# Essays on Information Technology, Human Capital, and the Future of Work

By

Sebastian Steffen

B.A., Princeton University, 2013

S.M., Massachusetts Institute of Technology, 2020

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2022



Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Department of Management

April 15, 2022

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Erik Brynjolfsson

Jerry Yang and Akiko Yamazaki Professor

Thesis Supervisor

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Sinan Aral

David Austin Professor of Management

Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Catherine Tucker

Sloan Distinguished Professor of Management

Professor, Marketing

Faculty Chair, MIT Sloan PhD Program

**Essays on Information Technology, Human Capital, and the Future of Work**

by

Sebastian Steffen

Submitted to the Department of Management
on April 15, 2022, in partial fulfillment of the
requirements for the DEGREE of
DOCTOR OF PHILOSOPHY IN MANAGEMENT

## Abstract

This dissertation contains three essays concerning the economics of information technology, human capital, and the future of work. In the first essay, 'Occupational Change: Automation and Reskilling Risks', I develop a methodology to study occupational skill demands and estimate the returns to skills, by leveraging novel data from over 200 million online job postings from 2010 until 2020. I find large heterogeneity in skill returns across industries and identify potential (re)skill investment opportunities for workers.

In the second essay, 'Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance', I build on the methodology and data from the previous chapter to measure how feasible it is for firms to shift their workforce to remote work. Using these data, I then causally identify how much remote work practices aided firms' resilience against the Covid-19 pandemic, as measures by sales, net income, stock market returns, and volatility. The findings highlight that firms need to strategically manage the labor composition and digitization of their organizations, and consider that work-from-home practices, besides their many other advantages, are an effective way to hedge against operational risks.

In the final essay, 'Treating the Symptoms or the Cause? Substantive and Symbolic Talent Acquisition in Response to Data Breaches', I use the data from the first chapter to study firms' hiring responses to data breaches. Advancing the theory of substantive and symbolic IT adoption to complementary human capital acquisitions, I find that firms significantly increase their hiring for cybersecurity as well as public relations and legal workers after suffering breach. I also find that public scrutiny can serve as an effective mechanism to shift firms' hiring investments toward substantive, rather than symbolic measures. Given the increase in the volume and severity of cyberattacks, these results provide important and timely insights into firms' responses and incentives to more substantively safeguard their data.

Thesis Supervisor: Erik Brynjolfsson
Title: Jerry Yang and Akiko Yamazaki Professor

Thesis Supervisor: Sinan Aral
Title: David Austin Professor of Management

# Acknowledgments

My PhD experience at MIT has been incredibly rewarding. With the invaluable guidance and inspiring research from my advisor, Erik Brynjolfsson, as well as my other dissertation committee members, Sinan Aral, Sonny Tambe, and John Horton, I found my own research agenda that I am extremely passionate about.

In addition, many other professors were instrumental in stimulating my research interests - David Autor, Daron Acemoglu, Josh Angrist, Victor Chernozhukov, Wanda Orlikowski, Regina Barzilay, Andrew Ng, Ben Golub, Matthew Jackson, Sendhil Mullainathan, and Marshall van Alstyne, among many others, who are not only top notch researchers but also incredible teachers.

I also want to thank David Verrill and the team from the MIT Initiative on the Digital Economy, Christie Ko, Susan Young, and the team from the Stanford Digital Economy Lab, as well as Chris Foglia and the MIT Center for Information Systems Research for their generous support throughout the entire PhD program. They made sure that things worked and enabled me to attend workshops, research consortia, present at the WISE, ICIS, CIST, AOM, ASSA, and AAAI conferences and meet and interact with so many fellow researchers. I hope to pay it forward.

I am also very grateful for the many fruitful industry research relationships I was able to build during my time at MIT. I want to thank Bledi Taska and the entire EMSI Burning Glass Technologies team, who were incredibly supportive and helpful as I learnt my way around their data, the LinkedIn Economic Graph Research team, from whom I learnt so much during and after my internship with them, as well as the MIT-IBM AI Lab, the Google AI Research and Innovation team, as well as the NBER.

I was very fortunate to meet with and become coauthors and friends with so many amazing people at MIT and beyond. I am proud to be a part of the Sloan IT PhD family, including Daniel Rock, Avi Collis, David Holtz, Michael Zhao, Guillaume Saint-Jacques, Alex Moehring, Zanele Munyikwa, Hong-Yi Tu Ye, Emma Van Inwegen, Lindsey Raymond, Mohammed Alsobay, .... You put the IT in MIT. Madhav Kumar, Stefan Dimitriadis, Soomi Kim, Heather Yang - thank you for keeping me on track. Thanks also to all of the MIT IDE and Stanford HAI PostDocs - Wang Jin, Daniela Scur, Ananya Sen, Sarah Bana, Georgios Petropoulos, Xiupeng Wang, Sagit Bar-Gill, Seth Benzell, Andrey Fradkin, Xiang Hui, Meng Liu. All of you rock and truly helped to make this journey more bearable or, dare I say, fun.

I want to thank all of the, more or less random, internet strangers, for your wisdom of the crowd - in aggregation, your knowledge through and work on StackOverflow, github, Wikipedia, documentation and IT infrastructure maintenance, Twitter threads, Youtube, Twitch and other channels were invaluable to me.

This list would not be complete without thanking the many other mentors I had along the way. I want to thank Henry Farber for awakening my interest in academic research, as well as Susan Athey and Markus Mobius for continuing to foster this interest at Microsoft Research.

During my path to the PhD all of my grandparents passed away, which was, and continues to be, hard to come to terms with. I have been, and am, exceedingly fortunate to have you in my life. You, my brother, and my parents were the best family one could have wished for and your love, guidance, and support will always be with me. I strive to make you proud. Finally, an (uncountably) infinite amount of gratitude and love goes to my wife, Linda Zhao, who is proof that I did at least one thing right so far.

Of course, the discerning reader may reject these claims of gratitude as disingenuous and mere signaling or cheap talk. I still appreciate you taking the time to read this and want to use this final paragraph to disabuse you: I meant every word.

# Contents

**1 Occupational Change: Automation and Reskilling Risks** **17**

1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

1.2 Literature . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

1.2.1 Returns to Skills . . . . . . . . . . . . . . . . . . . . . . . 19

1.2.2 Technological Change and Automation . . . . . . . . . . . . . 21

1.3 Data and Methodology . . . . . . . . . . . . . . . . . . . . . . . 23

1.3.1 Millions of Annotated Job Postings . . . . . . . . . . . . . . 24

1.3.2 Wages and other Occupational Data . . . . . . . . . . . . . . 26

1.3.3 Occupational Automation . . . . . . . . . . . . . . . . . . . 26

1.3.4 From Job Postings to a Panel of Occupational Skill Demands . 26

1.3.5 Measuring Changes in Occupational Skill Demands . . . . . . 27

1.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

1.4.1 The Implied Market Values of Skills . . . . . . . . . . . . . . 28

1.4.2 What is driving Occupational Change? . . . . . . . . . . . . . 30

1.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

**2 Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance** **35**

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

2.2 Hypothesis Development . . . . . . . . . . . . . . . . . . . . . . 42

2.2.1 Higher Resilience Through WFH practices during Crises, such as Covid-19 . . . . . . . . . . . . . . . . . . . . . . . . . . 43

2.2.2 WFH and Complementary IT Investment . . . . . . . . . . . . 44

2.2.3 Industry Differences (Essential vs. Non-essential, High-Tech vs. Low-Tech) . . . 44
2.3 Data Construction and Summary Statistics . . . 46
2.3.1 Data . . . 46
2.3.2 Summary Statistics . . . 49
2.4 Empirical Methodology & Results . . . 51
2.4.1 Empirical Methodology . . . 51
2.4.2 Empirical Results - Baseline . . . 52
2.4.3 Empirical Results - Industry Heterogeneity . . . 57
2.5 Conclusion . . . 59

**3 Treating the Symptoms or the Cause? Substantive and Symbolic Talent Acquisition in Response to Data Breaches 63**
3.1 Introduction . . . 64
3.2 Literature Review and Hypothesis Development . . . 68
3.3 Data and Summary Statistics . . . 74
3.3.1 Data on Firms' Hiring from Burning Glass Technologies (BGT) 74
3.3.2 Data on Firms' Data Breach Events from Privacy Rights Clearinghouse (PRC) . . . 75
3.3.3 Matching BGT Hiring Data with PRC Data Breach Events . . . 76
3.3.4 Hiring for Cybersecurity and Legal/PR as Substantive and Symbolic Adoption . . . 77
3.3.5 Data on Firms' Media and Public Attention from MIT Media-Cloud and Google Trends . . . 77
3.3.6 Summary Statistics . . . 78
3.4 Empirical Methods . . . 79
3.4.1 Difference-in-Differences Estimation . . . 79
3.4.2 Quarterly Dynamics . . . 80
3.5 Main Results . . . 81
3.5.1 Effect of Data Breaches on the Probability of Substantive Hiring 81

3.5.2 Effect of Data Breaches on the Probability of Symbolic Hiring 84
3.5.3 Placebo Tests on Non-Relevant Jobs . . . . . . . . . . . . . . . . 85
3.5.4 Demand for Related Skills . . . . . . . . . . . . . . . . . . . . 86
3.6 Additional Results and Effect Heterogeneity . . . . . . . . . . . . . . . . 88
3.6.1 Detailed Decomposition by Occupations . . . . . . . . . . . . . 88
3.6.2 Effect Heterogeneity by Data Breach Type . . . . . . . . . . . 89
3.6.3 Effect Heterogeneity by Industry Sector . . . . . . . . . . . . . 90
3.6.4 Effect Heterogeneity by Ownership Type . . . . . . . . . . . . 91
3.6.5 Discussion of Results . . . . . . . . . . . . . . . . . . . . . . . 92
3.7 Public Visibility as a Mechanism . . . . . . . . . . . . . . . . . . . . . 93
3.7.1 Measuring the Change of Visibility Related to Data Breaches . 93
3.7.2 Effect Heterogeneity by Visibility . . . . . . . . . . . . . . . . 94
3.8 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96

**4 Tables & Figures for Chapter 1 99**

**5 Tables & Figures for Chapter 2 111**

**6 Tables & Figures for Chapter 3 121**

**7 Appendices for Chapter 1 133**

**8 Appendices for Chapter 2 135**

8.1 Appendix A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
8.2 Appendix B . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 136
8.3 Appendix C . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 138
8.4 Appendix D . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
8.5 Appendix E . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 140
8.6 Appendix F . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
8.7 Appendix G . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 144
8.8 Appendix H . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149

**9 Appendices for Chapter 3** **151**

9.1 Appendix A - Further Results and Robustness Checks . . . . . . . . . . 151

9.2 Appendix B - Monthly Dynamics . . . . . . . . . . . . . . . . . . . . . 155

9.3 Appendix C - Spillover Effects on Competitors . . . . . . . . . . . . . 159

9.4 Appendix D - Preventative Effects of Cybersecurity Hiring for Future Breaches . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 161

9.5 Appendix E - Definitions of Occupational and Skill groups . . . . . . 163

9.5.1 Occupational Groups . . . . . . . . . . . . . . . . . . . . . . 163

9.5.2 Skill Groups . . . . . . . . . . . . . . . . . . . . . . . . . . . 163

**10 Bibliography for Chapter 1** **166**

**11 Bibliography for Chapter 2** **171**

**12 Bibliography for Chapter 3** **176**

# List of Figures


4-1 **Yearly Number of Job Postings scraped by BGT (2010-2018).** 105

4-2 **Logged Overall Skill Shares by 2010 Wage Tercile, using the 28 BGT Skill Cluster Families.** . . . . . . . . . . . . . . . . . . . . 106

4-3 **Logged Overall Skill Shares for Skill Clusters within the IT Skill Cluster Family by 2010 Wage Tercile.** . . . . . . . . . . . . 107

4-4 **The top 30 skill demand shares of Data Scientists (2010-2018).** 108

4-5 **The top 30 skill demand shares of Lumberjacks (2010-2018).** 108

4-6 **Occupational Cosine Similarity between 2010 and 2018 skill share vectors over 2010 wage percentile.** . . . . . . . . . . . . . . . 109

4-7 **Occupational Cosine Similarity between 2010 and 2018 skill share vectors over 2010 wage percentile.** . . . . . . . . . . . . . . . 110

5-1 **Distribution of WFH indices by Sector** . . . . . . . . . . . . . . . . 116

5-2 **Average Quarterly WFH Index by Firms with High vs. Low Pre-Covid-19 WFH Feasibility** . . . . . . . . . . . . . . . . . . . . . 117

5-3 **Sector Average WFH indices (2019 and 2020)** . . . . . . . . . . 118

5-4 **Dynamics of the treatment effect in the DID setting** . . . . . 119

5-5 **Dynamics of the treatment effects in the DID setting on IT-related Job Postings** . . . . . . . . . . . . . . . . . . . . . . . . . . 120

6-1 **Timely Notification Requirements by States' Data Breach Notification Laws** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128

6-2 **Effect Heterogeneity of Data Breach on Firm's Talent Acquisition** . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 129

6-3 **Effect of Media Attention on Probability of Posting** . . . . . . 130
6-4 **Effect of Public Attention on Probability of Posting** . . . . . . 130
6-5 **Effect on Probability of More Substantive Postings** . . . . . . 131

F1 **WFH Feasibility and Firm Performance across Sectors** . . . . 143
H1 **Herfindahl–Hirschman Index 2018-2020** . . . . . . . . . . . . . 149

B1 **Monthly Dynamic Effect on Substantive Hiring** . . . . . . . . . 157
B2 **Monthly Dynamic Effect on Symbolic Hiring** . . . . . . . . . . 158
C1 **Effect of a Data Breach on Competitors' Job Postings** . . . . 160

# List of Tables


4.1 **Panel Wage Regression - BGT Skills** . . . . . . . . . . . . . . . . 100

4.2 **Top and Bottom 5 Occupation Fixed Effects from BGT Skills Wage Panel Model 3.** . . . . . . . . . . . . . . . . . . . . . . . . . . 101

4.3 **Industry Fixed Effects from BGT Skills Wage Panel Model 3.** 101

4.4 **Panel Wage Regression - Deming Skills** . . . . . . . . . . . . . . . 102

4.5 **Top and Bottom 5 Occupation Fixed Effects from Deming Skills Wage Panel Model 3.** . . . . . . . . . . . . . . . . . . . . . . 103

4.6 Industry Fixed Effects from Deming Skills Wage Panel Model 3. . . . 103

4.7 **Occupational Change** . . . . . . . . . . . . . . . . . . . . . . . . . 104

5.1 **Variable Definitions and Summary Statistics** . . . . . . . . . . . 112

5.2 **WFH Feasibility and Financial Performance** . . . . . . . . . . . 113

5.3 **WFH Feasibility and Stock Market Reactions** . . . . . . . . . . 114

5.4 **WFH Feasibility and Financial Performance - Controls for Industry-level Demand Shocks** . . . . . . . . . . . . . . . . . . . 115

5.5 **WFH Feasibility and Stock Market Reactions - controlling the Demand Shock by Industry** . . . . . . . . . . . . . . . . . . 115

6.1 **Summary Statistics** . . . . . . . . . . . . . . . . . . . . . . . . . 122

6.2 **Effect of a Data Breach on Talent Acquisition** . . . . . . . . . 123

6.3 **Effect of a Data Breach on Whether Firms Hire More Substantive Than Symbolic Jobs** . . . . . . . . . . . . . . . . . . . . 124

6.4 **Effect of a Data Breach on Skill Demands** . . . . . . . . . . . . 125

6.5 **Effect of a Data Breach by Data Breach Type** . . . . . . . . . 126

6.6 Effect of a Data Breach on Demand for Specific Cybersecurity Occupations . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127

A1 WFH Feasibility and Firm Performance: State-Quarter Fixed Effects . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
B1 WFH and firm performance: matched sample analyses . . . . 137
C1 WFH Feasibility and firm performance: Continuous WFH proxy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 138
D1 Falsification Test (by Position of Firm Name by Alphabetic Order) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
E1 WFH Feasibility and Firm Performance: Essential vs. Non-Essential . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 141
E2 WFH Feasibility and Firm Performance: High-tech vs. Other Industries . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 142
G1 Industry definition at 2-digit NAICS Level. . . . . . . . . . . . 144
G2 High-Tech Industry definition at 4-digit NAICS Level according to [Decker et al., 2017]. . . . . . . . . . . . . . . . . . . . . . 145
G3 Essential Industry definition at 4-digit NAICS Level according to [Papanikolaou and Schmidt, 2020]. . . . . . . . . . . . . 146
G3 Essential Industry definition at 4-digit NAICS Level (Continued). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 147
G3 Essential Industry definition at 4-digit NAICS Level (Continued). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 148
G4 Aggregated Industry definition following [Autor et al., 2020]. 148

A1 Effect of Data Breach on Talent Acquisition Using Breach Events with the Largest Number of Breached Records . . . . 151
A2 Additional Robustness Checks for Effect of Data Breach on Talent Acquisition . . . . . . . . . . . . . . . . . . . . . . . . . 152
A3 Effect of Data Breach on Demanding Related Skills Using Breach Events with the Largest Number of Breached Records 153

A4 **Additional Robustness Checks for Effect of Data Breach on Skill Demands** . . . . . . . . . . . . . . . . . . . . . . . . . . . . 154
D1 **Effect of Cybersecurity Job Posting on Future Events** . . . . 162
E1 **Definitions of Cybersecurity SOC codes.** . . . . . . . . . . . . . . 164
E2 **Definitions of Legal and P&R SOC Occupations.** . . . . . . . . . 164
E3 **Definitions of Cybersecurity, Legal, and P&R Skills.** . . . . . . 165

# Chapter 1

# Occupational Change: Automation and Reskilling Risks


## Abstract

We derive a novel occupation-industry level panel of skill demands from the near-universe of tagged online job postings in the US for the last decade (2010-2018). We use this data to study how the skill demands of occupations have changed and how these changes affect the returns to skills. Low- and medium-wage occupations' skill demands changed more than those of high-wage ones. Thus, lower-wage workers face not only higher risks of technological displacement but also increased risks of reskilling in order to stay productive. We show that routine-biased technological change (RBTC) due to automation technologies such as ML can best explain these results, while skill-biased and (endogenously) directed technological change cannot. Technical skills, such as ML, Business, Software, and Data Skills have particularly high implied market values, as do Social Skills and Creativity. These therefore represent lucrative (re-)skill investment opportunities for workers, unlike writing and non-cognitive skills. Finally, there is significant heterogeneity in industry fixed effects with the Utilities, Mining, Management and IT Industries offering much higher returns than the Food and Retail industries, even after controlling for skills.

## 1.1 Introduction

Skills are investment assets and as such have value, face unexpected future returns, and thus incur risk of depreciation. After all, skills are also referred to as human *capital* ([Becker, 1964]). The social sciences have a long history of studying the returns to skills, albeit through proxies such as years of schooling or college attendance ([Mincer, 1974, Goldin and Katz, 2008, Goldin and Katz, 2009, Michaels et al., 2014, Beaudry et al., 2015], wages [Autor and Dorn, 2013], or test scores [Hanushek et al., 2015]). Most of these measures are static in time and are unable to capture dynamic short-term changes in the value of *actual* skills that may be exacerbated through the rapid adoption of a novel risk factor: artificial intelligence (AI), and specifically Machine Learning (ML).

ML's vast improvements to prediction tasks enables it to substitute, complement, and expand demand for occupations or skills within them ([Agrawal et al., 2019]). In particular, ML's ability to learn from large amounts of data make it exceptionally suitable to substitute for routine skills and complement non-routine skills ([Goos et al., 2014, Brynjolfsson et al., 2018]) - it is a form of Routine-Biased Technological Change (RBTC). Thus, the widespread adoption of ML may lead to a skill mismatch between the skill demands of jobs and workers' current skills and may lead to technological unemployment ([Restrepo, 2015]). Many firms plan to adopt ML capabilities to leverage data-driven decision-making but a large number lacks the necessary complements, including skilled workers, to do so ([Brynjolfsson and McElheran, 2016a]). While the adoption of ML has naturally led to a rise in the implied market value for ML-related skills, there has also been a depreciation of the value of routine skills. Thus, the relatively higher prevalence of routine skills in medium-wage occupations makes these jobs more susceptible to ML, such that ML may have been an important factor for occupational wage-polarization ([Autor and Dorn, 2013]). This polarization has led to a considerable rise in the service industry, in particular for workers in health services or last-mile services such as transporting people and packages ([Autor, 2019]). ML also increased the demand for 'ghost' workers who help to generate the massive labeled datasets necessary to train ML models ([Gray and Suri, 2019]).

In this paper we study the interactions between the implied market value of skills, changes in occupational skill demands, and ML by leveraging a novel panel dataset of occupation-level skill demands, which we derive from the near-universe of US online job postings of the last decade. Besides estimating returns to skills from this skill demand-side dataset, we also use it to understand what is driving the recent changes in occupational skill demands and which of the three leading theories of technological change can best explain them.

The paper proceeds as follows: section (1.2) gives an overview of the returns to skills and technological change and automation literature while section (1.3) explains our data and methodology that allows us to go from annotated job postings to yearly skill shares of occupation-industry cells. In the results section we first show that occupational skill demands are indeed changing and that low and medium-wage occupation-industry cells' skill shares changed more between 2010 and 2018 than those of high-wage ones. Then we present the results from our skill panel regression. As a robustness check we use two different types of skill classifications and show different sets of fixed effects models. Finally, we show the results (1.4) of our occupational change regression on proxy variables for each of the three leading technological change theories. Section (1.5) concludes.

## 1.2 Literature

### 1.2.1 Returns to Skills

Estimating the returns to skills goes back at least 70 years to Jacob Mincer's famous 'Mincer Earnings Regressions' ([Mincer, 1958]). In his seminal work he estimates:

$$lnw(s,x) = \alpha_0 + \rho_s s + \beta_0 x + \beta_1 x^2 + \epsilon \tag{1.1}$$

where w(s, x) is the wage at schooling level s and work experience x [1]. While the Mincer model was a work horse in the past, it has an important shortcoming which is incompatible with the increasing polarization of the wage distribution since the 1980s ([Firpo et al., 2011, Lemieux, 2006, Autor and Dorn, 2013]). In the canonical Mincer model, changes in wages can only occur due to changes in skill prices. Furthermore, *conditional on skills*, wages would be identical across occupations, which is unrealistic.

Another model, by [Welch, 1969] allows for the skill supply to be bundled, within workers, such that worker i's wage, $w_{it}$ can be estimated as:

$$w_{it} = \theta_t + \sum_{k=1}^{K} r_{kt} S_{ik} + u_{it} \tag{1.2}$$

, where $S_{ik}$ are the skill components of worker i, and $r_{kt}$ are the desired returns to skills coefficients. However, as discussed in [Firpo et al., 2011], this model still cannot account for unequal wages for identical skill bundles in different occupations - it requires that workers' skill bundles could be unbundled, which is not the case ([Gibbons et al., 2005]). To allow for wages to differ across occupations for similar skill bundles, [Firpo et al., 2011] estimate the following Welch-inspired model for wages of worker i in occupation j:

$$w_{ijt} = \theta_t + \sum_{k=1}^{K} r_{jkt} S_{ik} + u_{ijt} \tag{1.3}$$

This model is now general enough to account for technological change as well as offshoring. However, it still does not take industries into account, even though there is strong evidence for industry fixed effects.

Estimating the returns to skills from the skill supply side, i.e. from workers, is very challenging as the skill components of workers are generally latent. Schooling was the earliest proxy variable used by Mincer as well as Becker. However, with the significant rise in college-educated workers, measuring the return to an additional year of school, their proxy for the returns to skills, is no longer relevant. While the

[1] See ([Heckman et al., 2003]) for an excellent overview.

definition of the word skill is still hard to pin down, it has certainly moved away from being equivalent to attaining a college degree. Even test scores may only be poor indicators of skill. Thus, the more recent returns to skills estimates, which leverage the Programme for the International Assessment of Adult Competencies (PIAAC) test scores [Hanushek et al., 2015] may not be reliable either. Even if the test accurately measured skills around the time of test-taking, it becomes a worse skill measure the longer ago an individual took the test. In addition, there are other confounding hard-to-measure variables, such as grit [Duckworth and Gross, 2014], or latent, idiosyncratic preferences.

Due to the issues discussed above, we move away from skill supply side estimations and propose a *skill demand side estimation* in a similar style to [Firpo et al., 2011]. Thus, instead of estimating the returns that a worker gets from his skill bundle, we estimate how the market values skill bundles, by decomposing wages for each occupation into returns to the skills that are demanded for that occupation. By leveraging a large, and novel dataset of nearly all online US job postings of the last decade, we are able to gain additional data granularity and estimate not only an occupation-level panel but also an occupation-industry-level panel, thereby allowing for the theorized industry-fixed effects.

### 1.2.2 Technological Change and Automation

Artificial Intelligence is widely believed to be the next big General Purpose Technology [Brynjolfsson et al., 2018]. Automation, more generally, has the capacity to make labor more productive (labor-augmenting automation), to make automation itself more productive (automation at the intensive margin), to introduce new skills into the economy, or to displace a wider range of tasks (automation at the extensive margin)[Acemoglu and Restrepo, 2019]. This race between man and machine may lead to a rise of technological unemployment [Acemoglu and Restrepo, 2018]. No matter whether automation or (task) innovation 'wins', both forces lead to changes in occupations' underlying skill requirements and force workers to reskill to remain productive. However, given that automation and other IT capital are relative substitutes for workers who perform routine manual and cognitive tasks, but relative complements for workers who perform non-routine cognitive tasks, low-wage workers will face the brunt of the these changes [Autor et al., 2003, Michaels et al., 2014].

In fact, some of these changes have already manifested themselves. Some argue that the terms routine and non-routine characterize the relationship between tasks/skills and information technology (IT) and find that occupations have shifted towards requiring more analytical and interactive tasks and away from requiring cognitive-routine and manual-routine tasks [Spitz-Oener, 2006]. Skills, as a form of task-specific human capital, are an important source of individual wage growth [Gathmann and Schoenberg, 2010]. Thus, the relative loss of productivity of routine skills translates to lower wages and an overall more polarized wage and employment share distribution [Autor and Dorn, 2013].

However, medium- and high-wage occupations are not immune to occupational change either. Occupations that heavily rely on IT skills have been shown to change faster due to rapid software innovation [Hershbein and Kahn, 2018]. These fast obsoletion rates of specific software skills lead to relatively flatter earnings profiles for STEM workers [Deming and Noray, 2018]. Some have argued for a 'great reversal' in demand for cognitive skill and shown that more educated workers have begun to crowd out less educated workers, due to sorting and changes in relative productivity of workers and capital [Beaudry et al., 2015]. Automation and IT capital, such as Data-Driven Decision Making (DDD), have been rapidly adopted and have made plants more productive and efficient, requiring managers and other high-wage occupations to adapt to stay productive [Brynjolfsson and McElheran, 2016b, Bartel et al., 2007]. These results suggest that reskilling is both necessary as well as costly, in particular for low-wage workers, and that the dnynamics of occupational skill demands are an important phenomenon to study. This is especially true in light of IT enabling faster rates of technological adoption due to low margin costs and scalability.

In this paper we study the effects of skill demands and automation on three occupational outcomes: changes in (i) wages, (ii) employment shares, and (iii) occupational skill demands. In particular, we focus on occupational change and leverage a novel large data set of online job postings between 2010 and 2019 from which we derive a panel of occupational **skill share vectors**.

Our main results show that AI has multiple negative consequences for low and medium-wage occupations: not only do occupations that are highly susceptible to automation correlate with decreased employment shares, they also face much higher occupational change in terms of skills demanded. Thus, the 'lucky' low- and medium-wage workers that do manage to keep their job, still have to reskill relatively more in order to keep it. Notably, an occupation's suitability for machine learning (SML) does not correlate with decreased wages, perhaps due to workers being highly forward-looking and abandoning obsolete skills [Horton and Tambe, 2019].

We also find that certain skill groups correlate particularly strongly with occupational change. These include 'Information Technology' and 'Engineering', consistent with prior findings on the fast obsoletion rates of specific software skills in STEM occupations [Deming and Noray, 2018], as well as 'Maintenance and Repair' and 'Environment' skills, which is consistent with automation and IT displacing routine non-cognitive skills relatively more easily (RBTC) [Goos et al., 2014, Michaels et al., 2014].

## 1.3 Data and Methodology

Our methodology relies on several different data sources:

- Burning Glass Technologies (BGT),
- an analytics software company that scrapes and annotates job postings from nearly all online job posting sites and employment search engines,
- The US Bureau of Labor Statistics (BLS) Occupational Employment Survey (OES), which provides historic wages and employment shares at the occupation and occupation-industry level
- Automation-related proxies from several recent papers: Suitability for Machine Learning (SML) from [Brynjolfsson et al., 2018], AI Progress Scores from [Felten

et al., 2018], Cognitive Skill Fractions from [Alabdulkreem et al., 2018], and the AI, Software, and Robot Indices created by [Webb, 2019].

We will first describe these data sources in more detail before delving more into the methodology of going from job postings to yearly skill panels at the occupation, as well as at the occupation-industry level.

### 1.3.1 Millions of Annotated Job Postings

The BGT data covers about 600 million online job vacancy postings posted on over 40,000 distinct online job platforms in the United States between 2010 and 2018 and arguably covers the near-universe of job postings (Figure 4-1). Each vacancy posting is parsed, deduplicated, and annotated with the posting date, the SOC occupational code, the NAICS industry code, 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.

At the most detailed level, the BGT taxonomy includes ≈ 16,000 skills - these are nested within 658 skill clusters, which are nested within 28 skill cluster families. For example, *Python* is a skill within the *Scripting Languages* skill cluster, which itself falls into the *Information Technology* skill cluster family. The taxonomy was initially assembled from online resumes and is continuously updated through client feedback, research, and forums. K-Means clustering along with additional qualitative checks were employed to create meaningful skill clusters. Whenever new skills are added to the taxonomy, it is refit to the entire history of job postings data. This minimizes potential biases which may have arisen through BGT's time-varying ability to capture new skills.

Notably, this taxonomy is significantly more detailed than other skill taxonomies, such as the Bureau of Labor Statistics (BLS)' O*Net, which contains just 2 levels: it consists of 35 skills mapped into 6 skill groups. Furthermore job postings are scraped daily and are therefore able to capture changes in skill demands significantly faster. O*Net only undergoes yearly updates which usually only cover a subset of occupations.

However, broader taxonomies are still useful as they can be more interpretable, in particular in terms of the routine and non-routine skill distinction of the RBTC theory. Since the BGT taxonomy does not directly map to the O*Net taxonomy, we instead leverage the taxonomy built by [Deming and Noray, 2018] for additional robustness checks. Their taxonomy maps a subset of the BGT skills into 20 interpretable skill groups, that contain 'social skills', 'cognitive skills', and 'management skills' among others.

The BGT data is ideal for a panel study of occupational skill demands for several reasons. Given that each job can be viewed as a 'bundle of skills' [Restrepo, 2015, Deming and Kahn, 2018], each job posting represents a draw from the ground-truth skill distribution of the job's underlying occupation. Hundreds of thousands of job postings can thus pin down a very tight distribution of skill demands for a given occupation. However, some occupations are notoriously underrepresented in online job platforms, and thus in BGT's data. We therefore remove occupations with very few ($< 1,000$) job postings per year. To ensure that each aggregation cell contains enough job posting data points, we chose to use years, instead of months, as our temporary aggregation variable. This also minimizes potential seasonality effects. It furthermore allows us more granularity on the industry level. Since BGT also annotates job postings with 2-digit industry NAICS codes, we are able to derive a skills demand panel at the SOC6-NAICS2 level.

Besides the granularity enabled through its size, this data has another important feature that may be advantageous to study the returns to skills. Unlike most other papers which generally leverage a labor supply side panel with a static skill proxy, we rely on aggregated job postings, which represent the skill demand side. Previous studies face several significant sources of omitted variable biases, such as grit [Duckworth and Gross, 2014], and idiosyncratic preferences for non-pecuniary benefits [Katz and Autor, 1999], such as occupational choice, work hours, culture, among others.

