Trading on Talent: Human Capital and Firm Performance*

Anastassia Fedyk† and James Hodson‡

November 18, 2017

Abstract

How does a firm's human capital impact financial performance? By directly observing the employment and education trajectories of a significant proportion of U.S. public company employees from 1990 to the present, we explore the relationship between performance and two aspects of human capital: turnover and skills. First, we find that firms with higher employee turnover experience significantly worse future returns. A long-short strategy based on employee turnover with a three-month lag generates an excess compounded annual return of 14.3%. Second, firms with a larger emphasis on sales-oriented skills show better subsequent performance, whereas firms with more focus on administrative skills underperform. The effects of skills are heterogeneous across industries, with a larger premium on web development in Information, a higher premium on insurance in Manufacturing, and no benefit from sales-oriented skills in Finance.

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

*The authors are grateful to John Beshears, John Campbell, Lauren Cohen, Jeremy Stein, and seminar participants at Cubist Systematic and PanAgora Asset Management for insightful comments. This project received funding from the EU Research and Innovation programme Horizon 2020 under grant agreement No. 675044. This paper was awarded 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 research received a "not human subject research" determination from the Institutional Review Board at Harvard University (protocol #IRB17-0443).

†Harvard University, Department of Economics and Harvard Business School. Mail: Baker Library 244C, 25 Harvard Way, Boston, MA 02163. Email: afedyk@hbs.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 explore the relationship between firm performance and the turnover and composition of its employees using a comprehensive dataset of 37 million employment and education histories of U.S. public company employees between January 1990 and May 2017. We find that firms with more stable workforces experience significantly higher future stock market returns. The composition of human capital is also predictive of future performance: firms with higher percentages of employees with expertise in sales tend to outperform firms with less emphasis on this core business area, whereas firms with higher percentages of employees with administrative skills systematically underperform. Analysis sliced by industry reinforces the significance of human capital skills for firm performance, but also draws attention to important heterogeneity. On the one hand, a larger focus on managerial and administrative skills translates to worse firm performance across the board. On the other hand, benefits of sales-oriented skills are less universal: these skillsets no longer predict positive excess returns when we restrict our attention to the Finance & Insurance industry. Similarly, while financial skills are among the most important aspects of human capital for predicting future performance in the Manufacturing industry, these skills carry no positive premium in Information.

The increasing importance of human capital for modern firms has been stressed by practitioners and academics alike.[1] A number of studies consider the effects of Chief Executive Officers on a variety of corporate outcomes, including performance, strategy, leverage, mergers and acquisitions, and investment decisions.[2] While the documented CEO effects demonstrate the importance of firm leadership, a firm's human capital extends far beyond its uppermost executives. In this paper, we provide evidence that the composition of a firm's broad employee base is an important factor in driving firm performance.

[1] For an example of the latter, see Zingales (2000).

[2] 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.

We leverage techniques from machine learning on a large novel dataset of employment profiles to provide a detailed characterization of each firm's human capital. The dataset comes from a client relationship management and marketing compliance platform that aggregates information regarding individuals' employment and company events from a wide variety of sources including datasets for purchase, partner CRM databases, and online public records and resumes. The dataset covers more than 330 million individuals globally, with approximately 37 million 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). Approximately 77% of the employees are based in the U.S., with substantial coverage also in Europe, Asia, and South/Central America. Importantly, although our analysis centers on U.S. equities, we are able to identify employees of specific firms across borders, providing a more complete view of each individual firm's human capital.

We employ a mixture of textual analysis, existing knowledge-bases, and crowd-sourcing to structure and analyze the free-form employment data. Since the data come from a variety of sources, represent inconsistent formats, and largely feature unconstrained entry by individuals, we take a number of steps to reliably disambiguate each employee's educational attainment and link the names of employers to U.S. publicly traded companies. Of the approximately 4,800 employers active on the main U.S. equity markets (NYSE, NASDAQ, and AMEX), we are able to achieve significant coverage of human capital of almost 4,300 firms (89%). 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. This offers broader coverage than prior detailed 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 20 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. The 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 *Administration*, *Middle Management*, and *Banking & Finance*.

Mobility of human capital measured in our data reveals meaningful relationships between firms, which are not captured by traditional industry analysis. For example, we document that 58.6% of job transitions occur outside of two-digit NAICS classifications. There is a high degree of connectedness between different industries through the human capital channel; however, the extent of this connectedness is heterogenous across industries. Some industries, such as Information (NAICS code 51) and Professional, Scientific, & Technical Services (NAICS code 54) are quite versatile, and the bulk of their employees depart to other industries. Other industries, such as Manufacturing (NAICS codes 31-33) and Finance & Insurance (NAICS code 52) are more specialized, and the majority of their turnover stays within the same industry. This heterogeneity in turnover is paralleled by the diversity of skillsets reflected in different industries. For example, the most common skillsets in the Manufacturing industry are *Manufacturing* and *Electrical Engineering*, while the *Banking & Finance* skillset is far more common in firms that fall within the Finance and Insurance industry.

We analyze the relationship between employee mobility and the firms' financial performance. Our results indicate that firms with lower turnover rates systematically outperform

firms with higher turnover rates. In order to control for other factors known to affect returns, we compute abnormal turnover as our measure of employee turnover orthogonalized to firm size, book-to-market ratio, industry, and past returns. A 10% increase in abnormal turnover corresponds to 84 basis points lower returns during the following month, 78 basis points lower returns two or three months out, and 65 basis points lower returns six months out. We form a monthly long-short trading strategy based on this insight, going long a value-weighted index of firms in the top quintile by lagged turnover and short a value-weighted index of firms in the bottom quintile by lagged employee turnover. The long-short portfolio earns an excess monthly return of 1.12% when using employee turnover from three months prior and performs similarly using lags of one, two, or six months.

In addition to turnover in human capital, returns are related to the composition of human capital within each firm. We use the classification of employees into skillsets to evaluate the extent to which various skillsets correspond to better or worse financial performance. For each firm in each month, we compute the proportions of employees possessing each of the 44 skillsets. 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. This yields, for each firm in each month, a measure of abnormal focus on each skillset. We find that abnormal focus in the area of sales (skillsets such as *Sales Management* and *CRM*) corresponds to better future returns, while abnormal focus on managerial and administrative skillsets such as *Administration* and *Product Management* correlates negatively with future performance. In particular, for every additional 10% of employees specializing in *Sales Management*, firms experience 11 basis points higher returns three months later. By contrast, for every additional 10% of employees specializing in *Product Management*, firms see 32 basis points lower returns three months out.

While some skillsets are robustly predictive of returns across industries, other skillsets have more heterogenous effects. We look at the relationship between the percentage of

employees focusing on each skillset and subsequent returns in each of the three largest industries: Manufacturing, Finance & Insurance, and Information. Generic managerial and administrative skillsets such as *Product Management* and *Human Resources* are negatively predictive of returns across all three industries. For example, a 10% increase in the share of employees focusing on junior-level *Human Resources* corresponds to 26 basis points lower next-month returns in Information, 56 basis points lower monthly returns in Manufacturing, and 28 basis points lower next-month returns in Finance & Insurance. On the other hand, the importance of *Sales Management* shows in both Information and Manufacturing, positively predicting returns in these two industries, but not in Finance. *Web Development*, which is positively priced in Information, is not significantly associated with returns in either Manufacturing or Finance. Overall, the analysis sliced by industry reinforces the importance of the composition of human capital for firm performance, but also draws attention to the fact that different compositions are likely to be valuable for different types of firms.

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 indications of corporate culture and employee satisfaction,[4] non-compete agreements,[5] hiring,[6] and demographics such as age and education.[7] We expand this line of work by leveraging big data techniques to structure and analyze a large dataset with detailed information on 37 million employees of U.S. public companies. This allows us to characterize the composition of firms' human capital in terms of key skillsets across the broad landscape of U.S. public companies. Our findings indicate that the market prices not only executive, organizational, and demographic characteristics

[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).

of companies, but also the composition of skills across the full spectrum of a company's employees.

