Trading on Talent: Human Capital and Firm Performance*

Anastassia Fedyk† and James Hodson‡

July 30, 2019

Abstract

How does skilled human capital affect 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. Negative premia on social skillsets are likely driven by their risk profiles, as these skillsets display counter-cyclical performance. Meanwhile, negative premia on technical skillsets reflect patterns consistent with over-exuberance regarding contemporaneous popular technologies: IT and Mobile Network skillsets carry negative premia in early 2000s, while Data Analysis, Software Engineering, and Web Development display negative premia during the 2010s.

Keywords: return predictability, asset prices, human capital, skilled labor

*The authors are grateful to John Beshears, John Campbell, Lauren Cohen, David Laibson, Chris Malloy, Andrei Shleifer, Jeremy Stein, Wei Wu (discussant), Michael Weber (discussant), and seminar participants at Harvard University, Arrowstreet Capital, Cubist Systematic, PanAgora Asset Management, 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 sponsored by 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. Mail: 545 Student Services Building #1900, Berkeley, CA 94720. Email: fedyk@berkeley.edu.

‡Jožef Stefan International Postgraduate School, Artificial Intelligence Laboratory. Mail: Jamova 39, 1000 Ljubljana, Slovenia. Email: james.hodson@ijs.si.

1 Introduction

What is the value of a firm's human capital? We document that skillsets that traditionally carry a wage premium at the individual level fail to positively predict stock returns and operational variables at the firm level, showing an opposite effect. We focus on a set of technical skillsets such as *Software Engineering* and *Data Analysis*, and communication skillsets such as *Client Relationship Management* and *Social Media*. Abnormal concentration of a firm's employees in these skillsets forecasts negative future returns, supported by dampened future profitability (technical skillsets) or unexpected earnings (communication skillsets). In order to explore potential mechanisms behind these negative skill premia, we consider differential performance across time periods and market conditions. Communication skillsets perform relatively better during recession months and especially during the five-year period surrounding the Great Recession (2006-2010), indicating a potential risk-based explanation. Technical skillsets display patterns consistent with overvaluation of contemporaneous popular technologies. The main technical skillsets from the early 2000s (*Information Technology* and *Mobile Network*) see their negative premia concentrated in that time period, while more recently popular technical skillsets (*Data Analysis*, *Software Engineering*, and *Web Development*) display negative premia during the latest part of our sample, 2011-2016.

The increasing importance of human capital for modern firms has been stressed by practitioners and academics alike (see, for example, 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), mathematical 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 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 fields, does that predict

[1]See Bertrand and Schoar (2003), Malmendier and Tate (2005), Adams, Almeida, and Ferreira (2005), Malmendier and Tate (2008), Dyreng, Hanlon, and Maydew (2010), Graham, Harvey, and Puri (2013), and Perez-Gonzalez (2006), 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).

superior outcomes for the firm's operations or financial performance? Is there a premium for firms with more employees focused on social and communication skills? In this paper, 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 very distinct from the labor economics evidence on individual outcomes.

In order to characterize the skill focus of a given firm's workforce, we leverage techniques from machine learning on a large novel dataset of employment profiles. 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. Linked 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 at the end of our sample. This offers significantly broader coverage than what has been available from prior employment profile datasets, including LinkedIn (see Jeffers, 2017).

In addition to its coverage, a key advantage of our employment dataset is the inclusion of each employee's 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, across the 37 million profiles, the dataset features 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*), *Software Engineering* (most common skills: *java, sql, software development*), etc. 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." The most common skillset in our dataset is *Business Development*, followed closely by *Admin*, *Middle Management*, and *Banking & Finance*.

From the classification of individual employees into skillsets, we aggregate up to characterize the skill composition on the firm level. In particular, for each firm in each month, we compute the proportions of employees possessing each skillset. In order to screen out differences across industries and firm growth stages and account for known predictors of returns, we orthogonalize the skill proportions against firm characteristics such as size, book-to-market, industry, and past performance (firm momentum). This yields, for each firm in each month, a measure of abnormal focus on each skillset. Throughout our analysis, we focus on the five economy-wide skillsets classified as technical (*Data Analysis, Information Technology, Mobile Network, Software Engineering,* and *Web Development*) and the three skillsets classified as communication-based (*Client Relationship Management, Digital Marketing,* and *Social Media*). For comparison, 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*.

We show that firms with abnormal focus on technical and social skillsets earn systematically lower future abnormal returns. For example, for every additional 10% of employees having *Software Engineering* as either their primary or secondary skillset, the firm experiences a 7 basis points lower return in the following month (corresponding to -0.72% annual alpha), while every additional 10% of employees skilled in *Social Media* predict 21 basis points lower monthly returns (translating to -2.55% annual alpha). Instead, the skillsets that carry positive firm-level premia are core operational skills such as *Sales, Operations Management,* and *Industrial Management*.

The negative associations between firm performance and technical and communication skillsets are reinforced by our analysis of operational variables. Specifically, we find that most of these skillsets negatively forecast future operational performance, at different levels: technical skillsets tend to negatively predict even top-level items such as profitability, whereas communication skillsets increase profitability but (mostly) forecast lower standardized unexpected earnings. These patterns are consistent with 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 skill orthogonalization 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, with some evidence of better performance during recessions. The NBER recession sample is quite small, however, with only 26 months out of our 17-year period. 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. Technical skillsets do not display any analogous pattern, instead displaying more monotonic patterns over the three sub-periods.

Our final set of analyses supports the notion of the negative premia on technical skillsets reflecting over-pricing of currently popular skills. In particular, the three technical skillsets that have emerged as especially popular during the recent years – *Data Science*, *Software Engineering*, *Web Development* – all show strong negative relationships with future returns since 2010, but not during the 2000-2005 or 2006-2010 sub-periods. By contrast, *Information Technology* and *Mobile Network*, which were popular emerging skillsets during the early 2000s, had the strongest negative predictability for returns precisely during that period (2000-2005). Additional data on job openings and college graduates indicate support this pattern for the currently popular skillsets. 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 Infomation Science has risen by a modest 50% since 2010 (and are close to the same level as in 2005). The excess demand leads 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,[4] non-compete agreements,[5] hiring,[6] and demographics such as age and education.[7] 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 both on a large-scale (including overseas employees) and yet very granularly (taking advantage of specific individually-listed 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

[3]See, for example, Bertrand and Schoar (2003), Malmendier and Tate (2003), Galasso and Simcoe (2011), Kaplan, Klebanov, and Sorensen (2012), Yim (2013), Berger, Kick, and Schaeck (2014), and Benmelech and Frydman (2015).

[4]See Edmans (2011) and Guiso, Sapienza, and Zingales (2015)

[5]See, for example, Starr, Balasubramanian, and Sakakibara (2017) and Jeffers (2017).

[6]See Belo, Lin, and Bazdresch (2014) and Belo, Li, Lin, and Zho (2017).

[7]See Mukharlyamov (2016) and Kilic (2016).

individual level do not necessarily aggregate up to the level of the firm.

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 fact of the negative relationship between firm performance and technical and communication skillsets. Section 5 considers operational performance. Section 6 addresses potential mechanisms through sub-period analysis, and Section 7 concludes.

2 Data

Employment histories and demographic data are provided by Cognism, a global Client Relationship Management platform. Data on individual companies are merged with market data from the Center for Research in Security Prices and accounting data from Compustat. This section outlines the steps we take to structure the Cognism data and provides high-level information on coverage and descriptive mobility patterns.

2.1 Individual Profile Data

We observe the global employment market through the lens of a novel dataset of approximately 490 million individual employment and education records provided by Cognism, a platform for sales leads and customer relationship management. 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 has at least one active work experience record or education record between 1980 and 2019, 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 skills volunteered by the individual. We remove noisy profiles (profiles with improbable dates, too many empty fields, characters from outside the usual unicode ranges, etc.) from the analysis, leaving approximately 370 million profiles. For the current analysis, 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). This leaves us with 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.

Summary statistics of the demographic, educational, and employment characteristics of our sample are presented in Table 1. Most individuals in the sample are based in the U.S. (Panel 1). The average age is 36 years old, and the average (median) individual lists 3.6 (2) jobs (Panel 2). The average (median) number of skills is 10 (4), winzorized at the top and bottom 0.1%.

In order to explore the interplay of human capital and corporate performance, we process the individual profiles into a format that can be readily linked to financial data. The raw data contain a number of inconsistencies stemming from the diversity of unconstrained user input, and we take several steps to normalize and link the stated employer names to official trading names and relevant stock symbols for public companies (see Section 2.2).

Our main empirical analyses are based on human capital characteristics aggregated up from the individual employment profiles, which are then merged with industry classification, shares outstanding, revenues, cost of goods sold, assets, earnings, and book value from Compustat, as well as returns and market capitalization from CRSP. Over our primary sample period between 2000 and 2016, this yields an unbalanced panel with an average of 1,850 firms per month (from 1,175 at the start of the sample to 2,820 at the end), and a total of 3,094 firms (after pruning firms with fewer than 100 employees).

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 aggregates all online job openings based in the U.S. from over 40,000 sources including individual company websites and online job search websites.[8] 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 job

[8]This dataset is available to us through James Hodson's affiliation with Cognism, which has a restricted partnership with Burning Glass Technologies.

vacancies the U.S. economy.

