Human Capital Acquisition in Response to Data Breaches

Sarah Bana, Erik Brynjolfsson, Wang Jin, Sebastian Steffen, and Xiupeng Wang

Abstract: Do firms react to data breaches by investing in cybersecurity talent? Or, are they more likely to invest in talent that helps protect their public image or tackle the legal aftermath? In other words, do they treat the root cause of the vulnerability through substantive human capital investments or do they treat the symptoms through symbolic ones? Combining unique information on data breach events and detailed firm-level job posting data, we leverage a difference-in-differences (DiD) design and show that firms increase their hiring for both cybersecurity as well as public relations and legal workers after suffering a breach, with important heterogeneity by firm type. We further show that public scrutiny serves as an effective mechanism to incentivize substantive human capital investment and better align firms' incentives with those of the public. Gathering additional data on media and online search attention around data breaches, we find that increased public scrutiny due to data breaches shifts firms' investments toward substantive, rather than symbolic, measures. Given the increase in volume and severity of cyberattacks, our results provide important and timely insights into firms hiring responses and their incentives to more substantively safeguard their data.

Keywords: Cybersecurity, Substantive and Symbolic Adoption, Human Capital Acquisition, Media and Public Attention, Value of Data and Privacy

1 INTRODUCTION

The digitization of business activities has led to an ever-increasing stream of digital information and data. This transformation has been facing an increasing threat from cybercrimes that result in the loss of valuable data. A series papers have studied the effect of data breaches on stock market performance (Hilary et al. 2016), consumer behavior (Turjeman and Feinberg 2019), and litigation challenges (Romanosky et al. 2014). This paper fills the gap in the literature by studying firms' strategic investments in human capital in response to data breaches. In it, we show that breached firms increase their hiring for both substantive (i.e., cybersecurity-related workers) as well as symbolic (i.e., legal and public relations) talents. More importantly, by gathering data on media coverage and public search attention associated with the data breach events, this study finds that public scrutiny incentivizes breached firms to acquire relatively more substantive talents to target the root causes, instead of the symptoms, of the breach.

The World Economic Forum estimates that 463 exabytes of data will be created each day by 2025.[1] As a result, data protection has become a major responsibility for digitized firms. However, this responsibility is becoming a larger burden as the value of firms' data increases and hackers and cybercriminals launch more sophisticated attacks. Over the last 15 years, over 10,000 data breaches have been announced in the United States, which exposed trillions of individual records. The average cost per data breach is estimated at $8.64 million in the U.S. with an increasing number of incidents and scale over time.[2] With the rise of big data, the reliance on the cloud and software-as-a-service (SaaS) (August et al. 2014), as well as the increasing adoption of work-from-home practices during, and likely after, the Covid-19 pandemic (Bai et al. 2021, Barrero et al. 2021), firms are more vulnerable than

[1] See: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[2] See https://www.ibm.com/security/digital-assets/cost-data-breach-report (last visited on March 12, 2021)

ever to cybercrime.[3] Cybersecurity has therefore become an increasingly important domain for firms, policy makers, customers, and researchers alike.

Given the increasing importance of cybersecurity and the concurrent rise in the value of data as well as the number of cybercrimes, one would expect to see significant firm investment in cybercrime prevention. While proactive, preventative security investments have been found to be both cheaper and more effective in deterring data breaches (Kwon and Johnson 2014), the majority of cybersecurity investments in software and infrastructure, if they are taken at all, are taken retroactively (Kankanhalli et al. 2003) or insufficiently (Gordon et al. 2015a,b). Research on the effect of data breaches on firms' stock market performance (Acquisti et al. 2006, Amir et al. 2018, Hilary et al. 2016), consumer behavior (Janakiraman et al. 2018a, Turjeman and Feinberg 2019, Buckman et al. 2019), and litigation (Romanosky et al. 2014) reveals a similar level of inactivity and lack of care by investors, managers, and consumers. Richardson et al. (2019) provide a comprehensive review and conclude that investor reactions to data breaches are relatively small: public firms only suffer a short-term 0.3 % loss in cumulative abnormal returns after a breach - except for a few catastrophic incidents. In particular, and contrary to public belief, consumers have a tendency for inaction after receiving a data breach notification.[4] This is largely consistent with research findings in Turjeman and Feinberg (2019), which show that even for a breach of a matchmaking website, whose breached data could potentially embarrass customers and harm their personal relationships, initial reductions in usage and increases in image deletions were relatively short-lived. Athey et al. (2017) also report that despite claiming that privacy is very important, consumers often show behaviors contradicting such statements.

While there is some evidence that firms lack incentives to make proactive physical and IT capital investments to properly respond to data breaches, little is known about firms post-incident human capital investments. According to a recent Forbes article, "security

[3] See https://newsroom.ibm.com/2020-06-22-IBM-Security-Study-Finds-Employees-New-to-Working-from-Home-Pose-Security-Risk

[4] See https://www.idtheftcenter.org/data-breach-notice-research-by-the-identity-theft-resource-center-shows-consumers-dont-act-after-a-data-theft/.

starts with people", that cybersecurity talent is one of the most crucial investments that firms need to make to protect themselves in today's digital world.[5] Given the critical role of centralized IT governance in deterring and preventing data breaches (Liu et al. 2020), it is important to understand firms' human capital investment in cybersecurity after suffering data breaches. Following Angst et al. (2017), we apply neo-institutional theory to such human capital investments to further distinguish between *symbolic* and *substantive* adoption of protective organizational practices in hiring. Under this lens, substantive adoption represents hiring for workers that can help to treat the root cause of the cyber vulnerability, while symbolic adoption is only loosely coupled with such action and relatively more focused on the symptoms, or the aftermath, of a databreach. Given that hiring is often considered to be a strategic commitment, we argue that human capital investments, in the form of hiring for cybersecurity personnel and related skills, is one of the most substantive forms of cybersecurity investment, while hiring for human capital related to legal and public relations (PR) is relatively more symbolic.

However, in the past, large-scale firm-level data on cybersecurity-related human capital investments was hard to come by, especially for private firms, which often tend to be the most vulnerable to cyberattacks. Bringing together breach incident data from Privacy Rights Clearinghouse (PRC) with firm-level data on online job postings from Burning Glass Technologies (BGT), our work fills this gap by exploring the impact of data breaches on firms post-breach human capital investment. To the best of our knowledge, our paper is the first to explicitly address the effect of data breaches on firms' demand for cybersecurity workers, as well as other human capital and skills.[6] We view these human capital investments as important complements to firms' IT capital investments (e.g., Brynjolfsson and Milgrom 2013) and, following Angst et al. (2017), view them through the lens of substantive and symbolic adoption. While we cannot directly measure hiring, job postings provide a

[5]See https://www.forbes.com/sites/sap/2021/10/19/how-to-attract-cybersecurity-talent-and-build-a-culture-of-security/?sh=6242aba16b5f

[6]The closest papers are Say and Vasudeva (2020), Hilary et al. (2016) and Banker and Feng (2019), which explore management turnover in response to data breaches.

meaningful measure of hiring intent, especially in the short-term.

By adopting a staggered Difference-in-Difference design (Cheng and Hoekstra 2013), in which we consider suffering a data breach as the treatment, our results indicate that firms that suffer a data breach are two percentage points more likely to post cybersecurity-related jobs after a data breach. This substantive hiring effect is most pronounced three months after the announcement of the data breach and is isolated to incidents in which digital information was breached - the placebo test for breaches of physical (or analog) data shows no such effect. Taking advantage of the granularity of our occupation and skill-level data, our results also offer important managerial hiring insights into handling data breaches. We find that firms specifically target cybersecurity talent in information security analytics, computer system analytics, database administration, and network and computer systems administration, but not other IT occupations such as computer network support and computer network architecture, which suggests that threat detection and analysis, rather than prevention, are the most common response.

In addition, we further contribute to the literature by exploring firms' more symbolic hiring intents that primarily treat the symptoms of a data breach. Following the National Institute of Standards and Technology (NIST) framework (Petersen et al. 2020), we expand our analysis to include both hiring intent for legal as well as PR talent. Workers in legal occupations can help to resolve potential legal issues and assure compliance with applicable privacy laws, regulations, and constitutional requirements, while PR can counter negative media attention and manage the firm's image after suffering a data breach. We find a significant increase of two percentage points in symbolic hiring - a similar average magnitude as for substantive hiring. When comparing hiring of different types within firm, substantive hiring is larger than symbolic hiring and the difference is statistically significant.

Next, we conduct a series of tests to explore the heterogeneity of the treatment effect and its underlying mechanisms. First, we compare goods-producing industries with service-providing industries. Since firms in service-producing industries tend to rely much more on

customer trust due to dealing with more sensitive customer data (i.e. financial information), we hypothesized that firms in these industries need to hire relatively more substantive talent. Our results indicate that firms in these industries are indeed significantly more likely to hire both cybersecurity as well as legal and PR talent, while firms in goods-producing industries tend to be unresponsive.[7]

Second, we compare the data breach event responses across private firms and public firms. On the one hand, public firms face quarterly reporting requirements and are more likely to face regulatory attention and public scrutiny in general, which suggests that they should adopt more substantive measures. On the other hand, prior research shows that stock and customer responses to data breaches are limited as long as public attention and legal fallout are limited, which suggests that public firms have stronger incentives to adopt symbolic rather than substantive measures. A recent study further shows that public firms tend to choose the announcement dates of data breaches strategically to minimize public attention and stock market reactions (Foerderer and Schuetz 2022), especially for breaches of very sensitive, personal customer data such as healthcare and credential information. Together, this suggests that public firms may have stronger incentives to hire more symbolic human capital. Indeed, we find that while both public and private firms engage in significantly more symbolic hiring in responses to breaches, this effect is larger for public firms. For substantive hiring, we find a significant effect for private firms, while the picture is less clear for public firms.

Finally, we directly test the hypothesis that the increased public scrutiny specific to data breach events may serve as an effective mechanism to incentivize firms to adopt more cybersecurity (i.e. substantive) talent instead of PR and legal (i.e. symbolic) talent and thus help to realign firms' incentives with those of the public. We leverage search data from Google Trends and data on mentions in online news outlets from the MIT Media Cloud to

[7]Based on the Bureau of Labor Statistics (BLS) definition, the goods-producing sector includes construction and manufacturing while the service-providing sector include Information, Finance, Insurance Administrative and Support Services, as well as Retail Trade. For more details, please see https://www.bls.gov/iag/tgs/iag_index_naics.htm.

measure the public scrutiny that breached firms face. We show that a sharp rise in public scrutiny due to data breach events tends to increase firms' substantive hiring investments while this increase is much more subdued for their symbolic hiring investments. We also identify a consistent pattern of firms shifting away from symbolic towards substantive hiring after suffering a highly visible data breach. Overall, these results imply that public scrutiny can serve as an effective mechanism to incentivize firms to safeguard their data more seriously. However, since the public scrutiny of most data breaches remains relatively low, the question remains how to adequately educate customers to care more about their data privacy and security, such that they exert more public pressure on firms. As it stands, the responses by firms to hire for both substantive and symbolic human capital remain correspondingly low, which suggests that additional incentivization such as through policy intervention or, as we show, public scrutiny may be required.

The rest of this article is organized as follows. Section 2 discusses the relevant literature; Section 3 describes our data; Section 4 introduces the empirical methods; Section 5 and 6 present the main results and a series of effect heterogeneity; Section 7 investigates public visibility as mechanism for increased demand for substantive over symbolic adoption in response to data breaches; then section 8 concludes.

2 LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

A series of papers investigate how cybersecurity investments can help firms to better defend against and prevent cybercrime.[8] By surveying 63 Information System (IS) managers from various sectors of the economy, Kankanhalli et al. (2003) show that greater efforts and preventive measures lead to enhanced IS security effectiveness. This is particularly true for newer security technologies (Murciano-Goroff 2019), and such investments have also been

[8]For an excellent research curation on securing digital assets, see Hui et al. (2018).

linked to higher market values (Bose and Leung 2013, 2019, Aytes et al. 2006) as well as lower capital costs (Havakhor et al. 2020) and overall enhancements of firms' competitiveness.

Following neo-institutional theory, Angst et al. (2017) further differentiate IT security investments into substantive and symbolic adoption of IT security and find the latter to be less effective at preventing data breaches. That is, simply investing more in IT security infrastructure is not necessarily better - these investments tend to be more effective if they are directed towards substantive adoption aimed at achieving meaningful technical benefits instead of at symbolic, external signals of legitimacy. For example, a sticker for an alarm system may signal that a house is secured and is thus a symbolic investment, whereas the deployment of an actual alarm system is considered a substantive investment. In a similar vein, we argue that some types of hiring for human capital are more effective, or substantive, than others. Relatedly, Kwon and Johnson (2014) find that proactive security investments lower security failure rates and augment cost effectiveness much better than retroactive ones. Huang and Madnick (2020) further suggest several substantive types of actions firms should take to effectively respond to hacks, including investments in hiring cybersecurity professionals and enhancing internal cybersecurity capabilities, as well as several symbolic ones, such as voluntary public disclosure of breach incidents to limit the loss of consumer trust (Janakiraman et al. 2018b) and market value (Gordon et al. 2010). This is also consistent with the long-standing view of organizational complementarity (Bresnahan et al. 2002, Brynjolfsson and Milgrom 2013), where the return of IT capital investment depends on investments in complementary skilled labor. In this case, without investment in cybersecurity human capital, it becomes more challenging for firms to effectively improve their data and network security.

