State Deals, Hacker Meals: Do Government Contractors Attract the Attention of Cyber Threat Actors?

Jiaxing Huang, Aldo Musacchio, Debarshi Nandy, Anna Scherbina, Shenkai Wu

Brandeis University

October 6, 2026



Outline

1. Introduction
2. Data
3. Methodology
4. Future Plan

Motivation

1. **National Security:** Firms increasingly handle sensitive government data. A breach in a contractor can be a breach in national defense.
2. **Economic Espionage:** Foreign adversaries target the "soft underbelly" of the supply chain to steal intellectual property and trade secrets.
3. **Supply Chain Assessments:** SecurityScorecard reports find that 35% of attributable breaches in the federal supply chain are state-sponsored.
4. **Financial Impact:** Understanding if government contracts act as a "risk signal" helps investors and regulators price cyber risk accurately

Literature Review

- **Kamiya et al. (2021):** attacks where personal financial information is lost involve a significant shareholder wealth loss. If a firm experiences a cyberattack, its competitors will also experience a wealth loss.
- **Gong et al. (2026):** politically connected firms are more likely to experience data breaches. The association is stronger when political connections are highly visible and in more recent years.

Literature Review

- **Kemp (2021):** awareness of cyber security schemes is associated with better security practices, but implementing the recommended measures is not associated with lower victimization or fewer negative consequences.
- **Samuels (2021):** firms improve financial transparency to win contracts. This increased disclosure may unintentionally heighten the firm's visibility to external actors, including adversaries.

Data

- Data source
  - USAspending.gov: Source for all prime government contract awards (2000-2020).
  - Advisen Cyber Data: Comprehensive database of cyber events, widely used in academic finance.
  - Compustat: fundamentals data
- Period: Monthly data from January 2000 to December 2020
- Variables: Agency, Recipient, Contract size, Product Code, Event Date, Type (Breach/Ransomware), Actor, Industry (FF48).

Sample Data Description

Contract data:

Awardee: The Boeing Company

Contract description: ejector, engine, airframe, plate, transmission

Cyber incident data:

Attacked company: 7-Eleven, Inc.

Case description: A Russian man, Vladimir Drinkman, pleaded guilty on involvement in what authorities called the largest computer hacking scheme ever prosecuted in the United States, compromising more than 160 million credit card numbers.

Sample Selection

Merging Process:

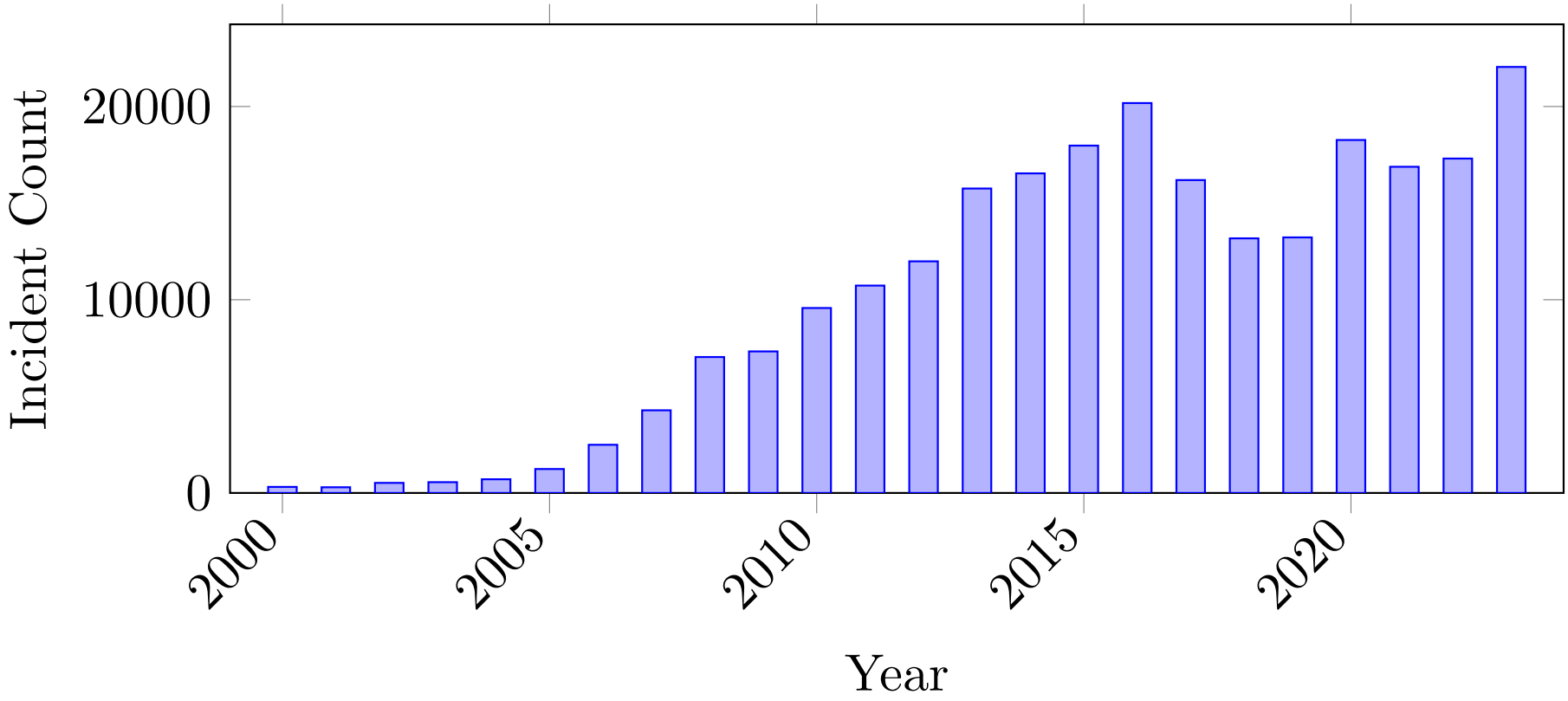
- Step 1: Perfect name match between USAspending recipients and Compustat firm names.
- Step 2: Identify the first year that a firm receives a contract to define the treatment cohort.
- Step 3: Perfect match of name and CIK between Compustat firm and Advisen.
- Step 4: Remove improper cyber incidents.
- Step 5: Re-classify to distinguish between pure cyberattack and lawsuit.
- Step 6: Count number of incidents for each matched firm with unique case ID.

