SoK: Quantifying Cyber Risk

Daniel W. Woods
University of Innsbruck
Innsbruck, Austria
daniel.woods@uibk.ac.at

Rainer Böhme
University of Innsbruck
Innsbruck, Austria
rainer.boehme@uibk.ac.at

2021 IEEE Symposium on Security and Privacy (SP) | 978-1-7281-8934-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/SP40001.2021.00053

***Abstract*—This paper introduces a causal model inspired by structural equation modeling that explains cyber risk outcomes in terms of latent factors measured using reflexive indicators. First, we use the model to classify empirical cyber harm studies. We discover cyber harms are not exceptional in terms of typical or extreme losses. The increasing frequency of data breaches is contested and stock market reactions to cyber incidents are becoming less damaging over time. Focusing on harms alone breeds fatalism; the causal model is most useful in evaluating the effectiveness of security interventions. We show how simple statistical relationships lead to spurious results in which more security spending or applying updates are associated with greater rates of compromise. When accounting for threat and exposure, indicators of security are shown to be important factors in explaining the variance in rates of compromise, especially when the studies use multiple indicators of the security level.**

***Index Terms*—cyber risk, security metrics, cyber harm, control effectiveness, science of security, causal model, structural equation modeling**

I. Introduction

Unsupported claims about the increasing risk of cyber attacks pervade introductions to security talks and papers. Organisations are expected to invest more in security even though research has inconsistently demonstrated how interventions reduce risk. This state of affairs leads to perceptions that cyber risk is more art than science.

With this in mind, our paper aims to systematise what is known about quantifying cyber risk. Risk estimates can justify additional resources for mitigation or be used to guide post-incident response. The term *cyber* will bristle with many in the security community. However, it is the concept of choice for policymakers and business leaders who make many of the decisions that security research should hope to influence. Such decisions are premised on foundational questions like:

RQ1 How much harm results from cyber incidents?
RQ2 Which security interventions effectively reduce harm?
RQ3 Have these answers changed over time?

Whereas security vendors scramble to provide self-interested answers with shaky methodologies [7, 82], this paper finds answers in empirical studies of real-world security outcomes. We systematise the literature using a causal model linking latent variables for security, exposure, and threat to security outcomes. The proposed model captures empirical cyber risk research ranging from machine learning models predicting web server compromise through to finance studies quantifying shareholder losses resulting from cyber incidents.

We focus on classifying studies quantifying cyber risk in organisations. The term *cyber risk* has two components, *risk* describes possible negative consequences (*harm*) weighted by the probability of occurrence. *Cyber* restricts our scope to incidents caused by logical (as opposed to physical) force [17]. Under this definition a fire (physical force) in a data centre (information harm) is not a cyber risk, whereas fire damage (physical harm) caused by compromised control systems (logical force) would be. *Incidents* within scope include denial of service attacks, machine and web-resource compromise, and organisational incidents. Associated harms range from lost shareholder value to ransomware payments to wasted time.

Our literature search first identified relevant works in top security conferences and the *Workshop on the Economics of Information Security*. We used backwards and forwards reference searches to identify additional relevant works until saturation was reached. Doing so captured relevant studies from disciplines including law, information systems, finance, and physics. We included studies that empirically measure real-world compromise or harm affecting organisations, which is a minority approach within the science of security [60, p. 12]. Studies providing promising ways of measuring security, exposure or threat were also included even when harm was not considered. We point readers to Anderson et al. [7] for aggregate estimates of cybercrime costs, and Dambra et al. [32] for cyber risk transfer research.

Section II introduces the causal model. Section III surveys harm studies speaking to **RQ1**. Section IV identifies mitigation studies that address **RQ2**. Temporal trends are identified throughout (**RQ3**). Section V discusses progress towards **RQ1–3**, model limitations, and future work.

II. A causal model of cyber risk

Risk is unobservable but we can indirectly measure its realisation as losses. Figure 1 uses artificial data to illustrate the stochastic relationship; the highest observed loss has multiple twins with similar security levels but much smaller losses.

We thank Stefan Laube and the anonymous reviewers for their thoughtful comments and the cited authors who responded to our request for feedback, especially Ben Stock, Camelia Simoiu, Kanta Matsuura, Martin Loeb, and Stefan Savage. The causal model grew out of Dagstuhl Seminar 16461, in particular the breakout group chaired by the second author. It was refined at the Empirical Cybersecurity Research Winter School in Obergurgl, Austria, 2019. The first author thanks Tom Verstraten for extolling the value of related work. The project is funded by the European Commission's call H2020-MSCA-IF-2019 under grant number 894700.

DOI 10.1109/SP40001.2021.00053

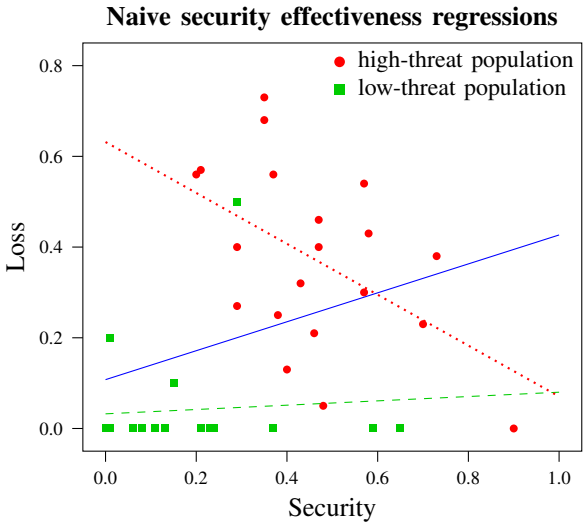


Fig. 1. The solid blue line fails to account for threat level, which may lead the high-threat population to under estimate the effectiveness of security.

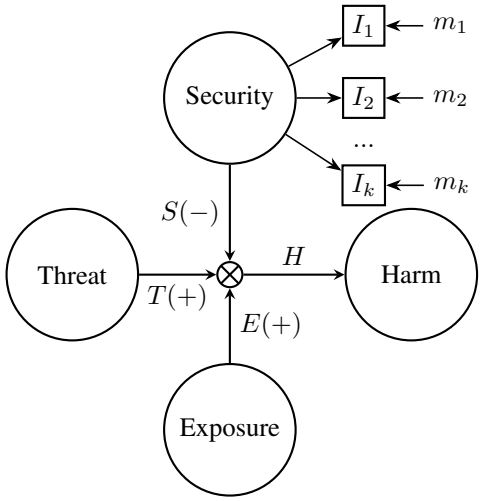


Fig. 2. A high-level causal model of cyber risk.

Regression analysis is designed to explore relationships in the presence of statistical noise. The Appendix contains a brief tutorial on regression analysis but the confident reader may continue. Putting measurement issues to one side, fitting a linear model in which security is the only explanatory variable (the blue line) suggests increasing security is associated with greater losses. This result has been found empirically—higher IT security budgets are associated with greater frequency of data breaches [105]. Research designs based on observational data are vulnerable to confounding variables and so we need to add relevant variables to the regression model.

Adding threat level leads to a better fit (see the Appendix for more detail) and provides insights into cross-dependencies. In Figure 1, the red dotted line slopes downwards while the green dashed line has a (not statistically significant) upward slope, which suggests security only reduces harm when it is implemented by the high-threat population. In fact, threat is the only necessary condition for harm to occur. As such, security should be conceptualised as *moderating* the relationship between threat and harm, such that more security translates into less expected harm.

The intuition that security effectiveness depends on the threat level is baked into risk management. If security is mitigation in risk management, then a third variable, *exposure*, is analogous to the amount of risk acceptance. More exposure means more vectors can be used to gain access (*surface exposure*) and a greater value of assets can be compromised (*asset exposure*), both of which amplify the effect of threat on expected harm. Figure 2 represents the relationship between threat level and expected harm as moderated by security and exposure. The notation $E(+)$ denotes a positive relationship in which more exposure amplifies the link between threat and harm, and $S(-)$ denotes a negative relationship. Many research designs fail to account for all three variables.

Measuring these abstract variables is challenging in practice. Reported losses ignore the full spectrum of harms [2] and harm is often avoided thanks to luck alone. The effect of security in moderating this stochastic relationship between threat and harm is even more difficult to measure. No single indicator captures the sum of preventive and reactive measures across an organisation's technology, processes and people. Modelling security as a *latent variable* overcomes this issue by linking noisy, observable indicators to the high-level concept.

Figure 2 shows this graphically as the latent variable for security has *reflexive indicators* I_1,...,I_k, which can be measured with m_i. The arrows flow from security to the indicators I_i because the indicators do not cause security, rather the security level influences the likelihood of a given measurement m_i. The latent variable must be inferred from these reflexive indicators.

Although Figure 2 describes the ideal research design, relationships like $S(-)$ can still represent statistical models comprised of *manifest variables* in which variables like security or threat are assumed to be directly measurable. This allows us to systematise a diverse body of literature on cyber risk quantification and show which factors determining risk outcomes are under consideration.

Very few studies directly link security interventions to harm outcomes. It is useful to introduce a mediating factor, *compromise*, which may or may not result in harm. Studies investigating the effectiveness of security controls tend to focus on an indicator of compromise without quantifying the resulting harm, whereas studies quantifying harm tend to sample exclusively from compromised entities ($C = 1$ in our model). Doing so cannot quantify how preventative security affects incident likelihood because $C = 1$ for all firms.

Often these studies explore how harm varies based not on the threat level, which is stochastic and difficult to measure, but based on the threat actor (e.g. ransomware gang) who caused an incident. While this provides some information about the threat level (e.g. a more sophisticated actor suggests a higher threat level), it cannot be measured for firms who were not compromised ($C = 0$). We denote this special case

$T|C$, namely the type of threat conditioned on compromise. Figure 3 shows the extended model using SEM notation.

To make this concrete, a 2017 study [116] is illustrated on Figure 3 using red arrows and indicators I_x. The authors argue that although indicators of security (e.g. hiding version numbers or the SSL configuration) correlate with less abuse when aggregated across web hosting providers, these variables do not directly cause security improvements. Rather the indicators are assumed to be *reflexive indicators* of an unobservable security level. The authors [116] construct four latent variables for preventive security S_p in order to explain website compromise C when controlling for surface exposure E_s. Table VI (in the Appendix) describes the corresponding technical indicators.

The rest of this paper systematises literature on cyber risk according to which of the relationships depicted in Figure 3 are explored. We focus on the statistical tests in the main contribution and ignore preliminary results or tables. This will be summarised in Table III. Our classification requires a fair amount of interpretation because assumptions are often unstated. For example, many data breach studies do not control for the size or industry of the victim, which we suggest is an implicit assumption that threat and exposure are constant across the analysed sample. We sent the paper to an author of each study and received no objections to our classification.

III. Cyber harm studies

The section speaks to the frequency and impact of cyber harm (**RQ1**). Classifying harm research using Figure 3 will reveal that these studies infrequently consider the moderating effect of security (S_p and S_r in Figure 3). With this in mind, a secondary goal is to identify which data sources could be used by mitigation studies in future work.

Table I gives an overview of empirical approaches to quantifying cyber harm. Section III-A considers data sources that collate public reports, whereas the studies in Section III-B rely on researchers collecting private reports. Studies in Section III-C extract data from publicly observable systems like courts proceedings or stock markets. Section III-D considers research into harms resulting from system wide events.

TABLE I
Overview of different approaches to quantifying cyber harm.