Importantly for our focus on firms' human capital hiring response to data breaches, there is evidence that while outsourcing may be cost-effective, it also induces a principal agent problem if both breach prevention and detection are outsourced to the same firm (Cezar et al. 2014). Huang et al. (2018) further suggest that investing in a firm-internal cybersecurity

team can partially alleviate misalignment of incentives as it allows firms to properly assign responsibilities and enable meaningful collaboration. Liu et al. (2020) corroborate that despite outsourcing, internal cybersecurity human capital is still necessary for data security, by showing in a large-scale sample of firms that centralized IT governance is more effective in reducing the risk of data breaches than outsourcing.

Despite this guidance from the literature on effective data breach management through physical and human capital investments in cybersecurity, firms likely still underinvest in cybersecurity (Telang 2015) and, importantly, may do so deliberately and rationally (Gordon and Loeb 2006, Foerderer and Schuetz 2022). One of the existing explanations for the hesitance to adopt more substantive measures is the intangible nature of many of the benefits of cybersecurity investments. The intangible benefits of this type of investments go far beyond the cybersecurity improvements or positive market returns of announcing such investments (Bose and Leung 2019), and are largely invisible in the short-run. Chief among these intangible benefits is the increased trust of consumers and business partners (Tanimura and Wehrly 2015, Solove 2007), which can easily be lost through the negative media attention after a data breach involving sensitive consumer data (Turjeman and Feinberg 2019). Therefore, many firms may myopically focus on improving public relation, resolving legal issues and dealing with other symptoms following a data breach, instead of targeting the cyber vulnerability - the root cause - of the problem.

Indeed, prior research suggests, with somewhat mixed evidence, that while there are direct financial costs associated with data breaches, these tend to be small and limited to specific types of firms (Ebrahimi and Eshghi 2022). An analysis of 266 breaches by Hilary et al. (2016) reveals that there were no significant persistent negative market reactions following a data breach. Similarly, Richardson et al. (2019) provide a comprehensive review on the consequences of data breaches and find very small short-term losses in returns on asset for breached firms, which diminish within days after the breach. Bolster et al. (2010) also report minimal impacts on firm values after data breaches, except when the breach gains

significant public exposure through newspapers or other media outlets, in which case the fallout can be substantial. More recently, Foerderer and Schuetz (2022) show that firms may strategically time the announcement of data breaches to coincide with busy news reporting in order to dilute attention and reduce potential stock market losses.

However, little is known about the impacts of data breaches on firms' human capital investments - largely due to significant data and transparency issues in the past. The consensus developed in prior literature regarding firms' inadequate incentives to invest in cybersecurity after suffering a breach thus lacks a more comprehensive picture, especially considering that hiring for cybersecurity professionals and skills is one of the most substantive and effective forms of cybersecurity investment. Building on prior literature, we start by exploring firms' demand for cybersecurity as well as other related skills after publicly announcing a data breach in accordance with state-mandated data breach notification laws.

To tackle data breach events and reduce the risk of future breaches, firms need to improve their cybersecurity infrastructure and practices. This may include temporarily closing all remote access, identifying breach source and vulnerability, installing and testing new security tools and infrastructure such as firewalls, malware detection software, intrusion detection, data loss prevention tools, and penetration tests, and eventually updating all of their security protocols (Huang et al. 2018). These procedures require significant investment as well as sophisticated skills and domain expertise. Generally, firms need to outsource and/or acquire new talent through hiring for such tasks.[9] This implies that we should expect to observe an increase in both the probability and number of job postings for cybersecurity occupations and skills posted by firms after being "treated", i.e., experiencing a data breach event. Our first hypothesis to test is therefore:

[9] An additional concern may be that if outsourcing or upskilling of internal workers represents a significant part of firms' responses to data breaches, the hiring effects we observe may be downward biased. However, these responses would imply significantly larger capital investments in cybersecurity after data breaches, which is inconsistent with prior findings. Furthermore, even firms which relied on outsourced cybersecurity investments or internal teams before suffering a breach may increase hiring their own cybersecurity talent afterwards, due to their needs to centralize and improve internal security (Huang et al. 2018, Liu et al. 2020). We therefore believe that it is unlikely that our results are significantly affected by these omitted variable bias.

Hypothesis 1 (Substantive Adoption - Cybersecurity Hiring) *To treat the cause of a data breach and to improve cybersecurity, firms increase substantive adoption after suffering a data breach by increasing their demand for cybersecurity workers and skills.*

However, prior literature has also shown that firms may lack incentives to adopt cybersecurity measures in a substantive manner (Gordon et al. 2015a,b). In fact, it has been documented that firms commonly treat data breach events "merely as public relations problems while continuing to use lax data security practices" (Manworren et al. 2016). Thus, firms are likely to focus on dealing with regulatory compliance, public relations, and subsequent potential legal fallout. For instance, firms are required to comply with data breach notification laws once they identify a breach. Additionally, the Sarbanes-Oxley Act (SOX) requires public firms to prove their cybersecurity credentials. If willfully failing to report truthfully, a CEO or CFO can be liable for maximum fines of up to $5 million and 20 years imprisonment.[10] Furthermore, firms may face class action lawsuits by their costumers, such as T-Mobile did after its data breach event in August 2021. Therefore, the demand for symbolic talent in public relations and legal occupations will likely increase due to these shocks:

Hypothesis 2 (Symbolic Adoption - PR and Legal Hiring) *To treat the symptoms of a data breach, firms increase symbolic adoption after suffering a data breach and increase their demand for public relations and legal workers and skills.*

While it is crucial to study firms' actual human capital investment responses after data breach events, large-scale empirical evidence is still lacking regarding what types of firms would be most motivated to tackle cybersecurity in a substantive manner and what could stimulate them in doing so. Firms generally lack incentives to react to data breaches, largely due to two asymmetric information problems. The first is that customers and investors cannot accurately observe, and thus evaluate, firms' investments in cybersecurity (Garcia

[10]Note however, that several of these laws may only be tangentially related or poorly enforced.

2013). This is akin to a moral hazard in which the lack of adequate monitoring by the public allows firms to shirk their responsibility and underinvest. Telang (2015) argues that firms lack the incentives to fully compensate customer losses and therefore underinvest in cybersecurity as they only minimize their own loss rather than the total social loss. Given that the negative impacts of data breach events on firms are found to be generally small and context-dependent (Richardson et al. 2019), the gap between what a benevolent social planner would want firms to do versus what they, myopically, choose to do appear to diverge significantly.[11]

The second reason is that customers often significantly underestimate the value of their own data. Customer data has important positive externalities that firms, such as insurance companies, advertisers or biotech companies, gain from having it, as customer data allows firms to build prediction models to extrapolate information - notably, not only for their customers, but also for non-customers (Choi et al. 2019). However, customers often do not understand how valuable their data is and, thus, even if they had more insights into firms' cybersecurity measures, their responses to data breaches and investments into cybersecurity may not be sufficient (Collis et al. 2021). The literature documents significant heterogeneity and inconsistencies in customers' views on privacy as well as in their responses to data breaches which vary with the sensitivity of the breached data (Turjeman and Feinberg 2019, Solove 2007). Turjeman and Feinberg (2019) show that the reductions in usage and increases in image deletions due to a data breach were relatively short-lived even for a matchmaking site, whose breached data could potentially embarrass customers and harm their personal relationships. Athey et al. (2017) find in a field experiment, that despite claiming that privacy is very important, consumers often show behaviors contradicting such statements while another survey shows that only 11% of respondents stopped dealing with a firm following a data breach (Ablon et al. 2016).

To overcome this lack of incentives and to shift towards more privacy-cognizant firm be-

[11] See Moore (2010) for a meaningful discussion on the economics of cybersecurity.

haviors, governments in the US and abroad have implemented regulation on data breaches. This includes the EU's implementation of the General Data Protection Regulation (GDPR) in 2018 as well as the data breach notification laws that all U.S. states adopted over time between 2002 and 2018. While government intervention has been shown to improve cybercrime prevention to some extent (Hui et al. 2017, Murciano-Goroff 2019, Romanosky et al. 2011), more action is required, given the lack of stringency, consistency, and enforceability of the laws across states as well as the surge of cybercrime activity in the US (Romanosky et al. 2014, Greenwood and Vaaler 2021).[12][13]

To further assess the mechanism behind firms' responses to data breaches of customer data, we follow the literature to assess firm and industry heterogeneity. First, we compare firms in service-providing industries with firms in goods-producing ones. The former tend to be more consumer-facing and their data breaches are more likely to contain sensitive customer information, which draws significantly more public and media attention.[14] Next, we compare public and private firms, with the former facing significantly higher regulatory and public scrutiny than the latter, including through mandated reporting and auditing (Richardson et al. 2019). Specifically, we test the following hypotheses to assess effect heterogeneity:

Hypothesis 3a (Heterogeneity by Industry) *Firms in service-providing industries react more strongly to data breaches in both substantive and symbolic hiring than firms in goods-producing industries.*

Hypothesis 3b (Heterogeneity by Listing Status: Public vs. Private Firms) *Public firms react more strongly to data breaches than private firms in both substantive and symbolic hiring.*

[12] Greenwood and Vaaler (2021) find no effect of decreasing in data breach incident counts or magnitudes following Breach Notification Law enactment.

[13] See https://www.security.org/resources/digital-privacy-legislation-by-state/

[14] One important challenge for our and many other cybersecurity-related papers is that the vast majority of data breaches remains undetected and unreported. This dark figure is particularly large for breaches of non-customer data, such as those from business to business (B2B) transactions which are most prevalent in goods-producing industries.

As we discuss above, data breaches may put firms under much greater public scrutiny, either because of the nature of the data (Hypothesis 3a) or the nature of the firm ownership (hypothesis 3b). In turn, firms may suffer significant financial and trust losses (Richardson et al. 2019). Thus, public scrutiny can incentivize firms to take data breaches and cybersecurity more seriously. To study this directly, we measure the public scrutiny associated with data breach events by collecting data on the occurrences of names of breached firms in media coverage (from the MIT MediaCloud) as well as in public searches (from Google Trends). Firms that face sharply elevated public scrutiny are likely under significantly higher pressure to improve and invest in substantive cybersecurity adoption. We test the following hypothesis on public scrutiny as an effective incentive-realigning mechanism:

Hypothesis 4 (Scrutiny Incentivizes Substantive adoption) *Firms that experience sharply elevated public scrutiny through media and public searches after a data breach, will increase substantive adoption relatively more than symbolic adoption.*

3 DATA AND SUMMARY STATISTICS

We combine several novel data sources to study firms' substantive and symbolic hiring responses to data breaches: (i) firms' online job postings from Burning Glass Technologies (BGT), (ii) data breach incidents from Privacy Rights Clearinghouse (PRC), (iii) firms' media attention from the the historic online news repository of the MIT Media Cloud project, and (iv) firms' public search attention from Google Trends. In the following subsections, we describe each of these data, the matching procedure, as well as provide summary statistics for our analytic sample.

3.1 Data on Firms Hiring from Burning Glass Technologies (BGT)

The BGT data covers about 200 million online job vacancy postings posted on over 40,000 distinct online job platforms in the United States between 2010 and 2020 and arguably covers

the "near-universe" of job postings. Each vacancy posting is parsed, deduplicated, and annotated with the posting date, an occupational code based on the Standardized Occupational Classification (SOC) system, an industry code based on the North American Industry Classification System (NAICS), the employer, and which skills were demanded, among several other variables. The skills data is annotated via BGT's industry-leading skill parser, which is rule-based and employs string searches as well as disambiguation rules. It maps each job postings' skills into a detailed skills taxonomy, which consists of 3 levels of granularity.[15]

At the most detailed level, the BGT taxonomy includes approximately 16,000 skills - these are nested within 658 skill clusters, which themselves are nested within 28 skill cluster families. For example, *Threat Analysis* and *Intrusion Detection* are both skills within the *Cybersecurity* skill cluster, which itself falls into the *Information Technology* skill cluster family.[16]

Given the unprecedented details information on skills within firms' job postings, the BGT data is ideal for studying the evolution of firms' skill demands over time (e.g., Deming and Kahn 2018 and Bana et al. 2020). Given the large number of postings, we are able to aggregate their count and their specific skill demands to the firm-month level. We create a balanced panel of firm-level skill demands for cybersecurity-related occupations[17] and skills.[18]

[15] A great overview of the skill details can be found in Lassébie et al. (2021).

[16] Notably, this taxonomy is significantly more detailed than other skill taxonomies, such as the Bureau of Labor Statistics (BLS)' O*NET skill taxonomy, which contains just two levels, with 35 skills mapped into six skill groups. Furthermore, BGT job postings are scraped daily and are therefore able to capture changes in skill demands on a monthly level. O*NET only undergoes yearly updates, which generally only cover a subset of occupations.

[17] Specifically, we include the following 2010 SOC occupations: Computer and Information Analysts (15-1120), Computer Systems Analysts (15-1121), Information Security Analysts (15-1122), Database and Systems Administrators and Network Architects (15-1140), Database Administrators (15-1141), Network and Computer Systems Administrators (15-1142), Computer Network Architects (15-1143), and Computer Network Support Specialists (15-1152). We also include the job postings if they in the following 2018 SOC occupations: Computer and Information Analysts (15-1210), Computer Systems Analysts2018 (15-1211), Information Security Analysts (15-1222), Computer Network Support Specialists (15-1231), Database and Systems Administrators and Network Architects (15-1240), Computer Network Architects (15-1241), Network and Computer Systems Administrators (15-1244), or Database Administrators and Architects (15-1245).

[18] Specifically, we include all skills that fall into the following BGT skill clusters: Cyber Security, Network Security, Technical Support, Database Administration, Data Management, Information Security, Application Security, and Internet Security.