Match rates by year

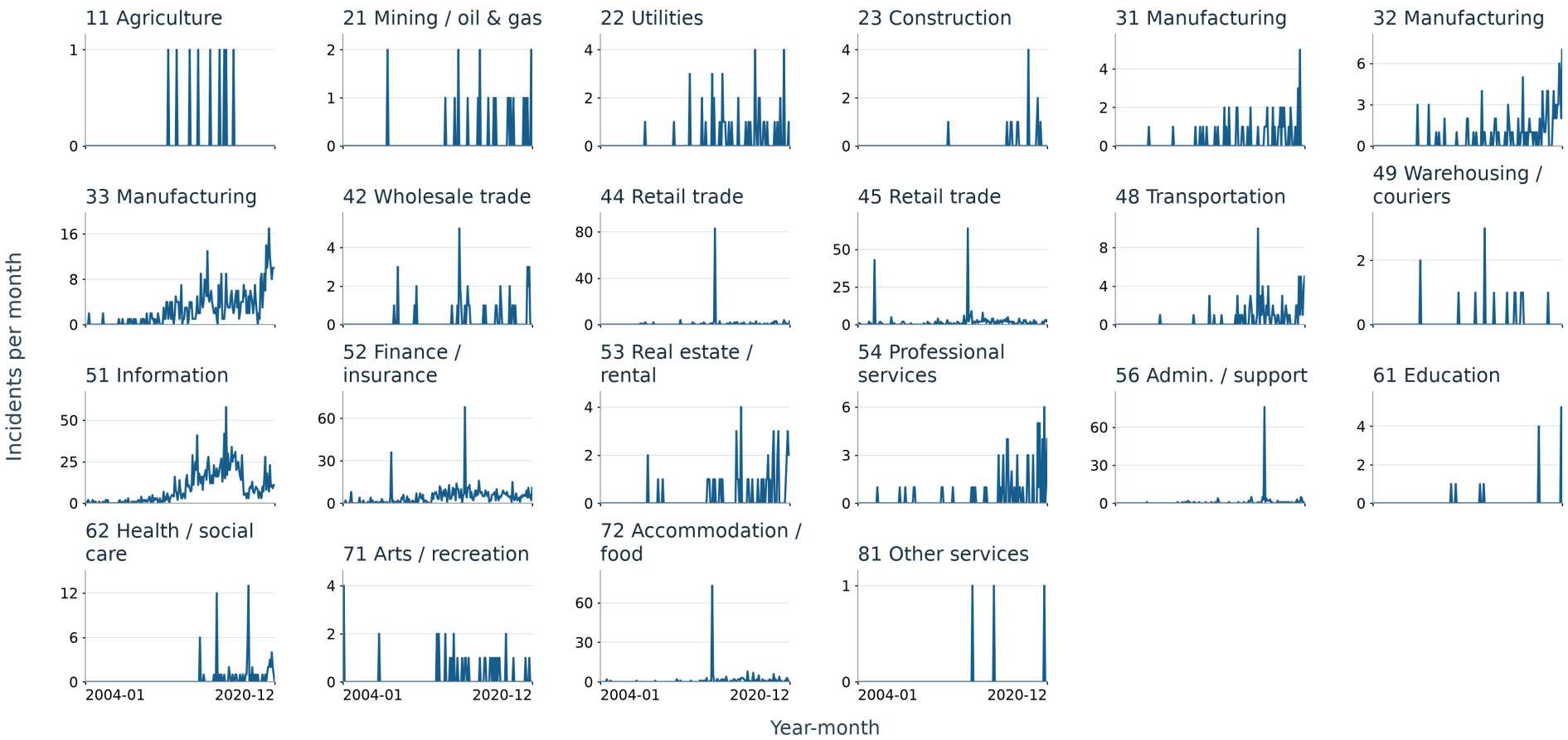
Year	Match Rate (%)	Non-Small Business (%)
2000	50.96	55.68
2001	51.07	55.84
2002	50.33	53.76
2003	52.88	48.88
2004	58.65	47.93
2005	58.13	49.99
2006	60.58	45.98
2007	60.55	50.72
2009	57.18	53.87
2010	57.26	53.02
2011	55.54	53.39
2012	54.20	53.45
2013	55.77	53.38
2014	60.07	54.16
2015	70.15	57.03
2016	68.56	57.44
2017	69.98	57.91
2018	68.11	55.55
2019	64.15	54.10
2020	64.54	52.19
2021	64.51	50.31
2022	58.02	50.27
2023	55.27	51.47