### 1.3.2 Wages and other Occupational Data

The Bureau of Labor Statistics (BLS) collects and publishes yearly data on employment shares and wages at different levels of aggregation. The BLS classifies occupations according to the Standard Occupational Classification System, SOC codes. The 2010 version defines 840 distinct detailed occupations, which are nested in 461 broad occupations, 97 minor groups, and 23 major groups. It also classifies industries according to the North American Industry Classification System, NAICS codes. The 2012 version defines 20 distinct industry sectors. We use their annual publications of wages and employment shares at the SOC level as well as at the SOC6 x NAICS2 level.

### 1.3.3 Occupational Automation

Besides these two main data sources we leverage several recent papers' occupation-level automation scores. We include automation as a proxy for the level of routineness of occupations, as routine tasks are more automatable [Autor and Dorn, 2013]. Routine tasks are particularly suitable for Machine Learning, which is why our preferred measure is the Suitability for Machine Learning (SML) metric from [Brynjolfsson et al., 2018]. We consider additional measures in the appendix. These include the AI scores from [Felten et al., 2018], which are based on the Electronic Frontier Foundation (EFF) AI Progress report, as well as the O*Net-derived cognitive skill fraction as defined in [Alabdulkreem et al., 2018]. Finally, we use the recent AI, software, and robot indices created by [Webb, 2019] from patent data.

### 1.3.4 From Job Postings to a Panel of Occupational Skill Demands

As previewed above, we aggregate the individual job postings to the occupation-year level, thereby viewing each of the 600 million job ads as a draw from the corresponding occupational skill distribution for occupation $i$ in year $t$. While the majority of the 840 different occupations listed in the SOC taxonomy are present in our data, we remove military occupations and occupations with fewer than 1,000 postings leaving us with 556 distinct occupations. We also noticed a 'buzzword bias' which led to inflationary usage of certain popular skill words within job postings, thereby misrepresenting actual skill demands. Thus, in our preferred specification we work with the more meaningful skill clusters - this way mentions of the skills such as 'AWS' and 'Azure' within the same posting are only counted once as the skill cluster 'Cloud Solutions'.

Since the total number of job postings increases from about 80 million in 2010 to 180 million in 2018, we normalize the raw skill cluster counts to derive the *skill share vectors* for each of the 556×9 occupation-year cells.

$$\boldsymbol{s_{ikt}} = \frac{1}{\sum_{j=1}^{S}(s_{i,t,j})}(s_{i,t,1}, \ldots s_{i,t,S}) \qquad \text{(Skill Shares)}$$

represents the share of skill k of all skill demands in occupation i at time t. To illustrate, figures 4-4 and 4-5 show the skill demand shares for Data Scientists and Lumberjacks, respectively, based on just the 30 most relevant skill clusters. These top 30 skill clusters cover about 12% and 66% of all skill demands for Data Scientist and Lumberjack occupations, respectively. The changes in skill demands over the past decade are clearly visibly and suggest that the implied market skills of skills vary considerably. This confirms that job seekers indeed face considerable risk in terms of how their skills are valued in the market but also that investment in the 'right' skills, i.e. via reskilling, may be very lucrative.

### 1.3.5 Measuring Changes in Occupational Skill Demands

To measure more rigorously how large these changes in occupational skill shares are across time, we apply Cosine Distance between the earliest (2010) and latest (2018) skill share vectors:

$$d_{cos}(\boldsymbol{s_{i,2010}}, \boldsymbol{s_{i,2018}}) = 1 - \frac{\boldsymbol{s_{i,2010}} \cdot \boldsymbol{s_{i,2018}}}{||\boldsymbol{s_{i,2010}}||\ ||\boldsymbol{s_{i,2018}}||} \qquad \text{(Cosine Distance)}$$

Notably, we calculate this distance over the entire skill share vectors and not just for the top 30 skills. There are many other suitable distance and similarity metrics to choose from [Cha, 2007], such as the Jensen-Shannon Divergence. However, the Euclidean distance is not one of them as the curse of dimensionality makes it meaningless in high dimensions, which is the case here due to the large number (658) of skills clusters.

Cosine Distance is one of the most widely used distance measures as it is fast to calculate and easy to interpret. It measures the angle between two vectors, with a magnitude of 1 implying perfect alignment, 0 indicating orthogonality, and -1 indicating perfectly opposite alignment. Occupational skill demands do not change by large magnitudes, such that none of the observed cosine distances falls below 0. For example, the cosine similarity between the 2010 and 2018 skill share vectors of the aforementioned Lumberjacks is 0.37 and is one of the larger changes in our data.

We calculate the Cosine Distance between the 2010 and 2018 skill share vectors for each SOC occupation in our sample. As can be seen in Figure 4-6, it appears that low wage and medium wage occupations changed significantly more than high wage occupations between 2010 and 2018. This is consistent with routine-biased technological change, as low and medium wage occupations tend to require more routine tasks. We will show the robustness of this relationship with regressions in the results section. Due to the large size of our data we can do the same for each occupation-industry combination, i.e. for each SOC6-NAICS2 cell (Figure 4-7). We again observe a similar trend. Low and medium-wage occupation-industry combinations changed more than high-wage ones. However, with this additional data granularity there seems to be an another point of inflection around the 70th wage percentile.

## 1.4 Results

### 1.4.1 The Implied Market Values of Skills

The returns to skills literature has previously faced challenging data problems:

> [Firpo et al., 2011] "Ideally, we would like to estimate the skill pricing parameters ($r_{jkt}$) using repeated cross sections from a large data set containing detailed information on wages, skills, and occupations. We could then look at the contribution of changes in occupational wage setting to the overall changes in the wage structure [...]. **Unfortunately, no such data set exists**."

While this remains true for skill supply side estimations based on worker panels, we believe that our panel data allows the estimation of skill pricing parameters from the skill demand side. In particular, we estimate the following:

$$w_{(ijt)} = \theta_i + \theta_j + \sum_{k=1}^{K} r_{ijk} s_{ijkt} + u_{ijt},$$

where $w_{ijt}$ is the wage of occupation i in industry j at time t, the $r_{ijk}$ are the returns to each skill k in occupation i and industry j, the $s_{ijkt}$ are the skill shares of each skill k in occupation i and industry j at time t, and $\theta_i$ and $\theta_j$ are the occupation and industry-specific fixed effects, respectively. Roughly speaking, by viewing each occupation as a bundle of skills, the wage paid to occupation i in industry j can be decomposed as a weighted average of the values of each skill and how much each skill is demanded for said occupation-industry combination. We again run these regressions for both the BGT-provided skill cluster families, as well as for the skill groups provided by Deming (2018). The results can be seen in tables (4.1) and (4.4). The corresponding industry-fixed effects can be seen in tables (4.3) and (4.6). For the corresponding occupation-fixed effects we only report the top 5 and bottom 5 in tables (4.2), and (4.5), as there are over 600 occupations.

The implied market values for the BGT skill cluster families 'Economic Policy', 'Analysis', and 'Design' were the three highest with values of over \$100 for a marginal percentage point. While there is no direct correspondence with the Deming skill groups, the BGT skill cluster families seem to roughly correspond with the highest-valued Deming skill groups. Those were 'ML, AI' as well as 'Business Systems', 'General Software' and 'Creativity'.

### 1.4.2 What is driving Occupational Change?

Before looking at occupational changes in skill shares, we present evidence for the (overall) demands for skills (Figure (4-2)) and, specifically for IT skills (Figure (4-3)). We can immediately see that there is considerable variation in the demands for different skills. For example, we can see the tremendous rise in demand for 'Big Data' and 'Artificial Intelligence' in high-wage occupations since about 2012, and 2016, respectively. These figures are on a log scale, so the demand for these increased by a magnitude of over 1000. Other important IT skills for high-paid jobs include Java, Javascript, and SQL which each made up almost 1 of every 1000 IT skills demanded in job postings. Considering that IT skills made about 5%, i.e. 1 in every 20 skills demanded in high-wage job postings, these skills appear to be particularly good investments. Another interesting observation, among others, is the shift of health skills from medium-wage occupation to low-wage ones.

Going one level deeper into the aggregation, we can study more detailed skill demands at the occupation level. Figures (4-4) and Figure (4-5) show the yearly skill share distribution for data scientists and lumberjacks, respectively. In each year column we can see the share that each skill made up of all skills demanded in job postings associated with these occupations in that year. Again, we can see the rise of 'Big Data' (dark blue) and 'Data Science' (blue) starting in about 2012 for Data Scientists. Naturally, the occupation of lumberjacks had very different skill demands, that include landscaping and, increasingly, agronomy and farming. It is apparent that the skill distribution for this occupation seems to have changed between 2010 and 2018 and we can measure this change with the aforementioned cosine distance.

We do not purport to know why, specifically, certain occupational skill demands changed or appear to be high. A priori, we would not have expected the occupation of lumberjacks to undergo such changes. Notably, these changes cannot be due to the BGT skills parser classifying skills inconsistently across time. If it indeed misclassified skill words, it would so *consistently* across time. Thus, sudden jumps in skill demands are due to actual changes in the language of the underlying job postings, for example due to the increased popularity of one word over its synonym. While these changes in language may not represent changes in actual skill demands, they are still important for workers as they can help them better match and be accepted for a job posting. Moreover, by only counting skills within the same skill cluster once instead of multiple times, we can at least partially account for some of these worries, as, for example similar skills such as 'AWS' and 'Azure' are both contained with the skill cluster 'Cloud' skills.

In Figure (4-6), we plot the cosine distance between 2010 and 2018 skill share vectors over the 2010 wage percentile (each point represents a different occupation). In the literature, the wage percentile is also referred to as the skill percentile, assuming that higher wage proxies higher skill [Autor and Dorn, 2013]. If this assumption holds, the figure supports the 'skill'-biased technological change theory. Lower-wage and medium-wage occupations incur higher changes to their skill share distribution than higher wage ones. These trends remain when the analyses are repeated at the occupation-industry level. A similar positive association between wage and skill share change can be see in Figure (4-7), which plots the same variables as Figure (4-6) but each point represents an occupation-industry combination rather than an occupation.

The above patterns support one leading theory of technological change (SBTC). To also test for the implications of and discern multiple leading theories of technical change, we turn to regression analyses. We regress the cosine distance on proxy variables for each of the three leading technological change theories: (1) SBTC, (2) RBTC, and (3) Directed Technological Change:

$$d_{cos}(s_{i,2010}, s_{i,2018}) = \beta_0 + \beta_1 SML_i + \beta_2 w_{i,2010} + \beta_3 w_{i,2010}^2 + \beta_4 wagebill_{i,2010} + \epsilon_i$$

**(1) SBTC**: To account for SBTC we follow the standard approach used for example by Autor, Dorn (2013) and use real wages from the earliest period, in this case 2010. Given that wages are polarized, we include squared wages in two of the columns of table (4.1). In the other columns we use wage tercile dummies to allow for more flexibility in the model.

**(2) RBTC**: To account for RBTC, we use the SML scores from Brynjolfsson, Mitchell, Rock (2018). The suitability for Machine Learning of an occupation is a good measure for this, because routine tasks are exactly the tasks that ML is particularly suitable for. We prefer this measure to the Routine Task Intensity (RTI) scores from Autor, Dorn (2013) since the latter relies on 5 non-maintained measures from the former Dictionary of Occupational Titles (now O*Net) [2].

**(3) Directed Technological Change**: According to Acemoglu (1998), technological change happens endogenously and its direction is determined by the size of the market for different inventions. Thus, if technological change is assumed to happen at the occupation level, it should correspond with the total wage bill of occupations, i.e. the product of wage and total employment of that occupation.

The results are shown in table (4.1). Notably, the SML score, our measure of RBTC, is positively correlated with occupational change, i.e. more routine occupations are associated with more changes in skill demands, ceteris paribus. SBTC is not significant. It is somewhat puzzling that the wage bill is significant but with a sign opposite of what Acemoglu (1998) would have predicted. Perhaps, an additional factor for the direction of technological change, besides the profitability of innovation, is the level of difficulty of innovation. This would be plausible if ideas are indeed getting harder to find, the more progress has been made [Bloom et al., 2020].

The fixed-effects models in columns (2) and (3) of table (4.1) show that the implied market values of skills differ considerably. The coefficients can be interpreted as the effective Dollar-amount associated with one additional percentage point in the skill share distribution of an occupation-industry cell. For example, an additional percentage point of 'Economics, Policy' skills is associated with an increase in wage of more than \$400. The fact that some skills have negative coefficients implies that these skills are not a good investment as the market does not value them. However, this does not mean that they are useless - given the bundling of skills into occupations,

[2] Specifically, manual tasks are defined as the DOT score for 'eye-hand-foot coordination', routine tasks are defined as the average of the 'Set limits, tolerances, and standards' and 'finger dexterity' DOT scores, and abstract tasks are defined as the average of 'direction control and planning' and the 'GED Math' score.

they may still be required to perform those jobs.

We present alternative estimates relying on the Deming (2018) definition of skills in table (4.4). The interpretation of these results is identical, except that the skill categories differ. We can immediately see the high implied market values for technical skills such as 'Machine Learning', 'AI', and 'General Software' skills. Conversely, non-cognitive skills are associated with negative values. This is in line with the theory of RBTC, as these types of skills are more routine and thus better relative substitutes for automation.

## 1.5 Conclusion

Technological change is essential to human progress. However, it also bears risks for some who may be left behind and displaced. Thus, GPTs like Machine Learning, which have the capability to drastically alter society and bring about immense progress, also induce the largest risks. We have shown that what is demanded of workers in terms of skills has changed over the past decade. The fact that low and medium-wage occupations' skill demands changed more than high-wage ones means that besides getting paid less, workers in the former also have to reskill more in order stay productive in and attempt to keep their jobs. However, reskilling may also offer opportunities for social mobility: some skills are highly valued by the market and may well be worth the time and effort investment - in fact, Machine Learning is one of them among other technical skills as well as creative and social skills. Not everyone will be able to acquire or benefit from these skills. In fact, it remains an open question which skills complement each other best in terms of productivity as well as learning.

There have been several theories on how technological innovation progresses - routine-biased (RBTC), wage-biased (SBTC), and endogenously-directed technological change. As automation is becoming more ubiquitous, the routineness of tasks seems so far to be one of the better indicators for where technological progress will progress fastest and therefore where corresponding changes in occupational skill demands will occur.

# Chapter 2

# Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance


## Abstract

Digital technologies and work-from-home practices can make jobs, firms, and industries more resilient to unanticipated shocks by reducing operational risks. We extract data from over 200 million U.S. job postings to construct a firm-level work-from-home (WFH) feasibility index by assessing firms' labor demand. Using a difference-in-differences (DiD) framework, we then demonstrate that firms who happened to have a high pre-pandemic WFH index had significantly better operational performance with higher sales, net incomes, stock market returns, and lower volatility during the pandemic compared to their industry peers. These results are particularly pronounced for firms in non-essential industries, where WFH feasibility was necessary to continue operation, as well as in low-tech industries, where WFH-enabled digital resilience led to larger comparative advantages over competitors. Lastly, confirming the complementarity between digital technologies and WFH practices, we find that firms with lower pre-pandemic WFH feasibility subsequently made proportionately greater software investments and hired more IT workers, in an attempt to catch up to their more digitally resilient competitors. Our results are robust to a stringent set of empirical specifications, industry-specific demand shocks and pre-trend controls, fixed effects, and falsification tests, suggesting that the identified effects of this natural experiment are likely causal. The findings in our study imply that firms need to strategically manage the labor composition and digitization of their organizations, and consider that work-from-home practices, besides their many other advantages, are an effective way to hedge against operational risks.

"We are finding we can reorganize our companies electronically very rapidly and that's the only type of organization that can begin to keep pace with the changing business conditions."

- Steve Jobs, 1990

## 2.1 Introduction

As digital technologies proliferate, work-from-home (WFH) practices have become an increasingly common practice and constitute an important dimension in the future of work.[1] A recent survey by Gallup estimated that more than 43% of workers reported working remotely at least once a week in 2017. Due to the Covid-19 pandemic, this number rose sharply during April and May of 2020 [Brynjolfsson et al., 2020] and is expected to increase significantly in the coming decades.[2] While the unanticipated coronavirus outbreak highlights the importance of work flexibility and has forced many firms to shift to WFH practices as the new work norm, firms varied greatly in their adoption of WFH practices before the onset of the outbreak. Thus, when the pandemic hit, some firms happened to be luckier and more prepared than their competitors. In this paper, we argue that due to the unanticipated nature of this shock, the pandemic constitutes a quasi-natural experiment - firms were unable to endogenously adjust their labor mix and operations beforehand and could only respond to it afterwards. Using over 200 million job postings from the near-universe of US online job postings from the Burning Glass Technology (BGT) database, we build a quarterly firm-level WFH feasibility index and show that firms with a higher pre-pandemic WFH feasibility index, as a measure of firms' resilience, weathered the pandemic significantly better than their low-WFH, within-industry peers on a wide range of financial and other firm-level outcomes. During the pandemic, these

[1]See U.S. Census report: https://www.census.gov/library/visualizations/2013/comm/home_based_workers.html.

[2]Chokshi, Niraj. Out of the Office: More People Are Working Remotely, Survey Finds, New York Times (February 15, 2017); https://www.nytimes.com/2017/02/15/us/remote-workers-work-from-home.html.

low-WFH peers attempted to catch up, by shifting a significantly higher percentage of their investments towards software capital as well as posting more job openings demanding IT-related skills to enable remote operation.

Of course, WFH practices already offered advantages before any pandemic(s). They afford workers increased flexibility in their work schedules, accommodate commuting, childcare, physical, and mental health needs, while allowing employer to retain, or even increase, worker productivity [Martin and MacDonnell, 2012, Bloom et al., 2015, Alipour et al., 2021, Sánchez et al., 2007, Martínez-Sánchez et al., 2008]. However, it wasn't until the pandemic hit that it became clear that WFH practices significantly lowered firms' operational risks [Alipour et al., 2021]. The feasibility of transitioning to and effectively implementing WFH practices in combination with digital technologies, such as Zoom and cloud computing, enabled many firms to quickly reorganize their organization, continue operating, in particular for multinational or cross-border firms, or reach relatively higher operational efficiency even while adhering to government-issued shutdowns and stay-at-home orders. In light of increasingly common extreme weather events, including wildfires, heatwaves, floods, and storms, among others, having flexible WFH and digital practices in place will offer additional value added. In fact, the resilience and ability to hedge against the extreme operational risks of these crisis events that WFH feasibility enables, are arguably even more important than the everyday work conveniences it can offer.

During the crisis, many tech companies introduced long-term, or even permanent, remote work policies[3]. This suggests that even outside of crisis events, WFH practices will be here to stay, at least in a hybrid or mixed-mode work model [Barrero et al., 2021], though further research will be needed to assess the long-term costs and benefits, as well as which types of workers can work remotely.

Not all jobs can feasibly be shifted to remote work. Only 37% of the roughly 900 occupations that the Bureau of Labor Statistics (BLS) defines in the Standard Occupational Classification (SOC) system can be fully performed remotely [Dingel

[3]Benveniste, Alexis. These companies' workers may never go back to the office. CNN (18 October, 2020); https://cnn.it/3jIobzJ McLean, Rob. These companies plan to make working from home the new normal. As in forever. CNN (25 June, 2020); https://cnn.it/3ebJU27

and Neiman, 2020]. This renders some industries inherently more vulnerable and potentially unable to operate during any crises that constrain labor mobility. Within industries, this also renders firms with a lower WFH feasiblity of their employment mix more vulnerable and rigid, and therefore unable to shift to WFH to continue operation during such crises (e.g., the Covid-19 pandemic). Thus, while virtually all firms would be incentivized to shift to WFH practices when faced by a crisis that constrains labor mobility and/or increases work-related hazards and risks, implementing these practices is going to be more feasible for some firms than for others. This variation in work-from-home feasibility exists even for competitors within the same industry, as their operations to produce similar outputs can differ considerably (See Figure 5-1).

Since early 2020, the current pandemic along with government-issued stay-at-home orders forced many firms to reduce basic operations or even cease operation entirely. This especially affected non-essential industries, leading to substantial, and potentially permanent, loss of sales, customers, and profits. Even when firms were allowed to continue their operations, the risk of Covid-19 and safety measures to reduce infections often significantly reduced the efficiency of on-premise employees and operations compared to WFH employees. Thus, all else equal, the feasibility of shifting employees to WFH should have helped firms become more resilient during the pandemic. In other words, the WFH feasibility of firms' pre-pandemic workforce can serve as an exogenous shifter of the impact of the pandemic on firms - a quasi-experimental design setting to explore the impact of WFH feasibility on firm resilience to crises. In addition, firms with a lower pre-pandemic WFH feasibility index, whose operations were significantly affected by the pandemic, may have had to shift their hiring mix and investments towards digital technologies and IT talent to hedge against the crisis and enable remote work, much more so than the ones that already had related human capital and infrastructure in place prior to the pandemic.

In the early stage of the current pandemic, government shutdowns prevented firms in non-essential industries from operating if they could not do so remotely. These measures provide an additional exogenous shock to firms within those industries, which allows us to further identify the causal effects of WFH feasibility on firm outcomes. If firms' WFH feasibility is truly associated with firms resilience to crises, a larger effect of WFH feasibility is expected for firms in industries which were deemed non-essential.

While maybe less obvious at first, the firms in low IT intensity industries might have a larger comparative advantage if they have a high pre-pandemic WFH feasibility index. While the aforementioned arguments certainly apparent in hindsight, the lack of firm-level WFH feasibility data prevented this type of study earlier on. Perhaps more importantly, the exact magnitudes of the performance differences (across and within industries) that WFH-enabled resilience made is very important to quantify for firms and managers alike to understand the costs and benefits.

To test our hypotheses, we merge the near-universe of U.S. job postings with the occupational WFH feasibility indicator from [Dingel and Neiman, 2020] to construct, for the first time, a measure of the percentage of firms' workforce that has the WFH option – a firm-level WFH feasibility index. We merge these data with financial and subsidary information for public firms from Compustat and Orbis using fuzzy name matching methods. We further augment these data with firm-level stock return and volatility information from the Center for Research in Security Prices (CRSP). We then estimate a standard Difference-in-Differences (DiD) regression model to examine whether and how much a firm's WFH feasibility *prior to* the Covid-19 pandemic influences its financial and stock market performance during this crisis by comparing the firms with high and low pre-pandemic WFH feasibility index.

Our main identifying assumption for causal estimates is that firms' pre-pandemic WFH index is orthogonal to the timing of the Covid-19 pandemic. In other words, firms were unable to anticipate the pandemic and were unable to rapidly and fully prepare for it by hiring for more WFH-feasible jobs or preemptively altering their business operation in the short-term [Ilut et al., 2018]. We believe this to be a reasonable assumption, given the unanticipated nature of the pandemic as well as the delayed governmental and state responses. To further alleviate identification concerns, we also control for industry-specific demand shocks and perform several other robustness checks.

Our key finding is that firms with high pre-pandemic WFH feasibility, and thus higher resilience to the pandemic, fared significantly better – with roughly 15% higher net incomes, 4% higher sales, and better stock market performance (measured by stock returns and volatility) compared to firms with lower pre-pandemic WFH indices. In absolute terms, firms with high pre-pandemic WFH indices on average retained more than $35 million higher net incomes and $392 million higher sales. In contrast, firms with lower pre-Covid-19 WFH indices only retained $13.2 million higher net incomes and $134 million higher sales. Compared to their high-WFH counterparts, these low-WFH firms also retained a significantly higher percentage of software investment as well as demand for IT-related workers, as they attempt to catch up in IT capabilities to facilitate remote work and to continue operation during the pandemic. These results are robust to empirical specifications that control for a host of firm-level characteristics, demand shocks, fixed effects, matching methods, and falsification tests, suggesting the identified effects are likely casual.

We also find that pre-pandemic WFH feasibility is associated with relatively better performance during the pandemic for (i) firms in non-essential industries (vs. those in essential ones) and (ii) for firms in low-tech (vs. those in high-tech ones). Resilience through WFH practices was most pronounced for firms in non-essential industries, because these were most adversely affected by Covid-19 and the associated government stay-at-home and shutdown orders. Firms in non-essential industries with lower pre-pandemic WFH indices also spent a significantly larger share on IT investments, while essential firms, which were legally guided to continue operation, did not. Moreover, the importance of WFH practices was also elevated in low-tech industries since high WFH scores, and thus resilience, were rare among firms in these industries, such that WFH feasibility led to larger comparative advantages over competitors. Conversely, in high-tech industries, a large share of firms were already able to support remote work, such that even the comparatively lower WFH firms in high-tech industries did not fare significantly worse than their high WFH peers. Our findings provide further evidence that WFH practices and complementary digital technologies were critical to continue operation [De' et al., 2020], especially for non-essential businesses which were forced to adjust.

Our study contributes to two strands of the literature. By studying the determinants of firm resilience during the Covid-19 pandemic, our paper joins a recent body of research that examines contributing factors to firms' differential performance and resilience during crisis episodes. [Albuquerque et al., 2020] and [Lins et al., 2017] find that firms with high environmental and social ratings tend to perform better during the Covid-19 pandemic and the 2007-2008 financial crisis, respectively. Other studies find evidence that firms with access to liquidity [Acharya and Steffen, 2020], high cash holdings [Ramelli and Wagner, 2020], more structural management practices [Schivardi et al., 2021], or a strong balance sheet [Ding et al., 2020] tended to perform better in the first quarter of 2020. During crises, firms also tend to accelerate the restructuring of their production towards routine-biased technologies, such as automation and robots, and their complementary workers [Hershbein and Kahn, 2018]. Our study expands the understanding of the determinants of firm resilience during a crisis period by taking a more labor-oriented focus. Specifically, our evidence suggests that digitally-enabled flexible work arrangements such as WFH can significantly reduce operational disruptions that firms experience in difficult times and can help with firms' resurgence in the aftermath of the crisis.[4]

Second, our study contributes to the recent literature that studies the shift of workplace norms towards more flexible work settings, such as working from home. Since the widespread rise of digital information and communication technologies and broadband internet access, remote work has become more and more accessible and adopted [alm, 2019]. In particular, [Brynjolfsson et al., 2020] survey a nationally-representative sample of the U.S. population and find that over 30 percent of workers

[4]Several other papers that examine the labor aspects during the Covid-19 crisis include the following: [Benzell et al., 2020] compare the actual closures of commercial locations to their recommendation of which ones should be closed first, and generate implications for the optimal sequence of re-openings when policymakers revive the economy. [Atkeson, 2020] analyzes the economic consequences of the Covid-19 pandemic and how they correlate with various assumptions about the ratio between the susceptible – infected – and recovered groups in the population. [Forsythe et al., 2020] and [Coibion et al., 2020] study the labor market implications of the Covid-19 pandemic. [Cui et al., 2021] find important gender inequality in academic productivity. [Choudhury et al., 2020] show that work-from-anywhere (WFA) may yield additional worker productivity gains beyond WFH measures.

switched to remote work due to the Covid-19 pandemic. This further supports our use of WFH feasibility as a measure of pandemic resilience. By conditioning on firms' ex ante WFH indices we mitigate endogeneity concerns that may arise from heterogeneity in firms' shift towards WFH practices *during* the pandemic. Thus, by investigating how differences in resilience lead to differential performance during the pandemic, we also provide further evidence for the value created by digitally-enabled WFH practices.

The remainder of the paper proceeds as follows. Section 2.2 develops a set of testable hypotheses. Section 2.3 discusses the data and summary statistics. We present our empirical findings in Section 2.4 and conclude in Section 2.5.

## 2.2 Hypothesis Development

A large body of prior literature explores the effect of WFH or remote work practices on work-related outcomes at the individual level including higher job satisfaction [Golden and Veiga, 2005, Golden, 2006], prolonged work hours and higher work efficiency [Madden and Jones, 2008, Bloom et al., 2015], lower worker stress [Gajendran and Harrison, 2007], and potentially lower turnover [Moen et al., 2011, Stavrou, 2005]. In contrast, relatively few studies focus on firm and organizational outcomes ([Bloom et al., 2015]; also see [Allen et al., 2015] for a review) possibly because large-scale firm-level data on the adoption of WFH practices is hard to come by. Among limited firm-level studies, the effect of WFH on firm performance is found to be positive [Meyer et al., 2001], but the magnitude tends to be small [Martin and MacDonnell, 2012] and contingent on complementary human resource management practices [Sánchez et al., 2007, Martínez-Sánchez et al., 2008]. However, despite the ability of WFH practices to improve worker performance and enable firm operations during crises and retain productivity, a recent study by [Yang et al., 2021] show that these practices caused collaboration networks among workers at Microsoft to become more static and siloed, which may lead to productivity losses in the long run, in particular for new trainee workers. Others also report on increased burnout and depression during extended periods of remote work [Hayes et al., 2020].

### 2.2.1 Higher Resilience Through WFH practices during Crises, such as Covid-19

Although the digital technologies that make WFH practices were known and implemented long before 2020, the global pandemic has drastically accelerated their adoption [Brynjolfsson et al., 2020]. The pandemic led the government, states, and local authorities to impose strict rules that limited mobility of employees through March and April (through stay-at-home orders). In addition, many workers voluntarily stayed home to minimize the risk of infection. In combination, firms had to aggressively apply digital tools and adjust towards WFH practices. These enabled many firms to continue operation during the pandemic. For instance, digital tools allowed firms' employees to maintain continuous communication with customers and suppliers while working remotely, thereby ensuring a less disrupted operation during the period of a lockdown. Digital tools and the WFH practices they enabled were therefore critical components of firm resilience against the pandemic.

However, remote work was more feasible for some firms and workers than others. While some industries, such as the information-services-providing industries, naturally had higher feasibility for WFH practices, there are still important within-industry differences between firms' adoption of these practices (See Figure 5-1). Firms rely on different production methods, management, technologies (e.g., levels of automation), or logistics and may thus have workforces that consist of very different occupations with varying degrees of WFH feasibility even in the same sector. Exploiting the variation of pre-pandemic WFH feasibility across firms (within industries), we therefore test our first hypothesis:

**Hypothesis 1** (WFH enabled higher pandemic resilience, lowered operational risks)**.** *Firms with high pre-pandemic WFH feasibility, were more resilient and fared significantly better during the pandemic on a wide range of firm-level outcomes compared to their low-WFH competitors.*

### 2.2.2 WFH and Complementary IT Investment

Digital technologies create the potential for increased resilience in firms' work arrangements. As early as the 1980s, [Olson, 1982] pointed out that the prospect of telecommuting work originated from the advancement of information and communication technology (ICT). Further development on storage, communication, and other transformational technologies unleashed the power of storage, transmission, and sharing of knowledge and information across time and space, facilitated coordination among geographically dispersed workers, and significantly transformed workplace practices [Leonardi and Bailey, 2008]. With the rise of the digital economy [Brynjolfsson and McAfee, 2014], a significantly higher percentage of jobs became suitable for WFH or remote work [Kizza, 2017]. Meanwhile, complementary ICT investment on both infrastructure and software (e.g. high-speed internet and Zoom) were on the rise, in particular due to drastically falling unit costs. Evidence for the rise of IT has been found in both case studies [Collins, 2005] and large-scale analyses [Oettinger, 2011]. As the pandemic suddenly forced firms to adopt WFH practices [Brynjolfsson et al., 2020, Yang et al., 2021], the laggards, namely firms with low pre-pandemic WFH feasibility, suffered significantly more, due to a lack of proper ICT infrastructure and software to support changes in operation. Thus, they likely had to sustain, if not increase, investment in ICT during the pandemic, to catch up and remain productive and operational.

