Northwestern Law

SEARLE CENTER
ON LAW, REGULATION,
AND ECONOMIC GROWTH

WORKFORCE SCIENCE PROJECT

Working PAPER SERIES • NO. 2015-02

Labor Market Peer Firms

Nan Li
Rotman School of Management
University of Toronto

WWW.LAW.NORTHWESTERN.EDU/SEARLECENTER

2015

Labor Market Peer Firms

Nan Li*

January 2014; this version, January 2015

Abstract

I show that labor market peer firms (LMPs) provide a new measure of economic linkages among firms. I identify LMPs by analyzing members' online viewing patterns on LinkedIn, a professional networking and corporate recruiting website. LMPs exhibit significant incremental power over standard output-market-based industry groupings to explain cross-sectional variation in base firms' stock returns and performance metrics. I create a measure of similarity across firms using LinkedIn members' self-reported labor market skills. When LMPs share more skills with the base firm, LMPs outperform standard industry groupings to explain return comovements.

*Nan Li is at Rotman School of Management, University of Toronto (nan.li@rotman.utoronto.ca). I thank members of my dissertation committee for their invaluable help and support: Douglas Skinner (Chair), Ray Ball, Gary Becker, Pingyang Gao, and Michael Minnis. I appreciate comments from and discussions with Philip Berger, Hans Christensen, Rui Cui, Pengjie Gao, Joseph Gerakos, João Granja, Zhiguo He, Yunzhi Hu, Erik Hurst, Filipe Lacerda, Christian Leuz, Meng Li, Mark Maffett, Sara Moreira, Valeri Nikolaev, Dan Nguyen, Elizabeth Pontikes, Haresh Sapra, Kelly Shue, Abbie Smith, Regina Wittenberg-Moerman, Wei Wu, Danxia Xie, Qiping Xu, Anastasia Zakolyukina, Sarah Zechman, Frank Shuang Zhou, and seminar participants at Chinese University of Hong Kong, Hong Kong University, Nanyang Technological University, National University of Singapore, New York University, Singapore Management University, UCLA, University of Chicago, University of Pittsburg, University of Toronto. I thank people at LinkedIn for helpful discussions. I acknowledge the generous financial support provided by the Accounting Research Center and PhD program at Booth School of Business, University of Chicago. All errors are my own.

1 Introduction

Understanding firms' economic linkages is central to economic analysis and policy making. Standard industry groupings are primarily based on product market relations and measure output but not necessarily input market shocks. Labor inputs, or human capital, play an important role in the production process (Becker [1962], Bresnahan et al. [2002], Lustig et al. [2011]), and potentially measure input market shocks. This study offers an innovative way to measure firms' economic linkages by identifying labor market peer firms (LMPs). I find that LMPs measure whether firms are exposed to common economic shocks, as reflected in earnings and stock return comovements.

Firms in the same product market encounter similar output market shocks that affect their revenues, whereas firms that use the same inputs, including common types of labor, experience similar input market shocks that impact their expenses. In addition, input and output market shocks (including market and industry risk throughout this paper) affect firms' stock prices.[1] When firms are exposed to common shocks, their stock prices and earnings comove. While it is difficult to identify or observe shocks, I rely on stock return and earnings comovement tests to infer the channels through which shocks affect asset prices. I use standard industry groupings to tease out common output market shocks. The incremental explanatory power of LMPs suggests that the labor market captures economically significant common input market shocks.

LMPs enable me to test the relation between the labor market and stock prices. Although each piece of evidence is potentially open to a different interpretation, together they support the idea that the labor market measures firms' exposure to common economic shocks. For instance, technology shocks to new industries likely change the outside options of a firm's human capital, shifting its labor costs and profits (Eisfeldt and Papanikolaou [2013]). Other studies further investigate the channels through which the labor market affects asset prices. Donangelo [2013] finds that firms in industries with highly mobile labor earn higher returns than those in industries with less mobile labor. Bazdrech et al. [2013] show that labor market activities react to both labor and capital adjustment costs in a dynamic neoclassical investment-based model and suggest that labor market frictions can have a significant impact on asset prices.