3.2 Data on Firms Data Breach Events from Privacy Rights Clearinghouse (PRC)

The Privacy Rights Clearinghouse (PRC) sources their data primarily from the state Attorneys General and the US Department of Health and Human Services offices. Their data contains over 9,000 data breach events between 2005 and 2019 subject to the compliance of state data breach notification laws. In total, these breach events exposed over ten billion individual records. Besides the announcement date and source, their data also contains the names and location of the breached organizations. The listed breaches cover a wide range of industries and types of organizations including business, non-profits, and government agencies. More importantly, they report a wide variety of breach incident types including breaches of digital information due to Hacks or Malware, Insider Trading or Credit Card Fraud, but also other breaches due to Unintended Disclosures or Physical Loss. Roughly a third of the breaches in the data are due to breaches of digital information, i.e. cyber-related.

3.3 Matching BGT Hiring Data with PRC Data Breach Events

Since the PRC and BGT databases do not share a common firm identifier, we take a multi-step approach to merge the two databases. Specifically, we first clean and standardize the firm name strings in both databases. Next, we use a combination of name and address fuzzy matching to construct a bridge between the PRC and BGT data. After algorithmic matching, we manually validate and check all possible high-quality matches to further increase the match rate. Overall, we identify and match over 50% of organizations from the PRC data in the BGT data. We further require that organizations (both in the control and treatment group) have at least 100 job postings in the entire BGT data (i.e. 2010 - 2020) to ensure sufficient quality of the job posting data - however, our results are robust to different cutoffs. Overall, our main sample contains a total of over 83,000 organizations and over 1,800 data

breach events.[19] On average, organizations in our sample have 9 job postings per month with about 0.25 of them related to cybersecurity occupations in a given month.

One limitation of this merged data for our estimation purpose is that cybersecurity hiring may occur through outsourcing or contracting through an outside firm. Specifically, a firm may hire a contractor or a third part vendor and/or conduct on the job training to improve their cybersecurity. This outsourcing behavior is not observable through job postings attributable to the firm. These data limitations are likely to create a downward bias on our estimates which we discuss in detail later in the results section.

3.4 Data on Firms Media and Public Attention from MIT MediaCloud and Google Trends

To further investigate the impact of public visibility of firms' data breaches on firms' hiring responses, we gather monthly data from two different sources: (i) the MIT Media Cloud Project and (ii) Google Trends.

The MIT Media Cloud project aggregates data from over 50,000 news sources and offers an Explorer API, which returns media attention for specific search phrases going back until 2011. We use this tool to derive both absolute and relative monthly media attention for the same search queries.[20]

Similarly, Google Trends offers an API to download monthly search indices for specific search phrases going back until 2004. Google normalizes these indices to range between 0 and 100 based on a representative sample of all Google searches within a specified geography and time frame. Given the large number of search Google handles every day, this data

[19] Many matched organizations in the PRC data are smaller local business (e.g., local clinics) and hence dropped due to the 100 job posting cutoff.

[20] One might be concerned that media attention is solely a function of the breach size. Indeed, the correlation between number of records breached and the number of media articles in the month of the breach is high. However, further analysis finds that this high correlation is only driven by the catastrophically large breaches. When excluding breaches with greater than one million records (21 breaches, and only 5 percent of our treatment sample), the correlation between media attention and records is 0.0884, suggesting that media attention is likely a distinct channel.

represents the public interest in a given topic very well and is particularly well-suited to identify spikes in interest (Baker and Fradkin 2017).[21] We therefore use Google Trends to get monthly US public interest since 2010 for the queries '<firm name>' as well as '<firm name> data breach' for all firms that suffered a data breach in our data.

3.5 Summary Statistics

Table 1 presents the summary statistics of the key variables in our main sample. Panel A presents summary statistics for the 'Never Treated', 'Treated', and 'Overall' samples respectively. Our final sample includes 1,435 firms that experienced any form of data breaches (i.e., "Treated"). Among them, 1,261 have posted at least one job demanding cybersecurity experts (as defined by the cybersecurity occupations listed above), and 1,118 have posted at least one job of legal or public relationship occupations. There are 87,628 firms in our sample never had data breach with 58,565 posted cybersecurity jobs (i.e., "Never Treated" - our control group). Among the breached firms, 96% of them are in service-providing sector where only 87% of the never breached firms on in service-providing sector.[22] Publicly traded firms take a larger proportion among the breached firms (7.6%) compared to the never breached firms (2.0%). Firms in the control group on average posted 10.6 jobs every month with 0.3 jobs looking for cybersecurity talents. Firms in the treated group posted 129.2 jobs a month with 3.2 jobs asking for cybersecurity. This is not surprising as the majority of the firms experienced data breaches are big firms. In contrast, never breached firms on average posted 0.07 jobs every month looking for legal and PR talents while treated firms posted 0.7 jobs in these areas.

Panel B of Table 1 presents the statistics of the treated firms comparing six months before the data breaches and the six months after. The average monthly job postings of breached firm increases 26% after data breaches and the monthly cybersecurity job postings

[21] See https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8

[22] As mentioned earlier, we follow the BLS definitions of goods-producing and service-providing industries.

increases 31%. To study the potentially amplifying effects of media and public attention on post-breach cybersecurity hiring, we also obtain data from MIT Media Cloud project, which tracks media mentions, as well as Google Trends, which tracks public search interest. As Panel B shows, using only the firm names as the key words, both the media coverage and the Google search trend do not change much after the data breach events. However, if we include the word "breach" in the key words, the media coverage and the google search trend nearly triple and septuple in magnitude respectively, implying a significant rise in public attention.[23]

4 EMPIRICAL METHODS

4.1 Difference-in-Differences Estimation

We apply a DiD research design to identify the impact of data breached on firms' human capital investment through our observational data - experimental settings with randomized controlled trials (RCT) would be unrealistic (concerning cost and ethic) to yield generalizable conclusions. We argue that the *timing* of data breaches is 'as-if' randomly assigned (i.e., firms can not perfectly predict the timing of the data breaches) and thus enables this type of quasi-natural experiment setting. This in turn allows us to leverage (DiD) analysis where we compare firms' hiring behavior for relevant occupations (and skills) before and after the breached events between Treated and Never Treated firms. There are no strategic benefits for firms to delay their investment responses after discovering a data breach.[24]. Thus, firms that experience a data breach (i.e., "treated firms") should exhibit posting patterns similar to those that don't (i.e., "control firms") prior to the breach date.

In a traditional DiD setting, there is a single time at which all treated units are treated.

[23]We will describe this data in more detail after presenting our main results.

[24]However, firms may have strategic benefits to delaying the data breach *announcement* to soften the media or shareholder impact. While the US data breach notification laws may mitigate these considerations, we explore potential timing heterogeneity around the breach announcement dates

Thus, if latent confounders coincided with the treatment, one could not disentangle the effect of the treatment from the confounders without additional data on untreated, or control, units. In our case, treatment, i.e. suffering a data breach, is staggered and happens at different times for different firms. This alleviates the concern of latent confounders, since they would have to coincide with multiple treatment times. Additionally, for organizations that experience multiple data breach events, we focus on just the first event in our data to avoid multiple treatment levels, which would otherwise violate the Stable Unit Treatment Values Assumption (SUTVA).

Formally, our baseline model is specified empirically as follows:

$$Job_{i,j,t} = \beta_0 + \beta_p D_{i,t} + \lambda_i + \lambda_y + \lambda_{m,j} + \varepsilon_{i,j,t} \tag{1}$$

where for each firm i, in two-digit NAICS industry j and month t, Job_{ijt} is a binary variable indicating whether the firm posts job vacancies for cybersecurity occupations or, in other specifications, for legal or PR occupations. In subsequent skill level analyses, it indicates whether a particular firm requests job vacancies that mention skills related to cybersecurity or legal or PR skills. $D_{i,t}$ is the indicator variable for whether the observed month is after the data breach events.[25] Additionally, we also inlcude the following fixed-effects in the regression:

- λ_i: firm-fixed effects (to control for time-invariant firm heterogeneity)
- λ_y: year-fixed effects (to control for changes in average posting behavior overtime)
- $\lambda_{m,j}$: calendar month of year by two-digit NAICS industry fixed effect (to capture time-varying industry-specific unobservable such as seasonality)

[25] Notably, in our preferred specification we exclude month 0, the month during which the breach event is announced

This is estimated as a fixed effects linear model.[26] Standard errors are clustered at the firm level. The coefficient β_1 thus captures the probability of posting additional jobs vacancies as a response after the data breach notification for firm i in industry j in month t. Both breached, i.e. treated, and control firms are in the sample, and the control firms help to estimate year-fixed effects (λ_y) and month-by-industry-fixed effects ($\lambda_{m,j}$) in the regression above.

One potential concern is whether firms would strategically delay announcement and thus affect announcement timing, which our identification strategy depends on. Although the US data breach notification laws could mitigate these considerations as shown in Figure 1, we further explore potential timing heterogeneity around the breach announcement dates.

4.2 Quarterly Dynamics

We also estimate the quarterly dynamics of cybersecurity hiring by comparing breached firms to the control group using the following strategy:

$$Job_{i,j,t} = \beta_0 + \sum_{p=-1}^{1} \beta_p D_{i,t+p} + \lambda_i + \lambda_y + \lambda_{m,j} + \varepsilon_{i,j,t} \tag{2}$$

where $D_{i,t+p}$ for $p \in [-1, 1]$ are indicator variables for the quarter relative to the data breach event. Quarter zero starts with the month of the announcement of the data breach event. The omitted category is two quarters before (-2) the data breach announcement. We chose to set the omitted category further back because we are currently using the notification date instead of the actual date of the breach events and there may be a delay between the two. Therefore, there lacks a clear cut on the treatment around the actual data breach events. To satisfy the parallel trends assumption of the DiD setting, firms that experience a data breach (i.e., "treated firms") should exhibit posting patterns similar to those that don't

[26] Results estimated using a fixed effects Poisson model are also reported.

(i.e., "control firms") prior to the breach date. We also performed additional regressions to mitigate potential timing and measurement issues, including 1) redefining the omitted category as the quarter immediately before the month of the data breach announcement; 2) estimating our specification with and without the announcement month or the preceding month (i.e., variation in announcement timing); 3) estimating a Poisson model to explicitly model the number of job postings as a non-negative count variable; 4) running a monthly estimation to examine the response timing more dynamically as well as parallel trends.

5 MAIN RESULTS

5.1 Effect of Data Breaches on the Probability of Substantive Hiring

We first examine whether firms respond to suffering a data breach by strengthening their cybersecurity workforce. Table 2 reports results using both posting probabilities (all columns excluding 3, 6, and 9) as well as counts as outcome variables (columns 3, 6, and 9) in linear regression models for easy interpretation.[27] In columns 1 and 2 of Table 2, the dependent variable is an indicator for whether the firm posted any cybersecurity jobs in that month. Both specifications focus on the time window consisting of the 6 months (2 quarters) before and after each data breach event. To mitigate potential confounding through firms' experience with data breaches as well as through latent firm heterogeneity, we use the first data breach event that firms ever suffered as the treatment and omit subsequent breaches of multiply-breached firms.[28] In both columns, we control for year-fixed effects, firm-fixed effects, and month-by-industry-fixed effects to account for general time-trends as well as potential time-invariant firm-level, and monthly-varying industry-level unobservables.[29]

[27] We also explore and report results from Poisson models in Appendix A

[28] As a robustness check, we also use the largest breach events, measured by the number of breached records, as the treatment and present these results in Table A1 in Appendix A.

[29] More specifically, we use calendar month by 2 digit NAICS fixed effects to address industry-specific seasonality.

Column 1 of Table 2 presents the result for a two period model in which the independent variables include an indicator for whether the observed month is before or after the data breach event became public, while in column 2, we decompose the time horizon to the quarterly level. The coefficient in column 1 indicates that firms increase their probability of posting cybersecurity jobs by 2.1 percentage points more after a breach, compared to the control group over the same time window. This effect is statistically significant at the 1% level.

Column 2 of Table 2 regresses the indicator variable for posting any cybersecurity jobs on quarterly indicator variables with the base case being two quarters prior to the data breach. This model allows us to better identify the timing of the treatment effect as well as test the pre-trend assumption necessary for a valid Difference-in-Differences model.[30] The coefficients show that before suffering a breach, breached firms do not hire significantly differently for cybersecurity, nor any other types, of job posting compared to non-breached firms. In the first quarter after the breaches (Quarter 0), the probability is 1.3 percentage points higher for breached firms than non-breached ones. It further increases to 3.4 percentage points in the second quarter (Quarter +1) after the breaches, which shows that the effect is most pronounced three months after the breach events. These results suggest a lag between firms' announcements of data breach events and the actual response of recruiting cybersecurity talent.

Thus far, we have reported the effect of data breaches on firms' cybersecurity-related hiring on the extensive margin, by reporting results with binary outcome indicator variables that equal one if firms posted any cybersecurity-related jobs and zero otherwise. We are also interested in the intensive margin of the effect.[31] We thus reported the results from the specification that employs the number of job postings of cybersecurity as the left-hand side

[30] We choose quarterly instead of monthly analysis to hedge against larger measurement error issues at the monthly level due to the job posting data. However, our results are robust when using the monthly-level analysis. These results are reported in the Appendix B.

[31] Our data limits our ability to observe and capture multiple hirings from single job postings or failed hiring attempts and hence substantially enlarge potential measurement error when measuring labor demand in absolute levels. Thus, our preferred specifications are those using binary outcome variables.

variable in column 3. The results indicate that the breached firms on average post 0.4 more jobs after the breach events compared to the non-breached firms.

