations of market discipline in addressing systemic cyber threats, advocating for macroprudential oversight and cross-

institutional contingency planning. Our analysis advances that agenda by producing a supervisory-relevant measure of bank-level cyber vulnerability that integrates both observable security posture and fundamental institutional characteristics. The resulting risk metric enables regulators to identify institutions most susceptible to future incidents, supporting targeted oversight and macroprudential risk assessment in an increasingly digital banking environment.

2. Data

Our analysis combines data from three primary sources: BitSight cybersecurity ratings, Zywave cyber incident reports, and regulatory bank-level financial filings from the FFIEC Call Reports. These datasets allow us to link observed security posture, realized cyber events, and institutional characteristics at the bank-quarter level. BitSight provides external assessments of cybersecurity conditions based on technical signals collected from internet-facing infrastructure. Zywave compiles detailed data on cyber incidents. Call Reports contain standardized regulatory disclosures on bank size, balance sheet composition, and profitability. Together, these sources form a panel suitable for modeling the determinants of cyber risk in the banking sector.

We use BitSight cybersecurity ratings to quantify each bank's external security posture based on a standardized set of risk vectors. The data include a range of technical signals related to network hygiene, configuration practices, and evidence of compromise, which we aggregate to the quarterly level. Specifically, available indicators include measures such as patching cadence, TLS/SSL configuration, web application headers, email authentication protocols (e.g., DKIM records), and botnet infection rates, among others. These signals are derived from externally observable behaviors and vulnerabilities and are designed to capture an institution's exposure to cyber threats across multiple dimensions.[2]

[2]See here for more details: https://help.bitsighttech.com/hc/en-us/articles/

Although BitSight cybersecurity ratings have been available for roughly a decade, we are aware of little independent analysis assessing their predictive value out of sample and none covering the banking sector. The study most comparable to ours is by Choi and Johnson (2021), who evaluate BitSight ratings in relation to hospital cyber incidents. Although there are many differences in statistical methodology and focus, the overall findings are similar: BitSight scores, including some more granular scores, can be effective for rank-ordering firms by incident risk when used in combination with industry-specific firm characteristics. Caro Rincon and Ordóñez (2023) document the relationship between BitSight's overall security rating, incident rates, and distance to default for public companies.[3] Instead, we focus on banks, many of which are not public, and find that more granular BitSight risk-indicators and firm characteristics are necessary to rank-order firms within the industry. We also assess out of sample predictive performance using Zywave incidents, as opposed to contemporaneous incidents provided by BitSight.

Figure 1 shows the distribution of BitSight cybersecurity risk vector scores separately for small ($<$ \$1 billion), mid-sized (\$1–10 billion), and large ($>$ \$10 billion) banks. Potential scores fall between 300 and 820, similar to FICO credit scores, with higher values indicating better performance (i.e., lower observed risk) for that dimension. Most indicators — such as botnet infections, insecure systems, and spam propagation — show distributions tightly clustered near the upper end of the scale, reflecting generally strong cybersecurity posture across banks. However, some metrics exhibit wider dispersion or pronounced lower tails, suggesting that certain institutions lag behind peers in specific security practices. For example, DKIM email authentication, Web Application Security, and SPF domains display notable variation among banks of all sizes. Overall, the distributions indicate that while most banks maintain relatively high technical security scores, there remain measurable and heterogeneous weaknesses across several risk vectors, even within the same size category.

360007320574-A-Guide-to-Navigating-and-Prioritizing-Bitsight-Risk-Categories-Risk-Vectors

[3]Moody's, which produced the study, acquired an ownership stake in BitSight in 2021.

Zywave provides detailed records of cyber incidents compiled from public records and news sources using a proprietary tracking system. Each record includes the incident date, a standardized incident type (e.g., data malicious breach, phishing, spoofing, social engineering, network/website disruptions, IT configuration or processing errors, and fraudulent use or account access), and a text description.[4] For a subset of observations, the dataset also reports the number of affected records or direct losses (e.g., from legal settlements), offering additional granularity on breach magnitude. Each incident is associated with a unique Zywave firm identifier as well as the domain name of the affected institution, which enables us to merge the data with BitSight cybersecurity ratings and regulatory filings at the bank level.

To align the incident data with other sources, we transform the incident-level dataset into a domain-by-quarter panel. We first restrict the sample to incidents falling into categories data malicious breach, phishing, spoofing, social engineering, network/website disruptions, IT configuration or processing errors, and fraudulent use or account access. Next we aggregate events by domain and calendar quarter. For each domain-quarter observation, we construct three binary indicators that serve as forward-looking outcomes in our empirical analysis: whether the domain experienced at least one qualifying cyber incident in the next quarter, in the next two quarters, or within the next four quarters.

Figure 2 plots the quarterly number of cyber incidents at U.S. banks by incident type from 2015 through 2024. Malicious data breaches dominate throughout the sample period, accounting for the vast majority of reported events and exhibiting substantial quarter-to-quarter volatility. Other categories—including phishing and social engineering, IT configuration errors, fraudulent account access, and network disruptions—occur far less frequently. The persistent prevalence of data breaches, combined with the steady presence of other incident types, highlights the ongoing exposure of the banking sector

[4] Zywave reports both an "accident date" and a potentially later "first notice date" reflecting disclosure for each incident. To ensure that our risk-indicators reflect vulnerabilities observable prior to the occurrence of an incident we use accident dates in our study.

to a range of cyber threats and reinforces the importance of forward-looking measures to monitor and mitigate these risks.

Figure 3 plots the percentage of banks experiencing at least one incident in each quarter, segmented by bank size. Incident frequency is consistently highest among banks with more than $10 billion in assets, peaking at over 20% of institutions in certain quarters. Mid-sized banks ($1–10 billion) show moderate and somewhat volatile incident rates, while small banks (less than $1 billion) rarely report incidents, remaining near or below 2% throughout. This distribution suggests that cyber risk is disproportionately concentrated among the largest institutions, consistent with their greater digital exposure, complexity, and visibility to attackers.

We obtain regulatory financial data for U.S. banks from the FFIEC Call Reports, which provide standardized quarterly information on bank balance sheets, income statements, and off-balance-sheet exposures. From these filings, we construct a set of control variables commonly used in bank risk analysis: total assets, asset growth, return on assets, deposits-to-assets, and loans-to-assets. These variables capture key dimensions of bank size, growth, profitability, and funding structure, and are included in all model specifications. The Call Reports also contain the domain name of each reporting institution, which we use to merge the financial data with external cybersecurity and incident records.

Table 1 compares the characteristics of banks included in our final analysis sample to those excluded due to missing domain linkages, separately for each size category. Across all size groups, in sample banks tend to be larger than those not covered. For banks with over $10 billion in assets or below $1 billion in assets, in sample banks have higher average asset growth than those not in-sample. Banks between $1 billion and $10 billion in assets have lower average asset growth in sample than not in sample.[5] Large banks have somewhat higher ROA on average in sample than not in sample, whereas small

[5]Table 1 reports means without winsorization, which we do prior to model estimation. Winsorizing asset growth and ROA at the 5% and 95% level for all banks renders in sample and not in sample means similar across all size categories.

banks have somewhat lower ROA on average in sample than not in sample. Other bank characteristics are similar in means across size categories whether in sample or not.

The final dataset is structured as a bank-quarter panel that integrates three sources of information. Each observation includes (i) current-quarter bank characteristics from Call Reports, (ii) current-quarter BitSight cybersecurity signals, and (iii) forward-looking incident indicators from Zywave, defined over one-, two-, and four-quarter horizons. This unified panel allows us to examine the relationship between observed security posture and future cyber incidents while controlling for underlying institutional characteristics.

3. Cross-Bank Metric of Cyber Risk Exposure

While BitSight provides an overall "Security Rating" intended to summarize an organization's cyber risk posture, we do not use this composite score as our primary predictive variable. For U.S. banks in our sample, the Security Rating alone exhibits limited discriminatory power for future incidents, and its proprietary weighting obscures the relative importance of underlying risk vectors. From a supervisory perspective, understanding which specific security dimensions contribute to elevated risk is essential for designing targeted interventions and monitoring emerging vulnerabilities. Our approach therefore focuses on modeling the granular BitSight feature set directly, allowing us to identify the most informative predictors and assess their stability over time.

This section develops and evaluates a predictive model of cyber incidents at the bank level. We adopt a random forest classifier as our primary modeling approach. This method is well-suited for capturing complex, non-parametric relationships and allows for flexible interaction structures without requiring explicit specification. Later, Section 4 characterizes the main drivers of our model.

3.1. Random Forest Model

To assess the predictive value of bank characteristics and externally observed cybersecurity signals, we estimate a sequence of quarterly random forest models with an expanding training window. We reserve the first two years of data (2015 Q1 - 2016 Q4) for initial training, with the first out of sample evaluation in 2017 Q1 and the last in 2024 Q4. Our main analysis predicts whether a bank will experience a qualifying cyber incident within the subsequent quarter, with longer prediction horizons evaluated for robustness. We evaluate three predictor sets: (i) bank characteristics alone, (ii) BitSight cybersecurity risk-indicators, and (iii) the combination of both. Out-of-sample performance is measured using the area under the receiver operating characteristic curve (AUC), with higher values indicating greater ability to distinguish between banks that will and will not experience a future incident.

Figure 4 presents the out-of-sample AUCs from models estimated using the three different sets of predictors, with dates corresponding to the end of the evaluation quarter and training ending the quarter prior. The results demonstrate that both bank characteristics and cybersecurity signals carry independent predictive content. BitSight-only models achieve relatively strong performance, with AUCs generally in the 80–90% range. Bank characteristics alone also yield consistent predictive power. Notably, the combined model outperforms the other two in nearly all quarters, with AUCs averaging around 89% and exhibiting less volatility over time. This pattern suggests that cyber incident risk reflects both latent institutional characteristics (e.g., size, growth, and financial structure) and observable security posture as captured by BitSight risk-indicators.

To clarify relative performance trends over time, we present trailing four-quarter moving averages of AUC for the same three models in Figure 5. The model combining bank characteristics with cyber risk-indicators persistently outperforms either of the alternatives on a moving average basis, with BitSight risk-indicators alone sometimes outper-

forming and sometimes underperforming bank characteristics. The performance improvement from combining both sets of variables highlights the value of integrating traditional supervisory data with external cyber risk signals when assessing banks' digital vulnerability.

