Trading on Talent: Human Capital and Firm Performance*

Anastassia Fedyk† and James Hodson‡

December 26, 2020

Abstract

How is skilled human capital reflected in firm performance? By directly observing the monthly career migration patterns of 37 million employees of US public companies, along with their education, demographics, and skills, we explore firm-level "skill premia." Our key empirical finding is that, contrary to the individual-level patterns documented by the labor economics literature, technical and social skillsets negatively forecast both financial and operational performance at the firm level. We explore several potential mechanisms for this finding. Social skillsets display counter-cyclical performance, suggesting that their negative premia reflect their risk profiles. Meanwhile, negative premia on technical skillsets show patterns consistent with over-exuberance regarding contemporaneous popular technologies: IT and Mobile Networks skillsets carry negative premia in the early 2000s, while Data Analysis, Software Engineering, and Web Development display negative premia during the 2010s.

Keywords: return predictability, asset prices, labor and finance, human capital, skilled labor, technology

*The authors are grateful to John Beshears, John Campbell, Lauren Cohen, Isaac Hacamo, Christopher Hrdlicka (discussant), David Laibson, Chris Malloy, Adrien Matray, Maria Semenova (discussant), Andrei Shleifer, Jeremy Stein, Annette Vissing-Jorgensen, Wei Wu (discussant), Michael Weber (discussant), and seminar participants at Arrowstreet Capital, Cubist Systematic, Harvard University, Norges Bank, PanAgora Asset Management, Washington University in St. Louis Olin, University of Utah Eccles, 8th International Moscow Finance Conference, Indiana Junior Finance Conference, UBS Quantitative Investment Conference, EFA 2020, ESSFM Gerzensee 2019, WFA 2019, and NFA 2018 for insightful comments. The authors thank Cognism Ltd. for providing the employment profiles data. This paper was awarded the 2018 Northern Finance Association Best Paper Award on Long Term Investments, the 2017 Jack Treynor Prize from the Q-Group (Institute for Quantitative Research in Finance), and the second place 2017 PanAgora Asset Management Dr. Richard A. Crowell Memorial Prize. This project received funding from the EU Research and Innovation programme Horizon 2020 under grant agreement No. 675044. This research received a "not human subject research" determination from the Institutional Review Board at Harvard University (protocol #IRB17-0443).

†UC Berkeley, Haas School of Business. Email: fedyk@berkeley.edu.

‡Jožef Stefan International Postgraduate School, Artificial Intelligence Laboratory and the AI for Good Foundation. Email: james.hodson@ijs.si. James Hodson concurrently servers as Chief Science Officer at Cognism Ltd.

1 Introduction

What are the returns to skilled human capital for firms? We document that skillsets that traditionally carry a wage premium at the individual level correspond to higher firm valuations but systematically lower future returns. We focus on a set of technical skillsets such as *Software Engineering* and *Data Analysis*, as well as communication skillsets such as *Client Relationship Management* and *Social Media*. Abnormal concentration of a firm's employees in these skillsets is associated with higher contemporaneous valuations but forecasts negative future returns, supported by dampened future earnings surprises. In order to explore potential mechanisms behind the negative skill premia, we consider differential performance of the technical and communication skillsets across different time periods and market conditions. Communication skillsets perform relatively better during recession months and during the five-year period surrounding the Great Recession (2006-2010), indicating that their negative returns may reflect a counter-cyclical risk profile. Technical skillsets display patterns consistent with overvaluation of contemporaneous popular technologies. The main technical skillsets from the early 2000s (*Information Technology* and *Mobile Networks*) see their negative premia (and higher contemporaneous firm valuations) concentrated in that time period, while more recently popular technical skillsets (*Data Analysis*, *Software Engineering*, and *Web Development*) display negative premia during the 2010s.

The increasing importance of human capital for modern firms has been stressed by practitioners and academics alike (Zingales, 2000). A number of studies consider the effects of senior management including Chief Executive Officers on corporate outcomes such as performance, strategy, capital structure, mergers and acquisitions, and investment decisions.[1] While this line of work demonstrates the importance of firm leadership, a firm's human capital extends far beyond its uppermost executives. Individual rank and file employees also drive the day-to-day operations of their employers. In order to explore the impact of these individual employees, we guide our analysis with findings from labor economics, which indicate that certain skills and characteristics – including leadership abilities (Kuhn and Weinberger, 2005), technical abilities (Ingram and Neumann, 2006), and social skills (Deming, 2017) – are important predictors of individual wages and career outcomes.[2] In this paper, we aggregate individual skill information up to the firm level and explore the question of whether "skill premia" translate from individual employees to their employer firms. If an abnormally high share of a firm's employees are skilled in technical or communication fields, does that predict

[1] See Bertrand and Schoar (2003), Malmendier and Tate (2005, 2008), Adams, Almeida, and Ferreira (2005), and Graham, Harvey, and Puri (2013), among others.

[2] For evidence on the rising importance of individual skills that are not directly measured by formal education and experience, see also Juhn, Murphy, and Pierce (1993).

higher valuations, superior operational performance, and positive future returns? We lay the foundation for investigating firm-level skill premia by documenting that the composition of a firm's broad employee base is an important factor in driving firm performance, in a way that is distinct from the labor economics evidence on individual outcomes.

A key challenge in conducting a comprehensive analysis of the impact of employee skills on firm performance has been the lack of employee-employer data with sufficient detail on individual employees' skills and backgrounds. In order to circumvent this challenge, we leverage techniques from machine learning on a large novel employer-employee dataset constructed from individual resumes. The dataset covers more than 490 million individuals globally, of which approximately 37 million represent employees of U.S. public companies since 1990. Each profile includes detailed information regarding employment, education, skills, geographic location, and demographic information (such as age, gender, and ethnicity). Importantly, although our analysis centers on U.S.-based firms, we are able to identify employees of specific firms across borders, providing a more complete view of each particular firm's aggregate human capital. We employ a mixture of textual analysis, existing knowledge-bases, and crowd-sourcing techniques to link the employer names appearing in the individual profiles to U.S. publicly traded companies. Merged with financial and accounting data, we obtain an unbalanced panel with a total of 3,094 firms and an average of 1,850 firms per month. We cover all of the public employer firms active with at least one hundred employees at the end of our sample. This offers significantly broader coverage than what has been available from prior employment profile datasets, including LinkedIn (see, for example, Jeffers, 2020).

In addition to its coverage, a key advantage of our employment dataset is the inclusion of individual employees' skills and abilities. However, the skills are self-reported, and each individual can specify his expertise in any manner he chooses, with no constraints. As a result, the 37 million profiles of employees of U.S. public firms feature more than 200 thousand unique strings denoting skills. In order to structure these self-reported skills into a limited number of easily interpretable skill*sets*, we employ a method from the topic modeling literature, Latent Dirichlet Allocation, and learn 44 latent topics (the skillsets) that would be most likely to generate the observed individually reported skills. These skillsets intuitively capture key areas of expertise in the modern workforce, including aspects such as *Business Development* (most common skills: *business strategy*, *marketing strategy*, *business development*) and *Software Engineering* (most common skills: *java, sql, software development*), among others. We then classify each individual as possessing two skillsets: we term the skillset likely to have generated the largest number of the individual's skills as his "primary skillset" and the next highest skillset as his "secondary skillset."

From the classification of individual employees into skillsets, we aggregate up to characterize the skill composition at the firm level. In particular, for each firm in each month, we compute the proportions of employees possessing each skillset. Throughout our analysis, we guide our focus on specific skillsets with prior evidence from the labor economics literature, which documents individual-level premia for technical and social skills (Ingram and Neumann, 2006; Deming, 2017). Specifically, we focus on the five generally applicable (rather than industry-specific) technical skillsets (*Data Analysis, Information Technology, Mobile Networks, Software Engineering,* and *Web Development*) and three communication skillsets (*Client Relationship Management, Digital Marketing,* and *Social Media*). For comparison, in our main return tests, we also consider core operational skillsets such as *Sales* and *Operations Management*. We leave out administrative and bureaucratic skillsets such as *Admin* and *Human Resources*, as well as industry-specific skillsets such as *Construction* and *Healthcare*.

Our main results show that firms with abnormal focus on technical and communication skillsets tend to have higher valuations but earn systematically lower subsequent returns. For example, an additional 10% of employees with *Software Engineering* as either their primary or secondary skillset corresponds to the firm having a 2.39% higher valuation as measured by Tobin's q but earning a 7 basis points lower return in the following month (corresponding to -0.72% annual alpha). Similarly, an additional 10% of the firm's employees being skilled in *Digital Marketing* corresponds to 0.53% higher Tobin's q and predicts 21 basis points lower monthly returns (translating to -2.55% annual alpha). Both the valuation and the return results are robust to including or excluding controls for industry and time fixed effects, firm size, book-to-market ratio, and past performance, consistent with the descriptive statistics on the employee skill variables, which show them to be largely unaffected by other firm characteristics. The return results are also not sensitive to the choice of risk model in calculating abnormal returns, with nearly identical estimates for raw excess returns, CAPM alphas, and Fama and French (1993) three-factor alphas. The negative return premia on technical and communication skillsets are present across a lengthy horizon, with comparable predictability one, two, three, or six months out, and across the major industrial sectors (manufacturing, information, and finance and insurance). Overall, this combination of results points towards technical and communication skills garnering higher contemporaneous valuations that subsequently reverse.

We further show that neither technical nor communication skillsets among employees enhance firm operations. Specifically, we analyze how changes in the skill composition of the firm's workforce relates to (i) changes in the firm's gross profitability and (ii) future earnings surprises. These tests are conducted at the quarterly frequency, and we estimate the relationship between quarterly changes in skilled labor and the impact on firm operations one, two, three, and four quarters later. Most of the technical and communication skillsets show zero relationship to firm profitability (a few exceptions are *Data Analysis* and *Information Technology*, which negatively forecast changes in profitability at some lags, and *Client Relationship Management*, which has a positive link to profitability at a one quarter lag). For standardized unexpected earnings, the impact of skilled labor is largely negative. For example, a one standard deviation increase in the inflow of *Data Analysis* employees predicts 1.49% lower standardized unexpected earnings in the next quarter and 2.89% lower earnings two quarters out, with comparable effects for skillsets such as *Software Engineering, Web Development*, and *Social Media*. The null effects on profitability and the negative effects on earnings surprises point towards the higher individual-level wage premia for employees with technical and communication skills increasing firms' personnel costs without improving operational efficiency.

We consider several potential mechanisms behind the observed negative firm-level skill premia. In order to test for whether our results reflect risk factors not captured by the firm-level controls and abnormal return calculations, we partition the sample based on market conditions. Using recession markers from the National Bureau of Economic Research, we find that all technical and communication skillsets perform poorly in non-recession months, but with evidence of better performance during recessions. The NBER recession sample is quite small, however, with only 26 months out of our 17-year period marked as recessions. To increase power, we augment the analysis using a coarser partition of time periods into 2000-2005, 2006-2010, and 2011-2016. We find that all three communication skillsets carry negative premia in both 2000-2005 and 2011-2016, but not in the five-year period around the Great Recession, 2006-2010, supporting the notion that communication skills may earn negative premia for their counter-cyclical risk profiles. Technical skillsets do not display an analogous pattern, instead reflecting more monotonic patterns over the three sub-periods.

Our final set of analyses supports the notion that the negative premia on technical skillsets reflect over-pricing of currently popular skills. First, we observe that each technical skillset is associated with higher firm-level valuations and negative subsequent returns precisely at the time when that skillset is especially popular. Specifically, the three technical skillsets that have emerged as popular during the recent years (*Data Analysis*, *Software Engineering*, and *Web Development*) all show strong positive relationships with valuations coupled with negative predictability for future returns since 2010, but not during the 2000-2005 or the 2006-2010 sub-periods. By contrast, *Information Technology* and *Mobile Networks*, which were popular emerging skillsets during the early 2000s, had the highest valuations followed by reversal (negative predictability for returns) precisely during that period (2000-2005), but not in recent years. Second, we use the limited reporting of wage bills by U.S. public firms to provide suggestive evidence that higher emphasis on technical skillsets is associated with higher wage expenses for firms – despite a lack of any positive impact of these skillsets on either profitability or earnings. Higher firm-level shares of employees with skills in*Data Analysis, Software Engineering,* and *Web Development* all correspond to large and statistically significant increases in the firm-level wage bill. Lastly, we bring additional data on job openings and college graduates to highlight the mismatch between the supply and demand of currently popular employee skills. Specifically, job postings in the Software Development area (which includes occupations ranging from Software Engineer to Web Developer) increased by over 200% between 2010 and 2016. However, the number of individuals receiving bachelors degrees in Computer Science or Information Science has risen by a modest 50% since 2010 (and are close to the same level as in 2005). Altogether, our combination of results is most consistent with excess demand leading to over-pricing of the currently popular technical skills.

Our paper contributes to the growing literature on the relationship between corporate outcomes and human capital. Most of the prior work looks at the human capital of top executives, including the role of CEO age and overconfidence.[3] Recent studies expand the focus beyond the chief executive, exploring survey indicators of corporate culture and employee satisfaction (Edmans, 2011; Guiso, Sapienza, and Zingales, 2015), non-compete agreements,[4] hiring and departures,[5] employee job functions (Rock, 2020), and demographics such as age and education (Mukharlyamov, 2016; Kilic, 2016). We open up this line of work to a detailed characterization of a given firm's employees in terms of their skills and abilities. By leveraging big data techniques on a large dataset with detailed information on millions of individual employees of U.S. public companies, we can characterize a firm's human capital on a large-scale basis (including overseas employees) and yet with very granular detail (taking advantage of specific individually-provided skills).

The ensuing findings link the literature on firm performance to the labor economics literature on skill premia. Specifically, our results indicate that the previously documented skill premia at the individual level do not necessarily aggregate up to the level of the firm. Rather, at the firm level, contemporaneously popular skills can be associated with high initial valuations but negative return premia in the future, suggesting an overpricing of skills that are in high demand but that do not necessarily translate into operational gains. Our results

[3]See, for example, Bertrand and Schoar (2003), Malmendier and Tate (2005), Galasso and Simcoe (2011), Kaplan, Klebanov, and Sorensen (2012), Yim (2013), and Benmelech and Frydman (2015).

[4]See, for example, Starr, Balasubramanian, and Sakakibara (2018) and Jeffers (2020).

[5]See Belo, Lin, and Bazdresch (2014), Belo, Li, Lin, and Zhao (2017), and Agrawal, Hacamo, and Hu (2020).

highlight the potential for over-exuberance about particular technologies to distort the valuation of firms' human capital, which complements contemporaneous work by Hombert and Matray (2020) documenting overvaluation (and subsequent poor performance) of technical skills during the tech boom from the perspective of individual employees.

The remainder of the paper proceeds as follows. Section 2 describes the large unstructured employment dataset, outlines the steps taken to structure the records and link firms to employees, and offers an overview of the broad labor mobility patterns visible in the data. Section 3 outlines the methodology for classifying self-reported skills into meaningful skillsets. Section 4 presents our main empirical findings: the negative firm-level return premia to technical and communication skillsets, coupled with higher current valuations but zero or negative effects on operational performance. Section 5 explores potential mechanisms behind our results, and Section 6 concludes.

2 Data

We introduce the detailed resume data provided by Cognism, a global Client Relationship Management platform, outline the steps we take to structure this dataset into matched employer-employee data, and present summary statistics. We also outline additional datasets, including job postings, which we match to firm-level measures from the Cognism data in subsequent analyses.

2.1 Individual Profile Data

We directly observe firms' human capital through the lens of a novel dataset of approximately 490 million individual employment and education records. The data provider, Cognism, pulls together partner Client Relationship Management databases, private feeds, and publicly available data from a variety of sources to maintain up-to-date information on the education, career, and notable events of individuals globally. Cognism's infrastructure ensures that records are independently verified and researched on a regular basis. Every profile in the starting dataset has at least one active work experience record or education record between 1980 and 2016, and the data include a broad sweep of job types from top executives to rural agricultural positions and factory workers.