2.2 Company Name Disambiguation

In this section, we detail the methodology used to disambiguate listed employer names and map them 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 2 offers examples of a few official company names matched to the listed names in the employment records.

We perform company name disambiguation using standard methods from entity disambiguation.[9] 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: (1) precision (# correct matches / # listed companies that get matched), which evaluates the extent to which the procedure avoids false positives; and (2) recall (# correct matches / # listed companies 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 2.

We begin by computing a weak baseline similarity measure between a listed company name and a candidate official company name using edit distance (see Damerau (1964)). For each listed 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 2, this baseline yields a highly precise set of matches (the precision is 85%). However, the recall is quite

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

poor, finding a match for only 14% of the listed company names that should be matched. To increase the procedure's recall, we strip out a set of endings 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 ends 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 2.

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 that we have determined to be extraneous to the matching procedure. 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 2. This is the final methodology used in the analysis.

We find that after cleaning and filtering, the dataset represents a comprehensive view of the employment landscape for a large proportion of the working population at U.S. public companies between 1990 and 2016.[10]

2.3 High-Level Descriptive Statistics on Labor Mobility

Although our primary analysis centers around public company employees, we validate our dataset on the human capital landscape by observing labor mobility patterns across the full economy. To do so, we also identify private company employees and link them to industry segments. This complete employment graph showcases a high degree of connectivity between the different industries via mobility of human capital.

Private company employers are more difficult to disambiguate than their publicly traded counterparts, since there are no comprehensive exchange lists for private companies. We make use of the Cognism's broad company database containing approximately 10 million global corporate entities, together with links to the OpenCorporates database, to identify candidate private firms in our data (following the same disambiguation procedure described in Section 2.2). For each identified firm that has a common industry classification schema

[10]Employee counts of NYSE, NASDAQ, and AMEX listed companies, as reported in 10Q filings to the SEC during 2016 provide an estimate of approximately 36 million global employees. As of May 2017, our coverage of currently active employees at these companies is approximately 14 million, representing 39%.

code, we provide a mapping into the NAICS taxonomy. Where no such code is available, we use Amazon's Mechanical Turk infrastructure to crowd-source 2-digit industry classifications. The total employment breakdown across industries (defined by two-digit NAICS industry codes) in 1996, 2006, and 2016 is presented in Table 3. Overall, we observe a number of industry shifts, including shrinking of the Manufacturing sector and growth in the Retail Trade sector.

The resulting industry codes across employers in our dataset give us the ability to understand migration patterns within and across the 2-digit NAICS classification over almost three decades of employment changes. 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 showcase 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 but less pertinent for more specialized industries such as Manufacturing and Finance & Insurance. In Appendix A, we also show that employee turnover is a meaningful predictor of performance at the level of individual firms: firms with higher monthly turnover show consistently lower returns up to six months later.

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 two steps. First, we use techniques from machine learning to condense hundreds of thousands of individually-entered skills into a 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 skiills

3.1 Identifying Skillsets

We use the Latent Dirichlet Allocation method for identifying topics in documents to identify common skillsets from the individuals' self-reported lists of skills.[11] We recover forty-four skillsets, which capture intuitive competencies ranging from product management to web development to healthcare professionals.

[11] For further details on Latent Dirichlet Allocation method, please see Blei, Ng, and Jordan (2003).

We begin by representing 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 profile. For example, an element d may be the following set: $d = \{\textit{MicrosoftOffice}, \textit{Melhoriacontinua}\}$. Let the vocabulary W consist of all skills that appear in at least one profile – for example, if there is a profile $d = \{MicrosoftOffice, Melhoriacontinua\}$ in the dataset, then *Microsoft Office* will appear in W, as will *Melhoria continua*. The size of the vocabulary W is denoted by V. We represent the elements d as unit vectors in the V-dimensional space of vocabulary terms.

The premise is that the observed profiles in D are generated from a latent set of skillsets, which we denote by T. Each element $t \in T$ is a unit vector in the k-dimensional skillset space, where the parameter k is the desired number of skillsets, specified by the researcher. In this paper, we consider $k = 44$ skillsets. We arrive at this number empirically by considering goodness-of-fit measures across values of $k \in [25, 60]$.

The Latent Dirichlet Allocation algorithm conceptualizes each individual's reported skills d as a sequence of terms drawn from a latent distribution $\mathcal{D}_d$ over the $k = 44$ possible skillsets. The distribution $\mathcal{D}_d$ is itself randomly determined for each individual: in particular, for each individual set of skills d, $\mathcal{D}_d$ is a multinomial distribution whose parameters are a random variable drawn from a pre-specified Dirichlet prior.

The generative process assumed by the Latent Dirichlet Allocation algorithm is as follows.

- Pre-specify model parameters: ξ, α, β.
- To construct each new profile of skills d:
  1. Choose the number of skills $N_d \sim Poisson(\xi)$.
  2. Choose a distribution over skillsets $\theta_d \in Dir(\alpha)$.
  3. Fill the N_d skills in the profile d by sequentially choosing each skill w_n as follows:
     (a) Choose a skillset $t_n \sim Multinomial(\theta_d)$.
     (b) Choose a skill w_n from $\mathbb{P}\{w_n | t_n, \beta\}$, the conditional probability distribution over skills in the vocabulary, conditional on the chosen skillset t_n.

We estimate the key parameters of the model, α and β, using a collapsed Gibbs sampling algorithm. Table 4 presents the results. For each of the $k = 44$ skillsets, we display the most frequent skills from that skillset – i.e., for each t_i, we list the terms w_j with the highest estimated values of $\hat{\beta}_{i,j}$.

The identified skillsets 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. 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 the economy, others – including *Sales*, *Information Technology*, and *Business Development* – 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 skills that have been highlighted as carrying wage and career premia at the individual level: mathematical (technical) skills and soft (communication) skills. For comparison, we also consider core operational skillsets such as *Sales* and *Operations Management*. The remaining excluded category consists of bureaucratic skills such as *Human Resources*, *Admin*, and *Middle Management* using generic tools such as the Microsoft Office suite (grouped into the set of administrative skillsets below).

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

- Technical skillsets: *Data Analysis; Information Technology; Mobile Network; 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 Employee Skillsets

Having identified the skillsets, we classify each employee as possessing the two skillsets that are most likely to have generated his particular combination of self-reported skills. The most prevalent skillsets across the 37 million employees of public U.S. companies are

Middle Management and *Business Development*, but there is substantial heterogeneity in skills across companies in different industries.

We classify individual employees into skillsets as follows. For each individual i with a non-empty listed skill profile d_i, we take the estimated distribution over skillsets, $\hat{\theta}_i$. Let $j(i)$ and $j'(i)$ denote the indices of the two largest values in the vector $\hat{\theta}_i$. Then individual i is deemed to possess the skills from the skillsets $j(i)$ (his primary skillset) and $j'(i)$ (his secondary skillset). That is, each individual is assigned to the two skillsets that best match his or her profile.

The most common skillsets are *Middle Management, Business Development*, and *Admin*, as can be seen from Figure 3. Other common skillsets include *Banking & Finance, Technical Product Management*, and *Manufacturing*. Skillsets covering *Education, Legal*, and *Personal Coaching* are the least common in our sample of public firm employees.

The landscape of skillsets looks quite different across industries. We provide two examples in Figures 4 and 5, which display the distribution of skillsets in Manufacturing (two-digit NAICS codes 31, 32, and 33) and Finance & Insurance (two-digit NAICS code 52). As can be seen from Figure 4, 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* ad *Business Development* and (see Figure 5). Different industries not only feature different compositions of skilled human capital, but also see the skillsets differently priced by the financial markets. We explore this heterogeneity in detail in Section 6.3.

4 Employee Skillsets and Firm Performance

We consider the relationship between firm performance and the distribution of skillsets possessed by the firm's employees. Contrary to the individual-level evidence, we find that firms with higher proportions of employees skilled in technical fields and communication skillsets tend to earn lower returns. Instead, higher alphas are seen for firms that focus more on core business areas such as *Sales* and *Operations management*.

4.1 Firm-Level Skill Measures

We evaluate the extent to which different skillsets are associated with differential firm performance by computing the prevalence of each skillset inside each firm's employee base, orthogonalized relative to other firm characteristics. We use lagged values of each skillset variable to predict the firm's future returns.

For each skillset, we begin by computing the percentage of employees whose employment profiles indicate that skillset as either their primary or their secondary skillset. Take a skillset j, and consider a firm i in each month t with a total of $N_{i,t}$ employees. Let $Skill^j_{i,t}$ denote the number of employees who are employed at firm i during month t and for whom the skillset j is either primary or secondary. We winzorize this variable at the top and bottom 1% across all firm-months.

In order to ensure that our measures of skillset prevalence are not affected by firm lifecycle stage, industry, or factors that have been shown to affect performance, we proceed as follows. 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 characteristics $X_{i,t}$ includes log market capitalization $logMC_{i,t}$, book-to-market ratio $BM_{i,t}$, industry dummies $Ind^k_{i,t}$ (where each k is a two-digit NAICS industry code), and momentum $Mom_{i,t}$, 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), i.e., the component of skillset prevalence that is unexplained by firm characteristics:

$$AbnSkill^j_{i,t} = Skill^j_{i,t} - PredSkill^j_{i,t}, \tag{2}$$

where $PredSkill^j_{i,t}$ is the predicted value for the percentage of firm i's employees with skillset j in month t from fitting specification (1). All results are also robust (and quantitatively larger) to using raw skill measures, without orthogonalizing to controls.