This paper is a joint test of two hypotheses. First, that the labor market measures common input market shocks to firm value so that LMPs have incremental power to explain stock return and earning comovements

[1] Peer firms might experience common exogenous supply and demand shocks, disaster shocks, and policy shocks that are one-time shifts to stock prices. Peer firms might also share common market or industry risk that is incorporated in stock prices. Stock price comovements can be driven by either exogenous one-time shocks, changes in risk factors, or a combination of shocks and risk. I do not disentangle shocks and risk and refer to all of them as shocks throughout this paper.

over output-market-based industry groupings. Second, that LinkedIn-implied LMPs are a good measure for labor market commonalities.

How do LMPs offer incremental power over standard industry groupings to measure firms' economic linkages that transmit economic shocks? First, LMPs measure demand and supply shocks to a firm's talent pool. Limits on specific labor could raise wage costs and impede a firm's production of goods. When firms hire common types of labor, they are subject to common economic shocks. Common shocks shift peer firms' labor costs and consequently their earnings. This commonality is reflected in stock return and earning comovements. Second, LMPs measure exposure to common non-labor input market shocks. For example, both chemical and consumer staple companies use crude oil as an input and both are exposed to oil price volatility. Identifying this type of commonality is difficult because chemical and consumer staple companies are in different product markets. However, this commonality could be detected through the labor market because they both hire chemical engineers to develop new products and oil analysts to hedge firms' oil price fluctuations.[2]

Identifying LMPs is challenging. Firms disclose little information about human capital (Aboody and Lev [1998], Lev and Zarowin [1999], Kanodia et al. [2004]). We only know the number of a firm's employees, R&D, SG&A expenses from its annual report, and the distribution of labor force skills at the aggregate industry level from government agencies. Existing literature examines how labor networks influence the transfer of knowledge and firm performance (Almeida and Kogut [1999], Cohen et al. [2010], Engelberg et al. [2013]). However, data on which firms hire common types of labor are not available.

This paper proposes a novel approach to identify LMPs using data from LinkedIn. LinkedIn is the world's largest professional network on the internet, with more than 200 million members. LinkedIn hosts company homepages for more than 3 million firms. Firms use LinkedIn as a recruiting channel. Often, and sometimes exclusively, firms post career-related news and new job openings on their LinkedIn company homepages. LinkedIn members view LinkedIn company homepages to learn about career opportunities. We can assume that a LinkedIn member's viewing history on LinkedIn company homepages reveals companies for which a member has the necessary work skills and for which the member might be interested in working. For each firm on LinkedIn, the site shows six other firms that members have "also viewed" (appear mostly frequently in members' adjacent views). For example, members who visited the LinkedIn homepage of the

[2] LMPs also measure output market shocks because firms in the same product market likely hire similar kinds of labor. I don't focus on LMPs' power to capture output market shocks as they are measured by product market peers.

electric car producer Tesla (SIC code 3711, Motor Vehicles) have “also viewed” the LinkedIn homepages of BMW (SIC code 3711), Ford (SIC code 3711), Solar City (SIC code 1700, Construction), Space Exploration Technologies (SIC code 3761, Guided Missiles and Space Vehicles), Google (SIC code 7375, Information Retrieval Services) and Apple (SIC code 3571, Electronic Computers). Obviously, Tesla’s LinkedIn “also viewed” firms are a mix of input and output markets peers. I classify the “also viewed” firms as the base firm’s LMPs.

The LinkedIn-implied LMPs detect firms’ commonalities that traditional industry groupings based on product market or production processes do not pick up. Seventy-two percent of S&P 1500 firms’ LMPs share the same two-digit GICS industry code with the base firm. In addition, a mere 34% of S&P 1500 firms’ LMPs have the same four-digit SIC industry code as the base firm. The fact that the six peer firms only have a 64% overlap with (on average) 282 peer firms in each four-digit GICS industry illustrates that LMPs measure commonalities beyond the output market.

LinkedIn-implied LMPs also measure the intensity of the relation between companies and can select groupings on the basis of closeness. LinkedIn members select a firm’s most relevant LMPs, whereas traditional industry groupings do not measure closeness among firms within a given industry.[3]

Consistent with my predictions, LMPs exhibit economically meaningful and significant incremental power to explain stock return comovements over standard industry groupings (SIC, NAICS, GICS) and text-based product market industry groupings (TNIC, Hoberg and Phillips [2010, 2013]). Standard industry groupings together explain 13.0% of cross-sectional return variation, whereas LMPs increase the explanatory power to 15.6%. Consistent with evidence from return-based comovement tests, LMPs also show incremental explanatory power over standard industry groupings to explain contemporaneous correlations in accounting-based fundamentals, including valuation multiples, profitability and expense ratios. If conventional industry groupings do a good job of measuring output market shocks, empirical evidence shows that LMPs capture economically important input market shocks to firm value.

