DOI: 10.71317/jgst.2.11.2026.968
Journal of Global Social Transformation
Journal homepage: https://rjsaonline.org/index.php/JGST

A Hybrid Deep Learning and Machine Learning Framework for Quantifying the Macroeconomic Impact of Cyberattacks: Threat Intelligence, Vulnerability Assessment, and Predictive Resilience Modeling in National Economic Systems

Hafiza Saadia Shams[1], Dr. Muhammad Usman Malik[2], Dr. Zulfqar Khan[3]
[1] National Institute of Intelligence & Security Studies, Quaid-i-Azam University, Islamabad 45320, Pakistan.
Email: saadiashams25@gmail.com
[2] National Institute of Intelligence & Security Studies, Quaid-i-Azam University, Islamabad 45320, Pakistan.
Email: usmanmalik836@gmail.com
[3] Professor, National Institute of Intelligence & Security Studies, Quaid-i-Azam University, Islamabad 45320, Pakistan.
Email: dirniiss@qau.edu.pk

ARTICLE INFO

Submission:
September 30, 2026
Revised:
October 12, 2026
Accepted:
October 24, 2026
Available Online:
November 04, 2026

Keywords:
Cyber risk; Critical infrastructure; Macroeconomic impact; Deep learning; Extreme value theory; Cyber resilience.

Corresponding author:
Hafiza Saadia Shams
Email:
saadiashams25@gmail.com

ABSTRACT

Cyberattacks on energy, finance, health and logistics systems now produce losses large enough to register at the level of national economies, yet the quantitative tools that link threat intelligence to macroeconomic exposure remain fragmented. This paper develops a hybrid framework that couples advanced econometrics with deep and machine learning across three modules: threat intelligence, vulnerability assessment and predictive resilience. Using a compiled dataset of published series for 2015–2025 (NVD/CVE disclosures, FBI IC3 complaints and losses, IBM/Ponemon breach costs), ITU Global Cybersecurity Index scores for 25 economies and 24 documented macro-relevant incidents, we estimate HAC and wild-bootstrap elasticities, ARDL bounds and DOLS long-run relations, Bayesian conjugate regressions, sup-F break and Granger tests, quantile, Tobit and fractional-logit models, and a Poisson–lognormal loss distribution with extreme-value diagnostics. A from-scratch LSTM deep ensemble, deep MLP, random forest, gradient boosting and an ARX–LSTM hybrid are evaluated by rolling-origin forecasting with Diebold–Mariano tests and exact Shapley attribution. Reported losses grew 35.6% a year against 22.4% for vulnerability disclosures; disclosures Granger-lead losses. Gradient boosting gave the lowest out-of-sample error (log RMSE 0.113 against 0.362 for a random walk); the LSTM ensemble reached 0.191. The hybrid projection puts US-reported losses at USD 52.0 billion by 2028 (80% interval 41.0–65.9). Incident severity is lognormal with a heavy right tail, and parameter uncertainty roughly doubles the 99% Value-at-Risk. Projected expected losses range from 0.15% of GDP in the United States to 0.26% in Nigeria, with Pakistan and Ukraine close behind.

Introduction

Cyber incidents have moved from being a firm-level operational nuisance to a recognised source of macroeconomic and systemic risk. The US Federal Bureau of Investigation recorded USD 16.6 billion of victim-reported internet crime losses in 2024, up from USD 1.07 billion in 2015 (Federal Bureau of Investigation [FBI], 2025), while the average cost of a data breach to a US organization passed USD 10 million for the first time in 2025 (IBM Security, 2025). The White House Council of Economic Advisers estimated that malicious cyber activity cost the US economy between USD 57 and 109 billion in 2016 alone (Council of Economic Advisers, 2018). Single events now carry national-scale consequences: the NotPetya wiper of 2017, attributed to the Russian military intelligence unit known as Sandworm, disrupted global shipping, pharmaceuticals and logistics with damages put at about USD 10

billion (Greenberg, 2019); the 2024 Change Healthcare ransomware attack disrupted claims processing across the US health system; and in 2025 the attack on Jaguar Land Rover halted production for about five weeks and was classified by the UK Cyber Monitoring Centre as a Category 3 systemic event with an estimated economy-wide cost of GBP 1.9 billion (Cyber Monitoring Centre, 2025).

For critical infrastructure protection, these developments raise a question that goes beyond the engineering of individual controls: how large is the macroeconomic exposure of a national economic system to cyberattack, how is it evolving, and which economies are least resilient to it? Existing evidence answers parts of this question with different instruments. Cost-of-cybercrime studies build bottom-up inventories of losses (Anderson et al., 2013, 2019); actuarial work characterizes the heavy-tailed severity of cyber events (Edwards et al., 2016; Eling & Wirfs, 2019; Maillart & Sornette, 2010); corporate-finance studies measure market reactions and supply-chain propagation (Crosignani et al., 2023; Florackis et al., 2023; Kamiya et al., 2021); and policy institutions have proposed scenario frameworks for systemic and financial-stability risk (Bouveret, 2018; Kopp et al., 2017; Welburn & Strong, 2022). Very few studies, however, bring threat intelligence, national vulnerability and predictive modelling together in one quantitative pipeline, and fewer still test whether modern deep learning adds forecasting value when the macro data are short and the events are rare (Eling et al., 2021; Woods & Böhme, 2021).

This paper addresses that gap. We propose and estimate a **hybrid deep learning and machine learning framework** with three linked modules. Module A (*threat intelligence*) relates vulnerability disclosures and attack activity to reported losses using a battery of small-sample time-series econometrics and benchmarks them against a deep-ensemble long short-term memory (LSTM) network, a deep multilayer perceptron, tree ensembles and an ARX–LSTM hybrid. Module B (*vulnerability assessment*) models national cyber-capability with quantile, censored and fractional-response regressions and maps economies into resilience clusters. Module C (*shock severity*) fits extreme-value and parametric severity models to documented macro-relevant incidents and builds a Poisson–lognormal loss distribution for tail-risk measurement. A *predictive resilience* layer then combines the projected loss intensity, national GDP and the vulnerability gap in a Monte Carlo simulation to rank economies by expected annual cyber loss as a share of GDP.

The paper makes four contributions to the critical infrastructure protection literature. First, it offers an integrated, reproducible pipeline from open threat-intelligence series to country-level macro exposure, with every input traceable to a public source. Second, it applies a deliberately demanding econometric toolkit wild-bootstrap inference, ARDL bounds testing, dynamic OLS, Bayesian model comparison, bootstrap structural-break tests, quantile and Tobit models, and extreme-value diagnostics that is designed for the short samples typical of cyber-macro data, and it reports where those tests lack power rather than over-claiming. Third, it subjects deep learning to an honest rolling-origin horse race with formal Diebold–Mariano tests and exact Shapley attribution, which shows when deep models help and when simpler learners dominate. Fourth, it translates the results into a resilience ranking that is directly usable for prioritizing national cyber-capacity investment, especially in emerging economies such as Pakistan, Nigeria and Ukraine.

The remainder of the paper reviews the literature and states the hypotheses, describes the data and the framework, presents the results, discusses implications and limitations, and concludes.

Literature Review and Hypotheses

Measuring the Economic Cost of Cyberattacks

Early attempts to put a price on cybercrime produced estimates that differed by orders of magnitude, largely because they mixed direct losses, defence expenditure and indirect costs. Anderson et al. (2013) proposed separating these components and showed that indirect and defence costs typically dwarf criminal revenue; their update (Anderson et al., 2019) found that the structure of costs had changed with the shift to ransomware and business e-mail compromise but that measurement remained patchy (Anderson et al., 2019). Romanosky (2016) used a large incident database to show that the median firm-level cost is modest relative to revenue, while Eling and Wirfs (2019) showed that operational-loss data contain a small number of extreme cyber events that account for most of the total. Gordon and Loeb (2002) provided the canonical microeconomic model of security investment, showing that optimal spending is a bounded fraction of the expected loss. At the macro level, Dreyer et al. (2018) at the RAND Corporation and the Council of Economic Advisers (2018) developed top-down approaches that scale firm-level loss rates by national output, an approach we adapt in the predictive resilience layer.

Heavy Tails, Systemic Propagation and Critical Infrastructure

A consistent empirical finding is that cyber losses are heavy-tailed. Maillart and Sornette (2010) documented power-law behaviour in the size of personal-data breaches; Edwards et al. (2016) showed that breach sizes are better described by a lognormal

distribution and that neither their frequency nor their size had trended upward in the way media reports suggested. Biener et al. (2015) concluded that cyber risk is insurable only with difficulty because of correlated, heavy-tailed losses. The systemic dimension is emphasized by studies of propagation: Crosignani et al. (2023) show that the NotPetya attack propagated through supply chains to customers of directly affected firms, and Welburn and Strong (2022) show that input–output linkages amplify direct losses into aggregate output effects. Attacks on physical critical infrastructure, such as the 2015 Ukrainian grid intrusion that cut power to about 225,000 customers (Lee et al., 2016), illustrate how cyber events cross into the real economy. For the financial sector, the IMF has proposed value-at-risk style frameworks for cyber losses (Bouveret, 2018; Kopp et al., 2017).

Market-Based Evidence and Firm-Level Cyber Risk

Corporate-finance research has measured the cost of cyberattacks through asset prices. Kamiya et al. (2021) find significant shareholder losses after successful attacks, especially where personal financial data are compromised, and Florackis et al. (2023) construct a text-based measure of firm cyber-risk exposure that is priced in the cross-section of returns. Jamilov et al. (2021) build a cyber-risk index from earnings calls and show that cyber-risk shocks spill over across firms and countries. These studies establish that cyber risk is economically material, but they rely on listed-firm data that are unavailable for most emerging economies, which is one reason a framework based on public macro series is useful.

Machine Learning and Deep Learning for Cyber-Risk Prediction

Recurrent neural networks, in particular the LSTM architecture (Hochreiter & Schmidhuber, 1997), are widely used for sequence forecasting, and hybrid models that combine a linear time-series component with a neural network for the nonlinear residual have a long history (Zhang, 2003). Deep ensembles provide a practical way to quantify predictive uncertainty (Lakshminarayanan et al., 2017), while tree ensembles such as random forests (Breiman, 2001) and gradient boosting (Friedman, 2001) remain strong baselines for tabular problems. Shapley-value attribution (Lundberg & Lee, 2017; Shapley, 1953) allows the drivers of black-box predictions to be inspected, and isolation forests (Liu et al., 2008) are effective for unsupervised anomaly detection. A recent systematization of cyber-risk quantification notes, however, that most applications of these methods use firm-level or network-telemetry data, and that their value for short macroeconomic series has rarely been tested against proper benchmarks (Woods & Böhme, 2021).