4.2 Firm-Level Skill Premia

To estimate the effect on firm performance from focusing on a given skillset j, we regress abnormal returns on lagged values of abnormal prevalence of skillset j. For abnormal returns, we use Fama and French (1993) three factor alphas. In untabulated results, we confirm that our findings are robust to alternative return constructions, including using raw excess returns.

Thus, for each skillset j, we estimate the following specification at four different lags:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : FF3alpha_{i,t} = \alpha + \beta AbnSkill^j_{i,t} + \epsilon_{i,t}, \tag{3}$$

where $FF3alpha_{i,t}$ is the difference between firm i's return during month t and the predicted return from the three Fama-French factors (market, size, and book-to-market ratio).

Table 5 reports the results for three groups of skillsets: (i) technical skillsets, following the evidence on the individual-level premia for technical and mathematical skills; (ii) communication skillsets, prompted by the individual-level premia on communication skillsets; and (iii) operational skillsets, as a group of comparison. All results 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%. We consider four different lags, looking at returns one month, two months, three months, and six months after the measurement of employee skills, displayed across the four columns of Table 5. Consistent with the composition of the workforce changing relatively slowly, the results are fairly similar across the lags.

Our key finding is that neither the technical skillsets nor the communication skillsets positively forecast future returns at the firm-level. In fact, both technical and communication skillsets are associated with negative future financial performance. 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 Network* 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 every additional 10% of employees (out of a total of 200%, accounting for each employee having both a primary and a secondary skillset) skilled in *Web Development*, 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 seems a 0.21% (2.55%) lower monthly (annual) alpha.

Instead, the skills that do display positive return predictability are operational skillsets that combine a less concentrated mix of technical and communication abilities to accomplish core business activities. *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, but not statistically significant. For example, for every additional 10% of employees skilled in *Operations Management* as either their primary or secondary skillset, the firm sees an additional 0.18% monthly (2.18% annual) alpha.

In Appendix A.2, 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 relationship between technically skilled employees and future returns is quite robust across the sectors of the economy, while communication skillsets carry differential premia in different industries, with both Manufacturing and Finance & Insurance having a less consistent negative relationship between future returns and these skills than Information.

In the next two sections, we explore the negative return premia on technical and communication skillsets further, by observing the link between these skillsets and firm operations and the variation of the negative premia across time.

5 Employee Skillsets and Firm Operations

We confirm that the negative return premia associated with technical and communication skillsets at the firm level are supported by the response of firm operations to corresponding changes in employee skill composition. We find that increases in technical focus, at the firm level, are associated with reduced future gross profitability. Communication skills are positively related to top-line profitability, but decrease earnings further down, resulting in a negative relationship between inflows of employees with communication skills and future earnings surprises.

In order to focus on changes in the employee composition, we define $SkillChange^j_{i,t}$ for firm i and skill j as the percentage difference in employees of firm i whose primary or secondary skillset is j over the course of quarter t. $AbnSkillChange^j_{i,t}$ is then computed by orthogonolizing $SkillChange^j_{i,t}$ to firm characteristics: industry, log market capitalization, and book-to-market ratio. Since changes in skill compositions can be quite noisy across firms and from month to month, we standardize each $AbnormalSkillChange^j$ variable by winsorizing at the top and bottom 1% and normalizing 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. In particular, we look at how quarterly changes in employee skill focus forecast future quarters' gross profitability, defined as the difference between revenues and cost of goods sold, scaled by assets. Specifically, we estimate the following predictive regression:

$$\text{For } L \in \{1\,q, 2\,q, 3\,q, 6\,q\} : Profitability_{i,t} = \alpha + \beta AbnSkillChange^j_{i,t} + \epsilon_{i,t}, \tag{4}$$

for lags varying from one to four quarters.

The results, which are presented in Table 6, support the negative impact of technical skills, but indicate that the negative premia on communication skills is not driven by decreases in top-level operations such as profitability. Increases in employee focus on most technical skills, with the exceptions of *Information Technology* and *Mobile Network*, negatively forecast future firm profitability, especially at longer horizons. The impact is immediate for increased focus on *Data Analysis*, which has a large and significant negative association with profitability throughout the following four quarters. The negative link between increased focus on *Web Development* and to a lesser extent *Software Engineering* is also present but delayed: most of these effects are noticeable starting 2-3 quarters out.

By contrast, communication skillsets such as *Client Relationship Management* (and to a lesser extent *Digital Media* and *Social Media*) are positively linked to gross profitability. An influx of employees skilled in these areas tends to forecast increased future revenues.

However, two of the three communication skillsets have a negative impact on downstream earnings, which we document by considering the link between increases in firm-level skill concentrations and future earnings surprises. Following the literature[12], we look at standardized unexpected earnings (SUE). $SUE_{i,t}$ 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 the quarterly change in the percentage of employees skilled in each of the six technical and three communication skillsets and future quarterly unexpected earnings through the following specification:

$$\text{For } L \in \{1\,q, 2\,q, 3\,q, 6\,q\} : SUE_{i,t} = \alpha + \beta AbnSkillChange^{j}_{i,t} + \epsilon_{i,t}, \tag{5}$$

for lags varying from one to four quarters. As before, each $AbnormalSkillChange^{j}$ variable is computed by orthogonolizing firm i's quarterly change in employees with skill j to firm-level controls (industry, log market capitalization, and book-to-market ratio), winsorizing at the top and bottom 1%, and normalizing to mean zero and standard deviation one.

The results, displayed in Table 7, indicate that both *Digital Marketing* and *Social Media* indeed negatively impact downstream earnings, albeit with a time lag. Increases in the percentage of employees possessing these skillsets forecast lower standardized unexpected earnings three or four quarters out. Most of the technical skillsets (with the exception of *Mobile Network*) are also negatively associated with future SUE, in addition to their negative relationship with future firm profitability.

Interestingly, *Client Relationship Management* is strongly positively predictive of both

[12] See, for example, Da, Gurun, Warachka (2014) and So and Wang (2014), among others.

future profitability and future unexpected earnings, but has a statistically insignificant negative relationship with future returns. As we show in the next section, the weak negative premium (or rather, the lack of a positive premium) on the *Client Relationship Management* skillset overall may be driven by its risk profile: this skillset is counter-cyclical, with a significant positive premium in recessions.

6 Potential Mechanisms for Negative Skill Premia

In this section, we further explore the negative firm-level premia on technical and communication skillsets through two additional analyses: partitioning the sample based on market conditions (U.S. recessions) and observing time trends over the course of our sixteen-year sample period.

The analysis across market conditions supports the notion of communication skillsets capturing counter-cyclical performance, with the lower returns reflecting a reduction in risk. The negative premia on technical skillsets, by contrast, is likely to be driven by over-excitement about emerging technologies: these skillsets tend to have the strongest negative associations with returns precisely during the sub-periods when they are first introduced and most in demand by employers.

6.1 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 (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) – 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. The results are robust to extending the lag to two, three, and six months.

Both technical and communication skillsets carry robustly negative 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 a full 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 (less 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. *Software Engineering* also has a positive point estimate, albeit insignificant in the small 26-month sample, and *Mobile Network*, *Web Development*, and *Social Media* all have imprecisely estimated zero coefficients.

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 8 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 2010-2016 sub-periods. However, all three of these skillsets show zero (or weakly positive) predictability for returns during the five-year subsample around the Great Recession, 2006-2010. This counter-cyclical pattern is not observed for any of the technical skillsets, which instead display time trends consistent with the corresponding popularity of each new skill – we explore this idea further in the next subsection.

6.2 Time Trends and Technical Skill Popularity

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. 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 Network*) 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 2010 to 2016 – and reestimate specification (3) within each of these sub-periods.

The results, displayed in Table 8, confirm our prediction that each technical skillset carries a negative premium precisely during the time of its popularity. Both *Information Technology* and *Mobile Network* had a significant negative relationship with 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 predicted 23 basis points (2.80%) lower monthly (annual) returns during the 2000-2005 period. The effect size for employees skilled in *Mobile Network* was similar, at -20 basis points per month (significant at the 1% level). Since then, the link between these two skillsets and returns has weakened substantially, approaching zero in the recent years. By contrast, the newly emerging skillsets *Data Analysis, Software Engineering*, and *Web Development* display the opposite pattern: these skillsets have insignificant (and even sometimes directionally positive) premia during the two earlier samples, 2000-2005 and 2006-2010. However, during the time when these skillsets became especially popular, the 2010s, the association between focus on these skillsets and subsequent performance has become significantly negative. An additional 10% of employees having *Web Development* as their primary or secondary skillset corresponds to 25 basis points (3.04%) lower monthly (annual) returns, significant at the 1% level. The effect sizes for *Software Engineering* and *Data Analysis* are milder at -12 basis points and -15 basis points, respectively, but still sizable and statistically significant.