To further explore firms’ labor market relationships, I create a skill-similarity measure that exploits a unique feature of my database. LinkedIn members post their skills to their online profiles. Each firm’s

[3]LinkedIn members’ viewing patterns likely identify firms’ most relevant peers and crucial linkages out of a potentially noisy and large sample of industry peers. For example, Discover Financial Services and Citibank are both financial firms with different business focuses. The LMPs of Discover Financial Services are Visa, American Express, Mastercard, Citibank, Capital One and J.P. Morgan Chase, whereas Citibank’s LMPs are HSBC, Goldman Sachs, Deutsche Bank, JP Morgan Chase, Morgan Stanley, and Bank of America. The differences in the LinkedIn-implied LMPs are consistent with the differences in the two firms’ business models: Discover’s LMPs are major credit card issuers, and Citi’s LMPs are large traditional banks.

LinkedIn company homepage summarizes the five top skills posted by their employees. For example, Deloitte's employees' top skills are: IFRS, Sarbanes-Oxley Act, Internal Controls, External Audit, and U.S. GAAP. For each skill, LinkedIn uses its database and a proprietary algorithm to arrive at a series of related skills. Using the top and related skills, I construct a database to proxy for each firm's employee skill set. I then calculate the ratio of common skills between any two firms in my sample. This ratio enables me to directly evaluate firms' labor market relatedness.

Using my skill-similarity measure, I find that the base firm shares more skills with LMPs than with non-peer firms, showing that my measure captures labor market relatedness. S&P 1500 firms on average have 11% of labor skills in common with their LMPs, but share only 6% of skills with non-peer firms that have at least one skill in common with the base firm. This suggests that LinkedIn-implied LMPs are a good measure for labor market commonalities.

I find that return comovements are stronger when the base firm shares more skills with its LMPs. LMPs explain 19.0% of the cross-sectional variation in returns when base firms share 21.3% of skills with their LMPs on average, whereas four-digit GICS peers explain only 14.0%. In contrast, when base firms share 1.3% of skills with LMPs, LMPs explain 3.5% of the cross-sectional variation in returns, whereas four-digit GICS peers explain 3.8% on average. Firms that share less labor skills with their LMPs have a high percentage of employees on LinkedIn, low turnover, high R&D expenses, high unionization rate, high growth, and are small.

The primary contribution of this paper is that I show LMPs provide a useful measure of firms' economic linkages, which are not easily measurable using traditional industry measures. I propose an innovative way to identify LMPs using members' viewing patterns on a professional social network, LinkedIn. To the best of my knowledge, this paper is the first to measure firms' common exposure to shocks through the labor market channel. I show that LMPs exhibit significant incremental power over output-market-based industry groupings to explain stock return and accounting-based performance metric comovements . I also show that LMPs outperform traditional industry groupings to explain contemporaneous stock returns when base firms share more skills with their peer firms. This measure has potential applications in diverse areas such as executive compensation, antitrust litigation, benchmarking in performance evaluation, among others.

This paper also contributes to the labor economics literature by providing a new method to proxy for skill sets at the firm level based on employees' self-reported skills. Labor market skills are an important dimension of production, although measuring them is challenging. The existing literature studies how firm and industry specific human capital affect wage fluctuations (Jaganathan and Wang [1996], Eiling [2013], Neal [1995], Weinberg [2001], Lazear [2009]), entrepreneurship and firm value (Lazear [2009]), IT skills and firm value (Tambe [2014]). To my knowledge, this paper is also the first to create a simple and intuitive skill-similarity measure across firms.