Research Gap and Hypotheses

The literature therefore leaves three gaps that matter for critical infrastructure policy: (i) the link between technical threat-intelligence indicators and macro losses has not been modelled with inference suited to short samples; (ii) deep learning has not been benchmarked rigorously against econometric and tree-based alternatives on such data; and (iii) evidence on vulnerability is rarely combined with severity and forecast models to produce country-level exposure estimates. We test five hypotheses:

- **H1** (threat–loss link): growth in vulnerability disclosures and attack activity is positively associated with growth in reported cyber losses, and disclosures lead losses in time.
- **H2** (stability): the threat–loss relationship has been stable over 2015–2024, with no structural break around the COVID-19 shock of 2020.
- **H3** (forecasting): deep learning models forecast reported losses more accurately than econometric and tree-ensemble benchmarks.
- **H4** (vulnerability): national cyber-capability rises with income and economic scale, and the income effect is strongest in the lower tail of the capability distribution.
- **H5** (severity): incident severity is heavy-tailed, such that tail-risk measures are dominated by a small number of extreme events and by parameter uncertainty.

Data and Methodology

Data

The analysis uses a compiled dataset in which every value is taken from a published source; nothing is simulated. It has four components. (a) **Global threat-intelligence and loss series, 2015–2025**: annual counts of published CVE identifiers from the NIST National Vulnerability Database and Gamblin's annual reviews (Gamblin, 2025; National Institute of Standards and Technology [NIST], n.d.); the number of complaints and the reported losses recorded by the FBI Internet Crime Complaint Center (FBI, 2025);

and the global and US average breach cost and breach lifecycle (days to identify and contain) from the IBM/Ponemon Cost of a Data Breach reports (IBM Security, 2025). The CVE and IC3 series are complete for 2015–2024; the IC3 report for 2025 was not included. (b) **Country resilience features** for 25 economies: the ITU Global Cybersecurity Index (GCI) 2020 score and rank (International Telecommunication Union [ITU], 2021), the GCI 2024 tier where publicly confirmed (ITU, 2024), World Bank income group and GDP for 2023 (World Bank, 2024), and the IBM 2025 breach cost where the country is covered. We define the *vulnerability gap* as 100 minus the GCI score. (c) **Major incidents**: 24 documented macro-relevant cyber events from 2015 to 2025 with host country, sector, attack type, attributed actor, scale and, for 13 events, an estimated economic cost with its basis (firm-reported, regulatory, insurer, government or economic-model estimate). One event the July 2024 Crowd Strike faulty update is non-malicious and serves as a control case. (d) **Regional breach costs** for 2025.

Three features of the data shape the methods. First, the annual series are short (ten complete observations), so inference relies on heteroskedasticity- and autocorrelation-consistent (HAC) standard errors, bootstrap and Bayesian methods rather than asymptotic approximations alone. Second, IC3 losses are US victim-reported losses only and understate the true total; we interpret them as a consistently measured *index* of loss intensity rather than as a global total. Third, incident costs mix estimation methods and some global events are assigned to an origin country, so they are used for severity-distribution modelling and not summed naively. Table 1 reports definitions and descriptive statistics.

Table 1: **Variable Definitions, Sources and Descriptive Statistics.**

Variable	Definition (source)	N	Mean	SD	Min	Max
CVE	CVE identifiers published per year (NVD; Gamblin)	10	19,390	10,103	6,447	40,009
COMP	IC3 complaints per year (FBI IC3)	10	588,768	266,535	288,012	880,418
LOSS	IC3 reported losses, USD bn (FBI IC3)	10	6.07	5.38	1.07	16.60
LPC	Loss per complaint, USD (derived)	10	8,828	5,054	3,715	19,313
IBM_G	Global avg. breach cost, USD m (IBM/Ponemon)	11	4.13	0.37	3.62	4.88
IBM_US	US avg. breach cost, USD m (IBM/Ponemon)	6	9.37	0.52	8.64	10.22
LIFE	Breach lifecycle, days (IBM/Ponemon)	6	270.0	17.2	241	287
GCI	ITU GCI 2020 score, 0–100 (ITU)	25	91.98	12.50	54.41	100.00
GAP	Vulnerability gap = 100 − GCI	25	8.02	12.50	0.00	45.59
GDP	GDP 2023, current USD bn (World Bank)	25	3,242	6,125	41	27,361
COST	Incident economic cost, USD bn (various)	13	2.10	2.95	0.0044	10.00

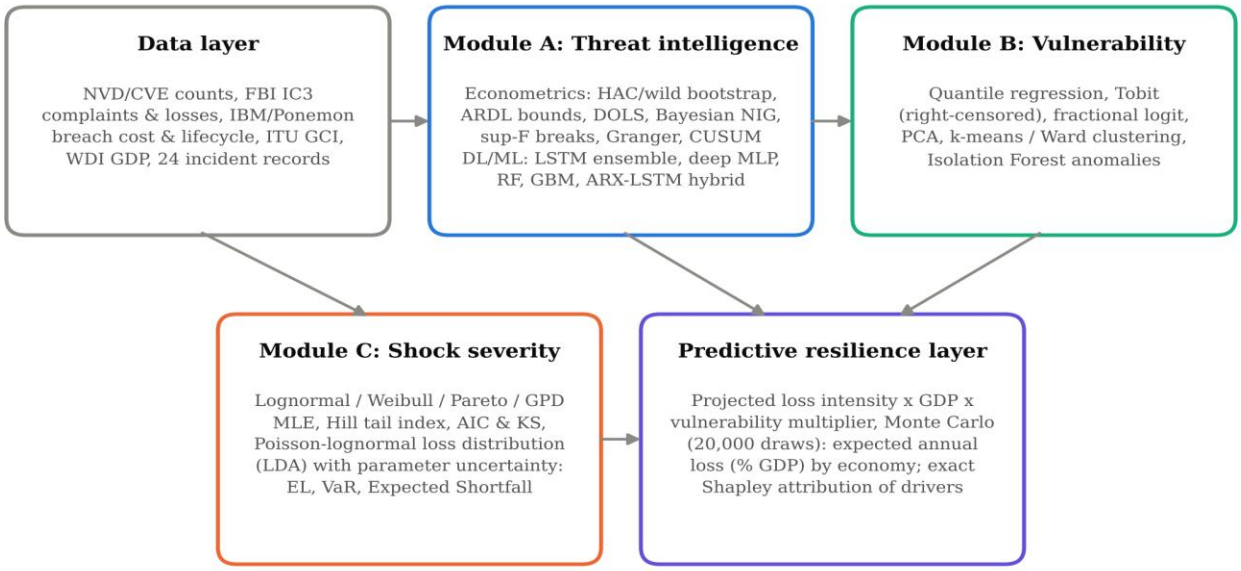
Note. Annual series cover 2015–2024 (CVE, COMP, LOSS, LPC), 2015–2025 (IBM_G) and 2020–2025 (IBM_US, LIFE). Country variables cover 25 economies; COST covers the 13 incidents with a published cost estimate (12 malicious plus the Crowd Strike control).

The Hybrid Framework

Figure 1 summarizes the architecture. The data layer feeds three analytical modules whose outputs are combined in a predictive resilience layer. The term *hybrid* has two meanings here. Within Module A, linear econometric models capture the persistent, interpretable part of the threat–loss relationship, while neural networks and tree ensembles are allowed to pick up nonlinear and lagged effects; the two are combined both additively (an ARX model with an LSTM fitted to its residuals, as in Zhang, 2003) and by equal-weight stacking. Across modules, statistical models of vulnerability and severity are combined with the machine-learning forecast to produce country-level exposure measures.

Figure 1

Architecture of the hybrid deep learning and machine learning framework. Module A (threat intelligence) links vulnerability disclosures and attack activity to losses; Module B (vulnerability assessment) models national capability; Module C (shock severity) characterizes incident tail risk; the predictive resilience layer combines them by Monte Carlo simulation.



Module A — Econometric Threat–Loss Models

Let L_t denote IC3 reported losses, V_t the number of CVEs and C_t the number of complaints in year t, with lower-case letters denoting natural logarithms. The baseline static elasticity models are

$$l_t = \alpha + \beta_V v_t + \beta_C c_t + \gamma t + \varepsilon_t$$

estimated in four nested versions (M1: v only; M2: c only; M3: v and c; M4: v, c and a linear trend). Standard errors are Newey–West HAC with one lag (Newey & West, 1987). Because ten observations are too few for reliable asymptotics, we also report restricted wild-bootstrap p-values with Rademacher weights and 4,999 replications (Davidson & Flachaire, 2008), and the usual diagnostics: Durbin–Watson, Jarque–Bera normality (Jarque & Bera, 1987), Breusch–Pagan heteroskedasticity (Breusch & Pagan, 1979), Ramsey RESET (Ramsey, 1969) and variance inflation factors.

Integration properties are examined with augmented Dickey–Fuller tests using MacKinnon finite-sample critical values (MacKinnon, 2010), complemented by a sieve bootstrap under the unit-root null. Long-run relations are examined three ways. The Engle–Granger residual test (Engle & Granger, 1987) checks for cointegration in M1. An autoregressive distributed lag model, ARDL(1,1),

$$l_t = a + \phi\, l_{t-1} + \theta_0\, v_t + \theta_1\, v_{t-1} + u_t, \quad \lambda_{LR} = (\theta_0 + \theta_1) / (1 - \phi)$$

delivers the long-run elasticity λ_{LR}, whose standard error is obtained by the delta method, and the corresponding unrestricted error-correction form is used for the Pesaran–Shin–Smith bounds F-test (Pesaran et al., 2001), which is valid for any mix of I(0) and I(1) regressors; we note that Narayan's (2005) small-sample critical values are stricter than the asymptotic ones. Finally, the dynamic OLS estimator of Stock and Watson (1993) augments the levels regression with leads and lags of Δv_t to correct for endogeneity and serial correlation in the cointegrating regression.

Parameter stability (H2) is tested with the Quandt–Andrews sup-F statistic for an unknown break in the level and slope of the log-loss trend (Andrews, 1993), using 30% trimming and a residual bootstrap for the p-value, a Chow test at 2020, and the CUSUM of recursive residuals (Brown et al., 1975). Predictive precedence (H1) is tested with Granger causality at lag one (Granger, 1969),

again with bootstrap p-values under the null. To quantify uncertainty about both parameters and model choice, M1–M3 are re-estimated with a conjugate Normal–Inverse-Gamma prior (Koop, 2003),

$$\beta \mid \sigma^2 \sim N(m_0, \sigma^2 V_0), \quad \sigma^{-2} \sim \text{Gamma}(a_0, b_0)$$

with centred regressors, m_0 equal to zero for slopes, $V_0 = 10I$, $a_0 = 2$ and $b_0 = 0.1$. Posterior moments are computed from 20,000 draws and posterior model probabilities from the closed-form marginal likelihoods.