Unit of analysis	# of studies	Econ loss	Sample size	Earliest study	Earliest sample
Public reports (Section III-A)					
Data breach	9	✗	600–6160	2008	2000
Operational loss	3	✓	341–1579	2015	<2003
Cyber incident	1	✓	2216	2016	2005
Private reports (Section III-B)					
Internal incident	2	✗	1800–23000	2010	1996
Insurance claim	1	✗	70	2019	2015
Crime reports	1	✓	7925	2020	2017
Firm survey response	3	✓	664–4209	2012	2012
Individual survey response	5	✓	1500–64287	2014	2010s
Externally observed (Section III-C)					
Legal case	2	✗	19–230	2011	1999
Legal case	1	✓	118	2017	2010
Bitcoin transaction	3	✓	10m	2014	2009
Criminal forum post	2	✓	13m	2007	2006
Insurance prices	1	✓	6828	2019	2007
Stock market reaction	19	✓	43–542	2003	1988
System-wide harm (Section III-D)					
Multi-party incident	1	✓	800	2019	2008

A. Publicly reported

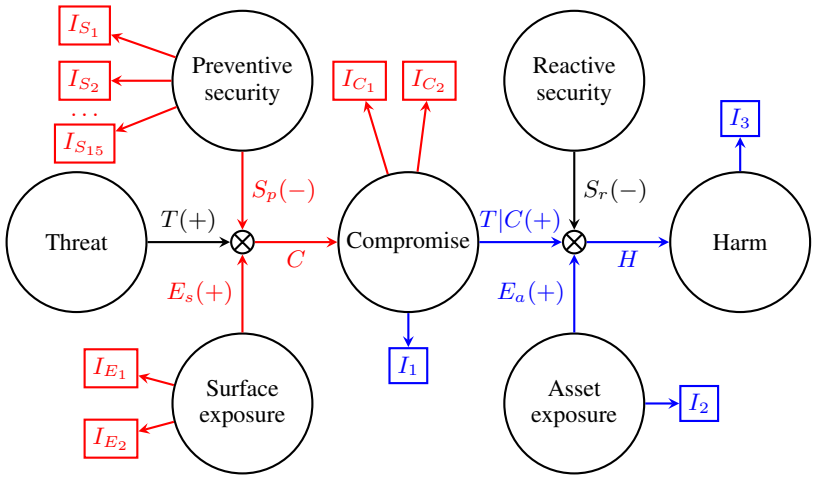
Organisations report cyber incidents to the public for both strategic reasons and compliance with reporting requirements [72]. Data brokers aggregate these reports to create pay-for-access databases, with some exceptions like Privacy Rights Clearinghouse providing free access. Large organisations are over-represented because their reports are more accessible.

Data breach studies Data breach studies only sample the sub-population of firms who have suffered a breach, which means harm is conditioned on a breach occurring. These studies estimate how the number of breached records is distributed. We do not count estimating the frequency of breaches across the entire US as investigating the probability of compromise since these estimates provide little information to organisations without knowing the population of possible victims [67]. Two studies [43, 127] addressed this by using the population of listed companies to estimate probability of breach, which is an indicator of C in Figure 3.

Using the same public reports means each study can only add data collected since the last study. Each researcher adopts more sophisticated methods to justify publication. Breach sizes were fitted with: just 1 parameter in 2010 [83], 2 and 3 in 2016 [36, 127], 6+ in 2018 [131], and the endlessly flexible regression trees in a 2020 study [43]. On the one hand, model sophistication identifies relationships that simple analyses cannot, such as Xu et al. [131] showing that the expected magnitude of the next breach increases with the time since the last breach. On the other hand, the proliferation of statistical tests leads to contradictory results (see Table II).

There is no consensus on whether breach frequency/size are stable over time (**RQ3**). They were shown to be decreasing/stable [40], stable/stable [36], increasing/stable [83, 131], and stable/increasing [127]. Many of the contradictions can be explained by how the data is sliced. Breach size was only found to be increasing in malicious breaches [128, 131] but never for negligent breaches. Frequency was only found to be increasing in the early years [30, 83] or in samples of malicious breaches [23, 131].

In terms of **RQ1**, the shape parameter in the distribution of breach size implies the expected number of breached records is infinite in some studies [43, 83] and finite in others [36, 131]. The possibility that the expected cost of a data breach is infinite raises two problems. First, the number of breached records is bounded by the number of records held [127]. Second, it is unclear how this maps to financial cost, which mandatory reporting laws do not require organisations to publish. The Jacobs Transform is frequently used to map the number of records to a financial cost [23, 36, 40, 43]. This



Description of latent factors

Threat
The motivation, capability and activity of adversaries.

Surface exposure
Factors increasing potential vectors of compromise.

Preventive security
Interventions reducing the ease of compromise.

Compromise
Violation of victim security goal.

Asset exposure
Factors increasing the value of what can be compromised.

Reactive security
Interventions reducing the impact of compromise.

Harm
Negative consequences resulting from compromise.

Fig. 3. Describing the causes and correlates of cyber harm. The red arrows depict the model used by Tajalizadehkhoob et al. [116]. The blue arrows describe a simpler model describing harm as indicated by I_3 in terms of the type of compromise (I_1) and an indicator of asset exposure (I_2).

"transform" was derived in a blog post in which the author warns the "amount of variance in the model is a serious challenge to adoption" [66].

The predictive power of data breach studies is questionable. In 2016, Edwards et al. [36, Fig. 9] estimated that the probability of seeing a breach of 200 million or more records in the next 3 years had a probability of around 0.1. Wheatley et al. [127] derived a *maximum* breach size of 200 million, growing by 50% in the five years following 2016. Yahoo! reported the loss of 3 billion customer records in the same year as both publications (albeit lost years earlier). What do we really know about data breaches when even methods designed for tail events like extreme value theory [127] set bounds that are exceeded by an order of magnitude within the same year (with multiple breaches exceeding 500 million in the last 3 years)? The same authors [127] who derived the maximum bound warned about "dragon kings" [108] emerging from complex systems that risk models cannot predict.

Operational loss Much like data breach studies, op loss studies only consider harm H but in terms of financial loss. Two studies [13, 41] control for indicators of exposure E_a like industry, revenue, and employee count. Surprisingly, cyber operational losses are *less* heavy tailed than non-cyber losses [13]. The mean loss is also smaller, which suggests idiosyncratic cyber risk is not exceptional in the class of operational losses. A 2019 study [41, p. 10] supported the finding by reporting that non-cyber losses had greater "mean, standard deviation, median, skewness, and kurtosis". However, the authors report that the "tail risk measure" from [86, p. 283] is higher for cyber losses. This provides another example of model sophistication leading to contradictory results.

The ORX database [102] used in these studies comprises publicly reported operational losses. Larger organisations are over-represented in the data as they are more likely to suffer losses exceeding the threshold ($100k) and more likely to have the loss reported by a "major media source" [102]. The keyword filters used to filter cyber losses introduces additional noise. Only 25% of the losses in the study [41] were classified as data breaches, so what are the rest? It is hard to say without access to the proprietary data, but the largest cyber loss ($14.4 billion) resulted from a "money-laundering incident at the Bank of China in February 2005" [41, f. n. 9]. If you squint hard enough in the 21st century, anything can be a cyber loss.

Manually compiled The proprietary dataset used by Romanosky [98] collected publicly reported incidents using automated and manual methods. The results suggest cyber incidents are "far less" [98] costly than losses like fraud, theft, and bad debt when comparing medians and averages. In terms of frequency, he observed that "Health Care and Retail industries, however, suffer extremely low incident rates of around 0.3% or less" [98, p. 125]. This is likely an underestimate because the denominator captures all firms in the US census, whereas the numerator under reports losses of small firms, which are unlikely to be reported publicly. Normalizing a sample of publicly reported incidents is challenging because reporting biases are unknown and thus so is the population from which the sample was drawn.

B. Privately reported

Privately reported data must be collected directly from the organisation, which creates the opportunity to collect a representative sample as in surveys. In contrast, case studies collect a convenience sample using a relationship with one firm, which calls into question how well the results generalise.

Case studies Time to repair following a system failure is an indicator of harm H. Franke et al. [46] estimate the distribution of times to repair and suggest exploring the factors influencing this as future work. Schroeder and Gibson [104] show that both time to repair and frequency of failure depend on system complexity, an indicator of surface exposure E_s. Both of these studies use internal data meaning $n = 1$ in terms of organisations studied. The lack of consideration for security

TABLE II
THE OFTEN CONTRADICTORY FINDINGS FROM DATA BREACH STUDIES.

				Breach frequency		Breach size		∞
Reference	**Type of data**	n	**Years**	**Distribution**	**Trend**	**Distribution**	**Trend**	**Moment**
Curtin et al. (2008) [30]	N + M (USA)	899	2005–07	?	↗	?	?	?
Maillart et al. (2010) [83]	N + M (USA)	956	2000–08	?	↗	Power law	→	Yes
Edwards et al. (2016) [36]	N + M (USA)	2253	2005–15	Negative binomial	→	Lognormal (M)	→	No
Wheatley et al. (2016) [127]	M (World)	5365	2007–15	Poisson gen LM	→ (USA)	DT power law (t)	↗	Yes*
Eling et al. (2017) [40]	N + M (USA)	2266	2005–15	Negative binomial	↘	Skew-normal	→	No
Xu et al. (2018) [131]	M (USA)	600	2005–17	ARMA/GARCH	↗	Gen Pareto (t)	→	No
Wheatley et al. (2019) [128]	N + M (USA)	1713	2005–17	Negative binomial	→	Pareto	→, ↗ (M)	?
Carfora et al. (2019) [23]	N + M (USA)	5724	2005–17	Negative binomial	↗ (M)	Skew/Lognormal	?	No
Farkas et al. (2020) [43]	N + M (USA)	6160	2005–19	Binomial	?	Lognormal (t)	?	Yes

N/M = Negligent/malicious breach, (t) = distribution of the tail, LM = linear model, DT = double truncated, * = without maximum, ? = not reported.

is unsurprising given many failures are not malicious, but one could imagine future work restricted to security failures.

Case studies of organisations who aggregate data across multiple firms may provide general insights. Axon et al. [9] analyse 70 insurance claims from one insurer and show that response services are the most common costs, which is explained by how insurers encourage insureds to use post-incident services [45, 129]. Axon et al. [9] provide no quantitative estimates, likely because insurers believe claims data constitutes a competitive advantage [129]. Public sector organisations may be more willing to share data. Simpson and Moore analyse 7 925 attempted wire transfer thefts reported to the FBI's Internet Crime Complaint Center and find results like "small thefts succeed less often" [107] and international thefts succeed more often.

Survey data In a survey, the researcher collects private reports directly. The UK Government commissioned a survey [121] quantifying the frequency and impact of cyber incidents according to the firm size and industry, which constitute simple estimates of $E_s \rightarrow C$ and $E_a \rightarrow H$, respectively. Heitzenrater and Simpson [58] combine the survey [121] with control effectiveness data to quantify the *return on security investment* for commercial products like anti-virus or firewalls.

Consumer surveys of cybercrime are too numerous to exhaustively survey. Riek et al. [96] identify the most important surveys [42, 56, 61, 97] in the US and the EU, which we use to characterise the kind of insights to be gleaned. Self-reported losses are used to indicate compromise [56, 61, 97], whereas the Eurobarometer [42] focuses on victimisation rates. Security information is collected, such as security spending [96], identity theft detection methods [56], or anti-virus installation [42], but not linked to harm outcomes. Estimates of expected harm or frequency of compromise C must be made with reference to the population from which the sample was drawn. Solving this issue with representative sampling results in victims comprising a small fraction of the sample [44]. Riek et al. [96] addressed both issues by over-sampling victims and accounting for this with a reverse-weighting.

In terms of **RQ1**, Riek et al. [96] show that "most victims report no losses, many lose little, and a few lose a lot" [96, p. 13]. Interestingly, Hernandez et al. [61] discover near identical victimisation rates in the UK as compared with a comparable US sample. Survey work emphasises time costs in dealing with the incident [96] and also maintaining security controls [58].

C. Observed externally

The remaining studies observe publicly accessible systems without interacting with the organisation, which leads to measurement bias towards what is observable.

