Construction and Analysis of a Large-Scale Firm-Level Cybersecurity Posture Dataset

Completed Research Full Paper

Abulfaz Hajizada
The University of Tulsa
abh7687@utulsa.edu

Matthew Adams
The University of Tulsa
mla5406@utulsa.edu

Tyler Moore
The University of Tulsa
tyler-moore@utulsa.edu

Abstract

Cybersecurity risk presents a significant and growing challenge for firms. Understanding better how firms are defending themselves can help answer important questions about which controls are more effective and whether firms are investing enough in their defenses. Unfortunately, data on firm-level cybersecurity investments have been difficult for researchers to obtain at large scale. This paper describes a method for constructing firm-level cybersecurity posture metrics by aggregating a selection of data on security products tracked in the SWZD Company Information database. Our exploratory analysis demonstrates this dataset's value in enriching cybersecurity research, offering novel perspectives that could shape sector-specific best practices and enable empirical evaluation of security controls.

Keywords

Cybersecurity metrics, security controls, firm-level security posture, security infrastructure analysis

Introduction and Background

Cybersecurity is essential for the functioning of both society and the economy. Cybersecurity incidents have been shown to negatively impact firms' reputation, financial stability, and customer retention, as well as a multitude of other consequences (Li et al. 2023; Schlackl et al. 2022; Woods and Boehme 2021). The average cost of a data breach has increased by 140% over 14 years (Schlackl et al. 2022). Meanwhile, cyber-attacks continue to proliferate despite increased spending on defensive countermeasures by firms.

A key research question is whether and how firm investment in cybersecurity reduces risk. To obtain an answer, we need better data on multiple fronts. We need reliable measures of how firms invest in cybersecurity and the frequency and impact of attacks. Woods and Boehme (2021) describe a more complete causal model for quantifying cyber risk. Their study identified substantial gaps in terms of datasets available to measure firm-level cybersecurity. This paper contributes a measure of firm-level cybersecurity posture that leverages a promising dataset on IT expenditures at organizations of all sizes.

Interest in how firms manage cybersecurity investments dates back decades. In a seminal paper, Gordon and Loeb (2002) construct a model for information security investment. Built around the assumption of decreasing marginal returns to security investment, the model identified an optimal budget based on exogenously assigned breach probability functions. The highly influential model demonstrated that there is a rational limit to firm spending on cybersecurity. Notably, the model abstracted away consideration of specific controls by considering only a financial representation of the security budget.

Around the same time, Anderson (2001) applied concepts from microeconomics to explain how firm-level investment can fall short of socially optimal. In particular, Anderson (2001) explained how externalities can cause underinvestment (e.g., by attacks that harm others more than the firm who chooses whether to spend on precautions) and how information asymmetries can lead to misallocated spending (e.g., because evaluating the effectiveness of security controls is often hard for buyers). These efforts spawned two decades of research in security economics (see Moore (2024) for a review of key results).

Interviews with chief information security officers (CISOs) reveal that most firms follow process-based guidelines in the form of frameworks such as National Institute of Standards and Technology (NIST) Cybersecurity Framework when making cybersecurity investment decisions (Moore et al. 2015). While the frameworks can help decide which widgets to buy, they do not help in setting an appropriate budget. The reality is that cybersecurity is often viewed as a cost center, and tight budgets can lead to adopting only a few of the recommended controls. Consequently, it is important to study differences in firm-level cybersecurity investment for several reasons. First, it can help determine if cybersecurity investment is lagging at individual firms, in specific sectors, or for the economy as whole. Second, differences in firm-level investment might help explain why some firms are successfully attacked and others are not. Today, we have limited evidence to link firm investment to changes in risk of attack.

There are multiple existing approaches to measure firm-level cybersecurity posture. Over the past decade, several researchers and companies have developed “outside-in” measurements of firm security. For example, Liu et al. (2015) gathered network scan data of misconfigurations and malicious traffic on firm-controlled networks and used machine learning to predict subsequent breaches. Others counted open ports on firm network (Nagle et al. 2017), while Sarabi et al. (2016) generated risk predictions using similar data at the sectoral level. Meanwhile, several firms now generate outside-in measurements to estimate firm-level cyber risk, such as Security Scorecard and Bitsight. An advantage is that they can generate wide coverage of firms based on direct observation. The disadvantage is that they do not observe firm cybersecurity controls directly. Inside-out approaches, by contrast, leverage internal data from enterprises. Few academic researchers have gained access to such data, but it is being collected primarily by the cyber insurance industry. The advantages and disadvantages mirror outside-in approaches: better insight into controls but much more limited coverage of firms that is typically not shared.