**Hypothesis 2** (WFH Laggards increase complementary IT investments to catch up)**.** *Firms with low pre-pandemic WFH feasibility were forced to invest relatively more in remote-work complementary IT capital at the onset of the pandemic in order to stay operational and productive.*

### 2.2.3 Industry Differences (Essential vs. Non-essential, High-Tech vs. Low-Tech)

Although the aforementioned resilience-improving effect of WFH is dispersed widely across industries, the extent to how much firms benefit from WFH feasibility will be conditional on how much their operations were disrupted by the pandemic. For instance, stay-at-home orders affected some firms and industries more than others. A natural way to think about the level of constraint that the government order has put on different industries is through the definition of *essential industries*. If an industry was deemed essential, such as the food, retail, or health care industries, it was allowed to continue operation on premise during the pandemic [Papanikolaou and Schmidt, 2020]. Thus, firms in these industries were less likely to depend on the WFH feasibility of their employees. In contrast, many firms in *non-essential industries* were forced to cease or modify their operation due to the risk of Covid-19. Therefore, firms' feasibility to have their workers work remotely was crucial in these industries. Digital resilience, as measured by ability to work remotely, is likely to be a greater determinant of success or even survival during the pandemic in non-essential compared to essential industries.

**Hypothesis 3** (WFH had a greater effect in Non-Essential than Essential Industries)**.** *WFH feasibility had a positive effect on firm performance in all industries, but was particularly crucial in non-essential industries.*

In a similar vein, there are likely important performance differences between high- and low-tech industries [Decker et al., 2017]. Firms in high-tech industries were much more likely to have high WFH feasibility and/or implemented WFH practices, for example through hybrid or mixed work modes, even before the pandemic - the summary statistics of our WFH index confirm this (see Figure 5-1). Therefore, relatively more firms in these industries were able to switch to remote work and continue operating during the pandemic than firms in low-tech industries (e.g., Information sector vs. Retail sector). Conversely, for firms in low-tech industries, a high pre-pandemic WFH index likely offered significant competitive advantages during the pandemic.

**Hypothesis 4** (WFH had a greater effect in Low-Tech than High-Tech Industries.)**.** *WFH feasibility was more important for firm resilience in low-tech industries, where these practices offered larger comparative advantages than in high-tech industries, where WFH practices were already widely adopted.*

## 2.3 Data Construction and Summary Statistics

### 2.3.1 Data

#### WFH Index

To construct a firm-level WFH index, we proceed in several steps: First, we obtain detailed job postings data from Burning Glass Technologies (BGT) - a high-quality data source with comprehensive coverage of online job posting portals beginning in 2010 and with increasing popularity in recent economic research [Hershbein and Kahn, 2018, Deming and Kahn, 2018, Azar et al., 2020, Das et al., 2020, Acemoglu et al., 2020].[5] BGT annotates each job posting with: (i) an occupational title, based on the roughly 900 6-digit codes defined in the Bureau of Labor Statistics (BLS)' Standard Occupation Classification (SOC), (ii) an industry code, based on the roughly 1,200 6-digit codes defined in the North American Industry Classification System (NAICS), employer information, job posting date, detailed skill demands, and many more. Notably, the occupational SOC codes as well as the industry NAICS codes are created as hierarchical taxonomies, which means that multiple 6-digit codes are nested within more broadly-defined 4-digit codes, which again are nested within even more broadly-defined 2-digit codes. We aggregate this data to the (employer, SOC (6-digit), time period) level to obtain a measure for how many job postings each employer posts in every month, quarter, or year for each SOC occupation - it is thus a panel of employer's detailed labor demands.

Second, we merge this data with the WFH feasibility data from [Dingel and Neiman, 2020]. They provide a binary index at the 8-digit Occupational Information Network (O*NET) code level,[6] as well as a continuous, aggregated index at the 6-digit Standard Occupational Classification (SOC) level, which takes employment

[5] BGT provides partial coverage for 2007, but does not provide any coverage of job postings in 2008 or 2009. We therefore use BGT data starting from 2010.

[6] O*Net is a free online database that contains hundreds of occupational definitions to help students, job seekers, businesses and workforce development professionals to understand today's world of work in the United States. It was developed under the sponsorship of the US Department of Labor/Employment and Training Administration (USDOL/ETA) through a grant to the North Carolina Employment Security Commission (now part of the NC Commerce Department) during the 1990s.

shares into account.[7]

Finally, we take the weighted average over each firm's occupational WFH indices, weighted by its number of job postings for each occupation. This procedure enables us to construct, for each firm, the percentage of its labor demand that has WFH feasibility. Ideally, one would like to have such a measure constructed over all existing employees, which would capture the actual prevalence of WFH feasibility within an organization. Since such data is not available, we take the next best alternative and use the average value of each firm's quarterly WFH index over the 2010-2019 period. Since we use a standard DiD for our main specification, we create an indicator variable from this continuous WFH measure for sharp contrast but also show that our results are robust when using continuous measures.[8]

Notably, our measure is a forward-looking demand measure and we do not observe whether firms manage to fill these positions. In principle, it could bias our results if employers (job seekers) were able to anticipate the pandemic and adjust the job postings (offers) and hire (accept) towards occupations with higher WFH feasibility. However, due to the unanticipated nature of the pandemic, this seems highly unlikely. In fact, while we observed significant changes in average quarterly WFH indices during the pandemic (relative to the average indices in the first quarter of 2019), there are *no* significant differences among the quarters prior to the pandemic in 2019 regardless of whether firms had high or low pre-pandemic WFH indices as can be seen in Figure 5-2.

A more serious concern might be that we do not observe the large number of layoffs and furloughs during the pandemic. Again, for our main question in which we are interested in the effect of pre-pandemic WFH feasibility (firm resilience), this is

[7]Technically, their measure is at the 6-digit SOC hybrid level defined by the Bureau of Labor Statistics for the Occupational Employment Survey (OES). The BLS implemented this slightly altered taxonomy as an interim solution between the switch from the 2010 SOC system to the 2018 SOC system, which will be fully used for the 2020 OES. The BGT data uses the 6-digit 2010 SOC system, for which a crosswalk to the 6-digit hybrid system is readily available on the BLS website: https://www.bls.gov/oes/soc_2018.htm.

[8]Also note that our main results are robust and consistent when using median as an alternative cutoff or when comparing firms in the top quartile against firms in the bottom quartile. Results are not reported to save space but are available upon request.

not a concern.[9] Therefore, in the subsequent sections, high WFH index always refers to a high *pre-pandemic* WFH index, unless otherwise specified.

### IT-related Hiring

We also use the BGT data to construct a variable to measure firms' total number of job postings that request IT skills. The BGT data are ideal for this task since they allow us to observe firms' demand for specific skills from their job postings. BGT defines a taxonomy of roughly 16,000 skills, which are nested within about 800 skill clusters, which themselves are nested within about 30 distinct skill cluster families. For example, the programming language skill 'Python' falls within the skill cluster 'Coding and scripting languages', which itself belongs to the 'Information Technologies' (IT) skill cluster family. We use BGT's IT skill cluster family to capture all the IT-related skill demands and generate this variable at the quarterly level for each firm. Besides Python, mobile app development skills including swift, mobile application development, and web development skills such as CSS, Django, and HTML5 are a few examples of other skills that fall into the BGT IT skill cluster family. This variable allows us to test firms' labor market responses to the Covid pandemic specifically for IT-related needs.

### Accounting Fundamentals and Stock Return Data

We obtain quarterly accounting information from 2019 Q1 to 2020 Q3 from Compustat and stock return data from the Center for Research in Security Prices (CRSP).

[9]However, this does matter for studying how firms dynamically adjust their hiring during the pandemic. Firms with low pre-pandemic WFH indices, particularly in non-essential industries, likely had stronger pressure to fire workers in occupations that were not WFH feasible or hire more workers in occupations that permit remote work to continue operation. Many firms were forced to alter their business operations entirely towards higher work-from-home feasibility as can be seen in schools or management consulting. Since we do not observe the number of layoffs or reductions in hours, our WFH index during the pandemic is a noisy signal of resilience. A decline in a firm's WFH index during the pandemic may even be a sign of recovery as the firm begins to rehire previously fired workers in non-WFH occupations. Indeed, we do observe these trends and further analyze the recovery dynamics in a parallel study. Despite the potential endogeneity of the WFH indices *during the pandemic*, the *pre-pandemic* WFH indices are still an exogenous shifter of firms' pandemic performance, because the pandemic was unanticipated and significant change in WFH feasibility for large public firms through hiring is highly unlikely in the short-term [Ilut et al., 2018].

This data provides us key outcome and explanatory variables including sales, net income, capital expenditures, stock returns, and total assets.

#### Merging Data from Compustat and Burning Glass Technologies

Since Compustat and BGT do not share a common firm identifier, we take a multi-step approach to merge the two databases. Specifically, we use a combination of name and address fuzzy matching to construct a bridge between Compustat and BGT data. We use several methods including Soundex and Levenshtein distance to ensure match quality.

In some cases, an employer name in the BGT data is a subsidiary of a Compustat firm but its name is distinct from its parent, thus the existing algorithm cannot recognize their connection. To resolve this problem, we follow [Campello et al., 2019] and match the remaining employers to the subsidiaries of Compustat firms using information extracted from historical Orbis data provided by Bureau van Dijk (BvD). Orbis traces the evolution of firms' organizational structure through time, maintaining the parent-subsidiary correspondence. This historical information is robust to subsidiary opening, closing, and ownership changes, which is crucial for accurate matching.[10] We manually check the links identified to ensure the accuracy of our matching.

### 2.3.2 Summary Statistics

Following the procedure in section 2.3.1, we are able to match over 3,800 unique firms in Compustat. For instances where a Compustat firm has multiple subsidiaries, our firm-level WFH index uses a weighted average of all the subsidiaries' WFH indices, where the weight is each subsidiary's number of job postings. After the cleaning process,[11] our final analytical sample includes 9,550 observations corresponding to 2,176 unique public firms. The average WFH indices across sectors based on firms in our sample in both 2019 and 2020 are presented in Figure 5-3. Information, finance,

[10] We refer interested readers to [Campello et al., 2019] for a more detailed description of this part of the matching exercise.

[11] We require all observations to have non-missing information for key variables for our empirical exercise and hence have fewer unique firms in our analytic sample than in the matched sample.

education, and professional services are among the sectors with the highest WFH indices while the retail, health care, and accommodation sectors have the lowest average WFH indices for both 2019 and 2020, consistent with the general presumption.[12]

Table 5.1 Panel A shows that the mean value of the WFH index is 0.575 with a standard deviation of 0.286, which indicates a sizeable amount of variation in the cross section. Panel B further breaks down the key variables into pre- and post-Covid-19 periods and contrasts their values between subsamples of firms in the top quarter pre-Covid-19 WFH index (HighWFH=1) and the rest (HighWFH=0). We observe a differential impact of Covid-19 on high- and low-WFH firms. For instance, while the two groups of firms have similar average quarterly stock returns before the pandemic (7.0% vs. 6.9%), the high-WFH firms experience a relatively higher cumulative return during the period after the breakout of the disease than the low-WFH firms (4.3% vs. -1.5%). Similarly, while the average net income of high-WFH firms increased modestly from $281.7 million in 2019 to $316.8 million in 2020, the measure for low-WFH firms only increased from $199.3 million to $212.5 million over the same period. This indicates a significantly higher growth of net income (12.5% vs. 6.6%) between the high-WFH firms and the rest of the firms during the same period. Although large in magnitude, these differences in means could be driven by the distribution of WFH indices within and/or across industry and therefore demand multivariate analyses, which we turn to in the next section.

[12] Note that the 2020 average WFH indices are less informative due to the fact that we can only observe hiring but not firing. This may matter particularly for low-tech industries, where many frontline workers were let go due to stay-at-home and shutdown orders but later rehired when these orders were rescinded. Through most of this paper we therefore focus on the *pre-pandemic* WFH index since the 2020 WFH index is more sensitive to firms' endogenous hiring activities.

## 2.4 Empirical Methodology & Results

### 2.4.1 Empirical Methodology

To test our hypotheses, we employ a difference-in-differences (DiD) research design[13]. Specifically, we estimate the following multivariate fixed effects regression:

$$Y_{it} = \beta(\text{HighWFH}_i \times \text{Covid-19}_t) + \gamma X_{i,t-1} + \upsilon_i + \tau_t + \epsilon_{i,t} \tag{2.1}$$

where subscript $i$ and $t$ index firm and time (i.e., quarter), respectively.[14] $Y_{it}$ specifies the firm-level outcome variables for both financial and stock market performance. In particular, we examine sales, net income, total capital investment, software capital investment, stock returns, and return volatility. $HighWFH_i$ is an indicator variable, which takes the value of one if firm $i$'s average WFH index calculated based on its annual job posting data during the pre-Covid-19 period (2010-2019) falls into the top quartile of the sample distribution and zero otherwise. $Covid-19_t$ is an indicator variable that is set equal to one for 2020 Q1-Q3 and zero otherwise. We chose to use an indicator variable instead of a continuous WFH index for our main empirical specification, because we were most interested in the differential impact on the most prepared firms and those are less prepared - the results remain significant and unchanged when using the continuous WFH index (see Appendix 8.3). The parameter of interest is $\beta$, which captures the differential impact of Covid-19 on firms with high versus the rest of the firms (with lower pre-pandemic WFH indices). X is an array of time-varying firm-level controls including firm size (total asset), cash holdings, leverage ratio, R&D indicator, and the dividend payout indicator. $\upsilon_i$ specifies firm-fixed effects, which controls for time-invariant firm-level characteristics. $\tau_t$ specifies the time-fixed effects. Note that the inclusion of both firm-fixed effects and time-fixed effects absorbs the main effect of *HighWFH* and *Covid-19*. Throughout our empirical

[13] Despite several recent developments on the DiD methodology [Callaway and Sant'Anna, 2020], most only pertain to non-traditional settings.

[14] In particular, by including firm and time fixed effects, we already control for individual firms' indicators for belonging to the top WFH quartile ($HighWFH_i$) as well as for the indicators for time periods during the pandemic ($Covid_t$).

analyses, robust standard errors are clustered at the firm level.[15]

### 2.4.2 Empirical Results - Baseline

In this section, we present regression results on the impact of pre-Covid-19 WFH on firm performance through the first three quarters of 2020.[16] We focus our attention primarily on firms' financial performance, IT investments and hiring, as well as stock market outcomes. In all specifications, we control for time and firm fixed effects to eliminate, as much as possible, common timing-varying transitory shocks across firms and unobserved time-invariant firm heterogeneity. Additional time-varying firm-level characteristics such as firm size, cash holdings, leverage ratio, and R&D are also controlled for in all specifications. In additional robustness checks, we also control for potentially heterogeneous demand shocks induced by the onset of the pandemic as well as a variety of other specifications.

#### Financial Performance

We first investigate whether high pre-pandemic WFH scores drive better financial performance during the pandemic (Table 5.2 columns 1-3), as hypothesized in Hypothesis 1. Results from our models indicate that firms that had high WFH indices prior to the Covid-19 pandemic earned 15.5% higher net incomes and 3.8% higher sales relative to their low-WFH peers (Columns 1 and 2, respectively) during the pandemic - both coefficients are significant at the 1% level, confirming Hypothesis 1. Not surprisingly, these firms also retained a significantly higher level of overall capital investment compared to their peers with lower pre-Covid-19 WFH indices. Conditional on our argument that the outbreak of the pandemic was unanticipated by firms and that hiring and labor compositions are slow to adjust, these results suggest that

[15] In an alternative specification, we augment the time-fixed effect with state-quarter fixed effects, which is motivated by the recent findings in [Brynjolfsson et al., 2020] that the Covid-19-induced switch to WFH is highly correlated with the incidence of the pandemic in each state. Our baseline results are fully retained after controlling for the interactions of time and state (based on firm headquarters) fixed effects.

[16] The analysis period is defined by the latest data available when the authors started the empirical analysis of this research project.

high WFH feasibility helped firms hedge against the sudden crisis that constrained labor mobility and that WFH practices led to better financial performance during the crisis.

### IT Investments

IT is essential to enable firms and their employees to work remotely. The recent pandemic has accelerated firms' adoption of WFH practices and IT for all firms, as both firms with high and low pre-pandemic WFH indices alike strove to continue their operation [Bloom et al., 2020]. Empirically, it is not obvious which firms might have invested a larger share in IT. On one hand, better performing, high WFH firms could further improve their IT infrastructure as well as their hiring for high WFH feasibility jobs and thereby enlarge their competitive advantages over their low WFH peers. Since larger firms tend to have higher WFH feasibility, this implies potentially large implications for market concentration and monopoly powers. On the other hand, low WFH firms may realize, in a potentially shocking and painful manner, the importance of WFH feasibility for their organization and employees given the operational and infection risks as well as the stay-at-home orders. As a remedy, they might invest more in their IT infrastructure and hiring for IT-related employees. Taking advantage of the richness of our analytic data which combines capital investment data from Compustat with detailed job posting and skill demands from BGT, we test this empirically using software capital investment and demand for IT-related skills (see Hypothesis 2).

Interestingly, the results in (Table 5.2 columns 4-5) show that high WFH firms allocated a smaller capital investment share towards software and posted significantly fewer IT-related jobs, confirming Hypothesis 2. These findings show that low WFH firms attempted to catch up in WFH feasibility by retaining a higher percentage of their investment in IT software and by hiring more IT-related human capitals. Firms with lower pre-Covid-19 WFH indices had to retain IT investment (both software and human capital) to continue their operation, while firms with high pre-Covid-19 WFH indices already happened to have these very valuable elements in place before the unprecedented pandemic, strongly suggesting that IT investments are complementary to WFH practices.[17]

### Stock Market Performance

Using a similar regression framework, we examine whether firms' stock market performance during the Covid-19 pandemic is correlated with their ex ante WFH feasibility. In March 2020, the Covid-19 pandemic and the oil price war between Russia and the Organization of the Petroleum Exporting Countries (OPEC) resulted in the most significant stock market crash in the last decade. In particular, between February 12, 2020 and March 23, 2020, all three major stock indices (i.e., the Dow Jones Industrial Average, the NASDAQ Composite, and S&P 500 Index) experienced larger than 20% declines.

Since reaching a bottom in March 2020, all three stock indices not only rebounded, but reached new heights by November 2020. This is due to a variety of reasons ranging from expectations of an effective vaccine, unprecedented liquidity injections by the Federal Reserve, to historic stimulus packages passed by Congress. At a micro-level, firms that were well-known to have high WFH feasibility (e.g. mainly firms in the Information sector) or that specialized in WFH-complementary technologies (e.g. Zoom and DocuSign) significantly outperformed traditional firms.

We examine specifically firm-level WFH feasibility on outcome variables such as stock returns and volatility, and report the results in Table 5.3. We find that, compared to low-WFH firms, firms with high levels of WFH prior to the Covid-19-induced shutdown had significantly higher returns and abnormal returns, and meanwhile, lower volatility and idiosyncratic volatility.[18] Concretely, our estimates in Column 2 of Table 5.3 suggest that, ceteris paribus, firms in the top pre-pandemic WFH quartile (i.e., *HighWFH=1*) earned a 4.3% higher abnormal return than other firms

[17] Note that there is a significant percentage of firms that reported missing capital investment on software in Compustat. We treat them as zeroes in the main analysis. However, our results are robust and consistent to removing those firms and relying only on firms with a non-missing history of reporting capital investments in software. Results are available upon request.

[18] Return volatility and idiosyncratic volatility are calculated as the standard deviation of daily returns over a quarter and that of residuals obtained from fitting a CAPM using daily returns in a quarter. Our construction of idiosyncratic volatility is in line with [Ang et al., 2006]. Table 5.1 Panel A details the variable definitions.

during the pandemic (controlling for fixed effects, as before). These results together suggest that high values of pre-pandemic WFH, as a proxy of firms' resilience to the pandemic, caused firms to perform significantly better.

**Robustness Tests**

We conduct several robustness checks to address additional concerns and obtain similar results as in the main findings. First, we examine the effect of WFH feasibility while controlling for industry specific demand shocks. The pandemic appears to have had uneven demand impacts across industries (e.g., IT industry v.s. accommodation) though this might be in part caused by firms ability to quickly transit to remote operations (i.e., WFH feasibility). Nevertheless, we test the robustness of our results by controlling for the quarterly average sales growth of the industries to which the focal firms belong as a proxy for industry demand shock.[19] As show by the results in Tables 5.4 and 5.5, the effect of pre-pandemic WFH feasibility remains similar to those in the baseline models reported in Tables 5.2 and 5.3.

An alternative concern could be that the identified effects may be confounded by time-varying heterogeneous geographical effects of the pandemic - high WFH firms may happen to operate in states that have less strict and/or shorter length stay-at-home orders or in states where the pandemic hit earlier or later than in other states. This would be a serious concern for small or even medium-sized businesses that operate exclusively in smaller local markets. However, the firms in our sample are large public firms that operate at a national level. To address this potential issue empirically, we re-estimate our baseline models controlling for time-varying state fixed-effects. Our results stay completely intact as shown in Tables A1A and A1B in Appendix 8.1.

Next, we explore the effect of WFH before and after the Covid-19 pandemic at the quarterly level to explore the potential existence of pre-trends. That is, the identified performance effect of high pre-Covid-19 WFH feasibility should only be evident in

[19] We calculate the industry-level quarterly average sales growth for firms in the Compustat data at the 6-digit NAICS level whenever possible.

the first quarter of 2020 and not in the quarters preceding the pandemic, as a mere continuation of a potential pre-trend. To address this concern, we examine the timing of the treatment effects and report our results in Figure 5-4 Panels A and B, and Figure 5-5. The specifications mirror our baseline models. The only modifications are that we explore the differences between firms with high vs. lower pre-Covid-19 WFH indices by quarter and control for potential industry demand shocks and geographical differences. These figures show that firms with high pre-Covid-19 WFH indices had either similar or somewhat worse performance prior to the pandemic with a sharp turning point between 2019 Q4 and 2020 Q1, which coincides with the timing of the pandemic and thus confirms that no significant pre-trends exist.

In addition, our identification strategy exploits the fact that the pandemic was unanticipated and that some firms happened to be more resilient than their peers due to their pre-Covid hiring. There is still a possibility that performance differences during the pandemic are driven by other, unobserved firm-level characteristics that our resilience measure does not capture. We thus perform both propensity score matching and coarsened exact matching to construct samples that consist of firms that are similar, based on pre-pandemic observables, but differ on their pre-pandemic WFH indices. More specifically, we match the firms using returns on asset and size prior to the pandemic within 2-digit NAICS-level industries based on each outcome variable.[20] Using these matched subsamples, we re-estimate our baseline specification with firm, time, and pair fixed-effects. The comparison is therefore between high and low WFH firms within the same sector with similar returns on asset and size before the pandemic. We find consistent and robust results which are similar to our main finding. The results are reported in Appendix 8.2.

In another robustness test we consider an alternative WFH proxy (WFH index) by using a continuous WFH index instead of an indicator as our key explanatory variable. As presented in Appendix 8.3, the results are consistent to those reported in the baseline model in Tables 5.2 and 5.3 as well.

[20] We match the treatment and control groups through propensity score matching using the nearest neighbor method imposing common support with 3 neighbors and 0.001 caliper.

Lastly, in order to rule out the possibility that our WFH indices are picking up random effects, we also perform a placebo test by constructing an alternative index using the rank of firms by firms' name in alphabetical order. The results are reported in Appendix 8.4. The mostly close-to-zero and insignificant coefficients in this test confirm that our baseline findings are not random.

As shown in all aforementioned robustness tests, our baseline results are fully retained when controlling for industry demand shocks or time-varying state heterogeneity, when using propensity score and coarsened exact matching samples, when using an alternative, continuous instead of indicator-based measure of WFH feasibility, and are not caused by random noise, as further confirmed by the firm name placebo test. Overall, these findings further demonstrate that our identified effects of WFH feasibility on firm performance during the crisis are of a causal nature and that WFH feasibility was critical for firm crisis operation and resilience.

### 2.4.3 Empirical Results - Industry Heterogeneity

So far, we reported the average effects of high WFH feasibility on firm performance using the entire analytic sample. In this section we explore specific groups of industries in more detail.

**Essential vs. Non-essential Industries**

In Table E1, we break down our sample and separately examine the impact of WFH in essential and non-essential industries based on 4-digit NAICS codes as in [Papanikolaou and Schmidt, 2020]).[21] Confirming Hypothesis 3, we find that high WFH firms in non-essential industries had 18.3% and 5.9% higher sales and net incomes, respectively, than their within-industry peers with lower pre-Covid-19 WFH indices. In contrast, for firms in essential industries, these differentials are generally much smaller and noisier. Meanwhile, we find more salient differences in results for stock market performance between the essential and non-essential industries. While high WFH

[21] For more details on the definition of the essential and non-essential industries (at 4-digit NAICS level), please see [Papanikolaou and Schmidt, 2020].

firms in non-essential industries enjoyed significantly higher returns and lower volatility than their peers, the effect of high WFH feasibility is more moderate in essential industries. These results support the argument that pre-pandemic WFH feasibility within non-essential industries played a crucial role for firms resilience and their ability to effectively continue their operation. In addition, our results indicate that the significant differences in capital investments on software primarily load on firms in non-essential industries where the continuation of operation hinged more strongly on such investments. This finding further corroborates the notion of complementarity between work-from-home practices and IT.[22]

### High-Tech vs. Low-Tech

We follow [Decker et al., 2017] to separate our sample into high-tech and low-tech industries and re-estimate our main regressions separately within these subsamples.[23] Conventional wisdom suggests that firms in high-tech industries are more able to adopt WFH due to the nature of their work, while firms in other industries generally do not have such flexibility and thus had, on average, lower WFH feasibility. Indeed, the mean and median of WFH indices for high-tech industries are 0.71 and 0.78 respectively with a standard error of 0.25 while the mean and median of WFH indices for other industries are 0.55 and 0.56 with a larger standard error of 0.29. The results of this exercise are presented in Table E2.

Panels A and B of Table E2 present the financial performance results for the high-tech and low-tech industries, respectively. Overall, we find that our results are consistently smaller and mostly insignificant in the high-tech industries, but are more pronounced in the low-tech industries. Panels C and D display a similar pattern for firms' stock market performance. While these results may appear somewhat

[22]We further explore the heterogeneous effects of WFH on firm resilience during the Covid-19 pandemic by breaking down our sample by sectors. Using the baseline specification similar to those in Tables 5.2 and 5.3, we estimate the marginal effects of the interaction term between the high WFH indicator and the pandemic indicator and plot them in Appendix 8.6 Figure F1. We find that the identified effect of WFH loads primarily on manufacturing, finance, insurance and real estate, and service sectors. The definitions of industry sectors are reported in Appendix G1.

[23]The high-tech industry is defined as in [Decker et al., 2017] and can be found in Appendix Table G2.

surprising at first glance, they are consistent with the notion in [Brynjolfsson and Milgrom, 2012]: once WFH practices are adopted, which is the case in the high-tech industry, further improving WFH adoption may not be feasible or may not generate additional comparative advantages in firm performance. Conversely, in low-tech industries, where WFH practices were not nearly as widely adopted, the early adoption of these practices offered significant competitive advantages.

#### Sector Variation

To further understand the cross-industry variation in effects of firms' WFH feasibility on firm performance during the pandemic, we split our analytic sample by major sectors and re-estimate the baseline specifications.[24] The results are presented in Figure F1 in Appendix 8.6.[25] The effect of pre-pandemic WFH feasibility appears more salient in the Manufacturing, Finance, Insurance, Real Estate, and Services sectors but were negligible in Wholesale and Retail Trade, Utilities, Transportation, and Others. Noticeably, the large confidence intervals due to much smaller sample sizes lead to much more imprecise estimates.

Although we find some industry and sector-level variation in the effects of the pre-pandemic WFH index on firm resilience during the pandemic, we note that our sample is comprised of large public firms that are likely operating in multiple industries simultaneously. This could lead to measurement errors, especially when firms are relatively evenly split among multiple industries.

## 2.5 Conclusion

The advancement of digital tools over the last several decades allows workers to work from home (or anywhere) and enables firms to continue their operation with remote labor force. Many potential benefits are reported when firms adopt these digital tools and keep high workforce capability to work from home. For instance, remote

[24] To preserve statistical power, we focus on sectors instead of more refined industries to ensure that we have a large enough number of firms.

[25] The estimated coefficients from these models are also available upon request.

work increases work schedule flexibility, reduces (time and financial) costs of commuting, accommodates childcare, physical, and mental health needs, while also allowing employers to retain, or even increase, worker productivity and moral during 'normal times'. These benefits are even more pronounced during challenging crisis times, such as the Covid-19 pandemic, wildfires, blizzards, storms and other increasingly common natural disasters. Such crises constrain labor mobility and increase risk of on-premise operations and worker commutes. WFH practices can help keep workers safe and productive and allow firms to hedge against the operational risks due to the unexpected crisis events. Thus, they are a strategic form of insurance against the potentially devastating loss and bankruptcy that may result from sudden shocks. By investing in work-from-home practices and complementary IT capital in preparation for unexpected crises, firms gain resilience and can avoid the potentially rushed, stressful, more costly, and demand-constrained implementation of these practices during times of widespread need.

Many firms might have survived the Covid-19 pandemic if they had invested in remote practices and IT capital, but large-scale evidence on the actual effectiveness and effect magnitude of WFH practices is scarce. In this paper, we exploit data from the near-universe of U.S. job postings to construct a novel firm-level WFH index to gain insight into the resilience that remote work provides. We then use the current pandemic as a natural experiment setting and demonstrate that firms with high pre-pandemic WFH index values before the unanticipated crisis performed significantly better during the crisis compared to their peers on several dimensions ranging from financial performance, such as sales and net income, to stock returns, and return volatility. Further, during the pandemic the WFH laggards attempted to catch up by investing more heavily in complementary IT capital.[26] The magnitude and significance of our results are robust to a range of robustness checks including tests for pre-trends, controlling for industry-specific demand shocks, implementing propensity score and coarsened exact matchings, using continuous and indicator measures for

[26]Though more research on firms' hiring responses will be needed, because job postings data during the pandemic is likely less reliable as it does not capture the firing, only the (re)hiring, of non-essential workers nor the increased heterogeneity in filling job postings

the WFH index, and the inclusion of state and quarter fixed effects. An additional falsification test also suggests that the identified effect of the WFH feasibility is unlikely to be caused by random patterns. Consistent with our hypotheses, we also find important industry heterogeneity - non-essential industries, which were more vulnerable to government mandated restrictions on traditional work arrangements, benefited significantly more from WFH practices compared to essential industries. Low-tech industries also benefited more from WFH practices compared to high-tech industries during the pandemic, due to larger comparative advantages and lower pre-pandemic adoption rates of such practices.

Overall, our study shows that the effect of WFH feasibility on firms' resilience during crises is large, positive, and statistically significant. This causal effect is more pronounced in industries where we expected higher benefits. Our study contributes to the literature by constructing, for the first time, a firm-level WFH feasibility index, providing some of the first evidence on how WFH practices can help firms cope with major adverse social and economic shocks, and quantifying the magnitude and significance of such effects within and across industries.

Further investigation into the exact mechanisms through which WFH practices are linked to performance and risk reduction is a fruitful area for future research. A deep understanding of these underlying issues is particularly valuable to ensure a more efficient and smooth adoption of and transition into WFH as employers and employees alike embrace the new reality in the aftermath of the crisis. Most likely a hybrid work arrangement will arise, which enables worker flexibility and safety through remote work, while still enabling the crucial interpersonal ties interactions of in-office work. In case of crises, switching to fully remote work will be significantly easier when hybrid work arrangements and complementary IT capital are already in place, as they insure against operational risk. Striking the right balance between cost efficiency and supporting the firms and individuals who suffered the most from the pandemic will be critical to ensure a faster and equitable recovery.

One troublesome pattern that we find in our analytic sample is that the pandemic appears to have further increased within-industry inequality and market concentration

- see Figure H1. Given the intuition that larger firms tend to be more digital with higher capabilities for remote work, and suffered significantly less, WFH feasibility may have contributed to increased industry concentration as well as monopoly powers of superstar firms, which requires further investigation.[27]

[27] We see this as an important but separate research opportunity and explore further on the issue in a parallel research project.