Legal cases Legal systems are reasonably transparent. Studies reveal factors determining the likelihood of breach litigation in the US [99], the costs of regulatory fines in the UK [25], and describe the evolution of the security requirements in the FTC's prosecutions [19]. The actual harm is suffered by a third-party but these studies investigate the defendant's harm, in terms of costs assigned by court.

Romanosky et al. [99] discover no clear trend in the absolute number of litigated data breaches from 2005 to 2010 **(RQ3)**. They identify a number of factors impacting the probability a reported data breach will be litigated, such as the number of records breached. In the UK, only a "small" fraction of public breaches leads to fines [25], which average £110k of the £500k limit that is now much higher due to GDPR. Such estimates are limited to costs assigned by courts and regulators. Further, legal cases take years to resolve which introduces logistical difficulties in linking mitigation measures to legal outcomes.

Cybercrime ecosystem can be studied to extract indicators of harm, such as the typical ransomware payment. Three studies [79, 92, 110] used this to estimate the rate of compromise related to the CryptoLocker ransomware campaign varies over time (T). Two studies find that a specific ransomware campaign displays significant temporal variance **(RQ3)**. Paquet et al. [92] include an additional 34 ransomware families, which allows them to link harm to type of compromise indicated by payment amount and the campaign. Such estimates are difficult to link to the characteristics of the victim who suffered the loss or the mitigation measures employed.

Although not speaking to harm to specific victims, research directly measuring threat actors can be used to estimate aggregate costs of cybercrime. Data breach harms to consumers can be observed at the point at which stolen data is sold, such

as by monitoring public channels [47, 118] or by infiltrating private forums [4]. These markets are noisy, which may lead to exaggerated cost estimates [59]. Diffuse harms related to spam [75], unlicensed pharmacies [73, 85] or ransomware-at-scale [63] can be more reliably quantified at the source, namely the criminal operation. Interested readers should refer to Anderson et al. [7] for a definitive survey.

Insurance prices A sub-population of insurers file their pricing schemes with a regulator [100]. Woods et al. [130] extract these prices and show cyber insurance premiums trend downwards from 2008–2018 **(RQ3)**. They also introduce a method using these prices to quantify expected loss (**RQ1**). The method, which is analogous to model stealing [120], infers a loss distribution based on how the quoted premium varies with changes in the amount of insurance.

Stock market reaction studies quantify harm to shareholders as indicated by abnormal returns. All studies control for exposure E_a via victim industry or size. In terms of **RQ1**, perceptions of the economic impact of data breaches on stock market value have been characterised as "Much Ado about Nothing" [95], but this has a temporal dimension (**RQ3**). Both Gordon et al. [53] and Gay [51] provide evidence market reactions are becoming less negative over time. Figure 4 shows the decreasing effect by means of a meta-study.

Later studies suggest that corporate leaders learned how to mitigate the negative stock market reaction after a breach had occurred. Board-level incentives mean costlier attacks are less likely to be disclosed [5] and, when they are disclosed, the negative reaction is offset by the strategic release of positive news [51]. Two studies provide evidence of insider trading [29, 80], which undermines the methodology because the *abnormal* trading following a breach is not concentrated in the event window following public disclosure.

Stock market reactions could lead corporate leaders to divert more resources to security following an incident. A reduced negative shock is associated with breach disclosures that commit to "action-oriented" measures to improve security [125] and faster breach discovery [68]. Perhaps more importantly, victims are more likely to increase board oversight of cyber risk post-incident [68], which may lead more resources to be assigned to security. Markets reward news about security investments regardless of whether a breach occurred. Displaying cybersecurity awareness [11] or certifying to international standards [33, 93] leads to positive returns.

D. Correlated risk

Focusing on individual losses ignores what is perhaps the most extreme aspect of cyber loss-correlation across firms. Events impacting popular software and cloud providers may cause losses across many firms. The Morris worm infected up to 10% of the devices connected to the Internet in 1988 [67]. More recently, the NotPetya attack exploited a flaw in Windows to cause an estimated $\$10$ billion of damage across hundreds of companies [28, 54].

Multi-party incidents An industry report [31] extracted over 800 multi-party cyber incidents causing $5\,437$ distinct losses from the same proprietary source as [98]. This approach focused on harm premised on a multi-incident party occurring and how this varied by industry. The median and 95$^{\text{th}}$ percentile of multi-party incident losses ($\$1$m and $\$417$m) were an order of magnitude larger than for single-party incidents ($\$77$k and $\$16$m), although these figures are not normalised by the number of affected firms. Curiously, their data shows a cluster of three losses at the maximum value in the sample.

E. Summary

Data breaches and stock market reactions have received the most research attention. Market reactions became less negative over time [51, 53] (see Figure 4) as firms learned how to manipulate announcements [5, 29, 51, 80]. Table II shows many contradictory results about data breaches depending on how the data is sliced and the analysis methodology. Even more worryingly for the data breach studies, Eling et al. [41] show the distribution of number of records does not transfer to that of financial costs [41].

A minority of studies [13, 41, 98] quantify financial costs and find typical cyber risks are *smaller* and less heavy tailed than non-cyber losses. Surveys of firms [12, 58] and individuals [96] reveal less alarming harm estimates. The maximum loss in a survey [58] of small UK businesses was £310k ($\$410$k), whereas an op loss database's mean was $\$43$m [41]. This points to jurisdictional differences and the most worrying aspect of this section—cyber harm estimates are not consistent across samples or statistical methods.

IV. Cyber risk mitigation studies

This section is concerned by empirical studies of how security controls affect outcomes in real systems. Inductive security proofs and attack papers that only demonstrate an attack is possible are out of scope.

We proceed by highlighting promising ways to quantify latent variables, known as measurement models. Measurement models for threat, security and exposure are covered in Section IV-B, Section IV-A, and Section IV-C respectively. Finally, Section IV-D identifies the holy grail—research investigating the structural links between these variables. We classify research using the causal model throughout, which is summarised in Section IV-E and Table III.

A. Measuring security

A measurement model reduces a set of indicators to a lower dimensional output that can be used to explore structural relationships between latent variables. This subsection covers security measurement models based on single indicators, self-reported indicators, and researcher intervention.

Single indicators Certifications are designed to reduce organisational security to a pass-fail test. Cybersecurity certifications were associated with positive stock market reactions [33, 93]. Yet no study demonstrates that certification is linked to better risk outcomes. Selection effects are pervasive as market incentives distort seemingly reliable security indicators. Firms look for auditors with the least stringent requirements

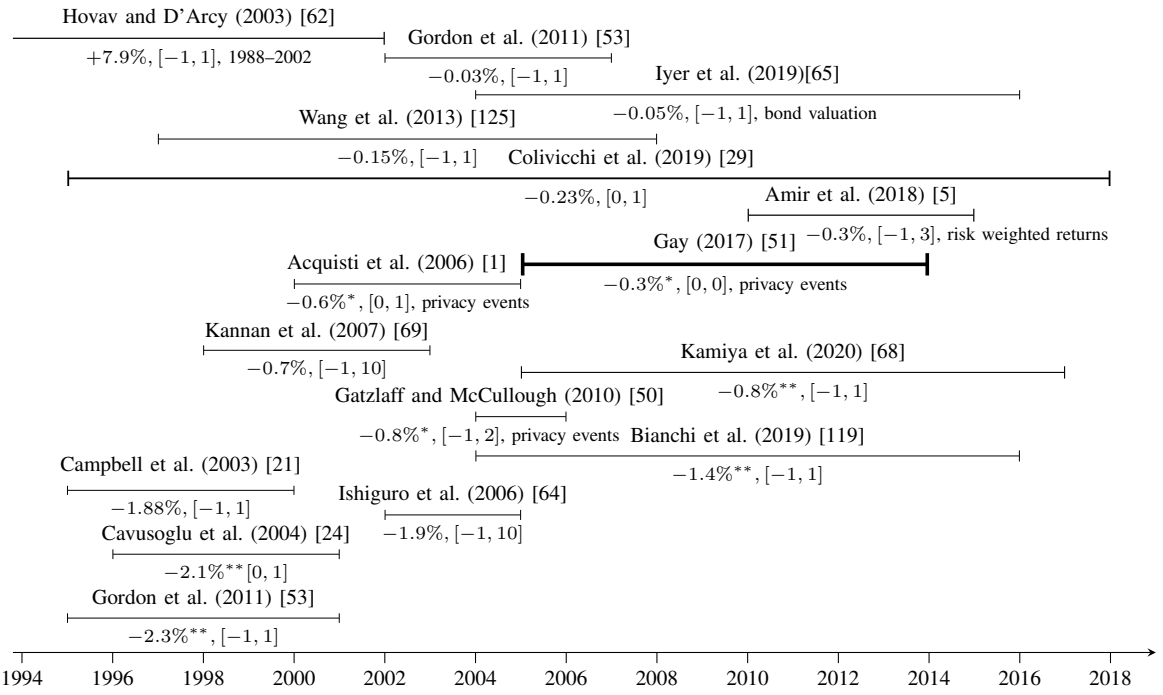


Fig. 4. Impact of security incident disclosure on firm stock market value. The effect is reported as cumulative abnormal returns (CAR), bar placement and thickness describe sample period and size, event window reported as [Days before, days after], statistical significance levels $^{*}\, p \leq 0.05,\ ^{**}\, p \leq 0.01$.

when certification is mandatory, which creates a race to the bottom [6, 74]. Optional certification is no better, websites certified by TRUSTe were shown to be more than twice as likely to be untrustworthy as uncertified sites [35]. More recently, Rahman et al. [94] showed that 86% of websites violated at least one of the requirements in the PCI-DSS standard they were certified to.

Looking elsewhere, one might expect security budgets to function as a crude indicator of security. We already identified how higher IT security budgets are associated with greater frequency of data breaches [105]. Security budget is likely tracking a hidden variable for risk exposure, such as organisation size since both breach frequency and size "scale with organisation size" [128, p. 11]. Even when controlling for firm characteristics in a logistic regression, Biancotti [12] find that defense expenditure in 2016 is positively correlated with the probability of experiencing an incident in 2017. Potential explanations for this result include: not controlling for threat, using noisy indicators of exposure, organisations spending resources inefficiently, or accounting tricks like re-assigning existing costs under the security budget.

Self-reported indicators Discovering one indicator of security with wide predictive power is unlikely, which motivates collecting multiple indicators. Egelman et al. [38] developed the Security Behavior Intentions Scale (SeBIS) in which a user's answers to 16 questions map onto four aspects of security behaviour with desirable psychometric properties. The sub-scales were shown to predict end-user behaviour [39, 103] but were not linked to harm outcomes. Sawaya et al. [103] show the scale does not "generalize" across cultures.

We are not aware of a similar scale for organisational security, though research from information systems uses questionnaire responses to explain security outcomes. In a seminal 1990 study, Straub [113] used a survey of $1\,211$ organisations to measure latent factors related to organisational commitment to security. The model showed organisational commitment to security correlates with better self-reported harm outcomes, such as the frequency and cost of incidents. Adding rival explanations like preventative measures did not improve the model, although the indicator—the number of security software packages in use—was weak. Organisations connecting networks to the Internet since this study enables direct measurements that avoid self-reported data [37, 81, 89, 101].

Researcher intervention The preceding studies simply observe security levels, whereas notification studies allow the researcher to randomly assign which subjects receive the intervention. Stock et al. [111] show that when notifications reach the website owner, the vulnerability has a 40% likelihood of remediation in the best case. The authors do not link the notification to harm or compromise outcomes, which is also true for studies notifying vulnerable name-servers [26], misaligned firewall policies [77], and HTTPS misconfiguration [132].