Additionally, we conduct a series of robustness tests concerning the announcement timing, breach event severity, an alternative regression model targeting categorical outcome variable - Poisson model. As shown in Figure 1, data breach notification laws in some states require firms to report such events within 30 to 60 days, or one to two months, of noticing their occurrence. Therefore, the treatment is not sharp at the month zero as there could be strategic timing decision by firms on when to announce the breach event.[32] In order to reduce the noise due to the ambiguity of the treatment time, in column 1 of Table 3, we drop the observations from month zero and the month before but include two more earlier months to the sample. The results in this column show the breached firms increase the probability of posting cybersecurity jobs by 2.2 percentage points, indicating that our result is robustness to the potential measurement error in the breach date. Column 2 of Table 3 presents the results from a similar regression but with the quarter prior to the data breaches as the omitted period. The result is robust here as well.

Another potential concern is that firms might respond more strongly to severe data breach events (i.e., larger number of records breached). To address for this issue, instead of looking at each firm's first breach event ever recorded, we employ the events with the most breached records and show that our findings are robust to this test. These results are reported in Appendix Table A1. Lastly, we also performed the Poisson regression and present the result in column 3 of Table 3. The result shows that breached firms post 1.1 more jobs after the breached events than non-breached firms. In the Poisson regression, we include the calendar month fixed effects and 2 digit NAICS industry fixed effects rather than the month-by-industry fixed effects as the latter exceeds our computing power.

Overall, the results in Table 2 are robust to a wide range of tests and support our

[32]We use the data breach announcement date as the treatment date for our baseline specification. While this is neither the exact date of the breach nor the date on which the firm notices, we consider that it is still the best date available.

Hypothesis 1, suggesting that breached firms do increase substantive adoption by hiring cybersecurity talents. Although statistically significant, it is worth noting that the economic magnitude of the identified effect seems small at 2.1 percentage points. This result suggests that although breach events force certain firms to take actions, the incentive might still be insufficient for others to take the substantive adoption of acquiring cybersecurity talents and treat the cause of data breach threats.

However, data breach incidents may substantially damage their public impression, cause class-action lawsuit, and deteriorate future performance. Therefore, firms might have large incentive to take actions targeting these challenges by acquiring legal and PR talents after a data breach. We further investigate this in the following section.

5.2 Effect of Data Breaches on the Probability of Symbolic Hiring

The loss of valuable data to outside parties can lead to major challenges for firms beyond those of a vulnerable cybersecurity infrastructure. This may include public scrutiny or lawsuits, which may be easier to deal with than the underlying cybersecurity issues. Therefore, besides, or perhaps instead of, strengthening their IT infrastructure and IT labor force, firms may demand public relations or legal talent in order to manage their brand image and counter negative media attention, or to assure compliance with applicable privacy laws, regulations, and constitutional requirements, as stated in Hypothesis 2. While these types of responses are laid out in the NIST framework, these types of symbolic, instead of substantive, adoption do not directly fix the cause, the cybersecurity vulnerability, but only the temporary symptoms of the latest data breach. In columns 4 to 6 in Table 2, we estimate the effect of data breaches on firms' symbolic adoption of legal and public relation talents.[33]. Following equation 1 and section 5.1, the dependent variable in columns 4 and 5 is an indicator for whether the firm posts any job vacancies for legal or public relation occupations.

[33]Similar to 5.1, we use the firms' first data breach events as the treatment and omits further breaches of multiple-breached firms. We also use the events with the most breached records as the treatment as a robustness test and present the results in Table A1 in Appendix A

The result in column 4 shows that breached firms are 2.1 percentage point more likely to hire legal or public relation talents. This finding does not just confirm our Hypothesis 2 that firms increase symbolic adoption to deal with the symptoms of a data breach but also indicate an effect with a magnitude that is similar to the substantive adoption of cybersecurity workers. In parallel to Column 2, column 5 shows that the timing for these symbolic hiring attempts are similarly delayed into the second quarter after the data breach events, potentially suggesting a joint of both substantive and symbolic responses. In particular, breached firms are 1.5 percentage points more likely to hire legal and PR talents during the first quarter (Quarter 0) immediately after the event, and are 4 percent points more likely to take the action during the second quarter (Quarter +1). However, although results presented in Columns 4 and 5 show a positive effect on extensive margin, the intensive margin of the effect on symbolic hiring is not statistically significant from zero. [34]

Similar to Section 5.1, we also performed additional robustness checks in Appendix A. Column 4 drops the observations from month zero and the month before but includes two more earlier months to the sample because the exact dates of the breach events may differ from the notification dates. It shows a similar result as Column 4 in Table 2. We also present the results from the quarterly specification while omitting the quarter immediately before the data breaches. Such results shown in Column 5 of Table 3 indicate a lagged action in the second quarter after the data breaches from the breached firms, which concludes a similar timing of action presented in Column 5 of Table 2.

5.3 Placebo Tests on Non-Relevant Jobs

Although we are able to provide supporting evidence that data breaches have a significant positive effect on firm's substantive and symbolic hiring, there still can be concerns that the identified increased demand might be caused by other random shocks and may not be limited to cybersecurity, legal and PR talent. Therefore, we also performed a set of placebo

[34] We also performed the Poisson regression and present the result in column 6 of Table 3. The Poisson regression also shows non-significance for symbolic hiring.

tests to investigate whether such effect also applies to occupations that are not relevant to either substantive adoption or symbolic adoption. We present the results in columns 7 to 9 in Table 2.[35]

Similar to Section 5.1 and 5.2, We also use the first data breach events as the treatment and omit the future events for firms with multiple breach events. When estimating the effect of data breaches on the probability of hiring non-relevant talents, both the two-period model (Column 7) and the quarterly specification (Column 8) show that breached firms do not increase their demand for talents that are not relevant with solving any issued after the data breaches.[36] The intensive margin presented in Column 9 also shows data breaches have no effect on firms' talent demand on non-relevant occupations. Overall, the results presented in these three columns show that, the increased demand on talents of cybersecurity, legal, and PR, as seen in the previous sections, are not driven by other firm-level shocks that might increased the overall human capital demand following the data breaches, suggesting the identified effects are likely causal.

5.4 Demand for Related Skills

An alternative and more elaborate way to measure firm's response to data breaches is through their change of demand for related skills instead of occupations. Therefore, in addition to SOC-based definitions of cybersecurity, legal, and PR job postings, we can also capture such demand through analogous skills that each job posting demands. There are 122 unique skills among the roughly 16,000 skills in the BGT taxonomy that are nested within the cybersecurity skill cluster. Therefore, in addition to SOC-based definitions of cybersecurity, legal, and PR job postings, we can also capture such demand through analogous skills that each job posting demands. Using this skill-based definition, we then investigate whether firms recruit more talent with cybersecurity skills, symbolic skills, or both after experiencing

[35] We define non-relevant hiring as any job posting from the firm that is not under cybersecurity occupations nor under legal or PR occupations.

[36] Using the events with the most breached records as the treatment, Column 3 of Table A1 concludes with a similar result.

a data breach. The results are presented in Table 5. Similar to Table 2 and related sections shown above, we also use the first breach events as the treatment.[37] Columns 1 and 2 present the effect of data breach on firm's demand of cybersecurity skills; Columns 3 and 4 present the effect on firm's demand of legal and PR skills; and Columns 5 and 6 present the effect on the demand of skills that do not require any relevant skills to data breach or cyber attacks.

The left-hand side variable of Columns 1 and 2 in Table 5 is the binary variable indicating whether the firm demanded any cybersecurity skills. We define the variable equals one if the firm posted any job that acquired any skill in the BGT cybersecurity skill cluster in that month, regardless the occupation code related with the job posting. The coefficients in these two columns are similar as those reported in Table 5. After the breach event treatment, firms are 1.6 percentage point more likely to demand any cybersecurity skills, as shown in column 1.[38] In addition, the quarterly analysis shows that the hiring reactions of breached firms are initially weak during the first quarter after the data breach event (1.2 percentage points and statistically significant at 10% level), but become stronger and more statistically significant during the second quarter after the breach (2.7 percentage points). This is consistent with the results reported in Section 5.1.[39]

Columns 3 and 4 of Table 5 present the results for PR and legal skills that follow the similar layout of columns 1 to 3. The results show that, firms who experience their first data breach events are 1.7 percentage points more likely to demand PR and legal skills, regardless the occupation requirements. This is similar to the magnitude of the effect on cybersecurity skills reported in Column 1. The quarterly dynamics presented in Column 5 also shows a similar pattern as cybersecurity skill in Column 2, where the effect of the data breach on the demand of PR and legal skills mainly started to show up in the second quarter after the

[37] For robustness check, results using breach events with the most lost records are presented in the Appendix and Table A2.

[38] The results are similar (e.g., 2.0 percentage point) if we use firms' data breaches with the largest number of breached records as the treatment (reported in Table A2).

[39] The robustness checks including omitting the month of breach event and one month before and omitting the quarter immediately before the announcement of the data breaches are also performed. Results are similar to those presented above and are shown in Columns 1 and 2 of Table A3.

events. However, the intensive margin presented in Column 6 does not show a statistically significant result, which is also consistent with the results reported in column 6 of Table 2.[40]

6 ADDITIONAL RESULTS AND EFFECT HETEROGENEITY

6.1 Detailed Decomposition by Occupations

Not all occupations related to cybersecurity receive the same attention from breached firm after suffering a breach event. For instance, according to the FBI, "The notion that you can protect your perimeter is falling by the wayside & detection is now critical," which suggests that firms focus more on detection rather than protection.[41] We utilize the richness of our data to identify demand heterogeneity for specific types of cybersecurity occupations. Table 6 presents the regression coefficients from model 1 for this test. The results show that breached firms attempt to acquire more information security analysts, computer system analysts, data administration experts, and network and computer systems administrators after the breach events. They are presented in Columns 1, 2, 3 and 4, respectively. In contrast, no significant effect is found for network support specialist and computer network architects as shown in Columns 5 and 6.

6.2 Effect Heterogeneity by Data Breach Type

Data provided by Privacy Rights Clearinghouse categorizes breach events into seven types. To further investigate how firms react to different types of data breach events, particularly those types that are related to cybersecurity, this study focuses on the ones that are most

[40] Robustness checks presented in Column 2 of Table A2 and Columns 4 to 6 of Table A3 show that the results are robust across different selection of treatments, different treatment time, or different omitted period.

[41] See more details at https://web.archive.org/web/20150420211301/http://blog.norsecorp.com/2015/03/12/fbi-official-says-prepare-for-more-damaging-cyber-attacks/

likely to affect firms' skill demands in cybersecurity. We first look at the effect on the probability of cybersecurity hiring after suffering a loss of digital data, which is captured by the breach types HACK (hacked by outside party or infected by malware) and CARD (fraud involving debit or credit cards not via hacking). The results are presented in column 1 of Table 7. It shows that the probability of breached firms to hire cybersecurity talents is 3.6 percentage points higher than that of non-breached firms. The magnitude of the coefficient is thus higher than the coefficient that appears in column 1 in Table 2.

For comparison, we also investigate how firms react to data breach events in which only physical data was breached, as captured by the breach types PORT (loss of portable devices) and PHYS (loss of physical documents). The result presented in Column 2 of Table 7 suggests that firms do not take actions to increase their probability of hiring cybersecurity talents after the non-cyber events.

We also look at the effect on the PR and legal jobs after these two types of data breach events as presented in Columns 3 and 4 in Table 7. The coefficient in Column 3 suggests that firms on average put 0.22 more job postings demanding PR and legal talent after a cyber event while no effect is present for non-cyber events as shown in Column 4. The results show that firms who suffered a loss of digital data through HACK or CARD are likely to hire more PR and legal talents, though the magnitude is lower compared to those for cybersecurity. Similarly, data losses through PORT and PHYS has no effect on firm's hiring of PR and legal talents.

These tests not only reveal the types of the data breach events that firms respond to, but also serve as another set of placebo test further strengthen our causal argument about hiring in response to data breaches.

6.3 Effect Heterogeneity by Industry Sector

In order to study how firms in different industry sectors react to data breaches, we follow the U.S. Bureau of Labor Statistics (BLS) to assign firms to two broadly defined supersector

groups: goods-producing industries[42] and service-providing industries.[43] The results of the coefficients are presented in Panel (a) of Figure 2. The coefficients plotted here are from eight separate regressions and the X-axis represents the change in probability of posting cybersecurity job postings after a data breach event. Results from substantive hiring (i.e., cybersecurity-related) and symbolic hiring (i.e., legal and public relation related) are plotted by black and grey lines, respectively.

In Panel (a), the coefficient in the top row is the benchmark from the regression model of Columns 1 and 4 from Table 2. The second row shows that the estimated coefficients for firms in goods-producing industries are negative and noisy, suggesting that these firms do not take any action on talent hiring after a data breach. In contrast, as shown in the third row, firms in service-providing industries do react to data breaches in both substantive and symbolic adoptions. Lastly, we find similar results when excluding firms from the information sector (NAICS 51) as it includes IT service providers who may react to the breach events incurred by other firms, as shown in the bottom row. Perhaps more interestingly, the results of symbolic hiring are almost identical to substantive hiring, suggesting that firms from service-providing industries are also more likely to respond to data breaches with increased symbolic hiring.

The results presented in Panel (a) of Figure 2 suggest that the reaction to data breaches is almost exclusively driven by firms in the service-providing industries in our sample. This is consistent with our argument that the nature of the data collected may be significantly different across industries. Data collected by firms in the service-providing industries typically consists of highly sensitive and private individual customer information, such as customer social security and credit card numbers. Exposure of such data can put breached firms' clients under significant public scrutiny, increase the chance of large class action lawsuits.

[42]See https://www.bls.gov/iag/tgs/iag06.htm (last accessed on Feb 17, 2021). The goods-producing industries include Agriculture, Forestry, Fishing and Hunting (NAICS 11), Mining, Quarrying, and Oil and Gas Extraction (NAICS 21), Construction (NAICS 23), and Manufacturing (NAICS 31-33).