# Chapter 3

# Treating the Symptoms or the Cause? Substantive and Symbolic Talent Acquisition in Response to Data Breaches


## Abstract

Do firms react to data breaches by investing in cybersecurity talent? Or, are they more likely to invest in talent to 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 hiring demand data, we leverage a difference-in-differences (DiD) design and find that firms increase their hiring for both cybersecurity as well as public relations and legal workers after a breach. We investigate whether public scrutiny serves as an effective mechanism to better align firms' incentives with those of the public. Gathering additional data on media and online search attention around data breaches, we find that public scrutiny may shift firms' investments toward substantive, rather than symbolic, measures. Given the increase in the volume and severity of cyberattacks, our results provide important and timely insights into firms' responses and incentives to more substantively safeguard their data.

## 3.1 Introduction

The digitization of business activities has led to an ever-increasing stream of digital information and data. 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 increasingly challenging 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 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. And yet, 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, Gordon et al., 2015b]. 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 [Ro-

[1] Source: 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)

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

manosky 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, and very rare, 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 for firms' lack of incentives and proper responses to data breaches, little is known about the effects of data breaches on firms' post-incident human capital investments. Following [Angst et al., 2017], we leverage neo-institutional theory to distinguish between the symbolic and substantive adoption of protective organizational practices in hiring. Under this lens, substantive adoption represents the adoption of practices that are tightly coupled with action, namely treating the cause of the cyber vulnerability, while symbolic adoption is only loosely coupled with such action. We argue that human capital investment, in the form of hiring for cybersecurity personnel, 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.[5] Indeed, according to a recent Forbes article, "security starts with people", implying that cybersecurity talent is one of the most crucial investments that firms need to make in today's digital world.[6] Given the critical role of centralized IT governance in deterring and preventing data breaches [Liu et al.,

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

[5]The closest papers are [Say and Vasudeva, 2020] and [Hilary et al., 2016], which explore management turnover in response to data breaches.

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

2020], it is important to understand firms' human capital investment in cybersecurity after suffering data breaches.

However, in the past, 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. We view these human capital investments as important complements to firms' IT capital investments and, following [Angst et al., 2017], view them through the lens of substantive and symbolic adoption.

Our results indicate that firms that suffer a data breach are two percentage points more likely to post cybersecurity-related job postings after the 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 skill and occupation-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, i.e. the aftermath, of a data breach instead of more substantive ones that directly treat the cause, the cyber vulnerability. Following the National Institute of Standards and Technology (NIST) framework [Petersen et al., 2020], we expand our analysis to include both hiring intents for legal talent, which resolves potential legal issues and assures compliance with applicable privacy laws, regulations, and constitutional requirements, as well as for public relations talent, which counters negative media attention and manages the firm's image after suffering a data breach. We find a significant increase in this kind of symbolic hiring, but that this increase is significantly smaller than that for substantive hiring.

Next, we conduct a series of mechanism tests to explore the heterogeneity of the treatment effect across industries and firms' listing status. We start our exploration by comparing goods-producing with service-providing industries and find that our results are largely driven by firms in service-providing industries, likely because firms in those industries tend to rely much more on customer trust, as they deal with more sensitive customer data, such as financial information. Our results indicate that firms in these industries are indeed more likely to hire both cybersecurity and other related human capital while firms in goods-producing industries tend to be unresponsive.[7]

Second, we compare the data breach event responses between 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, which suggests that they should adopt more substantive measures. On the other hand, prior research shows that the 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. Indeed, we find that while both public and private firms engage in significantly more symbolic hiring, this effect is even larger for public firms. For substantive hiring, we find a significant effect for private firms as well, while the picture is less clear for public firms.

Finally, we test the hypothesis that public scrutiny may serve as an effective mechanism to move firms more towards adoption of substantive cybersecurity mea-

[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.

sures and thus that public scrutiny can help to realign firms' incentives with those of the public. We leverage public search data from Google Trends and data on mentions in online news outlets from the MIT Media Cloud to measure the public scrutiny that breached firms face. We show that public scrutiny significantly increases firms' substantive hiring investments while not increasing their symbolic ones. Using outlier detection, we also identify data breach events with particularly high levels of public scrutiny and find suggestive evidence that firms shift even more strongly away from symbolic towards substantive hiring after suffering such a highly visible data security and trust crisis.

The rest of this article is organized as follows. Section 3.2 discusses the relevant literature; Section 3.3 describes the data; Section 3.4 introduces the empirical methods; Section 3.5 and 3.6 present the main results and a series of effect heterogeneity; Section 3.7 investigates public visibility as a mechanism for increased talent demand as a response to data breaches; then section 3.8 concludes.

## 3.2 Literature Review and Hypothesis Development

A series of papers investigate how cybersecurity investments can help firms 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 been linked to higher market values [Bose and Leung, 2013, Bose and Leung, 2019, Aytes et al., 2006] and lower capital costs [Havakhor et al., 2020] and can thus enhance firms' overall 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. Thus, more IT security investment is not necessarily better - these investments have to be directed towards

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

achieving meaningful technical benefits instead of being mere external signals of legitimacy (such as through the sticker of an alarm system instead of the deployment of an actual alarm system). Similarly, [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] suggest several substantive types of actions firms should take to effectively respond to hacks, including investments into hiring cybersecurity professionals, enhancing internal cybersecurity capabilities, as well 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].

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] 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. A recent study [Liu et al., 2020] provides empirical evidence from a large-scale sample that centralized IT governance is more effective in reducing the risk of data breaches than outsourcing.

Despite this guidance for effective data breach management, firms likely still underinvest in cybersecurity [Telang, 2015] and, importantly, may do so deliberately and rationally [Gordon and Loeb, 2006]. One of the existing explanations for this hesitance is the intangible nature of many of the benefits of cybersecurity investments, which go far beyond the tangible positive market returns of announcing such investments [Bose and Leung, 2019], and which are largely invisible in the short-run. Chief among these intangible benefits is the trust of consumers and business partners [Solove, 2007], which may be one of the key drivers of cybersecurity adoption as trust can easily be lost, such as through 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 - of the problem.

Indeed, prior research suggests that the direct financial costs that firms suffer after a data breach are small. 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.

The literature on cybersecurity is well developed for the financial impacts of data breaches. 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 must 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] We therefore 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., experienced a data breach event). Our first hypothesis to test is therefore:

**Hypothesis 5** (Substantive Adoption - Cybersecurity Hiring)**.** *To treat the cause of a data breach and improve cybersecurity, firms increase substantive adoption after suffering a data breach by increasing their demand for cybersecurity workers and skills.*

However, prior literature has shown that firms lack incentives to adopt cybersecurity measures in a substantive manner [Gordon et al., 2015a, Gordon et al., 2015b]. 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]. This suggests that 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. In addition, the Sarbanes-Oxley Act (SOX) requires public firms to prove their cybersecurity credentials[10]. 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. 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 talent in public relations and legal occupations may also change due to these shocks:

**Hypothesis 6** (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 PR and Legal workers and skills.*

[9] One particular concern here is that our identified effect would be significantly downward biased if most firms outsource their cybersecurity needs after a data breach. However, this would imply that we should observe significantly larger (capital) investments in cybersecurity after data breaches, which is inconsistent with the findings in prior studies. In addition, even firms which relied on outsourced cybersecurity investments before suffering a breach might increase hiring their own cybersecurity talent afterwards, due to their needs to increase the security effectiveness [Huang et al., 2018]. Therefore, it is unlikely that our results are significantly affected by omitted variable bias due to outsourcing.

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

While it is crucial to reveal firms' actual human capital investment responses after data breach events, there is still a lack of understanding regarding what types of firms would be most motivated to tackle cybersecurity in a substantive manner. Firms generally lack significant 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, 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, in part due to 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 customers, but also for non-customers [Choi et al., 2019]. Thus, even if customers had more insights into firms' cybersecurity measures, their responses to data breaches and underinvestments into cybersecurity may not be sufficient. The literature documents significant heterogeneity 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] further show that even for a breach of a matchmaking site, whose breached data could potentially embarrass customers and harm their personal relationships, the initial reductions in usage and increases in image deletions were relatively short-lived. In a field experiment, [Athey et al., 2017] discover that despite claiming that privacy is very important, consumers often show behaviors contradicting such statements.

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

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 transparency and to shift towards more privacy-cognizant behaviors, several governments implemented regulation on data breach. 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 supporting evidence has been provided that government intervention generated some positive impact on cybercrime prevention to both internationally and in the US [Hui et al., 2017, Murciano-Goroff, 2019, Romanosky et al., 2011], much effort is still needed given the lack of stringency and enforceability of the laws across states and the surge of the cybercrime activity in the US [Romanosky et al., 2014], leading to the inadequate investment in cybersecurity found in prior studies.[12]

To further assess the mechanism behind firms' responses to data breaches, we consider several splits of our data. 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 attentions.

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]. Finally, we directly measure public scrutiny of 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 faced sharply elevated public scrutiny are likely under significantly higher pressure to improve and invest in substantive cybersecurity adoption. We thus have the following hypothesises to assess effect heterogeneity and the public scrutiny mechanism:

**Hypothesis 7** (Heterogeneity by Industry)**.** *Firms in service-providing industries react more strongly to data breaches in both substantive and symbolic hiring than*

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

*firms in goods-producing industries.*

**Hypothesis 8** (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.*

**Hypothesis 9** (Public scrutiny (Media and Search Attention) 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.*

## 3.3 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 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 final sample.

### 3.3.1 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.[13]

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.[14]

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[15] and skills.[16]

### 3.3.2 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 ex-

[13] A great overview of the skill details can be found in [Lassébie et al., 2021].

[14] 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.

[15] Specifically, we include the following 2010 SOC occupations: Information Security Analysts (15-1122), Computer Systems Analysts (15-1121), Computer Network Support Specialists (15-1152), Database Administrators (15-1141), Network and Computer Systems Administrators (15-1142), Computer Network Architects (15-1143).

[16] 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.

posed 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.

### 3.3.3 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.[17] 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. 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).

One limitation of the BGT data for our estimation purpose is that cybersecurity

[17] Many matched firms in the PRC data are dropped due to the 100 job posting cutoff.

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.

### 3.3.4 Hiring for Cybersecurity and Legal/PR as Substantive and Symbolic Adoption

### 3.3.5 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.[18]

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 represents the public interest in a given topic very well and is particularly well-suited to identify spikes in interest [Baker and Fradkin, 2017].[19] We

[18] 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 a distinct channel.

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

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 firm names in our list, which suffered a data breach.

### 3.3.6 Summary Statistics

Table 6.1 presents the summary statistics of the key variables in our main sample. Panel A describes the full sample with both the treated (breached) and the control (never breached) firms. Our final sample includes 1,435 firms that experienced any form of data breaches. Among them, 1,261 have posted at least one job demanding cybersecurity experts (as defined by the cybersecurity occupations listed above). There are 87,628 firms in our sample never had data breach with 58,565 posted cybersecurity jobs. 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.[20] 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. For comparison, 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. Column three of Panel A presents the statistics of the key variables for the entire sample.

Panel B of table 6.2 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 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. Panel B shows, using only the firm names

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

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

## 3.4 Empirical Methods

### 3.4.1 Difference-in-Differences Estimation

The impact of data breaches on firms can only be studied through observational data - experimental settings with randomized controlled trials (RCT) would either be immoral or not realistic enough to yield generalizable conclusions. Thus, natural experiments and a difference-in-difference analysis between breached, i.e. treated, and non-treated, i.e. control, firms is the most suitable setting.

It is reasonable to assume that the timing of data breaches is 'as-if' randomly assigned and thus enables this type of natural experiment setting and leveraging Difference-in-Differences (DiD) analysis. There are no strategic benefits for firms to delay their investment responses after discovering a data breach.[21]. 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. Thus, if latent confounders coincided with the treatment, one could not disentangle the effects of the treatment and 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.

[21]However, firms may have strategic benefits to delaying the data breach *announcement* to soften the media or shareholder impact [Foerderer and Schuetz, 2022]. While the US data breach notification laws may mitigate these considerations, we still explore potential timing heterogeneity around the breach announcement dates

Formally, our baseline model is specified empirically as follows:

$$Job_{i,j,t} = \beta_0 + \beta_p D_{i,t} + \lambda_i + \lambda_y + \lambda_{m,j} + \varepsilon_{i,j,t} \tag{3.1}$$

where for each firm $i$, in two-digit NAICS industry $j$ and month $t$, $Job_{ijt}$ indicates whether the firm posts job vacancies for cybersecurity occupations or, in other specifications, for legal or PR occupations. Alternatively, it indicates whether a particular firm requests job vacancies that mention skills related to cybersecurity or legal or PR skills in the skill-level analysis. $D_{i,t}$ is the indicator variable for whether the observed month is after the data breach events.[22] Both breached, i.e. treated, and control firms are in the sample, and the control firms help to estimate the following fixed effects in the regression above:

- $\lambda_i$: firm-fixed effects (to control for firm heterogeneity)
- $\lambda_y$: year-fixed effects (to control for changes in overall posting behavior)
- $\lambda_{m,j}$: calendar month of year by two-digit NAICS industry fixed effect (to capture industry-specific seasonality)

This is estimated as a fixed effects linear model.[23] Standard errors are clustered at the firm level. The coefficient $\beta_1$ 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 $t$.

### 3.4.2 Quarterly Dynamics

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

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

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

$$Job_{i,j,t} = \beta_0 + \sum_{p=-1}^{1} \beta_p D_{i,t+p} + \lambda_i + \lambda_y + \lambda_{m,j} + \varepsilon_{i,j,t} \tag{3.2}$$

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 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. 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; 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.[24]

## 3.5 Main Results

### 3.5.1 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 6.2 reports results using both posting probabilities (columns 1, 2, 4, 5, 7, and 8) as well as counts as outcome variables (columns 3, 6, and 9) in linear regresion models.[25] In columns one and two of Table 6.2, 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

[24] These results are omitted in the manuscript due to space limitation but can be found in Appendix 9.1.

[25] We also explore and report results from Poisson models in the appendix.

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.[26] 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.[27]

Column one of Table 6.2 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 two, we decompose the time horizon to the quarterly level. The coefficient in column one 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.[28]

Column two of Table 6.2 regresses the indicator variable for posting any cybersecurity job postings 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. The coefficients show that before suffering a breach, breached firms do not hire significantly differently for cybersecurity, nor any other type, 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

[26]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 9.1.

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

[28]Additionally, in our data, we use the data breach announcement date as the breach date. While this is neither the exact date of the breach nor the date on which the firm notices, it is still the best date available. As shown in Figure 6-1, 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. In order to reduce the noise due to the ambiguity of the treatment time, in column 1 of Table A2, 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.

firms than non-breached ones. It further increases to 3.4 percentage points in the second quarter (Quarter +1) after the breaches.[29] These results suggest a lag between firms' announcements of data breach events and the actual response of recruiting cybersecurity talent. The effect is most pronounced three months after the breach events.[30]

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 employee the number of job postings of cybersecurity as the left-hand side variable and report results in column three. The results indicate that the breached firms on average post 0.4 more jobs after the breach events compared to the non-breached firms.[32] Overall, the results in Table 6.2 largely support our Hypothesis 5, suggesting that breached firms do increase the probability of taking the substantive adoption by hiring cybersecurity talents. Although statistically significant, the magnitude of the identified effect seems small at 2.1 percentage points. This result suggests that firms lack incentive to take the substantive adoption of acquiring cybersecurity talents to 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. Thus, as stated in

[29]Column two of Table A2 presents the results from a similar regression but with the quarter prior to the data breaches as the omitted period. The result is robust to this alternative specification.

[30]One potential concern is that firms might respond more strongly to larger data breach events. 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.

[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.

[32]Additionally, we also performed the Poisson regression and present the result in column three of Table A2. 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 two digit NAICS industry fixed effects rather than the month-by-industry fixed effects as the latter exceeds our computing power.

Hypothesis 6, firms should also increase symbolic adoption by acquiring legal and PR talents after a data breach. We further investigate this hypothesis in the following section.

### 3.5.2 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. 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 four to six in Table 6.2, we estimate the effect of data breaches on firms' symbolic adoption of legal and public relation talents.[33]. Following equation 3.1 and section 3.5.1, the dependent variable in columns four and five is an indicator for whether the firm posts any job vacancies for legal or public relation occupations.

The result in column four shows that breached firms are 2.1 percentage point more likely to hire legal or public relation talents, a similar magnitude as the effect on cybersecurity hiring presented in Column one. This confirms our Hypothesis 6 that firms increase symbolic adoption to deal with the symptoms of a data breach. Second, the magnitude of the coefficient shows that the increased probability of symbolic adoption of legal and PR workers is similar to the substantive adoption of cybersecurity workers. Similar to Column two, column five shows that the timing for these

[33] Similar to 3.5.1, 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 9.1

symbolic hiring attempts are similarly delayed into the second quarter after the data breach events. 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).[34] However, although results presented in Columns four and five show a positive effect on extensive margin, the intensive margin of the effect on symbolic hiring is not statistically significant from zero. [35]

### 3.5.3 Placebo Tests on Non-Relevant Jobs

Although by adopting a difference-in-difference strategy we are able to provide causal effect of data breach on firm's substantive and symbolic hiring, there still can be concerns that the breached firms are in general larger than the non-breached firms, or the increased talent demand may not be limited to cybersecurity, legal and PR. Therefore, we also performed a set of placebo 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 6.2. [36]

Similar to Section 3.5.1 and 3.5.2, 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. [37] The intensive margin presented

[34] Similar to Section 3.5.1, we also performed additional robustness checks in Appendix 9.1. Column four 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 four in Table 6.2. 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 A2 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 five of Table 6.2.

[35] We also performed the Poisson regression and present the result in column six of Table A2. The Poisson regression also shows non-significance for symbolic hiring.

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

[37] Using the events with the most breached records as the treatment, Column 3 of Table A1

in column 9 also shows data breaches have no effect on firms' talent demand on non-relevant occupations. 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 the overall increased human capital demand following the data breaches, which is reassuring.

### 3.5.4 Demand for Related Skills

An alternative way to measure firm's response to data breaches is through their change of demand for related skills instead of occupations. 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 a data breach. The results are presented in Table 6.4. Similar to Table 6.2 and related sections shown above, we also use the first breach events as the treatment.[38] Columns 1 to 3 present the effect of data breach on firm's demand of cybersecurity skills; columns 4 to 6 present the effect on firm's demand of legal and PR skills; columns 7 to 9 present the effect on the demand of skills that are not relevant to data breach or cyber attacks.

The left-hand side variable of columns 1 and 2 in Table 6.4 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 6.4. After the first breach event recorded in the Privacy Rights Clearinghouse data, firms are 1.6 percentage point more likely to demand any cybersecurity skills, as shown in

concludes with a similar result.

[38] For robustness check, results using breach events with the most lost records are presented in the Appendix and Table A3.

column one.[39] 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 shown in Section 3.5.1.[40]

Instead of a binary outcome, column 3 in Table 6.4 uses the number of job postings demanding any cybersecurity skills. The coefficient suggests that breached firms on average post 0.89 more jobs than non-breached firms after the loss of data. In terms of magnitude, this is more than twice the size of the coefficient for demanding cybersecurity occupations as presented in column seven of Table 6.2. This corroborates that breached firms hire more cybersecurity personnel after the breach, but, more importantly, that they might broaden the cybersecurity skill requirement to non-cybersecurity occupations as well.[41]

Columns 3 to 6 of Table 6.4 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 shows a similar magnitude as the effect on cybersecurity skills in column 1. The quarterly dynamics presented in column 5 also show 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 events. However, the intensive margin presented in column 6 does not show a statistically significant result, which is consistent with column 6 of Table 6.2.[42]

---

[39]Focusing instead firms' data breaches with the largest number of breached records, leads to a similar effect size of 2.0 percentage point, as shown in Table A3

[40]The robustness check of omitting the month of breach event and one month before and the robustness check on omitting the quarter immediately before the announcement of the data breaches are also performed, as shown in columns 1 and 2 of Table A4. And the results are similar.

[41]However,the Poisson regression in column 3 of Table A4 does not show any statistical significance, potentially due to larger measurement errors when measuring demand in absolute levels using the BGT data. Another potential reason is, instead of including the calendar month by industry fixed effects, we include the month fixed effects and the industry fixed effects due to computational complexity.

[42]Robustness checks presented in column 2 of Table A3 and columns 4 to 6 of Table A4 show that the results are robust across different selection of treatments, different treatment time, or different

Similar to Section 3.5.3, we also performed a placebo test on firm's demand of skills that are not relevant to data breaches. We consider a job that requires skills not relevant to data breaches if it does not acquire any skill either in cybersecurity skill cluster, or in PR or legal skills, according to the BGT taxonomy. Columns 7 and 8 of Table 6.2 show there is no effect on the demand of non-relevant skills at the extensive margin. However, columns 9 shows breached firms seem also to be associated with the increase of the number of the job postings that acquire skills not relevant to solving the cybersecurity challenge or resolving the PR or legal crisis.

## 3.6 Additional Results and Effect Heterogeneity

### 3.6.1 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.[43] We utilize the richness of our data to identify demand heterogeneity for specific types of cybersecurity occupations. Table 6.6 presents the regression coefficients from model 3.1 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 one, two, four and five, respectively. In contrast, no significant effect is found for network support specialist and computer network architects as shown in columns three and six.

omitted period.

[43]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/

### 3.6.2 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 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 one of Table 6.5. 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 one in Table 6.2.

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 two of Table 6.5 suggests that firms do not take actions to increase their probability of hiring cybersecurity talents after the non-cyber events. This is very plausible and can serve as a placebo test, since these kinds of events are not associated with cybercrimes and therefore do not provide firms with additional incentives to strengthen their cybersecurity teams.

We also look at the effect on the PR and legal jobs after these two types of data breach events as presented in columns three and four in Table 6.5. The coefficient in column three suggests 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 four. 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 when facing pressure from the public, though the magnitude is relatively lower and the coefficient is only statistically significant at five percent level. Similar to cybersecurity talents, data losses through PORT and PHYS has no effect on firm's hiring of PR and legal talents.

### 3.6.3 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[44] and service-providing industries.[45] The results of the coefficients are presented in Panel (a) of Figure 6-2. 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 column 1 and 4 from Table 6.2. The second row shows that firms in goods-producing industries 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. In the bottom row 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. 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 6-2 suggest that the reaction to data breaches is almost exclusively driven by firms in the service-providing industries in our sample while firms in the goods-producing industries do not appear to react. A potential explanation is that the nature of the data collected may be significantly

[44]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).

[45]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.

different across industries. Data collected by firms in the service-providing industries typically consists of highly sensitive and private customer information, such as customer emails and credit card numbers. Exposure of such data can put breached firms' clients under significant risk, increase the chance of large class action lawsuits, and 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 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.

### 3.6.4 Effect Heterogeneity by Ownership Type

Consumers may find it difficult to evaluate a firm's investment in cybersecurity, which may lead to underinvestment by the firm. This is consistent with findings in the prior literature on cybersecurity capital investments as well as our main results on cybersecurity hiring. Although a data breach is associated with a statistically significant effect on cybersecurity hiring, the magnitude of the effect is relatively small. To understand to what extent firms have more incentives 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 Compustat and BGT data through a fuzzy name matching algorithm, we identified 3390 public firms in our data.[46] While non-accurately estimated due to the sample size for public firms, Panel (b) of Figure 6-2 shows the probability of substantive adoption (i.e., posting cybersecurity jobs) increased following a data breach event for both public (though imprecisely measured), and private

[46] 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.

firms. 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 6-2 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 8.

### 3.6.5 Discussion of Results

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, this implies that firms would instead invest heavily in software and infrastructure investment, which is inconsistent with the findings of prior studies.

Additionally, our skill-level tests (instead of occupations) further address the concern with better granularity, better 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.[47][48]

[47]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.

[48]Prior research has also found some positive spillover effects on market valuation, performance, and cash holdings [Garg, 2020, Aytes et al., 2006] of breached firms' competitors - often through awareness in the form of 'That could have been us' moments [Samuelson, 2007]. This would also bias the estimates on the focal firms. Consistent with prior literature, as presented in Appendix 9.3, we find a positive and significant but small spillover effect of competitors' data breach events on other firms' cybersecurity hiring within a given industry. The timing of this effect helps explain the smaller and noisier effects on the focal firms in the first few months after the breach (e.g., the spillover effect appears in the first two months of the data breach event and hence bias the estimates on the focal firms downwards during the same period). However, the magnitude of the estimated coefficient on the focal firms is still economically small when accounting for this spill over effect, consistent with our hypothesis that firms lack incentives to allocate sufficient human capital resources to safeguard data.

Combined with insufficient financial investments in cybersecurity found in the earlier studies [Gordon et al., 2015a, Gordon et al., 2015b, Manworren et al., 2016], the relatively small effect that we identify on firms' human capital investment in cybersecurity may reflect market failure due to misaligned incentives and information asymmetries between consumers and firms [Moore, 2010]. We further dive into the mechanism by exploring the heterogeneous effects of data breaches by sectors. We report results in Figure B1 and show that the identified effect is isolated to the service-providing sectors (e.g. wholesale and retail trade) where most breached data consists of sensitive customer information (e.g. credit card information) and is more likely to draw significantly larger public and media attention, in contrast to operational data in goods-producing industries (e.g. manufacturing). We then move on to explore the heterogeneous effects of data breaches by firm type (e.g. public vs. private) and provide consistent evidence that the impact of data breaches on firms' cybersecurity hiring is much larger for public firms, which face more stringent regulation [Angst et al., 2017] and higher public scrutiny, than for private firms.

## 3.7 Public Visibility as a Mechanism

The heterogeneity suggests that the scrutiny placed on the firm may affect the firms' decision to hire substantively. To 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.

### 3.7.1 Measuring the Change of Visibility Related to Data Breaches

We are interested in assessing the effect of media and public attention on hiring behavior. Because firms may experience substantially different media and public attention unrelated to cybersecurity, we must control for pre-period visibility. In contrast to the comparison between public and private firms, these two 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).[49] 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 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 3.1, which limits the treatment group to only high visibility firms, and Equation 3.1, 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.

### 3.7.2 Effect Heterogeneity by Visibility

Figure 6-3 plots the change in the probability of substantive and symbolic postings 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 6-3. The effect of high visibility breaches on symbolic postings is smaller and insignificant, although the confidence intervals do not allow us to reject the null hypothesis that these two effects are equal.

While a low visibility breach leads to an increased probability of cybersecurity hiring, this effect is larger for high visibility breaches. The magnitude of this difference is economically significant: while low media attention breaches are associated with

[49] 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.

a 3.1 percentage point increase in the probability of a cybersecurity (substantive) posting, high media attention breaches are associated with a 7.0 percentage point increase. This is more than double the effect size.

This effect is smaller for symbolic postings. Though high visibility breaches are more likely to hire in symbolic roles, this effect is not statistically different from zero. In terms of magnitude, high visibility breaches are 4.0 percentage points more likely to post a symbolic job, while low visibility breaches are 1.8 percentage points more likely to post a symbolic job. This is suggestive evidence that media attention invites increased hiring of cybersecurity jobs, and less so of legal/PR jobs.

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 6-4. While it may be assumed that media shares and search shares are positively correlated, this correlation is not very high, supporting visibility coming from two different channels.

Figure 6-4, measuring visibility by Google searches, demonstrates that a similar visibility pattern for substantive postings exists for public attention. As with the media attention figure, the confidence intervals are large.

In this figure, we see that firms hire for substantive positions under both high and low visibility. However, we only see significant (non-zero) symbolic hiring under low visibility - under high visibility hiring for symbolic is not significant. This implies that visibility leads to firms hiring relatively more substantive than symbolic.

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 6-5, the relationship between these two outcomes is assessed by looking at 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.

## 3.8 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 unclear. This paper brings together job vacancy and skill demand data from Burning Glass Technologies 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 difference-in-differences 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 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 5 and Hypothesis 6 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 market failure due to information asymmetry is one of the potential channels that explains the observed under-investment behavior of breached firms. 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. Public attention such as Media coverage and public searches 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 results suggest that consumers, media, as well as government should work together to provide better incentives for firms to safeguard data, protect customers, and increase the social welfare in our increasingly digitized world.