Notifying subjects who have already been compromised allows researchers to quantify the impact of a form of reactive security S_r. Vasek et al. [123] show that notifying hosting providers reduces time to clean-up malware URLs from 153 days to 101 days. Similarly, Li et al. show "direct communication with webmasters increases the likelihood of cleanup by over 50% and reduces infection lengths by at least 62%" [78]. The authors additionally control for indicators of exposure like

site language or popularity and show that less popular sites are associated with longer infection periods.

Summary *Single indicators* like security budget or certification should in theory summarise organisational security and hence explain security outcomes. In actuality, they are vulnerable to selection effects and manipulation. *Self-reported indicators* successfully explain security outcomes [39, 113] but are costly to collect. Studies collecting technical indicators [37, 81, 89, 101] can be more easily scaled. These studies are described in Section IV-D as they investigate the full causal model. Notification studies allow the researcher to control the security level and more confidently identify causal effects.

B. Measuring threat

The presence of active adversaries is a unique aspect of security research [60]. We identify approaches to controlling for threat level that vary across: time, target, and researcher intervention.

Time Empirical observations of malicious activity can be aggregated over time to track the changing threat level [70, Fig. 2–3]. Alternatively, expert sentiments can be tracked over time [52]. This provides longitudinal insights but the aggregate index does not speak to heterogeneity across organisations.

Targets Studying attackers in-the-wild can identify variation in targeting by threat actors. Tajalizadehkhoob et al. [114] analysed around 150k Zeus malware configuration files collected by a managed service provider. They show that just 175 of 6 500 financial institutions were targeted, of which larger banks were disproportionately represented. Similar studies identify factors affecting victimisation rates for DDoS amplification attacks [90] and phishing emails [106]. Simoiu et al. [106] find that user adoption of 2-factor authentication or a recovery mechanism is positively associated with phishing targeting. The authors warn against a causal interpretation in which criminals seek out victims with greater security, and instead suggest that victims who are more likely to be targeted are also more likely to employ security measures.

An even more fine-grained measurement involves detecting denial of service (DoS) attacks in the *back-scatter* of internet traffic. Moore et al. [87] used this approach to estimate the frequency, severity, and duration of a subset of DoS attacks. The method identifies which exact IP addresses were targeted.

Researcher intervention Simulating the attacker as part of an experiment provides complete control over the threat level of each subject in a laboratory setting. For example, Cai and Yap [20] study Android anti-virus (AV) app effectiveness using 200 known malware strains. In the causal diagram, this experimental design investigates how compromise C is determined by the installed app's preventative security S_p when exposed to the same malware samples T.

Ecological validity is questionable in this research design because the authors only used "sufficiently old" malware samples that were "detected by at least 40 out of 57 AVs" [20]. This means the study over-samples *detectable* malware, whereas rational attackers deliberately use *undetectable* malware. This can be addressed by collecting malware samples via honeypots [15, 48, 49]. The question remains as to whether failing to detect a malware sample translates into harm or even meaningful compromise.

Summary A unifying approach to controlling for threat is unlikely to be found. Although bigger targets tend to face a greater threat, many DoS attacks on home machines constitute "relatively large, severe attacks with rates in the thousands of packets" [87, p. 133]. Research designs should consider the specific form of cyber attack when deciding how to control for varying threats.

C. Measuring exposure

Constructing a measurement model for exposure seems intuitively simple because exposed assets are also exposed to measurement. Selecting the unit of analysis and the right number of variables are challenging.

Unit of analysis Stone et al. [112] tried to shame careless hosting providers by creating a ranking of the amount of persistent maliciousness. Hosting providers were associated with an autonomous system (AS), which functions as the technical unit of analysis. Tajalizadehkhoob et al. [115] argue this is a bad approach because some providers share ASs and others operate multiple ASs. The authors provide an alternative way forward by building a costly mapping from IP addresses to 45 358 hosting providers [117].

Variables The number of IP addresses associated with a hosting provider has been used to control for exposure [112, 133], but is this enough? Tajalizadehkhoob et al. [117] show it can explain 20% of the variance in phishing abuse associated with each hosting provider. This rises to 84% when three additional variables related to the size and business model of the hosting providers are added to the model. The majority (77%) of the remaining 16% of variance can be explained by including variables related to pricing and the ICT index of the hosting provider. This leads the authors to ask: if so much can be explained by exposure alone, what are we studying when we study abuse?

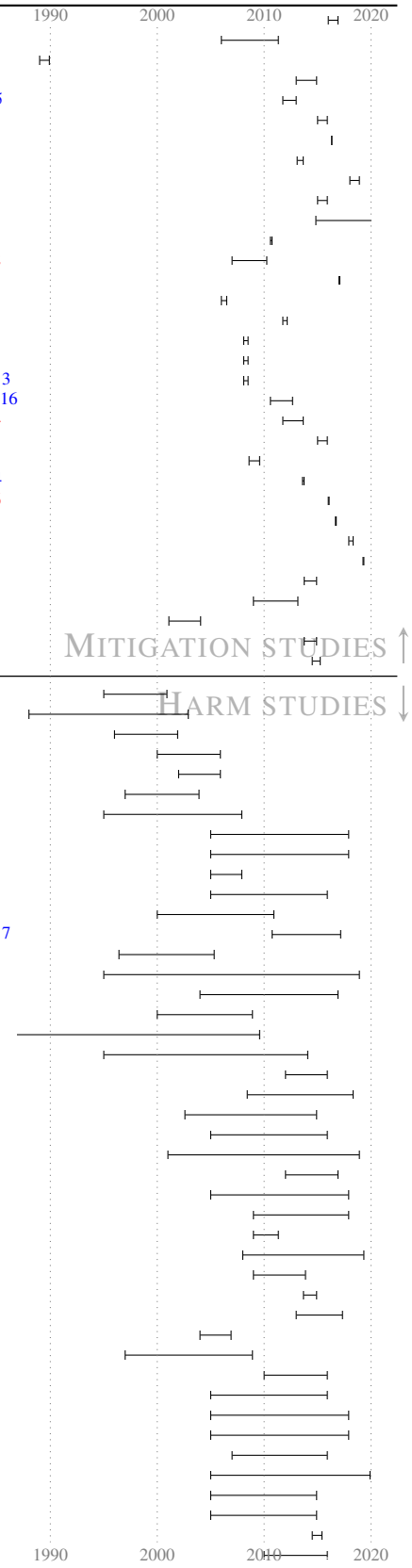
The explanatory power of exposure is further demonstrated by Soska and Christin [109]. They trained a classifier to predict whether a website will become malicious C, which achieves 66%/17% true/false positive rates. The features are all based on the website's content and traffic statistics, both of which represent indicators of exposure E_s. A powerful aspect of their research design is that features can be gathered after compromise has been observed thanks to "an archive of more than 391 billion web-pages saved over time"

Summary The explanatory power of exposure can be easily underestimated when omitting relevant variables or using the wrong unit of analysis. Going from one to four indicators of exposure in hosting providers led to a four fold increase in explanatory power [117]. Many of these variables were only available because the authors focused on hosting providers rather than relying on flawed proxies like measuring the number of IPs at the associated AS [115]. Beyond organisations, Canali et al. [22] show indicators of exposure like the amount or time of web-browsing impact compromise outcomes.

TABLE III

SYSTEMATISATION OF CYBER RISK QUANTIFICATION BY INVESTIGATED CONSTRUCT OF THE CAUSAL MODEL.

Survey of firms	Biancotti [12]	WEIS 2017	
Organisation incident	Sen and Borle [105]	JMIS 2015	
	Straub Jr [113]	ISR 1990	
	Sarabi et al. [101]	JCyS 2016	
Abuse study	Vasek et al. [122]	ITDSCM 2015	
	Edwards et al. [37]	arXiv 2019	
	Tajalizadehkhoob et al. [116]	CCS 2017	
	Zhang et al. [133]	NDSS 2014	
End-user	DeKoven et al. [34]	IMC 2019	
	Bilge et al. [14]	CCS 2017	
Threat index	Geer [52]	SnP 2010	
Cybercrime ecosystem	Levchenko et al. [75]	Oakland 2011	
	McCoy et al. [85]	USENIX 2012	
	Huang et al. [63]	Oakland 2018	
	Franklin et al. [47]	CCS 2007	
End-user	Lalonde Lévesque et al. [71]	CCS 2013	
AV effectiveness	Gashi et al. [48]	NCA 2009	
	Bishop et al. [15]	ISSRE 2011	
	Gashi et al. [49]	SafeComp 2013	
	Cai and Yap [20]	CODASPY 2016	
Abuse study	Soska and Christin [109]	USENIX 2014	
	Tajalizadehkhoob et al. [117]	TOIT 2018	
	Stone-Gross et al. [112]	ACSAC 2009	
End-user	Canali et al. [22]	AsiaCCS 2014	
Notification	Stock et al. [111]	USENIX 2016	
	Cetin et al. [26]	WEIS 2017	
	Zeng et al. [132]	WEIS 2019	
Compliance	Rahaman et al. [94]	CCS 2019	
Cybercrime ecosystem	Noroozian et al. [90]	RAID 2016	
	Tajalizadehkhoob et al. [114]	WEIS 2014	
IP backscatter	Moore et al. [87]	TOCS 2016	
Organisation incident	Liu et al. [81]	USENIX 2015	
Abuse study	Nagle et al. [89]	WEIS 2017	
Market reaction	Campbell et al. [21]	JCoS 2003	
	Hovav and D'Arcy [62]	RIMR 2003	
	Cavusoglu et al. [24]	JEC 2004	
	Acquisti et al. [1]	ICIS 2006	
	Ishiguro et al. [64]	WESSI 2006	
	Kannan et al. [69]	JEC 2007	
	Gordon et al. [53]	JCoS 2011	
	Iyer et al. [65]	FRL 2019	
	Kamiya et al. [68]	JFE 2020	
Data breach	Curtin and Ayres [30]	JIPLS 2008	
	Eling and Loperfido [40]	IME 2017	
Legal cases	Romanosky et al. [99]	JELS 2014	
	Ceross and Simpson [25]	SafeComp 2017	
Case study	Schroeder and Gibson [104]	TDSC 2010	
Market reaction	Colivicchi and Vignaroli [29]	JMS 2019	
	Tosun [119]	SSRN 2019	
Data breach	Maillart and Sornette [83]	EPJ 2010	
Operational loss	Biener et al. [13]	GPRI 2015	
	Eling and Wirfs [41]	EJOR 2019	
Survey of firms	Heitzenrater and Simpson [58]	JCyS 2015	
Insurance prices	Woods et al. [130]	WEIS 2019	
Market reaction	Park et al. [93]	JITS 2016	
	Deane et al. [33]	IFM 2016	
	Malliouris and Simpson [84]	WEIS 2019	
	Berkman et al. [11]	JAPP 2018	
Data breach	Xu et al. [131]	TIFS 2018	
Operational loss	Bouveret [18]	IMF 2018	
Case study	Franke et al. [46]	TOR 2014	
Manually compiled	Cyentia Institute [31]	Self 2019	
Bitcoin ledger	Spagnuolo et al. [110]	FC 2014	
	Liao et al. [79]	APWG 2016	
	Paquet-Clouston et al. [92]	JCyS 2019	
Market reaction	Gatzlaff and McCullough [50]	RIMR 2010	
	Wang et al. [125]	ISR 2013	
	Amir et al. [5]	RAS 2018	
Data breach	Edwards et al. [36]	JCyS 2016	
	Wheatley et al. [128]	GPRI 2019	
	Carfora et al. [23]	JOR 2019	
Data breach	Wheatley et al. [127]	EPJ 2016	
	Farkas et al. [43]	Self 2020	
Manually compiled	Romanosky [98]	JCyS 2016	
Market reaction	Gay [51]	JCyS 2017	
Notification	Li et al. [78]	WWW 2016	
Notification	Vasek et al. [123]	WISCS 2016	



Each study is classified according to our causal diagram (size due to space constraints) and then a category. We include the venue and year of publication in the fourth column. The fifth column describes the time period of the sample. Blue = Computer science, Red = Big four security conference, Orange = Inter-disciplinary CS, Green = Finance and Management, Grey = Miscellaneous. ◑ denotes the type of threat conditioned on compromise $T|C$.