[43]See https://www.bls.gov/iag/tgs/iag07.htm (last accessed on Feb 17, 2021). The service-providing industries consists of all industries that not included in goods-producing industries: NAICS 42 - NAICS 92.

The consequences can be catastrophic for these firms where customers' trust is particularly important and thus put the firms under pressure to fix the problem and regain customers' trust.

However, firms in the goods-producing industries are more likely to collect the data internally (e.g. production and inventory data) and/or on firm customers, and are therefore are less likely to face such public pressure. Such comparison between the service-providing industries and goods-producing industries suggests that when firms face external pressure, they are more likely to take actions.

6.4 Effect Heterogeneity by Ownership Type

To further understand whether public scrutiny can incentivize firms to respond to a data breach, we look at the effect heterogeneity between public and private firms. Public firms tend to be under significantly higher public scrutiny through investors and regulators as well as under a higher reporting burden (e.g., the SOX Act cybersecurity compliance), which likely puts significant pressure on them to take substantive as well as symbolic action after suffering a data breach event. By applying a crosswalk between data on public companies from the Compustat and BGT data through a fuzzy name matching algorithm, we identified 3390 public firms in our data.[44] While imprecisely estimated due to the small sample size of treatment group for public firms, Panel (b) of Figure 2 shows the probability of substantive adoption (i.e., posting cybersecurity jobs) increased following a data breach event for both public and private firms.[45] Similar patterns can be found in the right panel when exploring the posting probability of symbolic jobs (i.e., legal and public relations). Consistent with our expectation, the results in Figure 2 indicate that public firms are more likely to increase substantive and symbolic talent demand after data breach events compared to private firms and thus supports our Hypothesis 3b.

[44]We use the crosswalk between the BGT and Compustat database from Bai et al. (2021) to distinguish public firms from private firms in our sample.

[45]The estimated coefficient on the effect of substantive adoption for public firms is only statistically significant at 10% level because of the small sample size of treatment group.

6.5 Discussion

Although the identified effects on cybersecurity hiring are statistically significant in general, the economic magnitudes are generally small and may be far from socially optimal. One particular concern here is that our identified effect could be significantly downward biased if most firms outsourced their cybersecurity needs after the data breaches. However, although firms who outsource their cybersecurity personnel may not react to data breaches by investing in related human capital, they should potentially instead invest heavily in software and infrastructure investment, which is not supported by the findings of prior studies.

Additionally, our skill-level tests (instead of occupations) further address the concern with better granularity than conventional studies with IT labor data. Even if firms outsource their core cybersecurity tasks, our skill-level tests allow us to observe whether firms hire employees with any cybersecurity skills outside cybersecurity occupations. Arguably, firms are likely to still need to have internal employee(s) with certain cybersecurity talent to coordinate with their cybersecurity vendors, especially when data is essential to their operation or with high sensitivity.[46]

Combined with insufficient financial investments in cybersecurity found in the earlier studies (Gordon et al. 2015a,b, Manworren et al. 2016), the relatively small effect that we identify on firms' human capital investment in cybersecurity is more likely to reflect market failure due to misaligned incentives and information asymmetries between consumers and firms (Moore 2010) than from suffering a downward bias. We investigate the role of public scrutiny to highlight data breach events as an effective intervention next.

[46] Another particular concern that we cannot address due to data limitation is the probability of on-the-job cybersecurity training after the data breaches. However, this activity, if true and if large enough, would also increase firms investment level and should be captured by the estimates in the earlier studies.

7 PUBLIC VISIBILITY AS A MECHANISM

Our results highlighting differences between public and private firms may be related to public firms exhibiting more strategic disclosure behavior in response to a data breach, outlined in Foerderer and Schuetz (2022). Firms' strategic behavior demonstrates that media and public attention are important sources of attention. Indeed, Foerderer and Schuetz (2022) find that the magnitude of this effect is meaningful and economically significant: if news media is one standard deviation above the mean, implying the firm will receive less scrutiny, the loss of market capitalization for the median firm reporting a breach goes from $347 million to $85 million.

Potentially, media and public attention following the breach have direct effects on job posting behavior, just as they do on stock prices. To further test this mechanism formally, we estimate the main specification, subsampling using the level of media attention and Google Search Trends related to the breach as measures of public visibility. We first describe the measurement of these concepts, followed by the results.

7.1 Measuring the Change of Visibility Related to Data Breaches

To assess the effect of visibility on hiring behavior, we first need to construct the measures of media and public attention. Because firms may experience substantially different media and public attention unrelated to cybersecurity, we must control for pre-period visibility in each context. In contrast to the comparison between public and private firms, these measures allow us to construct the change of public visibility before and after the data breach. We thus define our breach visibility as the increased visibility in the month of the breach according to these sources compared to the average visibility in the pre-period (-6 to -1).[47] This is a time-invariant measure for each firm experiencing a breach.

In order to analyze the effect of visibility around data breaches on hiring behavior, we

[47]We can also define breach visibility in the media attention data using the raw number of articles containing firm name and breach. The results are similar and available upon request.

split the sample into two groups, a high visibility group and a low visibility group. The high visibility group consists of firms that have breach visibility above a certain threshold, while the low visibility group consists of firms that have breach visibility below this threshold.

With these groups, we estimate two separate regressions: Equation 1, which limits the treatment group to only high visibility firms, and Equation 1, which limits the treatment group to only low visibility firms. In both regressions, the control group is the same – firms that have not experienced a data breach. What differs between the two specifications is the treatment group. Because breach visibility is defined to be time-invariant, this specification allows us to retain firm fixed effects for posting behavior, as we know these are crucial to the assumptions made for identification. We estimate the effect of data breaches on firms' human capital demand for both high visibility and low visibility events separately.

7.2 Effect Heterogeneity by Visibility

Figure 3 plots the change in the probability of posting substantive and symbolic jobs by high and low visibility breaches. We hypothesized that high visibility breaches would lead to an increased probability of substantive postings following the breach. This is what we find in Figure 3: high media attention breaches are associated with a 6.6 percentage point increase in the probability of posting a substantive role while there is no significant relationship with symbolic postings. On the other hand, low media attention breaches are associated with small increases in both substantive and symbolic postings. One interpretation of Figure 3 is that high visibility breaches incentivize firms to shift their hiring efforts towards cybersecurity roles, while firms experiencing low visibility breaches have no such incentive.

We repeat the exercise with Google Trends data to investigate the effect on talent hiring from different public attention levels and present the results in Figure 4. While it may be assumed that media shares and search shares are positively correlated, this correlation is empirically not very high ($\rho = 0.1184$), supporting visibility coming from two different channels.

Figure 4, measuring visibility by Google searches, demonstrates that a similar visibility pattern exists for public attention. High media attention is associated with a 6.0 percentage point increase in the probability of posting a substantive role, with no significant effect on symbolic postings. Firms do appear to react to greater visibility by posting substantive roles. Low visibility breaches, by contrast, are followed by similarly sized increases in both symbolic and substantive roles.

These figures demonstrate the role of visibility on firms' probability of posting substantive or symbolic jobs, treating these types of postings as independent. In Figure 5, we create an indicator for more or the same number of substantive postings compared to symbolic postings at the monthly level. This allows us to assess the relationship between visibility and the probability of more substantive postings than symbolic postings. Both media and public attention display the same relationship – high visibility breaches are associated with an increased probability of more substantive postings than symbolic postings. These effects are much smaller for low visibility breach. This result demonstrates that visibility can encourage firms to invest in a way that treats the cause of the data breach, as opposed to the symptoms.

The exercise has also been performed with an alternative time window, where we redefine breach visibility based on the month immediately after the data breach (as opposed to the month of the data breach). These results are similar and available upon request.

Consistent with the discussion in Foerderer and Schuetz (2022), visibility around the data breach affects the firm. In our case, we show that firms experiencing high visibility breaches, defined by both media and public attention, are much more likely to post substantive roles than symbolic roles.

8 CONCLUSION

As our economy gets more digitized, an increasing amount of data is being generated, stored, and distributed. Many organizations value and rely on data as a critical resource. Meanwhile, cybercrime targeting these data are becoming more common and sophisticated. This puts customers' privacy and public safety under risk and threatens our digital society. Although policy makers have been urging both public and private entities to take actions to strengthen cybersecurity against potential cyberthreats, how firms invest in cybersecurity and safeguard their data remains understudied. This paper brings together job vacancy and skill demand data from the BGT with data on breach events from Privacy Rights Clearinghouse to provide empirical insights into the human capital investments that firms make after a breach. In particular, we contribute to the literature and study firms' substantive adoption of cybersecurity talent in the labor market to treat the root cause, the cyber vulnerability, as well as their symbolic adoption of legal and PR workers that only treat the symptoms.

We apply a DiD strategy to study the effect of data breaches on hiring for cybersecurity and relevant talent at the occupation level. The results show that breached firms demand significantly more cybersecurity than non-breached firms starting three months after the data breach is made public, though the economic magnitude of the effect is relatively small. Falsification tests using analog data breaches (such as physical loss), robustness checks using skill-level data, and confirming the non-existence of pre-trends using the monthly dynamic test corroborate the causality of the identified effect. Taking advantage of the granularity of our database, we are also able to capture that data breach events increase the demand for specific skills in information security, computer systems, and database administration, among other skills related to cybersecurity. In addition and perhaps more interestingly, a similar effect is also identified for symbolic hiring. That is, firms are more likely to increase their demand in legal and public relation talents after experiencing a data breach. Therefore, Hypothesis 1 and Hypothesis 2 are both supported.

Next, we further explore the effect heterogeneity across industries. We find that firms in

service-providing sectors, such as the Information, Finance, Insurance Administrative and Support Services, as well as Retail Trade industries react through hiring attempts, while those in goods-producing sectors, such as Construction and Manufacturing, do not. This finding is true for both substantive and symbolic adoption. We suspect that firms in these service-providing industries are precisely those firms that are more likely to deal with sensitive customer data, such as financial information, and in which customer trust matters the most. Abundant anecdotes also suggest that these firms are more likely to suffer from negative public relationships and class actions (e.g., Equifax, Capital One, and T-mobile).

Lastly, we conduct extensive mechanism tests to explore whether public scrutiny could incentivize firms to invest more in both substantive and symbolic human capital. We first compare private firms with public firms, as the latter are more likely to face regulatory and public scrutiny. Our results indicate that public firms are more likely to post (and post relatively more on average) cybersecurity jobs compared to their private peers. Next, we use data from Google Trends and the MIT Media Cloud to proxy for the public scrutiny that firms face before and after a data breach event. In so doing, we are able to capture the changes of public attention towards to firms with data breach events. We find suggestive evidence that firms with higher public scrutiny are more likely to respond to data breaches by acquiring cybersecurity talent, but less so for legal and PR talent, compared to their low public-scrutiny peers. This suggests that public scrunity can serve as an effective mechanism to realign firms' incentives with social interests, as it increases the substantive adoption that firms engage in. Taken together, our findings suggest that firms may lack incentives to allocate a socially optimal level of investment in human capital to secure their data, thereby causing a loss of general welfare. General scrutiny such as those faced by large public firms (compared to smaller private firms) are less effective in motivate firms to invest in substantive human capital (i.e., cybersecurity talents) since these firms can game the announcement dates and strategically downplay the negative impact of a given data breach. However, when public attention such as Media coverage and public searches are data-breach specific, they help

motivate firms to respond to data breaches through their hiring in cybersecurity talent. With an increase in the value of data as well as the number of cybercrimes targeting customer data, our findings suggest that consumers, media, as well as government should work together to better informed the public and provide better incentives for firms to safeguard data, protect customers, and increase the social welfare in our increasingly digitized world.

References

Ablon L, Heaton P, Lavery DC, Romanosky S (2016) Consumer attitudes toward data breach notifications and loss of personal information. Technical report, RAND Corporation.

Acquisti A, Friedman A, Telang R (2006) Is There a Cost to Privacy Breaches? *ICIS 2006 Proceedings* 94.

Amir E, Levi S, Livne T (2018) Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of Accounting Studies* 23(3):1177–1206, ISSN 13806653.

Angst CM, Block ES, D'Arcy J, Kelley K (2017) When do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches. *MIS Quarterly* 41(3):893–A8, ISSN 02767783.

Athey S, Catalini C, Tucker C (2017) The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. *NBER Working Paper Series* ISSN 1098-6596.

August T, Niculescu MF, Shin H (2014) Cloud implications on software network structure and security risks. *Information Systems Research* 25(3):489–510, ISSN 15265536.

Aytes K, Byers S, Santhanakrishnan M (2006) The Economic impact of information security breaches: Firm value and intra-industry effects. *AMCIS 2006 Proceedings* 6:3293–3300.

Bai JJ, Brynjolfsson E, Jin W, Steffen S, Wan C (2021) Digital resilience: How work-from-home feasibility affects firm performance. Technical report, National Bureau of Economic Research.

Baker SR, Fradkin A (2017) The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. *The Review of Economics and Statistics* 99(5):756–768, ISSN 0034-6535.

Bana S, Brynjolfsson E, Rock D, Steffen S (2020) job2vec: Learning a representation of jobs .

Banker RD, Feng CQ (2019) The Impact of Information Security Breach Incidents on CIO Turnover. *Journal of Information Systems* 33(3):309–329, ISSN 0888-7985.

Barrero JM, Bloom N, Davis SJ (2021) Why working from home will stick. Technical report, National Bureau of Economic Research.

Bolster P, Pantalone CH, Trahan EA (2010) Security breaches and firm value. *Journal of Business Valuation and Economic Loss Analysis* 5(1), ISSN 19329156.

Bose I, Leung ACM (2013) The impact of adoption of identity theft countermeasures on firm value. *Decision Support Systems* 55(3):753–763, ISSN 01679236.

