Federal Reserve Bank of Dallas

What Drives Cyber Losses at U.S. Banks? Potential Statistical Markers

Anthony Murphy, Michael L. Tindall, Kelly Klemme,
Joseph I. Suek and Seth J. Dunbar

Working Paper 2520 **May 2025**

Research Department

https://doi.org/10.24149/wp2520

What Drives Cyber Losses at U.S. Banks? Potential Statistical Markers*

Anthony Murphy†, Michael L. Tindall‡, Kelly Klemme§,
Joseph I. Suek± and Seth J. Dunbar°

May 2025

Abstract

Bank supervisors and regulators are keen to understand and mitigate bank cyber risks. We model average annual loss (AAL) rates from "attritional" cyber-attacks and other cyber events using new, individual bank level data from the CyberCube "analytics platform" combined with standard bank performance measures. We estimate a variety of regression models to robustly identify the systematic drivers of these loss rates. We find that cyber risk AAL loss rates are significantly U-shaped in bank size, contrary to the view these risks are declining in bank size. Bank cyber risk contains a large idiosyncratic component, so apart from bank size, the explanatory power of standard bank performance measures is limited. Controlling for bank size, more profitable and efficient banks have lower cyber related loss rates.

Keywords: Banks, cyber losses, econometric models.

JEL Codes: C21, C54, G21.

---

*The views expressed here are those of the authors and not those of the Federal Reserve Bank of Dallas or the Federal Reserve System. The authors thank the Cyber Risk Data Committee of the Federal Reserve Board for providing the data used in this analysis, CyberCube for technical assistance, Ben Munyan, our discussant Ping McLemore and other participants in the 2025 Interagency Risk Quantification Forum for helpful comments and suggestions.

†Corresponding author: Anthony Murphy, Research Department, Federal Reserve Bank of Dallas, anthony.murphy@dal.frb.org.

‡Corresponding author: Michael L. Tindall, Banking Supervision Department, Federal Reserve Bank of Dallas, michael.tindall@dal.frb.org.

§Kelly Klemme, Banking Supervision Department, Federal Reserve Bank of Dallas.

±Joseph I. Suek, Banking Supervision Department, Federal Reserve Bank of Dallas.

°Seth J. Dunbar, Banking Supervision Department, Federal Reserve Bank of Dallas.

1. Introduction

Cyber risks are a major concern to bank supervisors and regulators. Yet, to date, these officials have not had access to measures of potential bank-level cyber loss rates, or robust estimates of the relationship between these loss rates, bank size and standard measures of bank performance. Likewise, bank examiners, who must evaluate cyber risks, do not have at their disposal risk markers that would help them identify banks with heightened risk from cyber-attacks and other cyber incidents.

Leveraging recent bank-level estimates of average annual losses (AAL) due to cyber incidents, we examine whether AAL loss rates can be modeled using standard bank performance data familiar to policymakers and regulators, e.g., using some of the performance measures in Uniform Bank Performance Reports. Although a large fraction of the variation in cyber risk loss rates appears to be idiosyncratic and not explained by traditional measures of bank performance, a few statistical markers / systematic relationships are present in the data.

We find that cyber risk AAL loss rates are significantly U-shaped in bank size, contrary to the view these cyber risks are declining in bank size. Apart from bank size, the explanatory power of standard bank performance measures is limited. We find that, controlling for bank size, more profitable and efficient banks have lower cyber AAL loss rates. To the best of our knowledge, these are seminal results not found in prior research.

2. Cyber Risk Annual Average Losses

Cyber risk AAL stands for average annual loss or the amount of money that is expected to be lost on average each year due to cyber incidents. Insurance companies use AAL to assess different types of risk and determine the appropriate coverage amounts. AAL measures the immediate risk of cyber incidents rather than the likelihood of attack in the future.

We use the bank-level cyber risk AAL loss rates produced by CyberCube (https://www.cybcube.com), using its proprietary analytics platform for cyber risk scenario modeling.[1] CyberCube models cyber risk

[1] CyberCube defines AAL as "a widely used loss statistic that has a diverse range of applications in catastrophe risk management, primarily for actuarial pricing. ... [It} estimates the annual policy premium needed to cover losses from modeled cyber catastrophes over time, assuming the exposure remains constant."

and losses using AI methods to incorporate data from a variety of sources, and stochastic simulations involving a variety of potential cyber incidents.