We also look at the time trends in job openings and available talent (college graduates with corresponding majors) to further support our interpretation of the negative premia on technical skills reflecting over-exuberance about popular skillsets. 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.[13] The total annual job postings numbers, reported weekly, triples 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 has been stable throughout the 2010s.[14]

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 and before (we display the pre-trend of the number of graduates starting five years earlier, in 2005). The time series of Computer Science and Information Science majors ends in 2015. Beginning in 2016, the Digest of Education Statistics from NCES aggregates Computer Science with non-computer engineering, and we therefore leave 2016 out of the figure. The annual numbers of U.S. students graduating with bachelor's degrees in Computer Science and Information Science, computed from the Integrated Postsecondary Education Data System survey data, are plotted in red in Figure 8, with the axis on the left. These numbers vary from approximately 40,000 to 60,000 graduates per year, and the increase is not nearly sufficient to fill the rise in job openings. As firms compete for the limited pool of qualified technical employees, we are likely to observe not only rising wages but also less qualified candidates filling the same roles corresponding to these over-hyped skillsets, in order to meet demand. This can generate overpricing of the popular skillsets and correspondingly worse subsequent firm operations and returns.

7 Conclusion

We document that the employee skill composition of a firm can have important repercussions for its performance in financial markets, but in a substantially different pattern from

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

[14] For example, Liu and Wu, 2018, show that the number of publicly traded firms represented in Burning Glass data has remained unchanged from 2010 onwards

individual-level skill premia established by the labor economics literature. Specifically, technical skillsets and communication skillsets, both of which have been found to carry positive wage premia for individuals, 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, appears to capture over-excitement about novel emerging technologies at the height of their popularity.

Our work bridges together two literatures. On the one hand, we contrast and build upon a 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 do not necessarily aggregate up to analogous premia at the firm level.

References

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

[2] 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.

[3] Belo, Frederico, J. Li, X. Lin, and X. Zhao (2017). "Labor-force heterogeneity and asset prices: The importance of skilled labor." *The Review of Financial Studies,* forthcoming.

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

[5] Berger, A. N., T. Kick, and K. Schaeck (2014). "Executive board composition and bank risk taking." *Journal of Corporate Finance*, 28: 48-65.

[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] Da, Z., U. G. Gurun, and M. Warachka (2014). "Frog in the pan: Continuous information and momentum." The review of financial studies 27.7: 2171-2218.

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

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

[11] Donangelo, A. (2014). "Labor mobility: Implications for asset pricing." *Journal of Finance,* 69.3: 1321?1346.

[12] Dyreng, S. D., M. Hanlon, and E. L. Maydew (2010). "The effects of executives on corporate tax avoidance." *The Accounting Review,* 85.4: 1163-1189.

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

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

[15] Fama, Eugene F, and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33.1: 3?56.

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

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

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

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

[20] Jegadeesh, N., and S. Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance,* 48.1: 65?91.

[21] Jeffers, J. S. (2017). "The Impact of Restricting Labor Mobility on Corporate Investment and Entrepreneurship." Working Paper.

[22] 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.

[23] Liu, Y. and X. Wu (2018). "Labor Market Competitor Network and the Transmission of Shocks."

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

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

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

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

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

[29] Moskowitz, Tobias J, and Mark Grinblatt (1999). "Do Industries Explain Momentum?" *Journal of Finance,* 54.4: 1249?1290.

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

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

[32] Perez-Gonzalez, F. (2006). "Inherited control and firm performance." *American Economic Review,* 96.5: 1559-1588.

[33] Singhal, A. (2012). "Introducing the Knowledge Graph: Things, Not Strings". Official Blog (of Google).

[34] Starr, E., N. Balasubramanian, and M. Sakakibara (2017). "Screening spinouts? How noncompete enforceability affects the creation, growth, and survival of new firms." *Management Science,* forthcoming.

[35] 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.

[36] Vrandecic, D., and M. Krötzsch (2014). "Wikidata: a free collaborative knowledgebase." *Communications of the ACM*, 57.10: 78-85.

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

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

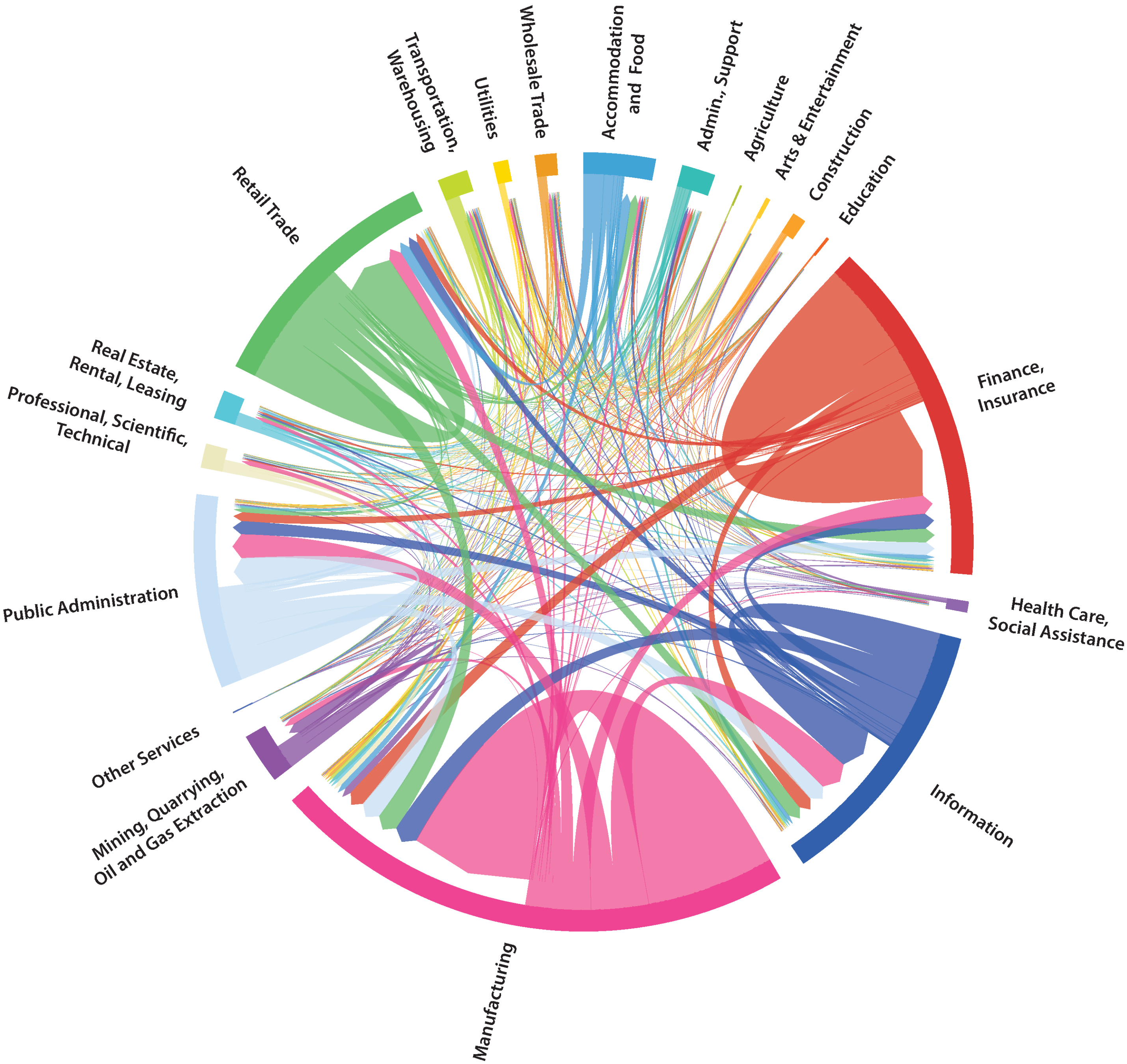

Figure 1: Cross-industry employee moves during the years 2010-2017. We consider all job changes that involve switching employer. Each arrows captures the prevalence of moves from one industry to another, following two-digit NAICS industry classification.