Aggregating model performance over time further illustrates these patterns. Figure 6 reports the mean out-of-sample AUC across all quarters for five model specifications, including one that uses lagged incidents—defined as the occurrence of a cyber incident at the bank within the past year—as predictors. The lag-incidents-only model attains a lower average AUC of around 79%, indicating that recent incident history alone has limited predictive value and alleviating concerns that our results are driven solely by persistence of incidents at the same institutions or by reporting concentration.[6] BitSight-only and bank-characteristics-only models achieve comparable average AUCs of around 87%, respectively, while combining them yields average AUC of around 89%. Adding lagged incidents to the combined model yields no material change, suggesting that most predictive power derives from bank characteristics and cybersecurity signals rather than incident history. These results confirm that the two information sets are complementary and that the joint specification delivers the most accurate forecasts of future cyber incidents.

Figure 7 provides a validation of the model's predictive power by examining realized incident rates as a function of out-of-sample predicted risk. For each prediction quarter, we aggregate banks into deciles based on their predicted probability of a cyber incident in the next quarter (as estimated by the random forest model), and then calculate the actual proportion of banks experiencing an incident within each probability bin. Observed incident rates rise sharply as predicted risk reaches the highest deciles. This pattern is especially pronounced for banks with assets greater than $10 billion. In the top decile of predicted probabilities, over 60% of large banks experience a cyber incident in the sub-

[6]For work on the propensity of firms to report cyber attacks, see Amir, Levi, and Livne (2018).

sequent quarter. Small-and-mid-size banks also show large relative increases in incident rates for institutions in the highest risk decile. By contrast, incident rates remain very low for banks of all sizes among institutions in the lowest risk decile. The clear separation across predicted risk bins and size groups underscores both the accuracy and practical utility of the model for supervisory surveillance. The ability to identify those banks most likely to suffer cyber events provides a basis for targeted monitoring and risk management interventions.

3.2. Robustness

One important consideration in assessing model performance is whether predictive accuracy varies systematically with bank size. Larger banks differ from smaller institutions in several dimensions relevant to cyber risk—including IT infrastructure complexity, regulatory scrutiny, and public visibility—which may influence both the likelihood of incidents and the observability of relevant security signals. To examine this, we evaluate AUC over time for our main model specification — combining bank characteristics and BitSight features — on three sub-samples: small banks (assets below $1 billion), mid-sized banks ($1–10 billion), and large banks (above $10 billion). This split allows us to test whether the model's predictive content is robust across the size distribution, rather than being driven disproportionately by one segment of the industry.

The results, shown in Figure 8, indicate that predictive performance is strong across all size categories, with AUCs generally between 70% and 90%.[7] For comparison, FICO scores typically achieve AUC of 60-70% in credit scoring applications (Di Maggio and Ratnadiwakara, 2026), so even the lowest observed AUCs from our model reflect good separation by risk. Large banks exhibit the highest and most stable AUCs, likely reflecting richer and more reliable external security signals as well as more frequent incident

[7] To reduce noise and simplify comparison across series, the figure plots one-year moving averages, equivalent to Figure 5.

reporting. For small banks, the model still achieves AUCs averaging around 80% and never below 70%, underscoring that the combined predictor set retains substantial forecasting power even when applied to institutions with fewer observable signals and potentially less complete incident reporting. Interestingly, performance is the weakest for mid-sized banks, although it remains good and generally tracks with small bank performance over time. Small and mid-sized bank AUC both fall post-pandemic and improve in the most recent data. Overall, AUC remains good across size segments and time, reinforcing the general applicability of the model as a supervisory tool for monitoring cyber risk throughout the banking sector.

We also test the robustness of our results to alternative prediction horizons, re-estimating the same combined random forest model using forward-looking windows of one quarter, two quarters, and one year. An embargo period equal to forecast horizon separates training and testing periods, e.g., predictions begin in 2017 Q1, Q2, and Q4 for one-quarter, two-quarter, and one-year horizons, respectively. Figure 9 shows that predictive performance remains strong, with average AUCs above 80% for all forecast horizons. One-quarter forecasts have the most volatile AUCs, as bank incident indicators are more likely to change from one quarter to the next than over a one-year window. However, the one quarter forecasts are also the most accurate, with AUCs averaging over 5% higher than the one-year forecasts. As we discuss in Section 4, there is some time-variation in which risk-indicators are the most important for forecasting. For longer forecast horizons, updates to model weights lag due to the embargo period and risk-indicators become relatively stale, so some reduction in forecast performance is to be expected. Nevertheless, performance remains good even at the one-year horizon, confirming that model forecasts are not transient, and enhancing their practical utility for supervisory monitoring across different planning and intervention timeframes.

3.3. Comparison with Textual Metrics

An alternative way to measure exposure to cyber risk is the analysis of related public disclosures by firms, on the basis that firms reveal some information about their exposure to such risk or their ability to prevent incidents. Two prominent examples of this approach are described in Florackis et al. (2023a) (FLMW), who analyze risk-factor information from 10-Ks, and Jamilov et al. (2023) (JRT), who analyze quarterly earnings calls. In this section we compare these metrics to our baseline model combining bank characteristics and BitSight risk-indicators, evaluating each approach against Zywave incidents on a uniform sample of banks. We confirm that both textual metrics have some ability to rank-order banks by incident risk, but our model performs much better and mostly subsumes information in the textual metrics.

Because our sample of banks is predominantly private whereas FLMW and JRT cover public firms spanning many industries, the number of firms in the merged sample is quite small. Scores from JRT match our quarterly frequency, whereas for FLMW we repeat annual scores for each quarter. In addition, overlap in time is limited: our merged sample runs from 2015 Q1 through 2017 Q4.[8] During this interval the set of banks available from all sources is quite stable, with between 92 and 104 banks depending on the quarter.

We compare FLMW and JRT to two versions of our model. Both follow the fitting and out-of-sample evaluation procedures described previously in this section, using bank characteristics in Table 1 the set of BitSight variables included in Figure 12, but applied to the merged sample including FLMW and JRT scores. The second model variant adds the scores of FLMW and JRT alongside other independent variables, to assess incremental performance gains.

Figure 10 shows the results, evaluating performance by quarterly AUC. Although

[8]BitSight coverage is limited during 2015, the first year for which it is available, so we begin out of sample evaluation in 2016 Q2 to allow for at least one quarter of training data with complete coverage. We obtain cyber scores for FLMW via the link provided in Florackis, Louca, Michaely, and Weber (2023b), and cyber scores for JRT from Rustam Jamilov at https://users.ox.ac.uk/~econ0628/Cyber_Risk_Data.zip.

FLMW and JRT achieve AUC above 50% in all quarters, confirming that the scores have predictive value, AUC is often below 60%. FLMW achieves a maximum AUC of around 62% in 2017 Q1 whereas JRT achieves a maximum AUC around 79% in 2017 Q3. In contrast, our baseline model performs better and more consistently, with minimum AUC of 89% and maximum AUC above 95%. Adding FLMW and JRT scores to our model leaves performance little changed, increasing it by about 0.5% on average.

The superior performance of our model may be unsurprising given that it is specialized to the banking sector whereas FLMW and JRT are cross-industry metrics. However, the comparison establishes that these textual metrics contain little information beyond what is already captured by the combination of firm characteristics and BitSight risk-indicators, at least for public banks.

4. Analysis of Explanatory Variables and Interpretation of Results

Having established the practicality of our model for out of sample cyber incident prediction, we now investigate the main drivers of the model in more detail. Which predictive variables are the most important in the banking sector? How do they relate to incidents and to each other? Since the focus of this section is interpretation rather than out of sample prediction, we investigate random forest models fitted to our full training sample – the same data available in the final quarter analyzed in Section 3.

4.1. Variable Importance

We begin by characterizing variable importance with all candidate variables included in the model, then consider whether relative importance changes in models where subsets of the variables are excluded, following the flow of Figure 11.

Figure 11a compares all independent variables according to two commonly used im-

portance metrics: the mean decrease in Gini impurity and mean decrease in classification accuracy due to permutation. Gini impurity is used internally by the random forest algorithm, to select the variable within a candidate set that maximally decreases heterogeneity of class among observations along each branch of the tree, i.e., splitting banks into those with incidents and those without in our case. Permutation importance captures the reduction in classification accuracy when the values of a given variable are randomly shuffled. For each metric, we normalize the sum of importance across variables to 100. In our case, relative variable importance is similar whether impurity or permutation is used, with some notable exceptions.

One result from Figure 11a is that bank characteristics are important to cyber risk. Total assets, deposits/assets, and loans/assets are the three most important variables for predicting incidents based on Gini impurity, a finding similar to that of Jiang et al. (2024) for public firm characteristics that we extend to private banks. Total assets, in particular, has importance roughly twice that of the most important BitSight variable, which is web application security. ROA and asset growth, the two remaining characteristics, are also in the top 10 candidate variables by impurity but are among the least important variables by permutation, constituting the two cases for which impurity and permutation differ the most. After web application security, the most important BitSight variables are TLS/SSL certificates, TLS/SSL configuration, DKIM records, and patching cadence. These variables are relatively more important when assessed via permutation: three of the five most important variables are BitSight risk-indicators according to permutation.

In Section 4, we evaluated models with bank characteristics and BitSight risk-indicators either separately or in combination. One possibility is that the relative importance of some variables changes greatly once correlated alternatives are included, e.g., that BitSight variables proxy for bank characteristics or vice versa. Figure 11b and Figure 11c show this is not the case, as the relative importance of characteristics and the top

five BitSight risk-indicators remains about the same for each model variant.[9]

Since information technology evolves rapidly, it is interesting to see whether the most important risk-indicators change over time. Figure 12 shows the five most important BitSight risk-indicators in each quarter of our evaluation sample, estimated using a trailing eight-quarter moving training window to reveal time-variation. DKIM records, TLS/SSL configuration, and web application security are important for most of the sample period. TLS/SSL certificates enters the top five around the middle of the sample and remains important. Other variables, such as potentially exploited, botnet infections, and spam propagation, are initially important but later drop out. Patching cadence and open ports enter later in the sample period. Other variables are of only sporadic importance, or are never important at all.[10]

To simplify exposition and reduce clutter in the plots, the remaining sections focus on the five bank characteristics and the five most important BitSight risk-indicators identified in Figure 11a: web application security, TLS/SSL certificates, TLS/SSL configuration, DKIM records, and patching cadence.

4.2. Variable Interactions

Figure 11b and Figure 11c suggested limited direct interaction between bank characteristics and risk-indicators: interactions across categories do not change relative variable importance. However, there are strong interactions within characteristics and within risk-indicators.

To investigate these interactions in more detail, Figure 13 shows network visualizations of variable interactions. Node size and color tint indicate Gini impurity, whereas edge linewidth and color tint indicate interaction importance according to the unnormal-

[9]Of course, the absolute importance of each variable does drop when the variable set is expanded: the figures normalize importance within each subset to 100, to emphasize relative values.