Table 1: Match rates of USAspending contractors and Compustat by year (percentage of contractors)

Cyber-attack incidents by year



Monthly cyberattacks by industry



Cyber incidents case type summary

Case type	Count	Cyber = 1 n (row %)
Data - Malicious Breach	11,353	8,720 (76.81%)
Privacy - Unauthorized Contact or Disclosure	9,738	0 (0.00%)
financial	3,757	1,934 (51.48%)
na	2,399	387 (16.13%)
Data - Unintentional Disclosure	2,150	0 (0.00%)
unknown	1,837	1,011 (55.04%)
Network/Website Disruption	1,660	1,486 (89.52%)
Data - Physically Lost or Stolen	1,452	0 (0.00%)
Privacy - Unauthorized Data Collection	1,132	17 (1.50%)
Cyber Extortion	1,031	1,027 (99.61%)

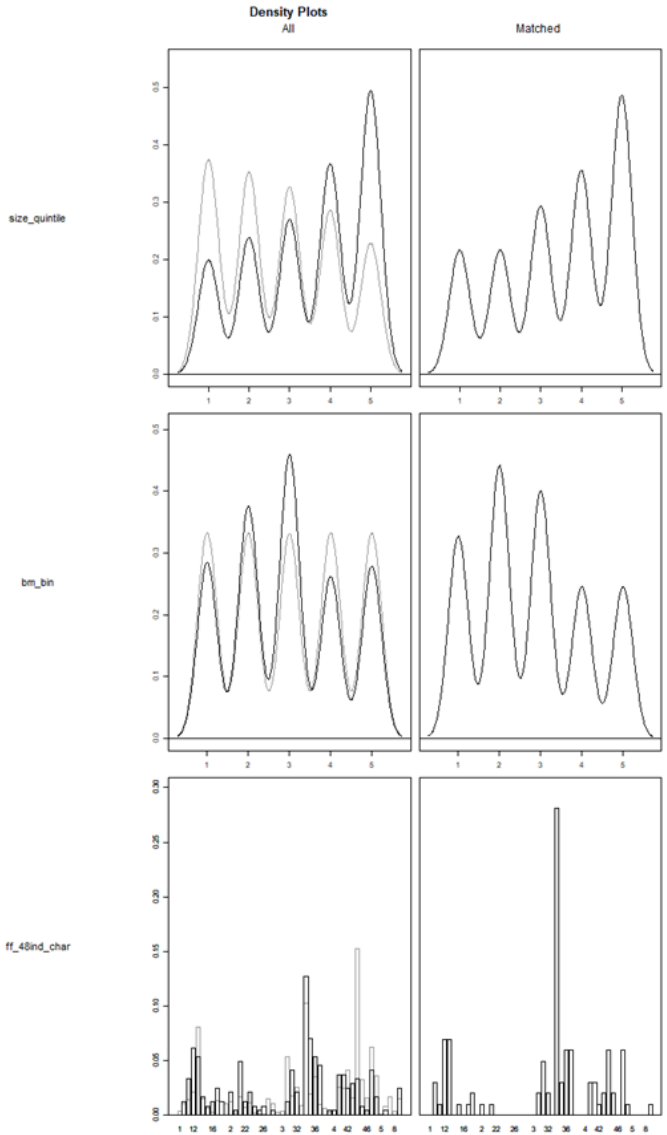
Table 2: Cyber incidents case type summary