CyberCube calculates bank-level AAL loss rates from routine, non-catastrophic, cyber incidents by averaging across 50 thousand loss simulations of 28 scenarios, using a lognormal loss severity distribution and a Poisson event arrival distribution.[2] The 28 scenarios include the following:

- Cash and data theft (online banking services, enterprise payroll provider, mobile point of sale provider, e-commerce platform).
- Denial of service (cloud services provider, leading DNS provider).
- Destructive malware (cloud services provider, endpoint operating system, server operating system).
- Extortion (point of sale vendor).
- Outages (content delivery network, mobile network provider, internet service provider, leading electricity utility).
- Malicious configuration (cloud services provider).
- Ransomware (cloud services provider, file sharing provider, content management system, email service provider, endpoint operating system, server operating system).

The modeled losses cover business interruption, contingent business interruption, data restoration, extortion payments, fund transfer fraud, investigation and response, legal liability and regulatory costs. Physical damage and intangible losses are excluded.

The CyberCube AAL loss rate projections are noisy and uncertain. On the one hand, the projections may understate bank cyber risk since they are based in part on historical insurance claim data, which may understate actual losses. They may not adequately capture the effects of novel, previously unseen, types of cyberattacks and incidents. (Simultaneous cyber incidents at multiple service providers or cyber incidents that spread from one service provider to another are not captured in the CyberCube attritional loss model.) On the other hand, the AAL loss rates may overstate cyber risk since banks and their service providers mitigate future risk and losses from past cyber incidents, both at their own firm and across the industry.

[2] Formally we model “attritional” cyber risk losses resulting from independent events that affect one bank. We will model “catastrophic” (CAT) cyber losses arising from events affecting multiple banks in a future version of this paper.

The CyberCube projections are computed using AI tools as well as formal, statistical models. Statistical test measures are not currently generated. In addition, since the data are taken from a variety of sources, data quality checks are difficult to perform. To some extent, the impact of measurement error in the AAL loss rates is reduced since the winsorized loss rate is always the dependent variable in our regressions.

Cyber risk AAL loss rates are generally quite small – one or two base points (bps) of revenue - as shown by the summary statistics in **Table 1.**[3] The mean loss rates for the 3,622 banks is 1.19 bps, with a median of 1.06 bps and a 90% percentile value of 1.50 bps. The mean loss rate on a portfolio basis ranges from 1.06 basis points for foreign banking organizations (FBO) banks to 1.93 bps for the systemically important Large Institution Supervision Coordinating Committee (LISCC) banks. The dispersion in 90th percentile AAL loss rates is even wider, ranging from 1.19% for FBOS to 2.72 bps for LISCC banks.

3. Bank Performance Data

The bank performance measures used in our paper include bank total assets, asset growth, the number of full-time employees, the return on assets (ROA) and equity (ROE), the common equity tier 1 ratio (CET1), a dummy for community banks not reporting CET1, net noncore funding dependence, the tier 1 leverage ratio, the efficiency ratio, the share of loans 90 days past due or non-accrual, outside data-processing expenses as a percent of income, a dummy for banks not reporting outside data-processing expenses because they were small, personnel expenses as a percent of income, and overhead expense as a percent of income[4]. Income refers to the sum of net interest and noninterest income, and the performance data are at the bank holding company or lead bank level for 2023 Q4.

We chose to winsorize all or our continuous data to reduce the impact of outliers. Summary statistics are set out in Table 2. When estimating the models in the next section, we generally pruned highly collinear, related variables (such as the efficiency ratio and the ratio of overhead expenses to income) since the second variable conveyed very little additional information beyond what was in the first variable. The exception is when we estimated double machine learning models.

[3] We refer to both banks and bank holding companies etc. as “banks”.
[4] The bank efficiency ratio is the ratio of non-interest operating costs to the sum of tax equivalent net interest income plus non-interest income.

We also choose to model the log of the AAL loss rate and follow convention by using log total assets as a measure of bank size. We also constructed various cubic splines in log total assets to explore the possibility of a non-linear relationships between the log AAL loss rate and log total assets. As shown in Figure 1, the unconditional relationship between the cyber risk AAL loss rate and size (log total assets) is highly nonlinear and U-shaped. The loss rate steadily declines with bank size, hitting a minimum around $0.87B in total assets, and then increases steadily albeit not reaching the same loss rate found as the smallest banks. The minimum falls between the 50th and 75th of total assets in our dataset, and corresponds to a small community banking organization.