This paper highlights the usefulness of, and opens a door for, analyzing the information content of employee disclosures. The existing literature extensively studies the information content of disclosure from the corporate perspective but remains silent on the employee perspective (Healy and Palepu [2001], Beyer et al. [2010], Berger [2011]). This paper utilizes employees' career-oriented online viewing patterns to identify LMPs. It also uses employees' disclosures in their online profile to proxy for each firm's skill set and to create a skill-similarity measure. This paper joins the growing literature on how information technology helps shape a firm's information environment (Blankespoor et al. [2013]), and shows the usefulness of the latent intelligence of people's Internet search patterns (Da et al. [2011], Lee et al. [2012]).[4]

Several caveats are in order. First, stock return and earnings comovement tests imply correlation, not causality. LMPs measure firms' economic linkages that likely transmit common shocks to firm value, but I cannot identify specific shocks in this paper. Second, the LinkedIn-implied LMPs identification method is limited by the availability of "people also viewed" firms LinkedIn displays on its company homepages, although they are likely a firm's most relevant peers.

I present related literature and develop hypotheses in section 2. Then I discuss the LinkedIn data, LMPs, and the skill-similarity measure in section 3. I report the main results in section 4. I conduct robustness tests in section 5 and conclude in section 6.

2 Hypothesis Development and Related Literature

2.1 Labor Market, Skill-Similarity and Cross-sectional Variation of Return Comovements

In this section, I provide the main intuition and propose a stylized model of LMPs and earnings comovement in the Appendix. When LMPs hire common types of labor, shocks to common talent pool move wage costs together. Therefore, LMPs' earnings and stock prices comove. The earning comovements are stronger when LMPs share more common types of labor.

[4]The LinkedIn-implied LMPs measure is not directly related to firms' disclosure choices, but it has potential to help understand how common business conditions in the input market affect firms' disclosure choices.

Several earlier studies suggest that human capital is a risk factor in asset pricing. Mayers [1973] introduces the idea of non-marketable assets (human capital) to the CAPM and suggests firms' systematic risk and the risk of the market portfolio should include risk attributable to the existence of human capital. Jaganathan and Wang [1996] allow for the inclusion of human capital, measured by the growth rate in per-capita labor income and show that the human-capital beta loads in a time-varying CAPM framework.[5]

Recent research builds micro foundations through which human capital demands a risk premium in a frictional world. The intuition behind some of these theoretical models is that human capital is not perfectly contractible and employees have the option to explore better outside opportunities, and such outside options covary with the state of the economy, thus human capital intensive firms demands a higher risk premium. Eisfeldt and Papanikolaou [2013] develop a model in which shareholders and key talent share cash flows that accrue to a firm. Key talent is essential to the efficiency of a firm's production and is in scarce supply. The sharing rule depends on the outside opportunity of the key talent. Technology shocks to a new firm increase the outside options of the key talent in the old firm and thus decrease the shareholders' share of the surplus. Therefore, shareholders require a risk premium for being exposed to value loss from the threat that key talent may exit.[6]

2.1.1 LMPs as A Measure for Labor Market Commonalities

First, we need to know whether the LinkedIn-implied LMPs capture firms' crucial labor market commonalities. When measuring LMPs, my methodology focuses on the most skilled (and presumably most important or most educated) segment of the labor market. Twenty-nine percent of LinkedIn members have a graduate degree (including professional and doctoral degrees) and 50% have college degrees (excluding graduate degrees). According to the most recent U.S. Census Bureau's Current Population Survey (CPS), only 19.8% of U.S. population has a college degree (excluding graduate degrees), while only 11.1% of the U.S. population has a graduate degree (including professional and doctoral degrees). These numbers indicate that the employees recruited on LinkedIn are likely the essential labor for the firm and the pool of talent in scarce supply.

[5] Labor market activities likely measure shocks to both the labor and the capital markets. Bazdrech et al. [2013] propose an investment-based asset pricing model with labor and capital stochastic adjustment costs to explain the negative correlation between hiring and risk premiums. Firms with relatively high hiring rates are expanding firms that face high adjustment costs. If the economy experiences a shock that lowers adjustment costs, these firms will benefit the most from these lower costs, allowing these firms to grow faster and make profits more quickly.

[6] Other research studies the relation between general labor and asset prices. Donangelo [2013] uses labor mobility to measure workers' response to outside opportunities. He finds that firms in highly mobile industries earn returns 5.3% higher than those in less mobile industries.