[10]We have experimented with feature selection, limiting the model to use only a subset of variables identified as important. This has small but slightly negative impact on the model's predictive accuracy.

ized H-statistic of Friedman and Popescu (2008). The H-statistic is an unsigned measure of variable interaction based on the mean squared difference between the model prediction when two variables are allowed to interact (joint prediction) or not (synthesis of uni-variate predictions). It takes values from zero, indicating exclusively uni-variate prediction, to one, indicating exclusively joint prediction. To reduce figure complexity and consistent with previously discussed results, we focus on interactions within bank characteristics (Figure 13a) and within BitSight risk-indicators (Figure 13b).

Figure 13a shows that the most important characteristics, such as total assets and loans/assets, also have the strongest interactions with other characteristics on average. By contrast, the smallest interactions are between asset growth and ROA and asset growth and deposits/assets. However, all characteristics interact, with even the smallest H-statistics around 0.3. Of note is the largest interaction, which is between total assets and asset growth, indicating that size and its first derivative are jointly important to cybersecurity risk.

Figure 13b shows interactions between selected BitSight variables. Since total assets is overall the most important variable, we also plot interactions between it and the BitSight variables. Such interactions are, however, very small. This gives us confidence that BitSight scores may be interpreted independent of bank size.

In contrast to the results in Figure 13a for characteristics, the most important BitSight interactions in Figure 13b do not generally originate from the most important variable, which is web application security, but rather from patching cadence. The only large interaction not involving patching cadence is between DKIM records and TLS/SSL certificates. Patching cadence interacts strongly with all four other BitSight variables selected during the final part of the evaluation period: web app. security, TLS/SSL configuration, TLS/SSL certificates, and DKIM records. A natural interpretation is that the risk of an incident increases when there is both unpatched software and a weakness in communication protocols.

4.3. Univariate Dependencies

We now investigate in detail the relationship between each of our selected variables and incident rates as uncovered by the random forest model, via partial dependency plots (PDP). Since random forests allow for general non-parametric dependencies, the relationship between each predictive variable and incidents is most easily summarized by plotting the model-implied incident probability conditional on the value of a selected variable, taking expectations over other variable values according to the sample distribution.

Figure 14 shows the PDP for each bank characteristic, with annual incident probabilities on a log scale on the y axes. In the top left plot, we see incident rates rising sharply with the log of total assets, particularly in the transition between mid-sized and large banks, consistent with the higher incident rate among large banks illustrated in Figure 3. However, the PDP plot reveals a relationship between total assets and incidents that is relatively flat within large banks and within small banks. The remaining PDP plots for bank characteristics are, by contrast, unambiguously non-monotonic. Deposits/assets and loans/assets are similar "smirks" varying over a wide range, falling from approximately 10% quarterly incident probability for very small x-values to approximately 1% for middle x-values before rising again to around 3% for large x-values. Asset growth and ROA share similar "smile" patterns, but vary over a narrow range of around 1-2% annual incident probability. In short, banks that are near the extremes in terms of their growth rate, asset composition, or liability composition are at higher risk of cyber incidents.

Figure 15 shows the PDP for each selected BitSight risk-indicator. First, we note that variation in incident probabilities is less for BitSight variables than we observed for bank characteristics, with the exceptions of ROA and asset growth. However, most BitSight variables have the expected relationship to incident probabilities – they rise as the scores fall – with incident probabilities increasing by at least 50% and in some cases doubling

conditional on very low scores. Very low scores are rare, but they are observed, and incidents are also uncommon for small-and-medium-sized banks.

However, Figure 15 includes a few curiosities worthy of remark, as not all relationships are monotonic or of the expected sign. Web application security and TLS/SSL certificates both show incident probabilities counterintuitively increasing with scores in the upper range. This could reflect maximum scores (typically 820) as default values when no information about the risk-indicator is available, such that incident probabilities match the higher unconditional mean for very high scores versus a lower conditional mean with measurably good-but-not-perfect scores. Regardless of the underlying mechanism, the random forest shines in dealing with such local non-monotonicity relative to more traditional linear models. Although we do not include results here, alternatives such as logistic regression generally find weaker predictive relationships, sometimes with point estimates of opposite sign, precisely because so many observations fall near the top end of the score range, where variation is either uninformative or, worse, blurs the distinction between confirmation of strong security and "no data." Since methods such as logistic regression must assign a single coefficient summarizing the global relationship, murky relationships in small but frequently observed score ranges may dominate.

Finally, there is one variable in Figure 15 where high scores are unambiguously and counterintuitively "bad news": DKIM records, in the top-right plot. Although we lack a conclusive explanation, we do find that the DKIM records score increases following an incident, perhaps because it is an easy risk-indicator for firms to improve in response. Since lagged-incidents are predictive of future incidents, this could explain why high DKIM scores are associated with increased probability of an incident.

4.4. Bivariate Dependencies

Before concluding, we revisit in more detail the most interesting set of interactions between BitSight variables. Figure 16 shows pair-wise interaction heatmaps for the five

most important BitSight variables on the upper triangle, reprises univariate PDP plots on the diagonal, and has scatterplots of observations colorized by model-implied incident probability in the lower triangle. The figure serves two main purposes. First, recall that patching cadence had the strongest interactions with other variables in Figure 13b, but the H-statistic is unsigned, leaving the nature of the interaction ambiguous. The heatmaps show that the model predicts the highest incident probabilities when patching cadence is low *and* TLS/SSL certificates, TLS/SSL configuration, or web application headers is low. This is consistent with our earlier interpretation that unpatched software combines with weak communication protocols to increase incident risk. Second, the scatterplots show that, although moderate and low scores are less common than high scores, they are observed not only for individual variables but also for several combinations of variables. Such combinations may be critical to identifying high incident risk.

5. Conclusion

This paper develops a sector-wide, forward-looking measure of cyber risk exposure for the U.S. banking industry, combining externally observed cybersecurity posture, realized incident data, and bank characteristics. By integrating BitSight's granular technical risk indicators with regulatory Call Reports and Zywave's incident records, we construct a bank-quarter panel covering institutions of all sizes from 2015 to 2024. Using a random forest model, we show that both cybersecurity posture and financial characteristics provide independent and complementary predictive power for cyber incidents over the subsequent year.

Our results are robust across bank sizes and prediction horizons, underscoring the model's applicability for supervisory monitoring throughout the sector. The analysis reveals that predictive accuracy does not rely only on persistence in which banks experience incidents, but instead reflects contemporaneous differences in network security

practices, IT configuration hygiene, and institutional fundamentals. Moreover, by identifying the specific technical risk vectors most closely associated with elevated incident probabilities—such as patching cadence, TLS/SSL configuration, and web application security—the framework yields actionable insights for targeted oversight and internal risk management.

From a policy perspective, the findings highlight the value of integrating commercial cyber risk indicators with supervisory data to generate timely, institution-specific measures of digital vulnerability. Such measures can support microprudential interventions at the bank level as well as macroprudential assessments of systemic exposure to cyber threats.

References

Ahnert, T., M. Brolley, D. A. Cimon, and R. Riordan (2024). Cyber risk and security investment. *Available at SSRN 4057505*.

Amir, E., S. Levi, and T. Livne (2018). Do firms underreport information on cyber-attacks? evidence from capital markets. *Review of Accounting Studies 23*(3), 1177–1206.

Berger, A. N., F. Curti, A. Mihov, and J. Sedunov (2022). Operational risk is more systemic than you think: Evidence from us bank holding companies. *Journal of Banking & Finance 143*, 106619.

Bouveret, A. (2018). Cyber risk for the financial sector: A framework for quantitative assessment. (18/143).

Caro Rincon, A. and G. Ordóñez (2023). The impact of cyber security management practices on the likelihood of cyber events and its effect on financial risk. Technical report, Moody's Analytics.

Chernobai, A., A. Ozdagli, and J. Wang (2021). Business complexity and risk management: Evidence from operational risk events in us bank holding companies. *Journal of Monetary Economics 117*, 418–440.

Choi, S. J. and M. E. Johnson (2021). The relationship between cybersecurity ratings and the risk of hospital data breaches. *Journal of the American Medical Informatics Association 28*(10), 2085–2092.

Di Maggio, M. and D. Ratnadiwakara (2026). Invisible primes: Fintech lending with alternative data. *Management Science*. Published Online Ahead of Print: May 13, 2026.

Duffie, D. and J. Younger (2019). *Cyber runs*. Brookings.

Eisenbach, T. M., A. Kovner, and M. Lee (2025). When it rains, it pours: Cyber vulnerability and financial conditions. *Economic Policy Review 31*(1), 1–24.

Eisenbach, T. M., A. Kovner, and M. J. Lee (2022). Cyber risk and the us financial system: A pre-mortem analysis. *Journal of Financial Economics 145*(3), 802–826.

Eling, M. and J. Wirfs (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research 272*(3), 1109–1119.

Florackis, C., C. Louca, R. Michaely, and M. Weber (2023a). Cybersecurity risk. *The Review of Financial Studies 36*(1), 351–407.

Florackis, C., C. Louca, R. Michaely, and M. Weber (2023b). Cybersecurity risk: The data. *Chicago Booth Research Paper* (23-01).

Friedman, J. H. and B. E. Popescu (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics 2*(3), 916–954.

Gogolin, F., I. Lim, and F. Vallascas (2021). Cyberattacks on small banks and the impact on local banking markets. *Available at SSRN 3823296*.

Heo, Y. (2024). Cybersecurity and bank fragility. *Available at SSRN 4660090*.

Huang, H. H. and C. Wang (2021). Do banks price firms' data breaches? *The Accounting Review 96*(3), 261–286.

Jamilov, R., H. Rey, and A. Tahoun (2023). The anatomy of cyber risk. Technical report, National Bureau of Economic Research.

Jiang, H., N. Khanna, Q. Yang, and J. Zhou (2024). The cyber risk premium. *Management Science 70*(12), 8791–8817.

Jin, Y. J., N. Li, S. Liu, and S. M. K. Nainar (2023). Cyber attacks, discretionary loan loss provisions, and banks' earnings management. *Finance Research Letters 54*, 103705.

Kamiya, S., J.-K. Kang, J. Kim, A. Milidonis, and R. M. Stulz (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics 139*(3), 719–749.

Kashyap, A. K. and A. Wetherilt (2019). Some principles for regulating cyber risk. In *AEA Papers and Proceedings*, Volume 109, pp. 482–487. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kopp, E., L. Kaffenberger, and C. Wilson (2017). Cyber risk, market failures, and financial stability. IMF Working Paper 17/185, International Monetary Fund.

Murphy, A., M. L. Tindall, K. Klemme, J. I. Suek, and S. J. Dunbar (2025, May). What drives cyber losses at u.s. banks? potential statistical markers. Working Paper 2520, Federal Reserve Bank of Dallas. Accessed: 2025-08-15.