The remainder of the paper proceeds as follows. Section 2 describes the large unstructured employment dataset and the steps taken to structure the records and link firms to employees. Section 3 presents the methodology for classifying self-reported skills into meaningful skillsets. Section 4 presents the empirical analysis of the relationship between employee turnover and firm performance, while Section 5 relates performance to the composition of employee skillsets. Section 6 concludes.

2 Data

Employment histories and demographic data are provided by 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 data for effective financial analysis.

2.1 Individual Profile Data

We observe the global employment market through the lens of a novel dataset of approximately 330 million individual employment and education records provided by an aggregator of employment profiles for sales and compliance. The aggregator 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, notable events, and marketing compliance status of individuals globally. The aggregator's infrastructure ensures that records are independently verified and researched on a regular basis.

For each individual in our sample we have a unique identifier, along with 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 through electronic resumes and online testing services. We remove empty profiles (profiles with no listed jobs, education, or skills) from the analysis, as well as any individual who has never worked for a U.S. publicly traded company (NYSE, NASDAQ, and AMEX exchanges). This leaves us with a sample of approximately 37 million U.S. public-company employees with employment information spanning the 1990's to May 2017.

Summary statistics of the demographic, educational, and employment characteristics of the individuals in our sample are tabulated 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 profiles contain a number of inconsistencies stemming from the diversity of unconstrained user input, as well as the variety of underlying data sources. We take the following steps to normalize the individual profile data: (1) disambiguation of employer names to official trading names and relevant stock symbols for public companies (outlined in Section 2.1.1) and (2) extraction of normalized degree and institution names from the education records, linking to educational institution rankings to identify individuals with degrees from top institutions (Section 2.1.2).

Our main empirical analyses are based on human capital characteristics constructed from the individual employment profiles merged with industry classification, shares outstanding, and book value from Compustat, as well as returns and market capitalization from CRSP. Over our sample period between January 1990 and May 2017, 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).

2.1.1 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. Since employer names in the records are entered free-form, the raw data contain a number of inconsistencies, alternative references, and misspellings.

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. Table 2 offers examples of the official company names matched to the listed names in the employment records.

We perform company name disambiguation using standard methods from entity disambiguation.[8] To evaluate the disambiguation procedure, we manually look up 1,000 employment records and match 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

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

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, the baseline yields a highly precise set of matching (the precision is 85%). However, the recall is quite 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 cataloging 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 augment the above procedure by processing each company name record to remove extraneous information (parenthetical statements, departments, locations, job titles, and descriptions of roles). Additionally, we use a manually compiled list of common departments, job roles, and miscellaneous terms that we have empirically determined to be extraneous to the matching procedure. That said, in cases when multiple candidate strings appear in our database of canonical names, we favor longer matches. 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 2017.[9]

[9]Employee counts of NYSE, NASDAQ, and AMEX listed companies, as reported in 10Q filings to the SEC during 2017 provide an estimate of approximately 36 million global employees. As of May 2017, our

2.1.2 Education Extraction and Classification

We use the educational records to identify the types of degrees obtained by the individuals, as well as whether the individuals attended a top academic institution. The degree types of interest are: high school, vocational associates, bachelors, masters (other than an MBA), MBA, and doctoral (including Ph.D., M.D., and J.D. degrees).

We extract and disambiguate each degree type and institution name, manually correcting mistakes such as "Oxford Brookes University" being recognized as "Oxford University" and the "Princeton Theological Seminary" being recognized as "Princeton University." We compile a list of "elite" institutions from merging the U.S. News and World Report National Universities Ranking 2016 and the Times Higher Education World University Rankings 2016, providing a list of 100 global institutions. Any individual holding at least one degree from one of these institutions is labeled as having attended an elite institution.

The summary statistics of the education data are summarized in Panel 4 of Table 1. Out of the 37 million public company employees in our sample, 21 million individuals (57% of the sample) list at least some form of post-secondary education. The vast majority of those, 90%, list a bachelors degree, and a sizable percentage of the population (20%) hold a masters other than an MBA. The incidence of MBAs and doctoral degrees is lower, at 8% and 4%, respectively. Roughly 6% of the sample hold at least one of their listed degrees from an elite institution.

2.2 Private Companies and Industry Codes

Although our primary analysis centers around public company employees, in this subsection we look at the human capital landscape across the full economy. We identify private company employees and link them to industry segments. We document a high degree of connectivity between the different industries via mobility of human capitla.

Private company employers are more difficult to disambiguate than their publicly traded

coverage of active employees at these companies is approximately 14 million, representing 39%.

counterparts, since there are no comprehensive exchange lists for private companies. We make use of the data provider'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 disambiguation procedure described in Section 2.1.1). For each identified firm that has a common industry classification schema 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 a shrinking of the Manufacturing sector and growth in the Retail Trade sectors.

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 also 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.

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 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.[10] We recover forty-four skillsets, which capture intuitive competencies ranging from product management to web development to healthcare professionals.

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: $d = \{\text{Microsoft Office, Melhoria continua}\}$. Let the vocabulary W consist of all skills that appear in at least one profile – for example, if there is a profile $d = \{\text{Microsoft Office, Melhoria continua}\}$ 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

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

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 given the chosen skillset t_n.

We estimate the key parameters of the model, α and β, using a collapsed Gibbs sampling algorithm. Table 2 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.*

For expositional clarity, we group the identified skillsets into seven broad categories. The first six categories capture key areas of focus relevant across the economy – managerial and administrative, sales-oriented, communications, financial, technical, and artistic. The seventh category includes areas that are more specific to particular industries. The skillsets are grouped into these areas as follows:

- Managerial and Administrative skillsets: *Administration; Middle Management; Business Development; Product Management; Technical Product Management; Operations Management; Industrial Management; Logistics; Recruiting; Human Resources (Jr.); Human Resources (Sr.).*

- Sales-Oriented skillsets: *Sales Management; Sales; CRM.*

- Communications skillsets: *Digital Marketing; Social Media; Public Policy.*

- Financial skillsets: *Accounting & Auditing; Banking & Finance; Insurance.*

- Technical skillsets: *IT Management; Web Development; Software Engineering; Electrical Engineering; Data Analysis; Mobile Telecommunications.*

- Artistic skillsets: *Visual Design; Graphic Design; Web Design; Musical Production; Video & Film.*

- Specialized skillsets: *Pharmaceutical; Oil, Energy & Gas; Construction; Manufacturing; Retail; Military; Legal; Education; Non-Profit; Healthcare; Hospitality; Real Estate; 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 $j(i), j'(i)$ denote the indices of the two largest values in the vector $\hat{\theta}_i$. Then the 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 *Administration*, 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 (two-digit NAICS code 52). As can be seen from Figure 4, Manufacturing tilts heavily towards the *Electrical Engineering* and *Manufacturing* skillsets. In contrast, the Finance & Insurance industry leans heavily on *Banking & Finance* and, to a lesser extent, *Administration, Business Development* and *Accounting & Auditing* (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 5.2.

4 Employee Turnover and Firm Performance

We analyze the relationship between the stability of a firm's workforce and the firm'ssubsequent 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. A long-short trading strategy with monthly rebalancing based on three-months lagged human capital earns abnormal monthly returns of 1.12%.

4.1 Regression Analysis

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:

$$Turnover_{i,t} = \frac{Join_{i,t} + Depart_{i,t}}{N_{i,t}} \tag{1}$$

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,[11] 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:

$$Turnover_{i,t} = \alpha + \gamma X_{i,t} + \epsilon_{i,t}, \tag{2}$$

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 regression (2), i.e., the

[11] 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.

component of employee turnover unexplained by firm characteristics:

$$AbnTurnover_{i,t} = Turnover_{i,t} - PredTurnover_{i,t}, \tag{3}$$

where $PredTurnover_{i,t}$ is the predicted value for firm i's turnover in month t from fitting regression (2).