Module A — Deep Learning and Machine Learning Forecasting

The forecasting task is one-step-ahead prediction of log loss growth Δl_t using the growth rates of the three threat series over the previous $W = 2$ years, $x_t = (\Delta v_{t-k}, \Delta c_{t-k}, \Delta l_{t-k})_{k=1,2}$. The LSTM (Hochreiter & Schmidhuber, 1997) processes this sequence through gated memory cells:

$$i_t, f_t, o_t = \sigma(W_x x_t + W_h h_{t-1} + b), \quad g_t = \tanh(\cdot)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, \quad h_t = o_t \odot \tanh(c_t), \quad \hat{y} = w_y' h_T + b_y$$

Because no deep learning library could be relied upon in the computing environment, the LSTM was implemented from first principles in NumPy with full back-propagation through time, Adam optimization (Kingma & Ba, 2015), gradient clipping, L2 regularization and a forget-gate bias of one. To stabilize estimation on a small sample and to quantify epistemic uncertainty, we train a *deep ensemble* (Lakshminarayanan et al., 2017) of 15 networks (30 for the final projection) with different random initializations and average their predictions. The other learners are a deep multilayer perceptron with three hidden layers (16–8–4 units, tanh activation, L2 penalty, 15-member ensemble), a random forest with 500 trees (Breiman, 2001) and a gradient-boosting machine with 300 depth-2 trees, learning rate 0.05 and subsampling of 0.8 (Friedman, 2001). Hyperparameters were fixed a priori (Appendix Table A1) rather than tuned on the test years.

Benchmarks are a random walk, a random walk with drift, a log-linear trend, Holt's linear exponential smoothing (Holt, 1957/2004) and an ARX econometric model, $\Delta l_t = a + b\,\Delta v_{t-1} + c\,\Delta c_{t-1} + e_t$. The *hybrid ARX–LSTM* adds a shrunken LSTM forecast of the ARX residual to the ARX forecast (Zhang, 2003), and the *stacked ensemble* averages the ARX, LSTM and gradient-boosting forecasts with equal weights fixed in advance. All models are evaluated by an expanding-window, rolling-origin design: for each test year 2021–2024, models are re-estimated using only information available up to the previous year. Accuracy is measured by the RMSE and MAE of log forecasts, the MAPE of levels and Theil's U relative to the random walk, and significance of accuracy differences is assessed with the Diebold–Mariano test (Diebold & Mariano, 1995) with the Harvey–Leybourne–Newbold small-sample correction (Harvey et al., 1997).

Interpretability is provided by exact Shapley values (Lundberg & Lee, 2017; Shapley, 1953). With six features, all 2^6 coalitions can be enumerated, so the attribution ϕ_j of feature j is computed exactly rather than approximated:

$$\phi_j = \Sigma_{S \subseteq F \setminus \{j\}} \, |S|!(|F|-|S|-1)!/|F|! \cdot [f(x_{S \cup \{j\}}) - f(x_S)]$$

with absent features set to their training means. An isolation forest (Liu et al., 2008) with 1,000 trees is applied to the annual growth vectors to flag anomalous threat regimes.

Module B — Vulnerability Assessment

National cyber-capability is modelled across the 25 economies as $GCI_i = \alpha + \beta \ln GDP_i + \delta HIGH_i + u_i$, where HIGH is a high-income indicator. Because the effect of income may differ between leaders and laggards (H4), we estimate linear quantile regressions (Koenker & Bassett, 1978) at $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ with pairs-bootstrap confidence intervals (999 replications). Because the GCI is bounded and the United States sits at the upper limit of 100, we also estimate a right-censored Tobit model (Tobin, 1958) and a Papke–Wooldridge fractional logit (Papke & Wooldridge, 1996) for GCI/100 with sandwich standard errors and average marginal effects. Resilience typologies are identified by principal component analysis of standardized GCI, log GDP, income group and log gap-weighted GDP exposure, followed by k-means clustering with the number of clusters chosen by the silhouette criterion (Rousseeuw, 1987) and checked against Ward hierarchical clustering (Ward, 1963).

Module C — Shock Severity and Loss Distribution

For the 12 malicious incidents with published cost estimates, we fit lognormal, Weibull and Pareto type I severity distributions by maximum likelihood and compare them by AIC and Kolmogorov–Smirnov tests; log normality is also checked with the Anderson–

Darling test on log costs. Tail heaviness is examined with the Hill estimator (Hill, 1975) across the number of upper order statistics k and with a generalised Pareto distribution fitted to exceedances above the median (Embrechts et al., 1997). Incident frequency is modelled as Poisson with intensity λ equal to the number of documented malicious incidents per year; the dispersion index of annual counts checks the Poisson assumption. The aggregate annual loss from catalogued macro-relevant incidents follows the loss distribution approach (LDA) of operational-risk measurement (McNeil et al., 2015):

$$S = \Sigma_{k=1}^{N} X_k, \quad N \sim \text{Poisson}(\lambda), \quad \ln X_k \sim N(\mu, \sigma^2)$$

simulated with 200,000 draws. We report the expected loss, Value-at-Risk $VaR_q = \inf\{s : P(S \le s) \ge q\}$ and Expected Shortfall $ES_q = E[S \mid S \ge VaR_q]$ at q = 0.95 and 0.99, both with plug-in parameters and with parameter uncertainty propagated by drawing (μ, σ, λ) from their sampling distributions.

Predictive Resilience Layer

The final layer converts the Module A projection into country exposure. The projected loss intensity ρ_t is the hybrid-ensemble forecast of reported losses divided by US GDP, grown at 4.5% a year in nominal terms from 2023. Expected annual loss for economy i in 2028 is

$$EAL_i = GDP_i \cdot \rho_{2028} \cdot (1 + \kappa \cdot GAP_i / 10)$$

where κ is the proportional loss uplift per 10 points of vulnerability gap. Because κ is not identified by the data, it is treated as a structural uncertainty and drawn from a uniform distribution on [0.05, 0.25], while ρ_{2028} is drawn from a lognormal distribution matched to the ensemble's 80% prediction interval; 20,000 Monte Carlo draws give the median and 80% interval of EAL_i and of EAL_i/GDP_i. Because IC3 losses are reported losses only, EAL should be read as a lower-bound, comparable index of exposure rather than a full estimate of economic cost.

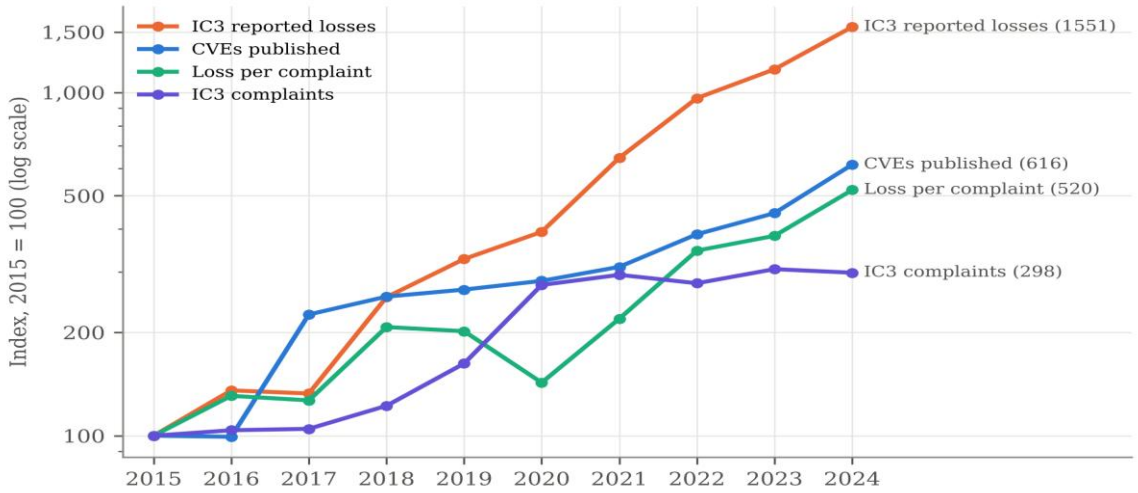
All analyses were run in Python 3 (NumPy, SciPy, scikit-learn, Matplotlib) with fixed random seeds; the scripts reproduce every table and figure from the source workbook.

Results

Descriptive Evidence

Between 2015 and 2024, IC3 reported losses rose from USD 1.07 billion to USD 16.6 billion, a compound annual growth rate of 35.6%. Over the same period, CVE disclosures grew by 22.4% a year, complaints by 12.9% and loss per complaint by 20.1% (Figure 2). Losses therefore rose much faster than the number of victims: the average loss per complaint increased from about USD 3,700 to USD 19,300, consistent with a shift towards higher-value fraud and extortion. Complaints jumped by 69% in 2020 and have since plateaued at around 800,000–880,000, whereas CVE disclosures accelerated, reaching 40,009 in 2024. The pairwise correlations of the log series are high (0.92 between log losses and log CVEs; 0.93 with log complaints), which, given common trends, motivates the integration and cointegration tests below rather than naive regression.

Figure 2: Threat-intelligence and loss indicators, 2015–2024, indexed to 2015 = 100 (log scale). Sources: NVD/CVE, FBI IC3.



The IBM/Ponemon data show a different, firm-level picture. The global average breach cost rose from USD 3.79 million in 2015 to a peak of USD 4.88 million in 2024 before falling to USD 4.44 million in 2025, while the US average reached USD 10.22 million in 2025, 2.30 times the global mean. The breach lifecycle fell from 287 days in 2021 to 241 days in 2025. Across 2020–2025 the lifecycle is negatively correlated with the US breach cost ($r = -0.80$, $p = 0.056$) and only weakly with the global cost ($r = -0.53$, $p = 0.28$). With six observations and opposing trends, this is not evidence that faster detection raises costs; it indicates that improvements in detection and containment have not offset the rising unit cost of incidents.

Integration, Elasticities and Long-Run Relations

Table 2: **Unit-Root and Cointegration Tests.**

Series / test	ADF τ (level, c)	5% CV	Bootstrap p	ADF τ (level, c+t)	5% CV	ADF τ (Δ, c)	5% CV
ln LOSS	-0.10	-3.23	0.576	-3.40	-4.01	-3.99**	-3.29
ln CVE	-0.91	-3.23	0.321	-2.13	-4.01	-3.38**	-3.29
ln COMP	-0.79	-3.23	0.449	-1.18	-4.01	-1.75	-3.29
Engle–Granger (M1 residual)	-2.16	-4.10	–	–	–	–	–
ARDL bounds F (k = 1)	1.22	I(0) 4.94 / I(1) 5.73	–	–	–	–	–

Note. ADF regressions with zero augmentation lags given T = 10. Critical values from MacKinnon (2010) finite-sample response surfaces; bootstrap p-values from a sieve bootstrap under the unit-root null (4,999 replications). Bounds critical values are Pesaran et al. (2001) Case III asymptotic values. ** denotes rejection of the unit-root null at 5%.