Ottonello, G. and A. E. Rizzo (2024). Do software companies spread cyber risk? *Available at SSRN*.

Ramírez, C. A. (2025). On equilibrium cyber risk. *Economics Letters 251*, 112307.

Sheneman, A. (2025). Cybersecurity risk and bank loan contracting. *Available at SSRN 3406217*.

The International Monetary Fund (2024, April). Cyber risk: A growing concern for macro-financial stability. In *Global Financial Stability Report*, Chapter 3. International Monetary Fund. Accessed: 2025-08-15.

Figure 1: BitSight Signal Distributions

This figure shows the 5th to 95th percentiles of BitSight cybersecurity indicator scores for U.S. banks, grouped by total asset size: less than $1 billion (blue), $1–$10 billion (green), and greater than $10 billion (red). Potential scores fall between 300 and 820, similar to FICO credit scores, with higher scores corresponding to stronger security along the evaluated dimension. Boxes are 25th to 75th percentiles, with ticks at the 50th percentile.

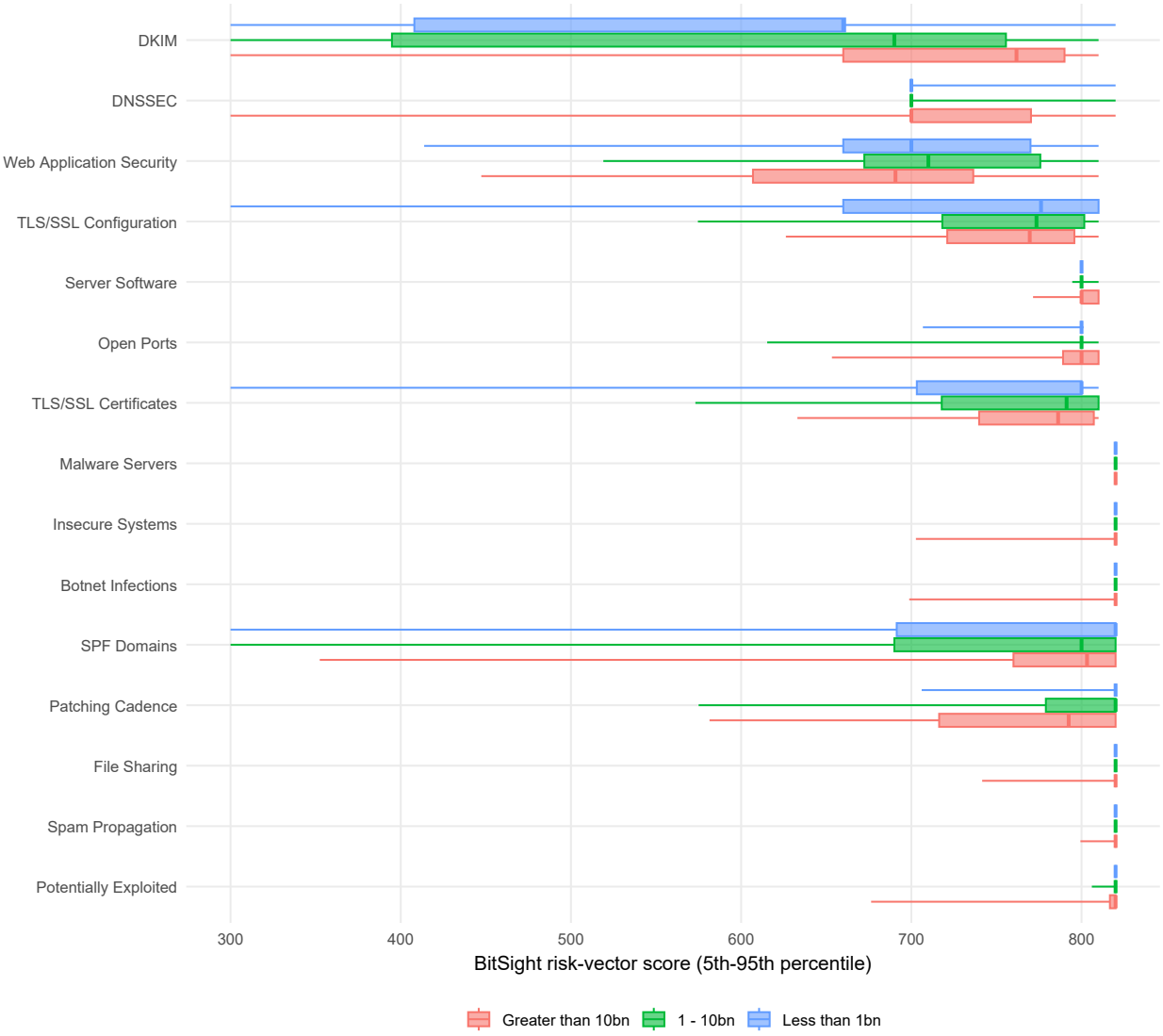


Figure 2: Number of incidents at banks

This figure shows quarterly counts of reported cyber incidents at U.S. banks from 2015 Q1 to 2024 Q4, disaggregated by incident category: data–malicious breach, identity–fraudulent use/account access, IT–configuration/implementation errors, IT–processing errors, network/website disruption, and phishing/spoofing/social engineering.

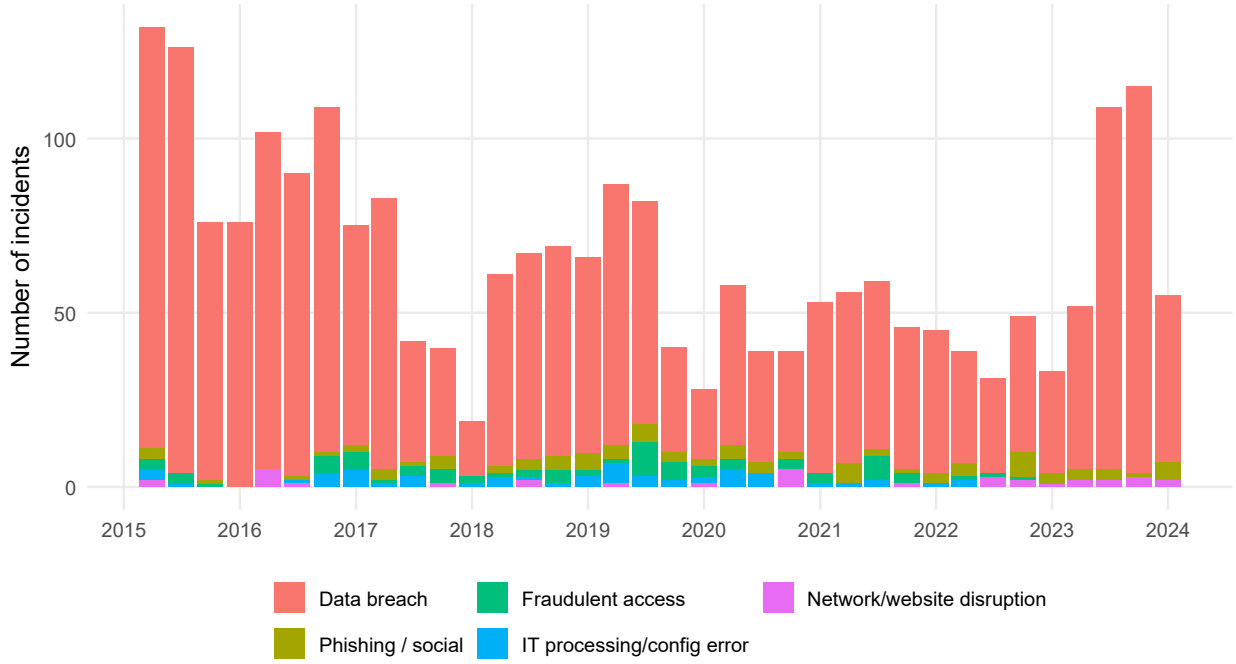


Figure 3: Percent of banks with incidents

This figure shows the time series of the percentage of banks experiencing at least one cyber incident from 2015 Q1 through 2024 Q4, segmented by total assets size: less than $1 billion (dashed blue), $1–10 billion (short dashed green), and greater than $10 billion (solid red).

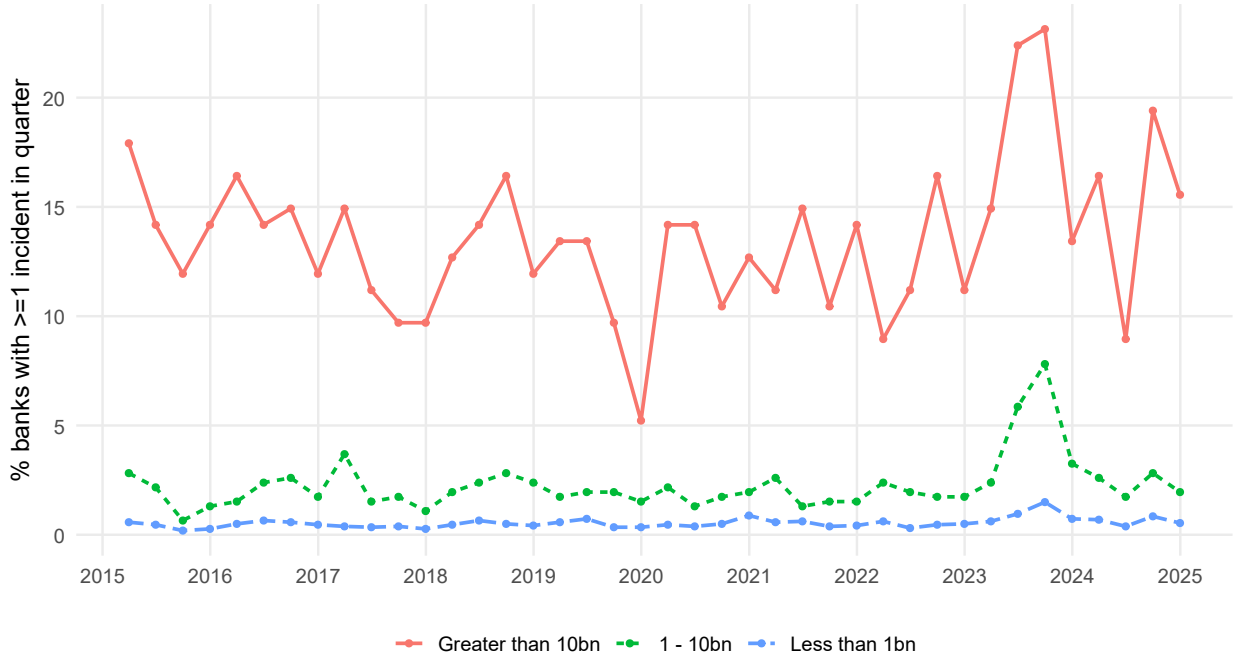


Table 1: Bank Characteristics

This table reports summary statistics for banks in and out of the estimation sample, grouped by asset size: greater than 10 billion, between 1 and 10 billion, and less than 1 billion. For each group, the table shows the number of banks (N), mean asset growth, deposits-to-assets ratio, loans-to-assets ratio, return on assets (ROA), and total assets (in thousands of USD).