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 financial performance. We use two types of returns: raw return $Ret_{i,t}$ defined as firm i's return during month t, and $AbnRet_{i,t}$ defined as the difference between firm i's return during month t and the market return during month t. We estimate each of these specifications for four different lags between turnover and returns:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : ReturnVar_{i,t} = \alpha + \beta TurnoverVar_{i,t-L} + \epsilon_{i,t} \tag{4}$$

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 5, indicate that a 10% increase in turnover during month $t-1$ corresponds to 70 basis points lower raw monthly return during month t and 66 basis points lower market-adjusted monthly return. The results are even stronger when the turnover variable is orthogonalized to firm characteristics such as size, book-to-market, industry, and month $t-1$ return. A 10% increase in abnormal turnover translates to 82-84 basis points lower returns during the following month.

The predictability extends well beyond the one-month lag, with employee turnover negatively predicting equity returns two, three, and six months out. The remainder of Table

5 displays the results of specification (4) for lags $L = 2$ months, 3 months, and 6 months. The predictability of abnormal returns from abnormal turnover declines but remains strong as we extend the horizon. Even at a six month lag, a 10% increase in a firm's abnormal employee turnover predicts a 65 basis points lower abnormal return. These findings suggest the potential for the formation of profitable trading strategies based on employee turnover, which we explore in the next subsection.

4.2 Portfolio Returns

We test the performance of a monthly long-short portfolio based on the relationship documented in the previous subsection. The long portfolio consists of all firms whose lagged abnormal employee turnover rates fall in the lowest quintile, while firms from the top quintile by lagged employee turnover form the short portfolio. Using a three-month lag, the long portfolio earns, on average, a 1.23% monthly return, while the short portfolio earns a substantially lower return of 0.10% per month, resulting in a total long-short return of 1.12% per month.

Each month, we sort firms into portfolios based on their lagged employee turnover. Thus, given a lag L, in month t, we form two portfolios: the high turnover portfolio and the low turnover portfolio. We use the lagged abnormal turnover variable, $AbnTurnover_{i,t-L}$, orthogonalized to each firm i's size, book-to-market, industry, and past performance. The high turnover portfolio consists of the top quintile of firms based on their lagged abnormal turnover $AbnTurnover_{i,t-L}$. The low turnover portfolio contains the bottom quintile of firms based on $AbnTurnover_{i,t-L}$. The long-short portfolio goes long the low turnover portfolio and short the high turnover portfolio.

Having constructed the portfolios using lagged employment turnover from month $t - L$, we compute the returns on these portfolios during the current month t. The returns are value-weighted and computed relative to the overall market returns. Thus, for example, the return on the high turnover portfolio during month t is the weighted average of the returns

on each firm in the portfolio, weighted by the firms' market capitalization, minus the market return during that month. The return on the low turnover portfolio is computed analogously, and the return on the long-short portfolio is equal to the return on the low turnover portfolio minus the return on the high turnover portfolio.

We evaluate the performance of the high turnover portfolio, the low turnover portfolio, and the long-short portfolio using lags of one, two, three, and six months. The average monthly returns and standard errors on each portfolio at each lag are presented in Table 6.

The low turnover portfolio substantially outperforms the portfolio of firms with high employee turnover. Using turnover rates from the previous month, a monthly rebalanced long-short strategy that loads positively on the low turnover portfolio and goes short the high turnover portfolio earns an excess return of 0.95% per month, strongly statistically significantly different from zero.

The long-short portfolio performance does not decay, and is even marginally stronger, as we allow for a larger lag between the employee turnover variable and the observed returns. Each month, a long-short portfolio formed based on employee turnover from two (three) months prior earns an excess return of 1.06% (1.12%) monthly. Performance is similar as the lag is extended to six months, with the long-short portfolio earning an excess return of 1.10% monthly. These are results are consistent with the patterns observed by Belo, Lin, and Bazdresch (2014) using quarterly hiring rates, and slightly stronger due to the finer granularity of the data. Overall, firm-level employee turnover is a strong predictor of subsequent stock market returns.

The performance of the turnover-based trading strategy is also remarkably consistent over time. Figure 6 displays the cumulative returns to the strategy using a three-month implementation lag: each month, the long and short portfolios are formed using the abnormal turnover values from three months prior. As can be seen from the figure, the cumulative returns are upward trending throughout the entirety of the sample period from April 2001 to December 2016, with an exception of flat performance in 2014-2015. The strategy does

particularly well during the crisis period of 2009, suggesting that the stability of the workforce plays an especially important role during harsh market conditions.

5 Employee Skillsets and Firm Performance

We now turn our attention to investigating the relationship between firm performance and the distribution of skillsets possessed by the firm's employees. We find that firms with higher proportions of employees skilled in sales tend to earn higher subsequent returns, whereas firms with more focus on administrative skillsets tend to underperform. While some skillsets (e.g., *Product Management*) carry similar effects across industries, others, such as *Web Design*, are valued differently in different segments of the economy.

5.1 Full-Sample Results

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.

We ensure that our measures of skillset prevalence are not affected by firm lifecycle stage, industry, or factors that have been shown to affect performance 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:

$$Skill_{i,t}^{j} = \alpha + \gamma X_{i,t} + \epsilon_{i,t}, \tag{5}$$

where the vector of characteristics $X_{i,t}$ includes market capitalization $Size_{i,t}$, book-to-market ratio $BM_{i,t}$, industry dummy $Ind_{i,t}$, and performance $Perf_{i,t}$.

Abnormal focus on skillset j is then defined as the residuals from the regression (5), i.e., the component of skillset prevalence that is unexplained by firm characteristics:

$$AbnSkill_{i,t}^{j} = Skill_{i,t}^{j} - PredSkill_{i,t}^{j}, \tag{6}$$

where $PredSkill_{i,t}^{j}$ is the predicted value for the percentage of firm i's employees with skillset j in month t from fitting regression (2).

To estimate the effect of focusing on a given skillset j on firm performance, we regress abnormal returns on lagged values of abnormal prevalence of skillset j. Thus, for each skillset j, we estimate the following specification at four different lags:

$$\text{For } L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta AbnSkill_{i,t}^{j} + \epsilon_{i,t}, \tag{7}$$

where $AbnRet_{i,t}$ is the difference between firm i's return during month t and the market return during the same month.

Table 7 reports the results grouped by the core areas of focus: (i) managerial and administrative skillsets, (ii) sales-oriented skillsets, (iii) communications skillsets, (iv) financial skillsets, (v) technical skillsets, and (vi) artistic skillsets.

Two areas stand out as robustly predictive of firm performance in either direction: sales-oriented skillsets and administrative and managerial skillsets. We see that a higher prevalence of sales-oriented skillsets in a company's employee base during a given month consistently predicts higher returns in the following month. An additional 10% of employees with *Sales Management* as a primary or secondary skillset corresponds to an additional 10 basis points

in the next month's return. On the other hand, more emphasis on managerial and administrative skillsets corresponds to significantly lower subsequent returns. An additional 10% of employees focusing on *Product Management* as their primary or secondary skillset predicts 32 basis points lower monthly return.

The other focus areas have more ambiguous relationships to firm performance. The communications category of skillsets is generally negatively associated with firm performance, similarly to the managerial and administrative skillsets but to a lesser extent. Skillsets across the financial and artistic categories consistently have zero association with firm performance in the full cross-section. The role of technical skills is most mixed. More focus on *Web Development* is positively associated with future performance, whereas higher emphasis on *Data Analysis* has a negative effect, and *Mobile Telecommunications* is neutral.

Consistent with the composition of the workforce changing relatively slowly, the results are almost identical at varying lags between the month where the prevalence of skillsets is measured and the performance outcome. As columns (2)-(3) of Table 7 show, the results at two, three, and six months lags mirror those with a lag of one month.