For each individual in our sample, the data include a unique identifier, city and country level location, and an approximate age derived from the individual's education history, where available. In addition, we observe the individual's education and employment history, as well as a set of self-reported skills. We remove noisy profiles (profiles with improbable dates, too many empty fields, etc.) from the analysis, leaving approximately 370 million profiles. Table 1 provides summary statistics of the demographic, educational, and employment characteristics of these 370 million profiles. Panel 1 shows the breakdown of profiles across geographic regions, with the majority of employees based in the United States. Panel 2 reveals that the average age of the employees in the data is 36 years old, that conditional on having any reported employment, the average (median) individual lists 3.6 (2) jobs, and that conditional on listing a non-empty set of skills, the average (median) individual reports 10 (4) skills. The descriptive statistics presented in Panel 2 for age, number of jobs, and number of listed skills are all winsorized at the top and bottom 0.1%. Panel 3 gives the breakdown by education for employees who specify at least one education experience; the majority of educated employees hold bachelor's degrees.

In order to be able to link human capital to measures of firm operations, valuations, and returns, we concentrate on individuals who have been employed by at least one U.S. publicly traded company (i.e., a company listed on the NYSE, NASDAQ, or AMEX). Linking Cognism resumes to publicly traded companies is nontrivial, since the raw data contain a number of inconsistencies stemming from the diversity of unconstrained user input. As such, we take several steps to normalize and link the stated employer names to official trading names and relevant stock symbols for public companies. The procedure is detailed in Appendix B.

The end result is a sample of approximately 37 million U.S. public-company employees with employment information spanning from 1990 to 2016. These employees provide us with more than 100 million observable employment transitions and span approximately 39% of the total employment of U.S. public firms at the end of the sample in 2016, including foreign-based workers employed by U.S. firms.[6] At the firm level, we require industry classification, shares outstanding, revenues, cost of goods sold, assets, earnings, and book value variables in Compustat and returns and market capitalization in Center for Research in Security Prices (CRSP), and we further exclude firms with fewer than 100 employees in Cognism to avoid noise from poor coverage. This yields an unbalanced panel with an average of 1,850 firms per month (from 1,175 firms at the start of the sample to 2,820 in 2016) and a total of 3,094 firms.

Table 2 gives a rough breakdown of the employee data across 2-digit NAICS industry sectors for three snapshots: as of 1996, 2006, and 2016. The distribution of employment across sectors is consistent with the breakdown of overall employment by U.S. public firms reported in Compustat. Overall, we observe a number of industry shifts, including shrinking

[6]Per Panel 1 of Table 1, coverage of international workers is lower than that of U.S.-based workers. Cognism's coverage of U.S.-based workers of U.S. public firms is approximately 67%.

of the Manufacturing sector and growth in the Retail Trade sector. Figures 1 and 2 show moves within and across industry as a proportion of employment change events in the 2010's and 1990's, respectively. The inter-industry moves captured in the figures highlight the importance of human capital to our understanding of firms. The majority of job change events, over 60%, occur outside of traditional industry lines as delineated by the broad two-digit NAICS codes. This is especially true for versatile industries such as Information and Public Administration. In Appendix A.2, we also show that employee turnover is a meaningful predictor of performance at the level of individual firms: firms with higher monthly turnover experience consistently lower stock price returns up to six months later.

2.2 Additional Datasets

In additional analyses, we employ three other datasets. We take the standard recession classification from the National Bureau of Economic Research. College major numbers come from the Integrated Postsecondary Education Data System, which is a system of interrelated surveys conducted annually by the National Center for Education Statistics (NCES). We access the data through the NCES Digest of Education Statistics. We also take advantage of a novel dataset on job postings from Burning Glass Technologies Inc., which provides comprehensive coverage of online job openings based in the U.S. from over 40,000 sources including individual company websites and online job search websites.[7] We work with the job opening numbers aggregated on a weekly basis. These data come with normalized occupations grouped into broad occupation groups and high-level career paths, as well as a large taxonomy of associated skills. The Burning Glass data span a more limited time range, from 2010 to present, a period during which job openings posted online have been fairly representative of the overall labor demand in the U.S. economy (Liu and Wu, 2019), covering approximately 60-70% of all job vacancies (including both online and offline postings) in the United States (Carnevale, Jayasundera, and Repnikov, 2014).

3 Methodology: Skilled Human Capital

In order to evaluate the relationship between a firm's performance and the composition of its human capital, we identify the key skillsets of each of the 37 million public company employees in our sample. This process occurs in several steps. First, we use techniques from machine learning to condense hundreds of thousands of employee-provided skills into a

[7]This dataset is available to us through James Hodson's affiliation with Cognism Ltd., which is engaged in an ongoing restricted partnership with Burning Glass Technologies.

manageable number of latent skillsets. Second, we classify each employee as possessing two skillsets – a primary and a secondary one – based on which of the identified skillsets best fit that employee's self-reported skills. Lastly, we aggregate the employee-level classification to the firm-level to assess firm-level skill premia.

3.1 Identifying Skillsets

We use the Latent Dirichlet Allocation method for structuring topics in documents to identify common skillsets from the individuals' self-reported lists of skills.[8] We recover forty-four skillsets, which capture intuitive competencies ranging from product management to web development to healthcare professionals.

In order to apply the topic modeling methodology, we represent the skills listed on each individual's profile as a set of terms from an overarching vocabulary of skills. Let D denote the set of individual profiles. Each element $d \in D$ represents one individual's set of skills, as reported on his or her profile. For example, an employee may list the set of skills $d = \{\textit{Microsoft Office}, \textit{Public Speaking}\}$. Let the vocabulary W consist of all skills that appear in at least one profile; for example, if there is a profile $d = \{\textit{Microsoft Office}, \textit{Public Speaking}\}$ in the dataset, then *Microsoft Office* will appear in W, as will *Public Speaking*. We represent all elements of D as unit vectors in the $|W|$-dimensional space of vocabulary terms (where $|W|$ is the size of the overall vocabulary of skills W). The premise of the Latent Dirichlet Allocation method is that the observed profiles in D are generated from a latent set of skillsets, which we denote by T. Each skillset $t \in T$ is a probability distribution over the terms in W. The output of the Latent Dirichlet Allocation model consists of (i) the individual skillsets (probability distributions over words) and (ii) a probability distribution over skillsets for each employee. In this paper, we consider $k = 44$ skillsets.[9] We arrive at this number empirically by considering goodness-of-fit measures across values of $k \in [25, 60]$.

The skillsets identified by the model, presented in Table 3, map intuitively onto common work skills; we label them for ease of exposition. For example, the *Web Development* skillset features *javacript, html,* and *java* as the most likely skills (i.e., the skillset *Web Development* is a probability distribution over skills that puts the highest weight on the skill *javascript*, followed by *html*, followed by *java*). On the other hand, the most likely skills in the *Personal Coaching* skillset are *coaching, public speaking,* and *sports.* While many of the skillsets – such as *Hospitality*, *Military*, and *Healthcare* – are relevant only in specific segments of

[8]For further details on Latent Dirichlet Allocation method, see Blei, Ng, and Jordan (2003). The Online Appendix in Fedyk (2020) provides a short summary of the method.

[9]More precisely, we train the final model with $k = 50$ skillsets and then group together skillsets that are duplicates of each other in foreign languages.

the economy, others – including *Sales*, *Information Technology*, and *Client Relationship Management* – are applicable across the broad spectrum of economic activity. We focus our analysis exclusively on the latter type of skillets.

To guide our analysis, we focus on the skillsets that have been highlighted as carrying wage and career premia at the individual level: technical skillsets and social/communication skillsets. For comparison, we also consider core operational skillsets such as *Sales* and *Operations Management*. The remaining excluded category consists of bureaucratic skillsets such as *Human Resources*, *Admin*, and *Middle Management* using generic tools such as the Microsoft Office suite and highly sector-specific skillsets such as *Non-Profit*, *Construction*, and *Healthcare*.

The full list of skillsets, grouped into the outlined categories, is as follows:

- Technical skillsets: *Data Analysis; Information Technology; Mobile Networks; Software Engineering; Web Development.*

- Communication skillsets: *Client Relationship Management; Digital Marketing; Social Media.*

- Operational skillsets: *Industrial Management; Logistics; Operations Management; Sales; Sales Management; Technical Product Management.*

- Administrative skillsets: *Admin; Business Development; Human Resources (Jr.); Human Resources (Sr.); Middle Management; Product Management (Generic); Recruiting.*

- Sector-specific skillsets: *Accounting & Auditing; Banking & Finance; Construction; Education; Electrical Engineering; Graphic Design; Healthcare; Hospitality; Insurance; Legal; Manufacturing; Military; Musical Production; Non-Profit; Oil, Energy & Gas; Pharmaceutical; Public Policy; Real Estate; Retail; Video & Film; Visual Design; Web Design; Personal Coaching.*

3.2 Classification of Individual Employees

Having identified the skillsets, we classify each employee as possessing the two skillsets that are most likely to have generated his or her particular combination of self-reported skills. Specifically, for each individual i with a non-empty set of self-reported skills, we take the estimated distribution over skillsets and let j and j' denote the indices of the two largest elements in this probability distribution. Then individual i is deemed to possess the skills from the skillsets j (his primary skillset) and j' (his secondary skillset). That is, each individual is assigned to the two skillsets that best match his or her profile.

The breakdown of primary and secondary skills across the full sample of 37 million employees of public U.S. companies is displayed in Figure 3, which shows that the most common skillsets include *Business Developmen* and *Middle Management*. *Education, Legal*, and *Personal Coaching* skillsets are the least common in our sample of public firm employees. The skillsets on which we focus our analysis, including technical skills such as *IT* and *Software Engineering* and communication skills such as *Digital Marketing*, tend to fall in the middle range, with sizable but not ubiquitous presence across public-firm employees.

There is substantial heterogeneity in skills across companies in different industries. Figures 4 and 5 provide the breakdown of employee skillsets across two example industries, Manufacturing (two-digit NAICS codes 31, 32, and 33) and Finance & Insurance (two-digit NAICS code 52). Manufacturing tilts heavily towards the sector-specific skillsets *Electrical Engineering* and *Manufacturing*. In contrast, the Finance & Insurance industry leans heavily on sector-specific skillsets *Banking & Finance* and *Accounting & Auditing*, as well as the more generic skillsets *Admin* and *Business Development*. Since different industries feature different compositions of skilled human capital, it may be the case that skill premia differ systematically across sectors. We confirm that this is not the case, and that our results are consistent across the major industry sectors, in Appendix A.3.

3.3 Firm-Level Skill Measures

We aggregate the individual employees' skillsets into firm-level measures of skilled human capital by computing, for each firm in each month, the percentage of employees skilled in a given area. Specifically, for each skillset j, firm i, and month t, we construct the measure $Skill^j_{i,t}$ as the number of employees who are employed at firm i during month t and for whom the skillset j is either primary or secondary, scaled by the total number of employees at the firm during that month, $N_{i,t}$. For the denominator $N_{i,t}$, we take the number of firm i's employees in the Cognism data with non-empty self-reported skill fields, rather than total employment numbers from Compustat, in order to avoid our skill measures being biased by differential data coverage across firms. We winsorize each skill focus variable $Skill^j_{i,t}$ at the top and bottom 1% across all firm-months.

Panel 1 of Table 4 presents the descriptive statistics of the firm-level skill measures for the set of skills in which we are interested: those in the technical and communication categories. The mean skill focus ranges from 1.3% of employees (*Data Analysis*) to 4.2% of employees (*Client Relationship Management*). Most of the skill variables are right-skewed, with minimum values of 0 for all skillsets and maximum values reaching as high as 48.7% for *Mobile Networks* and 40.7% for *Client Relationship Mangement*. Panel 2 presents the cross-correlations between the firm-level skill measures. The technical skillsets are positively correlated with each other, except for *Data Analysis*, which is not strongly related to any of the other skillsets. *Client Relationship Management* is positively correlated with technical skillsets (other than *Data Analysis*) and with *Digital Marketing*, which in turn is strongly positively correlated with *Social Media*.

Throughout our empirical analysis, we control for other factors known to predict firm performance. As such, for expositional clarity, Panel 3 of Table 4 presents the correlations between firm-level skill measures orthogonalized to these controls. Specifically, for each skillset j, we define abnormal prevalence as the residual from the regression of $Skill^j_{i,t}$ on firm-level controls:

$$Skill^j_{i,t} = \alpha + \gamma X_{i,t} + \epsilon_{i,t}, \tag{1}$$

where the vector of controls $X_{i,t}$ includes two-digit NAICS industry and month fixed effects, as well as firm size (log market capitalization), book-to-market ratio, and past performance (computed as the return for firm i from month $t-12$ to month $t-1$). Abnormal focus on skillset j is then defined as the residuals from the specification (1).

Encouragingly, the cross correlations between abnormal skill measures in Panel 3 are very similar to those between the raw skill measures in Panel 2. This helps assuage potential concerns that firms' endogenous choice to hire certain types of skilled labor reflects known firm characteristics such as industry, size, growth stage, and past performance, and we observe similarly consistent results across raw skill measures and regressions with controls throughout our empirical analyses.

4 Employee Skills and Firm Performance

In this section, we explore firm-level returns to employee skills. We begin by estimating the link between the composition of employee skillsets and firm valuations, and we then evaluate how employee skills predict firms' future returns. We find that skilled employees, especially for the technical skillsets, tend to garner higher contemporaneous valuations for the firms. But these high valuations subsequently reverse, with both technical and communication skillsets negatively predicting future returns. The negative future performance is also visible in firms' operations, as neither technical nor communication skillsets increase firms' profitability, and both technical and communication skillsets are actually associated with lower future earnings surprises.

4.1 Firm Valuations

We estimate firm valuations using Tobin's q following Chung and Pruitt (1994), as market value of the firm's equity and liabilities scaled by the book value of the firm's equity and liabilities. We link valuations to two groups of skills: (i) technical skillsets, following the evidence on the individual-level premia for technical and mathematical skills; and (ii) communication skillsets, prompted by the individual-level premia on communication skillsets.

Column 1 of Table 5 reports the results from estimating the following specification using Fama and MacBeth (1973) regressions for each technical and communication skillset:

$$Tobins\_q_{i,t} = \alpha + \beta Skill^{j}_{i,t} + X_{i,t} + \epsilon_{i,t}, \tag{2}$$

where the controls $X_{i,t}$ include industry fixed effects at the 2-digit NAICS level, size (log market capitalization), book-to-market ratio, and recent performance (stock return measured from month $t-12$ to month $t-1$). The results in column 2 exclude these controls. All results in Table 5 are reported per 10% change in the percentage of employees possessing a given skillset as either their primary or secondary skillset. Recall that every employee is classified as having two skillsets, so this change is computed relative to a total of 200%.

Overall, the valuation results in Table 5 indicate that the majority of the technical skills (*Data Analysis, Software Engineering*, and *Web Development*) correspond to higher firm valuations, consistent with prior evidence on the valuation of employees in firms' Engineering, IT, and Research departments (Rock, 2020). For example, an additional 10% of employees with *Data Analysis* as a primary or secondary skillset corresponds to a highly statistically significant 2.39% increase in Tobin's q. The only technical skillsets that do not display a positive relationship with valuations in the full sample are *Information Technology* and *Mobile Networks*, and as we show in Section 5, both of these skillsets do associate with higher valuations at the time of their relevance, namely in the early 2000s. Among communication skillsets, we observe a significant positive relationship between valuations and *Client Relationship Management* and *Digital Marketing* (although the latter is significant only without controls). The one communication skillset that does not display a link to valuations is *Social Media*.

Importantly, the results are very consistent across columns 1 and 2 of Table 5, with the coefficients on both technical and communication skillsets virtually unchanged by the inclusion of controls. This helps to confirm that the link between employee skills and valuations is not driven by firm size, growth stage, or recent performance.