Cyber incidents actor type summary

Actor type	Count	Cyber = 1 n (row %)
state-affiliated	212	212 (100.00%)
External - Hacktivist	701	687 (98.00%)
External - Criminal Organization	1,746	1,670 (95.65%)
External - Nation State	205	196 (95.61%)
activist	410	343 (83.66%)
organized crime	1,698	1,347 (79.33%)
unaffiliated	403	268 (66.50%)
External - Other	8,768	5,501 (62.74%)
system admin	297	150 (50.51%)
External - Criminal Individual	130	59 (45.38%)

Table 3: Cyber incidents actor type summary

Density Plots: Pre- and Post-Matching Balance



Black line: treated firms

Gray line: control firms

Summary Statistics

	Treated			Control			
Variable	**Median**	**Mean**	**SE**	**Median**	**Mean**	**SE**	**Difference**
log(incidents)	0.000	0.006	0.000	0.000	0.002	0.000	0.004***
Market Equity	520.066	8648.537	58.823	147.597	2862.314	13.477	5786.224***
B/M Ratio	0.455	10.574	1.821	0.544	5.591	1.273	4.983
Assets	509.114	17588.182	149.188	272.632	9320.964	79.148	8267.217***
Observations		737724			1271628		
# of Firms		3107			8289		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Treated vs. Control Group Characteristics

Cyberattack classification

Examples of Advisen dataset incidents:

1. "An employee of firm A lost a portable hard drive on bus."
   This is a physical lost/stolen, **NOT** cyberattack.
2. "Firm B faxing unwanted advertisements to people."
   This is an unauthorized contact, **NOT** cyberattack.
3. "Thieves use an external skimming device to duplicate ATM cards."
   This is skimming, **NOT** cyberattack.
4. "A man infiltrated a corporate Internet account, and attacked NASA and U.S. Department of Defense computers."
   This is a malicious breach into another computer through internet, so it **IS** cyberattack.

Cyberattack classification

We use AI tools to help us classify cybersecurity incidents:

Qwen 3.5, GPT-5.4, Claude Opus 4.6: They are used to classify whether a case is cyberattack or lawsuit, based on the case description.

SecureBERT: SecureBERT is a RoBERTa-based, domain-specific language model trained on a large cybersecurity-focused corpus. It is used to classify more cases whether they are cyberattack or lawsuit, and what the actor type is.

In the DID regression, we only use cyberattack incidents.

Human vs. AI

Is AI reliable?

Step 1: randomly select 40 cyber case descriptions

Step 2: 3 experts classify 40 cyber case descriptions, AI classify 20 randomly selected case samples

Step 3: 2 auditors blindly check if the classifications are consistent and correct

SecureBERT performed as well as human experts. The accuracy is 95%.