5.2 Heterogeneity across Industries

We explore heterogeneity in the performance-skillset relationship across industries, focusing on three primary industries: Information, Manufacturing, and Finance & Insurance. While some skillsets display consistent relationships with performance in all three industries, some skillsets are markedly more valuable in some industries than others.

The analysis in this section repeats the tests in specification (7) within three industries: 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 slowly, and the full-sample results do not depend on the lag between the measurement of human capital composition and returns (see Table 7), 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 8, reveal both similarities and differences across the industries. Focus on some skillsets, such as *Product Management*, is associated negatively with subsequent performance in all three of the main industries. Focus on other skillsets, such as *Insurance*, affects performance differently in different industries.

The main parallel across the industries is the effect of managerial and administrative skills. Across two out of three tabulated skillsets in this section (and holding also for untabulated skillsets such as *Operations Management*), a higher incidence of the skillset is associated with lower future performance. A 10% increase in *Human Resources (Jr.)* corresponds to 26 basis points lower next-month return in Information (significant at the 1% level), 56 basis points lower return in Manufacturing (significant at the 5% level) and 38 basis points lower return in Finance (significant at the 10% level). Similarly, a 10% increase in the percentage of employees with *Product Management* as one of their main areas of focus translates into 54-78 basis points lower next-month returns in all three industries.

The industries' valuations of the other skillsets are more heterogenous. Sales-oriented skillsets, which are most consistently associated with higher abnormal returns in the full-sample (see Table 7), are valued highly in both Information (for both *Sales* and *Sales Management*) and Manufacturing (for *Sales Management*), but not in Finance. Financial skillsets (*Insurance* and *Banking & Finance*), on the other hand, are valued positively in Manufacturing, where they are quite rare, negatively in Information, and neutrally in Finance, where these skillsets are possessed by the bulk of employees.

There is also considerable heterogeneity in terms of which technical skills are valuable in which industries. In the Information industry, higher focus on *Web Development* corresponds to higher future returns. The pattern is similar in Manufacturing, but not statistically significant. In Finance & Insurance, the relative valuations are quite different: a higher proportion of employees with *Web Development* skills is now associated with marginally

lower returns, albeit not statistically significant. The most valued technical skill in this industry is *Mobile Telecommunications*, which captures skills such as *telecommunications, wireless*, and *networking*. A larger focus on *Data Analysis*, on the other hand, is negatively predictive of future returns for financial companies.

Overall, the results suggest that there are certain – mostly common and generic – specialties that appear to be robustly predictive of lower returns across industries. However, the relationship is more nuanced for more specialized skills, highlighting the rich variation in how different skillsets are valued in different industries.

6 Conclusion

Our analysis suggests that the human capital structure of a firm can have important repercussions for its performance in financial markets. We motivate two channels for understanding the dynamics of company employment with respect to outcomes in capital markets. First, taking a macro view of the firm, we show that higher turnover, which is likely to stretch operational capacity, leads to significantly lower future returns. Second, on a more detailed level, the skills of individual employees and a firm's focus on particular *skillsets* can have meaningful and consistent effects on performance. A focus on sales-oriented skills intuitively provides greater pay-offs across many industries, whereas hiring too many employees into bureaucratic and administrative roles tends to have a negative effect.

Our work contrasts and builds 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. Individual employees' mobility and skillsets are only two aspects of the complex landscape of individual human capital. Detailed consideration of additional aspects – including hierarchical roles, demographic traits, and relationships across and outside the firm – might provide fruitful avenues for future research.

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

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

[10] 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.

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

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

[13] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[28] 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.

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

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

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

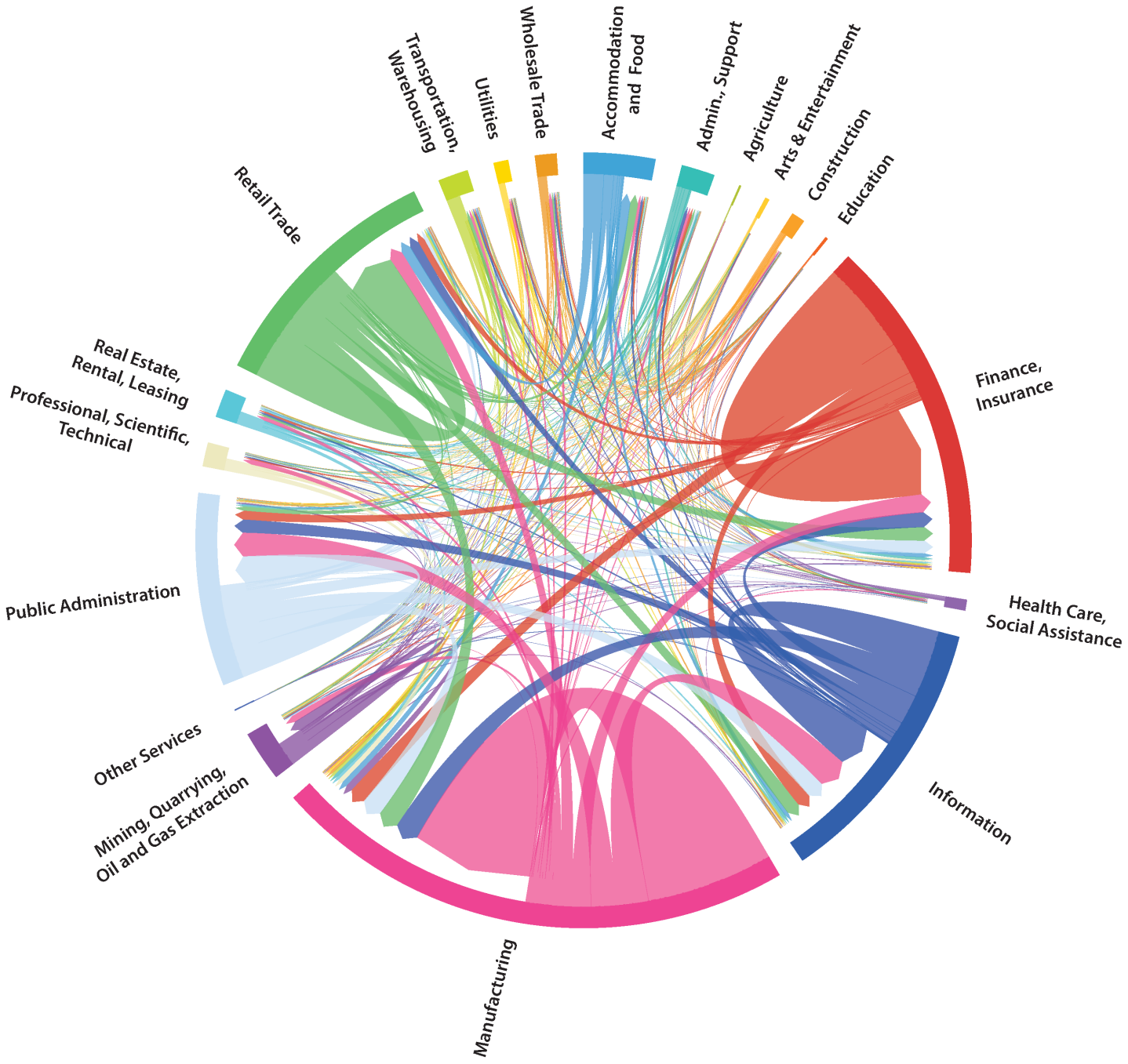

Figure 1: Cross-industry employee moves during the years 2010-2017. We consider all job changes that involve switching employer. Each arrows captures the prevalence of moves from one industry to another, following two-digit NAICS industry classification.