We are rather agnostic regarding the a priori contributions of the various bank performance measures to cyber risk, preferring to let the data speak for themselves. When considering our estimated models, one should be careful about interpreting statistically significant effects as causal effects since we don't have an underlying model of cyber risk. Instead, it may be best to just think of statistically significant results as possible risk markers.

Admittedly, some bank performance measures may have intuitive links to cyber risk. Regulatory authorities generally view banks with high asset growth rate cautiously. In such cases, regulatory concerns center on whether these banks have adequate risk-management systems in place and on credit-quality and liquidity issues. Inadequate risk-management systems would tend to be associated with high cyber risks. In addition, high asset growth achieved through mergers and acquisitions can lead to greater organizational complexity, expanding the potential for cyberattacks and increasing cyber risks. Conversely, banks with high asset growth may have greater financial resources to create improvements in cyber security, reducing cyber risks. Thus, the effect of high asset growth may hinge on its source.

Banks with high ROA/ROE may have high cyber risks because cyber criminals perceive them as having valuable data and financial resources that can be exploited. Likewise, banks with high asset levels may be perceived as high-value targets, increasing their cyber risks. These banks may also exhibit greater complexity making them more vulnerable to attack.

In the same vein, banks with high tier 1 common equity ratios may be perceived as higher-value targets, contributing to greater cyber risks. Or, alternatively, banks with a high ratio may simply offer cyber criminals a high payoff regardless of any value perceptions. Banks with high noncore funding may operate with reduced internet exposure, obtaining funds through brokered CDs and the fed funds market, thus decreasing cyber risk. Banks with high efficiency ratios have high operating expenses, which could lead to reduced spending on cyber security and greater cyber risk.

Banks may experience high data processing expenses because they are maintaining inefficient legacy systems which are more vulnerable to cyberattacks. In addition, these banks may have increased data handling activities and more complex systems, making them more vulnerable to cyberattacks and increasing cyber risks. High personnel expense can be associated with lower cyber risk if the spending is directed toward acquiring skilled cyber security staff and improving cyber security practices.

4. Modeling Cyber Risk AAL Loss Rates

(a) OLS Results

Some OLS regression results are presented in **Table 3**. In Model 1 the log of the AAL loss rate is linear in bank size whereas, in Model 2, bank size enters non-linearly as a restricted cubic spline (e.g., Carleton and McGee, 1970; Suits et al., 1978). The restricted cubic spline function is (a) linear in log assets before and after the bottom and top knots, (b) consists of piecewise cubic polynomials between adjacent knots and is (c) continuous and smooth at each knot, with continuous first and second derivatives.[5]

Model 1 suggests that cyber risk losses decline with size (log total assets), the most significant variable in the regression. Other statistically significant variables include the return on assets, noncore funding dependence, tier 1 leverage ratio, efficiency ratio and the ratio of data processing expense to income. As noted in the previous section, one can construct an intuitive justification for these variables.

However, once the linear term in bank size is replaced by a restricted cubic spline, bank size becomes even more significant and the bank performance variables – ROA, net noncore funding dependence, the tier 1 leverage ratio and the efficiency ratio - become far less significant (Model 2).

(b) Robustness of U-Shaped Bank Size Effect

---

[5] The cubic spline with three knots was selected using the Bayesian information criterion (BIC). The knots used the default (10^{th}, 50^{th} and 90^{Th} percentile) knot positions in Harrell (2001), Table 2.3.

The results in **Table 3** suggest that the most significant driver of the cyber risk AAL loss rate is bank asset size, with possibly minor roles for a handful of bank performance variables. To verify this finding, we estimate bank size effects using the double machine learning / cross-fit partialing out methods described in Belloni, Chernozhukov and Hansen (2014a, 2014b), and Belloni, Chernozhukov and Wei (2016), inter alios.

The double machine learning method controls for a range of additional right hand side variables and, unlike the Lasso and its variants (e.g., Hastie et al., 2009; Tibshirani, 1996), generates the coefficients and standard errors for a subset of the covariates, i.e. the three restricted cubic spline bank size / log total assets terms in this case. Loosely speaking, the results In **Table 4** have the added advantage of being estimates of values from the true model that generated the data being analyzed.