For example, Dow Chemicals' LMPs are determined by chemists rather than factory workers. Bresnahan et al. [2002] show that innovations in information technology, complementary workplace reorganizations, new products, and services constitute a significant skill-biased technical change affecting labor demand in the Unites States. Therefore, the LinkedIn-member sample likely represents economically important human capital and allows us to peek inside the firm to measure firms' crucial linkages.

Hypothesis 1: LMPs is a good measure of firms' crucial labor market commonalities if the base firm and peer firms share labor market skills.

2.1.2 Incremental Power of LMPs over Output-Market-Based Industry Measures

We need to show that LMPs provide incremental power to explain returns due to firms' commonalities in the labor market but not the product market. It is likely that firms in the same labor market are also in the same product market, and variation in earning comovements comes from variation in the product market, or vice versa. In reality, firms' input and output markets do not always overlap, and both input and output market shocks impact firms' stock returns and accounting-based performance metrics. Standard industry groupings are informative about output but not necessarily input market commonalities. While it is difficult to identify or observe the source of shocks, I rely on standard industry groupings to tease out earning variations from product markets (common output market shocks).

In addition to earnings, accounting-based performance metrics are a useful way to measure whether LMPs capture common shocks to firm value. If LMPs capture exposure to common shocks that affect firm value, I expect to observe significant associations in earnings between the base firm and its peer firms (Brown and Ball [1967], Bhojraj et al. [2003], Leary and Roberts [2013]). For example, if LMPs identify fluctuations in the input market that are not captured by product market peers, I expect LMPs to do an incremental job of explaining expense ratios, such as SG&A and R&D ratios, because SG&A expenses include general staff's wages and R&D expenses include research scientists' wages.

Hypothesis 2a: LMPs have incremental power over output-market-based peer portfolios to explain the base firm's stock return variation.

Hypothesis 2b: LMPs have incremental power over output-market-based peer portfolios to explain the base firm's accounting-based performance metrics variation.

Hypothesis 1 focuses on testing whether LMPs provide a useful characterization of firms' labor market linkages, and hypothesis 2 tests whether the labor market provides a channel to measure firms' exposure to crucial common shocks beyond the output market. My paper is a joint test of hypotheses 1 and 2.

2.1.3 Skill-Similarity and Cross-sectional Variation of Return Comovement

I create a labor market skill-similarity measure to directly examine what LMPs capture and measure firms' labor market relatedness.[7] It's important to know the labor skills of different occupations and firms (Lazear [2004, 2009]), however measuring skill sets at the firm level is challenging. In this paper, I create a database to proxy for each firm's labor market skill set based on LinkedIn members' self-reported skills, and also compute skill similarity among firms.

If LMPs do a better job of measuring input market shocks than product-market-based peer portfolios, I expect LMPs to have advantages over standard industry classifications when firms are closely related in the labor market. According to Proposition 1in the appendix and equation (**??**), when the overlap of labor skills between the base firm and peer firms is larger, the contribution of labor markets to the return covariance is stronger.

Hypothesis 3a: LMPs explain a larger portion of stock return variation when the base firm and within group peer firms share more labor market skills. LMPs outperform output-market-based peer portfolios to explain stock return variation when the base firm and within group peer firms share more labor market skills.

R&D constitutes a firm's unique strength that can create endogenous barriers to entry (Sutton [1991]). Hoberg and Phillips [2013] find R&D can create unique products that appeal to quality-sensitive consumers. They find that higher ex-ante R&D level is correlated with lower ex-post product market similarity. Producing R&D extensive products requires labor with unique skills. Therefore, to differentiate themselves in the product market, firms likely differentiate in labor force skills as well. Hence, R&D intensive firms likely share fewer skills and are likely to be less connected in the input and output markets.The skill-similarity measure enables us to peek inside firms' labor inputs and open the black box.

Hypothesis 3b: R&D intensity is negatively correlated with skill-similarity between the base firm and its LMPs. The return comovement between the base firm and LMPs is weaker when the base firm is R&D intensive.

[7] It is possible that LinkedIn members' online viewing patterns reflect firms mentioned together in news media rather than firms' labor market commonalities. If that's true, I do not expect the base firm to share a large set of common skills with LMPs.