	Greater than 10bn		1 - 10bn		Less than 1bn	
	Not in Sample	In Sample	Not in Sample	In Sample	Not in Sample	In Sample
N	43	97	198	461	1,818	2,610
Asset Growth (%)	2.244	3.150	3.789	3.014	1.937	2.285
Deposits/Assets	0.756	0.760	0.807	0.811	0.808	0.833
Loans/Assets	0.663	0.659	0.691	0.719	0.624	0.653
ROA (%)	1.128	1.350	1.249	1.308	1.462	1.196
Total Assets ($000)	85,014,978	124,533,291	2,460,908	2,766,632	227,927	285,468

Figure 4: Out-of-Sample Predictability - Quarterly

This figure plots quarterly out-of-sample predictability, measured by AUC, for three model specifications: Bank Characteristics (solid red line), BitSight + Bank Characteristics (short dashed green line), and BitSight Only (dashed blue line), over the period from 2017 Q1 to 2024 Q4.

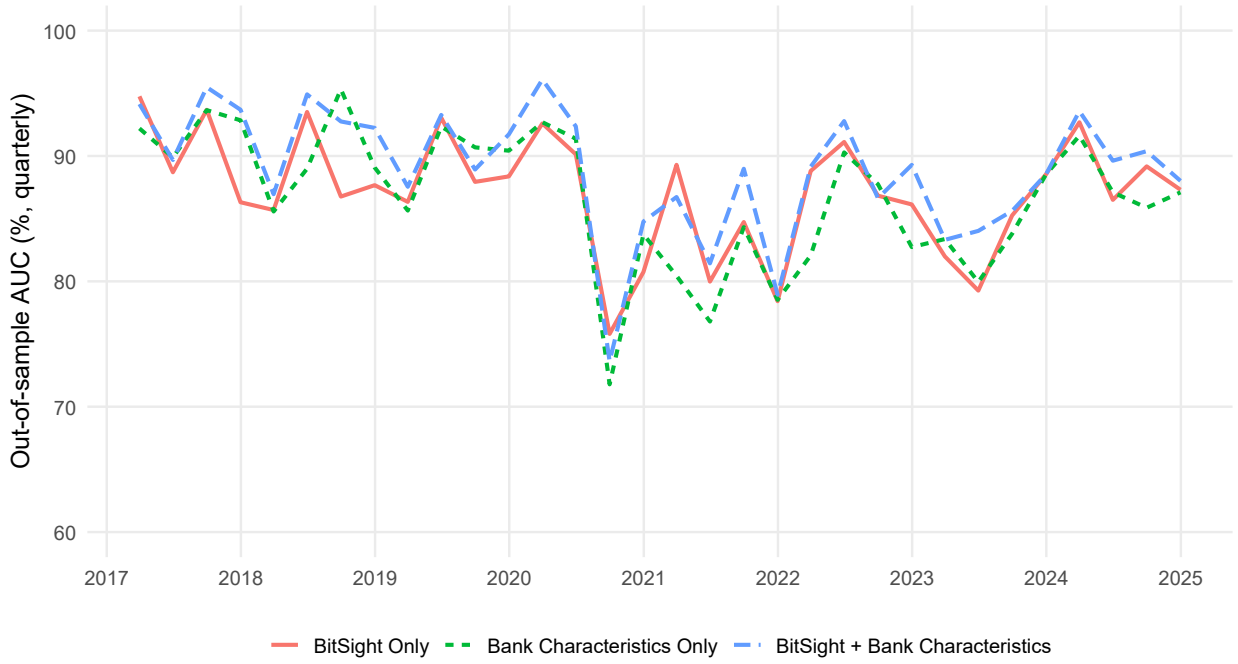


Figure 5: Out-of-Sample Predictability - Quarterly, Moving Average

This figure plots the trailing one-year moving average of quarterly out-of-sample predictability, measured by AUC, for three model specifications: Bank Characteristics (solid red line), BitSight + Bank Characteristics (short dashed green line), and BitSight Only (dashed blue line), over the period from 2017 Q4 to 2024 Q4.

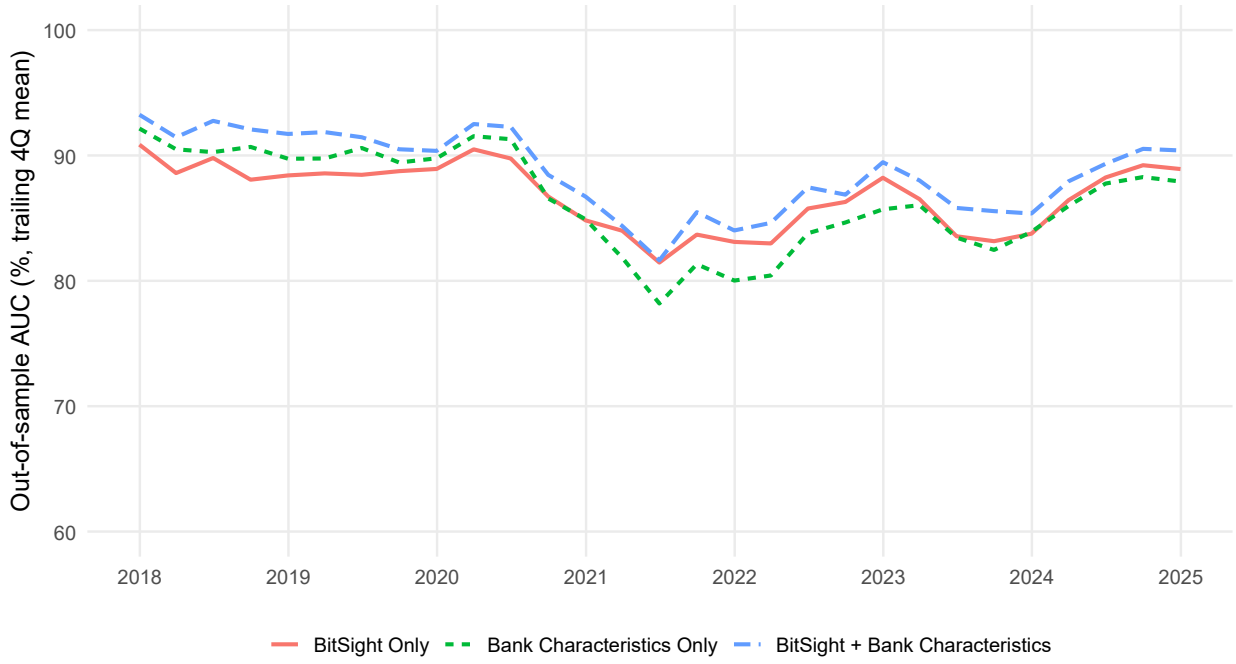


Figure 6: Out-of-Sample Predictability - Mean

This figure shows mean out-of-sample predictability (AUC) for five model specifications. Models using only lagged incidents achieve the lowest predictive accuracy, while those combining BitSight signals and bank characteristics perform substantially better. The inclusion of lagged incidents alongside BitSight and bank characteristics produces the highest mean AUC, marginally improving over the combined BitSight–bank characteristics model. These results provide a clear comparison of the relative contribution of different predictor sets to model performance.

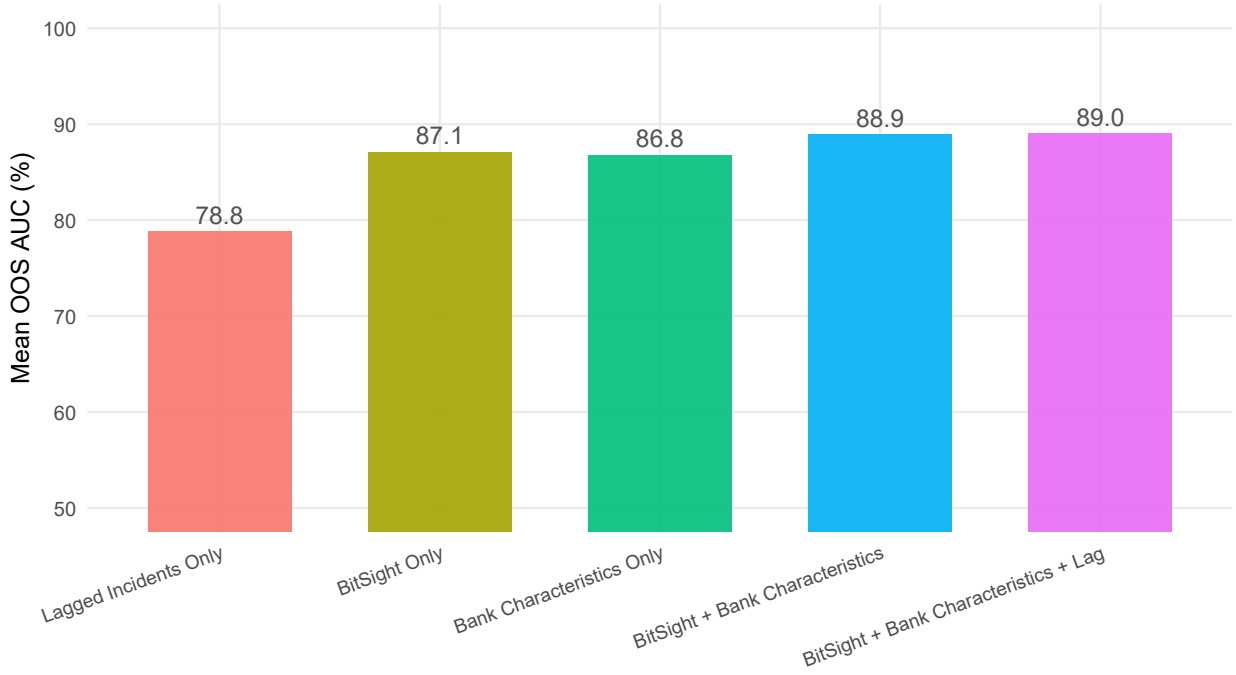


Figure 7: Out-of-Sample Predictability - Mean

This figure presents the distribution of actual cyber incidents across predicted probability deciles, segmented by bank size. The x-axis groups banks into deciles based on their out-of-sample predicted probability of a cyber incident, and the y-axis reports the observed percentage of banks within each decile that experienced an incident. Separate series are shown for banks with assets less than $1 billion (dashed blue), between $1 and $10 billion (short dashed green), and greater than 10 billion (solid red).
The plot allows for comparison of model calibration and discrimination across size categories, highlighting differences in incident prevalence between low- and high-risk deciles. It also facilitates visual assessment of how predicted risk translates into realized outcomes within each bank size class.