4.2 Firm Returns

Although higher shares of employees skilled in technical and communication areas tend to garner higher valuations for the firms, we find that these skillsets are associated with negative abnormal returns going forward.

To estimate the firm-level return premium to focusing on a given skillset j, we regress abnormal returns on lagged values of the share of employees with j as either their primary or secondary skillset:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}, \tag{3}$$

where $AbnRet_{i,t}$ is the abnormal return for firm i during month t. We use Fama and French (1993) three factor alphas for abnormal returns in the main specification. In columns 1 and 2 of Table A.1, we confirm that our findings are robust to alternative return constructions using raw excess returns (i.e., firm returns minus the weighted average market return in the same month) and CAPM alphas. As before, the controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, and past performance, and we show that the results are qualitatively similar, albeit somewhat stronger, with the exclusion of all controls in column 3 of Table A.1. For the main specification, we vary the lag L between the measurement of employee skill composition and future firm returns between one month and six months.

The results of the main return specification are reported in Table 6. Analogously to Table 5, all estimates are reported per 10% change in the percentage of employees possessing a given skillset as either their primary or secondary skillset. The results for the four different lags (looking at returns one month, two months, three months, and six months after the measurement of employee skills) are displayed across the four columns. Consistent with the composition of the workforce changing relatively slowly, the results are fairly similar across the lags.

Our key finding is that both technical skillsets and communication skillsets are associated with negative future returns. Directionally, the relationship between abnormal skill focus and future returns is negative for all six technical skillsets and all three communication skillsets, at all four considered lags. Statistically, the effect is significant, for at least one of the lags, at the 1% for *Mobile Networks* and *Digital Marketing*, at the 5% level for *Software Engineering* and *Social Media*, and at the 10% level for *Web Development* and *Client Relationship Management*. The negative relationship between future firm performance and abnormal focus on technical and communication skills is also economically meaningful. For example, for every additional 10% of employees skilled in *Web Development* as either their primary or secondary skillset, the firm experiences an average of 0.12% lower return in the following month, or a 1.45% lower return per year. Similarly, for every additional 10% of employees skilled in *Digital Marketing*, the firm sees a 0.21% (2.55%) lower monthly (annual) alpha. To the extent that employees with technical and communication skills may be more mobile than other employees, and labor mobility risk is priced in returns (Donangelo, 2014), the finding of *negative* return premia on these skillsets is even more striking.

Overall, the combination of results in Tables 5 and 6 suggests that technical and communication skills among employees garner higher valuations for their employers, but that these valuations subsequently reverse. By contrast, the skills that may be undervalued and that display positive return predictability are operational skillsets that leverage a less concentrated mix of technical and communication abilities to accomplish core business activities. We consider this group of skills as a comparison set, reported at the end of Table 6. *Sales*, *Operations Management*, *Technical Product Management*, and, to a lesser extent, *Industrial Management* all have positive and statistically significant relationships with future returns. The effects of *Sales Management* and *Logistics* are also positive, although not statistically significant.

In Appendix A.3, we consider the heterogeneity of the observed negative premia on technical and communication skillsets across industries, focusing on the three largest sectors (Manufacturing, Finance & Insurance, and Information). These additional results suggest that the negative skill premia are ubiquitous across the sectors of the economy, with some heterogeneity in exactly which technical and communication skillsets predict more significant negative returns in which industry. In the next subsection, we show that the negative return premia on both technical and communication skillsets are also accompanied by generally poorer operational performance.

4.3 Employee Skillsets and Firm Operations

The results so far suggest that employees with technical and communication skillsets are associated with higher firm valuations that subsequently reverse. We now provide evidence that the high valuations of these skillsets are not due to improvements in operational performance. Hiring of employees with technical and communication skills is associated with largely insignificant changes in firm profitability and mostly *negative* future earnings surprises.

We conduct the analysis of firm operations at the quarterly level using quarterly accounting variables. We focus on changes in gross profitability from quarter to quarter and the component of quarterly earnings that is unexpected relative to past earnings announcements. Correspondingly, we transform the employee skill measures into quarterly changes in the share of employees with a focus on each skillset. Specifically, we define $\Delta Skill^{j}_{i,t}$ for firm i and skill j as the percentage change, over the course of quarter t, in employees of firm i whose primary or secondary skillset is j, winsorized at the top and bottom 1% and standardized to mean zero and standard deviation one.

We begin the operational analysis by considering the relationship between firm-level increases in technical and communication skills and future firm profitability. Specifically, we estimate the following predictive regression:

$$\text{For } L \in \{1\,q, 2\,q, 3\,q, 4\,q\} : \Delta Profitability_{i,t} = \alpha + \beta \Delta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}, \qquad (4)$$

where $\Delta Profitability_{i,t}$ is the (percentage) change in firm i's profitability (defined as the difference between revenues and costs of goods sold, scaled by assets) from quarter $t-1$ to quarter t. We predict changes in profitability with lagged changes in employee skills using lags between one and four quarters. The controls $X_{i,t}$ include industry and quarter fixed effects, firm size, book-to-market ratio, and past performance, and in untabulated analyses we confirm that these results, too, are not sensitive to the exclusion of controls.

Table 7 presents the results for the four quarter lags across the four columns. For the majority of skillsets, there is no discernible relationship between increases in employees with technical or communication skills and future profitability. However, two of the technical skillsets display negative predictability for future profitability at some of the lags: a one standard deviation increase in employees skilled in *Data Analysis* translates into a 13% decrease in the gross profitability ratio four quarters later, and a one standard deviation increase in *Information Technology* employees corresponds to a 9% decrease in profitability two quarters out. By contrast, one of the communication skillsets (*Client Relationship Management*) has a significant positive relationship with gross profitability at a one quarter lag. The remaining skillsets, however, all show fairly precise zero effects on profitability.

We next turn to the impact of employees' technical and communication skills on firms' earnings surprises. In order to link skilled human capital to future earnings surprises, we look at standardized unexpected earnings (SUE). Following the literature (Da, Gurun, and Warachka, 2014; So and Wang, 2014), $SUE_{i,t}$ is defined as the difference in firm i's earnings per share in the current quarter t from the earnings per share four quarters ago ($t-4$), scaled by the standard deviation of this difference computed over the past eight quarters ($t-8$ through $t-1$). We estimate the relationship between quarterly changes in the percentage of employees skilled in each of the five technical and three communication skillsets and future quarterly unexpected earnings using the following specification:

$$\text{For } L \in \{1\,q, 2\,q, 3\,q, 4\,q\} : SUE_{i,t} = \alpha + \beta \Delta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}, \tag{5}$$

where $\Delta Skill^{j}_{i,t}$ is the change in the percentage of employees skilled in j at firm i from quarter $t-1$ to quarter t, winzorized at the top and bottom 1% and normalized to mean zero and standard deviation one. We consider lags between one and four quarters and control for industry and time fixed effects, firm size, book-to-market ratio, and past performance. Untabulated analyses confirm that the results are robust to the exclusion of these controls, as well as to the inclusion of an additional control for gross profitability.

The results, displayed in Table 8, indicate that most of the technical and communication skillsets are associated with negative future earnings surprises. Three of the five technical skillsets (*Data Analysis, Software Engineering,* and *Web Development*) and two of the three communication skillsets (*Digital Marketing* and *Social Media*) are associated with statistically significantly negative standardized unexpected earnings at least at one lag. For example, a one standard deviation increase in employees skilled in *Data Analysis* corresponds to 1.49% lower SUE one quarter later, 2.89% lower SUE two quarters later, and 1.51% lower SUE three quarters out. Only two skillsets, one technical (*Mobile Networks*) and one communication (*Client Relationship Management*) display a positive relationship with future standardized unexpected earnings.

Overall, the operational results show that employees with technical and communication skillsets, who are generally associated with higher firm valuations but subsequent return reversals, are *not* associated with better operational performance. The effect of skilled employees on profitability is largely zero and the effect on future standardized unexpected earnings is, if anything, negative. In the next section, we discuss potential mechanisms for this combined set of findings.

5 Mechanisms

We explore potential mechanisms for negative firm-level premia on technical and communication skillsets through three additional analyses: (i) partitioning the sample based on market conditions (U.S. recessions), (ii) observing time trends over the course of our seventeen-year sample period, and (iii) bringing supplementary evidence on wage expenses and job openings. The analysis across market conditions supports the notion of communication skillsets capturing counter-cyclical performance, with the lower returns reflecting a potential reduction in risk. The negative premia on technical skillsets, by contrast, are likely driven by over-excitement about emerging technologies: each technical skillset has the strongest pattern of high valuations and predictably negative returns precisely during the time when that technology is most in demand by employers. In-demand technical skills are also associated with higher wage bills and a mismatch between firms' demand and the supply of qualified graduates with corresponding expertise.

5.1 Skill Premia during Recessions

In order to classify our sample period into recession and non-recession market conditions, we use the business cycle recession indicators provided by the National Bureau of Economic Research (NBER). A total of 26 months out of the 204 months in our sample period (January 2000 through December 2016) are classified as recession months by the NBER.

Figure 6 displays the coefficient estimates from the relationship between abnormal skill focus and future returns (estimated through specification (3) with controls for industry and time fixed effects, firm size, book-to-market ratio, and recent performance) separately during recession and non-recession months. Panel 1 presents the results for the six technical skillsets, while Panel 2 displays the communication skillset results. For ease of exposition, coefficient estimates significant at the minimum of 10% level are highlighted in red (when negative) and blue (when positive); estimates that are not significant even at the 10% level are displayed as gray bars. Standard error bars are marked accordingly. The regressions are run incorporating one month lag between the calculation of abnormal skill focus and the subsequent Fama and French (1993) three factor alpha.

Both technical and communication skillsets carry robustly negative and significant return premia in good economic conditions (non-recessions). The estimates range from -9 basis points per month (*Information Technology*) to -26 basis points per month (*Web Development*) per 10% additional employees with a specific technical skillset as either their primary or secondary skillset. Similarly, next-months returns are 7 basis points lower for every additional 10% of employees skilled in *Client Relationship Management*, 23 basis points lower for every additional 10% of employees skilled in *Digital Marketing*, and 29 basis points per month lower for every 10% of employees skilled in *Social Media*. The results restricted to the recession months are less clear, due to the much smaller sample size (fewer than 13% of all months in our sample are classified as recession months by the NBER). Nonetheless, two of the considered skillsets, *Information Technology* and *Client Relationship Management*, are statistically significantly *positively* associated with future returns during poor market conditions.

We interpret the results sliced by market conditions as suggestive evidence that some of the negative premia on technical and communication skillsets may be driven by their risk profiles – specifically, skillsets such as *Client Relationship Management* may serve as a hedge for the firm's activities, with counter-cyclical performance. We explore this idea further with a coarser sub-sample analysis aimed at increasing power. Specifically, we consider broader five-year time windows of 2000-2005, 2006-2010, and 2011-2016, to see whether the returns on specific skillsets are different in the sub-period around the Great Recession (2006-2010) than in the preceding or following sub-periods.

The results on the bottom of Table 9 support the idea of communication skillsets capturing somewhat counter-cyclical risk profiles. All three of the communication skillsets display significant negative associations with future returns both in the 2000-2005 and the 2011-2016 sub-periods. However, all three of these skillsets show zero or weakly positive predictability for returns during the five-year sub-period around the Great Recession, 2006-2010. This counter-cyclical pattern is not observed for the technical skillsets, which instead display time trends consistent with the corresponding popularity of each technology – which we explore further in the next subsection.

5.2 Time Trends and Technical Skill Popularity

We take a closer look at whether the high valuations and subsequent predictable negative returns associated with technical skillsets reflect over-popularity of emerging technologies. At the firm level, the anticipated value of technical skillsets stems from streamlining more efficient product lines, automating repetitive and resource-intensive tasks, and providing a better understanding of operational efficiencies through software, data collection, and modeling. For example, in recent years, Data Science and Artificial Intelligence have been broadly praised as the future of business, with an implicit expectation that firms should increase their capabilities in these domains (Bughin, Seong, Manyika, Chui, and Joshi, 2018). In preceding decades, other technical skillsets were similarly touted as being critical to changing the economic landscape.

Given the combination of the firm-level push to hire certain technical skillsets and individual-level wage premia for these skillsets, we conjecture that the negative firm-level return premia are driven in part by inefficient over-investment in (and over-pricing of) these employees. To test this hypothesis, we break our sixteen-year sample into sub-periods and observe whether the negative relationship between focus on a specific technical skillset and returns is stronger at the times when that particular skillset is in greater demand.

Among the five technical skillsets, two (*Information Technology* and *Mobile Networks*) were relatively novel and in high demand at the beginning of our sample period, in early

2000s. The remaining three (*Data Analysis*, *Software Engineering*, and *Web Development*) emerged in great demand in the recent years, 2010s. We consider three sub-periods – 2000 to 2005, 2006 to 2010, and 2011 to 2016 – and reestimate specifications (2) and (3) within each of these sub-periods. The results, displayed in Tables 9 and 10, confirm our prediction that each technical skillset translates into higher valuations followed by reversal precisely during the time of its popularity. Table 10 shows the valuation results at five-year breakpoints (2001, 2006, 2011, and 2016), while Table 9 displays the return premia over 2000-2005, 2006-2010, and 2011-2016 subperiods.

Both *Information Technology* and *Mobile Networks* commanded high valuations followed by poor future returns in 2000-2005, during the time when these skillsets were emerging as popular technologies. An additional 10% of employees having *Information Technology* as their primary or secondary skillset corresponded to 8.70% higher Tobin's q in 2001 and predicted 23 basis points (2.80%) lower monthly (annual) returns during the 2000-2005 period, with the valuation result significant at the 1% level and the negative return premium significant at the 5% level. The effect sizes for employees skilled in *Mobile Networks* are similar: an additional 10% of employees with this skill were associated with 5.06% higher Tobin's q in 2001 and 20 basis points (2.43%) lower returns per month (per year), both significant at the 1% level. Since then, the link between these two skillsets and returns has weakened substantially, approaching zero in the recent years and even somewhat lower Tobin's q for firms with a higher focus on *Mobile Networks*.

By contrast, the currently popular skillsets *Data Analysis, Software Engineering*, and *Web Development* display the opposite pattern: these skillsets had insignificant (and even sometimes directionally positive) return premia during the two earlier samples, 2000-2005 and 2006-2010.However, during the time when these skillsets became especially popular, the 2010s, they began to carry higher valuations coupled with significant negative premia. For example, an additional 10% of employees having *Web Development* as their primary or secondary skillset corresponds to 2.23% higher valuations in 2011 and 25 basis points lower monthly returns (corresponding to -3.04% annual alpha) during 2011-2016, both significant at the 1% level. The effect sizes for *Software Engineering* and *Data Analysis* are milder but still sizable and statistically significant, with negative premia concentrated exclusively in the 2011-2016 sub-period.

5.3 Supply and Demand of Popular Skills

We look at two additional aspects to support our interpretation of the negative premia on technical skills reflecting over-exuberance about popular skillsets. First, we document that a larger focus on technical skillsets is associated with generally higher labor expenses at the firm level. Second, we look at the time trends in job openings and available talent to show that the demand for currently popular skills such as *Software Engineering* substantially outpaces the supply of qualified college graduates.

To begin with, in Figure 7, we report the coefficients and standard errors from the following regression for each technical skillset and (for comparison) each communication skillset:

$$WageBill_{i,t} = \alpha + \beta Skill^{j}_{i,t} + X_{i,t} + \epsilon_{i,t}, \tag{6}$$

where $WageBill_{i,t}$ captures the annual labor expenses reported by a subset of firms in Compustat (item XLR), and $Skill^{j}_{i,t}$ is the percentage of employees of firm i skilled in j, estimated at the annual level. The controls include industry and quarter fixed effects, as well as firm size, book-to-market ratio, and past performance. Since labor expenses are reported by fewer than 25% of all Compustat firms, the panel is much smaller than for the other analyses. Nonetheless, as can be seen from Figure 7, *Data Analysis, Software Engineering*, and *Web Development* are all significantly related to higher labor expenses. For comparison, the link between wages and communication skillsets is displayed on the right and does not show the same pattern – none of the communication skillsets are significantly related to labor expenses. Overall, technically skilled employees are significantly more expensive but do not appear to boost firm operations such as profitability and earnings surprises, as was demonstrated in Section 4.3.