Surveys provide another option. Insurers provide a detailed questionnaire to prospective customers during underwriting for cyber insurance, including detailed questions about the security controls in place (Nurse et al. 2020). Typically, these data are utilized only by insurers and not available to researchers. In one case, an insurer used questionnaire responses to study which controls may be more effective in avoiding later claims (Marsh McClennan 2023). Gandal et al. (2023) leverage a survey conducted by the Israeli Bureau of Statistics that included a question on which controls firms adopt. In addition to possible disconnects between stated and observed practices, data availability is often an issue with surveys.

Public regulatory filings offer another potential source of information on firm cybersecurity posture. Hajizada and Moore (2023) identify disclosed cyber incidents in SEC 10-K filings, noting a tendency toward underreporting. Most work has focused on estimating perceived cybersecurity threats to companies (Cheong et al. 2021; Gao et al. 2020; Li et al. 2018, 2023). While information on firm cybersecurity posture might be gleaned from these filings, no existing research has done so to date. Perhaps closest to the mark is the work by Florackis et al., (2023) which develops a measure of cybersecurity sophistication and exposure based on the occurrence of cyber-related terms in risk disclosures. Notably, some research has found a dilution in the informational value of these disclosures, raising concerns about generic disclosures that do not accurately reflect material risks (Li et al. 2018).

Some research has investigated the question of how firm posture affects cyber risk. Gandal et al. (2023) finds robust evidence that adopting security controls before the occurrence of a cyber incident reduces the likelihood of a data breach. Proactive security investment is seen to reduce security failure rates as well as being more cost effective than reactive investments (Kwon and Johnson 2014). Kwon and Johnson (2014) examine the cost effectiveness of proactive security investments in the healthcare sector and find that those who proactively invest can handle evolving cybersecurity threats as well as experience smaller breaches and lower breach notification costs than those who rely on reactive investment.

This paper builds on prior work by describing a way to construct a new empirical measurement of firm cybersecurity posture using an existing large-scale dataset. We then provide initial exploratory analysis that demonstrates its value in capturing differences in firm-level investment using characteristics such as firm size and sector. It is hoped that this dataset could be leveraged by the research community as a valuable input to studies comparing firm-level security investments to the risks of cyber incidents.

Dataset for Firm-Level Cybersecurity Posture

SWZD Company Intelligence Database

Spiceworks publishes a database called SWZD Company Intelligence, which reports on software installations for firms of all sizes (Forman 2005). The database's commercial purpose is to gather marketing data for firms selling software, providing business support, offering outsourcing services, and to help identify potential corporate clients (Johnston and Zhang 2018). Previously known as the Aberdeen or Harte Hanks Computer Intelligence Technology Database (CITDB), this database (hereafter referred to as simply "Spiceworks") has also been utilized in IT research for many years (Forman 2005).

In previous studies, the Spiceworks database has been used as the main resource for gathering enterprise-level firm data, particularly in the context of information technology (IT) and cybersecurity related software (Arora and Forman 2007; Huang et al. 2022; Li et al. 2023). Li et al. (2023) established a bidirectional relationship between cyber investment and data breaches, showing that security investment requires both countermeasure and threat awareness to be effective (Li et al. 2023). The authors used IT budget information to estimate cybersecurity spending levels. Similarly, Xue et al. (2021) used the database to study the association of commitment to IT infrastructure and engagement in real earnings management. Forman (2005) used it to obtain 3-digit Standard Industrial Classification (SIC) codes to control for industry effects. Other researchers use various categories to create firm-wide measures, such as IT capital stock based on market value of IT hardware and reported IT labor expenses (Huang et al. 2022) or estimating the total number of unique operating systems used by firms (Nagle 2019). Compared to prior work, we utilize security-specific fields in the database for the first time, as described next.

Methodology for Constructing Firm-Level Cybersecurity Posture Metric

We obtained three yearly snapshots of the SWZD Company Intelligence database, taken at the end of 2018, 2020, and 2022. For this paper, we focus our analysis on the 2022 data. In total, there are 19,323,203 records in 2022 covering 537,220 companies. In addition to firm demographics, the database reports on 10,278 total IT products. 3 of 24 product series are specific to security: Security Devices, IT Security Information, and Network Security Monitoring. 506 distinct products are included in this series.