D. Structural relationships

The previous subsections described different measurement models for latent factors. This section identifies research designs investigating the relationship between these latent factors. We point the readers back to previous descriptions of studies using latent models of security to explore structural relationships [113, 116] and turn to unidentified approaches, which break down into: between-subject, within-subject, and multiple indicator research designs.

Between-subject designs compare outcomes for subjects with differing levels of security. Edwards et al. [37] use this approach to study botnet infections across organisations with different security levels. They fit a linear model with variables like available network protocols and TLS configuration and certificate weaknesses. Training a separate model for each industry achieves the best balance between complexity and goodness of fit according to the chosen criteria. In some industries TLS certificate errors and misconfiguration were associated with *less* compromise [37]. The only consistent effect related to whether peer-to-peer file sharing was blocked.

At the level of web servers, Vasek et al. [122] use a case-control design to explore factors influencing the likelihood of a web server compromise. The authors discover evidence that running up to date software S_p "may actually put webservers at greater risk of being hacked" [122, p. 8]. This 'more security, more compromise' relationship likely resulted from sampling relatively many low-threat, low-security websites who only have low compromise rates because they are not targeted. Evidence in support of this is provided by restricting the sample to servers that have already been compromised, which is an indicator for high threat. After doing so, the authors observe a smaller fraction of updated websites (22.6%) are re-compromised than the fraction of sites that never update (33.5%). This suggests more security is associated with lower rates of compromise only in the high-risk population.

Within-subject designs track the same subject's security level over time using longitudinal data. Nagle et al. [89] fit a fixed-effect regression model using 33 million security events occurring at 480 enterprises collected by a security monitoring company. The number of open ports, which serves as an indicator of (the lack of) security management effort S_p, has a statistically significant effect on three of the four indicators of compromise C. The authors suggest the failure to establish an effect on the fourth indicator despite 33m observations results from the sparsity of observed malware infections. Such imbalances are common in samples collected from a population of firms *before* a breach has occurred—subsequent compromise and harm are (fortunately) rare exceptions.

Technical indicators A group of researchers used network scans to predict cyber risk outcomes in a cluster of related publications.The first study [133] identified a correlation between indicators of mismanaged networks S_p and malicious activities C emanating from the corresponding AS. The indicators of mismanagement are all normalised for exposure E_s. They also control for social and economic factors using a method designed to capture latent factors. The authors identify a statistically significant correlation between network mismanagement and network abuse. A metric aggregating all the individual symptoms had the strongest relationship [133, p. 8] highlighting the value of combining multiple noisy indicators.

A later publication [81] reformulates cyber risk forecasting network as a classification problem on an IP block. Blocks were labelled as breached using a set of a thousand incidents from various sources and then finding the victim's IP block. The data labelled as not breached is created by sampling from the remaining IP space, which is broken "into 2.9 million sets" [81, p. 1013]. The feature space includes indicators of security S_p and exposure E_s like mismanagement symptoms and the number of IP addresses. The authors argue "independence of the features from ground-truth data is maintained" [81, p. 1011] despite using the number of blacklisted IPs on the network as a feature. Arguably this is using an indicator of compromise to predict another indicator of compromise, but we interpret this as an indicator of increased threat level T.

A similar research design takes a sample of incidents, links these to a the victim's website domain, and labels these as breached domains [101]. The non-breached domains are sampled from "the largest publicly available directory of the Web" [101]. The studies achieve similar true and false positive rates (90%/10% [81] and 90%/11% [101]).

Both studies use an artificial case control by drawing labelled and unlabelled data from different populations; the cases labelled as breached are all drawn from the population of firms who have publicly reported a breach, mostly large corporations (see Section III). Whereas, the cases labelled as unbreached are drawn from a population of IP blocks or domain names that is not dominated by large corporations. The algorithms are likely detecting the difference between a large corporate network and a random web server, not the difference between large corporations according to the likelihood of breach. The fix, constructing control population from a *similar* population to the breached firms, is easier to state than to solve. A statistical twins approach was used to construct a homogeneous sample of hosting providers [117] but this must be done without ground-truth on the relevant dimensions of similarity.

End-user studies Although we have focused on organisational risk in this paper, research into individual devices and their users supports our narrative. For example, simple correlations reveal that end-users with 'computer expertise' [71] or that use the Tor browser [34] are associated with increased rate of compromise. The authors of both papers raise the possibility of confounding variables. Bilge et al. [14] include indicators of exposure in a model using random forests to predict device compromise and discover that applying security patches is the third most important feature (after two indicators of exposure).

Summary Applying between-subject research designs with single indicators of security lead to spurious results where more security is associated with more compromise [37, 122]. Adding control variables or using within-subject designs corrects the issue. The relative infrequency of compromise undermines statistical power leading to null results even with

33m observations of the security level [89].

Constructing latent factors for security provided more explanatory power than any single indicator in two studies [116, 133]. Although the learning representations are not explicitly latent, the success of applying random forests to predict incidents for organisations [81, 101] and machines [14] further supports our call to move away from explanations based on single indicators. Such models requires additional reporting to understand *how* security interventions affects the probability of compromise. Regression models became popular in the social sciences precisely because such effects are easily interpreted, even at the cost of predictive power.

E. Systematisation of cyber risk research

Table III summarises our systematisation. The first column visualises the relationships explored in the corresponding study, we see that the first block of diagrams, predominantly 'traditional' security research, have used relatively short sample windows. This can be contrasted with the harm studies (the second block) that explore longitudinal trends using databases aggregated by third parties.

The fourth column shows the diversity of venues for cyber risk research. Colour coding according to discipline shows cyber harm has mainly been explored in the finance (green) and interdisciplinary venues (orange). The top security conferences (red) focused on quantifying threat and security without considering structural relationships, putting aside a few recent exceptions. With the exception of Straub's seminal work in 1990 [113], research designs exploring multiple structural relationships have predominantly emerged in the last 6 years.

V. Discussion

We now return to each of our research questions.

A. RQ1: How much harm results from cyber incidents?

Data breaches in the US are the most studied incident because aggregated public reports are ripe for statistical analysis. Each study brings a new statistical approach leading to contradictory claims about the same dataset. This can be contrasted with experimental science in which each study collects additional data, applies similar statistical tests, and the field builds knowledge via meta-analyses. As a result, we have learned little about data breaches despite 10 years of analysis. We can at best agree that the number of records breached is heavy tailed, though this says little about financial cost [41].

Harm estimates are inconsistent across samples, reporting standards, and jurisdiction. The mean loss in a sample of global op losses extracted by text-mining [41] differs by an order of magnitude ($43m to $4.1m) when compared with a manually collected sample of public reports [98]. Estimates vary further across jurisdictions, only 0.1% of Italian firms suffered a loss greater than €200k in a 2016 survey [12]. This finding resulted from a stratified random sample collected by the Bank of Italy, which leads us to ask why so few independent statistical agencies employ their considerable expertise in collecting cybersecurity data?

Perhaps cyber risk is simply not *that* harmful [91]. Certainly when compared to the breaches reported in the media, typical breaches are smaller and less heavy tailed [43]. Cyber losses are less than fraud, bad debt, or retail theft [98], and cyber operational losses are both less on average and less heavy-tailed than non-cyber losses [13]. The lack of empirical support for the claim that cyber risk is exceptionally harmful casts doubt over the attention seeking assertions that pervade introductions to security papers and talks. These studies and our causal model are inadequate to provide evidence about systemic risk (alternatives are discussed in Section V-D).

B. RQ2: Which security interventions effectively reduce harm?

Our contribution is a framework to evaluate answers to this question. Actionable answers are unavailable based on current evidence. Simple statistical tests lead to spurious results like greater security budgets [12, 105], greater computer expertise [71] or updated software [122] being associated with greater frequency of compromise. The direction of such associations can be reversed by adding control variables [122].

Turning to the explanatory power of each latent factor, just using indicators of exposure can predict which websites will turn malicious [109] and explain most of the variance in abuse [117]. In contrast, indicators of security have little explanatory power alone. Liu et al. [81] re-train their model using each subset of the feature space alone and discover security mismanagement features "perform the worst" [81]. Yet when removing each from the full model, removing the subset of security indicators leads to the biggest decline in performance. This supports the fundamental intuition behind our causal model: security only explains harm outcomes when indicators of threat and exposure are added to the model.

Prioritising security interventions based on these studies is foolish. The best statistical models in terms of explanatory power measure security using multiple indicators [81, 113, 116]. Such approaches cannot isolate the effect of individual controls, let alone establish causality. Linking to policy, prescriptions in cybersecurity laws must be balanced against the lack of evidence on the effectiveness of specific prescriptions.

A promising development is notification studies [78, 122] in which security interventions can be randomly assigned outside a laboratory setting. Detected effects can reasonably be said to have been *caused* by the intervention. Adopting similar randomised control trial designs seems promising given their success in economics. With the power to randomly assign security interventions comes great ethical responsibility [88], which is compounded for researchers contemplating interventions related to threat actors [75, p. 9].

C. RQ3: Are these answers stable over time?

Harm studies have longer sample windows, approaching 20 years in some cases, than mitigation studies (see Table III). Data breaches are not increasing in frequency in general [36, 128] but they are increasing in both size and frequency if the sample is restricted to malicious breaches [128, 131]. The price of cyber insurance trended downwards from 2008–2018 [130],

although this has more to do with market dynamics than decreasing risk. In terms of shareholder value, the effect of breach disclosure seems to be decreasing over time. The timing of this shift (2001 [53] and 2005 [51]) is curiously close to when mandatory data breach notification laws came into effect. One explanation could be that post-2003 samples contain more inconsequential breaches that would not have been discovered beforehand and these drown out the effect of large breaches, which are shown to have the biggest impact on stock prices [5].

Sample windows in mitigation studies are too brief to learn about the effectiveness of security interventions over time (see Table III). For example, cyber incident forecasting performance holds when moving from a "one-month to a 12-month forecasting window" [81, p. 1019] but the researchers can test no further. This is partly explained by disciplinary norms around self-collected data and the availability of data brokers. Funding agencies might consider how to support institutionalised data collection and sharing as exemplified by the Cambridge Cybercrime Centre [27].

Balancing the admittedly limited evidence, there is little to suggest cyber harms are particularly unstable. This is consistent with similar studies of cybercrime in which global aggregate losses were in the same order of magnitude between 2012 [7] and 2019 [8] despite criminals innovating in methods.

D. Limitations

The causal model says little about other valuable approaches to security research, such as qualitative methods, that capture the subtleties of organisational security [10]. Within quantitative empirical research, limitations can be distilled into those of the model and more fundamental *unknowability*.