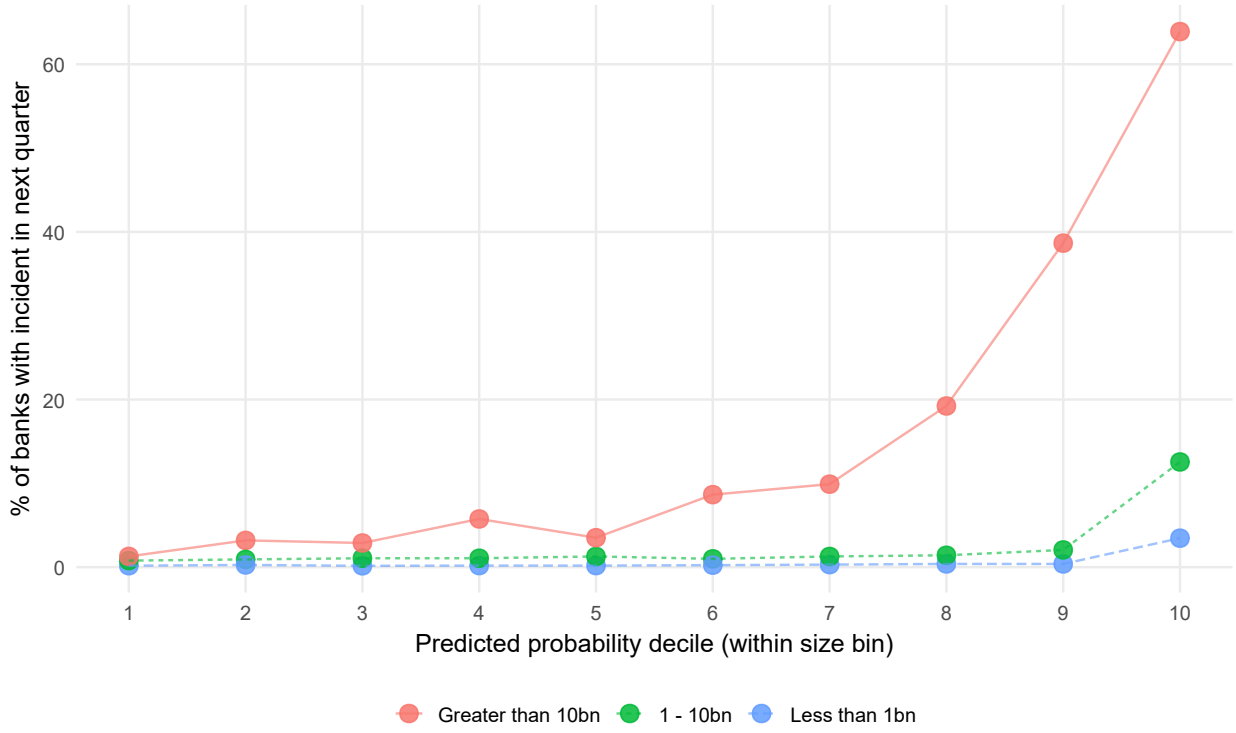


Figure 8: Out-of-Sample Predictability - By Size

This figure plots the trailing one-year moving average of quarterly out-of-sample predictability (AUC) from 2017 Q4 to 2024 Q4 separately for banks with total assets less than $1 billion (dashed blue), $1–10 billion (short dashed green), and greater than $10 billion (solid red).

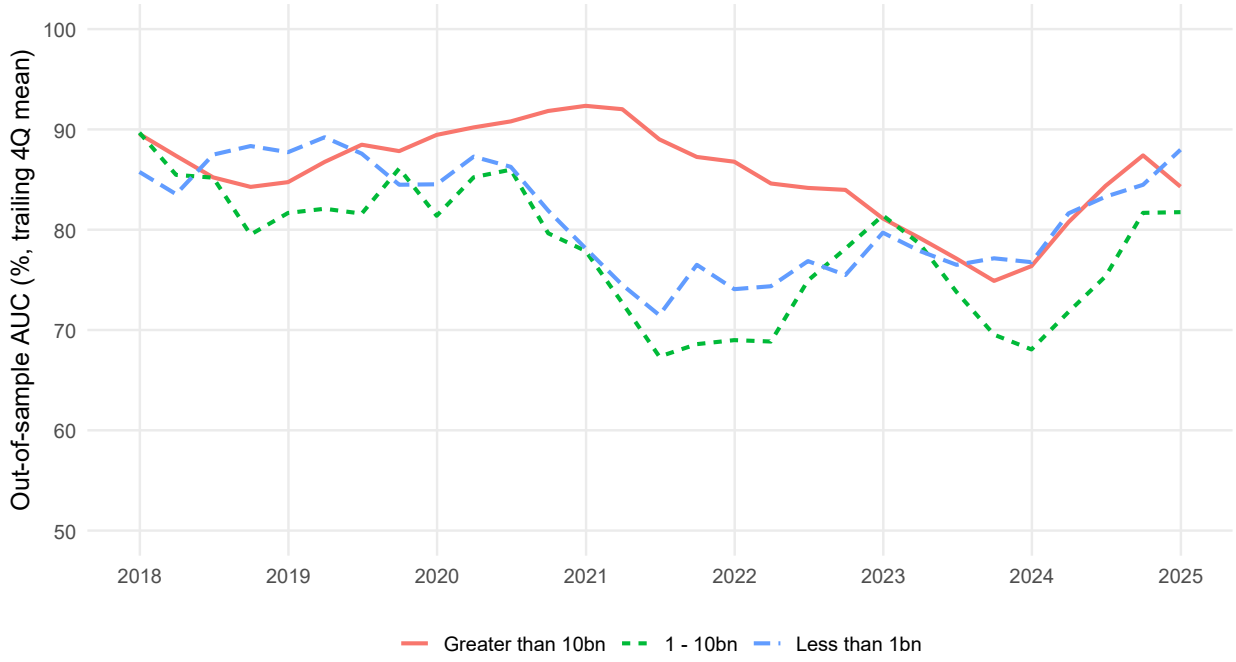


Figure 9: Out-of-Sample Predictability - By Incident Horizon

This figure plots out-of-sample predictability (AUC) from 2017 through 2024, broken down by prediction horizon. The solid red line corresponds to a one-quarter horizon, the short dashed green line to two quarters, and the dashed blue line to one year. The horizontal axis marks calendar years, while the vertical axis measures predictability in percentage AUC terms. An embargo period equal to forecast horizon separates training and testing periods: predictions begin in 2017 Q1, Q2, and Q4 for one-quarter, two-quarter, and one-year horizons, respectively.

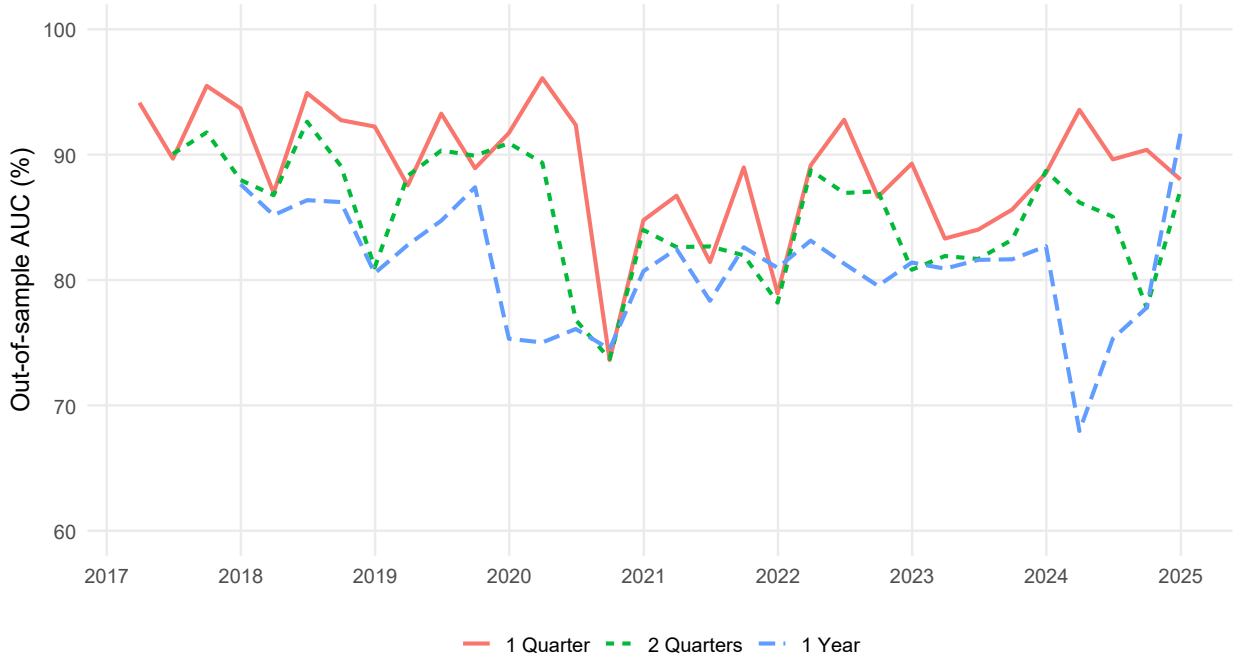


Figure 10: Out-of-Sample Predictability - Comparison with Text Metrics

The figure compares our baseline model, BitSight + Bank Characteristics, with two alternative metrics based on textual analysis, developed in Florackis et al. (2023a) (FLMW) and Jamilov et al. (2023) (JRT) respectively. Performance is assessed using AUC out-of-sample, based on incident occurrence over the following year. Also shown are results for the baseline model combined with textual metrics. Although FLMW and JRT achieve AUC of around 60% on average, indicating better than random performance, our baseline model performs much better, with average AUC above 90%. Adding textual metrics to our model leaves results about the same, improving AUC by about 0.5% on average.