Second, we consider the time trends in job openings and available talent (college graduates with corresponding majors). Using job postings data from Burning Glass Technologies, we observe annual numbers of job openings posted for the Software Development occupation group between 2010 and 2016. This includes occupations such as Software Developer / Engineer, Software QA Engineer/Tester, Web Developer, Computer Programmer, and Computer Scientist, and considers job postings for these roles from across the full spectrum of firms in the economy.[10] The total annual job postings numbers, reported weekly, triple from under 300,000 in 2010 to nearly 900,000 in 2016. This time series is displayed in blue in Figure 8 (with axis marked on the right) and shows a dramatic increase in Software Development job postings over the recent years (2014-2016). This is unlikely to be attributable to selection bias from Burning Glass data aggregating only online job postings, since their coverage, which is comprehensive at 60-70% of all U.S. job openings, has been stable throughout the

[10] Unfortunately, due to data limitations (job postings data from Burning Glass are only available starting in 2010), we are not able to conduct an analogous analysis for the two skillsets that were popular in the early 2000s, *Information Technology* and *Mobile Networks.*

2010s.[11]

At the same time, the amount of available talent, proxied by college graduates with majors in Computer Science and Information Science, has stayed relatively constant over this time period. The time series of Computer Science and Information Science majors comes from the NCES Integrated Postsecondary Education Data System survey. The red line in Figure 8 shows the annual numbers of U.S. students graduating with bachelor's degrees in Computer Science and Information Science from 2005 to 2015, with the axis on the left. We end the plot in 2015, as in 2016 NCES began aggregating computer science graduates with non-computer engineering graduates. The numbers vary from approximately 40,000 to 60,000 graduates per year, with the level in 2015 below that in 2005. As a result, graduates with degrees in Computer Science and Information Science have not increased nearly sufficiently to fill the growing number of Software Engineering positions. It is also unlikely that international graduates in CS-related disciplines are sufficient to close the gap; in light red, we plot the number of Computer Science graduates in the Cognism data (which has international coverage) to show that the total numbers of relevant college graduates do not appear to be rising either. As a result, as firms compete for the limited pool of qualified technical employees, the rising wages are accompanied by less qualified candidates filling the roles corresponding to these over-hyped skillsets, in order to meet demand. This likely contributes to the observed overpricing of the popular skillsets and correspondingly worse subsequent firm operations and returns.

6 Conclusion

We document that the employee skill composition of a firm can have important repercussions for the firm's valuation and subsequent performance, but that the stock return premia do not necessarily line up with the individual wage premia. Specifically, technical skillsets and communication skillsets, both of which have been found to carry positive wage premia for individuals in the labor economics literature, forecast negative firm-level returns and deteriorating operations. We interpret the negative premia on communication skillsets as likely reflecting these skills' counter-cyclical performance. Negative premia on technical skillsets, however, appear to capture over-excitement about novel technologies at the height of their popularity, whereby higher shares of technically skilled employees translate into higher current firm valuations but predictable reversals in the future.

Our work brings together two literatures. On the one hand, we contrast and build upon a

[11]For example, Liu and Wu (2019) show that the number of publicly traded firms represented in Burning Glass data has remained unchanged from 2010 onwards.

broad literature of CEO-centered firm analysis, with the prospect that a more granular look at individuals across the firm hierarchy can offer additional insights into the inner workings, efficiency, and ultimate success of the modern firm. On the other hand, we build on a long tradition in labor economics of estimating returns to education and skills. We show that skillsets that carry high returns for individuals can be overvalued by the market and therefore do not necessarily aggregate up to analogous return premia at the firm level.

References

[1] Adams, R. B., H. Almeida, and D. Ferreira (2005). "Powerful CEOs and their impact on corporate performance." *Review of Financial Studies,* 18.4: 1403-1432.

[2] Agrawal, A. K., I. Hacamo, and Z. Hu (2020). "Information dispersion across employees and stock returns." *Review of FInancial Studies,* forthcoming.

[3] Belo, F. X. Lin, and S. Bazdresch (2014). "Labor hiring, investment, and stock return predictability in the cross section." *Journal of Political Economy,* 122.1: 129-177.

[4] Belo, F., J. Li, X. Lin, and X. Zhao (2017). "Labor-force heterogeneity and asset prices: The importance of skilled labor." *Review of Financial Studies,* 30.10: 3669-3709.

[5] Benmelech, E. and C. Frydman (2015). "Military CEOs." *Journal of Financial Economics,* 117.1: 43-59.

[6] Bertrand, M., and A. Schoar (2003). "Managing with style: The effect of managers on firm policies." *Quarterly Journal of Economics,* 118.4: 1169-1208.

[7] Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). "Latent dirichlet allocation." *Journal of Machine Learning Research*, 3: 993-1022.

[8] Bughin, J., J. Seong, J. Manyika, M. Chui, and R. Joshi (2018). "Notes from the AI frontier: Modeling the impact of AI on the world economy," *McKinsey and Company Global Institute.*

[9] Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance." *Journal of Finance,* 52.1: 57-82.

[10] Carnevale, A. P., T. Jayasundera, and D. Repnikov (2014). "Understanding online job ads data." Georgetown University, Center on Education and the Workforce, Technical Report.

[11] Chung, K. H. and S. W. Pruitt (1994). "A simple approximation of Tobin's q." *Financial Management*, 23.3: 70-74.

[12] Da, Z., U. G. Gurun, and M. Warachka (2014). "Frog in the pan: Continuous information and momentum." *Review of Financial Studies,* 27.7: 2171-2218.

[13] Damerau, F. (1964). “A technique for computer detection and correction of spelling errors.” *Communications of the ACM*, 7.3: 659-664.

[14] Deming, D. J. (2017) “The growing importance of social skills in the labor market.” *Quarterly Journal of Economics,* 132.4: 1593-1640.

[15] Donangelo, A. (2014). “Labor mobility: Implications for asset pricing.” *Journal of Finance,* 69.3: 1321-1346.

[16] Edmans, A. (2011). “Does the stock market fully value intangibles? Employee satisfaction and equity prices.” *Journal of Financial Economics,* 101.3: 621-640.

[17] Fama, E. F, and K. R French (1992). “The cross-section of expected stock returns.” *Journal of Finance,* 47.2: 427?465.

[18] Fama, E. F, and K. R. French (1993). “Common risk factors in the returns on stocks and bonds.” *Journal of Financial Economics*, 33.1: 3?56.

[19] Fama, E. F., and J. D. MacBeth (1973). “Risk, return, and equilibrium: Empirical tests.” *Journal of Political Economy*, 81.3: 607-636.

[20] Fedyk, A. (2020). “Front page news: The effect of news positioning on financial markets.” Working paper.

[21] Galasso, A. and T. S. Simcoe (2011). “CEO overconfidence and innovation.” *Management Science,* 57.8: 1469-1484.

[22] Graham, J. R., C. R. Harvey, and M. Puri (2013). “Managerial attitudes and corporate actions.” *Journal of Financial Economics,* 109.1: 103-121.

[23] Guiso, L., P. Sapienza, and L. Zingales (2015). “The value of corporate culture.” *Journal of Financial Economics,* 117.1: 60-76: 2661-2700.

[24] Hombert, J. and A. Matray (2020). “Technology boom, labor reallocation, and human capital depreciation.” Working paper.

[25] Ingram, B. F. and G. R. Neumann (2006). “The returns to skill.” *Labour Economics,* 13.1: 35-59.

[26] Jeffers, J. S. (2020). “The impact of restricting labor mobility on corporate investment and entrepreneurship.” Working Paper.

[27] Juhn, C., K. M. Murphy, and B. Pierce (1993). “Wage inequality and the rise in returns to skill.” *Journal of Political Economy,* 101.3: 410-442.

[28] Liu, Y. and X. Wu (2019). “Labor links and shock transmissions.” Working paper.

[29] Kaplan, S. N., M. M. Klebanov, and M. Sorensen (2012). “Which CEO characteristics and abilities matter?” *Journal of Finance,* 67.3 (2012): 973-1007.

[30] Kilic, M. (2016). "Asset Pricing Implications of Hiring Demographics." Working paper.

[31] Kuhn, P. and C. Weinberger (2005). "Leadership skills and wages." *Journal of Labor Economics,* 23.3: 395-436.

[32] Malmendier, U. and G. Tate (2005). "CEO Overconfidence and Corporate Investment." *Journal of Finance,* 60.6:

[33] Malmendier, U., and G. Tate (2008). "Who makes acquisitions? CEO overconfidence and the market's reaction." *Journal of financial Economics,* 89.1: 20-43.

[34] Mukharlyamov, V. (2016). "Beyond the corner office: Employee characteristics and bank performance." Working paper, Georgetown University.

[35] Navigli, R. (2009). "Word sense disambiguation: A survey." *ACM Computing Surveys (CSUR)*, 41.2: 10:1-10:69.

[36] Starr, E., N. Balasubramanian, and M. Sakakibara (2018). "Screening spinouts? How noncompete enforceability affects the creation, growth, and survival of new firms." *Management Science,* 64.2:552-572.

[37] So, E. C. and S. Wang (2014). "News-driven return reversals: Liquidity provision ahead of earnings announcements." *Journal of Financial Economics,* 114.1: 20-35.

[38] Zingales, L. (2000). "In search of new foundations." *Journal of Finance,* 55.4: 1623-1654.

[39] Yim, S. (2013). "The acquisitiveness of youth: CEO age and acquisition behavior." *Journal of Financial Economics,* 108.1: 250-273.

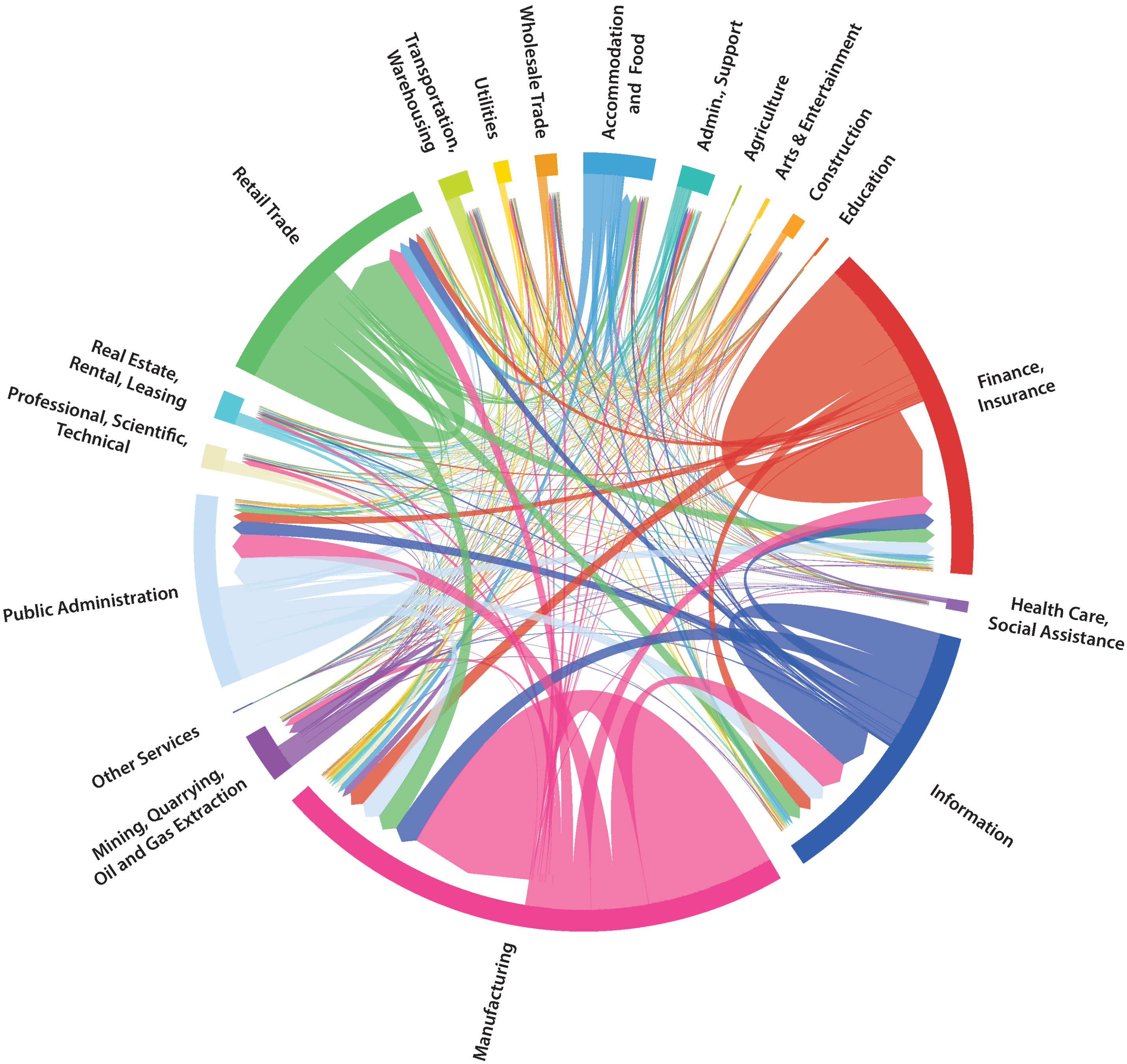


Figure 1: Cross-industry employee moves during the years 2010-2016. We consider all job moves that involve switching employers. Each arrow captures the prevalence of moves from one industry to another, with size in proportion to the number of transitions. The industries are delineated based on two-digit NAICS codes.

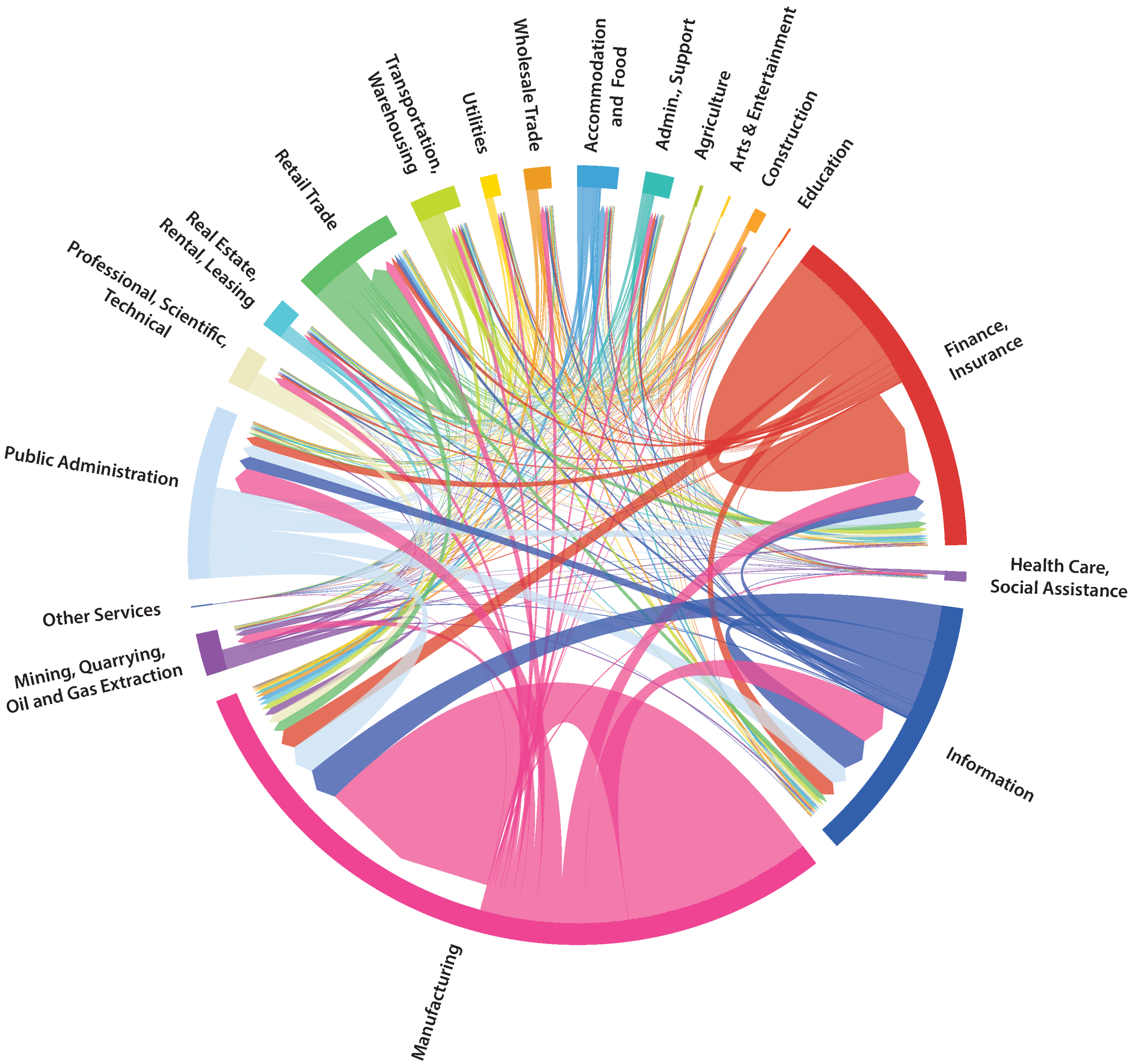


Figure 2: Cross-industry employee moves during the 1990s. We consider all job moves that involve switching employers. Each arrow captures the prevalence of moves from one industry to another, with size in proportion to the number of transitions. The industries are delineated based on two-digit NAICS codes.