These products are further categorized into 11 distinct product categories, as shown in Table 1. Since product offerings evolve over time, and the database has been tracking usage since 1995, some drift is inevitable. Hence, we manually reviewed all product categorizations to determine if the product are still assigned to the correct category. We found that 392 products matched their categories correctly. 53 products matched one of the other 11 categories better, so we updated its category. For example, Proofpoint was labeled as Anti-virus, when it is more accurately classified as Email Security. Because many security products have been re-branded and expanded functionality beyond anti-virus, we had to frequently update products in this category. Additionally, 10 products categorized as Security Information and Event Management (SIEM) were in fact Certificate Authorities selling SSL Certificates. We excluded these products entirely from our analysis, along with an additional 17 products that did not offer security capabilities (e.g., SAS Analytics, Quest Change Auditor, and IBM SPSS Modeler). Finally, we labeled 34 products in a new category called "Other Security". These products clearly offer security functionality but do not map onto any of the 11 existing product categories. This included application security testing products such as Veracode and Checkmarx and Nice Actimize, an anti-money laundering software.

A wide range of security products are tracked, ranging from advanced vulnerability assessment tools to DDoS mitigation solutions. The third column in Table 1 shows how many products belong to each of the 12 categories. Anti-virus products have the highest number of products (138), followed by SIEM (85). This suggests a high level of competition in these categories. The table also shows the total number of installations for products in each category, as well as the number and percentage of customers with at least one product installed. Product popularity varies greatly. Near the top of the list is Proofpoint, whose email security product is installed at over 100K sites. Overall, nearly half of all firms installed at least one email security product, which speaks to the importance of protecting against phishing and social engineering attacks. Identity and access management products are also popular, with over 600K installs covering nearly half of firms (Amazon's AWS Identity Management offering is the market leader). By contrast, some products are very unpopular, with a handful recording only a single installation.

Category	NIST Phase	# Products	# Installs	# Customers w/ Installs	% Customers w/ Installs
Anti-virus	Detect	138	203,746	79,783	17.6%
Email Security	Detect	7	332,869	211,236	46.7%
Endpoint Security	Detect	42	337,542	100,380	22.2%
IT Asset Management	Identify	8	20,142	16,523	3.7%
Other Security	Multiple	33	317,420	105,972	23.4%
Advanced Threat Protection	Protect	18	173,106	52,713	11.6%
Data Loss Prevention	Protect	4	27,842	22,717	5.0%
Firewall Software	Protect	35	265,249	121,983	26.9%
Identity Access Management Software	Protect	63	671,793	207,296	45.8%
VPN	Protect	17	199,027	138,134	30.5%
Disaster Recovery	Recover	27	88,742	52,341	11.6%
SIEM	Respond	85	509,651	120,486	26.6%

Table 1. Product Categorization and Installation Metrics.

Final Metric: NIST Cybersecurity Framework Coverage

This dataset helps quantify the cybersecurity priorities of organizations. Nonetheless, we are less interested in particular products than how firms adopt a suite of controls to manage cybersecurity risk. The Spiceworks categories, while informative, do not on their own represent a necessary or sufficient set of controls that firms must adopt. Consequently, we turn to the five "core" categories put forward in the NIST Cybersecurity Framework: Identify, Protect, Defend, Respond, and Recover (National Institute of Standards and Technology 2024). The framework drills down to subcategories with greater specificity, but a key point is that any comprehensive cybersecurity program should include controls covering all five categories. Not all controls are technical or match to security products. Nonetheless, there are technology-based controls for all five categories. Hence, we mapped the five NIST categories to the 12 Spiceworks categories as shown in the second column of Table 1. The number of NIST categories covered by a firm provides a reliable metric of how comprehensive its suite of cybersecurity controls is.

Dataset Limitations

While large and comprehensive, the Spiceworks database is by no means perfect. Prior researchers have reported an overrepresentation of firms with heavy investments in IT and larger firms within their sample space (Arora and Forman 2007; Nagle 2019). This overrepresentation is a common occurrence as larger organizations are more likely to suffer 'substantial' losses and, resultingly, more likely to be reported by media sources (Woods and Boehme 2021). Forman (2005) links the limitations on the Spiceworks database to its survey methodology, as not all companies respond to the survey annually, and those that do not respond may likely never respond, potentially resulting in nonresponse bias, additionally, due to the nature of the survey used, the data set underestimates specific categories due to firms' reporting habits (Nagle 2019). Nagle (2019) argued that the total amount of nonpecuniary OSS used by firms was underreported. Forman (2005) argues that the database is not a complete census of hardware and software as companies are likely to only report upon the technologies, they feel are most essential. Nonetheless, Forman (2005) states that the database is found to be "fairly representative along major dimensions such as size, industry composition, and geographic composition", and has consistently utilized the database in subsequent years (Arora and Forman 2007; Huang et al. 2022).

For our purposes, we recognize that there may be some security controls that are not tracked, particularly those that are not technology-based. We do expect that the figures presented here may underreport the totals due to data incompleteness. Additionally, for a dataset this large, it is inevitable that some reported data may be outdated. Nonetheless, the level of precision offered by this dataset is unparalleled and can form the basis of valuable metrics of cybersecurity posture.