Bose I, Leung ACM (2019) Adoption of identity theft countermeasures and its short- And long-term impact on firm value. *MIS Quarterly: Management Information Systems* 43(1):313–327, ISSN 21629730.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics* 117(1):339–376.

Brynjolfsson E, Milgrom P (2013) Complementarity in organizations. *The handbook of organizational economics* 11–55.

Buckman JR, Bockstedt JC, Hashim MJ (2019) Relative privacy valuations under varying disclosure characteristics. *Information Systems Research* 30(2):375–388, ISSN 15265536.

Cezar A, Cavusoglu H, Raghunathan S (2014) Outsourcing Information Security: Contracting Issues and Security Implications. *Management Science* 60(3):638–657, ISSN 00251909.

Cheng C, Hoekstra M (2013) Does strengthening self-defense law deter crime or escalate violence? evidence from expansions to castle doctrine. *Journal of Human Resources* 48(3):821–854.

Choi JP, Jeon DS, Kim BC (2019) Privacy and personal data collection with information externalities. *Journal of Public Economics* 173:113–124, ISSN 00472727.

Collis A, Moehring A, Sen A, Acquisti A (2021) Information frictions and heterogeneity in valuations of personal data. *SSRN Electronic Journal* .

Deming D, Kahn LB (2018) Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics* 36(S1):S337–S369.

Ebrahimi S, Eshghi K (2022) A meta-analysis of the factors influencing the impact of security breach announcements on stock returns of firms. *Electronic Markets* ISSN 1422-8890.

Foerderer J, Schuetz SW (2022) Data Breach Announcements and Stock Market Reactions: A Matter of Timing? *Management Science* ISSN 0025-1909.

Garcia ME (2013) The economics of data breach: Asymmetric information and policy interventions.

Gordon LA, Loeb MP (2006) Budgeting process for information security expenditures. *Commun. ACM* 49(1):121125, ISSN 0001-0782.

Gordon LA, Loeb MP, Lucyshyn W, Zhou L (2015a) Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity* 1(1):3–17, ISSN 20572093.

Gordon LA, Loeb MP, Lucyshyn W, Zhou L (2015b) The impact of information sharing on cybersecurity underinvestment: A real options perspective. *Journal of Accounting and Public Policy* 34(5):509–519, ISSN 18732070.

Gordon LA, Loeb MP, Sohail T (2010) Market Value of Voluntary Disclosures Concerning Information Security. *MIS Quarterly* 34(3):567–A2, ISSN 02767783.

Greenwood BN, Vaaler PM (2021) Do US State Breach Notification Laws Decrease Firm Data Breches?

Havakhor T, Rahman MS, Zhang T (2020) Cybersecurity Investments and the Cost of Capital.

Hilary G, Segal B, Zhang MH (2016) Cyber-Risk Disclosure: Who Cares? *SSRN Electronic Journal* 1–59.

Huang K, Madnick S (2020) A Cyberattack Doesn't Have to Sink Your Stock Price. *Harvard Business Review* .

Huang K, Siegel M, Madnick S (2018) Systematically Understanding the Cyber Attack Business : A Survey. *ACM Computing Surveys* 51(4).

Hui KL, Kim SH, Wang QH (2017) Cybercrime Deterrence and International Legislation: Evidence from Distributed Denial of Service Attacks. *MIS Quarterly* 41(2):497–A11, ISSN 02767783.

Hui KL, Vance A, Zhdanov D (2018) MIS Quarterly Research Curation on Securing Digital Assets. *MIS Quarterly Research Curations* .

Janakiraman R, Lim JH, Rishika R (2018a) The Effect of a Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer. *Journal of Marketing* 82(2):85–105, ISSN 00222429.

Janakiraman R, Lim JH, Rishika R (2018b) The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer. *Journal of Marketing* 82(2):85–105, ISSN 15477185.

Kankanhalli A, Teo HH, Tan BC, Wei KK (2003) An integrative study of information systems security effectiveness. *International Journal of Information Management* 23(2):139–154, ISSN 02684012.

Kwon J, Johnson ME (2014) Proactive Versus Reactive Security Investments in the Healthcare Sector. *MIS Quarterly* 38(2):1–3.

Lassébie J, Marcolin L, Vandeweyer M, Vignal B (2021) Speaking the same language: A machine learning approach to classify skills in Burning Glass Technologies data. *OECD Social, Employment and Migration Working Papers* (263).

Liu CW, Huang P, Lucas Jr HC (2020) Centralized it decision making and cybersecurity breaches: Evidence from us higher education institutions. *Journal of Management Information Systems* 37(3):758–787.

Manworren N, Letwat J, Daily O (2016) Why you should care about the target data breach. *Business Horizons* 59(3):257–266.

Moore T (2010) The economics of cybersecurity: Principles and policy options. *International Journal of Critical Infrastructure Protection* 3(3-4):103–117.

Murciano-Goroff R (2019) Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure?

Petersen R, Santos D, Smith MC, Wetzel KA, Witte G (2020) Workforce Framework for Cybersecurity (NICE Framework). *NIST Special Publication - U.S. Department of Commerce* .

Richardson VJ, Smith RE, Watson MW (2019) Much ado about nothing: The (lack of) economic impact of data privacy breaches. *Journal of Information Systems* 33(3):227–265, ISSN 15587959.

Romanosky S, Hoffman D, Acquisti A (2014) Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies* 11(1):74–104, ISSN 17401453.

Romanosky S, Telang R, Acquisti A (2011) Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management* 30(2):256–286.

Say G, Vasudeva G (2020) Learning from digital failures? the effectiveness of firms divestiture and management turnover responses to data breaches. *Strategy Science* 5(2):117–142.

Solove DJ (2007) "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review* 44(4):745–772, ISSN 00364037.

Tanimura JK, Wehrly EW (2015) The Market Value and Reputational Effects from Lost Confidential Information. *International Journal of Financial Management* 5(4):18–35

, ISSN 22295682.

Telang R (2015) Policy framework for data breaches. *IEEE Security and Privacy* 13(1):77–79, ISSN 15407993.

Turjeman D, Feinberg FM (2019) When the Data Are Out: Measuring Behavioral Changes Following a Data Breach. *Marketing Science* ISSN 1556-5068.

TABLES AND FIGURES

Table 1: Summary Statistics

Panel A: Full Sample

Variables	Never Treated	Treated	Overall
Average Number of Job Postings	10.584	129.171	12.495
Average Number of Cybersecurity Job Postings	0.296	3.168	0.342
Average Probability of Cybersecurity Job Postings	0.075	0.250	0.078
Average Number of Legal or PR Job Postings	0.066	0.746	0.077
Average Probability of Legal or PR Job Postings	0.069	0.261	0.072
Average Number of Non-relevant Job Postings	10.222	125.257	12.076
Average Probability of Non-relevant Job Postings	0.421	0.681	0.426
Good-Producing Industries (%)	12.512	4.476	12.339
Service-Providing Industries (%)	87.488	95.524	87.661
Private Firms (%)	97.986	92.358	97.895
Public Firms (%)	2.014	7.642	2.105
Number of Firms	87,628	1,435	89,063
Firms with Cybersecurity Postings	58,565	1,261	59,826

Panel B: Treated Only

Variables	Pre-treatment	Post-treatment
Average Number of Job Postings	164.407	206.731
Average Number of Cybersecurity Job Postings	3.709	4.858
Average Number of Legal or PR Job Postings	0.714	0.862
Average Number of Non-relevant Job Postings	159.984	201.011
Media Share of Firm Name	0.174	0.167
Media Share of Firm Name + Breach	2.266×10^{-4}	6.498×10^{-4}
Search Share of Firm Name	23.897	24.020
Search Share of Firm Name + Breach	0.255	1.714

Notes: Panel A describes the full sample for both treated (breached) and control (never breached) firms. We follow the BLS definitions of goods-producing industries and service-providing industries. Cybersecurity jobs are defined by the cybersecurity occupations listed in section 3.

Table 2: Effect of a Data Breach on Talent Acquisition

	Cybersecurity Jobs			PR and Legal Jobs			Other Jobs		
	Probability		Counts	Probability		Counts	Probability		Counts
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post Breach	0.021***		0.393***	0.021***		0.056	0.008		15.600
	(0.005)		(0.12)	(0.006)		(0.041)	(0.007)		(9.703)
Quarter (-1)		0.004			0.008			-0.002	
		(0.007)			(0.007)			(0.008)	
Quarter (0)		0.013*			0.015**			0.006	
		(0.007)			(0.008)			(0.008)	
Quarter (+1)		0.034***			0.039***			0.005	
		(0.008)			(0.008)			(0.009)	
Number of Firms	89,145								
R-squared	0.306	0.306	0.427	0.246	0.246	0.264	0.306	0.306	0.378

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome variable is listed at the top. This table shows the results for the first data breach in the 2010 to 2019 period, as documented by the PRC data. Number of firms: 89,063. Number of observations: 11,584,884. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Additional Robustness Checks for Effect of Data Breach on Talent Acquisition

	Cybersecurity Jobs			PR and Legal Jobs		
	Probability		Counts	Probability		Counts
	Drop Month 0, -1	Omit Quarter -1	Poisson	Drop Month 0, -1	Omit Quarter -1	Poisson
	(1)	(2)	(3)	(4)	(5)	(6)
Post Breach	0.024***		1.102**	0.024***		1.028
	(0.006)		(0.042)	(0.006)		(0.088)
Quarter (-2)		0.004			-0.008	
		(0.007)			(0.007)	
Quarter (0)		0.009			0.007	
		(0.006)			(0.006)	
Quarter (+1)		0.030**			0.031***	
		(0.007)			(0.007)	
R-squared	0.306	0.306		0.246	0.246	

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome variables are listed at the top. The specifications are similar to those that reported in 2. Number of firms: 89,063. Number of observations: 11,584,603. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Effect of a Data Breach on Whether Firms Hire More Substantive Than Symbolic Jobs

	Two Period Model (1)	Quarterly (2)
Post Breach	0.013**	
	(0.005)	
Quarter (-1)		0.003
		(0.006)
Quarter (0)		0.095
		(0.007)
Quarter (+1)		0.020**
		(0.008)
Number of Firms	89,063	
R-squared	0.284	0.284

Notes: The dependent variable in this table is the binary variable that equals one if the firm posts at least as many cybersecurity jobs as legal/PR jobs. Column 1 is a two period difference-in-difference model. Column 2 shows the coefficients on quarterly basis. All regressions control for firm-fixed effects, year-fixed effects, and calendar-month-of-year by two digit NAICS fixed effects. Number of firms: 89,063. Number of observations: 11,584,884. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of a Data Breach on Probability of Skill Demand

	Cybersecurity Skills		PR and Legal Skills		Not Relevant	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Post Breach	0.016***		0.017***		0.007	
	(0.006)		(0.006)		(0.007)	
Quarter (-1)		0.006		0.002		-0.007
		(0.007)		(0.007)		(0.007)
Quarter (0)		0.012*		0.008		0.003
		(0.007)		(0.008)		(0.009)
Quarter (+1)		0.027***		0.026***		0.003
		(0.008)		(0.009)		(0.009)
Number of Firms			89063			
R-squared	0.326	0.326	0.301	0.301	0.301	0.301

Notes: The dependent variables captures the probability of firms acquiring cybersecurity, PR, and legal skills instead of occupations. All regressions control for firm-fixed effects, year-fixed effects, and calendar-month-of-year by two digit NAICS fixed effects. Number of firms: 89,063. Number of observations: 11,584,884. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect on Demand of Specific Occupations under Cybersecurity

Variables	Information Security Analysts (1)	Computer Systems Analysts (2)	Database Administration (3)	Network and Computer Systems Administrators (4)	Computer Network Support Specialists (5)	Computer Network Architects (6)
Post Breach	0.008**	0.007*	0.010**	0.007*	0.001	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.003)
Firms	89,145	89,145	89,145	89,145	89,145	89,145
R-squared	0.216	0.237	0.204	0.188	0.128	0.191

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome is a binary variable for whether the firm posted a specific type of cybersecurity occupation following a data breach. All regressions control for firm fixed effects, year fixed effects, and calendar-month-of-year by two digit NAICS fixed effects. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Effect of a Data Breach by Data Breach Type

	Cybersecurity Jobs		PR and Legal Jobs	
Variables	Cyber (1)	Non-Cyber (2)	Cyber (3)	Non-Cyber (4)
Post Breach	0.036***	0.011	0.0215**	0.0103
	(0.009)	(0.010)	(0.010)	(0.0104)
Number of Firms	88064	88037	88064	88037
R-squared	0.305	0.305	0.245	0.245

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. In columns 1 and 2, the dependent variable is defined as a binary indicator for whether the firm posts any cybersecurity jobs. In columns 3 and 4 the dependent variable is defined as a binary indicator for whether the firm posts any PR or legal jobs. In columns 1 and 3, the types of breaches are limited to cyber breaches as defined by the Privacy Rights Clearinghouse: Fraud involving debit and credit cards not via hacking (skimming devices at point-of-service terminals, etc.), hacks by an outside party or Infections by malware. In Columns 2 and 4, the breaches are of non-cyber types, including: Loss of physical (paper documents that are lost, discarded or stolen) and portable devices (lost, discarded or stolen laptop, PDA, smartphone, memory stick, CDs, hard drive, data tape, etc.). All regressions control for firm fixed effects, year fixed effects, and calendar-month-of-year by two-digit NAICS industry fixed effects. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