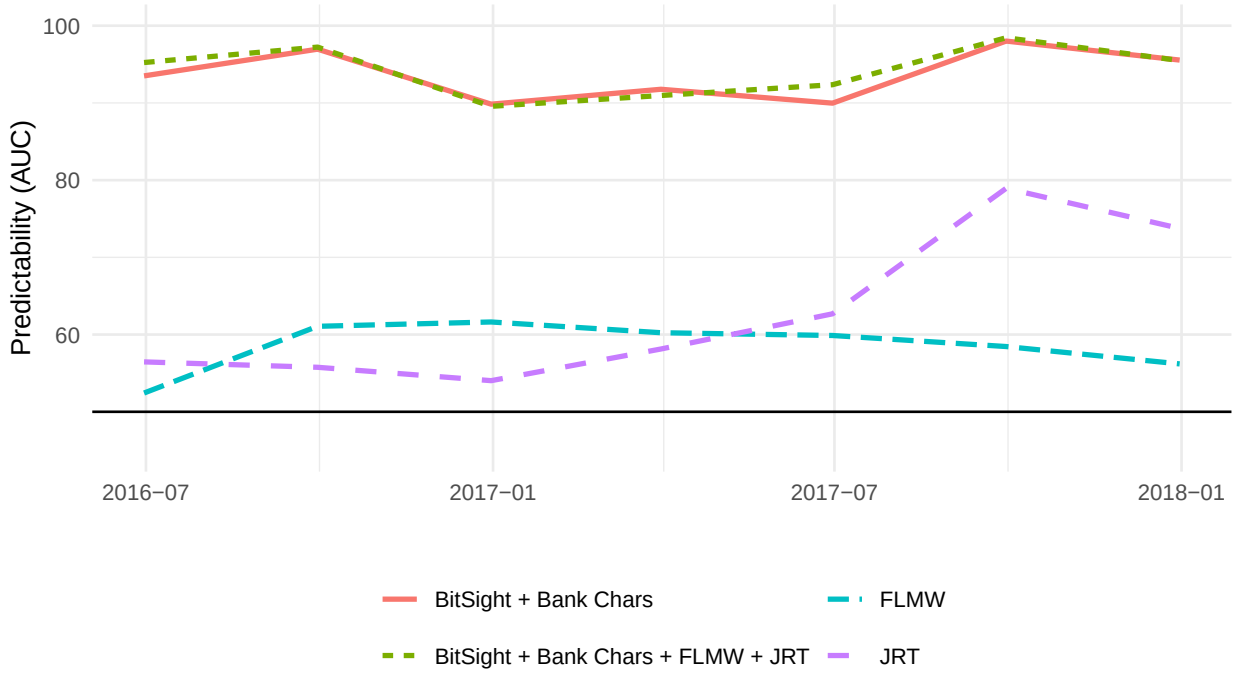
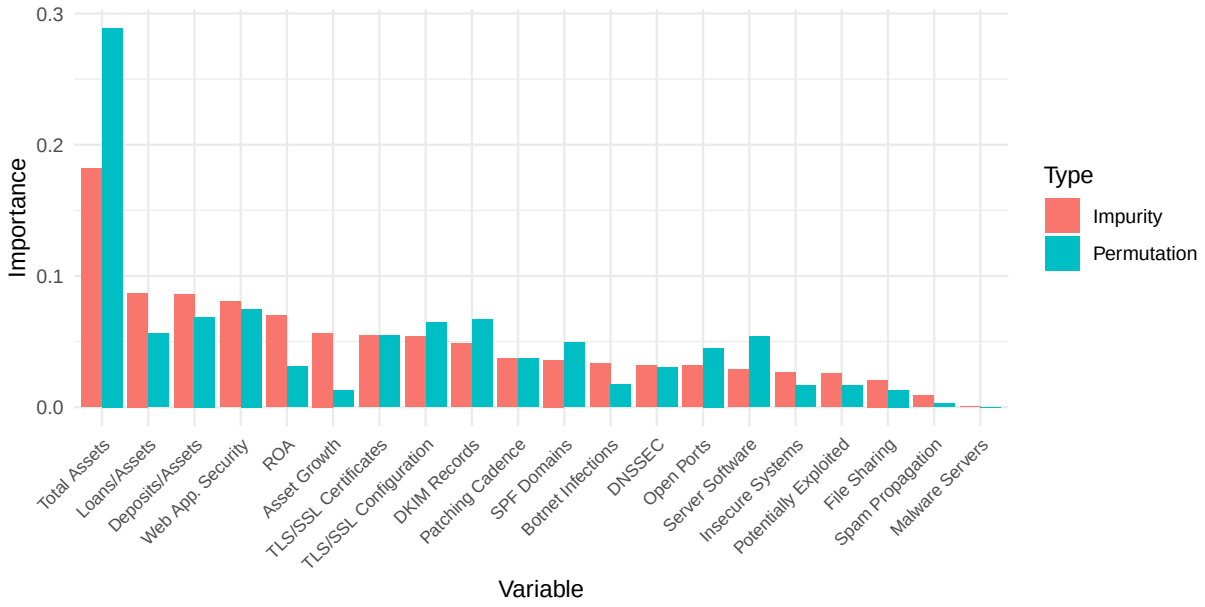


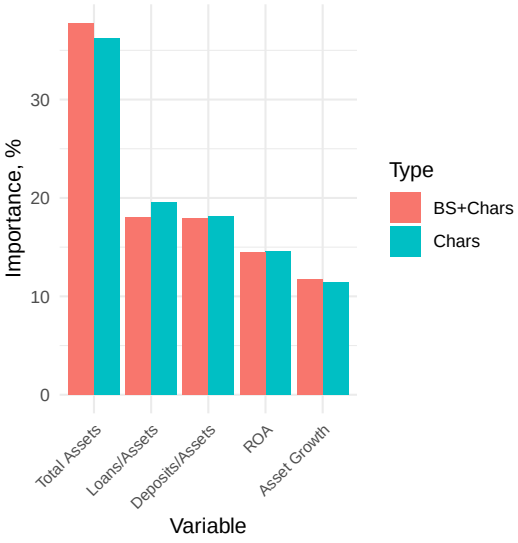
Figure 11: Importance of Independent Variables

This figure presents the variable importance results from the predictive model in three panels. Panel (a) shows the relative ranking of all predictors using both impurity-based and permutation measures. Panel (b) compares the contribution of firm characteristics alone with their contribution when combined with BitSight risk indicators, while Panel (c) compares the contribution of BitSight risk indicators alone with their contribution when combined with firm characteristics. Color coding distinguishes the different variable sets and importance measures across panels, providing a visual comparison of their relative influence in the model.

(a) All Variables



(b) Characteristics, Relative Importance Alone and with BitSight Risk-indicators



(c) BitSight Risk-indicators, Relative Importance Alone and with Characteristics

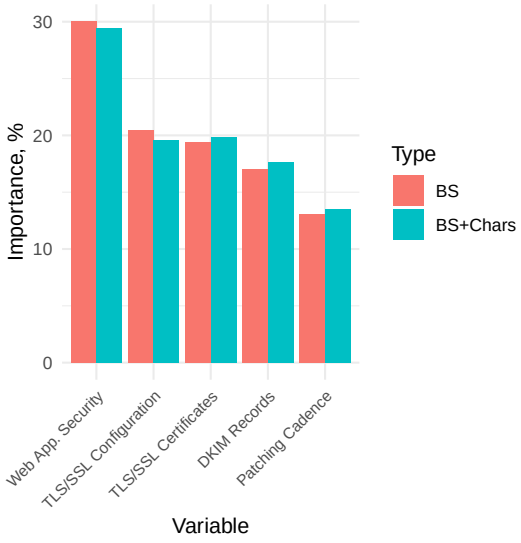


Figure 12: Selected Features

The figure shows BitSight cyber risk signals selected by the model each quarter between 2017 Q1 and 2024 Q4. Rows correspond to the fifteen candidate BitSight risk vectors and columns to test quarters; a filled cell indicates that the vector was among the top-five features retained by the backward selection procedure that quarter. The chart reveals which indicators are persistently informative and which enter only during specific intervals.

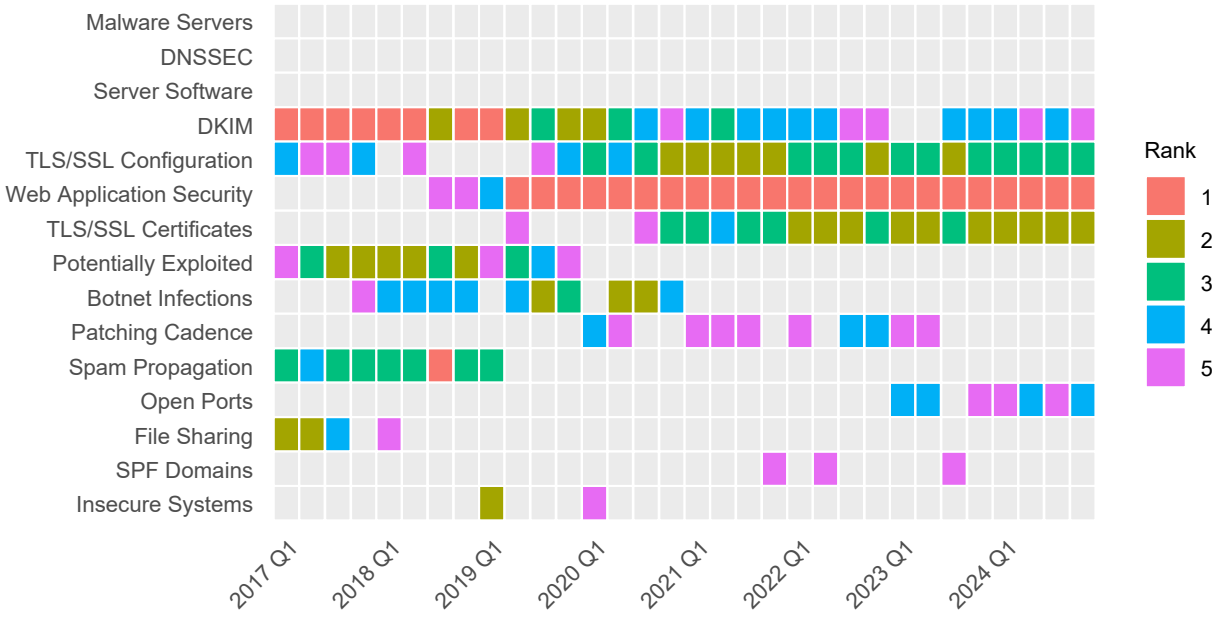
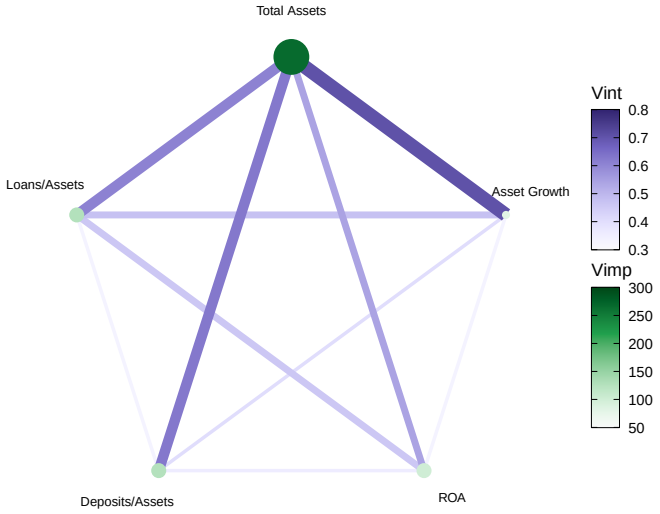


Figure 13: Variable Importance and Interaction Network

This figure presents network diagrams illustrating the importance and interactions of variables within two groups: bank characteristics (panel a) and BitSight risk indicators (panel b). Each diagram represents variables as nodes, with node size reflecting variable importance and edge thickness representing the strength of interaction between variables. The color gradients correspond to the magnitude of variable importance (Vimp) and interaction strength (Vint). Panel (a) displays the relationships among bank characteristics, while panel (b) shows the relationships among risk indicators, with both highlighting the most influential variables and strongest interactions in their respective groups.