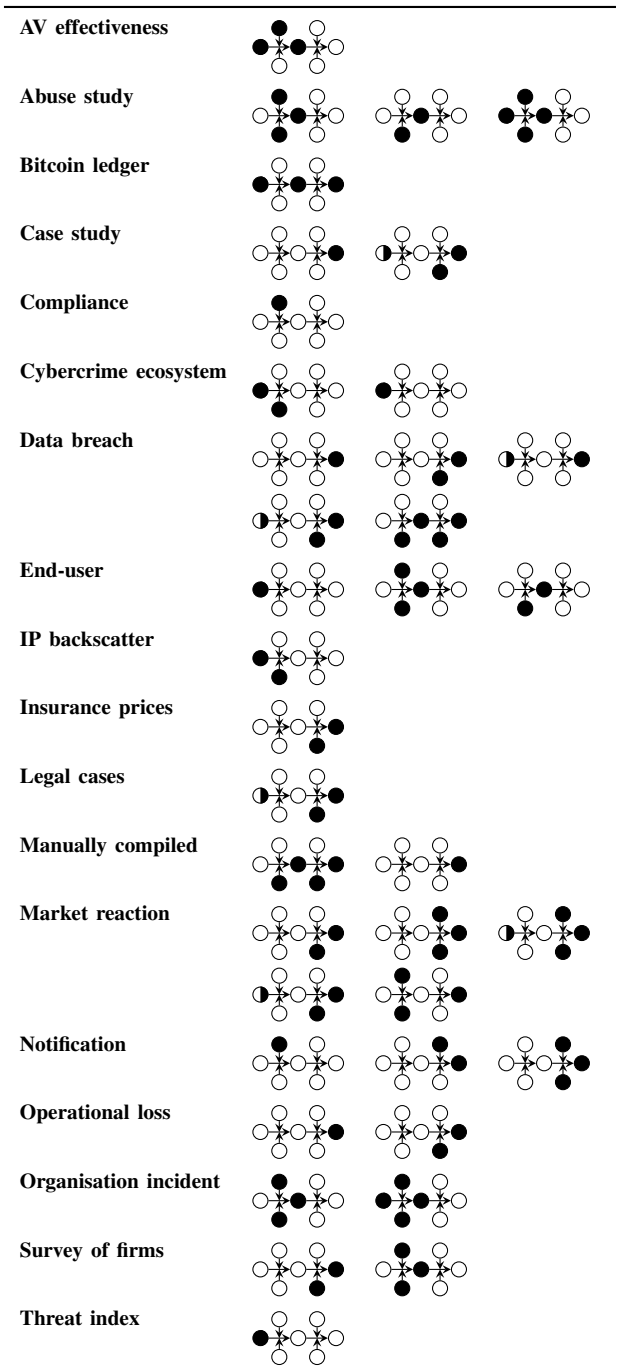
Model Limitations The causal model is intended for observational studies of cyber risk in organisations. This does not apply to research designs manipulating the security level as in notification studies. Law enforcement interventions cannot be studied by the model in its current form and must be treated as exogenous shocks impacting the threat level.

Our language often invokes linear relationships between variables, which does not reflect a naive belief that the world follows such models. Generalised linear models could be used to account for the non-linear distributions of harm identified in Section III.

Many authors opted for machine learning (ML) models instead of regressions. Although we suggest prediction rates are less interpretable than regression tables, the important properties of the causal model (e.g. variables for threat and exposure, multiple indicators) are present in ML studies.

Systemic cyber risk, however, requires a fundamentally different modelling approach because there are not enough observations for ML models *or* reduced form regressions. Knowledge about the loss generation process could be used to create structured models that require less data. For example, correlations in the attacks observed by Honeypots could parameterise correlations in risk models [16]. This topic is being considered by the finance community who consider how cyber risk poses a unique threat to financial stability [57, 126].

TABLE IV
SYSTEMATISATION OF CYBER RISK QUANTIFICATION BY CATEGORY.



Unknowability Creating knowledge about cyber harms and possible mitigation measures depends on available data. The size of a data-set is not everything as samples must also be representative of the broader population of interest. In terms of raw numbers, the surveyed studies analysed: 5 000 000 webpages [109]; 200 000 webservers [122]; 45 000 hosting providers [117]; 15 000 end-user devices [34]; 600 victims of malicious data breaches [131]; and 265 victims of data

breaches with financial cost [98]. Sub-components of complex systems like web page compromise are easier to study than emergent effects like firm-wide losses. This issue is particularly pressing for systemic risk, for which there are no empirical results. Detailed case-studies of the WannaCry and NotPetya incidents are an obvious starting point.

A second issue relates to social actors becoming aware of metrics. The signalling value of security certifications is eroded by market dynamics [6] and by selection effects [35]. Event window studies are undermined by strategically releasing positive news [51], withholding the most costly breaches [5], and by insider trading [80]. Such examples highlight Goodhart's law in which security metrics are optimised at the cost of actual security. A related problem is researcher measurements distorting other measurements, such as when network scans for research purposes are interpreted as an attacker probing for vulnerabilities [55].

Finally, data is political. Inferred causal relationships may not generalise beyond the population of study, such as across cultures [103], and this can lead to flawed (possibly harmful) recommendations. Harm estimates inevitably ignore certain victims and types of harm [76], such as individuals lacking the resources to quantify and communicate their harm. The 'cost of a data breach' skews towards direct costs to the firm as determined by accountants and not indirect harms suffered by victims of identity theft.

E. Future work

Throughout we have argued that the causal model (Figure 3) is the best statistical approach to quantifying cyber risk. However, this risks the naive takeaway that 'investigating more causal links is always better', which we do not endorse. Investigating the full causal model is an ambitious research design and often relies on prior work constructing measurement models for individual variables. Table IV is arguably more useful for funding agencies to distribute attention.

Our systemisation can both classify existing studies and show which studies are yet to be conducted. Table IV shows no data breach study has linked C or H to an indicator of security. There are reasons for this. Collecting data from sufficiently many breached firms before it is known which will be breached requires large samples, otherwise the sparsity of observed compromise undermines statistical tests [89]. A solution is to obtain explanatory variables after compromise has been observed. For example, Soska et al. [109] use the Internet Archive to collect historical website content.

More generally, future work should aim to quantify the relative effectiveness of different forms of security. Recent work identifying a statistical relationship between security measures and the prevalence of compromise marks progress since a 2009 critical review [124], but only a minority of these results speak to prioritisation. An example of the latter is evidence that hosting providers' security efforts "play a more significant role in fighting phishing abuse" [116, p. 13] than those of web masters. However, the authors warn against causally interpreting the effect of individual indicators.

VI. Conclusion

This paper systematises empirical research into cyber harm estimates and the effectiveness of security interventions. Inspired by structural equation models, we introduced a model explaining security outcomes using latent factors for security, exposure, and threat. The moderating role of security would ideally be measured using many reflexive indicators without necessarily identifying causality. Our survey of empirical cyber harm estimates finds little evidence that either the typical size or variance of cyber harm is particularly exceptional, but these studies do not consider the role of risk mitigation.

Applying the model to risk mitigation studies shows that threat level is often omitted. Indicators of exposure have good explanatory power in terms of cyber risk outcomes. Statistical tests that do not control for either factor lead to spurious results like increased security budgets leading to greater frequency of breach [105] or that applying software updates increases the likelihood of web-server compromise [122]. Studies that account for all attributes show security is a powerful determinant of cyber harm outcomes; indicators of network misconfiguration are the most important features in classifying whether an organisation will suffer a cyber incident [81].

Turning to the question of what risk science has to tell business leaders, firms should not underestimate the risk flowing from unnecessary exposure given its predictive power regarding multiple forms of compromise. In terms of risk mitigation, vendors promising simple solutions (single indicator explanations) should be ignored and security teams should be equipped with the resources to focus on the diversity of tasks that avert cyber harm. Policy makers' attention should be shifted away from typical losses, which are not exceptional, and towards systemic risk that we simply know nothing about.

References

[1] A. Acquisti, A. Friedman, and R. Telang. Is there a cost to privacy breaches? An event study. In *Proc. of the Int. Conf. on Information Systems*, pages 94–117, 2006.

[2] I. Agrafiotis, J. R. Nurse, M. Goldsmith, S. Creese, and D. Upton. A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity*, 4(1), 2018.

[3] L. Allodi and F. Massacci. Comparing vulnerability severity and exploits using case-control studies. *ACM Trans. on Inf. and System Security*, 17(1):1, 2014.

[4] L. Allodi, M. Corradin, and F. Massacci. Then and now: On the maturity of the cybercrime markets the lesson that black-hat marketeers learned. *IEEE Trans. on Emerging Topics in Computing*, 4(1):35–46, 2015.

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

[6] R. Anderson. Why information security is hard—An economic perspective. In *Proc. of the Computer Security Applications Conf.*, pages 358–365. IEEE, 2001.

[7] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. J. van Eeten, M. Levi, T. Moore, and S. Savage. Measuring the cost of cybercrime. In *Workshop on the Economics of Information Security*, 2012.

[8] R. Anderson, C. Barton, R. Böhme, R. Clayton, C. Ganán, T. Grasso, M. Levi, T. Moore, and M. Vasek. Measuring the changing cost of cybercrime. In *Workshop on the Economics of Information Security*, 2019.

[9] L. Axon, A. Erola, I. Agrafiotis, M. Goldsmith, and S. Creese. Analysing cyber-insurance claims to design harm-propagation trees. In *Int. Conf. on Cyber Situ. Aware., Data Analytics and Assess.* IEEE, 2019.

[10] A. Beautement, M. A. Sasse, and M. Wonham. The compliance budget: Managing security behaviour in organisations. In *Proc. of the New Security Paradigms Workshop*, pages 47–58. ACM, 2008.

[11] H. Berkman, J. Jona, G. Lee, and N. Soderstrom. Cybersecurity awareness and market valuations. *Journal of Accounting and Public Policy*, 37(6):508–526, 2018.

[12] C. Biancotti. The price of cyber (in)security: Evidence from the Italian private sector. In *Workshop on the Economics of Information Security*, 2018.

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

[14] L. Bilge, Y. Han, and M. Dell'Amico. Riskteller: Predicting the risk of cyber incidents. In *Proc. of the Conference on Computer and Communications Security*, pages 1299–1311. ACM, 2017.

[15] P. Bishop, R. Bloomfield, I. Gashi, and V. Stankovic. Diversity for security: A study with off-the-shelf antivirus engines. In *Int. Symp. on Software Reliability Engineering*, pages 11–19. IEEE, 2011.

[16] R. Böhme and G. Kataria. Models and measures for correlation in cyber-insurance. In *Workshop on the Economics of Information Security*, 2006.

[17] R. Böhme, S. Laube, and M. Riek. A fundamental approach to cyber risk analysis. *Variance*, 12(2):161–185, 2019.

[18] A. Bouveret. Estimation of losses due to cyber risk for financial institutions. *J. of Operational Risk*, 14(2), 2019.

[19] T. D. Breaux and D. L. Baumer. Legally "reasonable" security requirements: A 10-year FTC retrospective. *Computers & Security*, 30(4):178–193, 2011.

[20] Z. Cai and R. H. Yap. Inferring the detection logic and evaluating the effectiveness of Android anti-virus apps. In *Proc. of the Conf. on Data and Application Security and Privacy*, pages 172–182. ACM, 2016.

[21] K. Campbell, L. A. Gordon, M. P. Loeb, and L. Zhou. The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *J. of Computer Security*, 11(3):431–448, 2003.

[22] D. Canali, L. Bilge, and D. Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In *Proc. of the Symp. on Information, Computer and Communications Security (AsiaCCS)*, pages 171–182. ACM, 2014.

[23] M. Carfora, F. Martinelli, F. Mercaldo, and A. Orlando. Cyber risk management: An actuarial point of view. *Journal of Op. Risk*, 14(4):195–208, 2019.

[24] H. Cavusoglu, B. Mishra, and S. Raghunathan. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, 9(1):70–104, 2004.

[25] A. Ceross and A. Simpson. The use of data protection regulatory actions as a data source for privacy economics. In *Int. Conf. on Computer Safety, Reliability, and Security*, pages 350–360. Springer, 2017.

[26] O. Cetin, C. Ganán, M. Korczyński, and M. van Eeten. Make notifications great again: Learning how to notify in the age of large-scale vulnerability scanning. In *Workshop on the Econ. of Information Security*, 2017.

[27] R. Clayton. Cambridge Cloud Cybercrime Centre. *Available at: https://www.lightbluetouchpaper.org/2015/06/17/cambridge-cloud-cybercrime-centre/*, 2015. [Online; accessed 27-Aug-2020].

[28] A. Coburn, E. Leverett, and G. Woo. *Solving Cyber Risk: Protecting Your Company and Society*. Wiley, 2018.