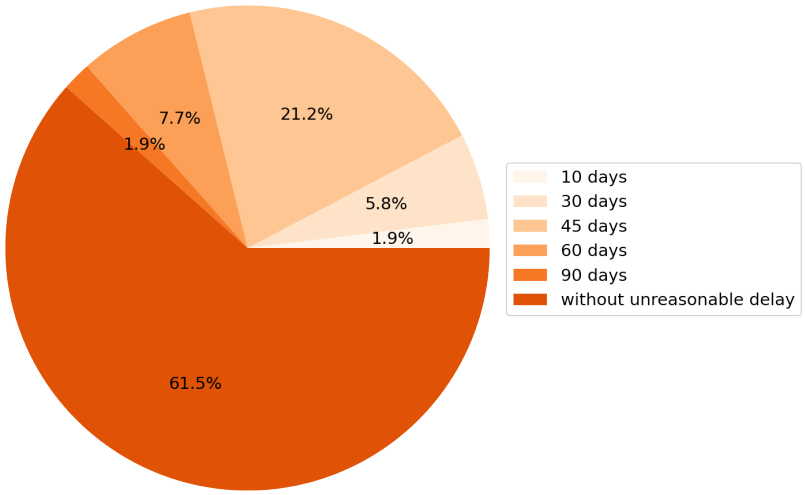
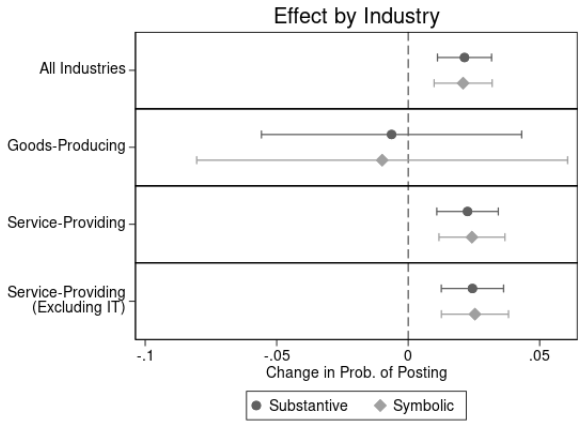


Figure 1: Timely Notification Requirements by States' Data Breach Notification Laws

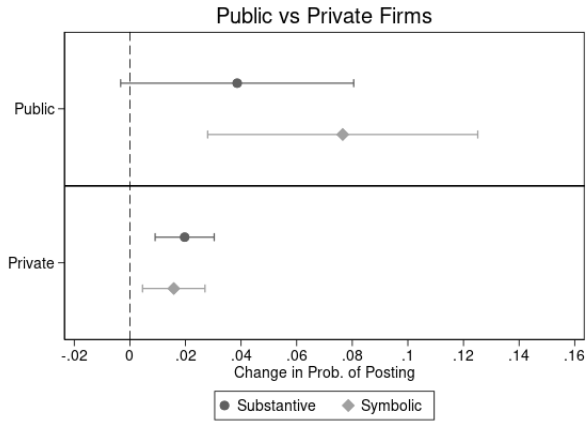
Notes: Figure derived from Perkins Coie Security Breach Notification Chart – Revised June 2020, available at https://www.perkinscoie.com/en/news-insights/security-breach-notification-chart.html. About 60% of US states require breached firms to notify the public as soon as they realize that they were breached. In total, 98% of all US states require firms to notify the public no longer than 60 days after suffering a data breach.

Figure 2: Effect Heterogeneity of Data Breach on Firm's Talent Acquisition

(a) By Industry

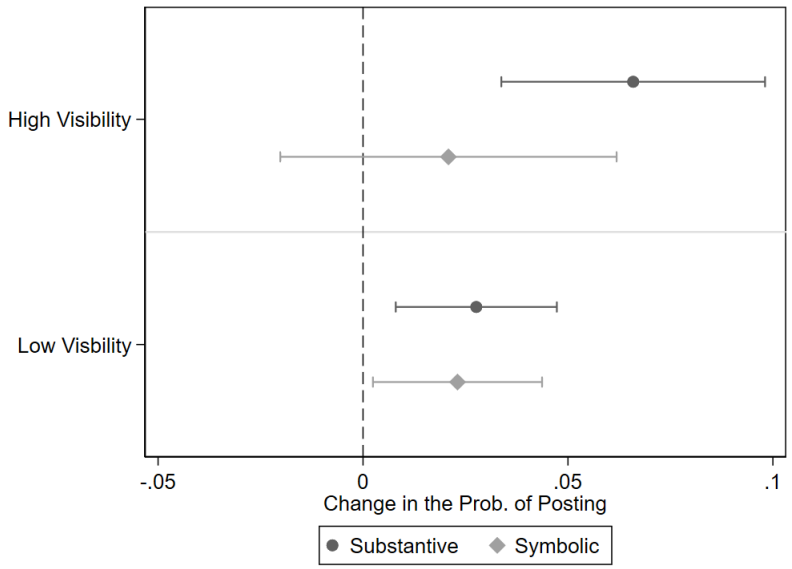


(b) By Ownership



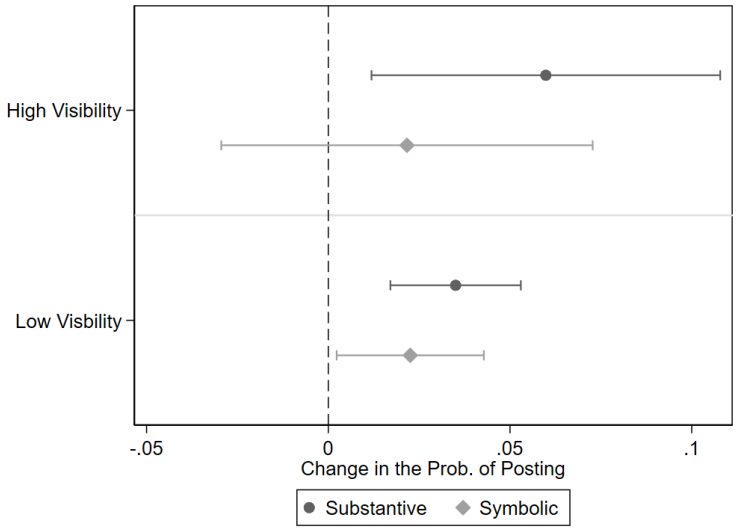
Notes: Each line represents the difference-in-differences coefficient from a regression where the outcome is an indicator for a firm posting substantive (dark grey) or symbolic (light gray) roles. The specification is outlined in Equation 1. (a) Goods-producing industries include agriculture (11), mining (21), utilities (22), construction (23), and manufacturing (31-33); Service-providing industries include all other industries that are not in the goods-producing industries: NAICS 42 - NAICS 81. (b) Public firms are identified through a crosswalk between Computat data and BGT data via the fuzzy name matching algorithm. There are 3,390 public firms identified in our data. The first breach is used, following Table 2. Lines represent 95% confidence intervals.

Figure 3: Effect of Media Attention on Probability of Posting



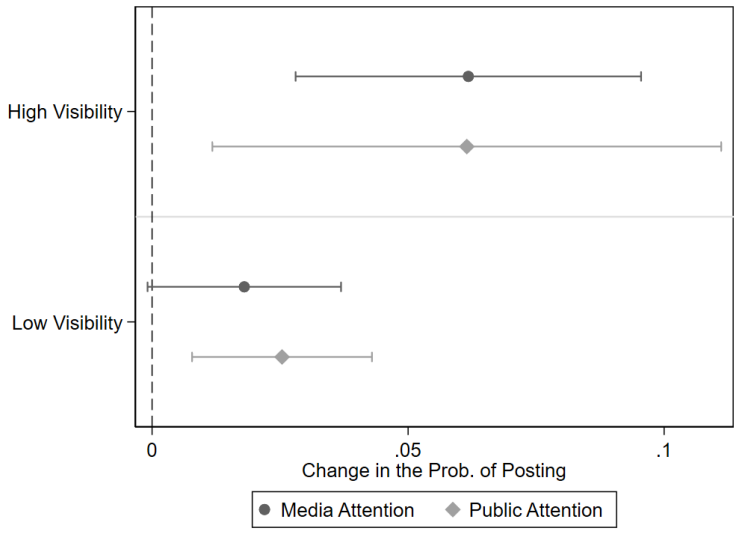
Notes: This figure displays the difference-in-differences coefficient associated with four regression specifications. The substantial lines, in darker gray, represent the coefficients for the post-breach period in a fixed effects regression of an indicator for having at least one cybersecurity posting on firm fixed effects, year fixed effects and month by industry fixed effects. The symbolic lines, in the lighter gray, represent the coefficients for the post-breach period in a fixed effects regression of an indicator for having at least one legal/public relations posting on firm fixed effects, year fixed effects and month by industry fixed effects. Each regression has an identical control group, namely firms that did not experience a data breach that was considered a CARD or HACK breach over the period. The treatment group differs – in the high visibility panel, it is firms with post-breach media share above or equal to 0.001. In the low visibility panel, it is firms with post-breach media share below 0.001. Lines represent 95% confidence intervals.

Figure 4: Effect of Public Attention on Probability of Posting



Notes: This figure displays the difference-in-differences coefficient associated with four regression specifications. The substantial lines, in the darker gray, represent the coefficients for the post-breach period in a fixed effects regression of an indicator for having at least one cybersecurity posting on firm fixed effects, year fixed effects and month by industry fixed effects. The symbolic lines, in the lighter gray, represent the coefficients for the post-breach period in a fixed effects regression of an indicator for having at least one legal/public relations posting on firm fixed effects, year fixed effects and month by industry fixed effects. Each regression has an identical control group, namely firms that did not experience a data breach that was considered a CARD or HACK breach over the period. The treatment group differs – in the high visibility panel, it is firms with post-breach search share above or equal to 0.1. In the low visibility panel, it is firms with post-breach search share below 0.1. Lines represent 95% confidence intervals.

Figure 5: Effect on Probability of More Substantive Postings



Notes: This figure displays the difference-in-differences coefficient associated with four regression specifications. The outcome variable is an indicator for having *more* or equal cybersecurity postings when compared to legal/PR postings in the same month for the same firm. The "Media Attention" lines, in the darker gray, represent the coefficients for the post-breach period in a fixed effects regression on firm fixed effects, year fixed effects and month by industry fixed effects. The "Public Attention" lines, in the lighter gray, represent the coefficients for the post-breach period in a fixed effects regression on firm fixed effects, year fixed effects and month by industry fixed effects. Each regression has an identical control group, namely firms that did not experience a data breach that was considered a CARD or HACK breach over the period. The treatment group differs – in the high visibility panel, it is firms with post-breach media or search share above or equal to a cutoff. In the low visibility panel, it is firms with post-breach media or search share below a cutoff. The cutoffs are defined as the same as the previous two figures. The Lines represent 95% confidence intervals.

Appendix A FURTHER RESULTS AND ROBUSTNESS CHECKS

Table A1: Effect of Data Breach on Talent Acquisition
Using Breach Events with the Largest Number of Breached Records

	Cybersecurity Jobs	PR and Legal Jobs	Other Jobs
	(1)	(2)	(3)
Post Breach	0.019***	0.019***	0.004
	(0.005)	(0.006)	(0.006)
R-squared	0.306	0.247	0.306

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome variables are listed at the top. The specifications are similar to those that reported in 2. The only difference is that the results reported here are based on the largest data breach in the 2010 to 2019 period for a given firm, as documented by the PRC data. Number of firms: 89,109. Number of observations: 11,585,516. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Effect of Data Breach on Demanding Related Skills
Using Breach Events with the Largest Number of Breached Records

	Cybersecurity Skills	PR and Legal Skills	Other Skills
	(1)	(2)	(3)
Post Breach	0.013**	0.014**	0.004
	(0.006)	(0.006)	(0.006)
R-squared	0.326	0.301	0.301

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome variables are listed at the top. The specifications are similar to those that reported in 5. The specifications are similar to those that reported in 5. The only difference is that the results reported here are based on the largest data breach in the 2010 to 2019 period for a given firm, as documented by the PRC data. Number of firms: 89,109. Number of observations: 11,585,516. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Additional Robustness Checks for Effect of Data Breach on Skill Demands

	Cybersecurity Skills		PR and Legal Skills	
	Drop Month 0, -1	Omit Quarter -1	Drop Month 0, -1	Omit Quarter -1
	(1)	(2)	(4)	(5)
Post Breach	0.020***		0.018***	
	(0.006)		(0.007)	
Quarter (-2)		-0.006		-0.002
		(0.007)		(0.007)
Quarter (0)		0.006		0.007
		(0.006)		(0.007)
Quarter (+1)		0.021***		0.024***
		(0.008)		(0.008)
R-squared	0.326	0.326	0.301	0.301

Notes: Each column estimates the difference-in-differences specification outlined in Equation 1. The outcome variables are listed at the top. Number of firms: 89,063. Number of observations: 11,584,603. Standard errors are clustered at the firm level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