# Chapter 4

# Tables & Figures for Chapter 1

### Table 4.1: Panel Wage Regression - BGT Skills

| | *Dependent variable:* | | |
|---|---|---|---|
| | Annual Wage | | |
| | (1) | (2) | (3) |
| Economics Policy | −586.46 (379.98) | 129.18** (55.69) | 437.15*** (49.96) |
| Analysis | 1,145.07*** (352.77) | 329.34*** (31.89) | 182.87*** (28.60) |
| Design | −631.36* (347.01) | −85.12*** (21.98) | 136.06*** (19.92) |
| Marketing, PR | −120.86 (340.70) | −56.53*** (13.58) | 66.50*** (12.29) |
| Manufacturing | 89.30 (341.08) | 42.02*** (11.46) | 62.99*** (10.27) |
| Engineering | −83.58 (346.34) | −10.66 (19.57) | 37.61** (17.54) |
| Business | 181.00 (340.48) | 56.20*** (11.01) | 31.11*** (9.88) |
| Health Care | 183.93 (339.47) | −44.04*** (9.55) | 19.88** (8.58) |
| Public Safety, Security | −25.33 (343.28) | −86.96*** (19.65) | 14.06 (17.65) |
| Architecture, Construction | 104.79 (340.95) | −67.77*** (12.63) | 8.57 (11.33) |
| HR | 462.34 (340.93) | 73.10*** (11.93) | −2.18 (10.72) |
| Environment | −9.29 (344.34) | −139.32*** (21.09) | −2.28 (18.92) |
| IT | 135.16 (340.98) | −72.05*** (11.43) | −4.02 (10.26) |
| Agriculture | −200.10 (350.44) | −44.59** (17.63) | −6.54 (15.80) |
| Energy, Utilities | 10.42 (345.05) | −57.59** (25.65) | −6.84 (22.98) |
| Media, Writing | −174.72 (349.68) | −169.68*** (23.19) | −7.24 (20.85) |
| Finance | −211.76 (341.37) | −90.02*** (11.59) | −8.07 (10.41) |
| Customer/Client Support | 218.25 (340.01) | −43.08*** (8.68) | −12.75 (7.79) |
| Supply Chain Logistics | 311.72 (340.32) | −44.34*** (8.87) | −30.62*** (7.94) |
| Sales | −6.32 (340.65) | −97.50*** (11.00) | −36.49*** (9.87) |
| Administration | 342.65 (339.95) | 3.02 (9.47) | −39.95*** (8.49) |
| Maintenance, Repair | 116.96 (340.79) | −73.31*** (9.42) | −46.06*** (8.44) |
| Personal Care | 146.16 (340.16) | −98.06*** (9.22) | −56.67*** (8.28) |
| Industry Knowledge | 90.68 (341.06) | −114.13*** (12.98) | −81.75*** (11.65) |
| Education, Training | 286.93 (343.29) | −186.47*** (14.07) | −96.82*** (12.64) |
| Science, Research | −373.19 (352.09) | −169.87*** (23.56) | −120.87*** (21.11) |
| Legal | −43.74 (344.72) | −222.45*** (20.71) | −199.83*** (18.55) |
| Occupation-Fixed Effects | *No* | *Yes* | *Yes* |
| Industry-Fixed Effects | *No* | *Yes* | *Yes* |
| Time-Fixed Effects | *No* | *No* | *Yes* |
| Observations | 6,658 | 58,678 | 58,678 |
| Adjusted $R^2$ | 0.99 | 0.97 | 0.98 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 4.2: Top and Bottom 5 Occupation Fixed Effects from BGT Skills Wage Panel Model 3.**

| SOC Title | SOC6 Code | FE Value |
|---|---|---|
| Anesthesiologists | 291061 | 249400.56 |
| Surgeons | 291067 | 243323.18 |
| Oral and maxillofacial surgeons | 291022 | 228877.01 |
| Obstetricians and gynecologists | 291064 | 224534.88 |
| Orthodontists | 291023 | 212590.82 |
| Tapers | 472082 | 22616.72 |
| Ushers, lobby attendants, and ticket takers | 393031 | 22206.52 |
| Food Processing Workers, All Other | 513099 | 21755.97 |
| Veterinary assistants and laboratory animal caretakers | 319096 | 19933.28 |
| Motion picture projectionists | 393021 | 19456.70 |

**Table 4.3: Industry Fixed Effects from BGT Skills Wage Panel Model 3.**

| Industry Title | NAICS2 Code | FE Value |
|---|---|---|
| Utilities | 22 | 13483.27 |
| Mining | 21 | 10903.34 |
| Management of Companies and Enterprises | 55 | 9880.43 |
| Professional, Scientific, and Technical Services | 54 | 8935.77 |
| Information | 51 | 8508.38 |
| Finance and Insurance | 52 | 7714.41 |
| Transportation and Warehousing | 48-49 | 4391.52 |
| Manufacturing | 31-33 | 4250.13 |
| Wholesale Trade | 42 | 4181.24 |
| Construction | 23 | 3294.98 |
| Educational Services | 61 | 0.00 |
| Real Estate and Rental and Leasing | 53 | -318.61 |
| Health Care and Social Assistance | 62 | -832.50 |
| Admin, Support, Waste Management, Remediation Services | 56 | -919.16 |
| Agriculture, Forestry, Fishing, and Hunting | 11 | -1016.43 |
| Other Services (except Public Administration) | 81 | -2015.06 |
| Arts, Entertainment, and Recreation | 71 | -2549.52 |
| Retail Trade | 44-45 | -3711.92 |
| Accomodation and Food Service | 72 | -4325.95 |

Table 4.4: Panel Wage Regression - Deming Skills

| | *Dependent variable:* | | |
|---|---|---|---|
| | Annual Wage | | |
| | (1) | (2) | (3) |
| ML, AI | 4,048.60*** (258.86) | 4,187.47*** (123.14) | 2,760.06*** (111.45) |
| Business Systems | −2,017.18*** (182.91) | 1,101.32*** (50.38) | 1,295.10*** (45.31) |
| General Software | 2,957.46*** (301.68) | 1,049.75*** (90.69) | 923.64*** (81.59) |
| Creativity | −526.78*** (115.18) | 25.85 (40.48) | 214.09*** (36.50) |
| Data Analysis | 1,121.32*** (303.66) | 622.56*** (88.50) | 104.70 (79.67) |
| Admin, Support | 55.19 (48.07) | 164.37*** (14.51) | 97.81*** (13.06) |
| Project Management | 21.09 (72.82) | 100.66*** (23.61) | 79.81*** (21.23) |
| Social | 88.11*** (25.19) | 114.33*** (10.02) | 69.39*** (9.05) |
| Customer Service | −0.92 (24.80) | 11.48 (7.74) | 36.03*** (6.96) |
| Engineering | −1,000.81*** (117.25) | −214.32*** (36.12) | 22.31 (32.53) |
| Cognitive | 309.49*** (50.36) | 39.49** (15.72) | −18.50 (14.16) |
| Finance | −990.31*** (72.72) | −108.17*** (17.36) | −20.06 (15.63) |
| Database | −355.12*** (126.67) | −157.31*** (40.17) | −79.58** (36.11) |
| Tech Support | −1,016.07*** (104.87) | −436.37*** (35.78) | −89.77*** (32.31) |
| Computer | −291.76*** (45.65) | −205.69*** (14.22) | −90.39*** (12.82) |
| Product Marketing | −43.47 (40.19) | −117.47*** (15.14) | −114.12*** (13.61) |
| Management | −378.05*** (70.64) | −214.38*** (26.33) | −120.48*** (23.68) |
| Non Cognitive | 525.45*** (35.44) | 59.06*** (12.03) | −135.84*** (10.95) |
| Writing | −477.47*** (83.78) | −329.63*** (27.82) | −188.97*** (25.03) |
| Occupation-Fixed Effects | *No* | *Yes* | *Yes* |
| Industry-Fixed Effects | *No* | *Yes* | *Yes* |
| Time-Fixed Effects | *No* | *No* | *Yes* |
| Observations | 6,662 | 58,873 | 58,873 |
| Adjusted $R^2$ | 0.99 | 0.97 | 0.98 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

**Table 4.5: Top and Bottom 5 Occupation Fixed Effects from Deming Skills Wage Panel Model 3.**

| SOC Title | SOC6 Code | FE Value |
|---|---|---|
| Anesthesiologists | 291061 | 249296.67 |
| Surgeons | 291067 | 242321.03 |
| Oral and maxillofacial surgeons | 291022 | 229119.17 |
| Obstetricians and gynecologists | 291064 | 223188.11 |
| Orthodontists | 291023 | 212841.38 |
| Sewing machine operators | 516031 | 21772.52 |
| Veterinary assistants and laboratory animal caretakers | 319096 | 20673.60 |
| Ushers, lobby attendants, and ticket takers | 393031 | 20399.14 |
| Motion picture projectionists | 393021 | 18711.47 |
| Food Processing Workers, All Other | 513099 | 17230.07 |

**Table 4.6:** Industry Fixed Effects from Deming Skills Wage Panel Model 3.

| Industry Title | NAICS2 Code | FE Value |
|---|---|---|
| Utilities | 22 | 13431.18 |
| Mining | 21 | 11152.31 |
| Management of Companies and Enterprises | 55 | 10096.24 |
| Professional, Scientific, and Technical Services | 54 | 8017.92 |
| Information | 51 | 7357.04 |
| Finance and Insurance | 52 | 7200.85 |
| Manufacturing | 31-33 | 5315.15 |
| Transportation and Warehousing | 48-49 | 5044.71 |
| Wholesale Trade | 42 | 4121.35 |
| Construction | 23 | 3364.98 |
| Educational Services | 61 | 0.00 |
| Real Estate and Rental and Leasing | 53 | -99.76 |
| Health Care and Social Assistance | 62 | -954.71 |
| Agriculture, Forestry, Fishing, and Hunting | 11 | -990.07 |
| Admin, Support, Waste Management, Remediation Services | 56 | -1055.74 |
| Other Services (except Public Administration) | 81 | -1727.06 |
| Arts, Entertainment, and Recreation | 71 | -2352.31 |
| Retail Trade | 44-45 | -2873.29 |
| Accomodation and Food Service | 72 | -4629.06 |

## Table 4.7: Occupational Change

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Occupational (2010-2018) Skill Change | | | |
| | (1) | (2) | (3) | (4) |
| SML Score | 0.08*** (0.02) | 0.10*** (0.03) | 0.06*** (0.02) | 0.08*** (0.03) |
| Log Wage | 0.32 (0.23) | 0.13 (0.30) | | |
| Log Wage$^2$ | −0.01 (0.01) | −0.01 (0.01) | | |
| Medium Wage Tercile | | | 0.02** (0.01) | 0.01 (0.01) |
| High Wage Tercile | | | 0.01(0.01) | −0.01(0.01) |
| Wage Bill | −0.00*** (0.00) | −0.00*** (0.00) | −0.00*** (0.00) | −0.00*** (0.00) |
| Constant | −1.96 (1.27) | 0.21 (1.75) | −0.10* (0.05) | 0.96 (0.64) |
| Skill Fixed Effects | *No* | *Yes* | *No* | *Yes* |
| Observations | 650 | 618 | 650 | 618 |
| Adjusted R$^2$ | 0.10 | 0.16 | 0.10 | 0.16 |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

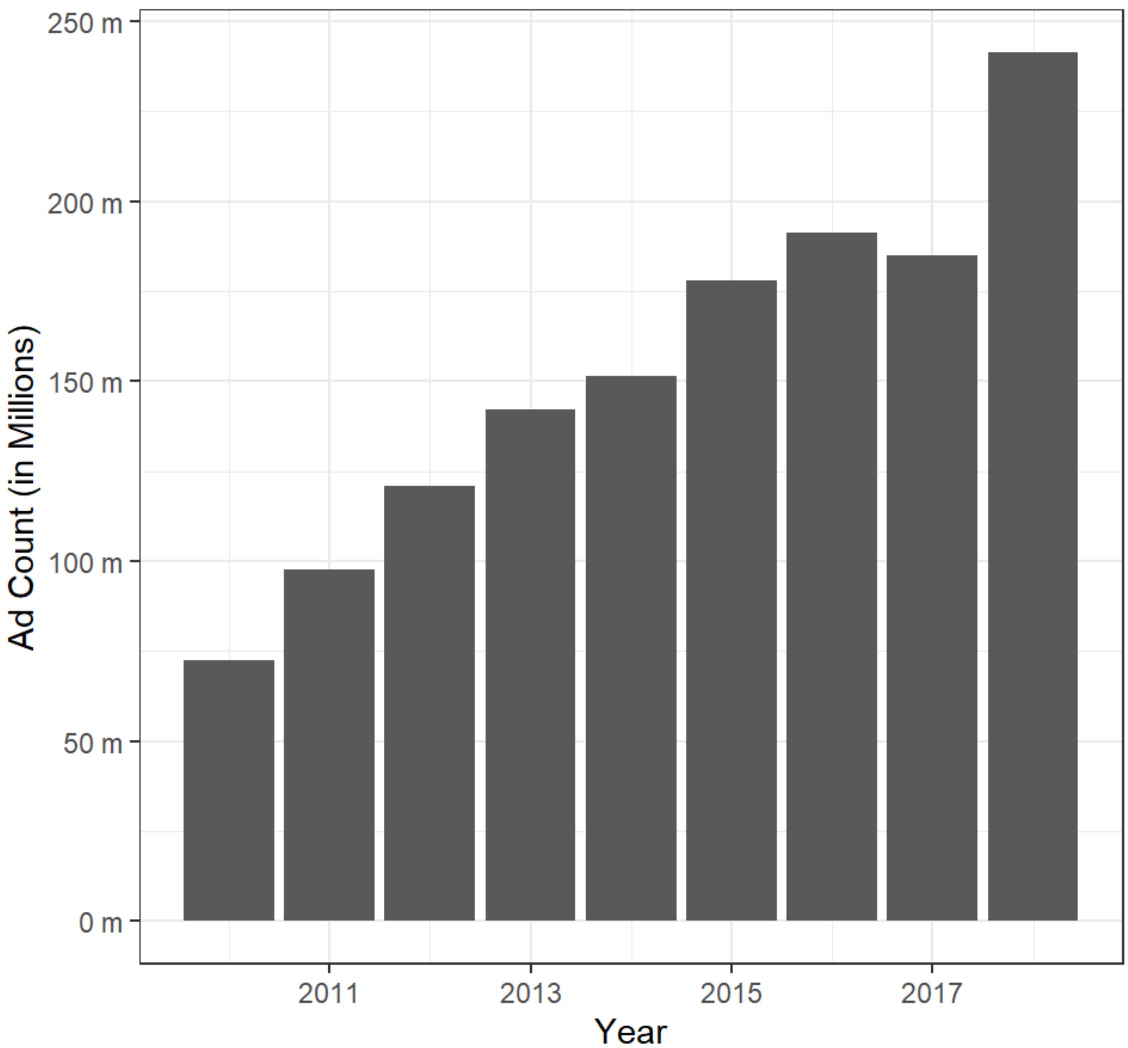



**Figure 4-1: Yearly Number of Job Postings scraped by BGT (2010-2018).**

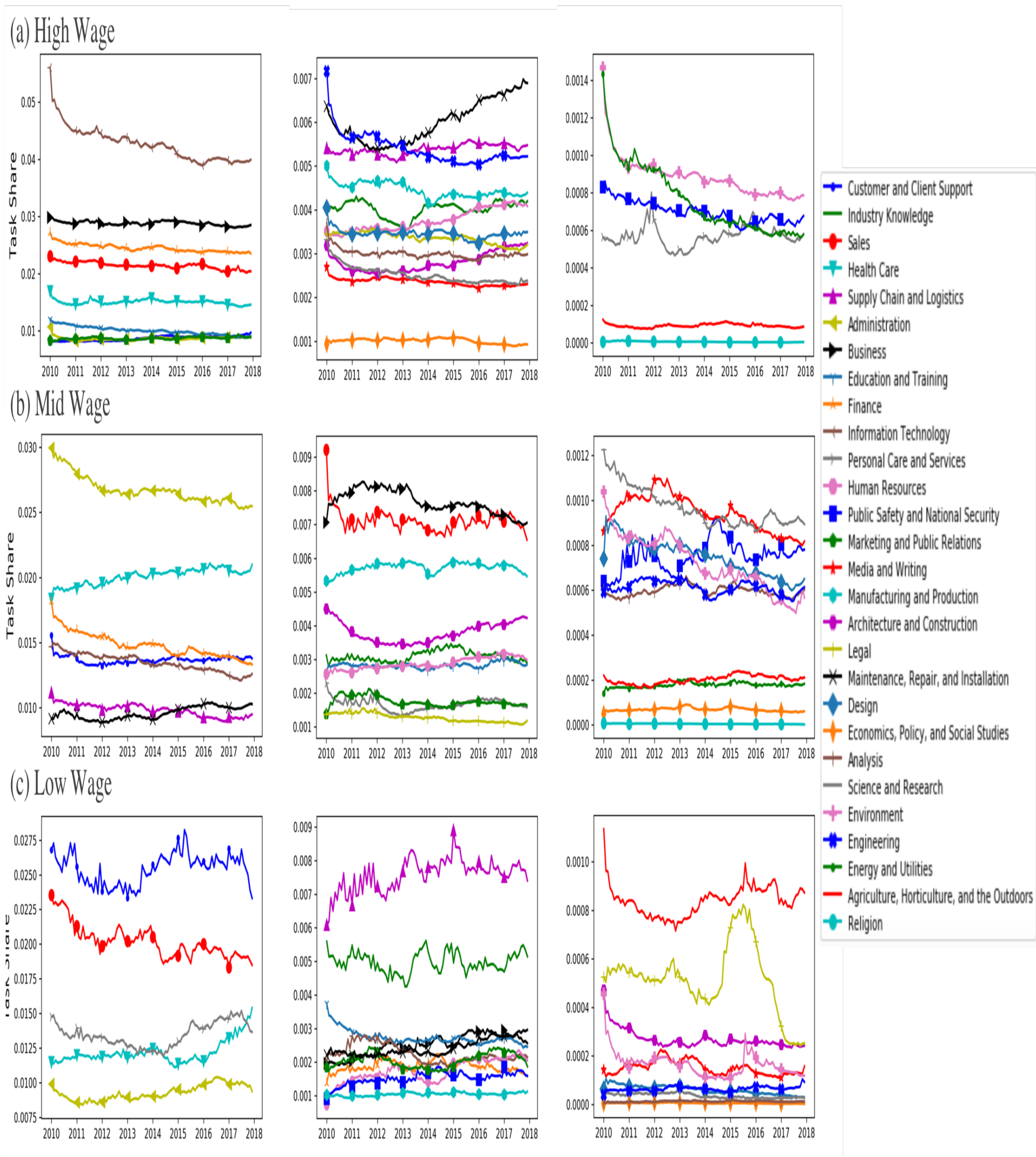



**Figure 4-2: Logged Overall Skill Shares by 2010 Wage Tercile, using the 28 BGT Skill Cluster Families.**

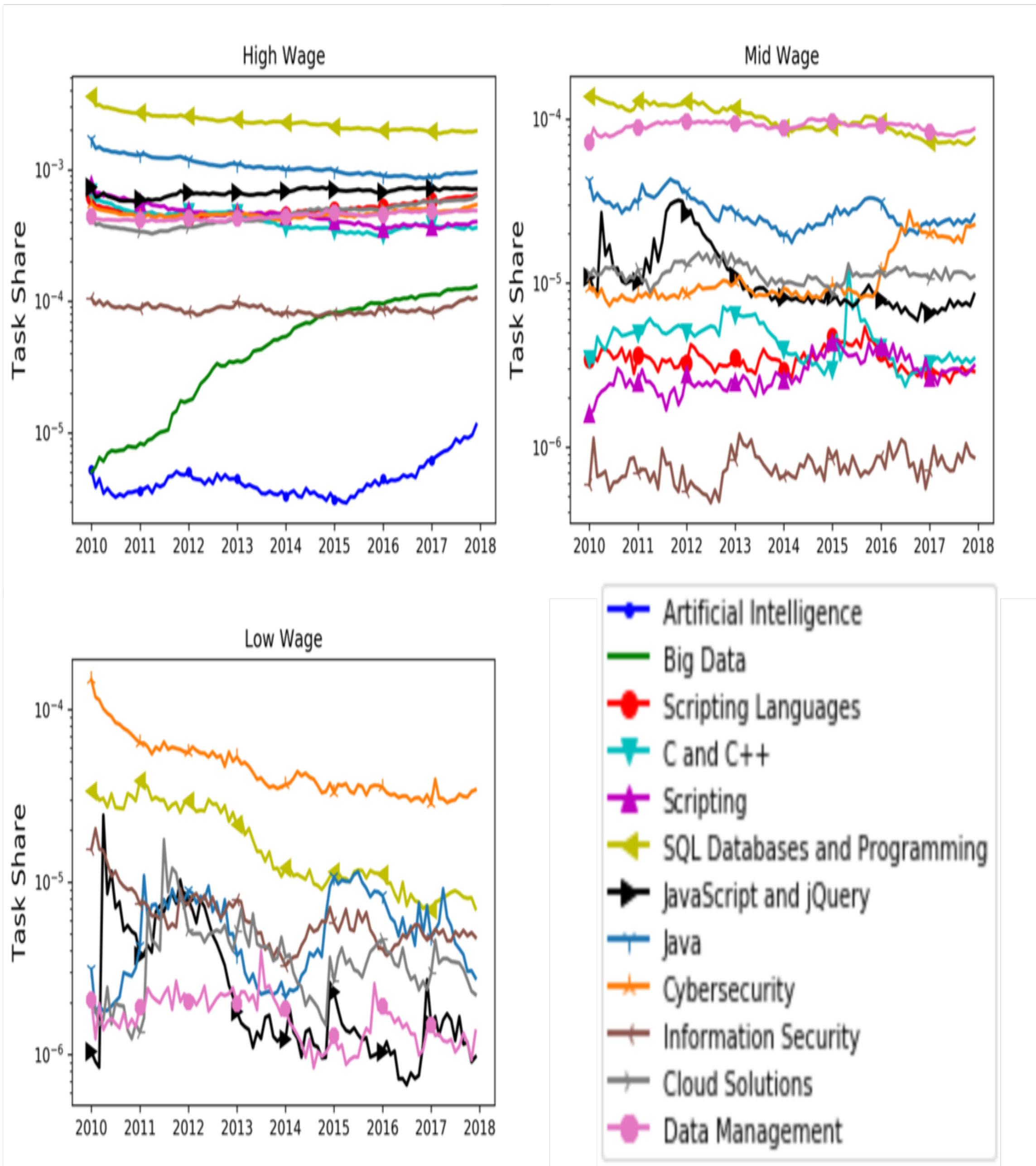



**Figure 4-3: Logged Overall Skill Shares for Skill Clusters within the IT Skill Cluster Family by 2010 Wage Tercile.**

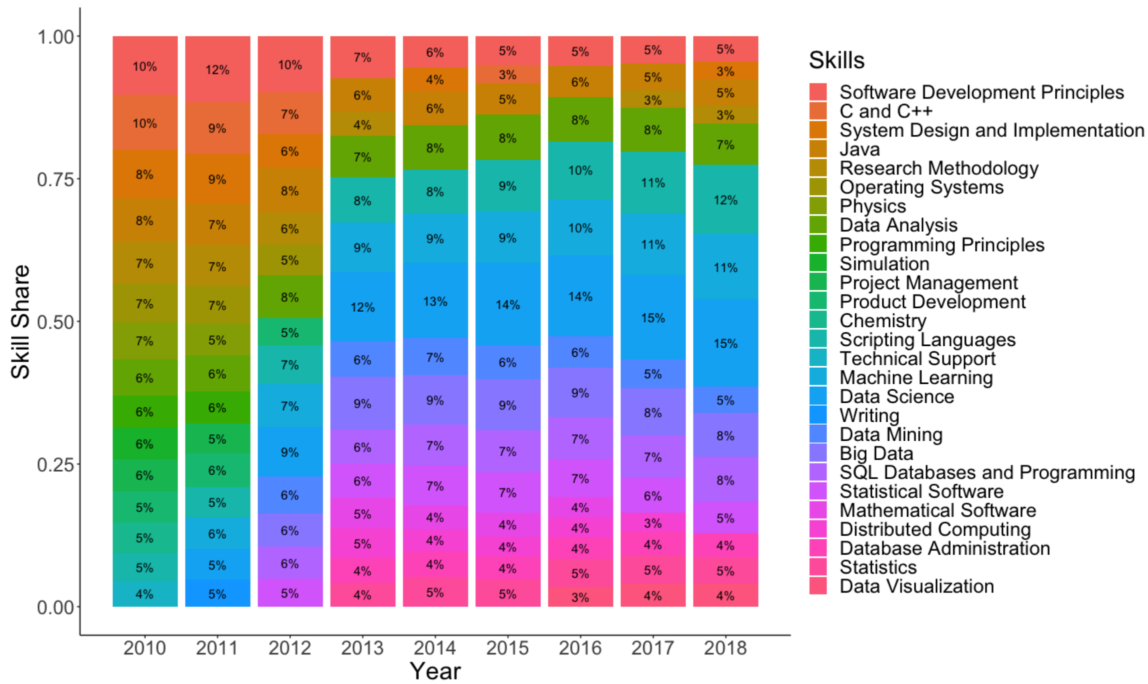



**Figure 4-4: The top 30 skill demand shares of Data Scientists (2010-2018).**

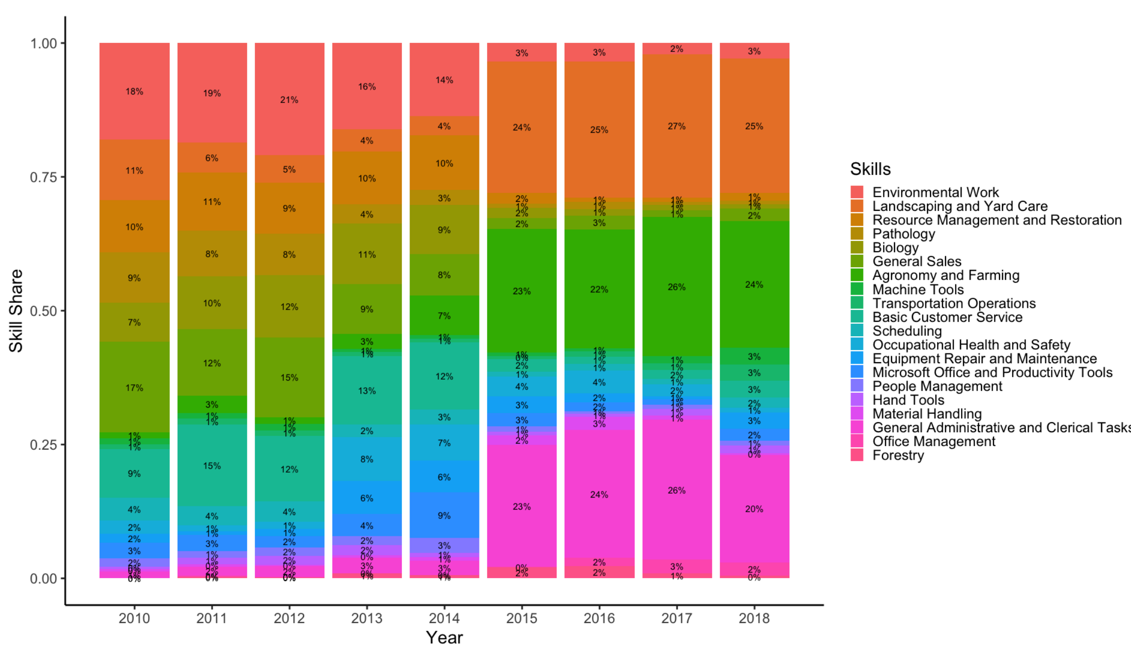



**Figure 4-5: The top 30 skill demand shares of Lumberjacks (2010-2018).**

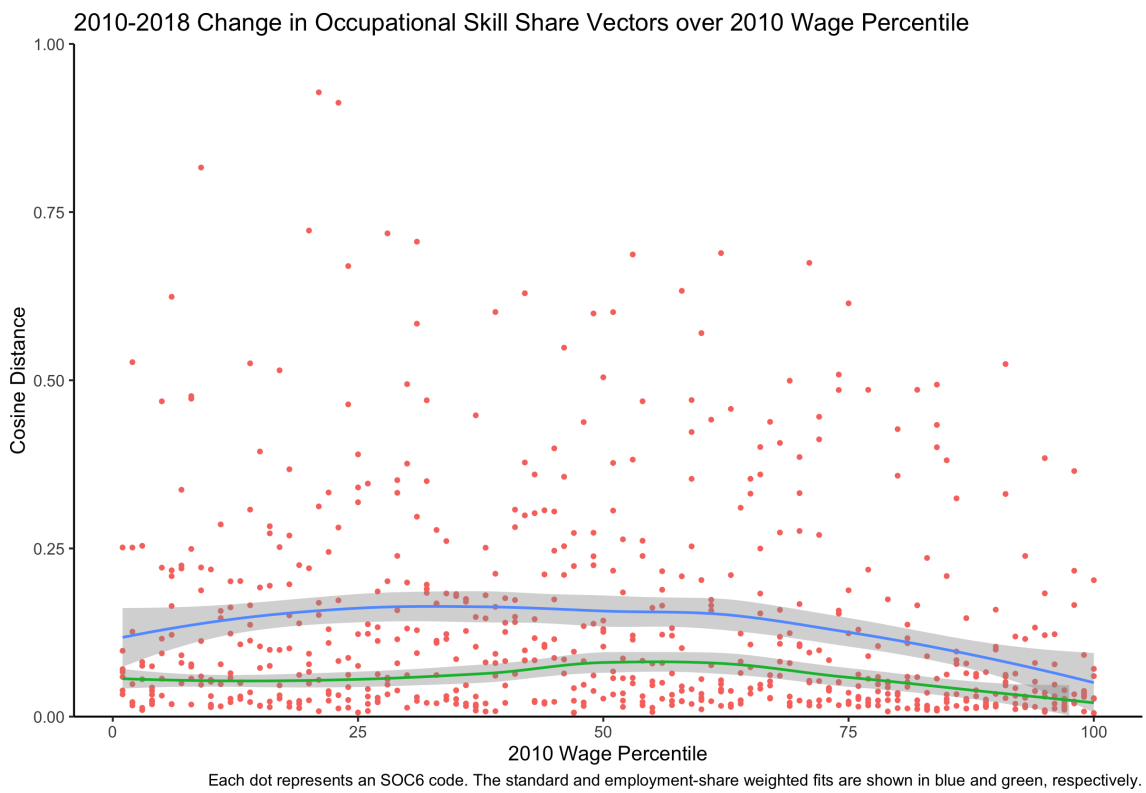



**Figure 4-6: Occupational Cosine Similarity between 2010 and 2018 skill share vectors over 2010 wage percentile.**

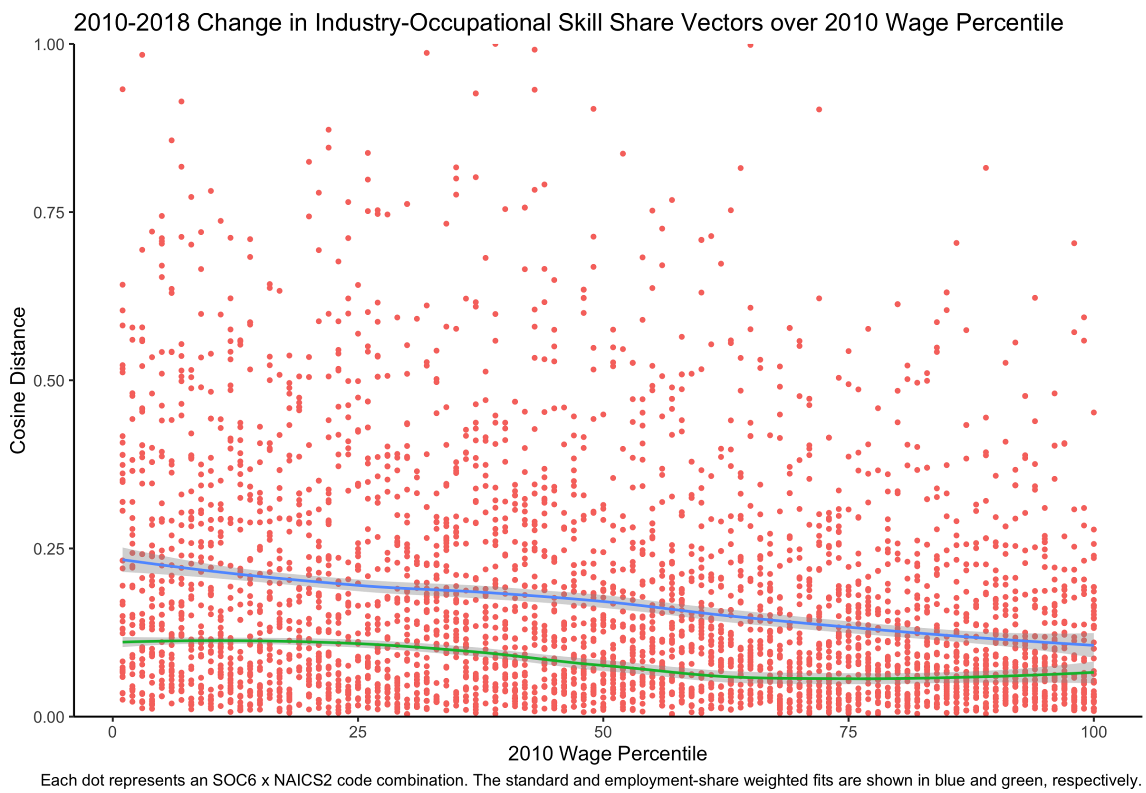



**Figure 4-7: Occupational Cosine Similarity between 2010 and 2018 skill share vectors over 2010 wage percentile.**

# Chapter 5

# Tables & Figures for Chapter 2

## Table 5.1: Variable Definitions and Summary Statistics

### (A) Full Sample

| Variables | Definition | Mean (SD) |
|---|---|---|
| WFH index | Average pre-Covid-19 WFH feasibility index (based on BGT job postings). HighWFH = 1 if firm's 2010-2019 WFH index is in top quartile of sample distribution (zero otherwise). | 0.575 (0.286) |
| Net Income | Firms' net income (Compustat) | 220.5 (912.0) |
| Sales | Firms' total sales (Compustat) | 2114 (7062) |
| Normalized Cap. Exp. | Firms' capital expenditure divided by total assets (Compustat) | 0.018 (0.036) |
| Normalized Software Exp. | Firms' capital expenditure on software divided by total assets (Compustat) | 0.002 (0.013) |
| Quarterly IT Hiring | Quarterly total number of job postings requested IT-related skills (BGT) | 56.83 (368.8) |
| Return | Sum of monthly returns per quarter (CRSP) | 0.045 (0.238) |
| Abn. Return | Sum of monthly abnormal returns per quarter. Monthly abnormal return is the difference between excess return and the CAPM beta times market excess return. CAPM beta is estimated using past 36 monthly returns (CRSP) | -0.018 (0.181) |
| Return Volatility | Standard deviation of daily returns per quarter multiplied by 2 [i.e. $\sqrt{4}$ quarters] to convert to annual basis (CRSP) | 0.057 (0.045) |
| Idio. Volatility | Standard deviation of the residuals obtained from fitting daily CAPM for every month for each firm and converted to annual basis (CRSP) | 0.044 (0.030) |
| Total Asset | Logarithm of total assets (Compustat) | 29,230 (3,285) |

### (B) Key variables in subsamples by pre-Covid-19 WFH Index and Time

| Pre-Pandemic WFH Feasibility | High (HighWFH = 1) | | Low (HighWFH = 0) | |
|---|---|---|---|---|
| Variables | Before Covid-19 | After Covid-19 | Before Covid-19 | After Covid-19 |
| WFH index | 0.887 | 0.874 | 0.503 | 0.504 |
| | (0.114) | (0.143) | (0.268) | (0.259) |
| Net Income | 281.7 | 316.8 | 199.3 | 212.5 |
| | (943.2) | (972.9) | (847.0) | (1001) |
| Sales | 1,913 | 2,305 | 2,083 | 2,217 |
| | (5,958) | (7,465) | (6,970) | (7,571) |
| Normalized Cap. Exp. | 0.014 | 0.009 | 0.021 | 0.013 |
| | (0.024) | (0.014) | (0.046) | (0.019) |
| Normalized Software Exp. | 0.006 | 0.005 | 0.001 | 0.001 |
| | (0.025) | (0.020) | (0.008) | (0.009) |
| Quarterly IT Hiring | 94.35 | 101.6 | 42.84 | 57.34 |
| | (502.5) | (438.9) | (322.2) | (365.9) |
| Return | 0.070 | 0.043 | 0.069 | -0.015 |
| | (0.189) | (0.331) | (0.166) | (0.326) |
| Abn. Return | -0.0004 | -0.001 | -0.005 | -0.054 |
| | (0.173) | (0.231) | (0.152) | (0.215) |
| Return Volatility | 0.045 | 0.082 | 0.043 | 0.084 |
| | (0.037) | (0.044) | (0.037) | (0.047) |
| Idio. Volatility | 0.037 | 0.058 | 0.036 | 0.060 |
| | (0.027) | (0.033) | (0.026) | (0.030) |
| Total Asset | 42,327 | 58,784 | 22,953 | 28,810 |
| | (235,800) | (294,900) | (136,000) | (175,700) |

*Notes:* Panel A reports variable definitions, data sources, and sample means and standard deviations for the key variables. The sample period contains 2019Q1-2020Q3. Panel B contrasts the means of key variables in pre- and post- Covid-19 subsamples (i.e., 2019 Q1-Q4 and 2020 Q1-Q3, respectively) and in the high and low pre-Covid-19 WFH subsamples. All continuous variables are winsorized at the 1st and 99th percentiles.