Table 2 shows that none of the log levels rejects a unit root, with or without trend, and bootstrap p-values between 0.32 and 0.58. First differences of log losses and log CVEs reject the null at 5% (τ = -3.99 and -3.38), so both behave as I(1); the differenced complaints series does not reject, which reflects the 2020 level shift and the low power of a test on nine observations rather than convincing evidence of I(2) behaviour. Neither the Engle–Granger statistic (-2.16 against a critical value of -4.10) nor the ARDL bounds F-statistic (1.22, below even the I(0) bound) supports a stable levels relationship. The static regressions reported below should therefore be read as reduced-form co-movement, not as structural long-run elasticities, and the forecasting models are specified in growth rates.

Table 3: Threat–Loss Regressions: Static, Dynamic and Long-Run Estimates (Dependent Variable: ln $LOSS_t$).

	M1	M2	M3	M4	ARDL(1,1)	DOLS
ln CVE_t	1.523***		0.762*	-0.419**	-0.167	2.693**
	(0.193)		(0.371)	(0.150)	(0.167)	(0.510)
	[0.049]		[0.068]	[0.178]	[0.546]	[0.484]
ln $COMP_t$		1.817***	1.073**	-0.144		
		(0.267)	(0.434)	(0.187)		
		[0.000]	[0.170]	[0.914]		
Trend				0.418***		
				(0.046)		

	M1	M2	M3	M4	ARDL(1,1)	DOLS
				[0.022]		
ln $LOSS_{t-1}$					0.745***	
					(0.123)	
					[0.044]	
ln CVE_{t-1}					0.627**	
					(0.208)	
					[0.436]	
Constant	-13.417***	-22.535***	-20.152***	5.509	-3.776*	-24.978**
	(1.874)	(3.423)	(2.820)	(2.866)	(1.846)	(4.912)
Long-run elasticity					1.804 (0.472)	2.693
R^2 / adj. R^2	0.841 / 0.821	0.869 / 0.853	0.934 / 0.915	0.991 / 0.986	0.979 / 0.966	0.971 / 0.941
T	10	10	10	10	9	7
Durbin–Watson	1.58	1.22	1.89	2.34	2.18	1.57
Jarque–Bera p	0.440	0.773	0.621	0.817	0.669	0.800
Breusch–Pagan p	0.595	0.136	0.798	0.092	0.399	0.220
RESET p	0.117	0.844	0.079	0.612	0.667	0.016
Max VIF	1.00	1.00	3.25	29.82	6.57	3.82
AIC	-15.96	-17.94	-22.78	-40.47	-29.64	-21.46

Note. Newey–West HAC standard errors (one lag) in parentheses; restricted wild-bootstrap p-values (Rademacher, 4,999 replications) in brackets. ***, **, * denote significance at 1%, 5% and 10% using HAC t-tests. DOLS includes one lead and the contemporaneous difference of ln CVE (coefficients not shown). Long-run ARDL elasticity with delta-method standard error.

Table 3 reports the elasticity estimates. In isolation, a 1% rise in CVE disclosures is associated with a 1.52% rise in reported losses (M1, HAC SE 0.19, wild-bootstrap p = 0.049), and a 1% rise in complaints with a 1.82% rise (M2, wild p = 0.0002). Both elasticities exceed one, so losses grow more than proportionally with the threat surface and with the volume of attacks. When both are included (M3), the elasticities fall to 0.76 and 1.07, the variance inflation factor is a moderate 3.2, and the wild-bootstrap p-values (0.068 and 0.170) are noticeably larger than the HAC p-values, which illustrates why bootstrap inference matters at T = 10. Adding a linear trend (M4) absorbs almost all of the variation: the trend coefficient implies growth of 41.8 log points a year and the threat variables lose significance under the bootstrap, while the VIF of 29.8 shows severe collinearity. The diagnostics are otherwise clean: residual normality, homoskedasticity and functional form are not rejected at 5% for M1–M4.

The dynamic specifications give the same message from a different angle. The ARDL(1,1) model shows strong persistence (ϕ = 0.745, wild p = 0.044) and a significant lagged CVE effect (θ_1 = 0.627, HAC p = 0.030), while the contemporaneous CVE coefficient is small and insignificant; the implied long-run elasticity is 1.80 (SE 0.47) and the error-correction speed is -0.255. The DOLS estimate of 2.69 is significant under HAC inference but not under the wild bootstrap (p = 0.484). Taken together, the estimates

support a positive, more-than-proportional and lagged association between vulnerability disclosures and losses, but the data are too short to pin down a unique long-run elasticity.

Bayesian Model Comparison, Stability and Causality

Table 4: Bayesian Normal–Inverse-Gamma Estimates and Posterior Model Probabilities.

Model	Coefficient	Posterior mean	Posterior SD	95% credible interval	P(β > 0)	P(β > 1)	P(model | data)
M1	ln CVE	1.474	0.213	[1.057, 1.897]	1.000	0.985	0.028
M2	ln COMP	1.741	0.241	[1.263, 2.218]	1.000	0.998	0.061
M3	ln CVE	0.786	0.261	[0.270, 1.305]	0.997	0.198	0.912
	ln COMP	1.007	0.305	[0.395, 1.616]	0.999	0.510	

Note. Regressors centred; prior $\beta \sim N(0, 10\sigma^2)$ for slopes, $a_0 = 2$, $b_0 = 0.1$; 20,000 posterior draws. Posterior model probabilities assume equal prior model probabilities.

The Bayesian results (Table 4) sharpen inference. The posterior probability that each elasticity is positive is above 0.99 in every model. The data strongly favour the two-variable model M3 (posterior probability 0.91), in which the CVE elasticity has a posterior mean of 0.79 (95% credible interval 0.27 to 1.30) and the complaints elasticity 1.01 (0.39 to 1.62). Vulnerability exposure and attack volume therefore carry separate information about losses, which supports using both in the forecasting models.

Table 5: Structural Stability and Granger-Causality Tests.

Test	Statistic	p-value	Bootstrap p	Decision (5%)
Quandt–Andrews sup-F, break in trend (estimated 2018)	3.44	–	0.308	No break
Chow test, break at 2020	0.156	0.859	–	No break
CUSUM of recursive residuals, M1	max |W|/bound = 2.49	–	–	Instability in static M1
Granger: ΔCVE → LOSS	6.22	0.047	0.073	Reject (asymptotic); 10% (bootstrap)
Granger: COMP → LOSS	1.62	0.250	0.324	Do not reject
Granger: LOSS → CVE	2.37	0.175	0.353	Do not reject
Granger: CVE → COMP	2.96	0.136	0.219	Do not reject

Note. sup-F with 30% trimming and residual-bootstrap p-value (4,999 replications). CUSUM bound at 5% (Brown et al., 1975); a ratio above one indicates a crossing. Granger tests use one lag of levels in logs; bootstrap p-values are generated under the restricted AR(1) null.

The stability tests (Table 5) show no break in the growth path of losses. The sup-F statistic peaks in 2018 but has a bootstrap p-value of 0.31, and the Chow test at 2020 is far from significant (p = 0.86): losses grew at roughly 31.9 log points a year before and after the pandemic, so the 2020 jump in complaints did not change the trajectory of losses. The CUSUM path of the static CVE-only model does cross its 5% bound, with recursive residuals that are persistently positive, which means a fixed elasticity on CVEs alone systematically under-predicts recent losses. This is consistent with the lack of cointegration and with the importance of complaint volume in M3. The Granger tests support the lead–lag part of H1: lagged CVE disclosures help predict losses (F = 6.22, p = 0.047;

bootstrap p = 0.073), whereas there is no evidence of reverse causality from losses to disclosures. H2 is therefore supported for the trend in losses but not for a static single-driver elasticity.

Forecasting Performance of the Hybrid DL/ML Models

Table 6: Rolling-Origin One-Step-Ahead Forecast Accuracy for IC3 Reported Losses, 2021–2024.

Model	RMSE (log)	MAE (log)	MAPE (%)	Bias (%)	Theil's U	DM vs RW (p)
Random walk	0.362	0.344	28.6	-28.6	1.000	–
RW + drift	0.137	0.117	11.2	-3.1	0.379	-2.89 (0.063)
Log-linear trend	0.130	0.108	10.1	-6.3	0.359	-2.96 (0.059)
Holt (damped-free)	0.117	0.091	8.5	-8.5	0.323	-2.89 (0.063)
ARX (econometric)	0.241	0.192	16.9	-11.5	0.664	-4.27 (0.024)
LSTM (deep ensemble)	0.191	0.166	15.4	-4.9	0.526	-3.35 (0.044)
Deep MLP	0.188	0.107	9.2	-7.6	0.518	-3.70 (0.034)
Random forest	0.121	0.091	8.9	0.2	0.335	-2.71 (0.073)
Gradient boosting	0.113	0.080	7.6	-3.4	0.312	-3.07 (0.055)
Hybrid ARX-LSTM	0.238	0.186	16.2	-13.6	0.656	-4.94 (0.016)
Stacked ensemble (ARX+LSTM+GB)	0.175	0.142	13.0	-6.8	0.483	-3.45 (0.041)

Note. Models re-estimated at each origin using data up to t − 1 only. Theil's U is the RMSE ratio to the random walk. DM: Diebold–Mariano statistic on squared log errors with the Harvey–Leybourne–Newbold correction; negative values favour the model over the random walk. With four forecast origins, DM tests have low power.

Figure 3: Out-of-sample forecast error by model family, 2021–2024 (rolling origin). Lower is better.