Exploratory Analysis

We now explore the data to demonstrate its value in measuring differences in firm cybersecurity posture.

Overall Firm-Level Investment

We first examine the firm-level investment overall. The rightmost column in Table 1 already reveals that there is wide variation in firm uptake of different controls. Email Security (47%) and Identity and Access Management (46%) are the most widely adopted control categories, while IT Asset Management and Data Loss Prevention see very low uptake (4 and 5% respectively).

The median number of specific security products adopted is 2, though the average is 7, which indicates that some firms adopt many more products. Figure 1 (left) plots the complementary cumulative distribution function (CDF) of the number of technological controls used by a firm. We can see from the solid black line that, overall, most firms report only a few controls, with a small minority of around 5% adopting over 30. Figure 1 (right) plots a complementary CDF of the number of Spiceworks product categories covered. Over half of firms adopt controls in just one category, with 20% adopting in all 5.

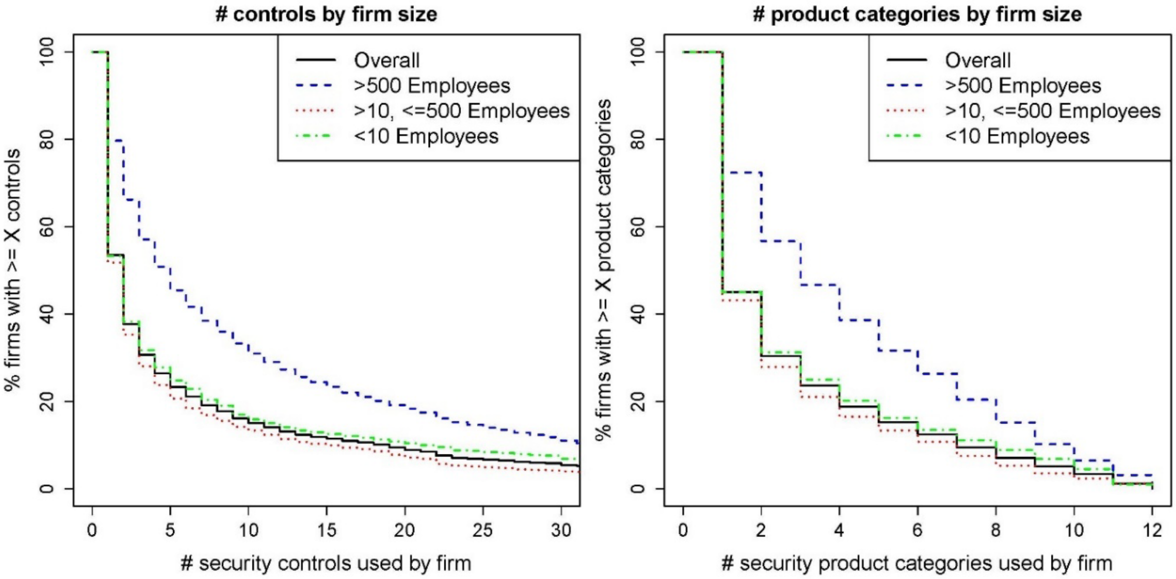

Figure 1. Distribution of controls adopted (left) and Spiceworks categories covered (right).

We now report on coverage of the NIST Cybersecurity Framework (CSF) categories. Overall, Protect and Detect have the best coverage, with 68% and 60% of firms respectively adopting at least one control in the category. Respond and recover receive less coverage, with Identify coming in last at 3.7%. This is concerning, given that the controls in Identify such as asset identification are fundamental inputs to controls in other categories.

Another way of viewing the data is to count how many NIST CSF categories are covered by firms. Nearly everyone covers at least one category, but it falls off quickly after that. 39% of firms have controls covering two categories, 21% cover 3. Just 2.4% of firms have controls covering all five categories, which is astounding since NIST recommends that all organizations adopt a baseline set of controls in all five categories. Of course, this only covers the controls tracked by Spiceworks. It is likely that some firms have greater coverage than reported here, but the overall picture remains troubling.

Impact of Firm Size on Cybersecurity Posture

We now cut the data by firm size. We divide firms into three categories. Small companies have less than ten employees. In the US these companies are classified as small office/home office (SOHO) (SBA, *Size standards* 2024). Medium companies have between 10 and 500 employees. These companies are considered Small and Midsize Enterprises (SMEs) based on their specific sector, on either their number of employees or earnings (SBA, *Size standards* 2024). For simplicity, we consider only employees, not earnings, when labeling a firm as an SME. Large companies are those with 500 or more employees. In total, we have 18,313 large firms, 255,637 SMEs and 178,454 SOHOs.