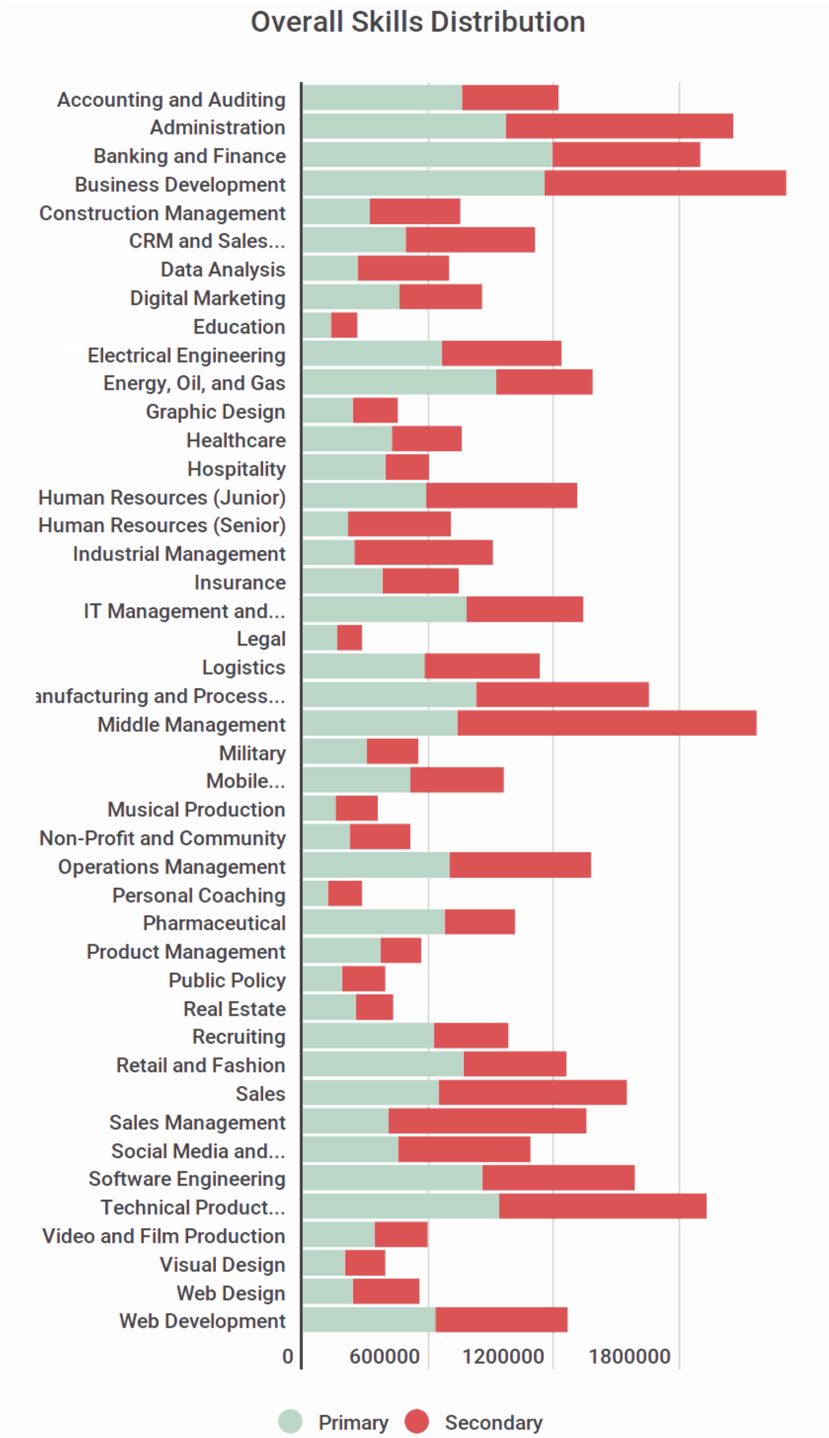


Figure 3: Frequency of skillsets across the full population of 37 million employees of public U.S. companies. Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the number of employees for whom that skillset is primary (in green) and the number of employees for whom that skillset is secondary (in red).

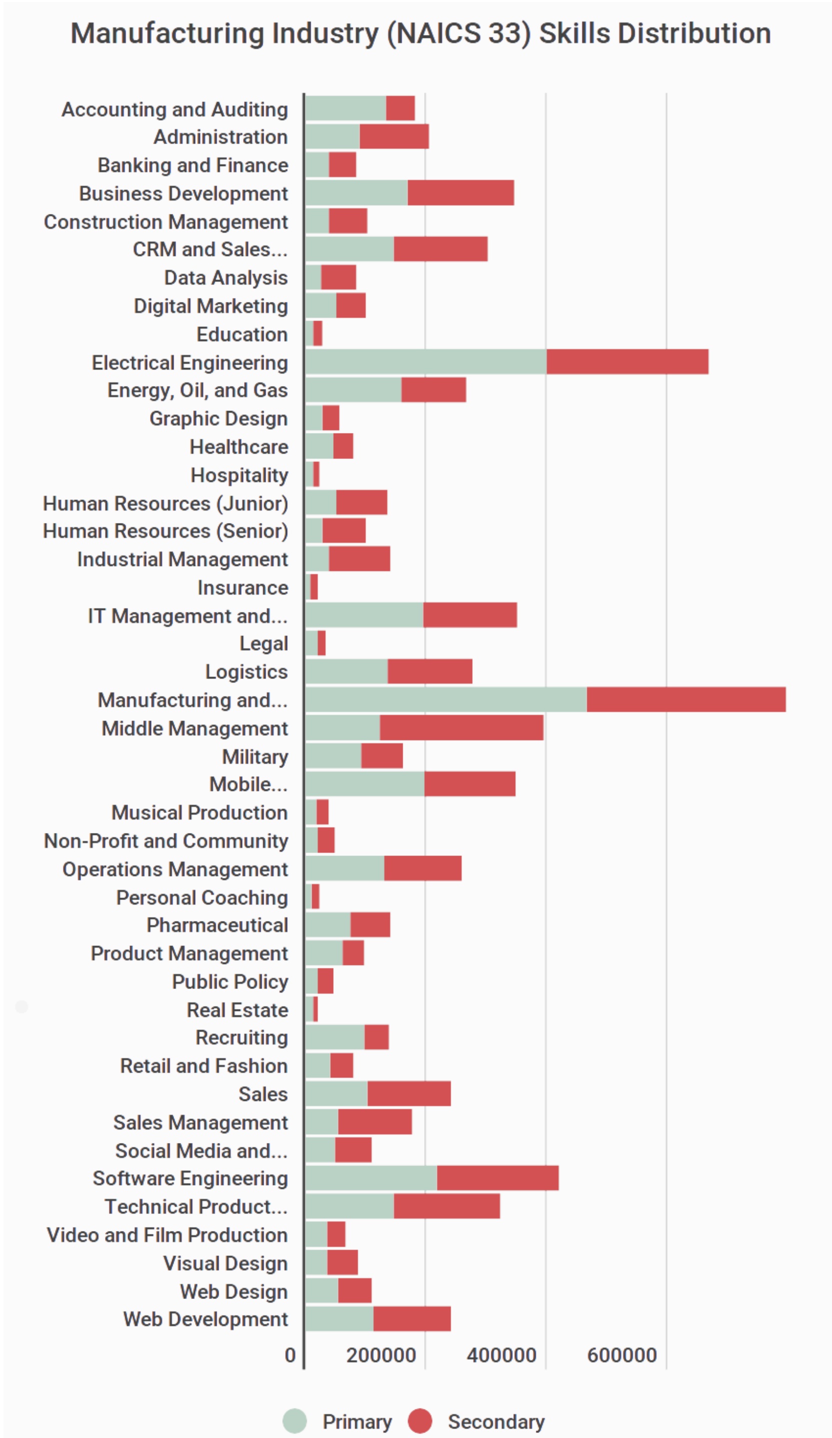


Figure 4: Frequency of skillsets across the employees of public U.S. companies in the Manufacturing industry (two-digit NAICS codes 31, 32, and 33). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the number of manufacturing industry employees for whom that skillset is primary (in green) and the number of employees for whom that skillset is secondary (in red).

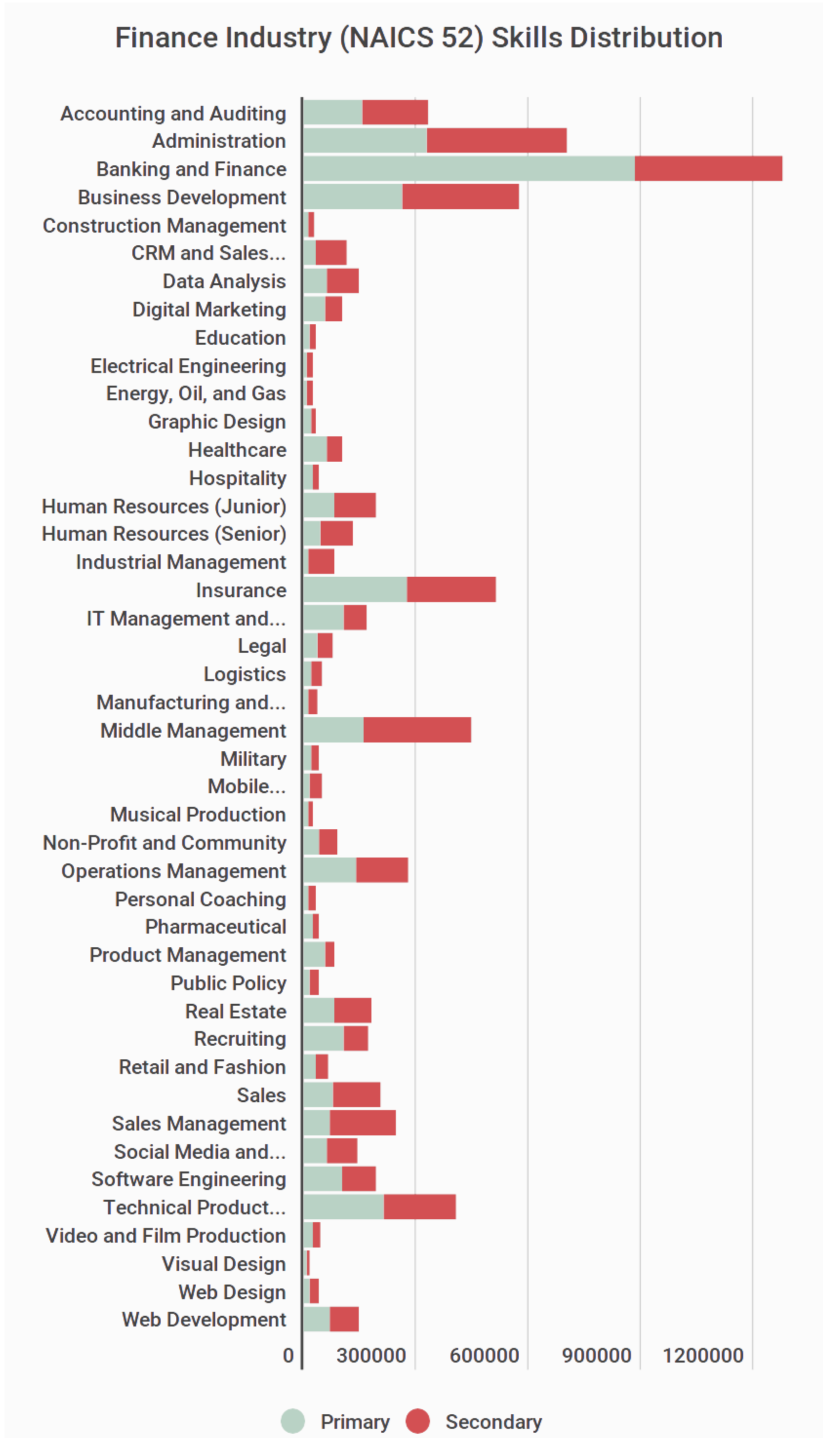
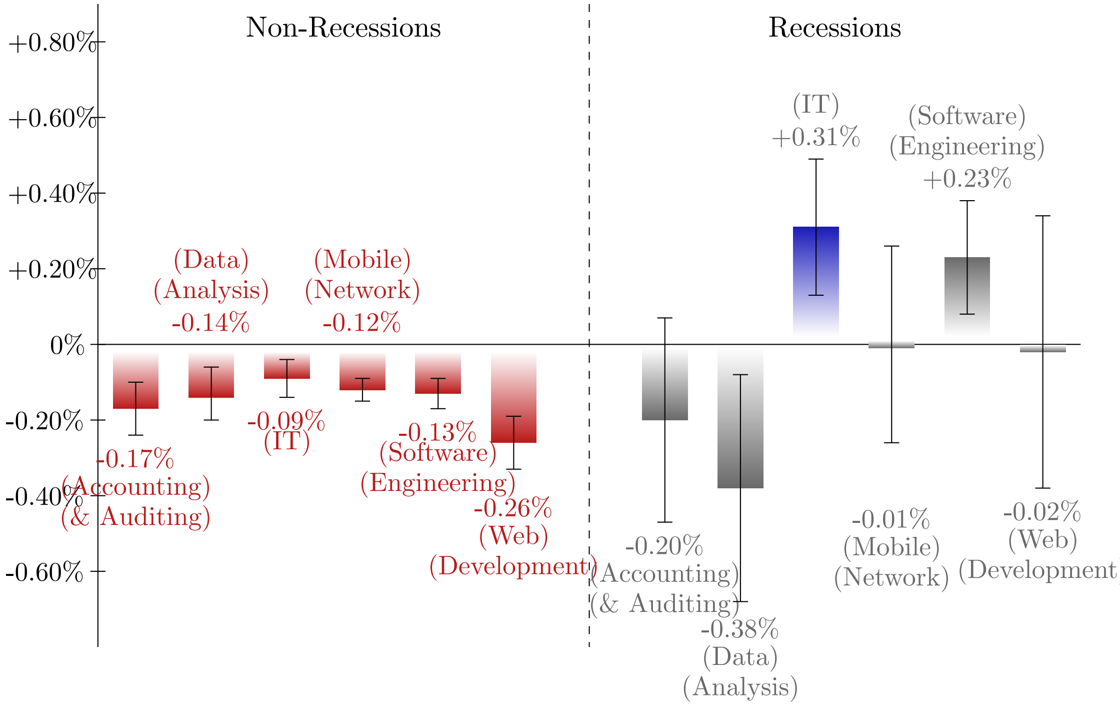


Figure 5: Frequency of skillsets across the employees of public U.S. companies in the Finance and Insurance industry (two-digit NAICS code 52). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the number of finance industry employees for whom that skillset is primary (in green) and the number of employees for whom that skillset is secondary (in red).