(a) Bank Characteristics



(b) BitSight Risk-indicators

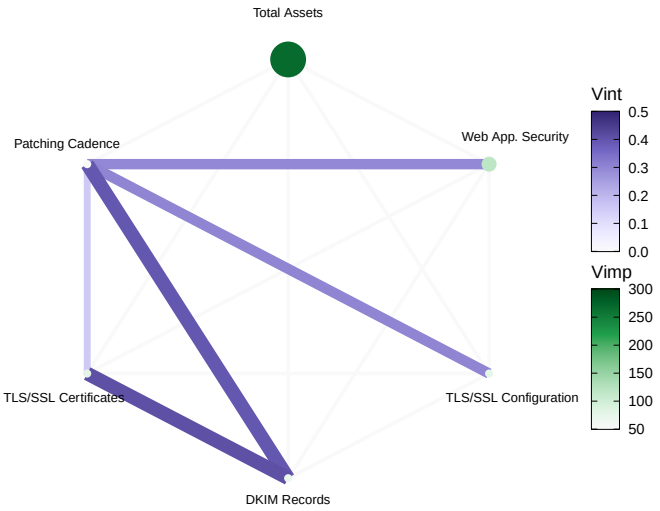


Figure 14: Partial Dependence Plots – Bank Characteristics

This figure presents partial dependence plots for the five bank characteristics used in the analysis. Each panel isolates the marginal effect of a single characteristic on the predicted outcome, holding other variables constant. The horizontal axis in each plot represents the range of observed values for the characteristic, and the vertical axis shows the corresponding change in the predicted probability scale. These plots illustrate the functional form of the relationship between each bank characteristic and the model's prediction.

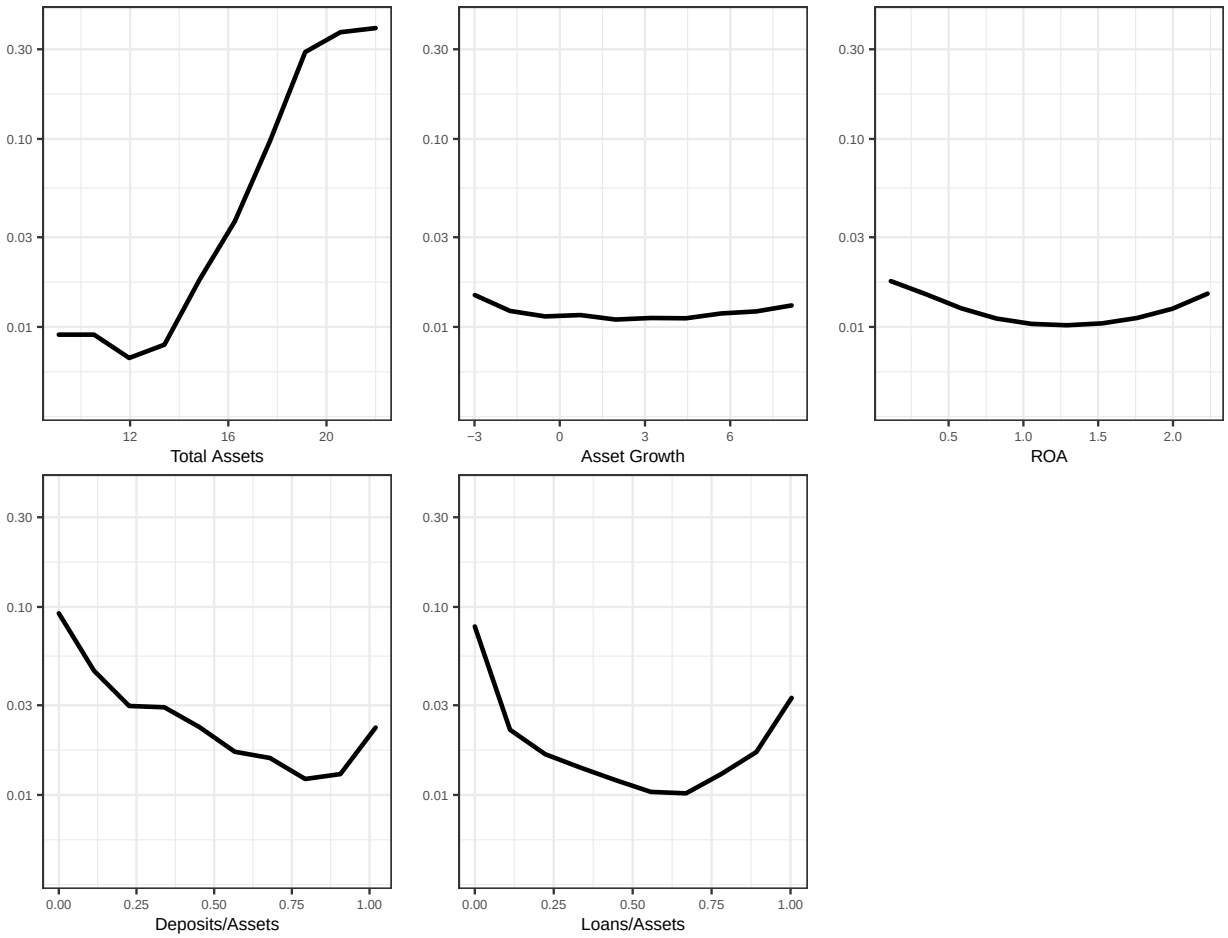


Figure 15: Partial Dependence Plots – BitSight Risk-indicators

This figure presents partial dependence plots for the BitSight risk-indicators. Each panel shows the relationship between a given risk-indicator and predicted incident likelihood, holding other variables constant. The horizontal axis displays the range of the respective indicator values, while the vertical axis shows the associated partial dependence.

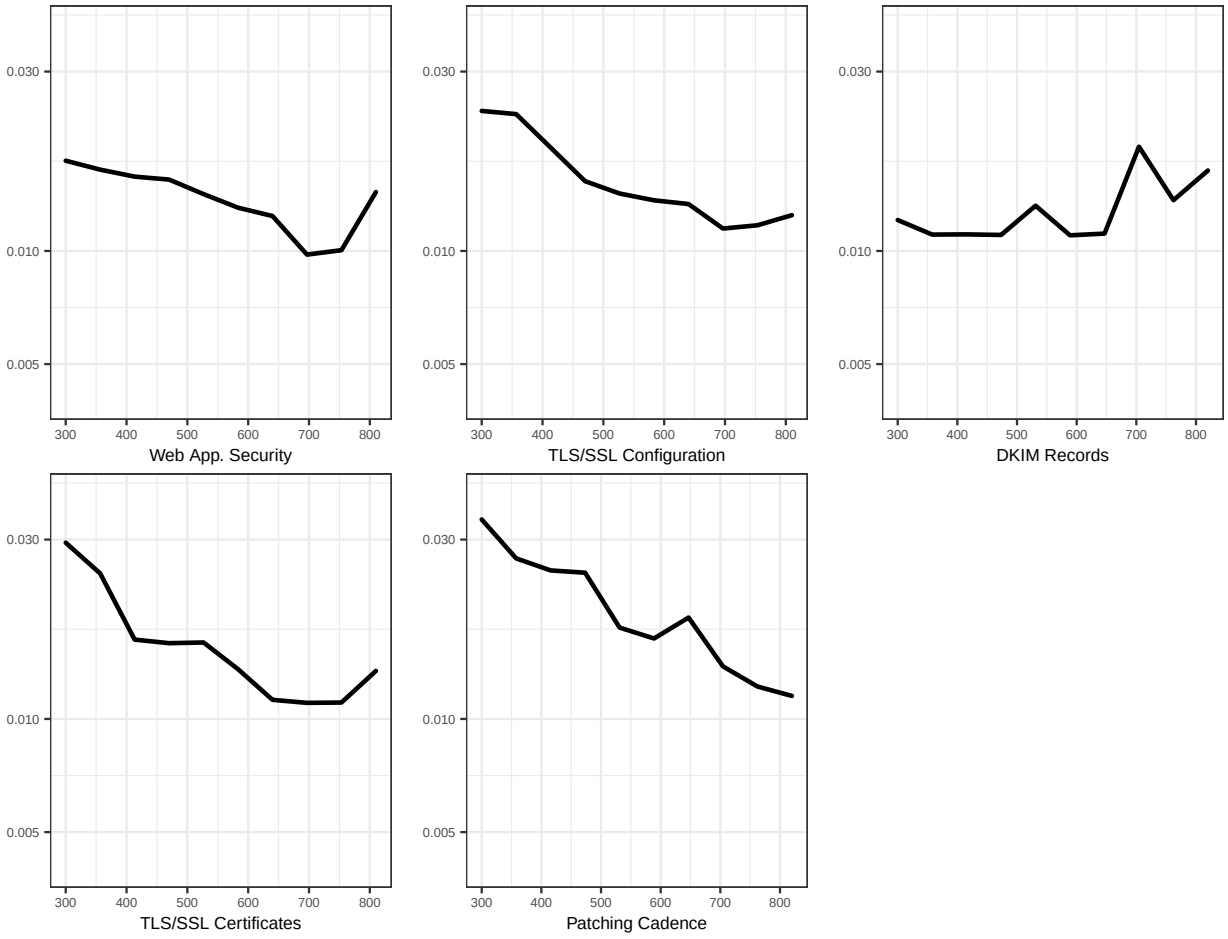


Figure 16: Pairwise Dependencies – Selected BitSight Risk-indicators

This figure displays pairwise partial dependence plots for selected BitSight risk-indicators. The diagonal panels show the individual partial dependence of each variable on the predicted probability. The upper triangle presents heatmaps of predicted probabilities for each pair of variables, while the lower triangle contains scatter plots of the underlying data points colored by predicted probability. Color gradients in the heatmaps range from blue (lower predicted probability) to red (higher predicted probability), illustrating how predicted outcomes vary across the joint distribution of the selected indicators.