The estimated bank size restricted cubic spline in **Table 4** is U shaped and, apart from a difference in means, is very similar to the estimated restricted cubic spline in Table 3, model 2, as shown in **Figure 3**. This similarity gives us confidence that, in our data, the cyber risk AAL loss rate – bank size relationship is indeed significantly U shaped.

(c) *Further Robustness Checks*

We now turn to consider some Autometrics results as an additional check. Autometrics is an algorithm for automatic model selection within the general-to-specific framework, also known as the 'Hendry' or 'LSE' methodology. It allows for extensive and efficient search over models, while avoiding overfitting and maintaining statistical congruence (e.g., Doornik, 2009; Hendry and Doornik, 2014, Castle, Doornik and Hendry, 2023).

We choose to estimate "small" models, a gauge or setting within Autometrics that reduces the expected number of falsely selected variables. We selected models with and without large outlier detection. Since we have a single cross section of possibly noisy AAL loss rate data, even after winsorizing, we believe it is worthwhile presenting results that account for large outliers.

Two estimated Autometrics models are set out in **Table 5**. The bank size restricted cubic spline terms are very significant in both models. Autometrics only retains a small number of bank performance variables in addition to the non-linear bank size effect.

In Model 3, three bank performance variables - the tier 1 leverage ratio, efficiency ratio and personnel expenses as a share of income – are retained. However, only one variable – the return on assets – is retained in model 4 when large outliers are controlled for / dummied out. Model 4 suggests that, controlling for bank size, the cyber risk AAL loss rate is lower the more profitable the bank – an intuitive finding.

(d) Discussion of U-Shaped Bank Size Effect

Mid-size banks might be less affected by cyber risk compared to small or large banks due to a combination of interrelated factors. First, economies of scale may play a role. Small banks often lack the financial resources to invest in robust cybersecurity infrastructure, making them more vulnerable to attacks. Mid-size banks typically have enough resources to invest in effective cybersecurity measures while not being overly burdened by the complexity of large-scale operations. Although large banks invest heavily in cybersecurity, their size and complexity can make them attractive targets for attackers, who aim for high-value returns.

Second, attractiveness to attackers may be influenced by asset size. Small banks may be perceived as "low-hanging fruit" by cybercriminals because they may have weaker defenses. Mid-size banks may not be as attractive as large banks because the potential payoff is smaller, and they are often better protected than small banks. Large banks may be seen as high-value targets because of their size, large customer base, and global presence.

Third, complexity of operations may have an impact. Small banks may have simpler operations which rely on outdated or less secure systems. Mid-size banks have more standardized and modernized systems compared to smaller banks, while avoiding the overwhelming complexity of large banks. Large banks, with their vast networks and integrations with multiple partners and platforms, can create more potential entry points for attackers.

Fourth, differences in regulatory and compliance pressures may create opportunities for cyber-attacks based on bank size. Small banks may struggle to meet stringent regulatory standards due to limited resources. Mid-size banks may be more likely to meet compliance requirements effectively without excessive strain on their resources. Large banks face significant regulatory scrutiny and may experience challenges in implementing uniform protections across their extensive operations.

Fifth, targeting trends and tactics by cybercriminals may create size-related effects. Cybercriminals often target institutions where they perceive the best balance between effort and reward. Mid-size banks may fall in a "sweet spot" where the effort to breach their systems may not justify the potential gains.

Finally, the ability to respond and recover from cyber-attacks may have a size component. Small banks have limited resources for recovery which can exacerbate the impact of an attack. Mid-size banks may have better crisis-management plans and can recover more effectively from cyber incidents. Large banks, while capable of significant recovery efforts, may be hampered by the scale of operations making them slower and less able to respond. By combining these factors, mid-size banks may naturally strike a balance between preparedness and reduced attractiveness to attackers, positioning them to be less affected by cyber risks.

5. Summary and Conclusion

We exploit a new dataset - CyberCube estimates of bank cyber risk average annual loss (AAL) rates combined with standard measures of bank performance - to explore the drivers of potential bank cyber losses. We estimate a variety of regression models of cyber-related AAL loss rates to robustly identify the systematic drivers of these losses.