Panel 1: Technical Skills



Panel 2: Communication Skills

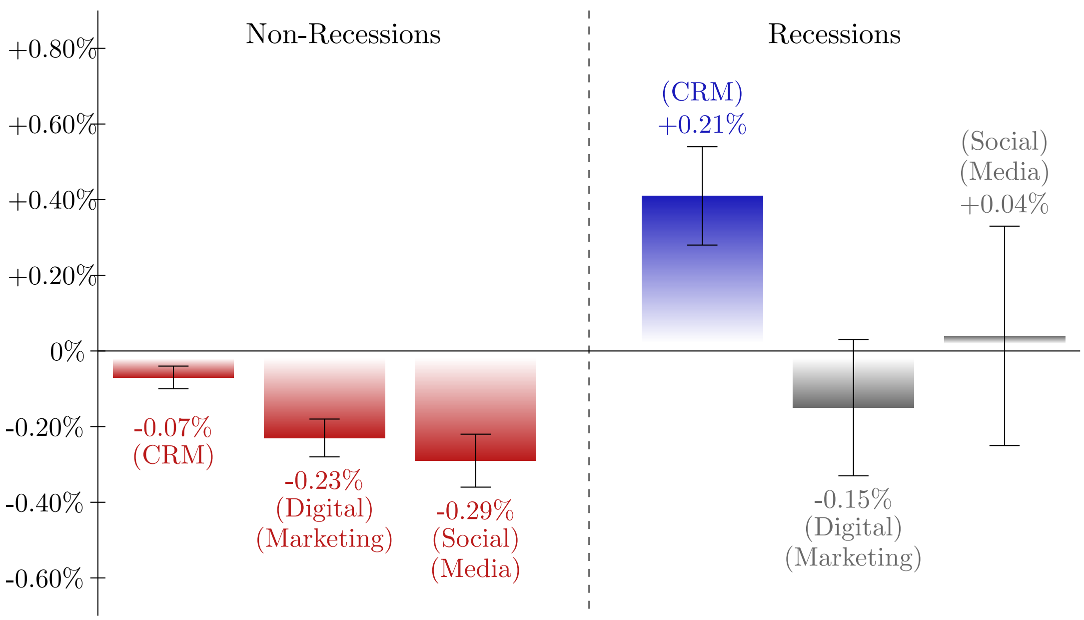


Figure 6: Coefficients from panel regressions of monthly three-factor alphas on the firm's percentage of employees with a given skillset as primary or secondary, computed during the previous month. Controls include industry and time fixed effects, firm size, book-to-market ratio, and past performance. The regression is estimated separately for months marked as recession months by the NBER and for non-recession months. **Panel 1** looks at technical skillsets, while **Panel 2** considers communication skillsets.

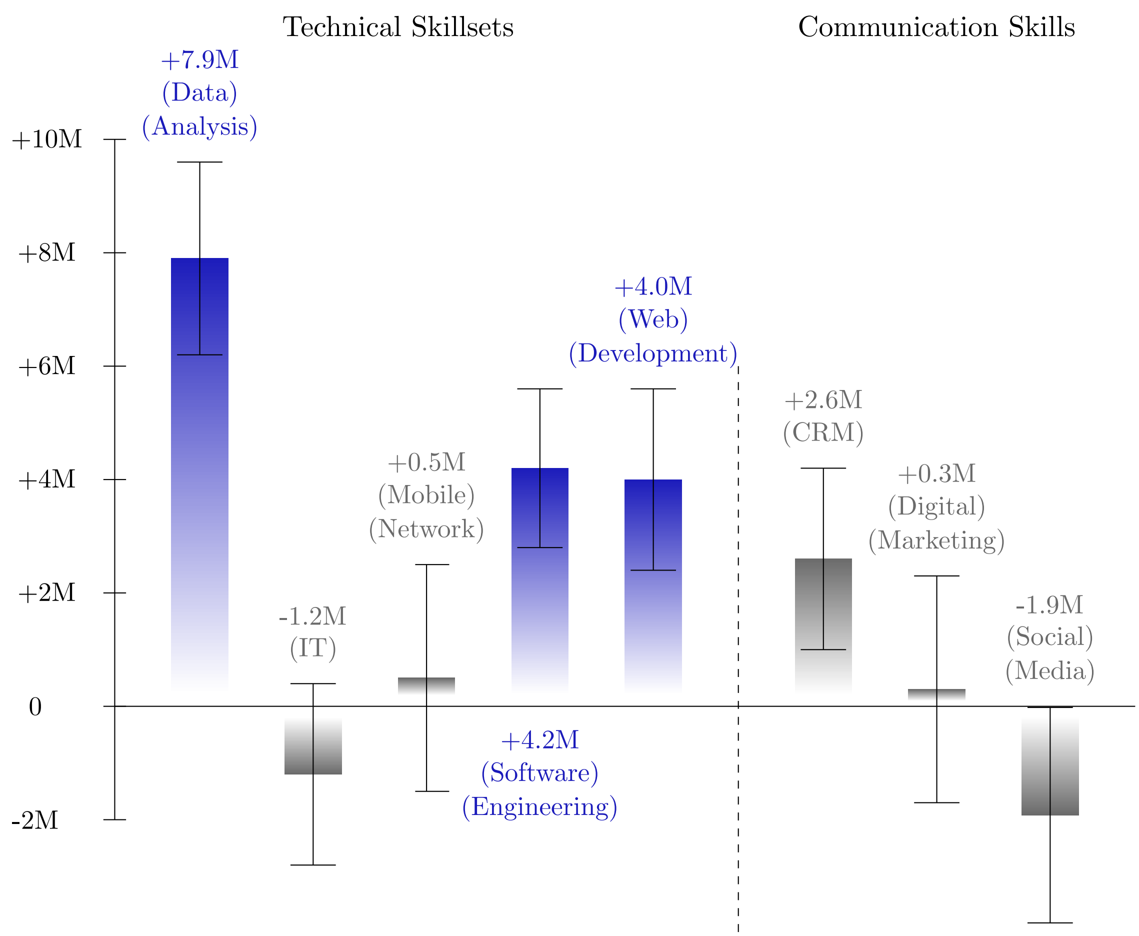


Figure 7: Coefficients from panel regressions of the annual firm-level wage bill (Compustat item XLR) on the firm's percentage of employees with a given skillset as primary or secondary, calculated over the course of the same year. Controls include industry and time fixed effects, firm size, book-to-market ratio, and past performance.

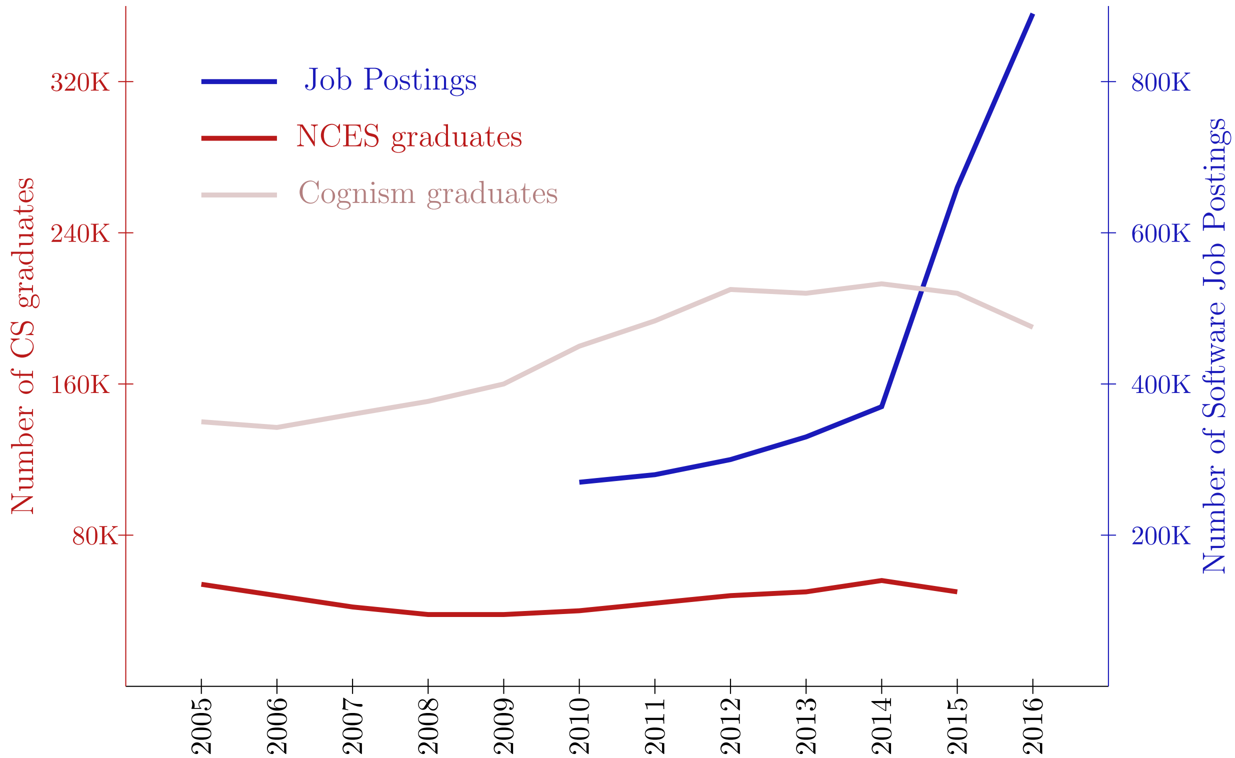


Figure 8: Time series of the number of software-related job openings (plotted in blue, axis on the right) and college graduates majoring in computer science or information science (in red, axis on the left). Job posting data is only available since 2010. Data on college graduates majoring specifically in computer science is available from the NCES through 2015 and in the Cognism data through 2016.

Table 1: Summary statistics of the resume data. **Panel 1** presents the breakdown by geographic region. **Panel 2** presents the distribution of experience in terms of age, jobs held, connections, and skills, calculated across users with non-empty information for each field. **Panel 3** shows the breakdown by education for those who report at least one degree.

Panel 1: Geography

United States	77.3%
Continental Europe	6.7%
Asia	4.6%
Central & South America	3.4%
United Kingdom	3.2%
Non-U.S. North America	1.8%
Middle East	1.3%
Oceania	1.1%
Africa	0.6%

Panel 2: Experience

	Mean	Minimum	25th %tile	Median	75th%tile	Maximum
Age[a]	36	18	30	28	64	74
# Jobs[b]	3.6	0	1	2	5	72
# Skills[c]	10	0	0	4	16	52

[a]Age is winsorized at top and bottom 0.1%.

[b]Includes job transitions within the same firm. The number of positions is winsorized at the top 0.1%; the reported maximum in the sample is 819.

[c]Numbers of skills are winsorized at the top 0.1%. The reported maximum in the sample is 886.

Panel 3: Education (for those reporting at least one)

% Reporting Vocational	3.97%
% Reporting Associates	1.74%
% Reporting Bachelors	89.56%
% Reporting Masters	20.37%
% Reporting MBA	7.94%
% Reporting Doctorate	3.65%

Table 2: Employment by NAICS two-digit industry code, as of January 1, 2016, January 1, 2006, and January 1, 1996.

Industry	2016		2006		1996	
Agriculture, Forestry, Fishing Hunting	36,205	0.1%	14,845	0.1%	5,763	0.2%
Mining, Quarrying, Oil & Gas Extract.	877,288	3.2%	316,291	2.8%	79,182	2.6%
Utilities	253,026	0.9%	107,028	1.0%	36,254	1.2%
Construction	234,854	0.9%	96,140	0.9%	19,696	0.7%
Manufacturing	8,636,437	31.7%	3,879,783	34.9%	1,175,268	39.0%
Wholesale Trade	318,250	1.2%	143,873	1.3%	39,758	1.3%
Retail Trade	2,836,364	10.4%	865,148	7.8%	173,430	5.8%
Transportation and Warehousing	560,056	2.1%	284,614	2.6%	109,168	3.6%
Information	3,665,602	13.5%	1,450,358	13.0%	363,539	12.1%
Finance and Insurance	4,514,713	16.6%	1,954,765	17.6%	520,407	17.3%
Real Estate and Rental and Leasing	439,400	1.6%	190,612	1.7%	55,932	1.9%
Professional, Scientific, Tech. Services	2,498,345	9.2%	956,351	8.6%	233,856	7.8%
Administrative, Support, etc. Services	463,610	1.7%	168,004	1.5%	33,042	1.1%
Educational Services	36,928	0.1%	11,454	0.1%	2,204	0.1%
Health Care and Social Assistance	169,899	0.6%	65,216	0.6%	15,763	0.5%
Arts, Entertainment, and Recreation	62,324	0.2%	16,222	0.1%	3,744	0.1%
Accommodation and Food Services	1,105,810	4.1%	348,543	3.1%	79,748	2.6%
Other Services	15,307	0.1%	6,888	0.1%	1,796	0.1%
Public Administration	491,926	1.8%	241,816	2.2%	66,624	2.2 %

Table 3: Labeled skillsets extracted from the self-reported skills and the five most likely terms to appear conditional on each skillset.

Skillset	Most Common Terms
Technical Skillsets	
Data Analysis	data analysis, research, statistics, microsoft office, spss
Information Technology	windows server, troubleshooting, active directory, networking, window
Mobile Networks	telecommunications, wireless, voip, networking, ip
Software Engineering	java, sql, software development, linux, agile methodologies
Web Development	javascript, html, java, css, sql
Communication Skillsets	
Client Relationship Management	business development, strategy, management, product management, crm
Digital Marketing	digital marketing, social media marketing, marketing, online advertising, online marketing
Social Media	social media, public relations, social media marketing, marketing, event management
Operational Skillsets	
Industrial Management	microsoft office, microsoft excel, sap, microsoft word, sap erp
Logistics	logistics, supply chain management, operations management, supply chain, purchasing
Operations Management	project management, change management, business analysis, it management, business process improvement
Sales	sales, sales management, account management, customer service, new business development
Sales Management	strategic planning, management, customer service, new business development, negotiation
Tech. Product Mgmt	business analysis, requirements analysis, sql, business intelligence, project management
Administrative Skillsets	
Admin	microsoft office, microsoft excel, microsoft word, powerpoint, customer service
Business Development	business strategy, marketing strategy, business development, management, market research
Human Resources (Jr.)	coaching, change management, training, leadership development, management
Human Resouces (Sr.)	teamwork, communication, microsoft office, customer service, time management
Middle Management	management, project management, leadership, strategic planning, process improvement
Product Mgmt (Generic)	microsoft office, strategic planning, negotiation, microsoft excel, microsoft word
Recruiting	recruiting, human resources, employee relations, talent acquisition, performance management

Table 3 (Continued): Labeled skillsets extracted from the self-reported skills and the five most likely terms to appear conditional on each skillset.