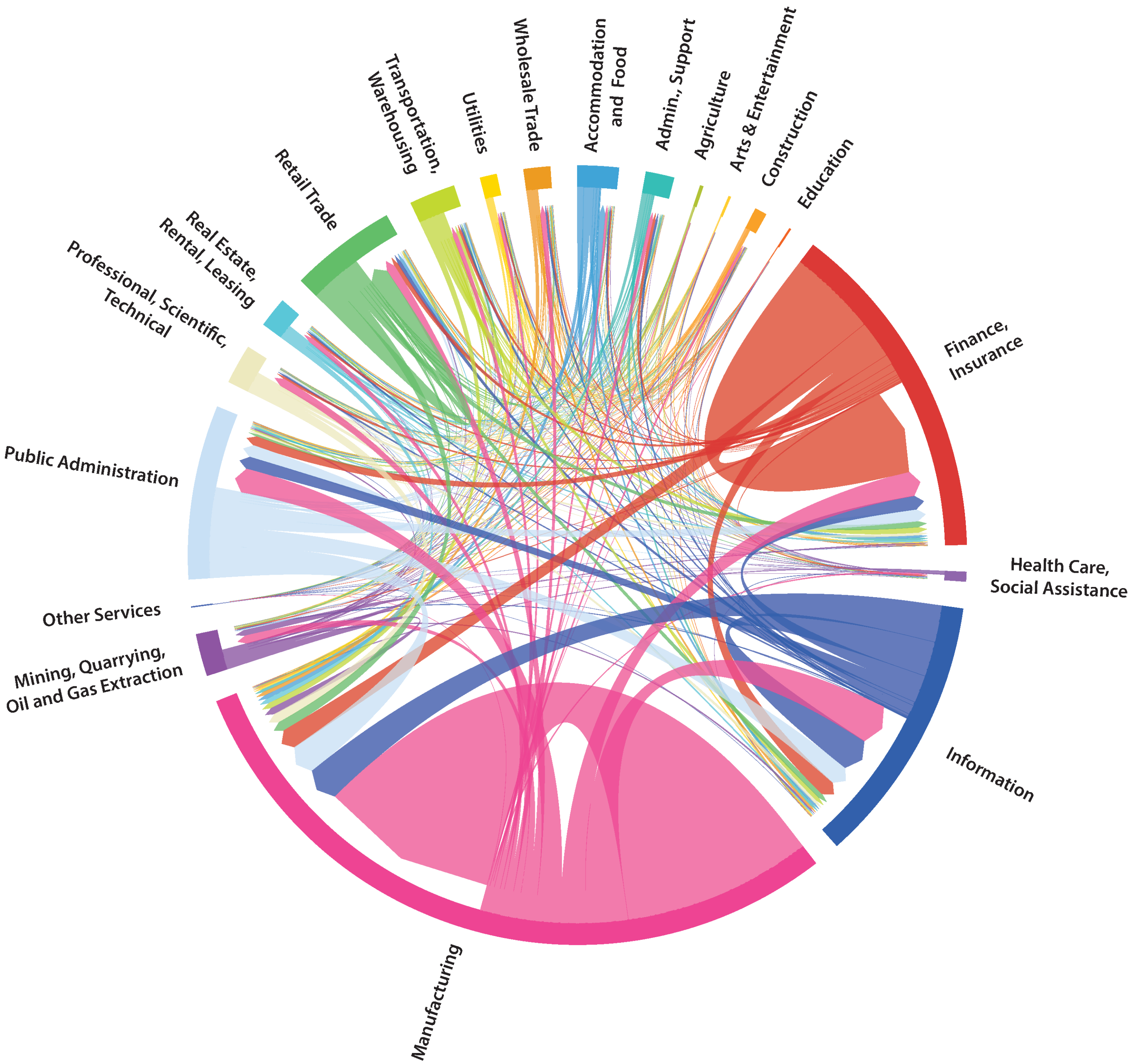

Figure 2: Cross-industry employee moves during the 1990s. We consider all job changes that involve switching employer. Each arrows captures the prevalence of moves from one industry to another, following two-digit NAICS industry classification.

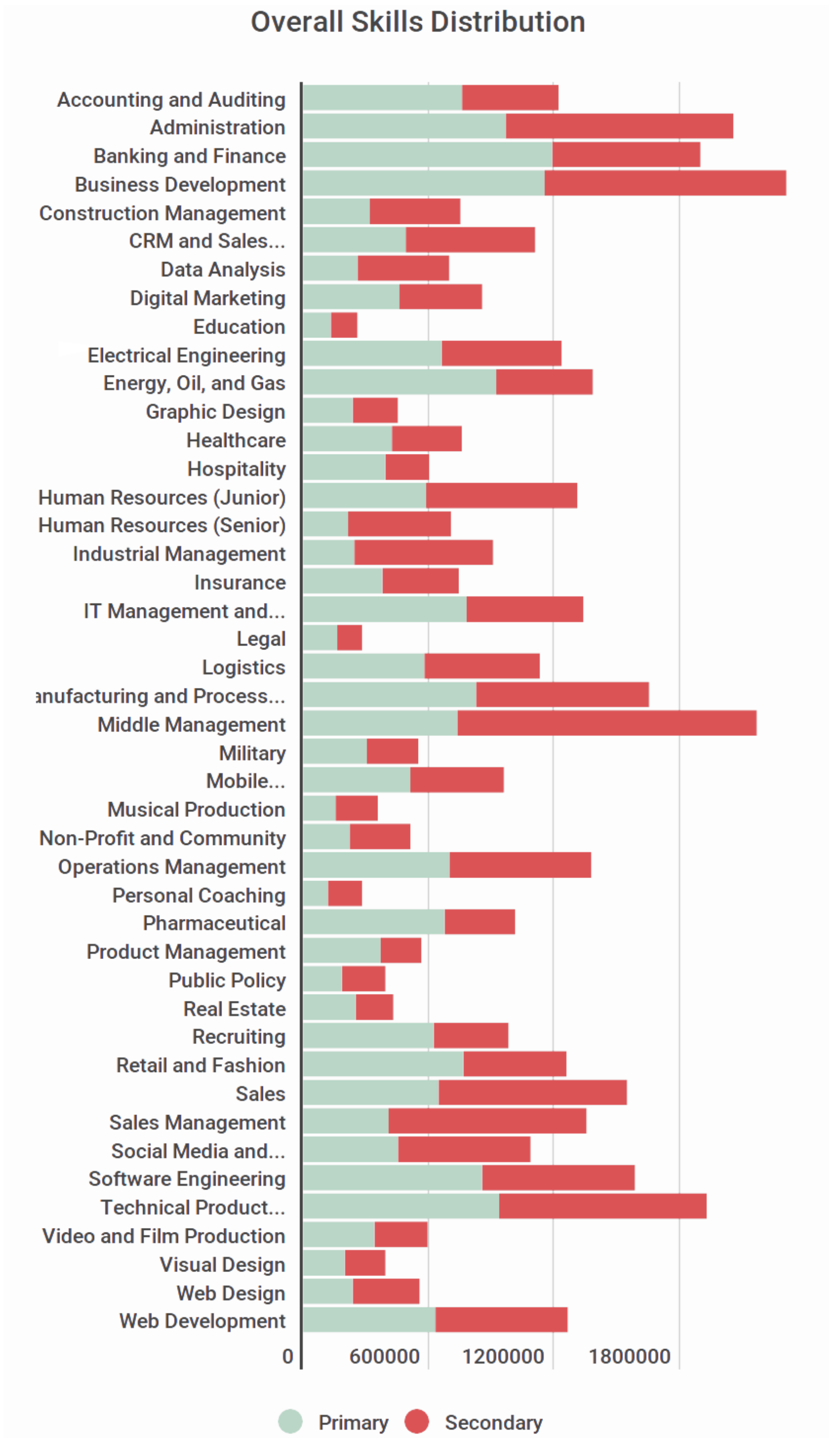

Figure 3: Frequency of skillsets across the full population of 37 million employees of public U.S. companies. Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the number of employees for whom that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

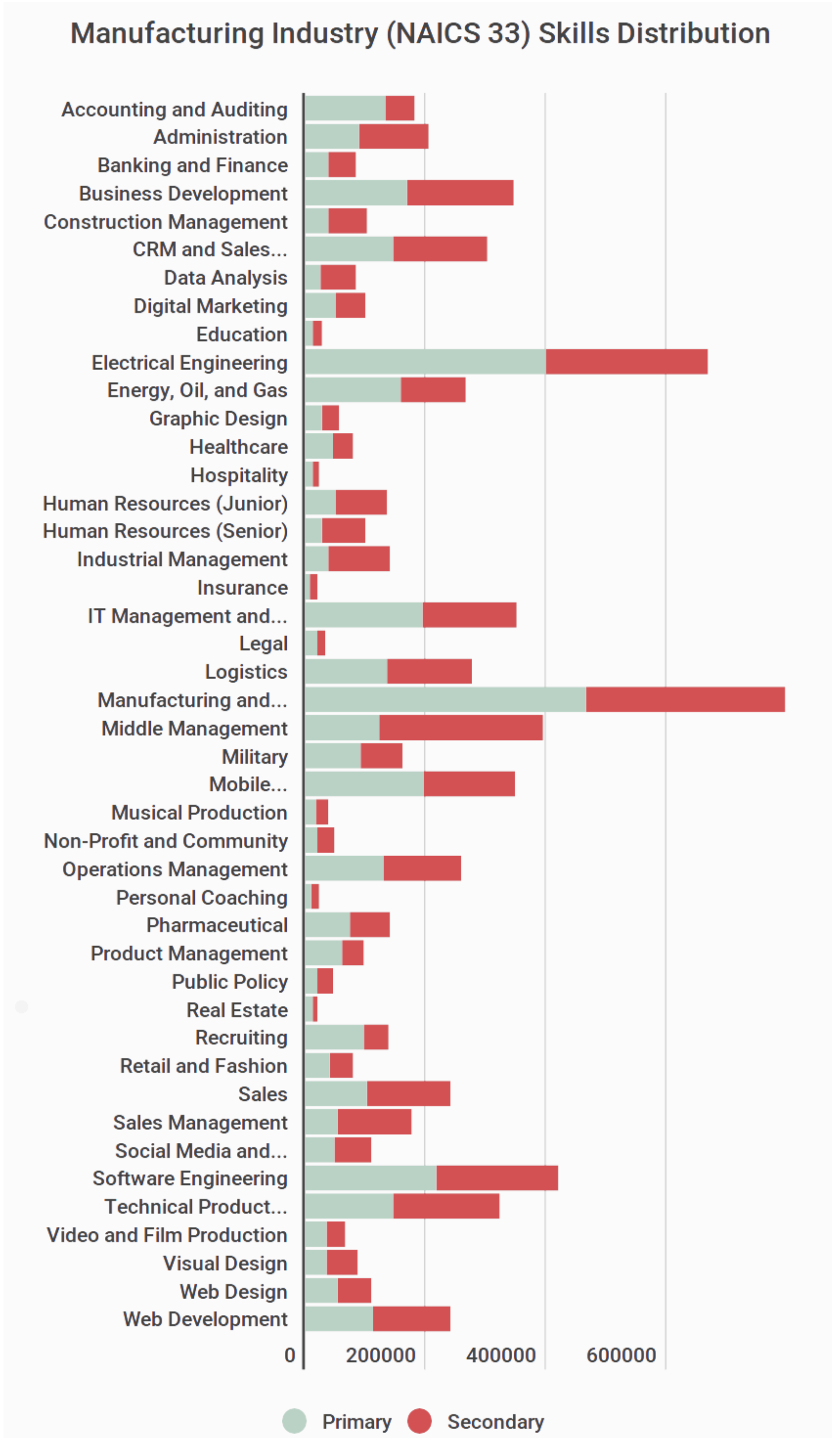

Figure 4: Frequency of skillsets across the employees of public U.S. companies in the Manufacturing industry (two-digit NAICS codes 31, 32, and 33). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows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