Methodology

We employ a Difference-in-Differences (DID) design with staggered adoption:

- Callaway & Sant' Anna : the most classic DiD model
- Stacked DID: construct cohort manually and run DiD, better for multiple periods

Methodology

First we employ a Difference-in-Differences (DiD) design with staggered adoption:

Callaway & Sant'Anna (2021)

$$\text{ATT}(g,t) = \mathbb{E}[Y_t - Y_{g-1} \mid G_g = 1] - \mathbb{E}[Y_t - Y_{g-1} \mid D_t = 0, G = \infty] \quad (1)$$

Y_t: $Log(1 + \text{number of incidents})$ of year t

g: The time period when a firm first becomes treated

t: 12, 24, 36 months event window after g

G_g: 1 if a firm is first treated in period g, 0 otherwise

$G = \infty$: Define as never treated group

$D_t = 0$: Define as not-yet treated group

Methodology

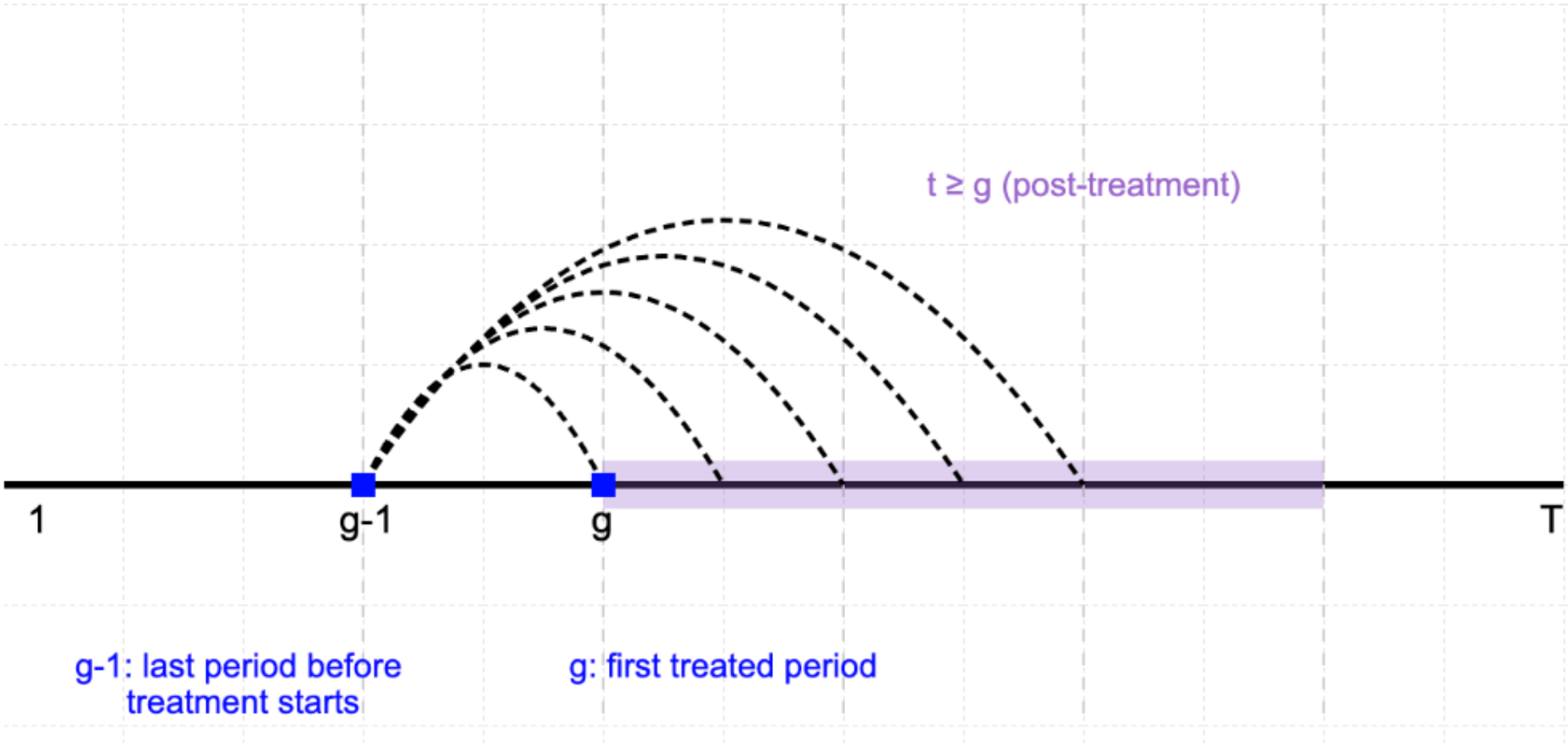


Figure 1: ATT for different time horizons

Treatment and Control Group

- Treatment Group: Firms are defined as "treated" if they have ever got a contract.
- Control Group (Matching): For each treated firm, control firms are matched by pre-1 year of Fama-French 48 industry, capital size quintile.