Skillset	Most Common Terms
Industry-Specific Skillsets	
Accounting & Auditing	accounting, financial reporting, financial analysis, auditing, financial accounting
Banking & Finance	banking, financial analysis, finance, risk management, portfolio management
Construction	construction, construction management, contract management, project planning, project management
Education	teaching, higher education, curriculum development, curriculum design, public speaking
Electrical Engineering	matlab, engineering, autocad, solidworks, c++
Graphic Design	graphic design, photoshop, photography, illustrator, adobe creative suite
Healthcare	healthcare, hospitals, healthcare management, clinical research, healthcare information technology
Hospitality	hospitality, customer service, hotels, hospitality management, food & beverage
Insurance	insurance, customer service, risk management, property & casualty insurance, general insurance
Legal	legal research, legal writing, litigation, civil litigation, corporate law
Manufacturing	manufacturing, continuous improvement, lean manufacturing, six sigma, product development
Military	military, security clearance, security, military experience, military operations
Musical Production	music, entertainment, music production, theatre, music industry
Non-Profit	nonprofits, public speaking, community outreach, fundraising, event planning
Oil, Energy & Gas	engineering, energy, petroleum, gas, project engineering
Pharmaceutical	pharmaceutical industry, biotechnology, molecular biology, life sciences, chemistry
Public Policy	research, international relations, policy analysis, policy, sustainability
Real Estate	real estate, investment properties, real estate transactions, residential homes, property management
Retail	retail, merchandising, customer service, sales, inventory management
Video & Film	editing, social media, video production, blogging, journalism
Visual Design	autocad, sketchup, interior design, photoshop, microsoft office
Web Design	photoshop, illustrator, web design, graphic design, indesign
Personal Coaching	coaching, public speaking, sports, wellness, nutrition

Table 4: Summary statistics of the main skillset variables. **Panel 1** summarizes the distributions of the firm-level skillset measures. **Panel 2** presents cross-correlations between these variables, while **Panel 3** displays the correlations between abnormal versions of the variables, orthogonalized to firm size, book-to-market ratio, and past performance, as well as industry and month fixed effects.

Panel 1: Individual Distributions

Skillset	mean	median	std	min	10%	25%	75%	90%	max
Technical Skillsets									
Data Analysis	1.3%	0.8%	1.8%	0.0%	0.0%	0.4%	1.5%	2.9%	11.9%
IT	4.3%	2.9%	4.9%	0.0%	1.2%	1.9%	4.6%	8.5%	33.2%
Mobile Networks	2.9%	0.7%	7.8%	0.0%	0.1%	0.4%	1.6%	4.7%	48.7%
Software Engineering	3.9%	0.9%	7.3%	0.0%	0.0%	0.3%	3.1%	12.9%	38.7%
Web Development	2.6%	1.3%	3.5%	0.0%	0.4%	0.7%	2.8%	6.2%	20.2%
Communication Skillsets									
CRM	4.2%	1.2%	7.8%	0.0%	0.2%	0.6%	3.4%	12.3%	40.7%
Digital Marketing	2.3%	1.1%	4.5%	0.0%	0.1%	0.5%	2.2%	4.6%	30.9%
Social Media	2.5%	1.5%	3.1%	0.0%	0.4%	0.8%	2.8%	5.6%	19.9%

Panel 2: Cross-correlations (raw variables)

	D.A.	I.T.	M.N.	S.E.	W.D.	C.R.M.	D.M.	S.M
Data Analysis	–	-0.09	-0.10	0.03	0.10	-0.05	0.04	0.02
IT		–	0.45	0.44	0.40	0.60	0.03	-0.12
Mobile Networks			–	0.32	0.14	0.44	-0.01	-0.09
Software Engineering				–	0.77	0.57	0.13	-0.11
Web Development					–	0.48	0.29	0.01
Client Relationship Management						–	0.19	-0.06
Digital Marketing							–	0.60
Social Media								–

Panel 3: Cross-correlations (abnormal variables)

	D.A.	I.T.	M.N.	S.E.	W.D.	C.R.M.	D.M.	S.M
Data Analysis	–	-0.11	-0.09	0.01	0.05	-0.04	0.01	-0.01
IT		–	0.39	0.32	0.25	0.51	-0.12	-0.19
Mobile Networks			–	0.21	-0.01	0.35	-0.12	-0.12
Software Engineering				–	0.71	0.45	-0.05	-0.21
Web Development					–	0.31	0.08	-0.14
Client Relationship Management						–	0.02	-0.14
Digital Marketing							–	0.59
Social Media								–

Table 5: Results from Fama and MacBeth (1973) regressions of monthly firm-level valuations (Tobin's q) on contemporaneous compositions of employee skillsets. We estimate the following specification for each skillset j:

$Tobin\_q_{i,t} = \alpha + \beta Skill^{j}_{i,t} + X_{i,t} + \epsilon_{i,t}$,

where $Tobin\_q_{i,t}$ denotes the ratio of the market value of firm i's equity and liabilities to the book value of equity and liabilities, measured using latest values at the end of month t. $Skill^{j}_{i,t}$ is the percentage of firm i's employees who possess skillset j in month t. The controls $X_{i,t}$ include industry fixed effects, as well as size, book-to-market ratio, and past performance in column 1 and no controls in column 2. The table reports the coefficient β corresponding to a 10% change in each skill variable.

Technical Skillsets	(1) With controls	(2) Without controls
Data Analysis		
Coefficient	1.15%***	1.27%***
(Standard error)	*(0.24%)*	*(0.24%)*
Information Technology		
Coefficient	0.43%	0.39%
(Standard error)	*(0.32%)*	*(0.27%)*
Mobile Networks		
Coefficient	-0.20%	-0.06%
(Standard error)	*(0.34%)*	*(0.29%)*
Software Engineering		
Coefficient	2.39%***	2.64%***
(Standard error)	*(0.40%)*	*(0.38%)*
Web Development		
Coefficient	2.31%***	2.28%***
(Standard error)	*(0.35%)*	*(0.40%)*

Communication Skillsets	(1) With controls	(2) Without controls
CRM		
Coefficient	2.59%***	2.61%***
(Standard error)	*(0.38%)*	*(0.31%)*
Digital Marketing		
Coefficient	0.53%	0.84%**
(Standard error)	*(0.46%)*	*(0.40%)*
Social Media		
Coefficient	-0.21%	0.10%
(Standard error)	*(0.26%)*	*(0.25%)*

Table 6: Results from panel regressions of monthly firm-level stock market returns on lagged composition of employee skillsets. We estimate the following specification at four lags for each skillset j:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Skill^j_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ denotes the return for firm i in month t in excess of that predicted by the Fama and French (1993) factors. $Skill^j_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the lagged month $t-L$, and the controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, and past performance. The table reports the coefficient β corresponding to a 10% change in each skill variable.

Skillset	(1) L = 1 month	(2) L = 2 months	(3) L = 3 months	(4) L = 6 months
Technical Skillsets				
Data Analysis				
Coefficient	-0.07%	-0.06%	-0.06%	-0.06%
(Standard error)	*(0.14%)*	*(0.14%)*	*(0.14%)*	*(0.14%)*
Information Technology				
Coefficient	-0.05%	-0.04%	-0.03%	-0.03%
(Standard error)	*(0.05%)*	*(0.05%)*	*(0.05%)*	*(0.05%)*
Mobile Networks				
Coefficient	-0.10%***	-0.10%***	-0.10%***	-0.08%***
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Software Engineering				
Coefficient	-0.07%**	-0.06%*	-0.06%*	-0.04%
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*
Web Development				
Coefficient	-0.12%*	-0.09%	-0.09%	-0.06%
(Standard error)	*(0.07%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Communication Skillsets				
CRM				
Coefficient	-0.06%*	-0.04%	-0.04%	-0.04%
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Digital Marketing				
Coefficient	-0.21%***	-0.18%***	-0.19%***	-0.21%***
(Standard error)	*(0.06%)*	*(0.06%)*	*(0.06%)*	*(0.06%)*
Social Media				
Coefficient	-0.17%**	-0.16%**	-0.16%**	-0.14%*
(Standard error)	*(0.08%)*	*(0.08%)*	*(0.08%)*	*(0.08%)*

Table 6 (Continued): Results from panel regressions of monthly firm-level stock market returns on lagged composition of employee skillsets. We estimate the following specification at four lags for each skillset j:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ denotes the return for firm i in month t in excess of that predicted by the Fama and French (1993) factors. $Skill^{j}_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the lagged month $t - L$, and the controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, industry, and past performance. The table reports the coefficient β corresponding to a 10% change in each skill variable.

Skillset	(1) L = 1 month	(2) L = 2 months	(3) L = 3 months	(4) L = 6 months
Operational Skillsets				
Industrial Management				
Coefficient	0.21%**	0.13%	0.16%*	0.10%
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.10%)*	*(0.10%)*
Logistics				
Coefficient	0.06%	0.05%	0.06%*	0.06%*
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*
Operations Management				
Coefficient	0.18%**	0.20%**	0.22%**	0.23%***
(Standard error)	*(0.09%)*	*(0.09%)*	*(0.09%)*	*(0.09%)*
Sales				
Coefficient	0.07%**	0.08%***	0.08%***	0.07%**
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Sales Management				
Coefficient	0.09%	0.06%	0.07%	0.09%
(Standard error)	*(0.08%)*	*(0.07%)*	*(0.08%)*	*(0.08%)*
Tech. Product Mgmt				
Coefficient	0.08%**	0.10%***	0.10%***	0.10%***
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*

Table 7: To evaluate the relationship between skill focus and gross profitability, we estimate the following specification for four quarter lags:
For $L \in \{1\,q, 2\,q, 3\,q, 4\,q\} : \Delta Profitability_{i,t} = \alpha + \beta \Delta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $\Delta Profitability_{i,t}$ is the percentage change in firm i's gross profitability from quarter $t-1$ to quarter t, defining gross profitability as the difference between revenues and cost of goods sold, scaled by assets. $\Delta Skill^{j}_{i,t-L}$ is the change in the percentage of firm i's employees who possess a skillset j during the lagged quarter $t-L$. The controls $X_{i,t}$ include industry and quarter fixed effects, firm size, book-to-market ratio, and past performance. The table reports the coefficient β corresponding to a one standard deviation move in the quarterly change in each skill variable.

Skillset	(1) L = 1 quarter	(2) L = 2 quarters	(3) L = 3 quarters	(4) L = 4 quarters
Technical Skillsets				
Data Analysis				
Coefficient	0.86%	5.78%	-0.63%	-13.09%***
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Information Technology				
Coefficient	0.64%	-8.81%**	0.04%	0.47%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Mobile Networks				
Coefficient	-1.08%	0.29%	-0.42%	-0.31%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Software Engineering				
Coefficient	0.08%	0.18%	2.01%	-0.15%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Web Development				
Coefficient	0.35%	-0.23%	0.02%	-0.69%
(Standard error)	*(3.63%)*	*(3.71%)*	*(0.80%)*	*(3.86%)*
Communication Skillsets				
CRM				
Coefficient	9.23%**	0.18%	-0.05%	-2.88%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Digital Marketing				
Coefficient	0.32%	1.33%	2.84%	1.28%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*
Social Media				
Coefficient	-0.20%	-0.32%	-0.57%	-0.11%
(Standard error)	*(3.63%)*	*(3.71%)*	*(3.80%)*	*(3.86%)*

Table 8: We estimate the relationship between skill changes and earnings surprises using the following specification for four quarter lags:
For $L \in \{1\,q, 2\,q, 3\,q, 4\,q\} : SUE_{i,t} = \alpha + \beta \Delta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $SUE_{i,t}$ is firm i's standardized earnings surprise in quarter t, defined as the difference between earnings in quarters t and $t-4$, scaled by the standard deviation of this difference over the preceding eight quarters. $\Delta Skill^{j}_{i,t-L}$ is the change in the percentage of firm i's employees who possess skillset j during the lagged quarter $t-L$. The controls $X_{i,t}$ include industry and quarter fixed effects, firm size, book-to-market ratio, industry, and past performance. The table reports the coefficient β corresponding to a one standard deviation move in the quarterly change in each skill variable.

Skillset	(1) L = 1 quarter	(2) L = 2 quarters	(3) L = 3 quarters	(4) L = 4 quarters
Technical Skillsets				
Data Analysis				
Coefficient	-1.49%**	-2.89%***	-1.51%**	-0.27%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Information Technology				
Coefficient	1.16%*	0.20%	0.48%	0.61%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Mobile Networks				
Coefficient	-0.12%	1.50%**	2.45%***	1.71%**
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Software Engineering				
Coefficient	-1.19%*	-1.23%*	-0.50%	1.07%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Web Development				
Coefficient	-0.19%	-1.86%***	-2.54%***	-1.26%*
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Communication Skillsets				
CRM				
Coefficient	0.73%	2.11%**	2.78%***	3.07%***
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Digital Marketing				
Coefficient	-0.69%	-0.66%	-1.35%**	-0.18%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*
Social Media				
Coefficient	-0.02%	-1.26%*	-0.17%	-2.08%***
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.68%)*

Table 9: Results from subsample analysis across three periods: 2000-2005, 2006-2010, and 2011-2016. We estimate the following specification for each sub-period:
$AbnRet_{i,t} = \alpha + \beta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ is the return for firm i in month t in excess of that predicted by the Fama and French (1993) factors. $Skill^{j}_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the lagged month $t-L$. The controls $X_{i,t}$ include industry and month fixed effects, as well as firm size, book-to-market ratio, and past performance. The table reports the coefficient β corresponding to an additional 10% of employees possessing each skillset as either their primary or secondary skillset.

Skillset	2000-2005	2006-2010	2011-2016
Technical Skillsets			
Data Analysis			
Coefficient	0.09%	-0.06%	-0.15%*
(Standard error)	*(0.18%)*	*(0.15%)*	*(0.10%)*
Information Technology			
Coefficient	-0.23%**	0.06%	-0.03%
(Standard error)	*(0.10%)*	*(0.09%)*	*(0.06%)*
Mobile Networks			
Coefficient	-0.20%***	-0.11%*	-0.06%
(Standard error)	*(0.03%)*	*(0.04%)*	*(0.05%)*
Software Engineering			
Coefficient	-0.05%	0.02%	-0.12%***
(Standard error)	*(0.10%)*	*(0.07%)*	*(0.04%)*
Web Development			
Coefficient	0.21%	0.20%	-0.25%***
(Standard error)	*(0.24%)*	*(0.17%)*	*(0.08%)*
Communication Skillsets			
CRM			
Coefficient	-0.11%*	0.01%	-0.08%*
(Standard error)	*(0.07%)*	*(0.06%)*	*(0.05%)*
Digital Marketing			
Coefficient	-0.34%***	0.05%	-0.26%***
(Standard error)	*(0.12%)*	*(0.09%)*	*(0.06%)*
Social Media			
Coefficient	-0.53%***	0.00%	-0.16%*
(Standard error)	*(0.19%)*	*(0.15%)*	*(0.08%)*

Table 10: Results from cross-sectional valuation regressions at four points in time: January 2001, 2006, 2011, and 2016. We estimate the following specification on each date:
$Tobins\_q_{i,t} = \alpha + \beta Skill^{j}_{i,t} + X_{i,t} + \epsilon_{i,t}$,
where $Tobins\_q_{i,t}$ is firm i's market value of equity and liabilities at the end of month t, scaled by the firm's book value of assets and liabilities at the same time. $Skill^{j}_{i,t}$ is the percentage of firm i's employees who possess skillset j in month t. The controls $X_{i,t}$ include industry and month fixed effects, as well as firm size, book-to-market ratio, industry, and past performance. The table reports the coefficient β corresponding to an additional 10% of employees possessing each skillset as either their primary or secondary skillset.