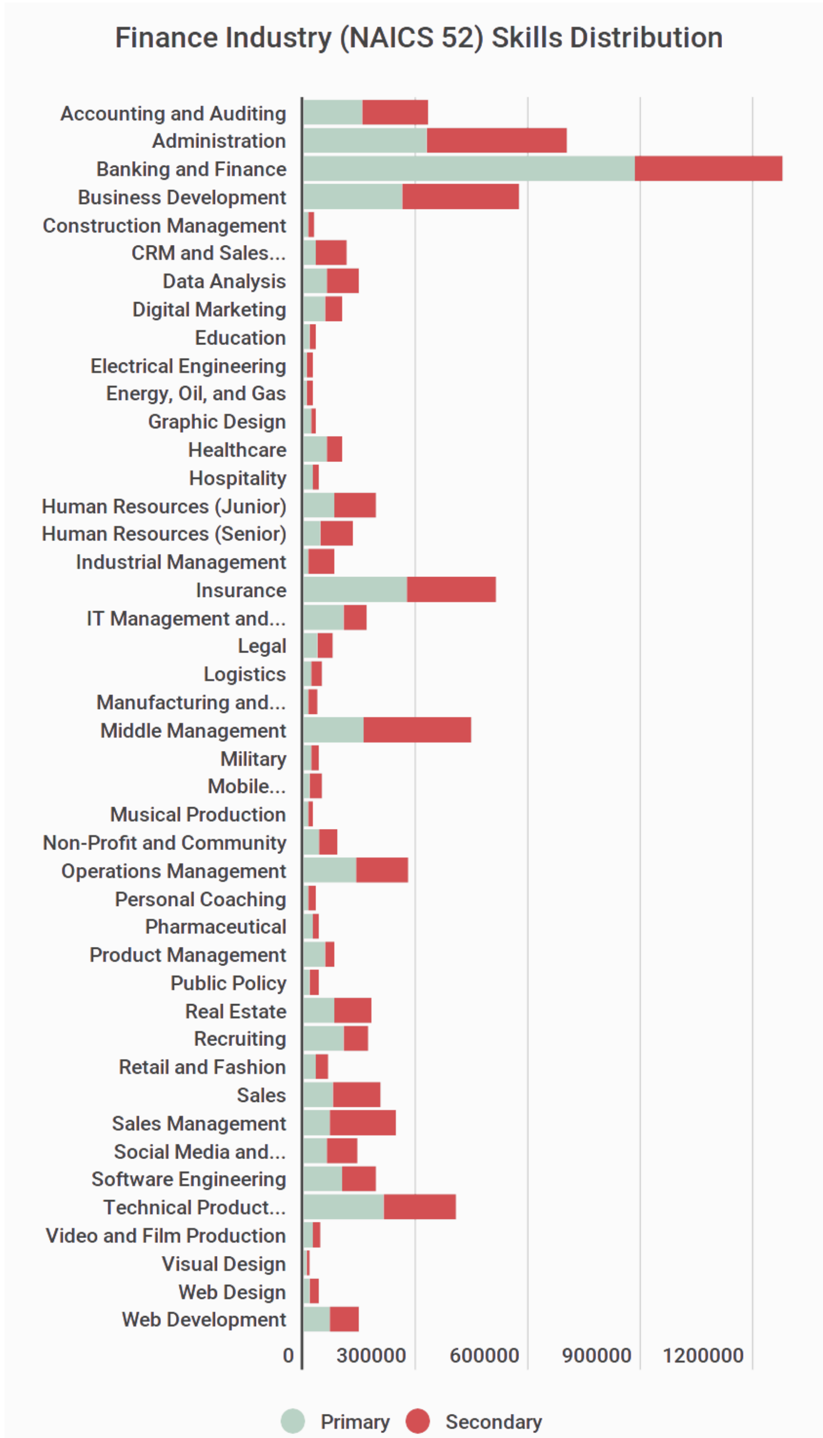

Figure 5: Frequency of skillsets across the employees of public U.S. companies in the Finance and Insurance industry (two-digit NAICS code 52). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the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

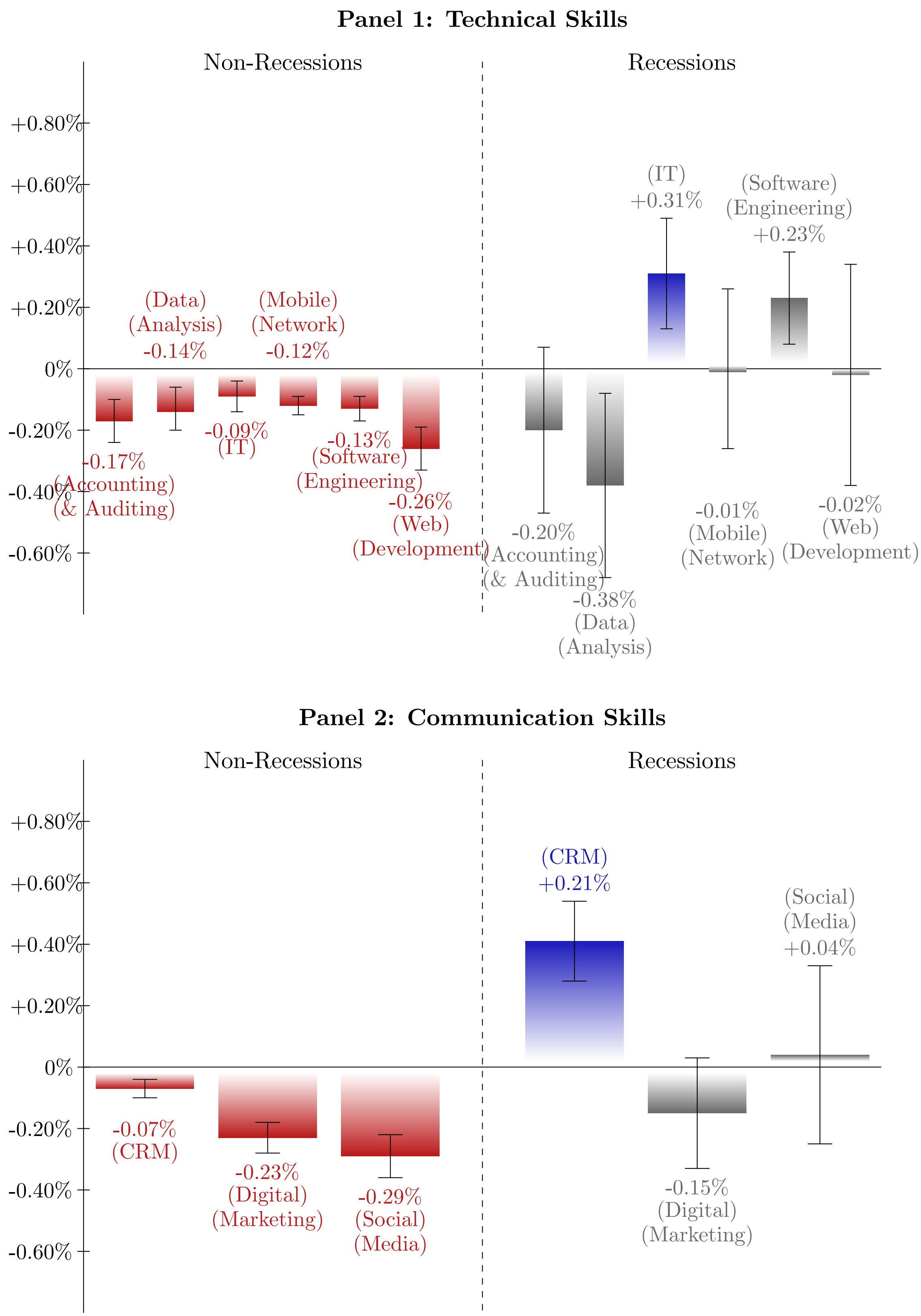

Figure 6: Relationship between skill composition and firm performance across market conditions. Panel 1 looks at technical skillsets, while Panel 2 considers communication skillsets.