We find that cyber risk AAL loss rates are markedly U-shaped in bank size, contrary to the view these cyber risks are declining in bank size. Since bank cyber risk has a large idiosyncratic component, it is probably unsurprising that, apart from bank size, the explanatory power of standard bank performance measures is limited. Controlling for bank size, more profitable and efficient banks have lower cyber AAL loss rates. To the best of our knowledge, these findings are novel.

References

Belloni, A., Chernozhukov, V., Hansen, C.B. (2014a), "High-dimensional methods and inference on structural and treatment effects", *Journal of Economic Perspectives*, 28, 29–50.

Belloni, A., Chernozhukov, V., Hansen, C.B. (2014b), "Inference on treatment effects after selection among high-dimensional controls", R*eview of Economic Studies*, 81, 608–650.

Belloni, A., Chernozhukov, V., Wei, Y. (2016), "Post-selection inference for generalized linear models with many controls", *Journal of Business & Economic Statistics,* 34, 606–619.

Carleton, W., & McGee, V. (1970), "Piecewise regression", *Journal of the American Statistical Association*, 65(131), 1109-24.

Castle, J.L., Doornik, J.A., Hendry, D.F. (2023), "Robust Discovery of Regression Models", *Econometrics and Statistics*, 26, 31-51.

Doornik, J. (2009), "Autometrics" in Castle, J., Shephard, N. (Eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of David F. Hendry*, Oxford, Oxford University Press.

Hastie, T., Tibshirani, R., Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*, New York: Springer.

Harrell, F.E. (2001), *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*, New York: Springer-Verlag.

Hendry, D.F., Doornik J.A. (2014), *Empirical Model Discovery and Theory Evaluation*, Cambridge: MIT Press.

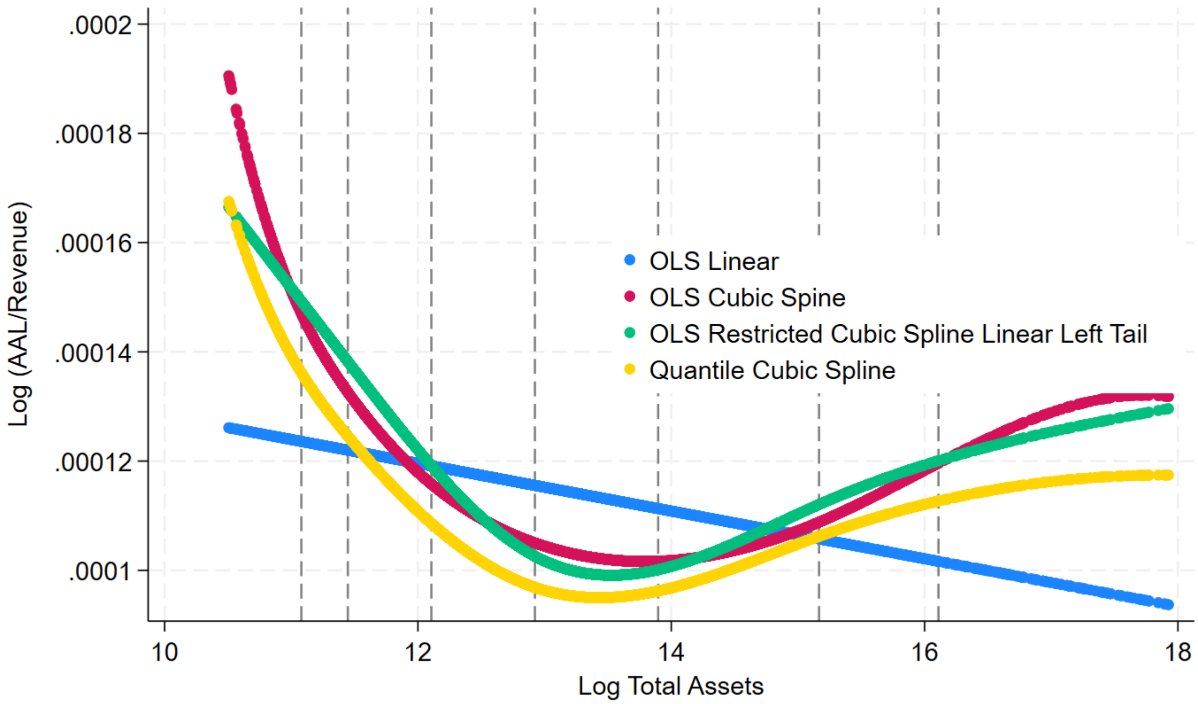
Schreft, S., Lee, S.J. (2023), "Systemic Cyber Risk in the Financial System: A First Look", memo, Federal Reserve Board of Governors.