## Table 5.2: WFH Feasibility and Financial Performance

| Models | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. | Log IT Hiring |
| HighWFH × COVID | 0.155*** | 0.038*** | 0.004** | -0.001** | -0.103* |
| | (0.047) | (0.011) | (0.001) | (0.000) | (0.058) |
| Size | 0.310** | 0.532*** | -0.026 | -0.003** | 0.218* |
| | (0.121) | (0.034) | (0.022) | (0.001) | (0.080) |
| Cash | -0.659** | -0.379*** | -0.001 | -0.002 | -0.072 |
| | (0.296) | (0.077) | (0.022) | (0.001) | (0.185) |
| Leverage | 0.590** | -0.312*** | 0.042 | 0.001 | -0.061 |
| | (0.258) | (0.066) | (0.028) | (0.002) | (0.167) |
| R&D | 0.449* | 0.059 | -0.011 | -0.001 | 0.084 |
| | (0.255) | (0.041) | (0.010) | (0.001) | (0.132) |
| Dividend | -6.209** | -2.318** | 0.111 | 0.010 | -5.44** |
| | (3.11) | (0.958) | (0.086) | (0.009) | (2.57) |
| Tobin's q | 0.184*** | 0.040*** | 0.003** | -0.001 | 0.041* |
| | (0.031) | (0.008) | (0.001) | (0.001) | (0.024) |
| ROE | 0.038 | 0.013 | 0.007* | -0.000 | 0.095* |
| | (0.084) | (0.021) | (0.004) | (0.000) | (0.057) |
| Firm FE | Y | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y | Y |
| Observations | 9568 | 9568 | 9568 | 9568 | 9568 |
| Adj. $R^2$ | 0.886 | 0.996 | 0.521 | 0.844 | 0.881 |

*Notes:* This table implements a difference-in-differences (DID) research design to examine the differential impact of Covid-19 on various performance metrics between firms with high and low WFH feasibility. The dependent variables across columns 1 to 4 are the logarithm of net income, logarithm of sales, total capital expenditure over total asset, and software expenditure over total assets, and software expenditure over total assets, respectively. The regression model is specified in Equation 2.1, and all variables are defined in Table 5.1 Panel A. Omitting Tobin's q and ROE in the specification provides similar results (available upon request). All estimations include firm and time fixed effects. Standard errors are clustered at the firm-level.

## Table 5.3: WFH Feasibility and Stock Market Reactions

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.051*** | 0.043*** | -0.006*** | -0.004*** |
| | (0.013) | (0.012) | (0.002) | (0.001) |
| Size | -0.120*** | -0.074** | -0.003 | -0.005** |
| | (0.031) | (0.030) | (0.004) | (0.003) |
| Cash | 0.180** | 0.112 | -0.005 | -0.002 |
| | (0.079) | (0.071) | (0.010) | (0.006) |
| Leverage | 0.105 | 0.031 | 0.022*** | 0.018*** |
| | (0.069) | (0.067) | (0.007) | (0.005) |
| R&D | 0.013 | -0.058 | 0.058*** | 0.014 |
| | (0.171) | (0.094) | (0.014) | (0.011) |
| Dividend | 0.048 | -0.948 | 0.341*** | 0.351*** |
| | (0.977) | (0.908) | (0.128) | (0.083) |
| Tobin's q | -0.110*** | -0.091*** | 0.002* | 0.001 |
| | (0.011) | (0.010) | (0.001) | (0.001) |
| ROE | 0.037 | 0.033 | -0.004 | -0.005** |
| | (0.029) | (0.026) | (0.004) | (0.002) |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9550 | 9550 | 9550 | 9507 |
| Adj. $R^2$ | 0.449 | 0.147 | 0.713 | 0.750 |

*Notes:* This table implements a difference-in-differences (DID) research design to examine the differential impact of Covid-19 on stock market reactions between firms with high and low WFH feasibility. The dependent variables are total return and abnormal return in Columns (1) and (2); and return volatility and idiosyncratic volatility in Columns (3) and (4), respectively. The regression specification is specified in Equation 2.1, and all variables are defined in Table 5.1 Panel A. Firm and time fixed effects are also included in the estimation. Standard errors are clustered at the firm level.

## Table 5.4: WFH Feasibility and Financial Performance - Controls for Industry-level Demand Shocks

| Models | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. | Log IT Hiring |
| HighWFH × COVID | 0.141*** | 0.029*** | 0.003** | -0.001* | -0.112* |
| | (0.048) | (0.011) | (0.001) | (0.000) | (0.060) |
| NAICS Growth | Y | Y | Y | Y | Y |
| Other Controls | Y | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y | Y |
| State-Quarter FE | Y | Y | Y | Y | Y |
| Observations | 9550 | 9550 | 9550 | 9550 | 9563 |
| Adj. $R^2$ | 0.887 | 0.996 | 0.538 | 0.837 | 0.882 |

*Notes:* This table re-estimates the difference-in-differences (DID) specification similar to those in Table 5.2 with one difference: we add the industry-average sales growth as an additional control to tease out industry transitory shock by the pandemic. The dependent variables across columns 1 to 4 are the logarithm of net income, logarithm of sales, total capital expenditure over total asset, and software expenditure over total assets, and software expenditure over total assets, respectively. Omitting Tobin's q and ROE in the specification provides similar results (available upon request). Standard errors are clustered at the firm-level. Our results are also robust when controlling sector-quarter fixed-effects that control for any time-varying sector-level shocks.

## Table 5.5: WFH Feasibility and Stock Market Reactions - controlling the Demand Shock by Industry

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.043*** | 0.035*** | -0.005*** | -0.003*** |
| | (0.013) | (0.012) | (0.002) | (0.001) |
| NAICS Growth | Y | Y | Y | Y |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9538 | 9538 | 9538 | 9495 |
| Adj. $R^2$ | 0.454 | 0.151 | 0.716 | 0.752 |

*Notes:* This table re-estimates the difference-in-differences (DID) specification similar to those in Table 5.3 The dependent variables are total return and abnormal return in Columns (1) and (2); and return volatility and idiosyncratic volatility in Columns (3) and (4), respectively. Omitting Tobin's q and ROE in the specification provides similar results (available upon request). Standard errors are clustered at the firm-level. Our results are also robust when controlling sector-quarter fixed-effects that control for any time-varying sector-level shocks.

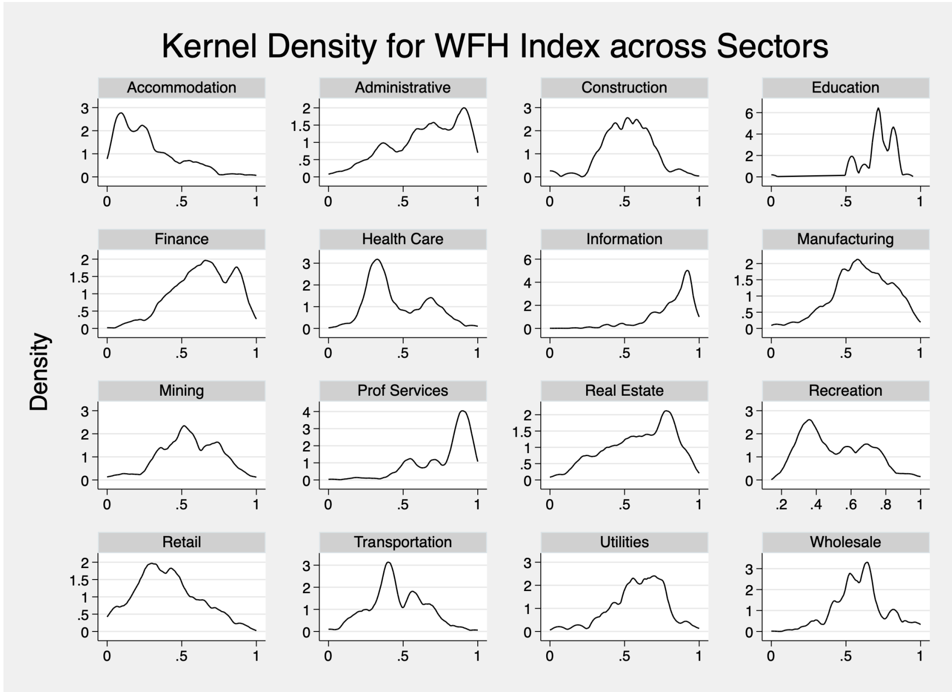



**Figure 5-1: Distribution of WFH indices by Sector**

*Notes:* This figure plots the distribution of the calculated WFH index by sector (2-digit NAICS from 2010-2019) for our sample.

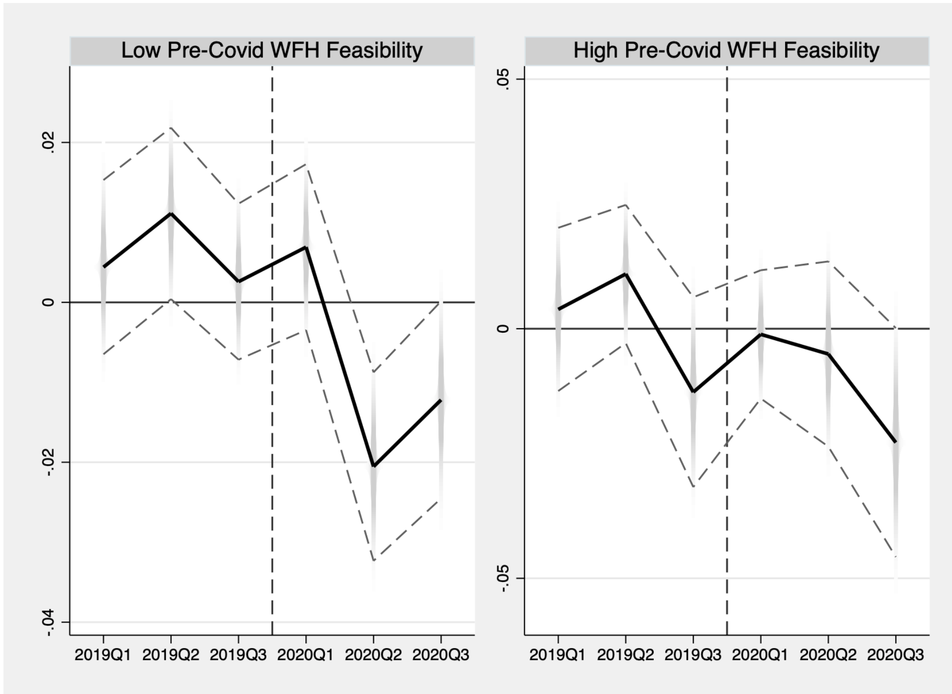



**Figure 5-2: Average Quarterly WFH Index by Firms with High vs. Low Pre-Covid-19 WFH Feasibility**

*Notes:* This figure plots the average quarterly WFH index for firms with high vs. low pre-Covid-19 WFH feasibility indices from 2019 Q1 to 2020 Q3. The values in 2019Q4 (as T-1) are used as the baseline group. Reported results are based on the specifications using the average quarterly WFH index and controlling for time-invariant firm unobservables and other firm controls. The WFH index is calculated based on our analysis sample. The source data comes from the Burning Glass Technologies (BGT) job vacancies data and the occupation-level WFH feasibility indicator by [Dingel and Neiman, 2020].

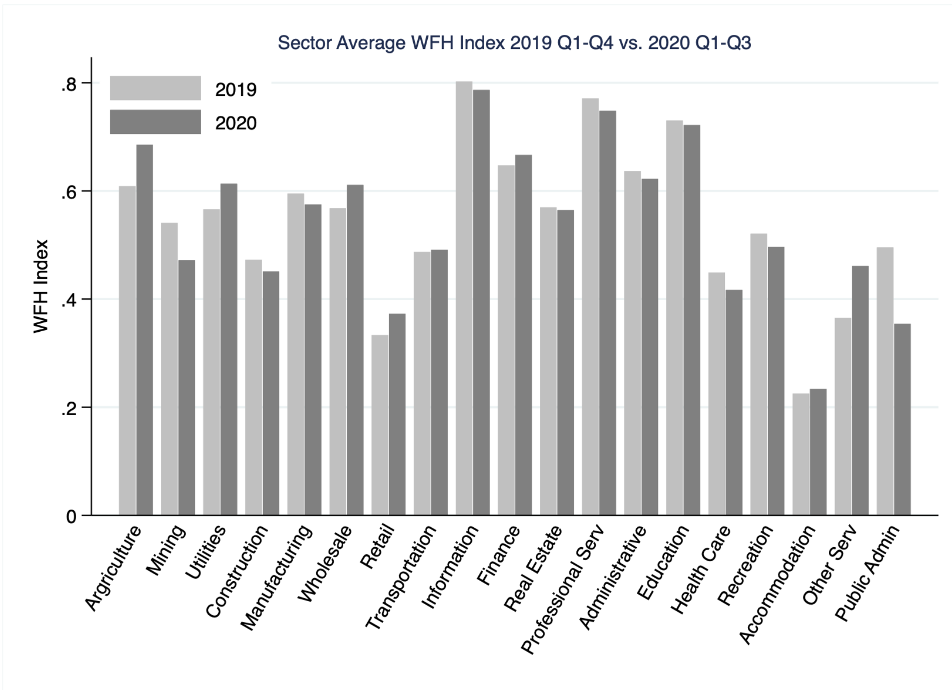
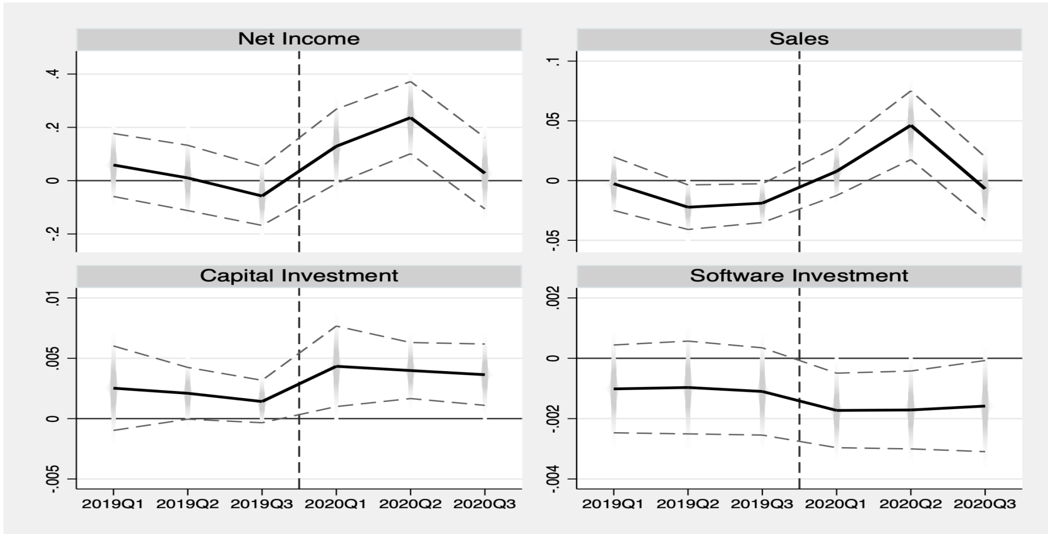
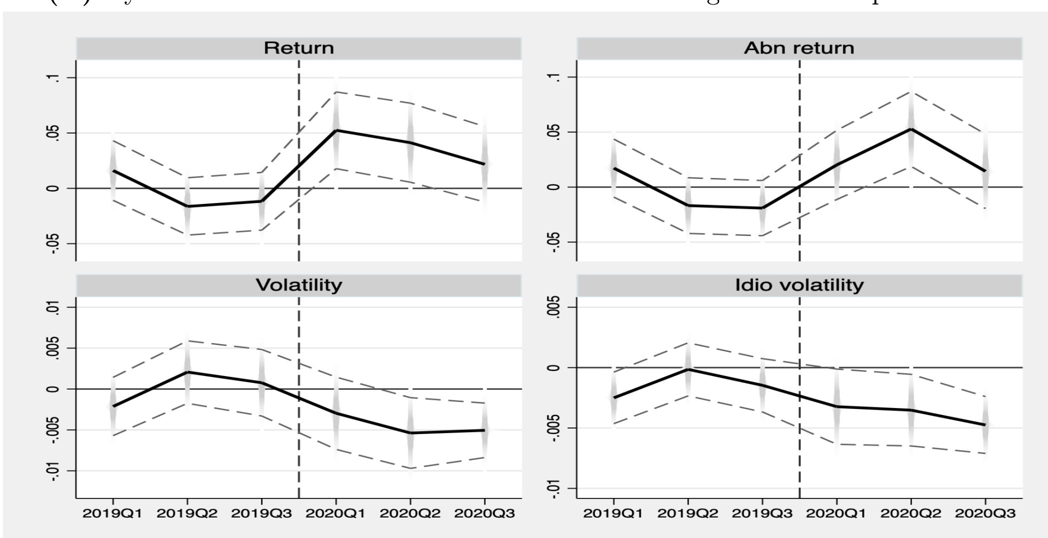



**Figure 5-3: Sector Average WFH indices (2019 and 2020)**

*Notes:* This figure plots the average WFH index by sector (2-digit NAICS) in both 2019 Q1-Q4 and 2020 Q1-Q3. The average WFH index is calculated based on our analysis sample. The source data comes from the Burning Glass Technologies (BGT) job vacancies data and the occupation-level WFH feasibility indicator by [Dingel and Neiman, 2020].




**(A)** Dynamics of the treatment effects in the DID setting on financial performance




**(B)** Dynamics of the treatment effects in the DID setting on stock returns and volatility

**Figure 5-4: Dynamics of the treatment effect in the DID setting**

*Notes:* Panels A and B plot the quarterly impact of HighWFH on firm outcome variables (A: financial performance; B: stock returns and volatility) for 2019Q1-2020Q3 with 90% confidence bands. The values in 2019Q4 (as T-1) are used as the baseline group. Results are based on specifications similar to the baseline models: we regress each outcome variable on the interaction of HighWFH and the quarter dummy alone with other firm controls. The values in 2019Q1 are used as the baseline for comparison.

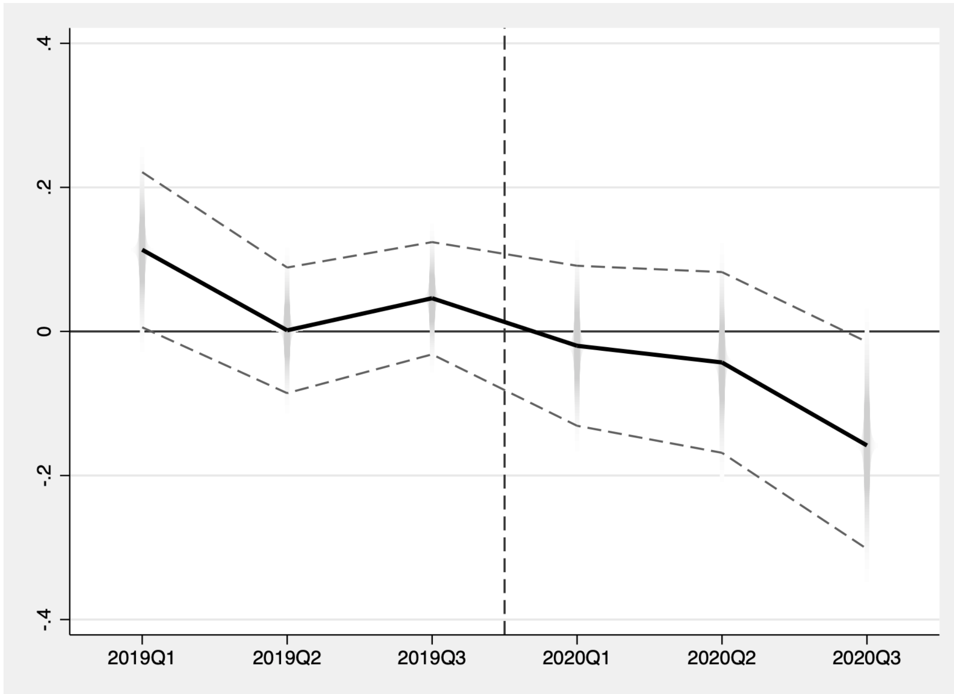



**Figure 5-5: Dynamics of the treatment effects in the DID setting on IT-related Job Postings**

*Notes:* This figure plots the impact of HighWFH on firms' IT-related in the period of 2019Q2-2020Q3 with the 90% confidence interval attached. Results are based on specifications similar to the baseline models: we regress each outcome variable on the interaction of HighWFH and the quarter dummy alone with other firm controls. The values in 2019Q4 (as T-1) are used as the baseline group. Results are based on specifications similar to the baseline models: we regress each outcome variable on the interaction of HighWFH and the quarter dummy alone with other firm controls. The values in 2019Q1 are used as the baseline for comparison.

# Chapter 6

# Tables & Figures for Chapter 3

**Table 6.1: Summary Statistics**

Panel A: Full Sample

| Variables | Never Treated | Treated | Overall |
|---|---|---|---|
| Avg. # of Job Postings | 10.584 | 129.171 | 12.495 |
| Avg. # of Cybersecurity Job Postings | 0.296 | 3.168 | 0.342 |
| Avg. Prob. of Cybersecurity Job Postings | 0.075 | 0.250 | 0.078 |
| Avg. # of Legal or PR Job Postings | 0.066 | 0.746 | 0.077 |
| Avg. Prob. of Legal or PR Job Postings | 0.069 | 0.261 | 0.072 |
| Avg. # of Non-relevant Job Postings | 10.222 | 125.257 | 12.076 |
| Avg. Prob. 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 |
| # 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 |
|---|---|---|
| Avg. # of Job Postings | 164.407 | 206.731 |
| Avg. # of Cybersecurity Job Postings | 3.709 | 4.858 |
| Avg. # of Legal or PR Job Postings | 0.714 | 0.862 |
| Avg. # of Non-relevant Job Postings | 159.984 | 201.011 |
| Media Share of Firm Name | 0.174 | 0.167 |
| Media Share of Firm Name + Breach | $2.266 \times 10^{-4}$ | $6.498 \times 10^{-4}$ |
| 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 3.3.

**Table 6.2: 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 3.1. 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 6.3: Effect of a Data Breach on Whether Firms Hire More Substantive Than Symbolic Jobs**

| | Two Period Model<br>(1) | Quarterly<br>(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 one is a two period difference-in-difference model. Column two 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 6.4: Effect of a Data Breach on Skill Demands**

| | Cybersecurity Skills | | | PR and Legal Skills | | | Not Relevant | | |
|---|---|---|---|---|---|---|---|---|---|
| | Probability | | Counts | Probability | | Counts | Probability | | Counts |
| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Post Breach | 0.016*** | | 0.885*** | 0.017*** | | -2.68 | 0.007 | | 17.18** |
| | (0.006) | | (0.280) | (0.006) | | (5.05) | (0.007) | | (6.79) |
| 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 | 0.414 | 0.301 | 0.301 | 0.349 | 0.301 | 0.301 | 0.352 |

*Notes:* This table mimics 6.2, except that the dependent variable captures 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 6.5: Effect of a Data Breach by Data Breach Type**

| | Cybersecurity Jobs | | PR and Legal Jobs | |
|---|---|---|---|---|
| | Cyber | Non-Cyber | Cyber | Non-Cyber |
| Variables | (1) | (2) | (3) | (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 3.1. In columns one and two, the dependent variable is defined as a binary indicator for whether the firm posts any cybersecurity jobs. In columns three and four the dependent variable is defined as a binary indicator for whether the firm posts any PR or legal jobs. In columns one and three, 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 two and four, 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

**Table 6.6: Effect of a Data Breach on Demand for Specific Cybersecurity Occupations**

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

*Notes:* Each column estimates the difference-in-differences specification outlined in Equation 3.1. 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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

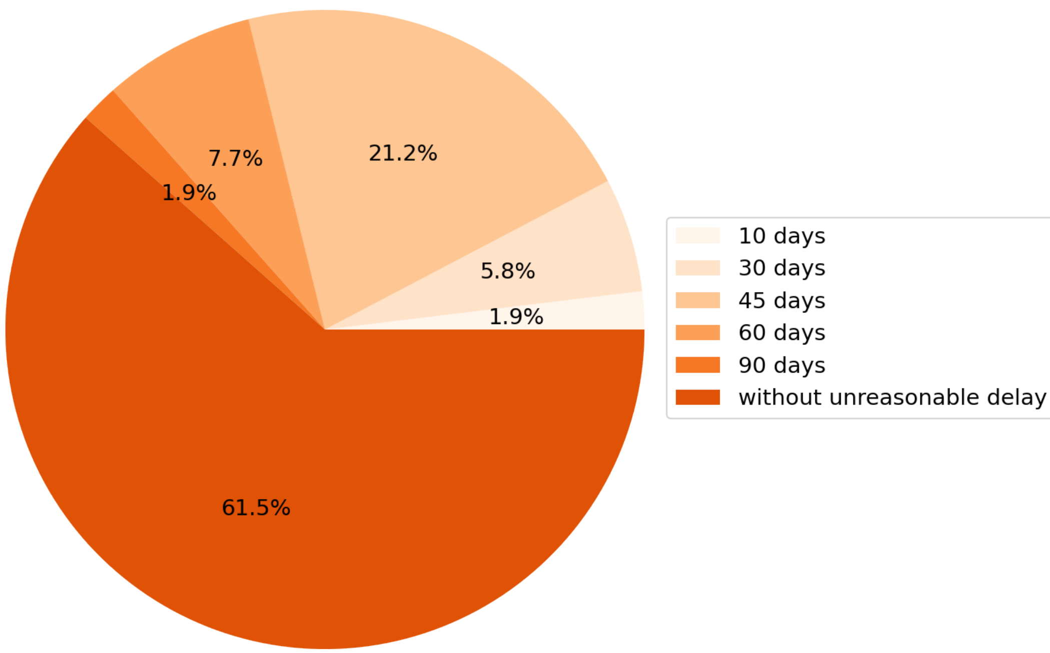
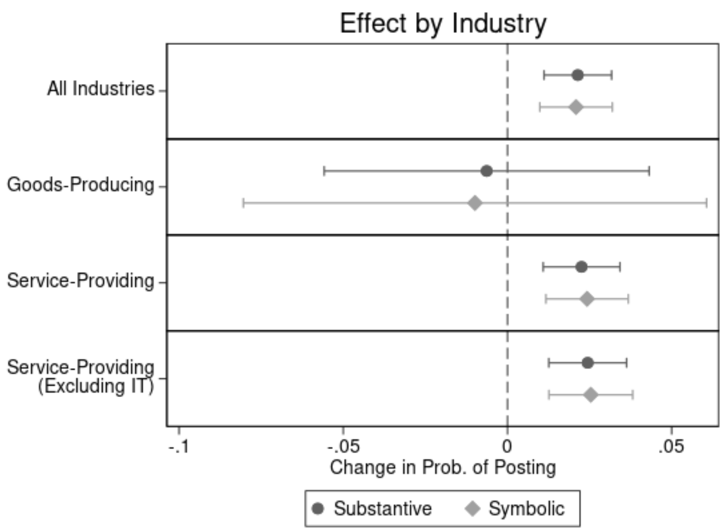
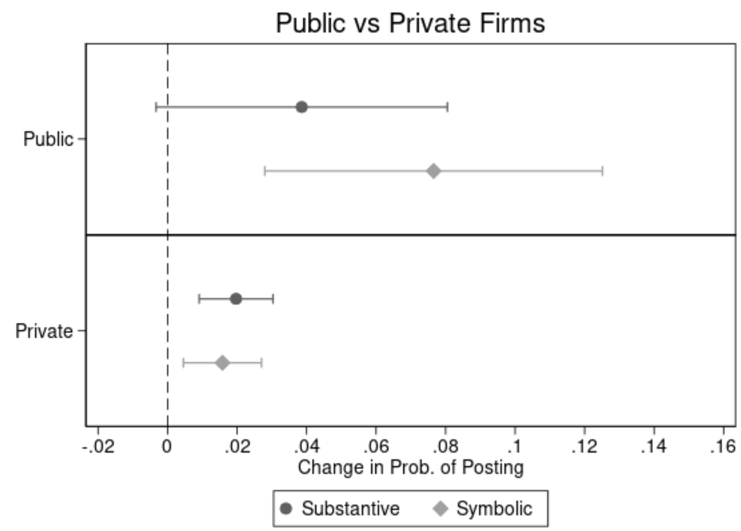



**Figure 6-1: Timely Notification Requirements by States' Data Breach Notification Laws**

*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.




**(A)** By Industry




**(B)** By Ownership

**Figure 6-2: Effect Heterogeneity of Data Breach on Firm's Talent Acquisition**

*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 3.1. (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 6.2.

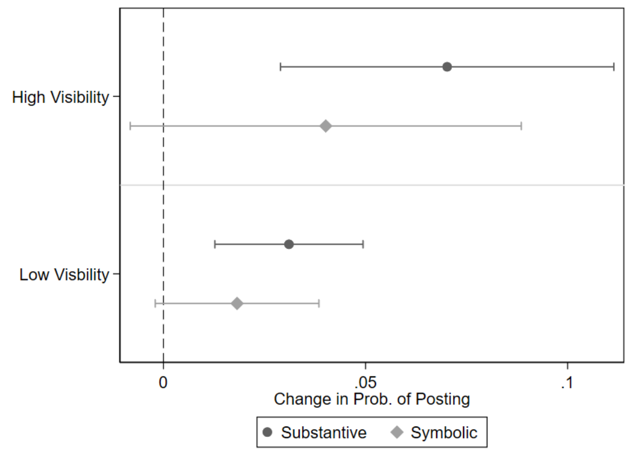



**Figure 6-3: Effect of Media Attention on Probability of Posting**

*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 0.005. In the low visibility panel, it is firms with post-breach media share below 0.005. Lines represent 95% confidence intervals.