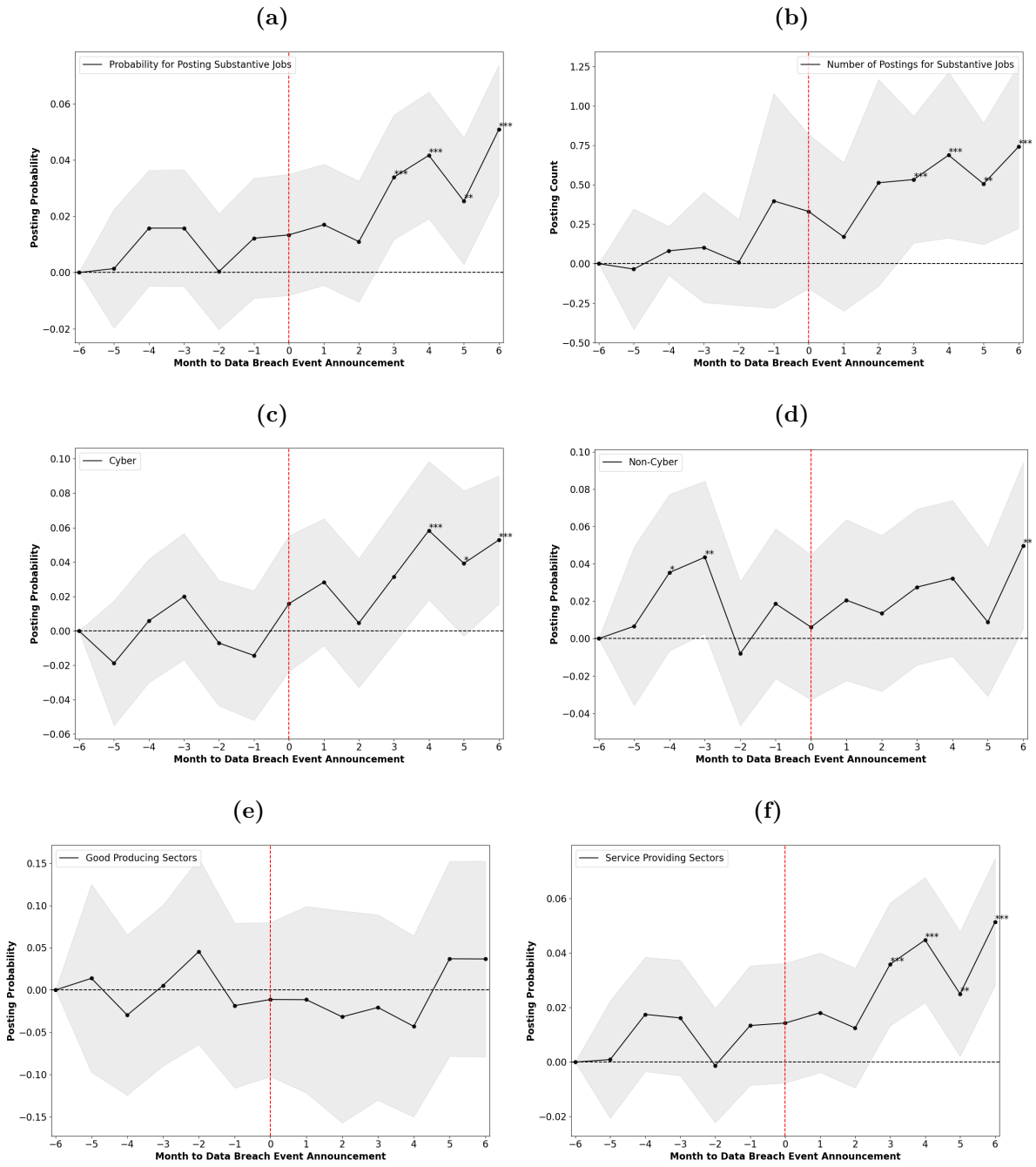
Appendix B MONTHLY DYNAMICS

To better understand firms' labor demand responses after suffering a data breach, we also take advantage of the temporal granularity of the BGT data and study substantive (i.e., Cybersecurity-related) and symbolic (i.e., Legal and PR-related) hiring at the monthly level. This analysis provides two major benefits: first, it serves as a parallel trend test and can help to rule out the existence of pre-trends; second, it allows the identification of firms' hiring response times.[48] The results are shown in Figure B1 and Figure B2. Figure B1 presents the results for substantive hiring and Figure B2 presents the results for symbolic hiring. All panels follow the same order in both figures: (a) presents the coefficients of the linear probability model for each month from six month prior to six month post the data breach events. The omitted month is the sixth month prior to the breach. The coefficients show that the breached firms are not different from the control firms in terms of the probability of hiring substantive or symbolic talents prior to the event and during the first two month after. However, they are about three percentage point more likely to post substantive or symbolic jobs starting from the third month after the breach, with an increasing trend through the later months. If we look at the number of job postings demanding substantive or symbolic skills, we see a similar pattern presented in panel (b). Again, the coefficients indicate that breached firms do not post more substantive or symbolic jobs than non-breached firms prior to the data breach events. But starting from month three after the data breach event, breached firms post, on average, significant more both substantive and symbolic jobs for than those than do not experience such events. The subsequent months also show similar increased substantive and symbolic job postings from the breached firms. These two plots highlight that breached firms do not appear to take immediate action, but instead only respond with about a 3-month delay, consistent with our quarterly analysis.

[48] This analysis can help to more precisely identify the pre-trend and the timing of the treatment effect compared to the quarterly test in columns 5 and 6 in Table 1 but also bears a risk of larger measurement error due to the unclear filling rates of each job postings.

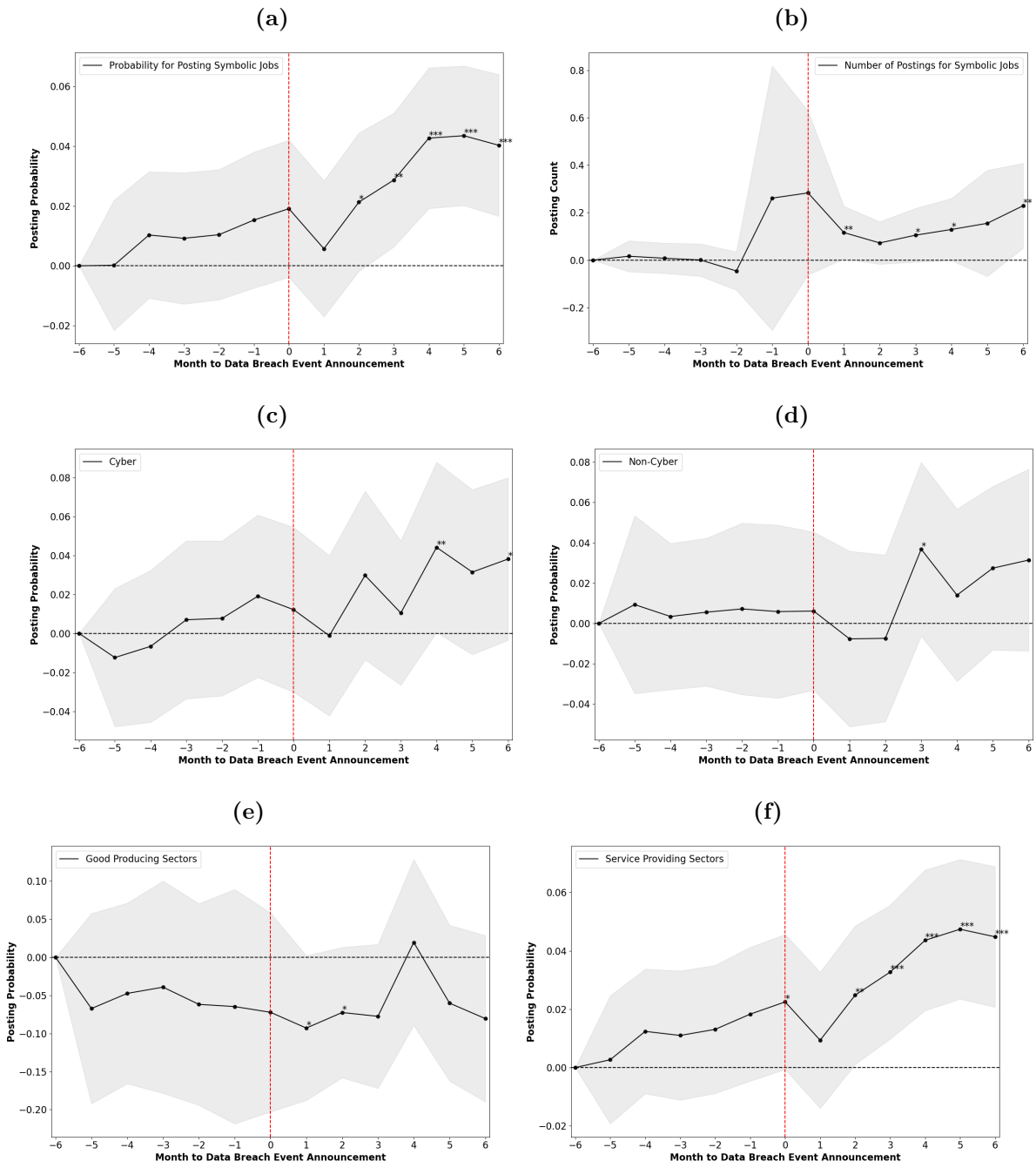
We also investigate how these monthly dynamics differ by the types of the breach events and by the industry that the firms operate in to examine the robustness of our earlier analysis. In Panel (c) we see that the probabilities of hiring both substantive and symbolic experts are higher for breached firms four months after the breach event through cyber attacks. In the placebo test for analog data losses, we observe much less cleaner patterns as shown in Panel (d). Additionally, consistent with the industry sector results from Figure 2, Figures B1 and B2 Panel (e) and (f) shows that only firms in service-providing industries increase their demand for both substantive and symbolic talents after data breach events while firms in good-producing industries take no such action. Thus, the monthly dynamics support the findings from Table 2 but highlight that there is about a three to four month delay in firms' hiring responses. One potential explanation is that it takes time for firms to decide their actual response to the breach events and to process the hiring procedure and job postings.

Figure B1: Monthly Dynamic Effect on Substantive Hiring



Notes: (a) Probability of posting cybersecurity jobs by month; (b) Number of cybersecurity postings by month; (c) Probability of posting cybersecurity jobs by month and for cyber attacks only (CARD + HACK); (d) Probability of posting cybersecurity jobs by month for non-cyber attacks (PHYS + PORT); (e) Probability of posting cybersecurity jobs by month in goods-producing sectors only; (f) Probability of posting cybersecurity jobs by month for service-providing sectors only. The grey areas are the 95% confidence intervals.

Figure B2: Monthly Dynamic Effect on Symbolic Hiring



Notes: (a) Probability of posting legal and PR jobs by month; (b) Number of legal and PR postings by month; (c) Probability of posting legal and PR jobs by month and for cyber attacks only (CARD + HACK); (d) Probability of posting legal and PR jobs by month for non-cyber attacks (PHYS + PORT); (e) Probability of posting legal and PR jobs by month in goods-producing sectors only; (f) Probability of posting legal and PR jobs by month for service-providing sectors only. The grey areas are the 95% confidence intervals.

Appendix C DEFINITION OF OCCUPATIONAL AND SKILL GROUPS

C.1 Occupational Groups

Burning Glass Technologies tags each job posting with one occupational code, based on the standard occupational classification (SOC) system. The Bureau of Labor Statistics (BLS) maintains and updates the SOC system. Recent updates to the occupational groupings occurred in 2010 and 2018 and were gradually rolled out into their data products (i.e. the CPS, OES, ...), as well as data products of private companies such as BGT, that rely on these classifications. We therefore define job postings for cybersecurity, which we use to identify substantive adoption, as those which were tagged with any of the 2010 or 2018 SOC codes specified in table C1.

Similarly, we define job postings for public relations (PR) and legal occupations, which we use to identify symbolic adoption, as those which were tagged with any of the SOC codes in table C2.[49]

C.2 Skill Groups

We leverage the Burning Glass Technology skill taxonomy, which identifies around 16,000 skills in online job postings. These skills are nested into nearly 900 unique skill clusters, which themselves are nested within 28 skill cluster families. Using their taxonomy, we classify the skills in table C3 as Cybersecurity, Legal, and PR skills, respectively.

[49] Note that these SOC codes were not updated between the 2010 and 2018 SOC taxonomies.

Table C1: Definitions of Cybersecurity SOC codes.

SOC Code	SOC Name	Notes
15-1120	Computer and Information Analysts	Cybersecurity; 2010 Code
15-1121	Computer Systems Analysts	2010 Code
15-1122	Information Security Analysts	Cybersecurity; 2010 Code
15-1140	Database and Systems Administrators and Network Architects	Cybersecurity; 2010 Code
15-1141	Database Administrators	Cybersecurity; 2010 Code
15-1142	Network and Computer Systems Administrator	Cybersecurity; 2010 Code
15-1143	Computer Network Architect	Cybersecurity; 2010 Code
15-1152	Computer Network Support Specialist	Cybersecurity; 2010 Code
15-1210	Computer and Information Analysts	Cybersecurity; 2018 Code
15-1211	Computer Systems Analysts	Cybersecurity; 2018 Code
15-1212	Information Security Analysts	Cybersecurity; 2018 Code
15-1231	Computer Network Support Specialists	Cybersecurity; 2018 Code
15-1240	Database and Network Administrators and Architects	Cybersecurity; 2018 Code
15-1241	Computer Network Architects	Cybersecurity; 2018 Code
15-1244	Network and Computer Systems Administrators	Cybersecurity; 2018 Code
15-1245	Database Administrators and Architects	Cybersecurity; 2018 Code

Table C2: Definitions of Legal and PR SOC Occupations.

SOC Code	SOC Name	Notes
23-1011	Lawyers	Legal
23-2011	Paralegals and Legal Assistants	Legal
11-2030	Public Relations and Fundraising Managers	PR
11-2032	Public Relations Managers	PR
27-3030	Public Relations Specialists	PR
27-3031	Public Relations Specialists	PR

Table C3: Definitions of Cybersecurity, Legal, and PR Skills.

Skill Group	List of Skills
Cybersecurity	Cybersecurity, Network Security, Technical Support, Database Administration, Data Management, Information Security, Application Security, Internet Security
Legal	Regulation and Law Compliance, Law Enforcement and Criminal Justice, Litigation, Legal Research, Intellectual Property, Labor Compliance, Forensics
Public Relations (PR)	Customer Relationship Management (CRM), General Marketing, Public Relations, Advertising, Brand Management, Investor Relations, Fundraising, Marketing Strategy, Corporate Communications, Media Strategy and Planning, Social Media, Concept Development