Suits, D. B., Mason, A., & Chan, L. (1978), Spline functions fitted by standard regression methods, *The Review of Economics and Statistics*, 60(1), 132-39.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society, Series B*, 58(1), 267–88.

Zou, H., Hastie, T. (2005), "Regularization and Variable Selection via the Elastic Net", *Journal of the Royal Statistical Society, Series B*, 67(2), 301-320.

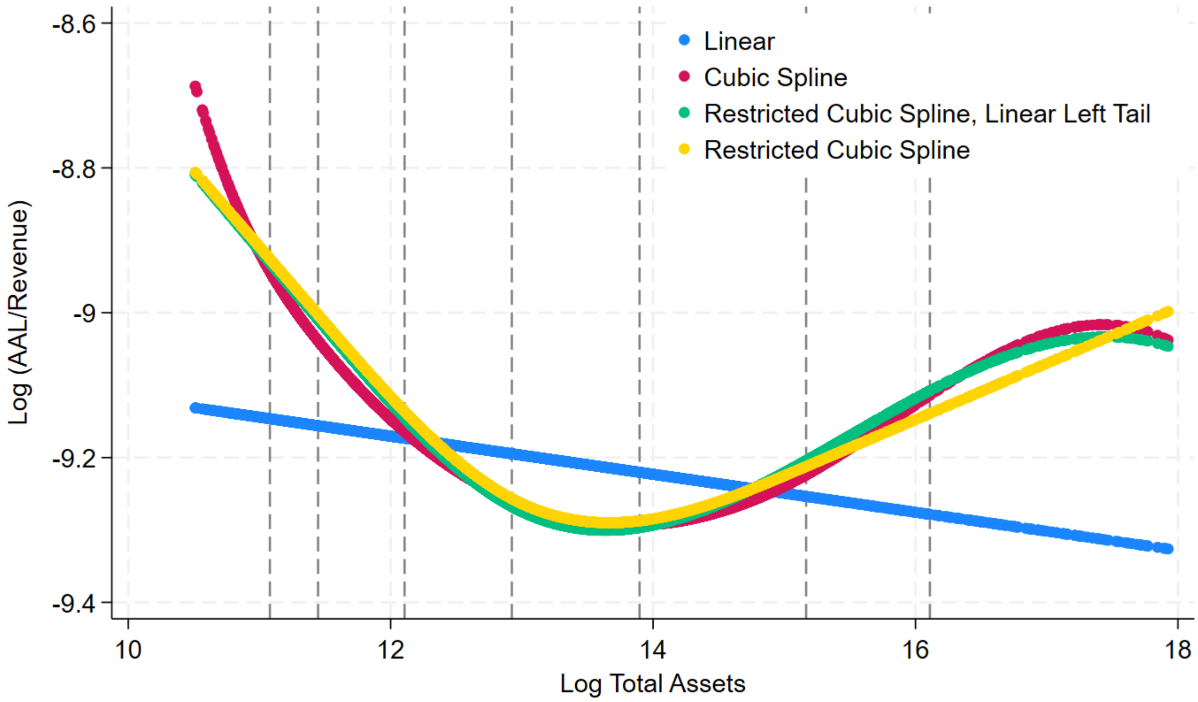
Figure 1: Regular Cyber Incident Loss Rates and Bank Size – Unconditional Relationship



Source: CyberCube and authors calculations.

Notes: The plot shows the fitted value of the OLS and quantile regressions of log (AAL / Revenue) on log Total Assets, a proxy for bank size, or cubic splines in log Total Assets. The plot shows the conditional relationship since no bank performance measures were included in the relationship. The variables are winsorized and the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of log Total Assets are shown as dotted vertical lines.