Hypothesis

Primary Hypothesis (H_1): Firms awarded government contracts will experience a statistically significant increase in the frequency of cyber attacks compared to matched non-contractor peers.

Alternative Hypothesis (H_0): Getting government contract has no effect on firms getting cyber attacks.

DID results by event year

Event year	N of firms	12m avg	t-stat	24m avg	t-stat	36m avg	t-stat
2000	507	-0.001**	(-2.201)	-0.003***	(-6.946)	-0.003***	(-7.103)
2001	256	0.000	(0.120)	0.002***	(3.333)	0.002***	(3.406)
2002	263	0.006***	(7.679)	0.001	(0.971)	0.000	(0.214)
2003	227	-0.006***	(-6.610)	-0.003***	(-3.061)	-0.002**	(-2.484)
2004	256	0.002**	(2.251)	0.002**	(2.546)	0.003***	(3.062)
2005	171	0.002	(0.828)	-0.009***	(-5.450)	-0.007***	(-5.093)
2006	140	-0.001	(-0.783)	0.007***	(4.567)	0.018***	(12.312)
2007	103	0.022***	(9.254)	0.000	(0.223)	-0.001	(-0.341)
2009	143	0.006***	(3.362)	0.008***	(5.014)	0.006***	(4.355)
2010	67	0.009***	(3.124)	0.003	(1.384)	0.003	(1.486)
2011	68	-0.006**	(-2.246)	0.002	(0.754)	-0.001	(-0.476)
2012	59	-0.004	(-1.183)	-0.004	(-1.484)	-0.008***	(-3.194)
2013	43	-0.013***	(-2.691)	-0.004	(-1.523)	-0.014***	(-5.004)
2014	56	0.009***	(3.053)	0.016***	(5.526)	0.052***	(15.869)
2015	71	0.000	(-0.174)	0.006***	(2.867)	0.010***	(4.496)
2016	47	0.032***	(8.788)	0.030***	(8.647)	0.033***	(10.250)
2017	49	0.028***	(11.879)	0.022***	(11.110)	0.023***	(11.334)
2018	42	-0.000	(-0.137)	-0.004	(-1.187)	-0.000	(-0.121)
2019	47	0.010*	(1.871)	-0.010**	(-2.029)	0.002	(0.482)
2020	52	-0.014***	(-4.844)	-0.001	(-0.256)	0.000	(0.214)
All Years	2667	0.003***	(7.047)	0.002***	(5.299)	0.003***	(10.558)

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 5: Difference-in-differences estimates by event year

Stacked DID Regression

Cengiz (2019)

$$\text{Incidents}_{ist} = \sum_{\substack{k=-12 \\ k\neq -1}}^{K} \delta_k \mathbf{1}\{t - g_s = k\} + \sum_{\substack{k=-12 \\ k\neq -1}}^{K} \beta_k D_{is} \times \mathbf{1}\{t - g_s = k\} + \alpha_{is} + \lambda_{st} + \varepsilon_{ist}. \tag{2}$$

Incidents_{ist}: Monthly incident count for firm i in stack s at time t.

g_s: Contract event month for stack s.

K: Post-event window length: 12, 24, or 36 months.

D_{is}: One for the treated firm; zero for its control.

$\mathbf{1}\{t - g_s = k\}$: One when the observation is in event month k.

α_{is}: Firm fixed effect.

λ_{st}: Time fixed effect.

Yearly DID results by industry

Table 6: Yearly DID Results by Industry

	(1)	(2)
Automobiles and Trucks (23)	−0.055***	0.148**
	(0.009)	(0.052)
Construction (18)	–	0.192**
		(0.022)
Insurance (45)	–	−1.224**
		(0.307)
Post years	1	2
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes

Monthly DID results by industry

	(1)	(2)	(3)	(4)
Pharmaceuticals (13)	–	–	0.001*	0.004*
Entertainment (7)	−0.026*	−0.024*	−0.030**	–
Business Services (34)	−0.003*	–	–	–
Post window	12m	24m	36m	60m
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes

Table 7: Monthly DID Results by Industry

Conclusion

- Using Callaway and Sant'Anna's DID method, in most years, the government contractors receive more cyber attacks than their peers.
- Firms in some industries seem to be more attractive to external cyber attackers after getting government contracts, while some are well prepared.

Future Plans

- Obtain additional data from other databases and incorporate them into the existing dataset.
- Explore other effects that getting a government contract brings to the company.