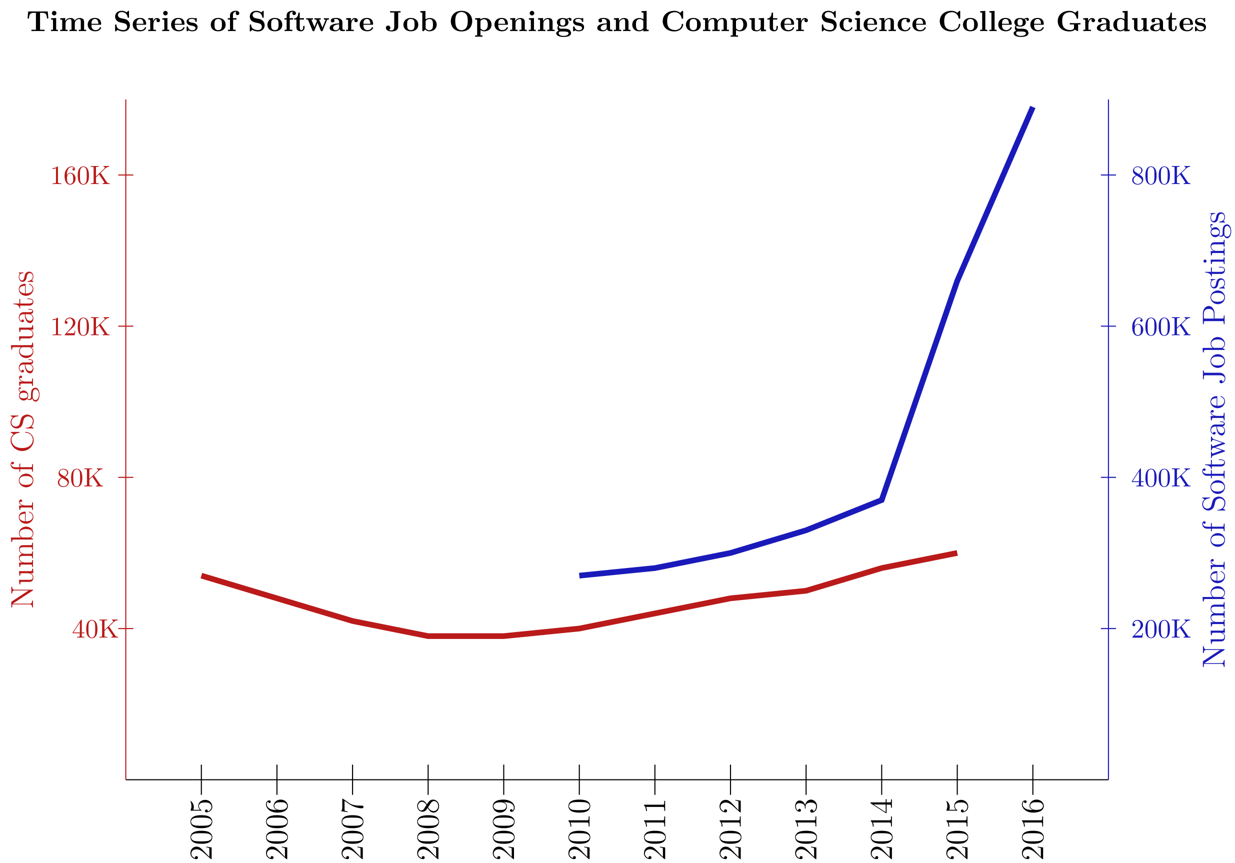

Figure 7: Time series of the number of software-related job openings (plotted in blue, axis on the right) and college graduates majoring in computer science or information science (in red, axis on the left). Job posting data is only available since 2010. Data on college graduates majoring specifically in computer science is available from the NCES Digest of Education Statistics through 2015.

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 (among users for whom age is calculated), jobs held, connections, and skills. **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 winzorized at top and bottom 0.1%.

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

[c]Numbers of skills are winzorized 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%
Elite institution	5.66%

Table 2: Public company name disambiguation. **Panel 1** presents examples of U.S. exchange-listed company names (NYSE and 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: (1) edit distance between listed strings and official names; (2) augmented by stripping out common endings such as "Inc." and "LP"; and (3) further augmented by accounting for abbreviations, parenthetical statements, and noisy details on either side of a potential match. **Panel 3** shows the proportion of employer entities listed on the major U.S. public exchanges (excluding OTC markets) that are covered by the data. Employer entities exclude funds and multiple share class listings.

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%**

Table 3: Employment by NAICS two-digit industry code, as of January 1, 2016, January 1, 2006, and January 1, 1996.

Panel 1: NAICS

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

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 Network	telecommunications, wireless, voip, networking, ip
Software Engineering	java, sql, software development, linux, agile methodologies
Web Development	javascript, html, java, css, sql
Communication Skillsets	
CRM	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 4 (Continued): Labeled skillsets extracted from the self-reported skills and the five most likely terms to appear conditional on each skillsets.

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 5: Results from panel regressions of monthly firm-level stock market returns on lagged composition of employee skillsets. We estimate the following specification for four lags:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : FF3alpha_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$,
where $FF3alpha_{i,t}$ the return for firm i in month t in excess of that predicted by Fama and French (1993) factors. $AbnSkill_{i,t-L}$, is the percentage of firm i's employees who possess a given skillset in the lagged month $t-L$, orthogonalized to size, book-to-market ratio, industry, and momentum of firm i. The table reports the coefficient β corresponding to 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)	*(0.08%)*	*(0.08%)*	*(0.08%)*	*(0.08%)*
Information Technology				
Coefficient	-0.05%	-0.04%	-0.03%	-0.03%
(Standard error)	*(0.05%)*	*(0.05%)*	*(0.05%)*	*(0.05%)*
Mobile Network				
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)	*(0.05%)*	*(0.05%)*	*(0.05%)*	*(0.05%)*
Social Media				
Coefficient	-0.17%**	-0.16%**	-0.16%**	-0.14%**
(Standard error)	*(0.07%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*

Table 5 (Continued): Results from panel regressions of monthly firm-level stock market returns on lagged composition of employee skillsets. We estimate the following specification for four lags: For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\}$: $FF3alpha_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$,
where $FF3alpha_{i,t}$ is the return for firm i in month t in excess of that predicted by Fama and French (1993) factors. $AbnSkill_{i,t-L}$, is the percentage of firm i's employees who possess a given skillset in the lagged month $t-L$, orthogonalized to size, book-to-market ratio, industry, and momentum of firm i.
The table reports the coefficient β corresponding to a 10% chage 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	0.16%*	0.12%	0.14%*	0.10%
(Standard error)	*(0.08%)*	*(0.08%)*	*(0.08%)*	*(0.08%)*
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)	*(0.07%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
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 6: To evaluate the relationship between skill focus and gross profitability, we estimate the following specification for four quarter lags:
For $L \in \{1\,q, 2\,q, 3\,q, 6\,q\} : Profitability_{i,t} = \alpha + \beta AbnSkillChange_{i,t-L} + \epsilon_{i,t}$,
where $Profitability_{i,t}$ is firm i's gross profibitability in quarter t, defined as the difference between revenues and cost of goods sold, scaled by assets. $AbnSkillChange_{i,t-L}$, is the change in the percentage of firm i's employees who possess a given skillset during the lagged quarter $t - L$, orthogonalized to size, book-to-market ratio, industry, and momentum. 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	-0.33%***	-0.38%***	-0.12%	-0.30%***
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.09%)*	*(0.09%)*
Information Technology				
Coefficient	0.11%	0.13%	0.12%	0.19%**
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.09%)*	*(0.09%)*
Mobile Network				
Coefficient	-0.07%	0.02%	0.12%	0.09%
(Standard error)	*(0.11%)*	*(0.11%)*	*(0.10%)*	*(0.11%)*
Software Engineering				
Coefficient	0.02%	-0.02%	-0.06%	-0.16%**
(Standard error)	*(0.09%)*	*(0.09%)*	*(0.08%)*	*(0.08%)*
Web Development				
Coefficient	-0.13%	-0.16%*	-0.22%***	-0.31%***
(Standard error)	*(0.08%)*	*(0.09%)*	*(0.08%)*	*(0.08%)*
Communication Skillsets				
CRM				
Coefficient	0.32%***	0.46%***	0.43%***	0.50%***
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.10%)*	*(0.10%)*
Digital Marketing				
Coefficient	0.03%	0.09%	0.10%	0.22%***
(Standard error)	*(0.09%)*	*(0.10%)*	*(0.09%)*	*(0.09%)*
Social Media				
Coefficient	0.09%	0.19%**	0.07%	0.09%
(Standard error)	*(0.10%)*	*(0.10%)*	*(0.09%)*	*(0.09%)*