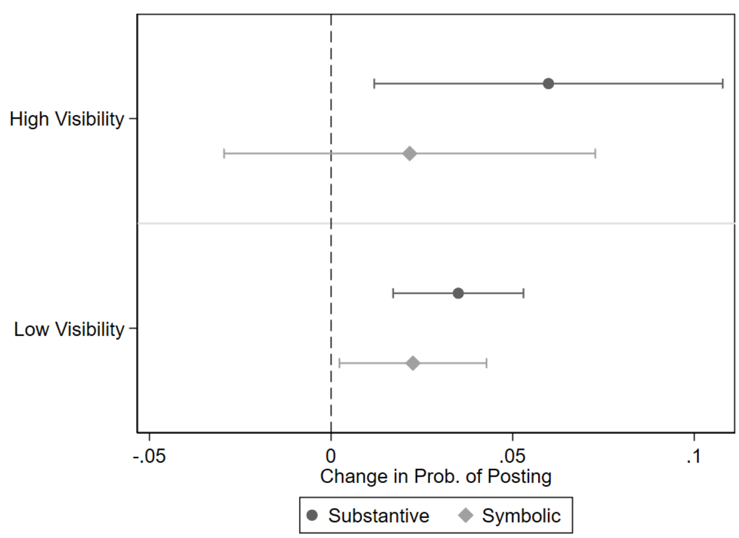



**Figure 6-4: Effect of Public Attention on Probability of Posting**

*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 5. In the low visibility panel, it is firms with post-breach search share below 5. Lines represent 95% confidence intervals.

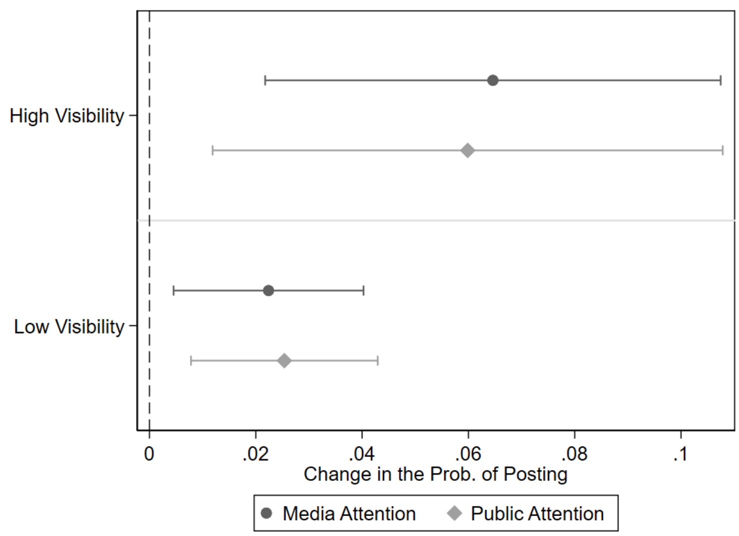



**Figure 6-5: Effect on Probability of More Substantive Postings**

*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.

# Chapter 7

# Appendices for Chapter 1

# Chapter 8

# Appendices for Chapter 2

## 8.1 Appendix A

**Table A1: WFH Feasibility and Firm Performance: State-Quarter Fixed Effects**

**(A)** Financial performance

| Models<br>Dependent Variable | (1)<br>Log Net Income | (2)<br>Log Sales | (3)<br>Norm. Cap. Exp. | (4)<br>Norm. Software Exp. |
|---|---|---|---|---|
| HighWFH × COVID | 0.151***<br>(0.048) | 0.033***<br>(0.011) | 0.002*<br>(0.001) | -0.001*<br>(0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| State-Quarter FE | Y | Y | Y | Y |
| Observations | 9550 | 9550 | 9550 | 9550 |
| Adj. $R^2$ | 0.886 | 0.996 | 0.538 | 0.837 |

**(B)** Stock Market performance

| Models<br>Dependent Variable | (1)<br>Return | (2)<br>Abn Return | (3)<br>Volatility | (4)<br>Idio. Volatility |
|---|---|---|---|---|
| HighWFH × COVID | 0.046***<br>(0.013) | 0.038***<br>(0.013) | -0.005***<br>(0.002) | -0.003**<br>(0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| State-Quarter FE | Y | Y | Y | Y |
| Observations | 9538 | 9538 | 9538 | 9495 |
| Adj. $R^2$ | 0.451 | 0.147 | 0.716 | 0.751 |

*Notes:* In Panels A and B, we repeat the analyses from Tables 5.2 and 5.3 using the same specifications except that we control for state-quarter fixed effects instead of quarter fixed effects to address the potential concern of geographic differences. All estimations include firm fixed effects. Standard errors are clustered at the firm level.

## 8.2 Appendix B

## Table B1: WFH and firm performance: matched sample analyses

### **(A)** Financial performance: matched sample analysis using PSM procedure

| Models<br>Dependent Variable | (1)<br>Log Net Income | (2)<br>Log Sales | (3)<br>Norm. Cap. Exp. | (4)<br>Norm. Software Exp. |
|---|---|---|---|---|
| HighWFH × COVID | 0.169*** | 0.030** | 0.001 | -0.001 |
| | (0.059) | (0.013) | (0.001) | (0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Pair FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 4290 | 4049 | 3509 | 3978 |
| Adj. $R^2$ | 0.873 | 0.995 | 0.712 | 0.815 |

### **(B)** Financial performance: matched sample analysis using CEM procedure

| Models<br>Dependent Variable | (1)<br>Log Net Income | (2)<br>Log Sales | (3)<br>Norm. Cap. Exp. | (4)<br>Norm. Software Exp. |
|---|---|---|---|---|
| HighWFH × COVID | 0.203*** | 0.031*** | 0.001* | -0.001* |
| | (0.051) | (0.011) | (0.001) | (0.000) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Pair FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 6905 | 6905 | 6905 | 6905 |
| Adj. $R^2$ | 0.883 | 0.996 | 0.740 | 0.826 |

### **(C)** Stock market performance: matched sample analysis using PSM procedure

| Models<br>Dependent Variable | (1)<br>Return | (2)<br>Abn Return | (3)<br>Volatility | (4)<br>Idio. Volatility |
|---|---|---|---|---|
| HighWFH × COVID | 0.063*** | 0.054*** | -0.007*** | -0.005** |
| | (0.014) | (0.014) | (0.002) | (0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Pair FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 5261 | 5261 | 5344 | 5242 |
| Adj. $R^2$ | 0.366 | 0.008 | 0.652 | 0.697 |

### **(D)** Stock market performance: matched sample analysis using CEM procedure

| Models<br>Dependent Variable | (1)<br>Return | (2)<br>Abn Return | (3)<br>Volatility | (4)<br>Idio. Volatility |
|---|---|---|---|---|
| HighWFH × COVID | 0.079*** | 0.068*** | -0.006*** | -0.004** |
| | (0.014) | (0.013) | (0.002) | (0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Pair FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 6892 | 6892 | 6891 | 6865 |
| Adj. $R^2$ | 0.475 | 0.155 | 0.708 | 0.717 |

*Notes:* Panels A and B report the matched sample results for financial performance proxies where the matching procedure is the propensity score matching (PSM) and coarsened exact matching (CEM), respectively. For results in Panel A, we estimate a probit model where the dependent variable is HighWFH and the controls include pre-Covid-19 return on asset (operating income / total asset), employment, and industry (2-digit NAICS) fixed effects using the pre-Covid-19 sample. We further implement a K2K nearest neighbor matching with no replacement and common support for each of the outcome variables. The results in Panel B are based on the coarsened exact matching approach using the same set of pre–19 variables. Panels C and D conduct the same matched sample analysis for stock performance. Standard errors are clustered at the firm level.

## 8.3 Appendix C

### Table C1: WFH Feasibility and firm performance: Continuous WFH proxy

**(A)** Financial performance

| Models<br>Dependent Variable | (1)<br>Log Net Income | (2)<br>Log Sales | (3)<br>Norm. Cap. Exp. | (4)<br>Norm. Software Exp. |
|---|---|---|---|---|
| HighWFH × COVID | 0.165*<br>(0.090) | 0.064***<br>(0.021) | 0.013***<br>(0.003) | -0.001*<br>(0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9568 | 9568 | 9568 | 9568 |
| Adj. $R^2$ | 0.886 | 0.996 | 0.522 | 0.844 |

**(B)** Stock Market performance

| Models<br>Dependent Variable | (1)<br>Return | (2)<br>Abn Return | (3)<br>Volatility | (4)<br>Idio. Volatility |
|---|---|---|---|---|
| HighWFH × COVID | 0.042*<br>(0.025) | 0.051**<br>(0.025) | -0.008**<br>(0.003) | -0.009***<br>(0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9550 | 9550 | 9550 | 9507 |
| Adj. $R^2$ | 0.448 | 0.145 | 0.713 | 0.750 |

*Notes:* Our baseline analysis is based on HighWFH, which takes the value of one if a firm's average WFH index calculated based on its annual job posting data during the pre-Covid-19 period (2010-2019) falls into the top quartile of the sample distribution and zero otherwise. In Panels A and B, we repeat the analysis using the raw, continuous WFH index. Firm and time fixed effects are also included in the estimation. Standard errors are clustered at the firm level.

## 8.4 Appendix D

### Table D1: Falsification Test (by Position of Firm Name by Alphabetic Order)

**(A)** Financial performance

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. |
| HighWFH × COVID | 0.005 | 0.009 | -0.001 | -0.000 |
| | (0.039) | (0.009) | (0.001) | (0.000) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9568 | 9568 | 9568 | 9568 |
| Adj. $R^2$ | 0.886 | 0.996 | 0.521 | 0.844 |

**(B)** Stock Market performance

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.016 | 0.010 | 0.002 | 0.002** |
| | (0.011) | (0.011) | (0.002) | (0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 9550 | 9550 | 9550 | 9507 |
| Adj. $R^2$ | 0.447 | 0.145 | 0.712 | 0.750 |

*Notes:* In Panels A and B, we repeat the analyses from Tables 5.2 and 5.3 to run a placebo test, in which instead of using the WFH index, we construct a firm index based on the alphabetical order of firms' names prior to the Covid-19 pandemic. All specifications are identical to the baseline specifications in Tables 5.2 and 5.3. Robust standard errors are clustered at the firm level.

## 8.5 Appendix E

## Table E1: WFH Feasibility and Firm Performance: Essential vs. Non-Essential

### (A) Essential Industries (Financial performance)

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. |
| HighWFH × COVID | 0.158** | 0.022 | 0.002* | -0.000 |
| | (0.070) | (0.018) | (0.001) | (0.000) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 4336 | 4336 | 4336 | 4336 |
| Adj. $R^2$ | 0.924 | 0.997 | 0.750 | 0.970 |

### (B) Non-Essential Industries (Financial performance)

| Models | (5) | (6) | (7) | (8) |
|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. |
| HighWFH × COVID | 0.183*** | 0.059*** | 0.007* | -0.001* |
| | (0.064) | (0.014) | (0.003) | (0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 5232 | 5232 | 5232 | 5232 |
| Adj. $R^2$ | 0.853 | 0.995 | 0.451 | 0.812 |

### (C) Essential Industries (Stock market performance)

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.028 | 0.028 | 0.001 | -0.001 |
| | (0.018) | (0.018) | (0.003) | (0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 4330 | 4330 | 4330 | 4323 |
| Adj. $R^2$ | 0.444 | 0.175 | 0.682 | 0.703 |

### (D) Non-Essential Industries (Stock market performance)

| Models | (5) | (6) | (7) | (8) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.062*** | 0.043*** | -0.010*** | -0.004*** |
| | (0.017) | (0.016) | (0.002) | (0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 5220 | 5220 | 5220 | 5184 |
| Adj. $R^2$ | 0.459 | 0.131 | 0.727 | 0.775 |

*Notes:* This table re-assesses the results reported in Tables 5.2 and 5.3 in two subsamples: essential vs. non-essential industries. The definitions of essential and non-essential industries are based on [Papanikolaou and Schmidt, 2020] and can be found in Appendix Table G3. Panels A and B report the results using financial performance proxies as the dependent variables. Panels C and D repeat the exercise with return and volatility as the dependent variables. The specifications in these panels are identical to those in Tables 5.2 and 5.3. Firm and time fixed effects are also included in the estimation. Standard errors are clustered at the firm level.

## Table E2: WFH Feasibility and Firm Performance: High-tech vs. Other Industries

**(A)** High-tech Industries (Financial performance)

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. |
| HighWFH × COVID | 0.095 | 0.030* | 0.000 | -0.002 |
| | (0.091) | (0.016) | (0.001) | (0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 1395 | 1395 | 1395 | 1395 |
| Adj. $R^2$ | 0.917 | 0.997 | 0.717 | 0.801 |

**(B)** Other Industries (Financial performance)

| Models | (5) | (6) | (7) | (8) |
|---|---|---|---|---|
| Dependent Variable | Log Net Income | Log Sales | Norm. Cap. Exp. | Norm. Software Exp. |
| HighWFH × COVID | 0.155*** | 0.036*** | 0.004** | -0.000 |
| | (0.057) | (0.014) | (0.002) | (0.000) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 8173 | 8173 | 8173 | 8173 |
| Adj. $R^2$ | 0.878 | 0.996 | 0.517 | 0.892 |

**(C)** High-tech Industries (Stock market performance)

| Models | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.041 | 0.019 | 0.001 | 0.001 |
| | (0.025) | (0.024) | (0.003) | (0.002) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 1395 | 1395 | 1395 | 1395 |
| Adj. $R^2$ | 0.380 | 0.156 | 0.695 | 0.776 |

**(D)** Other Industries (Stock market performance)

| Models | (5) | (6) | (7) | (8) |
|---|---|---|---|---|
| Dependent Variable | Return | Abn Return | Volatility | Idio. Volatility |
| HighWFH × COVID | 0.041*** | 0.031** | -0.004* | -0.002 |
| | (0.016) | (0.015) | (0.002) | (0.001) |
| Other Controls | Y | Y | Y | Y |
| Firm FE | Y | Y | Y | Y |
| Quarter FE | Y | Y | Y | Y |
| Observations | 8155 | 8155 | 8155 | 8112 |
| Adj. $R^2$ | 0.469 | 0.151 | 0.722 | 0.750 |

*Notes:* This table re-assesses the results reported in Tables 5.2 and 5.3 in two subsamples: high tech vs. other industries. The specifications in these panels are identical to those in Tables 5.2 and 5.3. The high-tech industries are defined as in [Decker et al., 2017] and can be reviewed in Appendix Table G2. Panels A and B report the results with financial performance proxies as the dependent variables. Panels C and D repeat the exercise with return and volatility as the dependent variables. Firm and time fixed effects are also included in the estimation. Standard errors are clustered at the firm level.

## 8.6 Appendix F

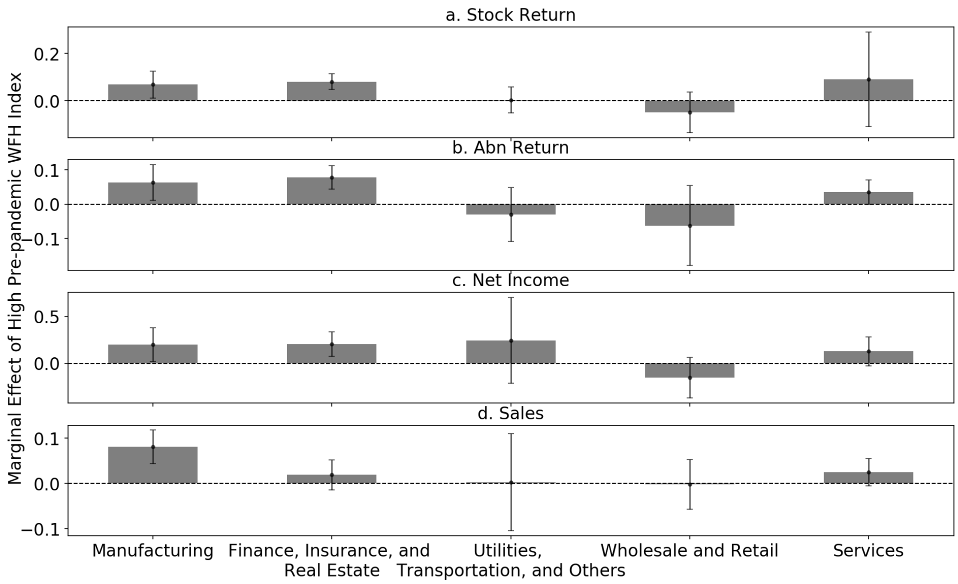



**Figure F1: WFH Feasibility and Firm Performance across Sectors**

*Notes:* This figure plots the marginal impact of HighWFH on firm outcome variables including financial performance and stock returns by sector with 90% confidence intervals. Reported results are based on the same specifications as in the baseline models: we regress each outcome variable on the interaction of HighWFH and the COVID dummy alone with other firm controls controlling for firm and quarter fixed-effects. We aggregate industry sectors following [Autor et al., 2020], which can also be found in Appendix G4. Please see the U.S. Census for the definition of 2-digit NAICS industries in Appendix G1.

## 8.7 Appendix G

**Table G1: Industry definition at 2-digit NAICS Level.**

| Sector (2-Digit NAICS) | Description |
|---|---|
| 11 | Agriculture, Forestry, Fishing and Hunting |
| 21 | Mining, Quarrying, and Oil and Gas Extraction |
| 22 | Utilities |
| 23 | Construction |
| 31-33 | Manufacturing |
| 42 | Wholesale Trade |
| 44-45 | Retail Trade |
| 48-49 | Transportation and Warehousing |
| 51 | Information |
| 52 | Finance and Insurance |
| 53 | Real Estate and Rental and Leasing |
| 54 | Professional, Scientific, and Technical Services |
| 55 | Management of Companies and Enterprises |
| 56 | Administrative and Support and Waste Management and Remediation Services |
| 61 | Educational Services |
| 62 | Health Care and Social Assistance |
| 71 | Arts, Entertainment, and Recreation |
| 72 | Accommodation and Food Services |
| 81 | Other Services (except Public Administration) |
| 92 | Public Administration |

**Table G2: High-Tech Industry definition at 4-digit NAICS Level according to [Decker et al., 2017].**

| 4-Digit NAICS | Description |
|---|---|
| 3341 | Computer and peripheral equipment manufacturing |
| 3342 | Communications equipment manufacturing |
| 3344 | Semiconductor and other electronic component manufacturing |
| 3345 | Navigational, measuring, electromedical, and control instruments manufacturing |
| 3254 | Pharmaceutical and medicine manufacturing |
| 3364 | Aerospace product and parts manufacturing |
| 5112 | Software publishers; 5161 Internet publishing and broadcasting |
| 5179 | Other telecommunications |
| 5181 | Internet service providers and Web search portals |
| 5182 | Data processing, hosting, and related services |
| 5413 | Architectural, engineering, and related services |
| 5415 | Computer systems design and related services |
| 5417 | Scientific research-and-development services |

**Table G3: Essential Industry definition at 4-digit NAICS Level according to [Papanikolaou and Schmidt, 2020].**

| 4-Digit NAICS | Description |
|---|---|
| 1111 | Oilseed and Grain Farming |
| 1112 | Vegetable and Melon Farming |
| 1113 | Fruit and Tree Nut Farming |
| 1119 | Other Crop Farming |
| 1121 | Cattle Ranching and Farming |
| 1122 | Hog and Pig Farming |
| 1123 | Poultry and Egg Production |
| 1124 | Sheep and Goat Farming |
| 1129 | Other Animal Production |
| 1141 | Fishing |
| 1142 | Hunting and Trapping |
| 1151 | Support Activities for Crop Production |
| 1152 | Support Activities for Animal Production |
| 2121 | Coal Mining |
| 2122 | Metal Ore Mining |
| 2123 | Nonmetallic Mineral Mining and Quarrying |
| 2131 | Support Activities for Mining |
| 2211 | Electric Power Generation, Transmission and Distribution |
| 2212 | Natural Gas Distribution |
| 2213 | Water, Sewage and Other Systems |
| 2373 | Highway, Street, and Bridge Construction |
| 3111 | Animal Food Manufacturing |
| 3112 | Grain and Oilseed Milling |
| 3113 | Sugar and Confectionery Product Manufacturing |
| 3114 | Fruit and Vegetable Preserving and Specialty Food Manufacturing |
| 3115 | Dairy Product Manufacturing |
| 3116 | Animal Slaughtering and Processing |
| 3117 | Seafood Product Preparation and Packaging |
| 3118 | Bakeries and Tortilla Manufacturing |
| 3119 | Other Food Manufacturing |
| 3121 | Beverage Manufacturing |
| 3254 | Pharmaceutical and Medicine Manufacturing |
| 3256 | Soap, Cleaning Compound, and Toilet Preparation Manufacturing |
| 3261 | Plastics Product Manufacturing |
| 3312 | Steel Product Manufacturing from Purchased Steel |
| 3313 | Alumina and Aluminum Production and Processing |
| 3331 | Agriculture, Construction, and Mining Machinery Manufacturing |

**Table G3: Essential Industry definition at 4-digit NAICS Level (Continued).**

| 4-Digit NAICS | Description |
|---|---|
| 3391 | Medical Equipment and Supplies Manufacturing |
| 4242 | Drugs and Druggists' Sundries Merchant Wholesalers |
| 4245 | Farm Product Raw Material Wholesalers |
| 4413 | Automotive Parts, Accessories, and Tire Stores |
| 4441 | Building Material and Supplies Dealers |
| 4451 | Grocery Stores |
| 4452 | Specialty Food Stores |
| 4453 | Beer, Wine, and Liquor Stores |
| 4461 | Health and Personal Care Stores |
| 4471 | Gasoline Stations |
| 4523 | General Merchandise Stores, including Warehouse Clubs and Supercenters |
| 4539 | Other Miscellaneous Store Retailers |
| 4541 | Electronic Shopping and Mail-Order Houses |
| 4812 | Nonscheduled Air Transportation |
| 4841 | General Freight Trucking |
| 4842 | Specialized Freight Trucking |
| 4851 | Urban Transit Systems |
| 4852 | Interurban and Rural Bus Transportation |
| 4853 | Taxi and Limousine Service |
| 4859 | Other Transit and Ground Passenger Transportation |
| 4861 | Pipeline Transportation of Crude Oil |
| 4862 | Pipeline Transportation of Natural Gas |
| 4885 | Freight Transportation Arrangement |
| 4911 | Postal Service |
| 4921 | Couriers and Express Delivery Services |
| 4922 | Local Messengers and Local Delivery |
| 4931 | Warehousing and Storage |
| 5173 | Telecommunications Resellers |
| 5179 | Other Telecommunications |
| 5211 | Monetary Authorities-Central Bank |
| 5221 | Depository Credit Intermediation |
| 5222 | Nondepository Credit Intermediation |
| 5223 | Activities Related to Credit Intermediation |
| 5231 | Securities and Commodity Contracts Intermediation and Brokerage |
| 5232 | Securities and Commodity Exchanges |
| 5239 | Other Financial Investment Activities |
| 5241 | Insurance Carriers |
| 5242 | Agencies, Brokerages, and Other Insurance Related Activities |
| 5251 | Insurance and Employee Benefit Funds |

Table G3: **Essential Industry definition at 4-digit NAICS Level (Continued).**

| 4-Digit NAICS | Description |
|---|---|
| 5259 | Other Investment Pools and Funds |
| 5411 | Legal Services |
| 5412 | Accounting, Tax Preparation, Bookkeeping, and Payroll Services |
| 5621 | Waste Collection |
| 5622 | Waste Treatment and Disposal |
| 5629 | Remediation and Other Waste Management Services |
| 6111 | Elementary and Secondary Schools |
| 6211 | Offices of Physicians |
| 6214 | Outpatient Care Centers |
| 6215 | Medical and Diagnostic Laboratories |
| 6216 | Home Health Care Services |
| 6219 | Other Ambulatory Health Care Services |
| 6221 | General Medical and Surgical Hospitals |
| 6223 | Specialty (except Psychiatric and Substance Abuse) Hospitals |
| 6231 | Nursing Care Facilities (Skilled Nursing Facilities) |
| 6233 | Continuing Care Retirement Communities and Assisted Living Facilities for the Elderly |
| 9211 | Executive, Legislative, and Other General Government Support |
| 9221 | Justice, Public Order, and Safety Activities |
| 9231 | Administration of Human Resource Programs |
| 9241 | Administration of Environmental Quality Programs |
| 9251 | Administration of Housing Programs, Urban Planning, and Community Development |
| 9261 | Administration of Economic Programs |
| 9271 | Space Research and Technology |
| 9281 | National Security and International Affairs |

**Table G4: Aggregated Industry definition following [Autor et al., 2020].**

| Aggregated Industry Sector | 2-Digit NAICS Code |
|---|---|
| Manufacturing | 31, 32, 33 |
| Finance, Insurance, & Real Estate | 52, 53 |
| Utilities, Transportation, and Others | 11, 21, 22, 23, 48, 49, 99 |
| Wholesale and Retail trade | 42, 44, 45 |
| Services | 51, 54, 55, 56, 61, 62, 71, 72, 81 |

## 8.8 Appendix H

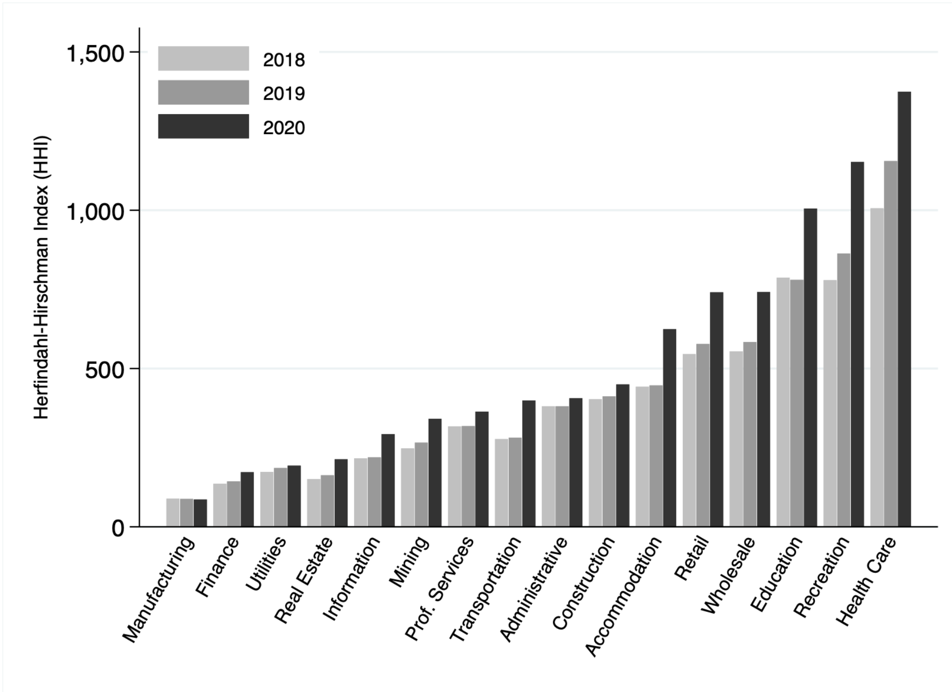



**Figure H1: Herfindahl–Hirschman Index 2018-2020**

*Notes:* This figure plots the conventional Herfindahl–Hirschman Index (HHI) between 2018 to 2020 across the major sectors (2-digit NAICS) calculated based on the reported revenue using the firms in the Compustat data.

# Chapter 9

# Appendices for Chapter 3

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

*Notes:* 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: 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:* 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 A3: Effect of Data Breach on Demanding Related Skills Using Breach Events with the Largest Number of Breached Records**

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

*Notes:* 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 A4: Additional Robustness Checks for Effect of Data Breach on Skill Demands**

| | Cybersecurity Skills | | | PR and Legal Skills | | |
|---|---|---|---|---|---|---|
| | 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 | 0.020*** | | 1.026 | 0.018*** | | 0.775 |
| | (0.006) | | (0.038) | (0.007) | | (.215) |
| 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:* 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

## 9.2 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.[1] 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.

We also investigate how these monthly dynamics differ by the types of the breach

[1] 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 five and six in Table A2 but also bears a risk of larger measurement error due to the unclear filling rates of each job postings.

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 cyberattacks. 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 6-2, 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 6.3 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.[2]

[2] Additionally, we investigate the spillover effect on breached firms' competitors in Appendix 9.3 and find that the spillover effect can potentially mask the immediate reaction of breached firms.

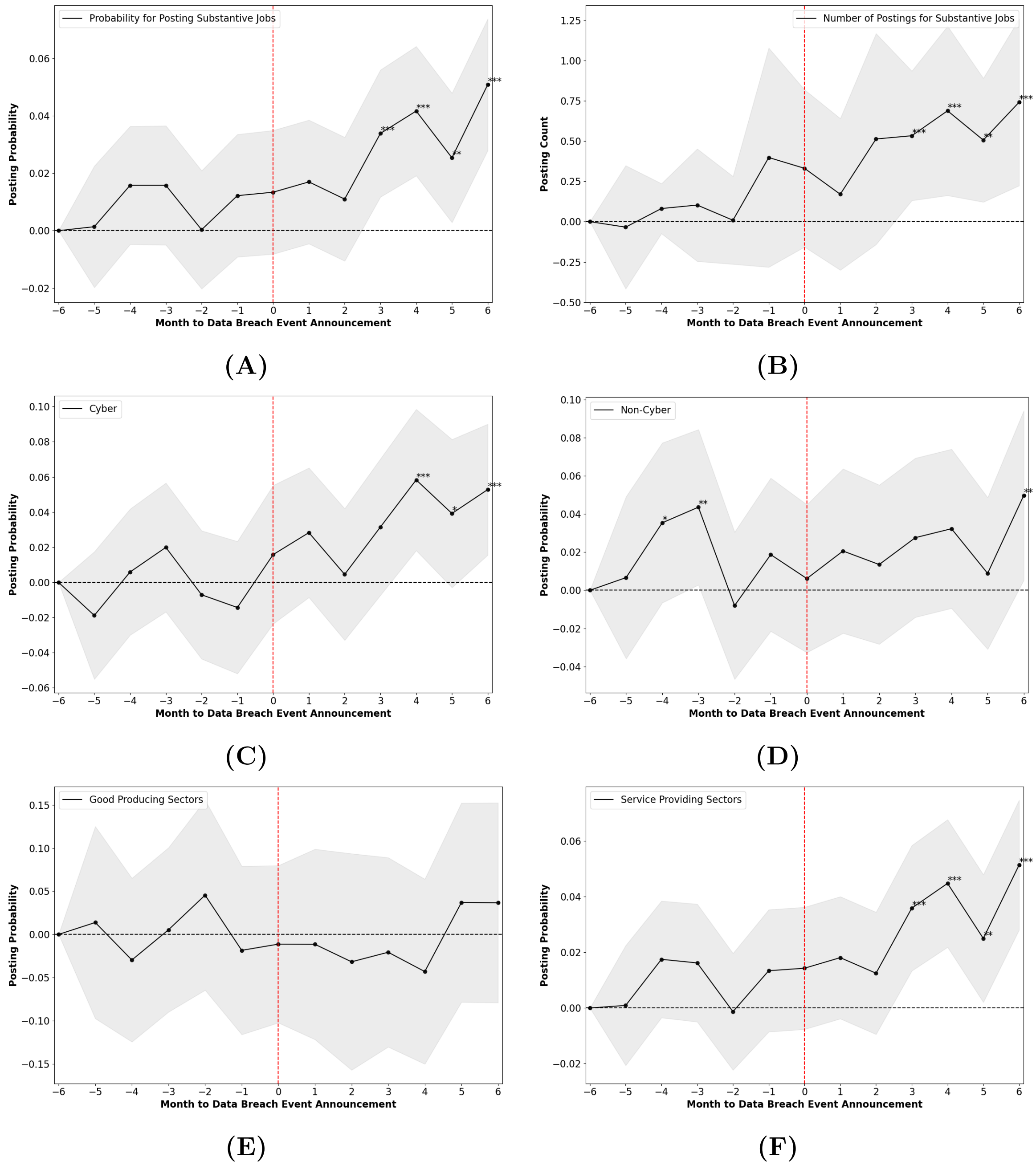



**Figure B1: Monthly Dynamic Effect on Substantive Hiring**

*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.