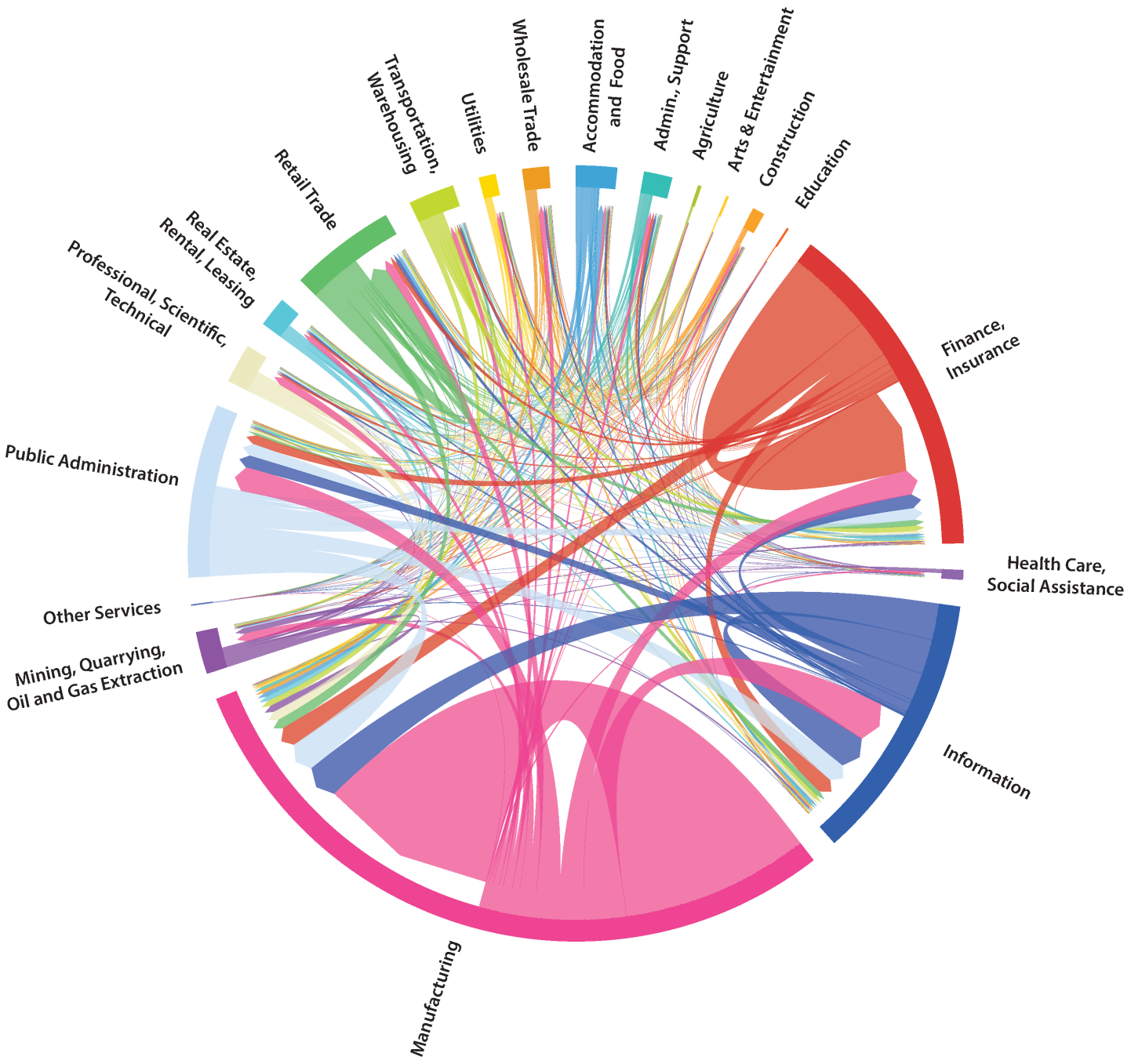

Figure 2: Cross-industry employee moves during the 1990s. We consider all job changes that involve switching employer. Each arrows captures the prevalence of moves from one industry to another, following two-digit NAICS industry classification.

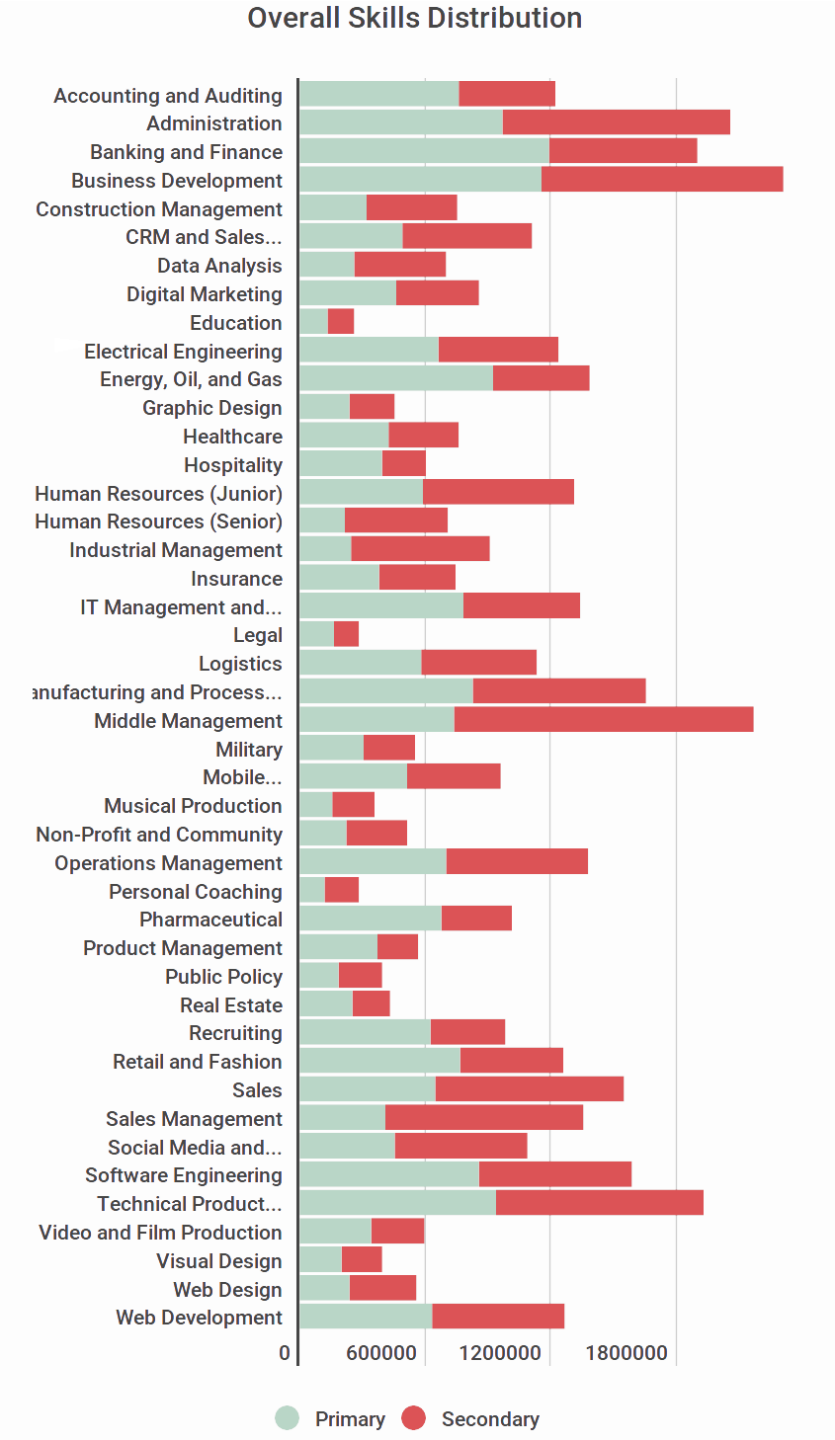

Figure 3: Frequency of skillsets across the full population of 37 million employees of public U.S. companies. Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the number of employees for whom that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