We anticipate that large companies would adopt more cybersecurity controls, and this is what we have found. Figure 1 also provides CDFs split by size. As expected, firms adopt more controls and cover more Spiceworks categories. Medium and small firms have fewer controls, but there is one surprise. SMEs actually adopt slightly fewer controls than SOHOs.

We observe a similar pattern when examining NIST CSF control coverage in Table 2. Large organizations outperform in all five NIST categories. The share of large firms that include Protect and Detect controls rose significantly, but jumps in the other categories are even bigger. The share of firms adopting Respond controls is nearly twice as large for large forms (47% vs 27%). A similar effect can be observed for Recover (24% vs 12%). Mid-size and very small firms lag considerably, with mid-size again faring slightly worse.

	Overall (%)	Firm Size (%)			Critical Infrastructure (%)	
		>500 Employees	<500 Employees ≥10	<10 Employees	No	Yes
Identify	3.7%	9.7%	3.4%	3.4%	2.7%	3.9%
Protect	67.6%	81%	65.2%	69.5%	66%	67.9%
Detect	60.3%	72.6%	60.5%	58.9%	55.3%	61.4%
Respond	26.6%	46.9%	24.0%	28.4%	23.1%	27.3%
Recover	11.6%	24.3%	9.8%	12.8%	8.4%	12.2%
Uncategorized	23.4%	39.6%	21.7%	24.2%	20.2%	24.1%
≥ 1 NIST Category	97.9%	99.1%	97.9%	97.7%	96.6%	98.1%
≥ 2 NIST Categories	38.6%	65.7%	36.4%	39.0%	34.2%	39.5%
≥ 3 NIST Categories	21.4%	42.6%	18.8%	23.0%	16.4%	22.4%
≥ 4 NIST Categories	9.5%	20.7%	7.6%	11.0%	6.7%	10.0%
5 NIST Categories	2.4%	6.4%	2.2%	2.3%	1.6%	2.6%

Table 2. Controls adopted by NIST CSF categories, broken down by firm size and sector.

When we examine the number of CSF categories covered, large firms again do best, with 43% covering at least 3 categories. Even so, just 6.4% of large firms report controls covering all five categories. This reflects the fact that nearly everyone does a poor job of adopting Identify controls.

How Does Cybersecurity Posture Differ for Critical Infrastructure Firms?

Since Presidential Decision Directive 63 was issued in 1998 (United States 1998), special focus has been placed on securing critical infrastructures (CIs) against attack. One might expect firms in these sectors would invest more in cybersecurity than others. To test that hypothesis, we leverage a mapping of the 16 critical infrastructure sectors to a list of 18 essential critical infrastructure workers (ECIW), along with their corresponding North American Industry Classification System (NAICS) codes (Billock et al. 2022). Based on the 6-digit NAICS codes, organizations were split into critical and non-critical sectors for comparative analysis. Perhaps surprisingly, a significant majority of firms (376,612 of 452,406) fall under one of the CI sectors. Overall, CI organizations surpass their counterparts in all five categories, but not by much. The right-most column in Table 2 breaks down coverage of NIST CSF controls for all CI firms together. These firms do only a bit better overall (not surprising with over 80% of firms classified as CI).

We can also look at the performance of specific sectors, as shown in Table 3. The second column displays the average number of controls adopted by the organizations in each sector and the third column reports

CI Sector	Avg # NIST	≥3 NIST (%)	Identify (%)	Protect (%)	Detect (%)	Respond (%)	Recover (%)	Other (%)
Chemical	1.63	17.6%	1.5%	65.7%	63.3%	22.8%	9.2%	14.8%
Commercial Facilities	1.86	26.4%	6%	71.7%	64%	30.1%	13.8%	32.9%
Communications & IT	1.71	20.9%	4.5%	74.5%	47.8%	32.7%	11.8%	25.7%
Critical Manufacturing	1.52	14.4%	3.5%	60%	62.3%	17.4%	8.7%	16.3%
Defense Industrial Base	3.17	62.8%	44.3%	84.2%	75.9%	70%	42.3%	66.4%
Education	1.51	14.4%	2%	70.6%	48.9%	21.2%	7.9%	17.3%
Energy	2.18	35.9%	8.5%	71.1%	73.4%	38.9%	25.6%	33.2%
Financial Services	2.07	34.5%	5.1%	72%	69.6%	39.2%	21.1%	33.1%
Food & Agriculture	1.91	31.3%	3.1%	74.1%	65.6%	40.3%	7.8%	36.6%
Hazardous Materials	1.46	10.8%	4.8%	61.4%	63.9%	8.4%	7.2%	27.7%
Healthcare/ Public Health	1.68	20.5%	3.5%	63.5%	62.8%	25.5%	12.4%	17.7%
Hygiene Products & Services	1.59	16.6%	2.3%	67.7%	54.2%	23.9%	10.5%	20.2%
Law Enforcement	1.44	11.9%	3.2%	55.3%	63.4%	15.5%	6.9%	12.1%
Other Government Operations	1.46	12.9%	3.1%	66.7%	52%	16.4%	7.3%	14.1%
Public Infrastructure	1.49	14.7%	1.5%	58.2%	65.3%	15.1%	8%	14.5%
Housing Services	1.44	12%	2.3%	59%	62.1%	13.8%	7.2%	12.9%
Transportation & Logistics	1.67	21.8%	2.5%	69.6%	56.6%	27%	11.7%	26.3%
Water and Wastewater	1.53	14.3%	4.2%	58.3%	63.6%	17.1%	9.7%	14.4%