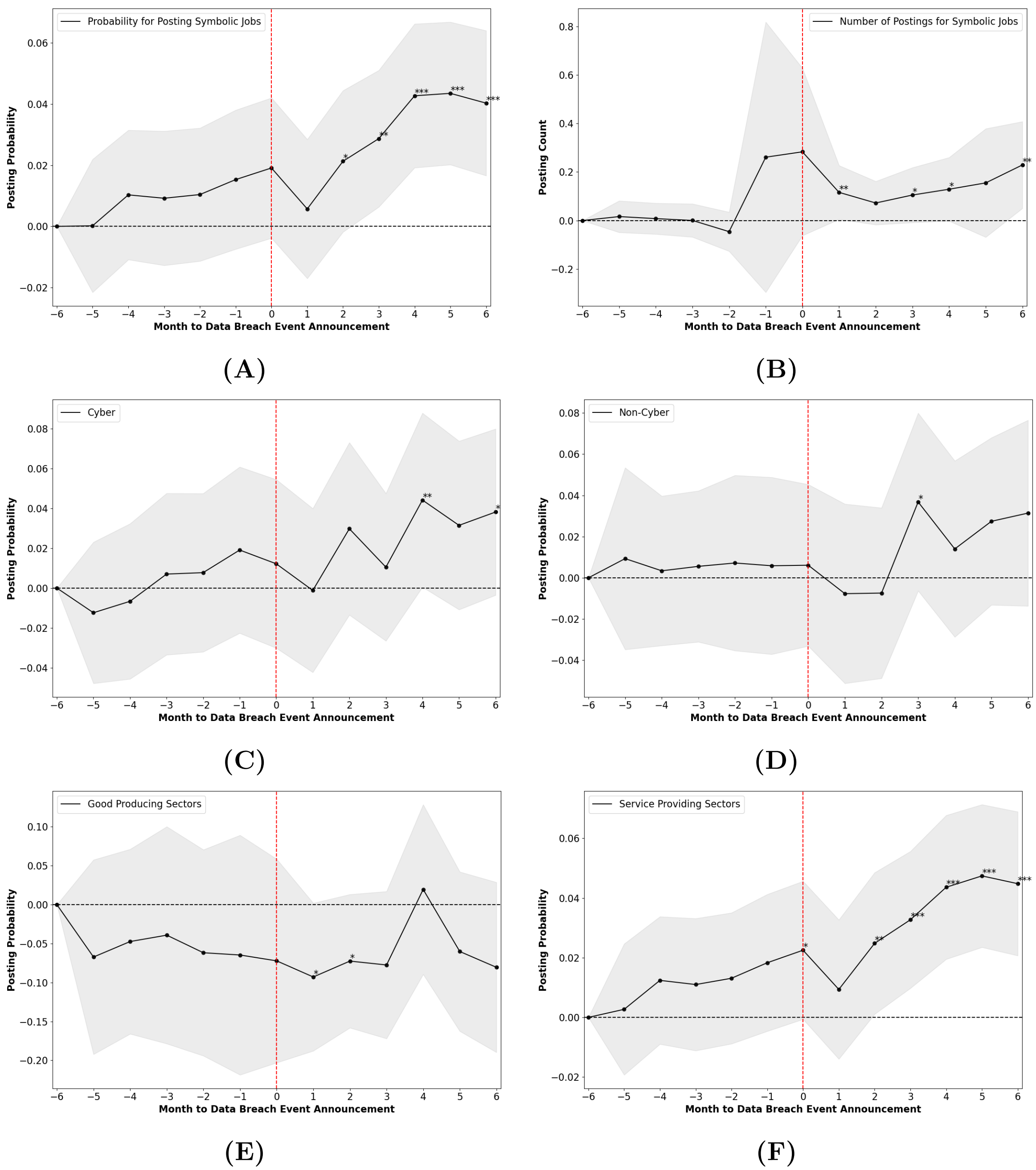



**Figure B2: Monthly Dynamic Effect on Symbolic Hiring**

*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.

## 9.3 Appendix C - Spillover Effects on Competitors

After examining the effects of data breaches on firms' own postings, we move to spillover effects on their peers. Spillovers from a competitor's data breach notification are estimated using a fixed effects regression model. The key econometric assumption is that competitors' data breaches within the same industry are exogenous to firms that did not experience a breach. Specifically, we estimate:

$$Postings_{i,j,t} = \alpha_0 + \sum_{p=-5}^{6} \alpha_p D_{i,t+p} + \lambda_i + \lambda_y + \lambda_{m,j} + \varepsilon_{i,j,t} \tag{9.1}$$

estimated using each firm $i$ in four-digit NAICS industry $j$, limited to industries that have one or two non-overlapping data breaches over the 2010 to 2020 time period. As before, $Postings_{ijt}$ is the number of cybersecurity related postings or non-cybersecurity related postings. $Breach_{i,t+p}$ for $p \in [-5, 6]$ is an indicator variable equal to one in month $t + p$ if a data breach occurred exactly $p$ months before in another firm in the same four-digit industry and zero otherwise. The omitted category is again six months before. The following fixed effects are included:

- $\lambda_i$: firm fixed effect (to control for firm heterogeneity)
- $\lambda_y$: year fixed effect (to control for changes in overall posting behavior)
- $\lambda_{m,j}$: calendar month of year by two-digit NAICS industry fixed effect (to capture industry-specific seasonality)

This is initially estimated as a fixed effects linear model.[3] Standard errors are clustered at the four-digit NAICS industry level.

The spillover effects are estimated using the sample of four-digit NAICS industries that experience two non-overlapping data breaches.[4] This is a substantially different sample, as some industries experience very frequent breaches. However, we make this restriction to allow event isolation and to get more accurate estimation of spillover

[3] Results estimated using a fixed effects Poisson model are available upon request.

[4] Non-overlapping is defined as no data breach occurring within the same four-digit NAICS within thirteen months of one another because our event window is 6 months on each side.

effects. A total of 50 four-digit industries fall into this category. Furthermore, firms that experience data breaches are excluded from this analysis, since they are directly treated.

The event study coefficients on the spillover effects are displayed in Figure C1. There appears to be a small but statistically significant effect of a competitors' data breach on cybersecurity postings. A data breach in the same industry is associated with additional postings for cybersecurity personnel, controlling for firm-fixed effects, and seasonality in posting behavior and year.[5] As a falsification test, we check whether non-cybersecurity postings increase as well. Reassuringly, there is no effect on non-cybersecurity postings.

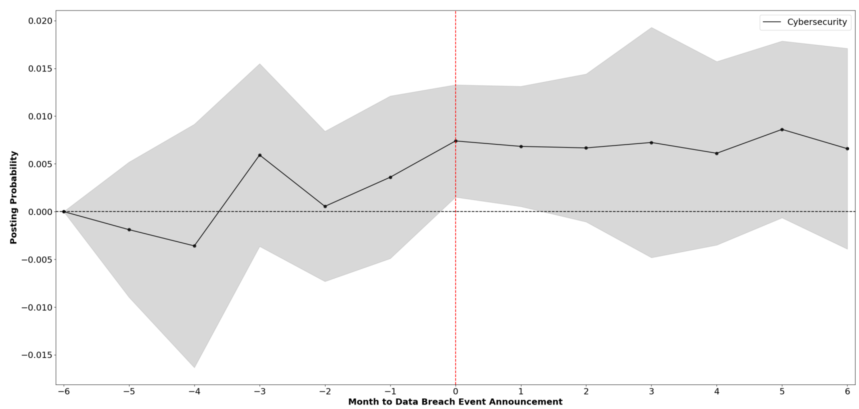



**Figure C1: Effect of a Data Breach on Competitors' Job Postings**

*Notes:* Coefficients from a fixed effects regression on the binary variable indicating cybersecurity related job postings, following a competitors' data breach. The omitted category is six months before the breach notification date. The sample is limited to firms in four-digit NAICS industries that only have one or two non-overlapping data breaches. Furthermore, it is limited to firms who have at least 100 postings from January 2010 to April 2020. Regressions control for firm fixed effects, year fixed effects, and calendar-month-of-year by NAICS two digit industry fixed effects. Standard errors are clustered at the four-digit NAICS industry level. Shaded areas are 95 confidence intervals.

[5] Currently, the window is limited to six months before and after, although for the spillover analysis, it may be valuable to extend the window as firms' HR processes to post a job can be burdensome and the strongest effect appears to be at the end of the window.

## 9.4 Appendix D - Preventative Effects of Cybersecurity Hiring for Future Breaches

Does hiring cybersecurity professionals effectively prevent future data breaches? To answer this question, we look at the probability of experiencing a data breach more than once, and its relationship with previous cybersecurity postings. There are a total of 518 firms experiencing cyber-related data breaches in our sample. Only 87 of these firms experience a second data breach within 24 months of the initial breach. Moreover, the 57 of these 87 breaches happen within three months of the first breach.

For the firms who have experienced a breach, we regress the probability of a second breach three to six months following the initial breach on a dummy variable for posting a cybersecurity role during the three months following the breach, and a control for total number of postings from the firm during the sample period, as a proxy for firm size. These results are displayed in Table D1 Panel A. There is no significant effect of hiring on preventing a future data breach. In fact, the coefficient is positive as opposed to negative. We repeat the exercise in Column 6, lengthening the possible second breach window. Again, there is no significant effect of a posting on a future breach. Columns 7 and 8 change the window associated with the cybersecurity investment to six months. Column 7 allows the breach window to be six to twelve months following the initial breach, while Column 8 allows the breach window to be six to eighteen months following the breach. The sign on the coefficient of interest flips; however, the sign is still significant.

Panel B repeats the exercise for breaches that are cyber breaches. In this case, the coefficients in Column 5 and 6 are significant and positive. This suggests that firms that post cybersecurity jobs are more likely to experience a second breach. This regression is inevitably subject to more endogeneity concerns than our main specification, which controls for industry. However, our small sample and the rarity of a second breach limits the more nuanced design for this analysis. In Columns 7 and 8, the coefficients drop in size substantially, and are now statistically insignificant.

## Table D1: Effect of Cybersecurity Job Posting on Future Events

| | All Types of 2nd Events | | | | Cyber Attacks Only | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Variables | mon 3-6 | mon 3-12 | mon 6-12 | mon 6-18 | mon 3-6 | mon 3-12 | mon 6-12 | mon 6-18 |
| PC Postings in Months 0-3 | 0.018 | 0.017 | | | 0.018* | 0.026** | | |
| | (0.012) | (0.016) | | | (0.011) | (0.012) | | |
| PC Postings in Months 0-6 | | | -0.003 | -0.004 | | | 0.006 | 0.015* |
| | | | (0.009) | (0.017) | | | (0.004) | (0.008) |
| Total Postings | Y | Y | Y | Y | Y | Y | Y | Y |
| Industry FE | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 518 | 518 | 518 | 517 | 518 | 518 | 518 | 517 |
| R-squared | 0.030 | 0.025 | 0.011 | 0.011 | 0.038 | 0.035 | 0.011 | 0.017 |

*Notes:* In all columns, we only investigate the effect of cybersecurity hiring following cyber attacks. The dependent variables in columns one to four is whether to have a second data breach event with in the studied windows while in columns five to 8 we only focus on the second cyber attacks. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

## 9.5 Appendix E - Definitions of Occupational and Skill groups

### 9.5.1 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 E1.

Similarly, we define job postings for public relations (P&R) and legal occupations, which we use to identify symbolic adoption, as those which were tagged with any of the SOC codes in table E2.[6]

### 9.5.2 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 E3 as Cybersecurity, Legal, and P&R skills, respectively.

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

**Table E1: 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 E2: Definitions of Legal and P&R 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 | P&R |
| 11-2032 | Public Relations Managers | P&R |
| 27-3030 | Public Relations Specialists | P&R |
| 27-3031 | Public Relations Specialists | P&R |

**Table E3: Definitions of Cybersecurity, Legal, and P&R 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 (P&R) | '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' |

# Chapter 1 - Bibliography


[Acemoglu and Restrepo, 2018] Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

[Acemoglu and Restrepo, 2019] Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30.

[Agrawal et al., 2019] Agrawal, A., Gans, J. S., and Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, 33(2):31–50.

[Alabdulkreem et al., 2018] Alabdulkreem, A., Frank, M. R., Sun, L., Alshebli, B., Hidalgo, C., and Rahwan, I. (2018). Unpacking the polarization of workplace skills. *Science Advances*, 4.

[Autor, 2019] Autor, D. (2019). Work of the Past, Work of the Future. *National Bureau of Economic Research Working Paper Series*, No. 25588.

[Autor and Dorn, 2013] Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *American Economic Review*, 103(5):1553–1597.

[Autor et al., 2003] Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

[Bartel et al., 2007] Bartel, A., Ichniowski, C., and Shaw, K. (2007). How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *Quarterly Journal of Economics*, 122(4):1721–1758.

[Beaudry et al., 2015] Beaudry, P., Green, D. A., and Sand, B. M. (2015). The Great Reversal in the Demand for Skill and Cognitive Tasks. *Journal of Labor Economics*, 34(S1):S199–S247.

[Becker, 1964] Becker, G. S. (1964). *Human capital; a theoretical and empirical analysis, with special reference to education.* National Bureau of Economic Research.

General series: no. 80. New York, National Bureau of Economic Research; distributed by Columbia University Press, 1964.

[Brynjolfsson and McElheran, 2016a] Brynjolfsson, E. and McElheran, K. (2016a). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, 106(5):133–139.

[Brynjolfsson and McElheran, 2016b] Brynjolfsson, E. and McElheran, K. (2016b). The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*, 106(5):133–139.

[Brynjolfsson et al., 2018] Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, 108:43–47.

[Cha, 2007] Cha, S.-H. (2007). Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307.

[Deming and Kahn, 2018] Deming, D. and Kahn, L. B. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

[Deming and Noray, 2018] Deming, D. J. and Noray, K. (2018). STEM Careers and Technological Change. *NBER Working Paper*.

[Duckworth and Gross, 2014] Duckworth, A. and Gross, J. J. (2014). Self-Control and Grit. *Current Directions in Psychological Science*, 23(5):319–325.

[Felten et al., 2018] Felten, E. W., Raj, M., and Seamans, R. (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108:54–57.

[Firpo et al., 2011] Firpo, S., Fortin, N. M., and Lemieux, T. (2011). Occupational Tasks and Changes in the Wage Structure. *IZA Discussion Paper Series*, (5542).

[Gathmann and Schoenberg, 2010] Gathmann, C. and Schoenberg, U. (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, 28(1):1–49.

[Gibbons et al., 2005] Gibbons, R., Katz, L. F., Lemieux, T., and Parent, D. (2005). Comparative Advantage, Learning, and Sectoral Wage Determination. *Journal of Labor Economics*, 23(4):681–723.

[Goldin and Katz, 2009] Goldin, C. and Katz, L. F. (2009). The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005.

[Goldin and Katz, 2008] Goldin, C. D. and Katz, L. F. (2008). The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005.

[Goos et al., 2014] Goos, M., Manning, A., and Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, 104(8):2509–2526.

[Gray and Suri, 2019] Gray, M. L. and Suri, S. (2019). *Ghost work : how to stop Silicon Valley from building a new global underclass.* Boston : Houghton Mifflin Harcourt, 2019.

[Hanushek et al., 2015] Hanushek, E. A., Schwerdt, G., Wiederhold, S., and Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73:103–130.

[Heckman et al., 2003] Heckman, J. J., Lochner, L. J., and Todd, P. E. (2003). Fifty Years of Mincer Earnings Regressions. *NBER Working Paper*.

[Hershbein and Kahn, 2018] Hershbein, B. and Kahn, L. B. (2018). Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. *American Economic Review*, 108(7):1737–1772.

[Horton and Tambe, 2019] Horton, J. J. and Tambe, P. (2019). The Death of a Technical Skill. *Working Paper*.

[Katz and Autor, 1999] Katz, L. and Autor, D. (1999). Changes in the Wage Structure and Earnings Inequality. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics, vol. 3A*, pages 1463–1555.

[Lemieux, 2006] Lemieux, T. (2006). The 'Mincer Equation' Thirty Years after Schooling, Experience, and Earnings. In Grossbard, S., editor, *Jacob Mincer: A Pioneer of Modern Labor Economics*, pages 127–145. New York:, U British Columbia.

[Michaels et al., 2014] Michaels, G., Natraj, A., and Van Reenen, J. (2014). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years. *Review of Economics and Statistics*, 96(1):60–77.

[Mincer, 1958] Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302.

[Mincer, 1974] Mincer, J. (1974). *Schooling, experience, and earnings.* Human behavior and social institutions: 2. New York, National Bureau of Economic Research; distributed by Columbia University Press, 1974.

[Restrepo, 2015] Restrepo, P. (2015). Skill Mismatch and Structural Unemployment. *Working Paper*.

[Spitz-Oener, 2006] Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, 24(2):235–270.

[Webb, 2019] Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. *NBER Working Paper*.

[Welch, 1969] Welch, F. (1969). Linear Synthesis of Skill Distribution. *Journal of Human Resources*, 4(3):311–327.

# Chapter 2 - Bibliography


[alm, 2019] (2019). *Telework in the 21st century : an evolutionary perspective*. The ILO future of work series. Edward Elgar Publishing, Cheltenham, UK ;.

[Acemoglu et al., 2020] Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2020). AI and Jobs: Evidence from Online Vacancies. *National Bureau of Economic Research Working Paper Series*, No. 28257.

[Acharya and Steffen, 2020] Acharya, V. V. and Steffen, S. (2020). The Risk of Being a Fallen Angel and the Corporate Dash for Cash in the Midst of COVID. *The Review of Corporate Finance Studies*, 9(3):430–471.

[Albuquerque et al., 2020] Albuquerque, R., Koskinen, Y., Yang, S., and Zhang, C. (2020). Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. *The Review of Corporate Finance Studies*, 9(3):593–621.

[Alipour et al., 2021] Alipour, J.-V., Fadinger, H., and Schymik, J. (2021). My Home Is my Castle – The Benefits of Working from Home During a Pandemic Crisis Evidence from Germany. *Journal of Public Economics*.

[Allen et al., 2015] Allen, T. D., Golden, T. D., and Shockley, K. M. (2015). How Effective Is Telecommuting? Assessing the Status of Our Scientific Findings. *Psychological Science in the Public Interest*, 16(2):40–68.

[Ang et al., 2006] Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61(1):259–299.

[Atkeson, 2020] Atkeson, A. G. (2020). What Will be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. *Research Department Staff Reports (Federal Reserve Bank of Minneapolis)*, pages 1–25.

[Autor et al., 2020] Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, 135(2):645–709.

[Azar et al., 2020] Azar, J., Marinescu, I., Steinbaum, M., and Taska, B. (2020). Concentration in US labor markets: Evidence from online vacancy data. *Labour Economics*, 66.

[Barrero et al., 2021] Barrero, J. M., Bloom, N., and Davis, S. J. (2021). Why working from home will stick. Technical report, National Bureau of Economic Research.

[Benzell et al., 2020] Benzell, S. G., Collis, A., and Nicolaides, C. (2020). Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. *Proceedings of the National Academy of Sciences*, 117(26):14642 LP – 14644.

[Bloom et al., 2020] Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are Ideas Getting Harder to Find? *American Economic Review*.

[Bloom et al., 2015] Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015). Does Working from Home Work? Evidence from a Chinese Experiment*. *Quarterly Journal of Economics*, 130(1):165–218.

[Brynjolfsson et al., 2020] Brynjolfsson, E., Horton, J. J., Ozimek, A., Rock, D., Sharma, G., and TuYe, H. (2020). Covid-19 and Remote Work: an Early Look At Us Data. *NBER Working Paper*, pages 1–30.

[Brynjolfsson and McAfee, 2014] Brynjolfsson, E. and McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

[Brynjolfsson and Milgrom, 2012] Brynjolfsson, E. and Milgrom, P. (2012). Complementarity in Organizations. In Gibbons, R. and Roberts, J., editors, *The Handbook of Organizational Economics*, chapter 1, pages 11–55. Princeton University Press.

[Callaway and Sant'Anna, 2020] Callaway, B. and Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*.

[Campello et al., 2019] Campello, M., Gao, J., and Xu, Q. (2019). Personal Income Taxes and Labor Downskilling: Evidence from 27 Million Job Postings. *Kelley School of Business Research Paper*.

[Choudhury et al., 2020] Choudhury, P., Foroughi, C., and Larson, B. (2020). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal*, November:655–683.

[Coibion et al., 2020] Coibion, O., Gorodnichenko, Y., and Weber, M. (2020). Labor Markets During the COVID-19 Crisis: A Preliminary View. *NBER Working Paper 27017*.

[Collins, 2005] Collins, M. (2005). The (not so simple) case for teleworking: A study at Lloyd's of London. *New Technology, Work and Employment*, 20(2):115–132.

[Cui et al., 2021] Cui, R., Ding, H., and Zhu, F. (2021). Gender inequality in research productivity during the COVID-19 pandemic. *Manufacturing & Service Operations Management*.

[Das et al., 2020] Das, S., Steffen, S., Clarke, W., Reddy, P., Brynjolfsson, E., and Fleming, M. (2020). Learning Occupational Task-Shares Dynamics for the Future of Work. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, pages 36–42, New York, NY, USA. Association for Computing Machinery.

[De' et al., 2020] De', R., Pandey, N., and Pal, A. (2020). Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International Journal of Information Management*, 55(Impact of COVID-19 Pandemic on Information Management Research and Practice: Editorial Perspectives).

[Decker et al., 2017] Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2017). Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown. *American Economic Review*, 107(5):322–326.

[Deming and Kahn, 2018] Deming, D. and Kahn, L. B. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

[Ding et al., 2020] Ding, W., Levine, R., Lin, C., and Xie, W. (2020). Corporate Immunity to the COVID-19 Pandemic. *National Bureau of Economic Research*.

[Dingel and Neiman, 2020] Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189.

[Forsythe et al., 2020] Forsythe, E., Kahn, L. B., Lange, F., and Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of public economics*, 189:104238.

[Gajendran and Harrison, 2007] Gajendran, R. S. and Harrison, D. A. (2007). The Good, the Bad, and the Unknown About Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences. *Journal of Applied Psychology*, 92(6):1524–1541.

[Golden, 2006] Golden, T. D. (2006). The role of relationships in understanding telecommuter satisfaction. *Journal of Organizational Behavior*, 27(3):319–340.

[Golden and Veiga, 2005] Golden, T. D. and Veiga, J. F. (2005). The Impact of Extent of Telecommuting on Job Satisfaction: Resolving Inconsistent Findings. *Journal of Management*, 31(2):301–318.

[Hayes et al., 2020] Hayes, S., Priestley, J. L., Ishmakhametov, N., and Ray, H. E. (2020). "i'm not working from home, i'm living at work": Perceived stress and work-related burnout before and during covid-19.

[Hershbein and Kahn, 2018] Hershbein, B. and Kahn, L. B. (2018). Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. *American Economic Review*, 108(7):1737–1772.

[Ilut et al., 2018] Ilut, C., Kehrig, M., and Schneider, M. (2018). Slow to Hire, Quick to Fire: Employment Dynamics with Asymmetric Responses to News. *Journal of Political Economy*, 126(5):2011–2071.

[Kizza, 2017] Kizza, J. M. (2017). *Ethical and social issues in the information age.* Texts in computer science. Cham, Switzerland : Springer, 2017.

[Leonardi and Bailey, 2008] Leonardi, P. M. and Bailey, D. E. (2008). Transformational Technologies and the Creation of New Work Practices: Making Implicit Knowledge Explicit in Task-Based Offshoring. *MIS Quarterly*, 32(2):411–436.

[Lins et al., 2017] Lins, K. V., Servaes, H., and Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *The Journal of Finance*, 72(4):1785–1824.

[Madden and Jones, 2008] Madden, M. and Jones, S. (2008). Networked Workers (PEW Report). Technical report.

[Martin and MacDonnell, 2012] Martin, B. H. and MacDonnell, R. (2012). Is telework effective for organizations? A meta-analysis of empirical research on perceptions of telework and organizational outcomes. *Management Research Review*, 35(7):602–616.

[Martínez-Sánchez et al., 2008] Martínez-Sánchez, A., Pérez-Pérez, M., Vela-Jiménez, M. J., and De-Luis-Carnicer, P. (2008). Telework adoption, change management, and firm performance. *Journal of Organizational Change Management*, 21(1):7–31.

[Meyer et al., 2001] Meyer, C. S., Mukerjee, S., and Sestero, A. (2001). Work-Family Benefits: Which Ones Maximize Profits? *Journal of Managerial Issues*, 13(1):28–44.

[Moen et al., 2011] Moen, P., Kelly, E. L., and Hill, R. (2011). Does Enhancing Work-Time Control and Flexibility Reduce Turnover? A Naturally Occurring Experiment. *Social Problems*, 58(1):69–98.

[Oettinger, 2011] Oettinger, G. S. (2011). The Incidence and Wage Consequences of Home-Based Work in the United States, 1980-2000. *Journal of Human Resources*, 46(2):237–260.

[Olson, 1982] Olson, M. (1982). *The Rise and Decline of Nations.* Yale University Press.

[Papanikolaou and Schmidt, 2020] Papanikolaou, D. and Schmidt, L. D. W. (2020). Working Remotely and the Supply-side Impact of Covid-19. *National Bureau of Economic Research Working Paper Series*, No. 27330.

[Ramelli and Wagner, 2020] Ramelli, S. and Wagner, A. F. (2020). Feverish Stock Price Reactions to Covid-19. *CEPR Working Paper No. DP14511.*

[Sánchez et al., 2007] Sánchez, A. M., Pérez, M. P., Carnicer, P. d. L., and Vela Jiménez, M. J. (2007). Teleworking and workplace flexibility: a study of impact on firm performance. *Personnel Review*, 36(1):42–64.

[Schivardi et al., 2021] Schivardi, F., Patnaik, M., Linarello, A., and Lamorgese, A. (2021). Management practices and resilience to shocks: Evidence from covid-19.

[Stavrou, 2005] Stavrou, E. T. (2005). Flexible work bundles and organizational competitiveness: a cross-national study of the European work context. *Journal of Organizational Behavior*, 26(8):923–947.

[Yang et al., 2021] Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., and Teevan, J. (2021). The effects of remote work on collaboration among information workers. *Nature Human Behaviour*.

# Chapter 3 - Bibliography

[Ablon et al., 2016] Ablon, L., Heaton, P., Lavery, D. C., and Romanosky, S. (2016). Consumer attitudes toward data breach notifications and loss of personal information. Technical report, RAND Corporation.

[Acquisti et al., 2006] Acquisti, A., Friedman, A., and Telang, R. (2006). Is There a Cost to Privacy Breaches? *ICIS 2006 Proceedings*, page 94.

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

[Angst et al., 2017] Angst, C. M., Block, E. S., D'Arcy, J., and 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.

[Athey et al., 2017] Athey, S., Catalini, C., and Tucker, C. (2017). The Digital Privacy Paradox: Small Money, Small Costs, Small Talk. *NBER Working Paper Series*.

[August et al., 2014] August, T., Niculescu, M. F., and Shin, H. (2014). Cloud implications on software network structure and security risks. *Information Systems Research*, 25(3):489–510.

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

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

[Baker and Fradkin, 2017] Baker, S. R. and 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.

[Bana et al., 2020] Bana, S., Brynjolfsson, E., Rock, D., and Steffen, S. (2020). job2vec: Learning a representation of jobs.

[Barrero et al., 2021] Barrero, J. M., Bloom, N., and Davis, S. J. (2021). Why working from home will stick. Technical report, National Bureau of Economic Research.

[Bolster et al., 2010] Bolster, P., Pantalone, C. H., and Trahan, E. A. (2010). Security breaches and firm value. *Journal of Business Valuation and Economic Loss Analysis*, 5(1).

[Bose and Leung, 2013] Bose, I. and Leung, A. C. M. (2013). The impact of adoption of identity theft countermeasures on firm value. *Decision Support Systems*, 55(3):753–763.

[Bose and Leung, 2019] Bose, I. and Leung, A. C. M. (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.

[Buckman et al., 2019] Buckman, J. R., Bockstedt, J. C., and Hashim, M. J. (2019). Relative privacy valuations under varying disclosure characteristics. *Information Systems Research*, 30(2):375–388.

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

[Choi et al., 2019] Choi, J. P., Jeon, D. S., and Kim, B. C. (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*, 173:113–124.

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

[Foerderer and Schuetz, 2022] Foerderer, J. and Schuetz, S. W. (2022). Data Breach Announcements and Stock Market Reactions: A Matter of Timing? *Management Science*.

[Garcia, 2013] Garcia, M. E. (2013). The economics of data breach: Asymmetric information and policy interventions.

[Garg, 2020] Garg, P. (2020). Cybersecurity breaches and cash holdings: Spillover effect. *Financial Management*, 49(2):503–519.

[Gordon and Loeb, 2006] Gordon, L. A. and Loeb, M. P. (2006). Budgeting process for information security expenditures. *Commun. ACM*, 49(1):121–125.

[Gordon et al., 2015a] Gordon, L. A., Loeb, M. P., Lucyshyn, W., and Zhou, L. (2015a). Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity*, 1(1):3–17.

[Gordon et al., 2015b] Gordon, L. A., Loeb, M. P., Lucyshyn, W., and 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.

[Gordon et al., 2010] Gordon, L. A., Loeb, M. P., and Sohail, T. (2010). Market Value of Voluntary Disclosures Concerning Information Security. *MIS Quarterly*, 34(3):567–A2.

[Havakhor et al., 2020] Havakhor, T., Rahman, M. S., and Zhang, T. (2020). Cybersecurity Investments and the Cost of Capital.

[Hilary et al., 2016] Hilary, G., Segal, B., and Zhang, M. H. (2016). Cyber-Risk Disclosure: Who Cares? *SSRN Electronic Journal*, pages 1–59.

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

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

[Hui et al., 2017] Hui, K.-L., Kim, S. H., and Wang, Q.-H. (2017). Cybercrime Deterrence and International Legislation: Evidence from Distributed Denial of Service Attacks. *MIS Quarterly*, 41(2):497–A11.

[Hui et al., 2018] Hui, K.-L., Vance, A., and Zhdanov, D. (2018). MIS Quarterly Research Curation on Securing Digital Assets. *MIS Quarterly Research Curations*.

[Janakiraman et al., 2018a] Janakiraman, R., Lim, J. H., and 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.

[Janakiraman et al., 2018b] Janakiraman, R., Lim, J. H., and 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.

[Kankanhalli et al., 2003] Kankanhalli, A., Teo, H. H., Tan, B. C., and Wei, K. K. (2003). An integrative study of information systems security effectiveness. *International Journal of Information Management*, 23(2):139–154.

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

[Lassébie et al., 2021] Lassébie, J., Marcolin, L., Vandeweyer, M., and 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 et al., 2020] Liu, C.-W., Huang, P., and Lucas Jr, H. C. (2020). Centralized it decision making and cybersecurity breaches: Evidence from us higher education institutions. *Journal of Management Information Systems*, 37(3):758–787.

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

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

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

[Petersen et al., 2020] Petersen, R., Santos, D., Smith, M. C., Wetzel, K. A., and Witte, G. (2020). Workforce Framework for Cybersecurity (NICE Framework). *NIST Special Publication - U.S. Department of Commerce*.

[Richardson et al., 2019] Richardson, V. J., Smith, R. E., and Watson, M. W. (2019). Much ado about nothing: The (lack of ) economic impact of data privacy breaches. *Journal of Information Systems*, 33(3):227–265.

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

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

[Samuelson, 2007] Samuelson (2007). Security breach notification laws: Views from chief security officers. Technical report.

[Say and Vasudeva, 2020] Say, G. and 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, 2007] Solove, D. J. (2007). "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review*, 44(4):745–772.

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

[Turjeman and Feinberg, 2019] Turjeman, D. and Feinberg, F. M. (2019). When the Data Are Out: Measuring Behavioral Changes Following a Data Breach. *Marketing Science*.