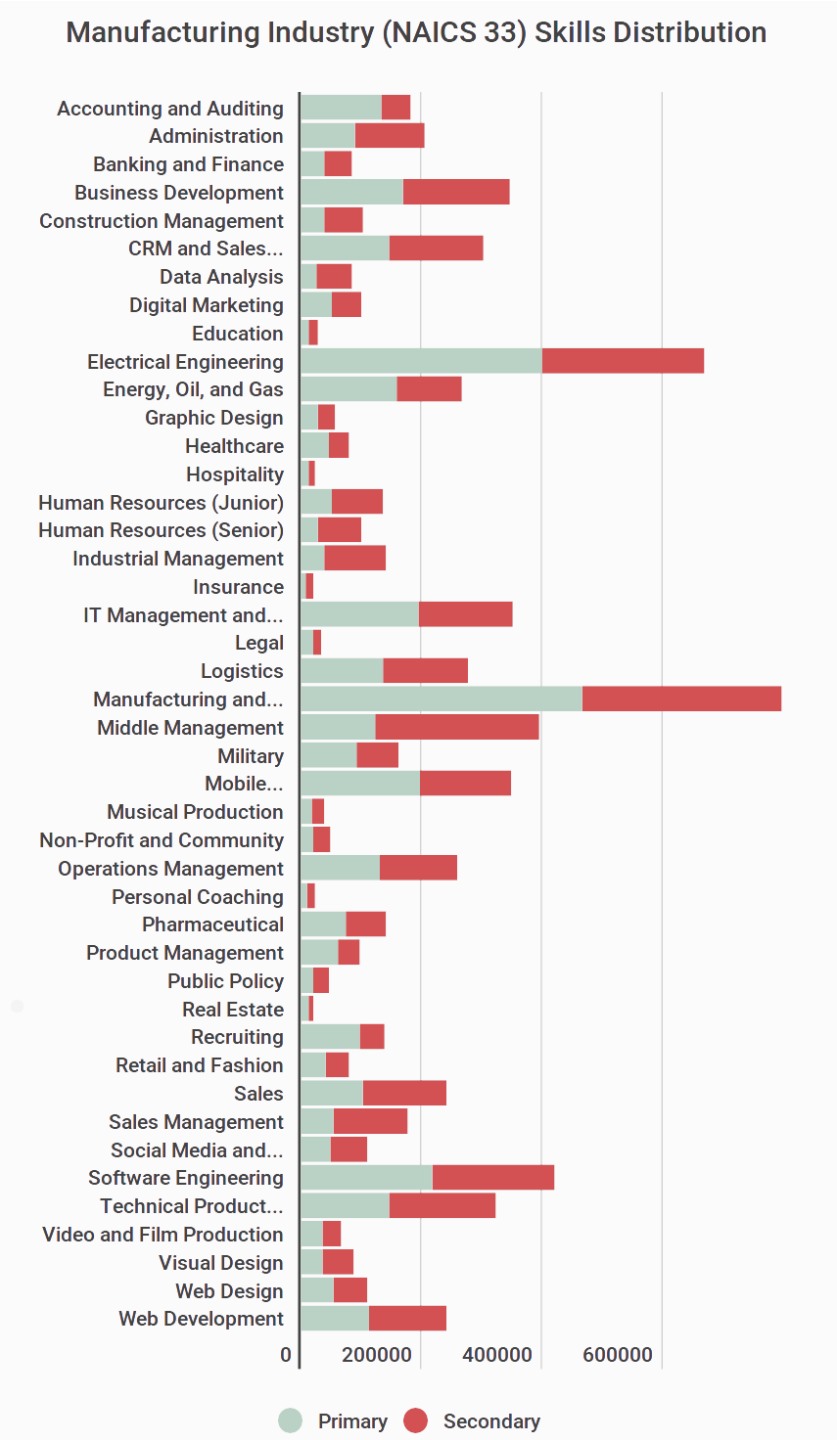

Figure 4: Frequency of skillsets across the employees of public U.S. companies in the Manufacturing industry (two-digit NAICS codes 31, 32, and 33). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows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

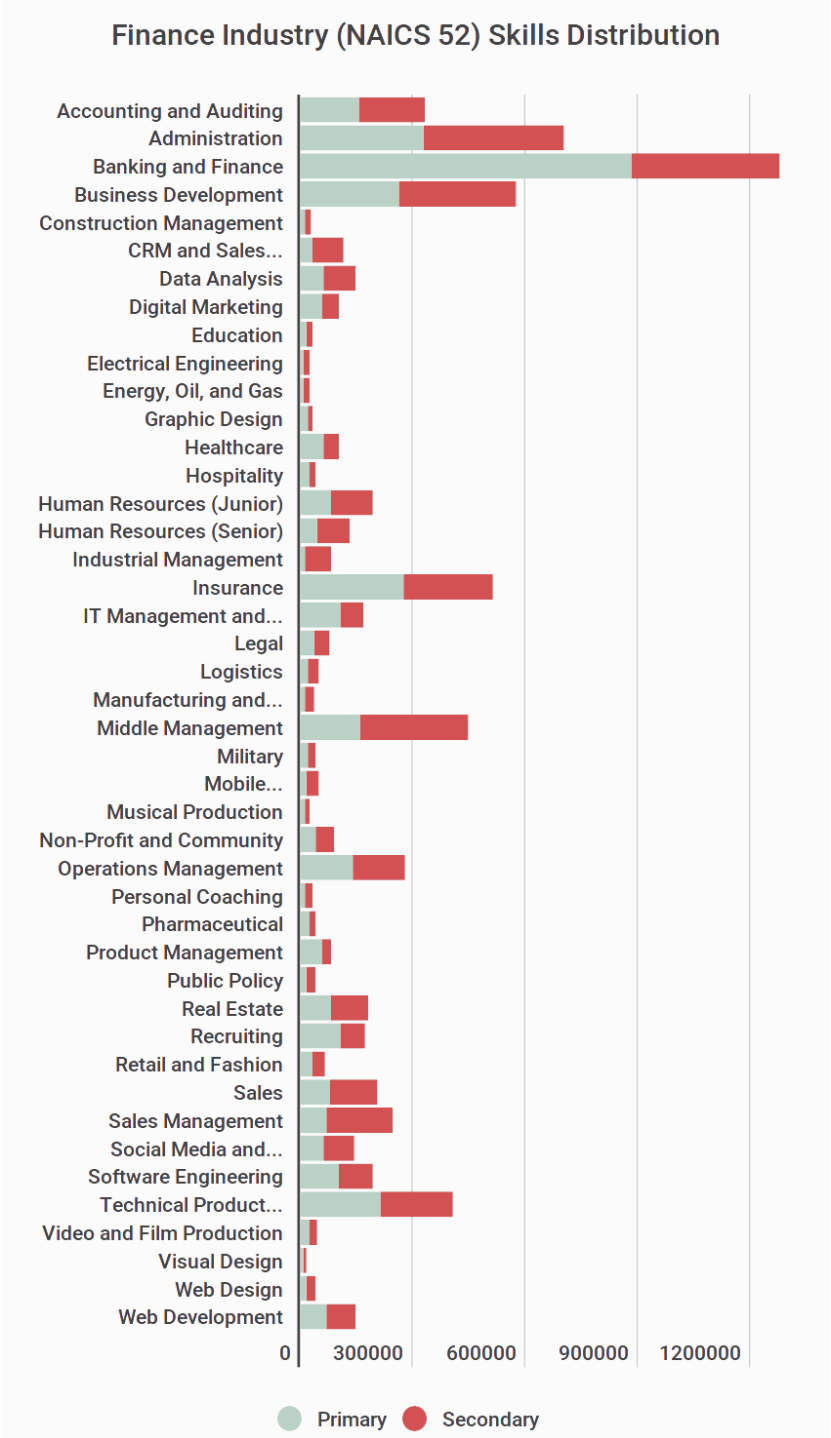

Figure 5: Frequency of skillsets across the employees of public U.S. companies in the Finance and Insurance industry (two-digit NAICS code 52). Each employee whose profile contains self-reported skills is classified as having a primary and a secondary skillset. For each skillset, the figure shows the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