[29] I. Colivicchi and R. Vignaroli. Forecasting the impact of information security breaches on stock market returns and VaR backtest. *Journal of Mathematical Finance*, 9 (3):402–454, 2019.

[30] M. Curtin and L. T. Ayres. Using science to combat data loss: Analyzing breaches by type and industry. *Journal of Law and Policy for the Inf. Society*, 4:569–601, 2008.

[31] Cyentia Institute. Ripples across the risk surface. *Available at: https://www.riskrecon.com/ripples-across-the-risk-surface*, 2019. [Online; accessed 2-Jan-2020].

[32] S. Dambra, L. Bilge, and D. Balzarotti. SoK: Cyber insurance—Technical challenges and a system security roadmap. In *Proc. of the Symp. on Security and Privacy*, pages 293–309. IEEE, 2020.

[33] J. K. Deane, D. M. Goldberg, T. R. Rakes, and L. P. Rees. The effect of information security certification announcements on the market value of the firm. *Inf. Technology and Management*, 20(3):107–121, 2019.

[34] L. F. DeKoven, A. Randall, A. Mirian, G. Akiwate, A. Blume, L. K. Saul, A. Schulman, G. M. Voelker, and S. Savage. Measuring security practices and how they impact security. In *Proc. of the Internet Measurement Conference*, pages 36–49. ACM, 2019.

[35] B. Edelman. Adverse selection in online "trust" certifications and search results. *Electronic Commerce Research and Applications*, 10(1):17–25, 2011.

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

[37] B. Edwards, J. Jacobs, and S. Forrest. Risky business: Assessing security with external measurements. *Available at arXiv: http://arxiv.org/abs/1904.11052*, 2019.

[38] S. Egelman and E. Peer. Scaling the security wall: Developing a security behavior intentions scale (SeBIS). In *Proc. of the Conf. on Human Factors in Computing Systems*, pages 2873–2882. ACM, 2015.

[39] S. Egelman, M. Harbach, and E. Peer. Behavior ever follows intention? A validation of the security behavior intentions scale (SeBIS). In *Proc. of the Conf. on Human Factors in Computing Systems*, pages 5257–5261. ACM, 2016.

[40] M. Eling and N. Loperfido. Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, 75:126–136, 2017.

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

[42] Eurobarometer Special 480. Europeans' attitudes towards internet security. *European Commission*, 2019.

[43] S. Farkas, O. Lopez, and M. Thomas. Cyber claim analysis through generalized Pareto regression trees with applications to insurance. *Available at HAL: https://hal.inria.fr/hal-02118080/*, 2020.

[44] D. Florêncio and C. Herley. Sex, lies and cyber-crime surveys. In *Work. on the Econ. of Inf. Security*, 2011.

[45] U. Franke. The cyber insurance market in Sweden. *Computers & Security*, 68:130–144, 2017.

[46] U. Franke, H. Holm, and J. König. The distribution of time to recovery of enterprise IT services. *IEEE Trans. on Reliability*, 63(4):858–867, 2014.

[47] J. Franklin, A. Perrig, V. Paxson, and S. Savage. An inquiry into the nature and causes of the wealth of internet miscreants. In *Proc. of the Conference on Computer and Communications Security*, pages 375–388. ACM, 2007.

[48] I. Gashi, V. Stankovic, C. Leita, and O. Thonnard. An experimental study of diversity with off-the-shelf antivirus engines. In *Proc. of the Int. Symp. on Network Computing and Applications*, pages 4–11. IEEE, 2009.

[49] I. Gashi, B. Sobesto, V. Stankovic, and M. Cukier. Does malware detection improve with diverse antivirus products? An empirical study. In *Proc. of the Int. Conf. on Computer Safety, Reliability, and Security*, pages 94–105. Springer, 2013.

[50] K. M. Gatzlaff and K. A. McCullough. The effect of data breaches on shareholder wealth. *Risk Management and Insurance Review*, 13(1):61–83, 2010.

[51] S. Gay. Strategic news bundling and privacy breach disclosures. *J. of Cybersecurity*, 3(2):91–108, 2017.

[52] D. E. Geer. An index of cybersecurity. *IEEE Security & Privacy*, 8(6):96–95, 2010.

[53] L. A. Gordon, M. P. Loeb, and L. Zhou. The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security*, 19(1):33–56, 2011.

[54] A. Greenberg. *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers*. Doubleday, 2019.

[55] D. Guido. The DBIR's 'Forest' of Exploit Signatures. *Available at: https://blog.trailofbits.com/2016/05/05/the-dbirs-forest-of-exploit-signatures/*, 2016. [Online; accessed 27-Aug-2020].

[56] E. Harrel. Victims of identity theft. *Bureau of Justice Statistics*, 2016. [Online; accessed 2-Apr-2020].

[57] J. Healey, P. Mosser, K. Rosen, and A. Tache. The future of financial stability and cyber risk. *The Brookings Institution Cybersecurity Project, October*, 2018.

[58] C. D. Heitzenrater and A. C. Simpson. Policy, statistics and questions: Reflections on UK cyber security disclosures. *Journal of Cybersecurity*, 2(1):43–56, 2016.

[59] C. Herley and D. Florêncio. Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy. In *Workshop on the Economics of Information Security*, 2009.

[60] C. Herley and P. C. Van Oorschot. SoK: Science, security and the elusive goal of security as a scientific pursuit. In *Proc. of the Symp. on Security and Privacy*, pages 99–120. IEEE, 2017.

[61] J. Hernandez-Castro and E. Boiten. Cybercrime prevalence and impact in the UK. *Computer Fraud & Security*, 2014(2):5–8, 2014.

[62] A. Hovav and J. D'Arcy. The impact of Denial-of-Service attack announcements on the market value of firms. *Risk Management and Insurance Review*, 6(2): 97–121, 2003.

[63] D. Y. Huang, M. M. Aliapoulios, V. G. Li, L. Invernizzi, E. Bursztein, K. McRoberts, J. Levin, K. Levchenko, A. C. Snoeren, and D. McCoy. Tracking ransomware end-to-end. In *Proc. of the Symp. on Security and Privacy*, pages 618–631. IEEE, 2018.

[64] M. Ishiguro, H. Tanaka, K. Matsuura, and I. Murase. The effect of information security incidents on corporate values in the Japanese stock market. In *Workshop on the Econ. of Securing the Information Infrastructure*, 2006.

[65] S. R. Iyer, B. J. Simkins, and H. Wang. Cyberattacks and impact on bond valuation. *Finance Research Letters*, 33(101215), 2019.

[66] J. Jacobs. Analyzing Ponemon cost of data breach. *Available at https://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/*, 2014. [Online; accessed 27-August-2020].

[67] E. Jardine. Mind the denominator: Towards a more effective measurement system for cybersecurity. *Journal of Cyber Policy*, 3(1):116–139, 2018.

[68] S. Kamiya, J.-K. Kang, J. Kim, A. Milidonis, and R. M. Stulz. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, In press, 2020.

[69] K. Kannan, J. Rees, and S. Sridhar. Market reactions to information security breach announcements: An empirical analysis. *Int. J. of Electronic Commerce*, 12(1):

69–91, 2007.

[70] P. Kotzias, L. Bilge, P.-A. Vervier, and J. Caballero. Mind your own business: A longitudinal study of threats and vulnerabilities in enterprises. In *Network and Distributed System Security Symp.* Internet Society, 2019.

[71] F. Lalonde Lévesque, J. Nsiempba, J. M. Fernandez, S. Chiasson, and A. Somayaji. A clinical study of risk factors related to malware infections. In *Proc. of the Conference on Computer & Communications Security*, pages 97–108. ACM, 2013.

[72] S. Laube and R. Böhme. Strategic aspects of cyber risk information sharing. *ACM Computing Surveys*, 50(5): 1–36, 2017.

[73] N. Leontiadis, T. Moore, and N. Christin. Pick your poison: Pricing and inventories at unlicensed online pharmacies. In *Proc. of the Conference on Electronic Commerce*, pages 621–638. ACM, 2013.

[74] J. Lerner and J. Tirole. A model of forum shopping. *American Economic Review*, 96(4):1091–1113, 2006.

[75] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage. Click trajectories: End-to-end analysis of the spam value chain. In *Proc. of the Symp. on Sec. and Priv.*, pages 431–446. IEEE, 2011.

[76] K. Levy and B. Schneier. Privacy threats in intimate relationships. *J. of Cybersecurity*, 6(1), 2020.

[77] F. Li, Z. Durumeric, J. Czyz, M. Karami, M. Bailey, D. McCoy, S. Savage, and V. Paxson. You've got vulnerability: Exploring effective vulnerability notifications. In *Proc. of the USENIX Security Symp.*, pages 1033–1050. USENIX, 2016.

[78] F. Li, G. Ho, E. Kuan, Y. Niu, L. Ballard, K. Thomas, E. Bursztein, and V. Paxson. Remedying web hijacking: Notification effectiveness and webmaster comprehension. In *Proc. of the Int. Conf. on World Wide Web*, pages 1009–1019. ACM, 2016.

[79] K. Liao, Z. Zhao, A. Doupé, and G.-J. Ahn. Behind closed doors: Measurement and analysis of Cryptolocker ransoms in Bitcoin. In *APWG Symp. on Electronic Crime Research*, pages 1–13. IEEE, 2016.

[80] Z. Lin, T. R. Sapp, J. R. Ulmer, and R. Parsa. Insider trading ahead of cyber breach announcements. *Journal of Financial Markets*, 50(100527), 2019.

[81] Y. Liu, A. Sarabi, J. Zhang, P. Naghizadeh, M. Karir, M. Bailey, and M. Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *Proc. of the USENIX Sec. Symp.*, pages 1009–1024. USENIX, 2015.

[82] P. Maass and M. Rajagopalan. Does cybercrime really cost $1 trillion? *Propublica*, 2012.

[83] T. Maillart and D. Sornette. Heavy-tailed distribution of cyber-risks. *The European Physical Journal B*, 75 (3):357–364, 2010.

[84] D. D. Malliouris and A. C. Simpson. The stock market impact of information security investments: The case of security standards. In *Workshop on the Economics of Information Security*, 2019.

[85] D. McCoy, A. Pitsillidis, J. Grant, N. Weaver, C. Kreibich, B. Krebs, G. Voelker, S. Savage, and K. Levchenko. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *Proc. of the USENIX Security Symp.*, pages 1–16. USENIX, 2012.

[86] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools (Revised Edition)*. Princeton University Press, 2015.

[87] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage. Inferring internet Denial-of-Service activity. *ACM Trans. on Comp. Systems*, 24(2):115–139, 2006.

[88] T. Moore and R. Clayton. Ethical dilemmas in take-down research. In *Int. Conf. on Financial Cryptography and Data Security*, pages 154–168. Springer, 2011.

[89] F. Nagle, S. Ransbotham, and G. Westerman. The effects of security management on security events. In *Workshop on the Econ. of Information Security*, 2017.

[90] A. Noroozian, M. Korczyński, C. H. Gañan, D. Makita, K. Yoshioka, and M. van Eeten. Who gets the boot? Analyzing victimization by DDoS-as-a-service. In *Int. Symp. on Research in Attacks, Intrusions, and Defenses*, pages 368–389. Springer, 2016.

[91] A. Odlyzko. Cybersecurity is not very important. *Ubiquity*, 2019:1–23, 2019.

[92] M. Paquet-Clouston, B. Haslhofer, and B. Dupont. Ransomware payments in the Bitcoin ecosystem. *Journal of Cybersecurity*, 5(1), 2019.

[93] J. Park, W.-J. Jung, and B. Kim. The effect of information security certification announcement on the market value of firms. *Journal of Information Technology Services*, 15(3):51–69, 2016.

[94] S. Rahaman, G. Wang, and D. Yao. Security certification in payment card industry: Testbeds, measurements, and recommendations. In *Proc. of the Conf. on Computer and Communications Security*, pages 481–498. ACM, 2019.