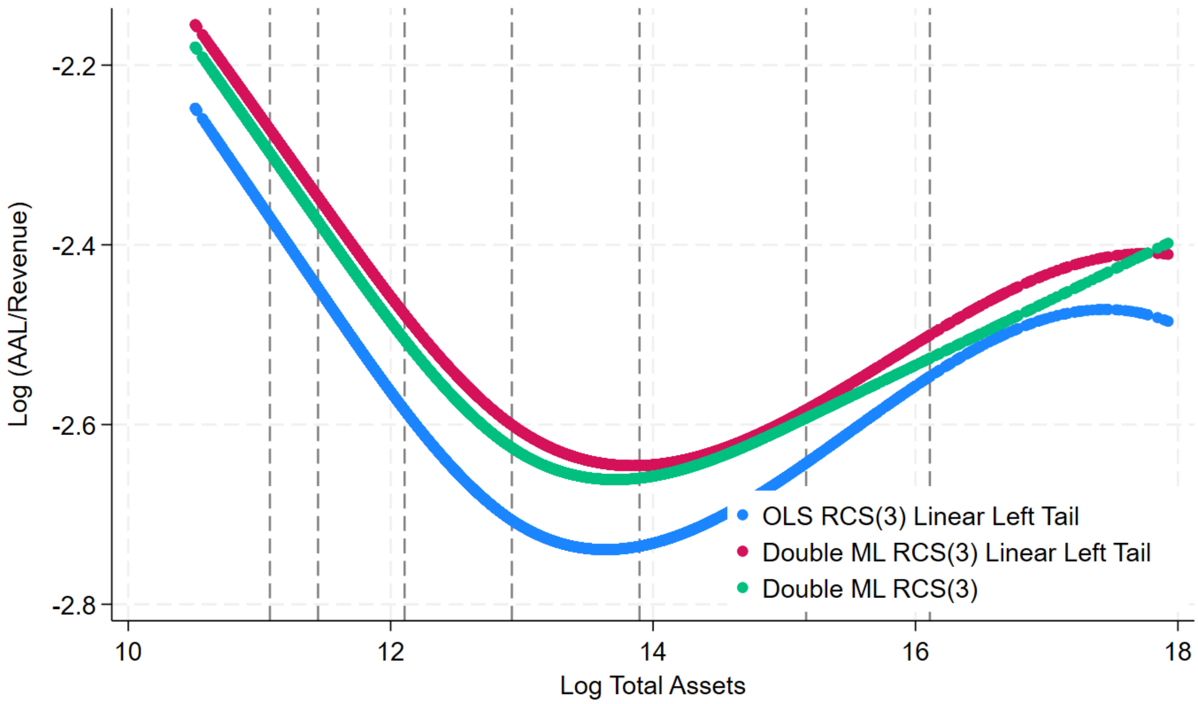
Figure 2: Regular Cyber Incident Loss Rates and Bank Size – Conditional Relationship



Source: CyberCube and authors calculations.

Notes: The bank size measure is proxied by log total assets. The estimated linear and restricted cubic splines, with three knots, are based on the OLS estimates in Table 3. All the continuous variables in Table 3 are winsorized. The 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of log Total Assets are shown as dotted vertical lines.

Figure 3: Comparison of Estimated Restricted Cubic Splines – OLS vs Double Machine Learning



Source: CyberCube and authors calculations.

Notes: The bank size measure is proxied by log total assets. The estimated restricted cubic splines are based on the OLS estimates in Table 3 and the double machine learning estimates in Table 4. All the continuous variables are winsorized. The 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of log Total Assets are shown as dotted vertical lines.

Table 1: Regular Cyber Incident AAL Loss Rates By Federal Reserve Bank Portfolio (Basis Points)

Portfolio	N	Mean	SD	Percentiles					
				10th	25th	50th	75th	90th	99th
CBO	3,380	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
RBO	145	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FBO	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
LBO	18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.182	[illegible]
LISCC	12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Other	20	[illegible]	1.028	0.860	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Total	3,622	[illegible]	[illegible]	[illegible]	0.935	[illegible]	[illegible]	[illegible]	[illegible]

Source: CyberCube and author's calculations. Notes: The AAL loss rates are expressed in basis points of revenue / net operating income. The portfolio categories are community, regional, foreign, large and Large Institution Supervision Coordinating Committee (LISCC) banking organizations and a combination of three savings and loan holding companies and 17 institutions not supervised by the Federal Reserve.