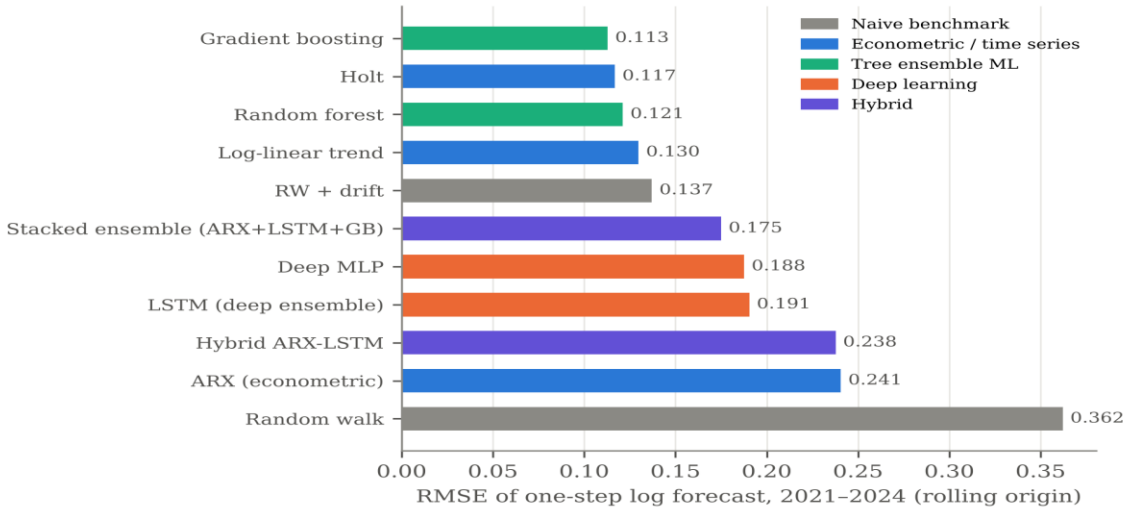


Table 6 and Figure 3 report the horse race. The random walk is clearly the worst model (log RMSE 0.362), because losses trended up in every test year. All other models improve on it, with Theil's U between 0.31 and 0.66. The best performer is the **gradient-boosting machine** (RMSE 0.113, MAPE 7.6%), followed closely by Holt smoothing (0.117) and the random forest (0.121). The deep learning models are competitive but not superior: the LSTM deep ensemble has an RMSE of 0.191 with a small bias of -4.9%, and the deep MLP has an MAE (0.107) close to Holt and the random forest but one large miss in 2022. The ARX–LSTM hybrid improves only marginally on its ARX base (0.238 against 0.241), and the stacked ensemble (0.175) sits between its components, as equal-weight combinations usually do. Every model beats the random walk at the 10% level on the Diebold–Mariano test, and the ARX, LSTM, deep MLP, hybrid and stacked models do so at 5%, but with only four origins none can be distinguished from the others statistically.

H3 is therefore **not supported** in its strong form. With ten annual observations, the inductive bias of shallow tree ensembles and exponential smoothing is better matched to the data than the larger hypothesis space of recurrent networks, even when the latter are regularized and ensembled. The value of the LSTM in the framework lies instead in its low bias, in the uncertainty information provided by ensemble dispersion and in its complementary pattern of errors, which is why it is retained in the projection ensemble.

Table 7: Projected IC3 Reported Losses (USD Billion), 2025–2028.

Year	CVE (projected)	Complaints (projected)	ARX	LSTM ensemble	Hybrid ensemble	80% interval	95% interval
2025	46,215	887,611	23.1	24.5	22.9	20.3–25.8	19.0–27.5
2026	55,501	905,763	29.6	31.4	29.9	25.3–35.3	23.0–38.7
2027	66,655	924,286	38.7	41.3	39.3	31.9–48.3	28.6–53.8
2028	80,049	943,187	50.6	56.8	52.0	41.0–65.9	36.3–75.2

Note. Recursive projections trained on 2015–2024. Exogenous CVE counts follow a log-linear trend fitted to 2015–2024; complaints follow the 2020–2024 trend. The hybrid ensemble averages ARX, LSTM (30-member deep ensemble) and gradient-boosting growth forecasts. Intervals combine ARX residual variance with LSTM ensemble dispersion (10,000 simulated paths).

Figure 4: **Actual IC3 reported losses (2015–2024), rolling one-step forecasts of the gradient-boosting and LSTM models (2021–2024), and the hybrid ensemble projection with 80% and 95% prediction intervals (2025–2028).**

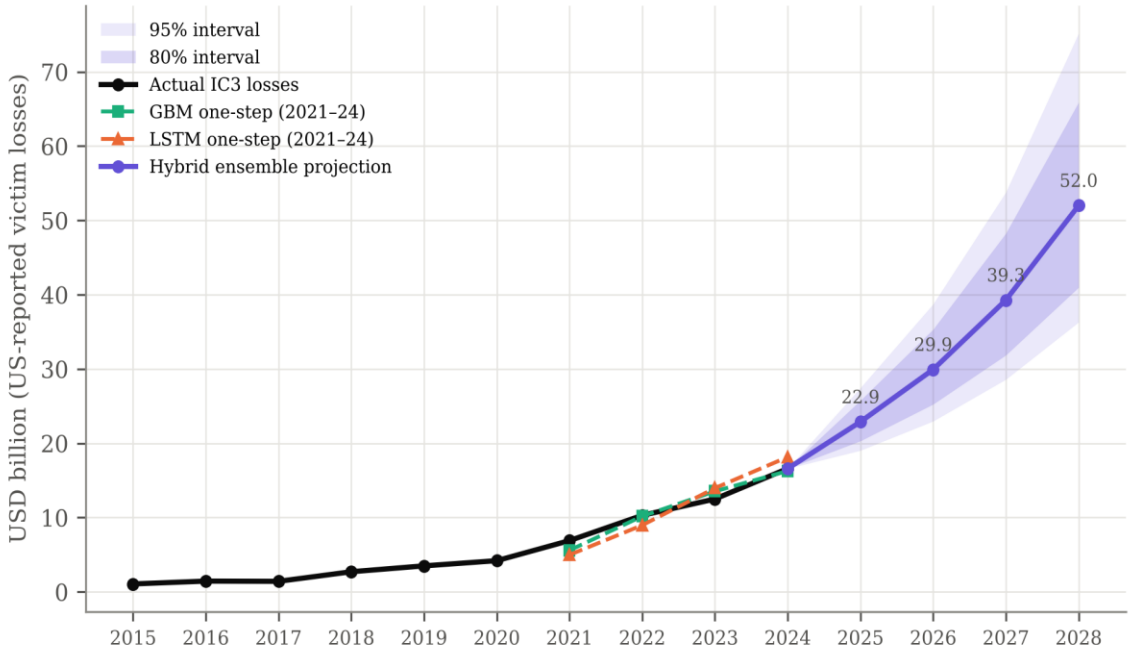
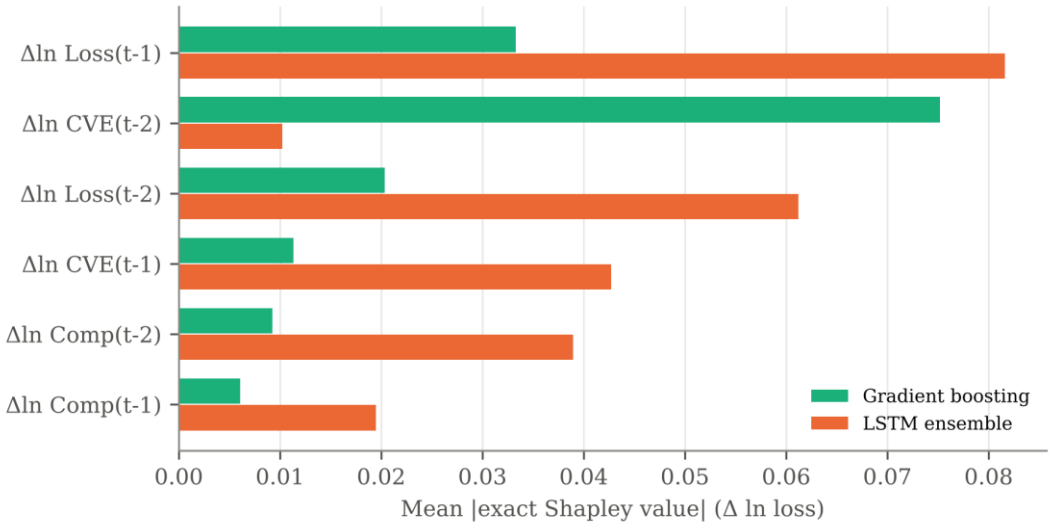


Table 7 and Figure 4 give the recursive projection. With CVE disclosures projected to reach about 80,000 a year by 2028 and complaint volume growing slowly, the hybrid ensemble projects US-reported losses of USD 22.9 billion in 2025 rising to USD 52.0 billion in 2028 (80% interval 41.0–65.9), an implied growth rate of about 33.0% a year — slightly below the historical 35.6%. The LSTM path is the most pessimistic (USD 56.8 billion in 2028) and the ARX path the most conservative (USD 50.6 billion). These projections are conditional on the continuation of current threat trends and on IC3 reporting practices remaining unchanged.

Explainability and Anomaly Detection

Figure 5: Global feature importance from exact Shapley values for the gradient-boosting and LSTM-ensemble forecasters (training sample).



The exact Shapley attributions (Figure 5) show how the two best non-linear learners use the threat-intelligence inputs. For gradient boosting, the dominant driver is CVE growth two years earlier (mean $|\phi|$ = 0.075), more than twice the next feature, lagged loss growth (0.033). The LSTM spreads importance more evenly, with the largest weight on the loss history (0.082 and 0.061 for the first and second lags) and substantial weight on the most recent CVE growth (0.043) and on lagged complaint growth (0.039). Both models agree with the Granger and ARDL results that vulnerability disclosures lead losses by one to two years — the time it takes for newly disclosed weaknesses to be weaponized and exploited at scale. The isolation forest flags two anomalous threat regimes: 2017, when CVE publication more than doubled (+127%) reflecting the expansion of CVE Numbering Authorities, and 2020, when complaints rose by 69% during the pandemic. Neither anomaly produced a break in the loss trend, which is consistent with the stability tests.

National Vulnerability and Resilience Typologies

Table 8: Determinants of National Cyber-Capability (Dependent Variable: ITU GCI 2020 Score).

Estimator	ln GDP	High income	Constant	Fit / notes
OLS (HC3 SE)	2.502 (1.446)	14.308*** (4.851)	65.74 (12.29)	R^2 = 0.469
Tobit, right-censored at 100	3.133*** (1.191)	14.761*** (3.695)	61.46 (8.43)	σ = 8.96; log L = -87.02
Fractional logit (AME, points)	3.170	15.487	–	Coef. 0.482 / 2.355; SE 0.138 / 0.397
Quantile regression, τ = 0.10	0.369 [-0.84, 9.79]	31.578 [10.42, 43.75]	62.73	Bootstrap 95% CI

Estimator	ln GDP	High income	Constant	Fit / notes
Quantile regression, τ = 0.25	0.487 [-0.46, 6.98]	30.352 [3.04, 42.22]	63.40	Bootstrap 95% CI
Quantile regression, τ = 0.50	0.717 [-0.44, 5.78]	15.765 [1.09, 32.25]	76.91	Bootstrap 95% CI
Quantile regression, τ = 0.75	0.370 [-0.44, 1.81]	2.463 [0.42, 18.14]	93.76	Bootstrap 95% CI
Quantile regression, τ = 0.90	0.080 [-0.86, 1.53]	2.347 [0.34, 7.31]	96.84	Bootstrap 95% CI

Note. N = 25. Standard errors in parentheses; 95% pairs-bootstrap confidence intervals (999 replications) in brackets. *** p < 0.01 (asymptotic). AME: average marginal effect on the 0–100 scale. The United States (GCI = 100) is treated as censored in the Tobit model.