Table 3. Control adoption variation by critical infrastructure sector.

what percentage of organizations adopted controls in three or more CSF categories. The rest of the columns represent five major NIST cybersecurity framework controls and other security products.

Several interesting trends are present in the table when examined in detail. The Defense Industrial Base sector shows a higher average number of adopted NIST categories across the board. This further confirms the prioritization of comprehensive cybersecurity practices by each organization in that sector. There are also some other notable sectors that display this type of responsibility such as Food and Agriculture, Energy, and Financial Services sectors. They seem to prioritize Protect, Detect, and Respond categories more compared to others. This may reflect the overarching regulatory pressures on these sectors.

A potential gap in cybersecurity readiness can also be seen by looking closely at sectors like Law Enforcement, Public Safety, and Other First Responders and Residential/Shelter Facilities and Housing and Real Estate, and Related Services. These all have below-average control adoption. The Law Enforcement, Public Safety, and Other First Responders sector has the worst performance overall in two of the five NIST categories, Protect and Recover, as well as in average number of controls adopted.

Identifying "Best" and "Worst" Performing Sectors

Another way to study cybersecurity posture across sectors is to investigate them all and identify those that perform best and worst. NAICS is the federal standard code for classifying businesses in sectors. The

NAICS code is hierarchical in nature, beginning with a two-digit code to register a high-level sector, and as more numbers are added the more niche the sector becomes. An example being that NAICS code 92 is “Public Administration”, NAICS code 926 is “Administration of Economic Programs,” and NAICS code 92612 is “Regulation and Administration of Transportation Programs.” Here we focus on the NAICS three-digit code since it strikes the best balance between general and specialized sectors.

NAICS3	≥3 NIST (%)	Avg #NIST	Name
238	6.5%	1.32	Specialty Trade Contractors
624	7.5%	1.37	Social Assistance
712	7.5%	1.36	Museums, Historical Sites, and Similar Institutions
321	9.2%	1.43	Wood Product Manufacturing
337	9.3%	1.44	Furniture and Related Product Manufacturing
713	10%	1.40	Amusement, Gambling, and Recreation Industries
484	11%	1.52	Truck Transportation
332	11.1%	1.46	Fabricated Metal Product Manufacturing
925	11.9%	1.51	Administration of Housing Programs, Urban Planning, and Community Development
611	12.0%	1.46	Educational Services
	29%	*1.96*	*Average*
518	48.9%	2.62	Data Processing, Hosting, and Related Services
721	49%	2.58	Accommodation
722	49.4%	2.36	Food Services and Drinking Places
524	50.2%	2.57	Insurance Carriers and Related Activities
923	54.7%	2.70	Administration of Human Resource Programs
517	55.9%	2.77	Telecommunications
452	57.1%	2.68	General Merchandise Stores
447	57.4%	2.88	Gasoline Stations
926	60.7%	2.83	Administration of Economic Programs
928	67.2%	3.31	National Security and International Affairs

Table 4. Best and worst performing sectors.

Table 4 presents the 10 highest and lowest performing sectors regarding the percentage of companies that have implemented at least 3 NIST controls. For this comparison we focused on NAICS3 codes that have at least 100 firms with over 100 employees, excluding these smaller firms from this analysis. Of the total 99 NAICS3 codes, 79 met these criteria. The data is presented as follows: the first column lists the NAICS 3-digit code for each sector, the second column indicates the percentage within that sector that have implemented at least 3 NIST controls, and the third column shows the average number of NIST controls adopted by entities within that sector. We include as benchmark three NIST CSF categories, as only 29% of firms within these sectors manage to achieve this on average.