Table 2: Summary Statistics (Winsorized)

Variable	N	Mean	SD	Percentiles				
				5th	25th	50th	75th	95th
ln (AAL / Revenue)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
log Assets	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Assets per FT Employee	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
No of FT Employees	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Asset Growth (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROA (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ROE (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CET1 (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CET1 Not Reported (0/1)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Noncore Funding (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Tier 1 Leverage Ratio	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Efficiency Ratio	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Loans 90 DPD / Non-Accrual (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Data Processing Expense / Income (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Date Processing Expense Not Reported (0/1)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Personnel Expense / Income (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Overhead Expense / Income (%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Sources: CyberCube, Call and FR Y-9C reports and authors calculations. Note: The top and bottom 1% of the continuous variables, i.e. the variables other than the two “data not reported” dummy variables, are winsorized.

Table 3: OLS Regression Results - Bank Size Is Main Determinant of Cyber AAL Rates

Dependent Variable = log (AAL / Revenue)	Model 1 Linear in Size / Log Total Assets		Model 2 Restricted Cubic Spline in Size	
	Coefficient	t-stat	Coefficient	t-stat
Intercept	-8.8554	122.53	-6.5972	62.67
Size = log total assets	-0.0263	8.19	-0.2092	20.82
Additional Restricted Cubic Spline Size Term	-	-	0.2414	27.42
Asset growth	-0.00003	0.07	0.0007	1.69
ROA	-0.0322	3.14	-0.0192	2.06
CET1	-0.0021	1.00	0.0001	0.09
CET1 NR dummy	-0.0014	0.95	0.0032	0.12
Net noncore funding dependence	-0.0014	5.88	-0.0006	2.50
Tier 1 leverage ratio	0.0084	4.51	0.0038	2.21
Efficiency ratio	0.0032	4.88	0.0019	1.48
Data processing expense / income	0.0038	2.15	0.0007	0.44
Data processing expense / income NR dummy	0.0705	4.60	-0.0102	0.72
Personnel expense / income	-0.0048	5.61	-0.0006	0.75
Adjusted R^2	0.09		0.25	
No. of observations	3,510		3,510	

Notes: The largest and smallest 1% of the values of all the continuous variables are winsorized. Income = net operating income. RCS = restricted cubic spline with three knots chosen using the 10^{th}, 50^{th} and 90^{th} percentile of log Total Assets, the default knot positions in Harrell (2001), Table 2.3. NR = not reported, see text for details.

Table 4: Double Machine Learning Estimates of Bank Size Effects - Restricted Cubic Spline With Linear Left Tail

Dependent Variable = log (AAL / Revenue)	Restricted Cubic Spline in Size / Log Total Assets	
	Coefficient	t-statistic
Bank Size / Log Total Assets (TA)		
Log TA	-0.1462	-9.03
$(\log TA – Knot_1)^3$	0.0072	11.02
$(\log TA – Knot_2)^3$	0.0076	-5.27
$(\log TA – Knot_3)^3$	-0.0042	-2.27
Wald $\chi^2(3)$	371.0	
No. of Observations	3504	

Notes: All continuous variables are winsorized using a 1% cutoff. The three knots are the 10th, 50th and 90th percentile of log Total Assets (TA). Ten folds were use for cross fitting in the double machine learning / cross-fit partialing out procedure. Eighteen controls were used. On average thirteen controls were used for partialing-out.

Table 5: Restricted Cubic Spline in Bank Size Autometrics Model Selections Results – More Efficient and Profitable Banks Appear To Have Lower Cyber Loss Rates

Dependent Variable = log (AAL / Revenue)	Small Model		Small Model With Large Outlier Detection	
	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	-6.6342	71.70	-6.4241	91.40
Size = log Total Assets	-0.2124	31.50	-0.2182	38.10
Additional RCS term in Size	0.2439	29.60	0.2464	34.50
ROA	-	-	-0.0240	4.47
Tier 1 leverage ratio	0.0038	2.63	-	-
Efficiency ratio	0.0013	5.06	-	-
Large outlier dummies (N =47)	No		Yes	
Adjusted R^2	0.25		0.43	

Notes: All continuous variables are winsorized using a 1% cutoff. The restricted cubic spline (RCS) in log total assets uses three knots. The three knots are the 10th, 50th and 90th percentile of log Total Assets (TA). The 47 large residuals / outliers are not reported.