Figure 6: **Quantile-regression process for the determinants of the GCI score, with 95% bootstrap bands and the OLS estimate for comparison.**

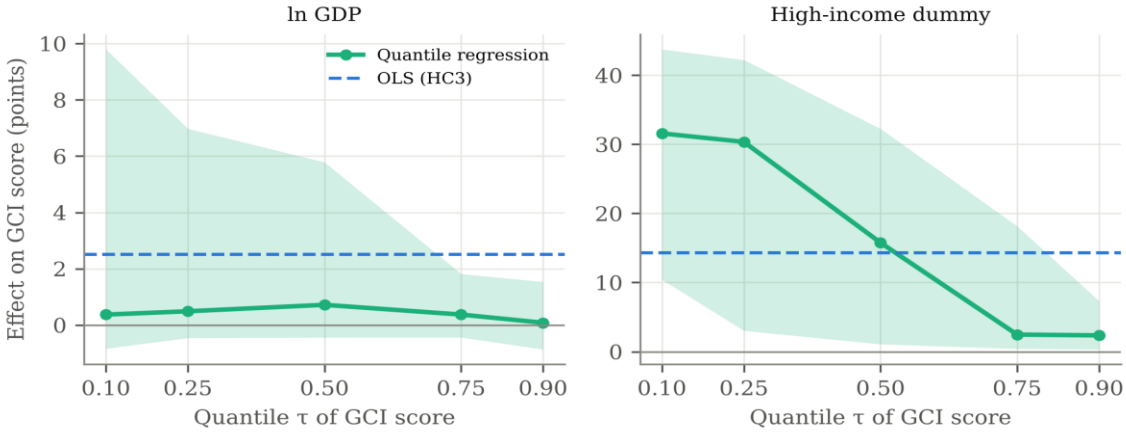
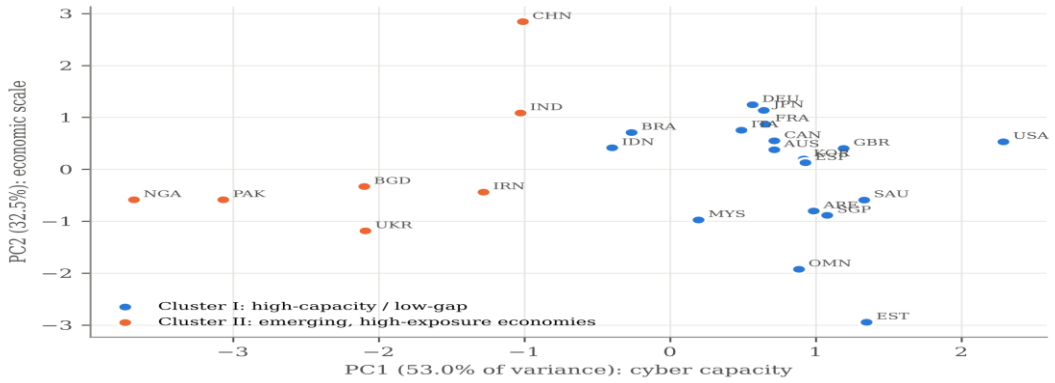


Table 8 and Figure 6 address H4. In the conditional-mean models, high-income status raises the GCI score by about 14–15 points (OLS 14.3; Tobit 14.8; fractional-logit AME 15.5), and each log point of GDP adds around 2.5–3.2 points, an effect that becomes significant once the censoring of the United States is accounted for in the Tobit model. The quantile process shows the income effect is highly heterogeneous: it is 31.6 points at τ = 0.10 and 30.4 at τ = 0.25, but only 2.5 at τ = 0.75 and 2.3 at τ = 0.90, with confidence bands that exclude zero throughout. Economic scale, by contrast, has no significant effect at any quantile. H4 is supported for income, which matters most among the laggards, but not for scale: large economies such as China, India and Brazil are not automatically better protected.

Figure 7: **Principal-component map of 25 economies on cyber-capacity and economic-scale dimensions, colored by k-means resilience cluster.**



The principal components (Figure 7) separate a *cyber-capacity* dimension (PC1, 53.0% of variance, loading on GCI and income) from an *economic-scale* dimension (PC2, 32.5%, loading on GDP and gap-weighted exposure). The silhouette criterion selects two clusters (silhouette 0.422), and the k-means partition agrees substantially with Ward hierarchical clustering (adjusted Rand index 0.68). Cluster I contains 18 high-capacity, mostly high-income economies with vulnerability gaps of about 5 GCI points or less. Cluster II contains seven emerging economies — Nigeria, Pakistan, Ukraine, Bangladesh, Iran, China and India — that combine lower income with either large vulnerability gaps or very large gap-weighted GDP exposure. China's position is driven by scale: with a gap of 7.5 points, its gap-weighted exposure (USD 1.33 trillion) is half of the 25-economy total.

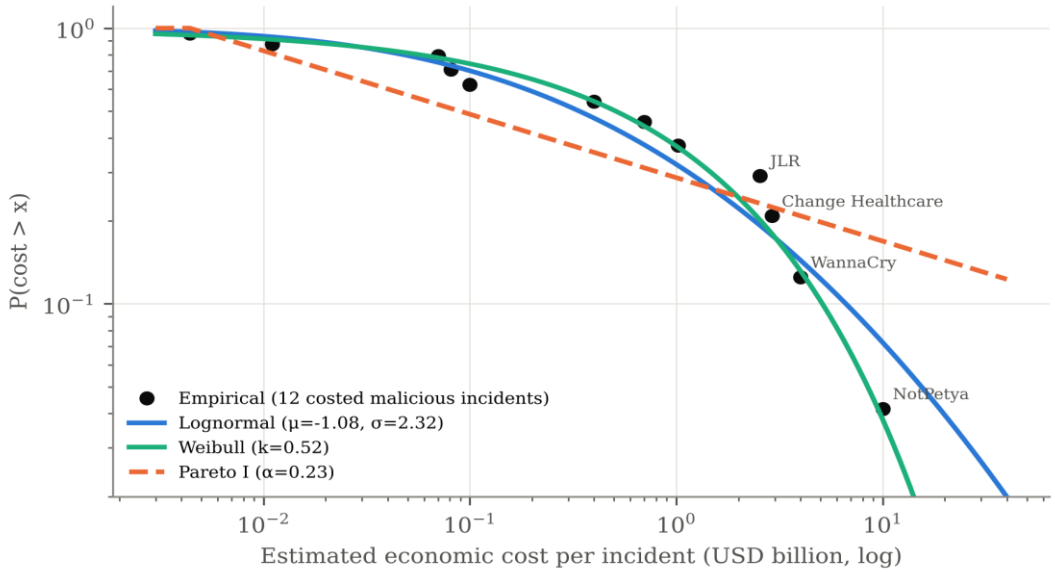
Incident Severity, Tail Risk and the Loss Distribution

Table 9: Severity-Distribution Fits and Loss-Distribution Risk Measures for Documented Malicious Incidents.

Component	Estimate	Fit / diagnostic	Value
Lognormal (μ, σ)	-1.081, 2.321	AIC; KS p	32.31; 0.944
Weibull (shape k, scale)	0.521, 1.029	AIC; KS p	31.59; 0.873
Pareto I (α, x_m)	0.230, 0.0044	AIC; KS p	35.31; 0.175
GPD above median (ξ, β)	0.083, 2.735	Threshold u (USD bn)	0.55
Hill tail index $\hat{\alpha}(k)$	1.54 (k = 3) to 0.24 (k = 10)	Anderson–Darling, ln cost	0.261 (5% CV 0.698)
Frequency λ (per year)	2.091	Dispersion index	0.426
Plug-in LDA: expected loss	10.60	With parameter uncertainty	29.82
Plug-in LDA: VaR 95%	36.75	With parameter uncertainty	47.33
Plug-in LDA: VaR 99%	146.94	With parameter uncertainty	281.45
Plug-in LDA: ES 95%	140.39	With parameter uncertainty	520.55
Plug-in LDA: ES 99%	429.66	With parameter uncertainty	2,176.55
P(no costed incident in a year)	–	Simulated	0.135

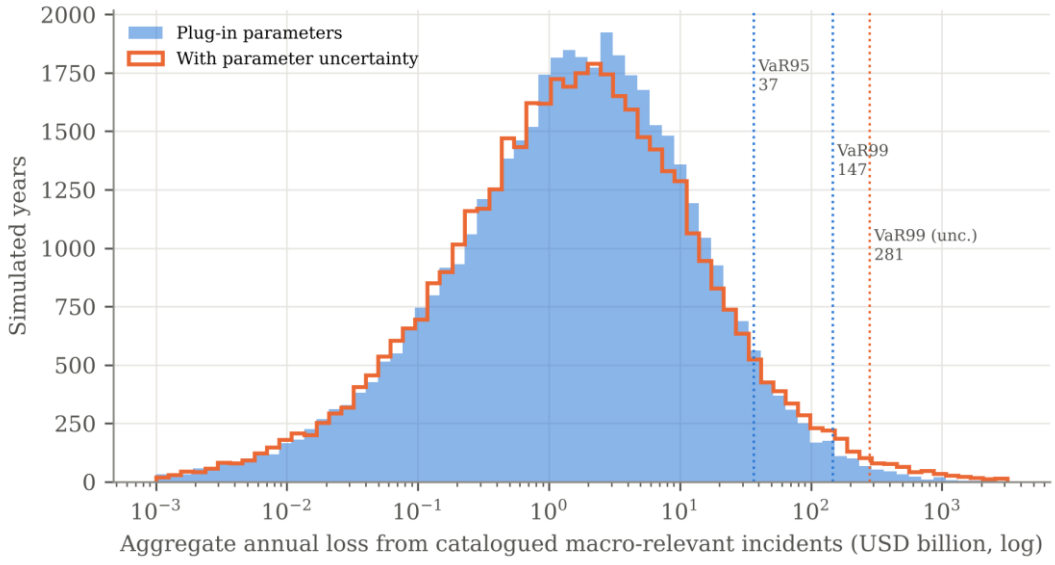
Note. Severities are published cost estimates (USD billion) for 12 malicious incidents, 2016–2025. LDA: compound Poisson–lognormal, 200,000 simulated years. Parameter uncertainty draws μ from its normal sampling distribution, σ^2 from a scaled χ^2 and λ from its Gamma posterior. Risk measures in USD billion.

Figure 8: Empirical complementary CDF of incident costs with fitted lognormal, Weibull and Pareto I distributions (log–log scale).



Module C results are in Table 9 and Figures 8 and 9. The costed malicious incidents range from USD 4.4 million (the Colonial Pipeline ransom) to about USD 10 billion (NotPetya), with a median of USD 0.55 billion and a mean of USD 1.82 billion — a mean-to-median ratio above three that signals strong right skew. The lognormal and Weibull distributions fit well (KS p = 0.94 and 0.87; AIC within one point of each other) and the Anderson–Darling test does not reject normality of log costs. A pure Pareto fit is clearly worse (AIC 35.3) and implies an implausible tail index of 0.23; the Hill estimator is unstable across k, falling from 1.54 to 0.24, and the GPD shape above the median is a modest ξ = 0.08. The evidence, in line with Edwards et al. (2016), therefore favours a lognormal body with a heavy but not infinite-mean tail. Annual counts are under-dispersed (dispersion index 0.43), so a Poisson frequency model is conservative.