	2001	2006	2011	2016
Technical Skillsets				
Data Analysis				
Coefficient	0.39%	1.73%*	0.53%	1.79%***
(Standard error)	*(2.19%)*	*(0.94%)*	*(0.46%)*	*(0.31%)*
Information Technology				
Coefficient	8.70%***	-1.18%	-0.08%	0.16%
(Standard error)	*(1.69%)*	*(0.86%)*	*(0.44%)*	*(0.35%)*
Mobile Networks				
Coefficient	5.06%***	-0.15%	-0.64%	-0.80%**
(Standard error)	*(1.34%)*	*(0.81%)*	*(0.41%)*	*(0.38%)*
Software Engineering				
Coefficient	8.77%***	2.25%**	1.69%***	1.57%***
(Standard error)	*(2.34%)*	*(0.97%)*	*(0.44%)*	*(0.38%)*
Web Development				
Coefficient	2.73%	2.80%**	2.23%***	1.92%***
(Standard error)	*(2.67%)*	*(1.20%)*	*(0.34%)*	*(0.34%)*

Appendix A Additional Analyses

A.1 Robustness

In Table A.1 we evaluate the robustness of our main finding that higher shares of firm employees with technical and communication skills predict lower subsequent returns. Specifically, we estimate alternative specifications of the baseline regression (3) reported in Table 6. We consider alternative abnormal returns in column 1 (using raw excess returns) and column 2 (using one-factor Capital Asset Pricing Model alphas). We also estimate our main specification with the Fama and French three factor alphas but without any controls in column 3. The results are very consistent across these alternative specifications.

A.2 Employee Turnover and Firm Performance

We analyze the relationship between the stability of a firm's workforce and the firm's subsequent stock market returns. Firms with higher employee turnover perform significantly worse than firms with low turnover, controlling for other firm characteristics including size, book-to-market ratio, industry, and past performance.

We compute monthly firm turnover as the sum of departing and new incoming employees during the given month, scaled by the firm's total number of employees. In particular, consider firm i in each month t, with a total of $N_{i,t}$ employees. Let $Join_{i,t}$ denote the number of employees who join firm i during month t and $Depart_{i,t}$ denote the number of employees who leave firm i during month t. Then the turnover variable is defined as:

$$Turnover_{i,t} = \frac{Join_{i,t} + Depart_{i,t}}{N_{i,t}} \tag{7}$$

We winsorize this variable at the top and bottom 1% across all firm-months. Importantly, our measure captures total turnover – in both positive and negative directions – rather than directional hiring or outflows, as has been done in prior or contemporaneous work (Belo, Lin, and Bazdresch, 2014; Agrawal, Hacamo, and Hu, 2020).

In Table A.2, we investigate the relationship between abnormal turnover and subsequent firm performance by estimating a predictive regression of monthly stock returns on lagged turnover:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta TurnoverVar_{i,t-L} + X_{i,t} + \epsilon_{i,t} \tag{8}$$

We use Fama and French (1993) three factor alphas for abnormal returns in the main specification (columns 1 and 3) and also consider raw excess returns in columns 2 and 4. The controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, and past performance in columns 3 and 4, and we show that the results are robust to excluding these controls in columns 1 and 2. We vary the lag L between the measurement of employee skill composition and future firm returns between one month and six months.

Firms with higher turnover in a given month experience significantly lower stock market returns, starting after a lag of 2-3 months. For example, a 10% increase in a firm's employee turnover predicts a 52 basis points lower monthly Fama and French (1993) three factor alpha three months out (corresponding to an annual alpha of -6.42%) and a 31 points lower three factor alpha six months out (mapping to an annual alpha of -3.84%). These findings suggest the potential for the formation of profitable trading strategies based on employee turnover.

A.3 Heterogeneity across Industries

We explore heterogeneity in the firm-level skill premia across sectors of the economy, focusing on the three largest industries: Information, Manufacturing, and Finance & Insurance. We find generally consistent negative relationships between skilled employees and future returns across all three sectors.

The analysis in this section repeats the tests in specification (3) within the following three industry definitions: Information (two-digit NAICS code 51), Manufacturing (two-digit NAICS codes 31, 32, and 33), and Finance (two-digit NAICS code 52). Since the prevalence of skillsets changes gradually, and the full-sample results do not vary depending on the lag between the measurement of human capital composition and returns (see Table 6), we present all results in this subsection with a one month lag. The results are robust to extending the lag to two, three, and six months.

The results, presented in Table A.3, reveal ubiquitous negative links between skilled employees and returns, with some heterogeneity across sectors in exactly which skills are associated with the negative premia. For example, while abnormal firm-level focus on *Mobile Networks* has a negative and significant link with future returns across all three of the main industries, *Information Technology* has a directionally positive (although insignificant) return premium in the Financial & Insurance industry. Similarly, both *Digital Marketing* and *Social Media* carry negative premia in Information and Finance & Insurance, but not in Manufacturing, while *Client Relationship Management* is negatively predictive of future returns in both Information and Manufacturing, but not in Finance & Insurance.

Appendix B Company Name Disambiguation

We detail the methodology used to disambiguate employer names in the noisy self-reported resumes in order to match employees to official company names and stock symbols.

Individuals in the sample are not constrained in the names that they use to describe current and past employers. As such, any particular firm may be referenced by a variety of alternative names, and these names may be corrupted by misspellings, missing qualifiers, or employee misunderstandings. For example, employees of Banana Republic, the apparel brand, may refer to their employer as "Banana Republic," when in fact they are employees of "The GAP Inc.," and Banana Republic is one of several brands in the firm's portfolio. Similarly, the vast majority of employees of Alphabet Inc. list that they work for the main Alphabet subsidiary, Google. Furthermore, abbreviations ("GE", "IBM"), missing suffixes ("Inc.", "Corp."), and a variety of other inconsistencies complicate the problem of reliably linking records to companies. Panel 1 of Table B.1 offers examples of a few employee-reported company names matched to official company names and tickers.

We perform company name disambiguation using standard methods from entity disambiguation.[12] To evaluate the disambiguation procedure, we create a training dataset by manually tagging 1,000 employment records and matching them to official company names and market identifiers. For a comprehensive list of publicly traded companies and their stock tickers, we use stock symbols from NASDAQ (covering the NYSE, AMEX, and NASDAQ equity exchanges) matched to official and "trading as" names from Investor Guide (investorguide.com), CRSP, Wikipedia, and Google Finance. The disambiguation procedure is evaluated against the manually tagged training set along two dimensions: (i) precision (# correct matches / # employee-reported company names that get matched), which evaluates the extent to which the procedure avoids false positives; and (ii) recall (# correct matches / # employee-reported company names that should be matched), which captures the avoidance of false negatives. Precision and recall for the various stages of the disambiguation process are presented in Panel 2 of Table B.1.

We begin by computing a weak baseline similarity measure between an employee-reported company name and a candidate official company name using edit distance (Damerau, 1964). For each employee-reported company name that has at least one match with edit distance of 0.25 or lower, we take the match with the lowest distance. As can be seen in Panel 2 of Table B.1, this baseline yields a rather precise set of matches (the precision is 85%). However, the recall is quite poor, finding a match for only 14% of the employee-reported company names that should be matched. To increase the procedure's recall, we strip out a set of endings

[12] See Navigli (2009) for a survey on entity disambiguation.

that commonly appear in company names, such as “Inc.” and “L.P.” The list of common endings is compiled by taking the set of one-, two-, and three-word combinations at the end of the company names and cataloguing those endings that appear in the data more than 10 times. Running the edit distance matching procedure on the names stripped of common endings yields a precision of 82% and a recall of 63%, cataloged in Panel 2 of Table B.1.

We further augment the above procedure by processing each company name record to remove extraneous information (parenthetical statements, departments, locations, job titles, and descriptions of roles). To do this, we use a manually compiled list of common departments, job roles, and miscellaneous terms. In cases when multiple candidate strings appear in our database of canonical names, we favor longer matches. Furthermore, in order to increase precision, we compile an exhaustive list of company name aliases to limit the potential for erroneous matches based on edit distance. Using these additional steps in the matching procedure, where available, leads to an increase in precision to 94% and an increase in recall to 82%, as detailed in Panel 2 of Table B.1. This is the final methodology used in the analysis.

Table A.1: Robustness for panel regressions of monthly firm-level returns on lagged composition of employee skillsets. We estimate the following specification at four lags for each skillset j:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Skill^{j}_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ is the raw excess return in column 1, the CAPM alpha in column 2, and the Fama and French (1993) three factor alpha in column 3. $Skill^{j}_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the lagged month $t-L$. In columns 1 and 2, the controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, industry, and past performance. Column 3 is estimated without controls. The table reports the coefficient β corresponding to a 10% change in each skill variable.

Skillset	(1) excess return	(2) CAPM α	(3) three-factor α (no controls)
Technical Skillsets			
Data Analysis			
Coefficient	-0.03%	-0.01%	-0.22%*
(Standard error)	*(0.14%)*	*(0.14%)*	*(0.13%)*
Information Technology			
Coefficient	-0.01%	-0.04%	-0.11%**
(Standard error)	*(0.06%)*	*(0.07%)*	*(0.05%)*
Mobile Networks			
Coefficient	-0.09%***	-0.11%***	-0.12%***
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*
Software Engineering			
Coefficient	-0.01%	-0.07%*	-0.13%***
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.03%)*
Web Development			
Coefficient	-0.03%	-0.13%**	-0.24%***
(Standard error)	*(0.07%)*	*(0.07%)*	*(0.07%)*
Communication Skillsets			
CRM			
Coefficient	-0.03%	-0.08%**	-0.07%**
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.03%)*
Digital Marketing			
Coefficient	-0.18%***	-0.21%***	-0.16%***
(Standard error)	*(0.06%)*	*(0.06%)*	*(0.05%)*
Social Media			
Coefficient	-0.21%**	-0.24%***	-0.17%**
(Standard error)	*(0.09%)*	*(0.09%)*	*(0.08%)*

Table A.1 (Continued): Robustness for panel regressions of monthly firm-level returns on lagged composition of employee skillsets. We estimate the following specification at four lags for each skillset j:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Skill^j_{i,t-L} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ is the raw excess return in column 1, the CAPM alpha in column 2, and the Fama and French (1993) three factor alpha in column 3. $Skill^j_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the lagged month $t-L$. In columns 1 and 2, the controls $X_{i,t}$ include industry and month fixed effects, firm size, book-to-market ratio, industry, and past performance. Column 3 is estimated without controls. The table reports the coefficient β corresponding to a 10% change in each skill variable.

Skillset	(1) excess return	(2) CAPM α	(3) FF3 α (no controls)
Operational Skillsets			
Industrial Management			
Coefficient	0.16%**	0.23%**	0.22%***
(Standard error)	*(0.11%)*	*(0.11%)*	*(0.09%)*
Logistics			
Coefficient	0.06%	0.08%**	0.08%**
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.03%)*
Operations Management			
Coefficient	0.27%***	0.20%**	0.09%
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.08%)*
Sales			
Coefficient	0.05%*	0.08%**	0.10%***
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*
Sales Management			
Coefficient	-0.00%	0.12%	0.32%***
(Standard error)	*(0.09%)*	*(0.08%)*	*(0.07%)*
Tech. Product Mgmt			
Coefficient	0.11%***	0.09%**	0.03%
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*

Table A.2: Results from panel regressions of monthly firm-level returns on lagged employee turnover. We estimate the following specification at four lags:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta Turnover_{i,t-L} + X_{i,t} + \epsilon_{i,t}$
where $AbnRet_{i,t}$ is the Fama and French (1993) three factor alpha in columns 1 and 3 and the raw excess return in columns 2 and 4. $Turnover_{i,t-L}$ is firm i's employee turnover (computed as the sum of the number of employees joining and the number of employees departing the firm, scaled by total firm employees in the Cognism data) during the lagged month $t-L$. Columns 1 and 2 estimate the regression without controls, while columns 3 and 4 control for industry and month fixed effects, as well as firm size, book-to-market ratio, and past performance. The table reports the coefficient β corresponding to an additional 10% of employee turnover.

Lag	(1) FF3 alpha	(2) Excess return	(3) FF3 alpha	(4) Excess return
L = 1 month				
Coefficient	-0.23%**	-1.21%***	-0.16%	-0.09%
(Standard error)	*(0.09%)*	*(0.11%)*	*(0.11%)*	*(0.12%)*
Industry, time FE			X	X
Size, B/M			X	X
Past returns			X	X
L = 2 months				
Coefficient	-0.54%***	-1.10%***	-0.52%***	-0.43%***
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.11%)*	*(0.12%)*
Industry, time FE			X	X
Size, B/M			X	X
Past returns			X	X
L = 3 months				
Coefficient	-0.62%***	-1.20%***	-0.52%***	-0.42%***
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.12%)*	*(0.12%)*
Industry, time FE			X	X
Size, B/M			X	X
Past returns			X	X
L = 6 months				
Coefficient	-0.27%***	-0.79%***	-0.31%**	-0.24%**
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.12%)*	*(0.12%)*
Industry, time FE			X	X
Size, B/M			X	X
Past returns			X	X

Table A.3: Results from separate regressions of returns on lagged employee skillsets across three industries. We estimate the following specification separately for the Information industry (two-digit NAICS code 51), Manufacturing (two-digit NAICS codes 31-33), and Finance and Insurance (two-digit NAICS code 52):
$AbnRet_{i,t} = \alpha + \beta Skill^{j}_{i,t-1} + X_{i,t} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ is the return for firm i in month t in excess of that predicted by Fama and French (1993) factors. $Skill^{j}_{i,t-L}$ is the percentage of firm i's employees who possess skillset j in the previous month $t-1$. The controls $X_{i,t}$ include industry and month fixed effects, as well as firm size, book-to-market ratio, and past performance. The table reports the coefficient β corresponding to an additional 10% of employees possessing each skillset as either their primary or secondary skillset.

Skillset	Information (NAICS 51)	Manufacturing (NAICS 31-33)	Finance & Insurance (NAICS 52)
Technical Skillsets			
Data Analysis			
Coefficient	-0.07%	0.13%	-0.29%
(Standard error)	*(0.08%)*	*(0.15%)*	*(0.32%)*
Information Technology			
Coefficient	-0.05%	-0.12%*	0.28%
(Standard error)	*(0.05%)*	*(0.07%)*	*(0.17%)*
Mobile Networks			
Coefficient	-0.10%***	-0.14%***	-0.74%*
(Standard error)	*(0.03%)*	*(0.04%)*	*(0.47%)*
Software Engineering			
Coefficient	-0.07%*	-0.18%***	0.02%
(Standard error)	*(0.04%)*	*(0.06%)*	*(0.15%)*
Web Development			
Coefficient	-0.11%*	-0.11%	-0.19%
(Standard error)	*(0.07%)*	*(0.18%)*	*(0.27%)*
Communication Skillsets			
CRM			
Coefficient	-0.06%*	-0.23%***	0.31%
(Standard error)	*(0.03%)*	*(0.06%)*	*(0.21%)*
Digital Marketing			
Coefficient	-0.20%***	0.04%	-0.31%
(Standard error)	*(0.05%)*	*(0.13%)*	*(0.38%)*
Social Media			
Coefficient	-0.17%**	0.14%	-0.69%*
(Standard error)	*(0.07%)*	*(0.15%)*	*(0.42%)*

Table B.1: Public company name disambiguation. **Panel 1** displays examples of U.S. exchange-listed company names (listed on the NYSE, AMEX, or NASDAQ) from the employment data, matched to official company names and tickers. **Panel 2** presents the precision and recall of three company name disambiguation methods: (i) edit distance between employee-reported strings and official names; (ii) augmented by stripping out common endings such as "Inc." and "LP"; and (iii) further augmented by accounting for abbreviations, parenthetical statements, and noisy details on either side of a potential match.

Panel 1: Example Firm Disambiguations

Employer Name	Official Name	Ticker
Lehman (in administration)	Lehman Brothers	LEH
Ameris	Ameris Bancorp	ABCB
CLG	Cambium Learning Group, Inc.	ABCD
Advisory Board (US) Co.	The Advisory Board Company	ABCO
Abiomed	ABIOMED, Inc.	ABMD
Arbor	Arbor Realty Trust Inc.	ABR
Abbott Labs	Abbott Laboratories	ABT
Google	Alphabet Inc.	GOOG
TJ Watson Research Center	International Business Machines Corporation	IBM

Panel 2: Performance of Firm Disambiguation Procedure

Approach	Precision	Recall
Baseline	85%	14%
Strip Endings	82%	63%
Augmented Matching	**94%**	**82%**