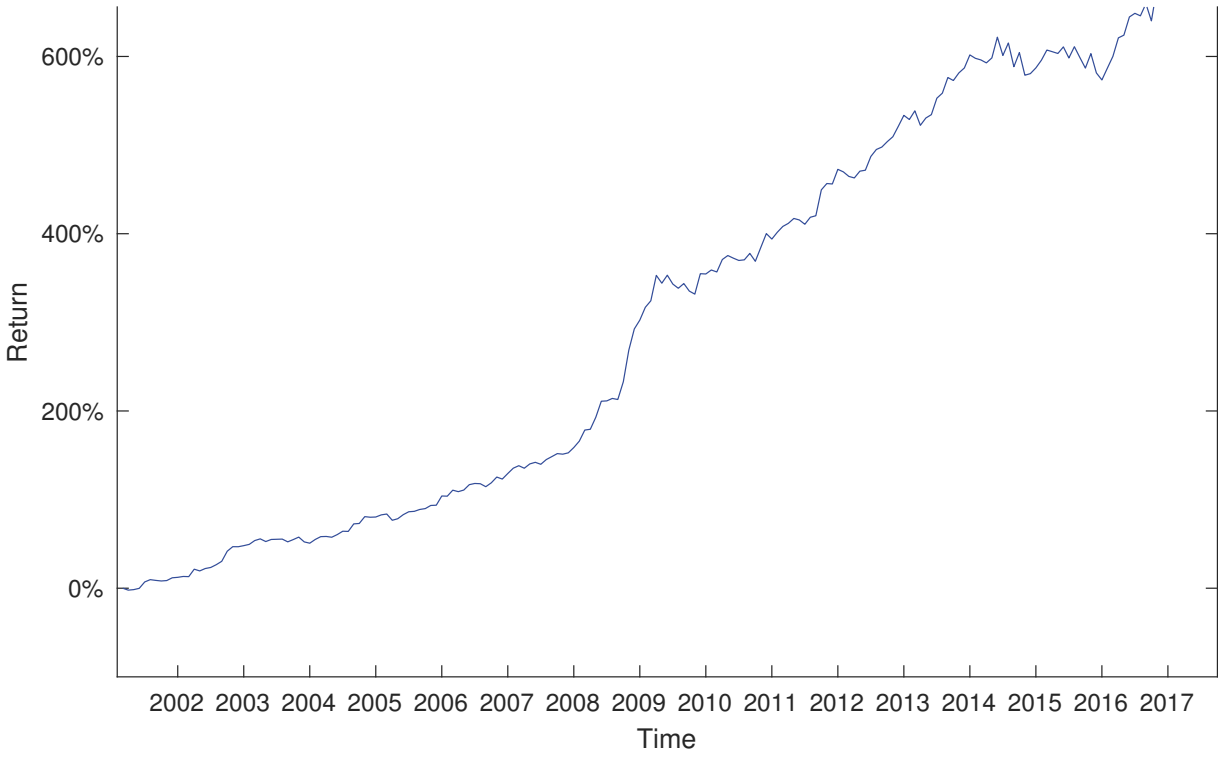

Figure 6: Cumulative returns on the trading strategy that goes long the bottom quintile of firms based on abnormal turnover and short the top quintile. Returns are computed monthly from April 2001 to December 2016, using abnormal turnover lagged by three months. The returns on the long and short portfolios are value-weighted across all firms in the given quintile.

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%**

Panel 3: Coverage of U.S. Equity Markets

Exchange	Employers	Coverage	Proportion
NYSE	2,033	1,865	92%
NASDAQ	2,523	2,196	87%
AMEX	240	206	86%
Overall	**4,796**	**4,267**	**89%**

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
Accounting & Auditing	accounting, financial reporting, financial analysis, auditing, financial accounting
Administration	microsoft office, microsoft excel, microsoft word, powerpoint, customer service
Banking & Finance	banking, financial analysis, finance, risk management, portfolio management
Business Development	business strategy, marketing strategy, business development, management, market research
Construction	construction, construction management, contract management, project planning, project management
CRM	business development, strategy, management, product management, crm
Data Analysis	data analysis, research, statistics, microsoft office, spss
Digital Marketing	digital marketing, social media marketing, marketing, online advertising, online marketing
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
Human Resources (Jr.)	coaching, change management, training, leadership development, management
Human Resouces (Sr.)	teamwork, communication, microsoft office, customer service, time management
Industrial Management	microsoft office, microsoft excel, sap, microsoft word, sap erp
Insurance	insurance, customer service, risk management, property & casualty insurance, general insurance
IT Management	windows server, troubleshooting, active directory, networking, window
Legal	legal research, legal writing, litigation, civil litigation, corporate law
Logistics	logistics, supply chain management, operations management, supply chain, purchasing
Manufacturing	manufacturing, continuous improvement, lean manufacturing, six sigma, product development
Middle Management	management, project management, leadership, strategic planning, process improvement
Military	military, security clearance, security, military experience, military operations
Mobile	telecommunications, wireless, voip, networking, ip
Musical Production	music, entertainment, music production, theatre, music industry

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
Non-Profit	nonprofits, public speaking, community outreach, fundraising, event planning
Oil, Energy & Gas	engineering, energy, petroleum, gas, project engineering
Operations Management	project management, change management, business analysis, it management, business process improvement
Personal Coaching	coaching, public speaking, sports, wellness, nutrition
Pharmaceutical	pharmaceutical industry, biotechnology, molecular biology, life sciences, chemistry
Product Management	microsoft office, strategic planning, negotiation, microsoft excel, microsoft word
Public Policy	research, international relations, policy analysis, policy, sustainability
Recruiting	recruiting, human resources, employee relations, talent acquisition, performance management
Real Estate	real estate, investment properties, real estate transactions, residential homes, property management
Retail	retail, merchandising, customer service, sales, inventory management
Sales	sales, sales management, account management, customer service, new business development
Sales Management	strategic planning, management, customer service, new business development, negotiation
Social Media	social media, public relations, social media marketing, marketing, event management
Software Engineering	java, sql, software development, linux, agile methodologies
Tech. Product Mgmt	business analysis, requirements analysis, sql, business intelligence, project 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
Web Development	javascript, html, java, css, sql

Table 5: 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 raw firm-level return during month t, $Ret_{i,t}$, in column (1) and abnormal return, $AbnRet_{i,t}$, of firm i over the market in columns (2)-(3). 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 contemporaneous stock market return during month $t-L$. 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) Raw Return Raw Turnover	(2) Abn. Return Raw Turnover	(3) Abn. Return Abn. Turnover	(4) Abn. Return Abn. Turnover
L = 1 month				
Coefficient	-0.70%***	-0.66%***	-0.82%***	-0.84%***
(Standard error)	*(0.08%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 2 months				
Coefficient	-0.34%***	-0.60%***	-0.76%***	-0.78%***
(Standard error)	*(0.08%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 3 months				
Coefficient	-0.49%***	-0.60%***	-0.77%***	-0.78%***
(Standard error)	*(0.08%)*	*(0.08%)*	*(0.08%)*	*(0.08%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X
L = 6 months				
Coefficient	-0.17**	-0.48%***	-0.64%***	-0.65%***
(Standard error)	*(0.08%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Turnover orthogonal to:				
Size, B/M			X	X
Industry			X	X
Past returns				X

Table 6: Portfolio returns based on lagged employee turnover rates. In each month t, we sort firms based on their lagged abnormal employee turnover, $AbnTurnover_{i,t-L}$ into two portfolios: the high turnover portfolio consists of the firms in the top quintile based on the abnormal turnover measure, while the low turnover portfolio contains the bottom quintile of firms based on abnormal turnover. We compute the value-weighted returns of each portfolio during the current month t, as well as the return on the long-short portfolio that goes long the low turnover portfolio and short the high turnover portfolio.
The results are displayed using abnormal employee turnover lagged by one, two, three, and six months. Abnormal turnover is orthogonalized to firm size, book-to-market ratio, industry, and the lagged monthly returns.