Table 4 summarizes this data showing several notable cases. NAICS 238, "Specialty Trade Contractors," exhibits the lowest percentage within its sector, with only 6.53% achieving three NIST categories. Meanwhile, NAICS 611, "Educational Services," has the tenth lowest percentage. Educational services handle a substantial volume of confidential information concerning current and former students, in addition to data on parents or guardians. A similar observation applies to NAICS 713, "Amusement, Gambling, and Recreation Industries," due to its association with extensive guest information. Researchers could potentially use these results to identify vulnerabilities within these sectors. The remaining sectors among the bottom ten exhibit characteristics which are to be expected, due to their relatively low levels of risk from cybercrime. For instance, NAICS 337, "Furniture and Related Product Manufacturing," and NAICS 332, "Fabricated Metal Product Manufacturing," while still vulnerable to cybercrimes, do not pose a substantial risk of affecting external stakeholders adversely.

This contrasts with sectors that have higher levels of associated risk, particularly those handling private and confidential information. An illustrative example being NAICS 928, "National Security and International Affairs," which has the highest percentage at 67.2%. This sector's results are reflective of the confidential data that it handles, as it concerns both high level staff and operational affairs. Other sectors in this context include NAICS 926, "Administration of Economic Programs"; NAICS 923, "Administration of Human Resource Programs"; and NAICS 518, "Data Processing, Hosting, and Related Services." These sectors are linked by the high levels of sensitive information that they possess, highlighting the imperative for robust cybersecurity controls. Some surprising entries among the best make sense on closer inspection. Gasoline Stations, General Merchandise Stores and Food Services and Drinking Places fare well. What these sectors have in common is the handling of payment card information which makes them subject to PCI/DSS Compliance rules. The stratification among sectors suggests that sectoral metrics and benchmarks based on this data could be valuable moving forward.

Conclusion

Measuring firm-level cybersecurity investment is important for improving the effectiveness of those investments, as well as tracking societal progress in increasing cybersecurity capabilities. Yet researchers have long lacked robust data sources. In this paper, we have presented a way to distill selected data entries from an important private data source on IT spending, SWZD Company Intelligence, to construct firm-level measurements of cybersecurity posture based on adherence to the NIST Cybersecurity Framework. The exploratory analysis presented demonstrates the strong potential of this dataset for use by researchers. We have shown how that posture tends by firm size and sector, and that overall, most firms could improve their performance. We are optimistic that this measure can be utilized to answer further questions relating firm cybersecurity investment to control effectiveness and risk reduction.

Acknowledgements

We gratefully acknowledge support from the US National Science Foundation Award No. 2147505.

REFERENCES

Anderson, R. 2001. "Why Information Security Is Hard – An Economic Perspective," *Seventeenth Annual Computer Security Applications Conference (ACSAC)*, pp. 358-365. (https://doi.org/10.1109/ACSAC.2001.991552).

Arora, A., and Forman, C. 2007. "Proximity and Information Technology Outsourcing: How Local Are IT Services Markets?," *Journal of Management Information Systems* (24:2), pp. 73–102. (https://doi.org/10.2753/MIS0742-1222240204).

Billock, R. M., Haring Sweeney, M., Steege, A. L., Michaels, R., and Luckhaupt, S. E. 2022. "Identifying Essential Critical Infrastructure Workers during the COVID-19 Pandemic Using Standardized Industry Codes," *American Journal of Industrial Medicine* (65:7), pp. 548–555. (https://doi.org/10.1002/ajim.23361).

Cheong, A., Yoon, K., Cho, S., and No, W. G. 2021. "Classifying the Contents of Cybersecurity Risk Disclosure through Textual Analysis and Factor Analysis," *Journal of Information Systems* (35:2), pp. 179–194. (https://doi.org/10.2308/ISYS-2020-031).

Florackis, C., Weber, M., Michaely, R., and Louca, C. 2023. "Cybersecurity Risk," *The Review of Financial Studies* (36:1), pp. 351–407. (https://doi.org/https://doi.org/10.1093/rfs/hhac024).

Forman, C. 2005. "The Corporate Digital Divide: Determinants of Internet Adoption," *Management Science* (51:4), pp. 641–654. (https://doi.org/10.1287/mnsc.1040.0343).

Gandal, N., Moore, T., Riordan, M., and Barnir, N. 2023. "Empirically Evaluating the Effect of Security Precautions on Cyber Incidents," *Computers and Security* (133). (https://doi.org/10.1016/j.cose.2023.103380).

Gao, L., Calderon, T. G., and Tang, F. 2020. "Public Companies' Cybersecurity Risk Disclosures," *International Journal of Accounting Information Systems* (38), Elsevier Inc., p. 100468. (https://doi.org/10.1016/j.accinf.2020.100468).