Figure 9: Simulated distribution of the aggregate annual loss from catalogued macro-relevant incidents (compound Poisson–lognormal), with and without parameter uncertainty.



The loss distribution shows why point estimates of cyber cost are misleading. With plug-in parameters the median year sees catalogued losses of about USD 1.2 billion, but the expected loss is USD 10.6 billion and the 1-in-100-year loss (VaR 99%) is USD 147 billion, with an Expected Shortfall of USD 430 billion. Once uncertainty about the severity parameters is propagated unavoidable

with 12 observations the VaR 99% nearly doubles to USD 281 billion and the expected loss almost triples to USD 29.8 billion. H5 is supported: tail-risk measures are dominated by rare extreme events and by uncertainty about how extreme they can be. The Crowd Strike control case is instructive here. Its estimated USD 5.4 billion cost for Fortune 500 firms would sit at the 92th percentile of the malicious-incident distribution, which shows that the macro-relevant quantity for infrastructure resilience is the scale of disruption, not whether the cause is malicious.

Normalizing by host GDP gives a sense of national shock size. The NotPetya cost, attributed to Ukraine as the origin and most-affected country, equals 5.6% of Ukraine's GDP indicative only, because most of the damage fell on multinational firms outside Ukraine while the JLR incident equals 0.076% of UK GDP. For a small or emerging economy, a single NotPetya-scale event is a macroeconomic shock comparable to a severe natural disaster.

Predictive Resilience: Expected Annual Cyber Loss by Economy

Table 10: Projected 2028 Expected Annual Cyber Loss (EAL) by Economy, Ranked by Share of GDP.

Economy	GCI 2020	Gap	GDP 2023 (USD bn)	EAL (USD bn)	80% interval	EAL (% GDP)	80% interval
Nigeria	54.41	45.59	363	0.92	0.67–1.26	0.255	0.185–0.347
Pakistan	64.88	35.12	338	0.78	0.58–1.05	0.231	0.172–0.310
Ukraine	65.93	34.07	179	0.41	0.31–0.55	0.229	0.170–0.306
Iran, Islamic Rep.	81.06	18.94	401	0.78	0.60–1.02	0.195	0.150–0.253
Bangladesh	81.27	18.73	437	0.85	0.66–1.10	0.195	0.150–0.253
China	92.53	7.47	17,795	30.16	23.68–38.47	0.169	0.133–0.216
Indonesia	94.88	5.12	1,371	2.25	1.77–2.86	0.164	0.129–0.209
Oman	96.04	3.96	109	0.18	0.14–0.22	0.162	0.128–0.205
Italy	96.13	3.87	2,255	3.64	2.87–4.62	0.161	0.127–0.205
Brazil	96.60	3.40	2,174	3.48	2.75–4.43	0.160	0.126–0.204
Germany	97.41	2.59	4,456	7.06	5.58–8.96	0.158	0.125–0.201
Australia	97.47	2.53	1,724	2.73	2.16–3.46	0.158	0.125–0.201
India	97.49	2.51	3,550	5.62	4.44–7.13	0.158	0.125–0.201

Economy	**GCI 2020**	**Gap**	**GDP 2023 (USD bn)**	**EAL (USD bn)**	**80% interval**	**EAL (% GDP)**	**80% interval**
France	97.60	2.40	3,031	4.79	3.78–6.08	0.158	0.125–0.201
Canada	97.67	2.33	2,140	3.38	2.67–4.29	0.158	0.125–0.200
Japan	97.82	2.18	4,213	6.63	5.24–8.42	0.157	0.124–0.200
United Arab Emirates	98.06	1.94	504	0.79	0.62–1.00	0.157	0.124–0.199
Malaysia	98.06	1.94	400	0.63	0.50–0.80	0.157	0.124–0.199
Singapore	98.52	1.48	501	0.78	0.62–0.99	0.156	0.123–0.198
Korea, Rep.	98.52	1.48	1,713	2.67	2.11–3.39	0.156	0.123–0.198
Spain	98.52	1.48	1,581	2.46	1.95–3.13	0.156	0.123–0.198
Estonia	99.48	0.52	41	0.06	0.05–0.08	0.154	0.121–0.195
United Kingdom	99.54	0.46	3,340	5.13	4.05–6.50	0.153	0.121–0.195
Saudi Arabia	99.54	0.46	1,068	1.64	1.30–2.08	0.153	0.121–0.195
United States	100.00	0.00	27,361	41.70	32.97–52.90	0.152	0.120–0.193

Note. EAL = GDP × ρ_{2028} × (1 + κ·GAP/10), with ρ_{2028} drawn from the hybrid-ensemble projection and κ ~ U(0.05, 0.25); 20,000 Monte Carlo draws. EAL is scaled on US-reported (IC3) losses and is therefore a lower-bound, comparable exposure index rather than a full estimate of economic cost.

Figure 10: **Projected 2028 expected annual cyber loss as a share of GDP (median and 80% interval), by economy.**

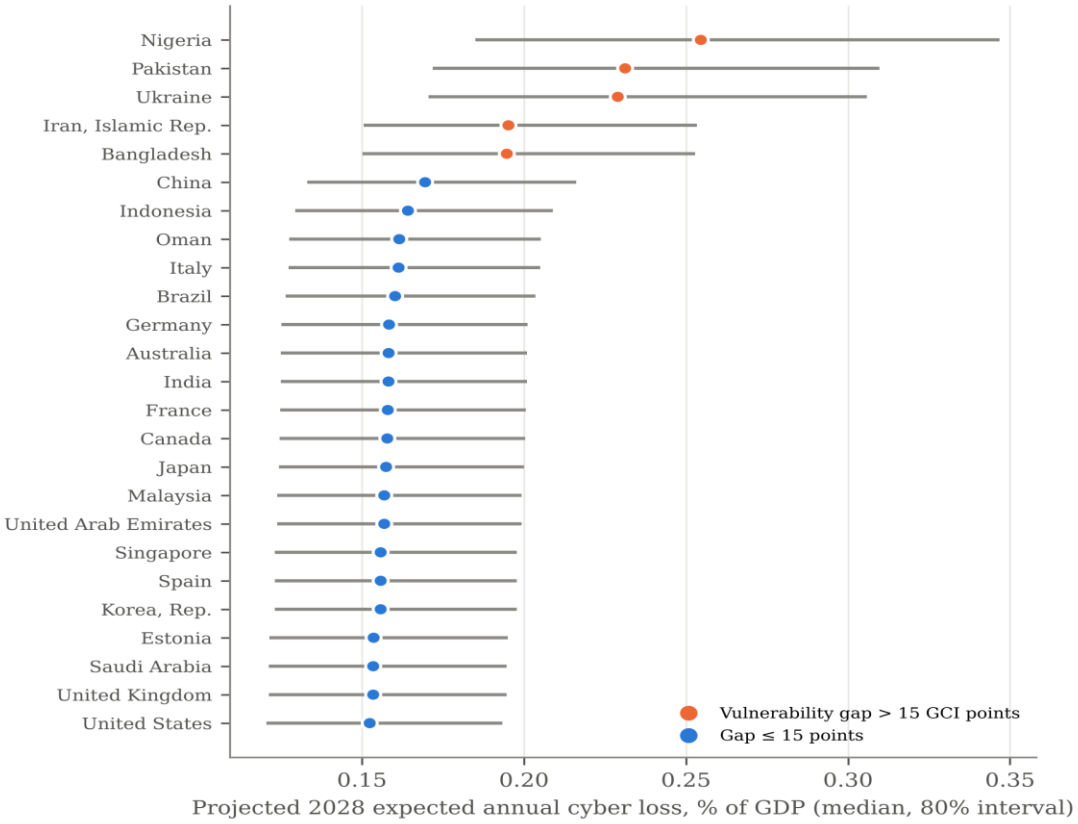


Table 10 and Figure 10 combine the three modules. The projected loss intensity rises from 0.058% of GDP in 2024 to 0.153% in 2028 for an economy with no vulnerability gap. Economies with large gaps face a markedly higher burden: the median EAL is 0.255% of GDP in Nigeria, 0.231% in Pakistan and 0.229% in Ukraine, against 0.152% in the United States an excess of about 50–67%. In absolute terms, exposure is concentrated in the largest economies (median EAL of USD 41.7 billion for the United States and USD 30.2 billion for China), and the 25 economies together face a median reported-loss-equivalent EAL of about USD 130 billion in 2028. Because κ is treated as uncertain, the 80% intervals for the high-gap economies overlap only partially with those of the leaders; the ranking is robust, the precise level is not. The practical implication for critical infrastructure policy is that closing the capability gap rather than scale or income alone is the lever that most reduces the expected macro cost of cyberattacks in emerging economies.

Discussion

Summary of Hypothesis Tests

Table 11 summarizes the evidence on the five hypotheses.

Table 11: Summary of Hypothesis Tests.

Hypothesis	Evidence	Verdict
H1 Threat–loss link and lead	Elasticities > 1 (M1, M2); P(β > 0) > 0.99; CVE Granger-leads losses; Shapley lag-2 CVE dominant	Supported
H2 Stability	No trend break (sup-F boot p = 0.31; Chow 2020 p = 0.86); static CVE-only elasticity unstable (CUSUM); no cointegration	Partly supported

Hypothesis	Evidence	Verdict
H3 DL superiority	GBM best (RMSE 0.113); LSTM 0.191; no significant DM differences among learners	Not supported
H4 Income and scale	High-income effect 31.6 pts at $\tau = 0.10$ vs 2.3 at $\tau = 0.90$; GDP not significant in QR	Supported for income only
H5 Heavy-tailed severity	Lognormal $\sigma = 2.32$; mean/median > 3; parameter risk doubles VaR 99%	Supported

Implications for Critical Infrastructure Protection and Policy

Three implications follow for the protection of critical infrastructure. First, the lead of vulnerability disclosures over losses gives national CERTs and regulators a one-to-two-year early-warning window. Disclosure volumes, weighted by the share affecting operational technology and widely deployed enterprise software, should be monitored as a macro-prudential indicator in the same way that credit growth is monitored for financial stability (Bouveret, 2018; Kopp et al., 2017). The patch-management failures behind Equifax, the MOVEit mass exploitation and the CDK Global outage in the incident set illustrate how quickly disclosed weaknesses turn into economy-wide losses.