Lag	(1) High Turnover Portfolio	(2) Low Turnover Portfolio	(3) Long-Short (Low - High)
L = 1 month			
Coefficient	0.12%	1.07%	0.95%***
(Standard error)	*(0.02%)*	*(0.15%)*	*(0.15%)*
Turnover orthogonal to:			
Size, B/M	X	X	X
Industry	X	X	X
Past returns	X	X	X
L = 2 months			
Coefficient	%	%	1.06%***
(Standard error)	*(0.08%)*	*(0.07%)*	*(0.17%)*
Turnover orthogonal to:			
Size, B/M	X	X	X
Industry	X	X	X
Past returns	X	X	X
L = 3 months			
Coefficient	0.10%	1.23%	1.12%***
(Standard error)	*(0.02%)*	*(0.16%)*	*(0.16%)*
Turnover orthogonal to:			
Size, B/M	X	X	X
Industry	X	X	X
Past returns	X	X	X
L = 6 months			
Coefficient	0.11%	1.10%	1.10***
(Standard error)	*(0.02%)*	*(0.14%)*	*(0.14%)*
Turnover orthogonal to:			
Size, B/M	X	X	X
Industry	X	X	X
Past returns	X	X	X

Table 7: 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 $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$,
where $AbnRet_{i,t}$ is the return for firm i in month t in excess of the market return. $AbnSkill_{i,t-L}$, 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 month $t - L$ performance of firm i. The table reports the coefficient β, scaled so as to correspond to the change in returns for every additional 10% of employees possessing a given skillset.

Skillset	(1) L = 1 month	(2) L = 2 months	(3) L = 3 months	(4) L = 6 months
Managerial & Administrative				
Business Development				
Coefficient	-0.08%***	-0.08%***	-0.08%***	-0.08%***
(Standard error)	*(0.02%)*	*(0.02%)*	*(0.02%)*	*(0.02%)*
Product Management				
Coefficient	-0.32%***	-0.32%***	-0.32%***	-0.33%***
(Standard error)	*(0.11%)*	*(0.11%)*	*(0.11%)*	*(0.11%)*
Administration				
Coefficient	-0.09%***	-0.08%***	-0.08%***	-0.06%**
(Standard error)	*(0.02%)*	*(0.02%)*	*(0.02%)*	*(0.02%)*
Human Resources (Jr.)				
Coefficient	-0.32%***	-0.32%***	-0.32%***	-0.31%***
(Standard error)	*(0.07%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Human Resources (Sr.)				
Coefficient	-0.08%	-0.07%	-0.06%	-0.06%
(Standard error)	*(0.06%)*	*(0.06%)*	*(0.06%)*	*(0.06%)*
Sales-Oriented Skills				
Sales				
Coefficient	0.07%**	0.07%**	0.07%**	0.07%**
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
CRM				
Coefficient	0.06%**	0.06%**	0.06%**	0.06%**
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Sales Management				
Coefficient	0.10%***	0.10%***	0.11%***	0.11%***
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Communications Skills				
Digital Marketing				
Coefficient	-0.00%	-0.00%	-0.01%	-0.01%
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*
Social Media				
Coefficient	-0.08%**	-0.08%**	-0.07%*	-0.07%*
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*
Public Policy				
Coefficient	-0.10%**	-0.10%**	-0.10%**	-0.10%*
(Standard error)	*(0.05%)*	*(0.05%)*	*(0.05%)*	*(0.06%)*

Table 7 (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 $L \in \{1\,mo, 2\,mo, 3\,mo, 6\,mo\} : AbnRet_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$, where $AbnRet_{i,t}$ is the return for firm i in month t in excess of the market return. $AbnSkill_{i,t-L}$, 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 month $t-L$ performance of firm i. The table reports the coefficient β, scaled so as to correspond to the change in returns for every additional 10% of employees possessing a given skillset.

Skillset	(1) L = 1 month	(2) L = 2 months	(3) L = 3 months	(4) L = 6 months
Financial Skills				
Accounting & Auditing				
Coefficient	-0.01%	-0.00%	-0.00%	-0.00%
(Standard error)	*(0.02%)*	*(0.02%)*	*(0.02%)*	*(0.02%)*
Banking & Finance				
Coefficient	-0.01%	-0.00%	0.00%	0.00%
(Standard error)	*(0.02%)*	*(0.02%)*	*(0.02%)*	*(0.02%)*
Insurance				
Coefficient	-0.05%*	-0.05%*	-0.05%*	-0.05%
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Technical Skills				
Web Development				
Coefficient	0.14%**	0.14%**	0.13%**	0.13%**
(Standard error)	*(0.07%)*	*(0.07%)*	*(0.07%)*	*(0.07%)*
Mobile				
Coefficient	0.02%	0.02%	0.02%	0.02%
(Standard error)	*(0.03%)*	*(0.03%)*	*(0.03%)*	*(0.03%)*
Data Analysis				
Coefficient	-0.19%*	-0.19%**	-0.18%**	-0.17%**
(Standard error)	*(0.09%)*	*(0.09%)*	*(0.09%)*	*(0.09%)*
Artistic Skills				
Web Design				
Coefficient	0.02%	0.02%	0.02%	0.01%
(Standard error)	*(0.08%)*	*(0.08%)*	*(0.08%)*	*(0.08%)*
Graphic Design				
Coefficient	0.01%	0.01%	0.00%	0.00%
(Standard error)	*(0.06%)*	*(0.06%)*	*(0.06%)*	*(0.06%)*
Video & Film				
Coefficient	0.00%	0.00%	0.00%	0.00%
(Standard error)	*(0.04%)*	*(0.04%)*	*(0.04%)*	*(0.04%)*

Table 8: Results from separate regressions of returns on lagged employee skillsets across three industries. We estimate the following specification for each industry:
$AbnRet_{i,t} = \alpha + \beta AbnSkill_{i,t-L} + \epsilon_{i,t}$
$AbnRet_{i,t}$ the return for firm i in month t in excess of the market return. $AbnSkill_{i,t-L}$, 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 month $t-L$ performance of firm i. The table reports the coefficient β, scaled so as to correspond to the change in returns for every additional 10% of employees possessing a given skillset.

Skillset	Information (NAICS 51)	Manufacturing (NAICS 31-33)	Finance & Insurance (NAICS 52)
Managerial & Administrative			
Administration			
Coefficient	-0.04%	0.03%	-0.05%
(Standard error)	*(0.03%)*	*(0.06%)*	*(0.03%)*
Product Management			
Coefficient	-0.54%***	-0.78%***	-0.65%***
(Standard error)	*(0.16%)*	*(0.23%)*	*(0.16%)*
Human Resources (Jr.)			
Coefficient	-0.26%***	-0.56%*	-0.38%***
(Standard error)	*(0.09%)*	*(0.23%)*	*(0.10%)*
Sales-Oriented Skills			
Sales			
Coefficient	0.07%**	0.04%	0.04%
(Standard error)	*(0.03%)*	*(0.05%)*	*(0.03%)*
Sales Management			
Coefficient	0.13%***	0.26%***	-0.02%
(Standard error)	*(0.04%)*	*(0.07%)*	*(0.03%)*
Financial Skills			
Banking & Finance			
Coefficient	-0.16%	0.26%***	0.00%
(Standard error)	*(0.10%)*	*(0.08%)*	*(0.01%)*
Insurance			
Coefficient	-0.09%***	0.62%**	-0.05%†
(Standard error)	*(0.03%)*	*(0.30%)*	*(0.03%)*
Technical Skills			
Web Development			
Coefficient	0.16%***	0.17%	-0.07%
(Standard error)	*(0.06%)*	*(0.15%)*	*(0.18%)*
Mobile Telecommunications			
Coefficient	-0.01%	-0.05%	0.93%***
(Standard error)	*(0.03%)*	*(0.05%)*	*(0.31%)*
Data Analysis			
Coefficient	-0.19%*	0.03%	-0.42%***
(Standard error)	*(0.10%)*	*(17%)*	*(0.18%)*