[95] V. Richardson, M. W. Watson, and R. E. Smith. Much ado about nothing: The (lack of) economic impact of data privacy breaches. *J. of Information Systems*, 2019.

[96] M. Riek and R. Böhme. The costs of consumer-facing cybercrime: an empirical exploration of measurement issues and estimates. *J. of Cybersecurity*, 4(1), 2018.

[97] J. Riekmann and M. Kraus. Tatort Internet: Kriminalität verursacht Bürgern Schäden in Milliardenhöhe. *DIW-Wochenbericht*, 82(12):295–301, 2015.

[98] S. Romanosky. Examining the costs and causes of cyber incidents. *J. of Cybersecurity*, 2(2):121–135, 2016.

[99] S. Romanosky, D. Hoffman, and A. Acquisti. Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, 11(1):74–104, 2014.

[100] S. Romanosky, A. Kuehn, L. Ablon, and T. Jones. Content analysis of cyber insurance policies: How do carriers price cyber risk? *J. of Cybersecurity*, 5(1), 2019.

[101] A. Sarabi, P. Naghizadeh, Y. Liu, and M. Liu. Risky

business: Fine-grained data breach prediction using business profiles. *J. of Cybersecurity*, 2(1):15–28, 2016.

[102] SAS Institute Inc. Oprisk global data: A comprehensive database of operational loss information, 2015. [Online; accessed 27-April-2020].

[103] Y. Sawaya, M. Sharif, N. Christin, A. Kubota, A. Nakarai, and A. Yamada. Self-confidence trumps knowledge: A cross-cultural study of security behavior. In *Proc. of the Conf. on Human Factors in Computing Systems*, pages 2202–2214. ACM, 2017.

[104] B. Schroeder and G. Gibson. A large-scale study of failures in high-performance computing systems. *IEEE Trans. on Dependable and Secure Computing*, 7(4): 337–350, 2010.

[105] R. Sen and S. Borle. Estimating the contextual risk of data breach: An empirical approach. *Journal of Management Information Systems*, 32(2):314–341, 2015.

[106] C. Simoiu, A. Zand, K. Thomas, and E. Bursztein. Who is targeted by email-based phishing and malware? measuring factors that differentiate risk. In *Proc. of the Internet Measure. Conf.*, page 567–576. ACM, 2020.

[107] G. Simpson and T. Moore. Empirical analysis of losses from business-email compromise. In *APWG Symp. on Electronic Crime Research*, 2020.

[108] D. Sornette and G. Ouillon. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *The European Physical J. Special Topics*, 205(1):1–26, 2012.

[109] K. Soska and N. Christin. Automatically detecting vulnerable websites before they turn malicious. In *Proc. of the USENIX Security Symp.*, pages 625–640. USENIX, 2014.

[110] M. Spagnuolo, F. Maggi, and S. Zanero. Bitiodine: Extracting intelligence from the Bitcoin network. In *Int. Conf. on Financial Cryptography and Data Security*, pages 457–468. Springer, 2014.

[111] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes. Hey, you have a problem: On the feasibility of large-scale web vulnerability notification. In *Proc. of the USENIX Security Symp.*, pages 1015–1032. USENIX, 2016.

[112] B. Stone-Gross, C. Kruegel, K. Almeroth, A. Moser, and E. Kirda. Fire: Finding rogue networks. In *Computer Security Applications Conf.*, pages 231–240. IEEE, 2009.

[113] D. W. Straub Jr. Effective IS security: An empirical study. *Inf. Systems Research*, 1(3):255–276, 1990.

[114] S. Tajalizadehkhoob, H. Asghari, C. Gañán, and M. J. van Eeten. Why them? Extracting intelligence about target selection from Zeus financial malware. In *Workshop on the Economics of Inf. Security*, 2014.

[115] S. Tajalizadehkhoob, M. Korczyński, A. Noroozian, C. Ganán, and M. van Eeten. Apples, oranges and hosting providers: Heterogeneity and security in the hosting market. In *Network Operations and Management Symp.*, pages 289–297. IEEE, 2016.

[116] S. Tajalizadehkhoob, T. Van Goethem, M. Korczyński, A. Noroozian, R. Böhme, T. Moore, W. Joosen, and M. van Eeten. Herding vulnerable cats: A statistical approach to disentangle joint responsibility for web security in shared hosting. In *Proc. of the Conf. on Computer and Communications Security*, pages 553–567. ACM, 2017.

[117] S. Tajalizadehkhoob, R. Böhme, C. Ganán, M. Korczyński, and M. van Eeten. Rotten apples or bad harvest? What we are measuring when we are measuring abuse. *ACM Trans. on Internet Tech.*, 18(4):1–25, 2018.

[118] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson. Trafficking fraudulent accounts: The role of the underground market in Twitter spam and abuse. In *Proc. of the USENIX Security Symp.*, pages 195–210. USENIX, 2013.

[119] O. K. Tosun. Cyber attacks and stock market activity. *Avail. at SSRN: https://ssrn.com/abstract=3190454*, 2019.

[120] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart. Stealing machine learning models via prediction APIs. In *Proc. of the USENIX Security Symp.*, pages 601–618. USENIX, 2016.

[121] UK Department for Business, Innovation and Skills. Information security breaches survey, 2015. [Online; accessed 27-March-2020].

[122] M. Vasek, J. Wadleigh, and T. Moore. Hacking is not random: A case-control study of webserver-compromise risk. *IEEE Trans. on Dependable and Secure Computing*, 13(2):206–219, 2015.

[123] M. Vasek, M. Weeden, and T. Moore. Measuring the impact of sharing abuse data with web hosting providers. In *Proc. of the Workshop on Inf. Sharing and Collaborative Security*, pages 71–80. ACM, 2016.

[124] V. Verendel. Quantified security is a weak hypothesis: A critical survey of results and assumptions. In *Proc. of the Workshop on New Security Paradigms*, pages 37–50. ACM, 2009.

[125] T. Wang, K. N. Kannan, and J. R. Ulmer. The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, 24(2):201–218, 2013.

[126] P. Warren, K. Kaivanto, and D. Prince. Could a cyber attack cause a systemic impact in the financial sector? *Bank of England Quarterly Bulletin*, 58(4):21–30, 2018.

[127] S. Wheatley, T. Maillart, and D. Sornette. The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*, 89(1):7, 2016.

[128] S. Wheatley, A. Hofmann, and D. Sornette. Addressing insurance of data breach cyber risks in the catastrophe framework. *Geneva Papers on Risk and Ins. -Issues and Prac.* , In press, 2020.

[129] D. W. Woods and T. Moore. Does insurance have a future in governing cybersecurity? *IEEE Security & Privacy*, 18(1):21–27, 2020.

[130] D. W. Woods, T. Moore, and A. C. Simpson. The county fair cyber loss distribution: Drawing inference

from insurance prices. In *Workshop on the Economics of Information Security*, 2019.

[131] M. Xu, K. M. Schweitzer, R. M. Bateman, and S. Xu. Modeling and predicting cyber hacking breaches. *IEEE Trans. on Inf. Forensics and Security*, 13(11):2856–2871, 2018.

[132] E. Zeng, F. Li, E. Stark, A. P. Felt, and P. Tabriz. Fixing HTTPS misconfigurations at scale: An experiment with security notifications. In *Workshop on the Economics of Information Security*, 2019.

[133] J. Zhang, Z. Durumeric, M. Bailey, M. Liu, and M. Karir. On the mismanagement and maliciousness of networks. In *Network and Distributed System Security Symp.* Internet Society, 2014.

APPENDIX

Throughout model parameters are estimated to minimise squared residuals. Regressing losses L on the security level S in the artificial data in Table V reveals that

$$L \approx 0.11 + 0.32^{*} S \tag{1}$$

A positive coefficient for S means increasing security level is associated with an increase in loss. Further, the coefficient is the slope of the solid blue line in Figure 1. The * means the coefficient is statistically significant at the $p \leq 0.05$ level. Security is associated with greater losses because the high-threat population spend more on security *and* also suffer greater losses. However, controlling for threat by re-estimating the model in the high-threat population ($T = 1$), we see that

$$L \approx 0.63 - 0.59^{*} S \tag{2}$$

Security now has a negative coefficient, as we would expect, and is statistically significant. Conversely, security has no significant effect in the low-threat population ($T = 0$).

$$L \approx 0.03 + 0.05\ S \tag{3}$$

The coefficients in (2) and (3) correspond to the slopes of the dotted red and dashed green line in Figure 1 respectively. It is evident that the threat level causes losses, and security moderates the effect of threat on losses.

This effect can be captured using an interaction between S and T so that

$$L \approx 0.03 + 0.6^{***}\ T + 0.05\ S - 0.6^{*}\ T \times S \tag{4}$$

In Model 4, threat is strongly significant ($p \leq 0.001$) and positively associated with losses. Increasing security in the overall population leads to a small *increase* in losses, although this relationship is not statistically significant. Whereas, the same increase in the high-threat population leads to a ten-fold larger *decrease* in loss ($-0.56 + 0.05 = -0.50$ vs $+0.05$), and this is statistically significant. The interaction term $T \times S$ captures the intuition that security *moderates* the relationship between threat and losses.

We can evaluate the fit of each model using the coefficient of variation R^2, which describes the proportion of the variance explained by the model. Model 1 only explains 10% of the variance, whereas Model 2 and Model 4 can explain 21% and 58% respectively. Model 3 explains nothing. R^2 values are adjusted for the varying number of parameters. It is impossible to explain all of the variance ($R^2 = 1$) with a linear model, given that the underlying data generation process is non-linear.

Structural equation modeling generalises this logic to allow for multiple moderating factors. Moreover, it considers explicit measurement models which estimate these structural relationships between latent constructs interpolated from multiple noisy indicators.

TABLE V
ARTIFICIAL DATA USED FOR THE EXAMPLE IN FIG. 1

Low-threat ($T = 0$)		High-threat ($T = 1$)	
Security S	Loss L	Security S	Loss L
0.06	0.00	0.37	0.56
0.08	0.00	0.57	0.54
0.65	0.00	0.35	0.73
0.01	0.20	0.73	0.38
0.29	0.50	0.38	0.25
0.13	0.00	0.47	0.40
0.59	0.00	0.47	0.46
0.00	0.00	0.57	0.30
0.37	0.00	0.90	0.00
0.23	0.00	0.70	0.23
0.00	0.00	0.35	0.68
0.15	0.10	0.21	0.57
0.00	0.00	0.43	0.32
0.08	0.00	0.20	0.56
0.21	0.00	0.29	0.40
0.01	0.00	0.40	0.13
0.11	0.00	0.29	0.27
0.00	0.00	0.58	0.43
0.01	0.00	0.48	0.05
0.24	0.00	0.46	0.21

TABLE VI
TECHNICAL INDICATORS [116] CORRESPONDING TO I_x IN FIGURE 3

	Technical indicator
I_{C_1}	# domains in phishing blacklist
I_{C_2}	# domains in malware blacklist
I_{E_1}	# IPs on shared hosting
I_{E_2}	# domains on shared hosting
I_{S_1}	HTTP server version
I_{S_2}	SSL version
I_{S_3}	Admin panel version
I_{S_4}	PHP version
I_{S_5}	OpenSSH version
I_{S_6}	CMS version
I_{S_7}	HttpOnlyCookie
I_{S_8}	X-Frame-Options
I_{S_9}	X-Content-Type-Options
$I_{S_{10}}$	Mixed-content inclusions
$I_{S_{11}}$	Secure cookie
$I_{S_{12}}$	Content-Security-Policy
$I_{S_{13}}$	HTTP Strict-Transport-Security
$I_{S_{14}}$	SSL-stripping vulnerable form
$I_{S_{15}}$	Browser XSS protection