Second, the severity results argue for a tail-risk approach to cyber resilience. Because expected losses are dominated by rare, extreme events many of them, like NotPetya and Change Healthcare, propagating through supply chains (Crosignani et al., 2023; Welburn & Strong, 2022) resilience planning should be organised around stress scenarios at the 99th percentile rather than around average incident cost. The near-doubling of VaR under parameter uncertainty also implies that public backstops for cyber insurance and sectoral reserve requirements should incorporate explicit model-risk buffers (Biener et al., 2015).

Third, the vulnerability analysis shows that the capability gap, not economic size, distinguishes the most exposed economies. For lower-middle-income economies such as Pakistan, Nigeria and Bangladesh, the quantile results imply that institutional investments captured by the GCI pillars legal frameworks, technical incident-response capacity, organizational strategy, capacity development and cooperation yield the largest marginal gains precisely where capability is weakest. Pakistan's reported movement to Tier 1 in the GCI 2024 (ITU, 2024) is an encouraging sign; the framework provides a way to track whether such gains translate into lower projected loss intensity. Because the energy and public-revenue systems of these economies have already been targeted (the 2020 K-Electric ransomware and 2021 Federal Board of Revenue attacks in Pakistan; the grid and telecom attacks in Ukraine), protecting a small number of systemically important operators would address a large share of the exposure.

Finally, the forecasting results carry a methodological lesson. Deep learning should not be assumed to be the best tool for macro-cyber forecasting. With short annual series, regularized tree ensembles and exponential smoothing match or beat recurrent networks, and the real advantage of a hybrid design is robustness and diagnostic richness econometric tests to establish what the data can and cannot identify, machine learning to capture nonlinear lags, and Shapley values to keep the result interpretable for decision makers. As higher-frequency threat-intelligence data (monthly CVE and ransomware leak-site counts, insurance claims) become available, the balance is likely to shift towards deep sequence models.

Limitations and Directions for Future Research

The study has several limitations. The annual loss series contains only ten complete observations and the forecast evaluation only four origins, so the power of unit-root, cointegration and Diebold–Mariano tests is low; we have mitigated this with bootstrap and Bayesian inference, but the conclusions about long-run elasticities remain tentative. IC3 losses are US victim-reported losses and understate the true global cost; extending the framework to other national reporting systems (for example, Action Fraud in the United Kingdom or national CERT statistics in Pakistan) would broaden its coverage. Incident costs mix firm-reported, insurer, government and model-based estimates, and two global events are attributed to a single country, so severity parameters carry measurement error in addition to sampling error. Several GCI and GDP entries in the source workbook are flagged for verification against primary ITU and World Bank releases, and the GCI 2024 tier is available for only some economies. The vulnerability multiplier κ is a calibrated scenario parameter rather than an estimated one. Future work should estimate κ from cross-country loss data, introduce input–output propagation to capture indirect losses, and test the framework on panel data with higher frequency.

Conclusions

This paper has proposed and estimated a hybrid deep learning, machine learning and econometric framework for quantifying the macroeconomic impact of cyberattacks on national economic systems. Applied to a fully sourced dataset of threat-intelligence series, national cyber-capability indicators and documented macro-relevant incidents, the framework yields five main findings. Reported cyber losses have grown at about 36% a year, faster than either vulnerability disclosures or attack volume, and vulnerability disclosures lead losses by one to two years. The growth path of losses has been stable through the pandemic shock, but no stable single-driver long-run elasticity can be identified from the available data. In small-sample forecasting, gradient boosting outperforms an LSTM deep ensemble, and the value of deep learning lies in the hybrid combination rather than in stand-alone accuracy; the hybrid ensemble projects US-reported losses of about USD 52 billion by 2028. Incident severity is lognormal and heavy-tailed, and parameter uncertainty roughly doubles the 99% Value-at-Risk. Finally, high-income status raises national cyber-capability most among the weakest performers, and economies with large capability gaps Nigeria, Pakistan and Ukraine among them face the highest projected cyber loss burden relative to GDP. For critical infrastructure protection, these results support treating vulnerability-disclosure trends as an early-warning indicator, planning for tail scenarios rather than averages, and directing capacity-building investment to the economies where the capability gap is largest.

Appendix A: Model Hyperparameters

Table A1: Hyperparameters of the Deep Learning and Machine Learning Models (Fixed a Priori).

Model	Architecture / settings	Training
LSTM deep ensemble	Input 3 features × 2 time steps; 1 LSTM layer, 6 hidden units; linear output; forget-gate bias 1	BPTT, Adam (lr 0.01), 400 epochs, L2 = 0.01, gradient clip ±5; 15 members (30 for projection); standardized inputs and target
Deep MLP	Hidden layers 16–8–4, tanh activation	Adam (lr 0.01), L2 α = 0.1, max 4,000 iterations; 15-member ensemble
Random forest	500 trees, min. leaf size 1, bootstrap sampling	Flattened 6-feature input
Gradient boosting	300 trees, depth 2, learning rate 0.05, subsample 0.8	Squared-error loss
ARX–LSTM hybrid	ARX (ridge 10^{-3}) + 0.5 × LSTM residual forecast	LSTM settings as above, fitted to ARX residuals
Isolation forest	1,000 trees, contamination 0.2	4-dimensional annual growth vectors
k-means / Ward	$k \in \{2,\ldots,6\}$ by silhouette; 50 initializations	Standardized features

References

1. Anderson, R., Barton, C., Böhme, R., Clayton, R., Ganán, C., Grasso, T., Levi, M., Moore, T., & Vasek, M. (2019). *Measuring the changing cost of cybercrime* [Paper presentation]. 18th Workshop on the Economics of Information Security (WEIS), Boston, MA, United States.
2. Anderson, R., Barton, C., Böhme, R., Clayton, R., van Eeten, M. J. G., Levi, M., Moore, T., & Savage, S. (2013). Measuring the cost of cybercrime. In R. Böhme (Ed.), *The economics of information security and privacy* (pp. 265–300). Springer.
3. Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica, 61*(4), 821–856.

4. Biener, C., Eling, M., & Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance – Issues and Practice, 40*(1), 131–158.

5. Bouveret, A. (2018). *Cyber risk for the financial sector: A framework for quantitative assessment* (IMF Working Paper No. WP/18/143). International Monetary Fund.

6. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

7. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294.

8. Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological), 37*(2), 149–192.

9. Council of Economic Advisers. (2018). *The cost of malicious cyber activity to the U.S. economy.* Executive Office of the President.

10. Crosignani, M., Macchiavelli, M., & Silva, A. F. (2023). Pirates without borders: The propagation of cyberattacks through firms' supply chains. *Journal of Financial Economics, 147*(2), 432–448.

11. Cyber Monitoring Centre. (2025, October). *Statement on the Jaguar Land Rover cyber incident.* https://cybermonitoringcentre.com/

12. Davidson, R., & Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics, 146*(1), 162–169.

13. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263.

14. Dreyer, P., Jones, T., Klima, K., Oberholtzer, J., Strong, A., Welburn, J. W., & Winkelman, Z. (2018). *Estimating the global cost of cyber risk: Methodology and examples* (Report No. RR-2299-WFHF). RAND Corporation.

15. Edwards, B., Hofmeyr, S., & Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity, 2*(1), 3–14.

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

17. Eling, M., McShane, M., & Nguyen, T. (2021). Cyber risk management: History and future research directions. *Risk Management and Insurance Review, 24*(1), 93–125.

18. Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance.* Springer.

19. Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*(2), 251–276.

20. Federal Bureau of Investigation. (2025). *Internet crime report 2024.* Internet Crime Complaint Center. https://www.ic3.gov/AnnualReport/Reports

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

22. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

23. Gamblin, J. (2025, January 5). *2024 CVE data review.* https://jerrygamblin.com/2025/01/05/2024-cve-data-review/

24. Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security, 5*(4), 438–457.

25. Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438.

26. Greenberg, A. (2019). *Sandworm: A new era of cyberwar and the hunt for the Kremlin's most dangerous hackers*. Doubleday.
27. Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291.
28. Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics, 3*(5), 1163–1174.
29. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
30. Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*(1), 5–10. (Original work published 1957)
31. IBM Security. (2025). *Cost of a data breach report 2025*. IBM Corporation. https://www.ibm.com/reports/data-breach
32. International Telecommunication Union. (2021). *Global cybersecurity index 2020*. ITU.
33. International Telecommunication Union. (2024). *Global cybersecurity index 2024* (5th ed.). ITU.
34. Jamilov, R., Rey, H., & Tahoun, A. (2021). *The anatomy of cyber risk* (NBER Working Paper No. 28906). National Bureau of Economic Research.
35. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review, 55*(2), 163–172.
36. Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics, 139*(3), 719–749.
37. Kingma, D. P., & Ba, J. (2015). *Adam: A method for stochastic optimization* [Paper presentation]. 3rd International Conference on Learning Representations (ICLR), San Diego, CA, United States.
38. Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33–50.
39. Koop, G. (2003). *Bayesian econometrics*. John Wiley & Sons.
40. Kopp, E., Kaffenberger, L., & Wilson, C. (2017). *Cyber risk, market failures, and financial stability* (IMF Working Paper No. WP/17/185). International Monetary Fund.
41. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems 30* (pp. 6402–6413). Curran Associates.
42. Lee, R. M., Assante, M. J., & Conway, T. (2016). *Analysis of the cyber-attack on the Ukrainian power grid*. Electricity Information Sharing and Analysis Center; SANS Industrial Control Systems.
43. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413–422). IEEE.
44. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.
45. MacKinnon, J. G. (2010). *Critical values for cointegration tests* (Queen's Economics Department Working Paper No. 1227). Queen's University.
46. Maillart, T., & Sornette, D. (2010). Heavy-tailed distribution of cyber-risks. *The European Physical Journal B, 75*(3), 357–364.
47. McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools* (Rev. ed.). Princeton University Press.
48. Narayan, P. K. (2005). The saving and investment nexus for China: Evidence from cointegration tests. *Applied Economics, 37*(17), 1979–1990.

49. National Institute of Standards and Technology. (n.d.). *National vulnerability database*. Retrieved November 1, 2026, from https://nvd.nist.gov/

50. Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*(3), 703–708.

51. Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*(6), 619–632.

52. Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16*(3), 289–326.

53. Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 31*(2), 350–371.

54. Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity, 2*(2), 121–135.

55. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.

56. Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games* (Vol. 2, pp. 307–317). Princeton University Press.

57. Stock, J. H., & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica, 61*(4), 783–820.

58. Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*(1), 24–36.

59. Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

60. Welburn, J. W., & Strong, A. M. (2022). Systemic cyber risk and aggregate impacts. *Risk Analysis, 42*(8), 1606–1622.

61. Woods, D. W., & Böhme, R. (2021). SoK: Quantifying cyber risk. In *2021 IEEE Symposium on Security and Privacy (SP)* (pp. 211–228). IEEE.

62. World Bank. (2024). *World development indicators: GDP (current US$)* [Data set]. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

63. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing, 50*, 159–175.



2026 by the authors; Journal of Global Social Transformation (JGST). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).