Table 7: We estimate the relationship between skill changes and earnings surprises using the following specification for four quarter lags:
For $L \in \{1\,q, 2\,q, 3\,q, 6\,q\}$: $SUE_{i,t} = \alpha + \beta AbnSkillChange_{i,t-L} + \epsilon_{i,t}$,
where $SUE_{i,t}$ is firm i's standardized earnings surprise in quarter t, defined as the difference between earnings in euarter t and $t-4$, scaled by the standard deviation of this difference over the preceding eight quarters. $AbnSkillChange_{i,t-L}$, is the change in the percentage of firm i's employees who possess a given skillset during the lagged quarter $t-L$, orthogonalized to size, book-to-market ratio, industry, and momentum. 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	-1.35%**	-2.46%***	-0.55%	0.55%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.67%)*	*(0.69%)*
Information Technology				
Coefficient	0.84%	-0.22%	-0.15%	-1.53%**
(Standard error)	*(0.73%)*	*(0.71%)*	*(0.71%)*	*(0.71%)*
Mobile Network				
Coefficient	0.13%	1.91%**	2.22%***	1.16%
(Standard error)	*(0.69%)*	*(0.70%)*	*(0.70%)*	*(0.71%)*
Software Engineering				
Coefficient	-1.13%**	-1.41%**	-0.43%	1.11%
(Standard error)	*(0.67%)*	*(0.68%)*	*(0.68%)*	*(0.69%)*
Web Development				
Coefficient	0.01%	-1.79%***	-2.32%***	-1.23%
(Standard error)	*(0.65%)*	*(0.66%)*	*(0.67%)*	*(0.70%)*
Communication Skillsets				
CRM				
Coefficient	0.41%	2.18%**	2.54%***	2.75%***
(Standard error)	*(0.74%)*	*(0.75%)*	*(0.75%)*	*(0.75%)*
Digital Marketing				
Coefficient	0.52%	-0.83%	-1.76%**	-0.46%
(Standard error)	*(0.69%)*	*(0.70%)*	*(0.70%)*	*(0.71%)*
Social Media				
Coefficient	0.05%	0.82%	-0.36%	-1.84%***
(Standard error)	*(0.68%)*	*(0.68%)*	*(0.69%)*	*(0.71%)*

Table 8: Results from subsample analysis across three periods: 2000-2005, 2006-2010, and 2011-2016. We estimate the following specification for each sub-period:

$$FF3alpha_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$$

$FF3alpha_{i,t}$ the return for firm i in month t in excess of that predicted by Fama and French (1993) factors. $AbnSkill_{i,t-L}$, is the percentage of firm i's employees who possess a given skillset in the lagged month $t-L$, orthogonalized to size, book-to-market ratio, industry, and momentum of firm i. The table reports the coefficient β corresponding to an additional 10% of employees possession each skillset as either their primary or secondary skillset.

Skillset	2000-2005 (NAICS 51)	2006-2010 (NAICS 31-33)	2010-2016 (NAICS 52)
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 Network			
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%)*

Appendix A 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.

In order to access whether firms with higher employee turnover underperform firms with more stable workforces, we compute abnormal monthly turnover for each firm as the portion of its turnover unexplained by other firm characteristics. We use lagged values of this variable to predict the firm's future returns, and find a strong negative relationship.

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:

$$Turnover_{i,t} = \frac{Join_{i,t} + Depart_{i,t}}{N_{i,t}} \tag{6}$$

We winzorize this variable at the top and bottom 1% across all firm-months.

In order to screen out the effect of other firm characteristics that have been shown to affect performance,[15] we define abnormal turnover as the component of turnover that is orthogonal to a number of firm-level controls. To do so, we first regress $Turnover_{i,t}$ on firm characteristics:

$$Turnover_{i,t} = \alpha + \gamma X_{i,t} + \epsilon_{i,t}, \tag{7}$$

where the vector of characteristics $X_{i,t}$ includes $Size_{i,t}$, defined as the natural logarithm of firm i's market capitalization, $BM_{i,t}$, defined as firm i's book-to-market ratio using book value from the most recent quarter-end preceding month t, and $Ind_{i,t}$, defined as the firm's industry classification. In some specifications, we include also $Perf_{i,t}$, defined as firm i's market-adjusted return during month t.

Abnormal turnover is then defined as the residuals from the specification (7), i.e., the component of employee turnover unexplained by firm characteristics:

$$AbnTurnover_{i,t} = Turnover_{i,t} - PredTurnover_{i,t}, \tag{8}$$

[15] For predictability of returns from firm characteristics such as size, book-to-market, and past performance, see Fama and French (1992, 1993), Jagadeesh and Titman (1993), and Moskowitz and Gringlatt (1999), among others.

where $PredTurnover_{i,t}$ is the predicted value for firm i's turnover in month t from fitting specification (7).

We investigate the relationship between abnormal turnover and subsequent firm performance by estimating a linear regression of monthly stock market returns on lagged turnover. For the explanatory variable, $TurnoverVar_{i,t}$, we consider both raw turnover and two types of abnormal turnover: orthogonal to size, book-to-market, and industry, and orthogonal to size, book-to-market, industry, and contemporaneous momentum, defined as firm i's return from month $t-12$ to month $t-1$. We use two types of returns: raw return $Ret_{i,t}$ defined as firm i's excess return during month t, and $FF3alpha_{i,t}$ defined as the difference between firm i's return during month t and the predicted return from the Fama and French (1993) three-factor model. We estimate each of these specifications for four different lags between turnover and returns:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : ReturnVar_{i,t} = \alpha + \beta TurnoverVar_{i,t-L} + \epsilon_{i,t} \tag{9}$$

Firms with higher turnover in a given month experience significantly lower stock market returns in the following month. The estimates of the coefficient β, presented at the top of Table A.1, indicate that a 10% increase in abnormal turnover (orthogonalized to firm characteristics such as size, book-to-market, industry, and momentum) during month $t-1$ corresponds to 22 basis points lower three-factor alpha during month t (corresponding to an annual alpha of -2.67%).

The return predictability from employee turnover extends well beyond the one-month lag. The remainder of Table A.1 displays the results of specification (9) for lags $L = 2$ months, 3 months, and 6 months. The predictability of abnormal returns from abnormal turnover is actually largest at a lag of 2-3 months. For example, a 10% increase in a firm's abnormal employee turnover predicts a 65 basis points lower abnormal return three months out, corresponding to an annual alpha of -8.08%. These findings suggest the potential for the formation of profitable trading strategies based on employee turnover.

A.1 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. Technical skillsets display generally consistent negative relationships with returns across all three industries, while communication skillsets show differential premia across the considered 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). As can be seen from Table 3, these are the three largest industries in terms of numbers of employees. 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 5), 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.2, reveal largely consistent patterns for technical skillsets, and some differences across industries in terms of premia on communication skillsets. Abnormal firm-level focus on *Mobile Network* and *Web Development* is directionally negative across all three of the main industries, with comparable statistical significance across the board. The only potentially meaningful cross-industry heterogeneity is the lack of negative premia on *Software Engineering* and *Information Technology* in the Financial & Insurance industry.

The different industries' valuations of the communication skillsets are more heterogenous. Both *Digital Marketing* and *Social Media* carry negative premia in Information and Finance & Insurance, but not in Manufacturing. Similarly while having an abnormal share of employees with skills in *Client Relationship Management* is negatively predictive of future returns in both Information and Manufacturing, this skillset is actually positively (albeit not statistically significantly) related to future returns in Finance & Insurance. Overall, while the communication skillsets carry a generally negative premia in Information, they are less predictive of poor future returns in Manufacturing and Finance & Insurance.

Table A.1: Results from panel regressions of monthly firm-level stock market returns on lagged employee turnover. We estimate the following specification for four lags:
For $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : ReturnVar_{i,t} = \alpha + \beta TurnoverVar_{i,t-L} + \epsilon_{i,t}$
The outcome variable, $ReturnVar_{i,t}$, is firm-level excess return during month t, $Ret_{i,t}$, in column (1) and Fame and French (1993) three factor alpha, $FF3alpha_{i,t}$, in columns (2)-(4). In column (1)-(2), the explanatory variable, $TurnoverVar_{i,t-L}$, is raw turnover, $Turnover_{i,t-L}$, of firm i during the lagged month $t - L$. In column (3), the explanatory variable is abnormal turnover, $AbnTurnover_{i,t-L}$, orthogonalized to size, book-to-market ratio, and industry of firm i. In column (4), the abnormal turnover variable is also orthogonalized to the firm's momentum (return from month $t-L-12$ to month $t-L-1$). The table reports the coefficient β, scaled so as to correspond to the change in returns for every additional 10% in employee turnover.

Lag	(1) Excess Return Raw Turnover	(2) FF3 alpha Raw Turnover	(3) FF3 alpha Abn. Turnover	(4) FF3 alpha Abn. w/ Mom
L = 1 month				
Coefficient	-1.21%***	-0.23%**	-0.22%**	-0.22%**
(Standard error)	*(0.11%)*	*(0.09%)*	*(0.09%)*	*(0.09%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 2 months				
Coefficient	-0.39%***	-0.54%***	-0.56%***	-0.56%***
(Standard error)	*(0.11%)*	*(0.10%)*	*(0.10%)*	*(0.10%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 3 months				
Coefficient	-0.87%***	-0.62%***	-0.65%***	-0.65%***
(Standard error)	*(0.11%)*	*(0.10%)*	*(0.10%)*	*(0.10%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 6 months				
Coefficient	-0.11	-0.27%***	-0.27%***	-0.27%**
(Standard error)	*(0.12%)*	*(0.11%)*	*(0.11%)*	*(0.11%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X

Table A.2: Results from separate regressions of returns on lagged employee skillsets across three industries. We estimate the following specification for each industry:

$FF3alpha_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$

$FF3alpha_{i,t}$ the return for firm i in month t in excess of that predicted by Fama and French (1993) factors. $AbnSkill_{i,t-L}$, is the percentage of firm i's employees who possess a given skillset in the lagged month $t-L$, orthogonalized to size, book-to-market ratio, industry, and momentum of firm i. The table reports the coefficient β corresponding to an additional 10% of employees possession 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 Network			
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%)*