Gordon, L. A., and Loeb, M. P. 2002. "The Economics of Information Security Investment," *ACM Transactions on Information and System Security* (5:4), pp. 438–457. (https://doi.org/10.1145/581271.581274).

Hajizada, A., and Moore, T. 2023. "On Gaps in Enterprise Cyber Attack Reporting," *Proceedings - 8th IEEE European Symposium on Security and Privacy Workshops*, Euro S&PW 2023, pp. 227–231. (https://doi.org/10.1109/EuroSPW59978.2023.00030).

Huang, P., Ceccagnoli, M., Forman, C., and Wu, D. J. 2022. "IT Knowledge Spillovers, Absorptive Capacity, and Productivity: Evidence from Enterprise Software," *Information Systems Research* (33:3), pp. 908–934. (https://doi.org/10.1287/isre.2021.1091).

Johnston, J. A., and Zhang, J. H. 2018. "Information Technology Investment and the Timeliness of Financial Reports," *Journal of Emerging Technologies in Accounting* (15:1), pp. 77–101. (https://doi.org/10.2308/jeta-52066).

Kwon, J., and Johnson, M. E. 2014. "Proactive Versus Reactive Security Investments in the Healthcare Sector," *Paper Knowledge. Toward a Media History of Documents* (38:2), pp. 451–472.

Li, H., No, W. G., and Wang, T. 2018. "SEC's Cybersecurity Disclosure Guidance and Disclosed Cybersecurity Risk Factors," I*nternational Journal of Accounting Information Systems* (30:June), pp. 40–55. (https://doi.org/10.1016/j.accinf.2018.06.003).

Li, W., Leung, A., and Yue, W. 2023. "Where Is IT in Information Security? The Interrelationship among IT Investment, Security Awareness, and Data Breaches," *MIS Quarterly* (47:1), pp. 317–342. (https://doi.org/10.25300/misq/2022/15713).

Liu, T., Sun, Y., Liu, Y., Gui, Y., Zhao, Y., Wang, D., and Shen, C. 2015. "Abnormal Traffic-Indexed State Estimation : A Cyber – Physical Fusion Approach for Smart Grid Attack Detection," *Future Generation Computer Systems* (49), Elsevier B.V., pp. 94–103. (https://doi.org/10.1016/j.future.2014.10.002).

Marsh McLennan. 2023. Using data to prioritize cybersecurity investments. Marsh McLennan.

Moore, T. 2024. "Security Economics Knowledge Guide," *The Cyber Security Body Of Knowledge* (1), pp. 0–24.

Moore, T., Dynes, S., and Chang, F. R. 2015. "Identifying How Firms Manage Cybersecurity Investment," *15th Workshop on the Economics of Information Security (WEIS)*, pp. 1–32.

Nagle, F. 2019. "Open Source Software and Firm Productivity," *Management Science* (65:3), pp. 1191–1215. (https://doi.org/10.1287/mnsc.2017.2977).

Nagle, F., Ransbotham, S., and Westerman, G. 2017. "The Effects of Security Management on Security Events," *Workshop on the Economics of Information Security (WEIS)*, pp. 1–18.

National Institute of Standards and Technology. 2024. The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29). https://doi.org/10.6028/NIST.CSWP.29

Nurse, J. R. C., Axon, L., Erola, A., Agrafiotis, I., Goldsmith, M., and Creese, S. 2020. "The Data That Drives Cyber Insurance : A Study into the Underwriting and Claims Processes," International Conference on Cyber Situational Awareness, Data Analytics and Assessment, pp. 1–8. (https://doi.org/10.1109/CyberSA49311.2020.9139703).

Sarabi, A., Naghizadeh, P., Liu, Y., and Liu, M. 2016. "Risky Business: Fine-Grained Data Breach Prediction Using Business Profiles," *Journal of Cybersecurity* (2:December), pp. 15–28. (https://doi.org/10.1093/cybsec/tyw004).

Schlackl, F., Link, N., and Hoehle, H. 2022. "Antecedents and Consequences of Data Breaches: A Systematic Review," *Information and Management* (59:4), Elsevier B.V., p. 103638. (https://doi.org/10.1016/j.im.2022.103638).

United States. 1998. Critical Infrastructure Protection (Presidential Decision Directive 63). Federation of American Scientists.

U.S. Small Business Administration. 2024. "Size standards," U.S. Small Business Administration, February 20 (available at https://www.sba.gov/federal-contracting/contracting-guide/size-standards; retrieved February 25, 2024).

Woods, D. 2021. "Systematization of Knowledge: Quantifying Cyber Risk," *IEEE Symposium on Security and Privacy* (S&P) (894700). (https://doi.org/10.1109/SP40001.2021.00053).