2.2 Research on Existing Industry and Peer Firm Groupings

Meaningful groupings of firms play a crucial role in many resource-allocation decisions. Business professionals and academic researchers identify peer firms for many purposes, including equity valuation (Brown and Ball [1967], King [1966], Fama and French [1997], Bhojraj and Lee [2002], DeFranco et al. [2012]), executive compensation (Albuquerque [2009]), corporate financial policy (Rauh and Sufi [2012], Leary and Roberts [2013]), information transfer (Foster [1981], Ramnath [2002], Cohen and Frazzini [2008]), accounting comparability (DeFranco et al. [2011]), risk management, and so forth. Industry groupings are also essential for understanding vertical and horizontal integration in industrial organizations (Fan and Lang [2000], Hoberg and Phillips [2010]).

I compare LMPs to a comprehensive set of output-market-based industry classifications. The Standard Industry Classification (SIC) System was developed by the Central Statistical Board in the 1930s and last updated in 1987. SIC is built on a demand-based conceptual framework, in that establishments are grouped into industries according to the similarity of their product markets. The North American Industry Classification System (NAICS) replace the SIC method in 1997 and is built on a supply-based conceptual framework in that establishments are grouped into industries according to similarities in the production processes. NAICS and SIC come from similar sources and research shows that they have similar power to explain firms' stock return comovement and financial multiples.[8]

The Global Industry Classification Standard (GICS), developed by Standard & Poor's and MSCI, offers a market-oriented industry classification. According to the GICS guidebook, companies are classified on the basis of their principal business activities. Specifically, a company's sources of revenue and earnings, as well as market perceptions revealed by investment research reports play important roles in the classification. Text Network Industry Classification (TNIC), developed by Hoberg and Phillips [2013], applies textual analysis to a firm's 10-K product descriptions, and groups firms into industries based on product similarities.[9] Bhojraj

[8]NAICS consider labor input commonalities in the industry classification process and is related with my LinkedIn-implied LMP measure more strongly than SIC. However, I use SIC instead of NAICS as one of my main comparison measure because SIC is more commonly used in the literature and the explanatory power of NAICS and SIC are similar in my sample.

[9]TNIC is a new product market oriented industry classification using text-based analysis of product descriptions in firms' 10-K reports. The idea of this measure is that firms operating in the same market have strong tendency to use similar product market vocabulary. Hoberg and Philips assign each firm a spatial location based on product words, generating a Hotelling-like product location space for each publicly traded U.S. firms. Then they calculate how similar each firm is to every other firm by calculating firm-by-firm pairwise word similarity scores, and classify peer firms based on the similarity scores. This classification system is like an intransitive network and each firm has its own set of distinct competitors. The LinkedIn-implied LMP measure is analogous to the network like TNIC measure in the sense that there is no static definition of industry. However, the LinkedIn-implied LMP is unique because it is based on the latent intelligence of LinkedIn members' online viewing patterns, and is a labor-market-oriented classification.

et al. [2003] compare different industry classifications and find GICS outperforms SIC, NAICS, the Fama and French 49 industry groupings (Fama and French [1997]) to explain contemporaneous correlations in stock returns.

In addition to grouping firms based on large scale industry classification schemes, some recent papers identify peer firms through Internet co-searches. Lee et al. [2012] identify peer firms using internet traffic patterns on the EDGAR website and find that firms appearing in chronologically adjacent searches by the same individual are fundamentally similar on multiple dimensions. EDGAR-traffic-based peer firm groupings outperform peer firms based on GICS to explain cross-sectional variation in base firms' stock returns and valuation multiples. Leung et al. [2012] identify peer firms using online search behavior of individuals who visit the Yahoo! Finance website. Building on the intuition behind Da et al. [2011] on agents' online search behavior and attention, these authors view the Yahoo! search peers as small subsets of stocks (habitats) on which investors focus their attention (Barberis et al. [2005]). They find that the base firm's stock returns comove with its search clusters. The cluster composition and comovement patterns change as investor attention shifts.

These internet-traffic-based measures are based on investors' perceptions, and probably work through a mix of fundamental and investor sentiment channels. In fact, the relatively frequent change of peer firms' composition is perhaps driven by such perception changes. In contrast, my LinkedIn-implied LMPs are connected clearly through the labor market and capture a new important dimension of the production functions beyond the output market.

A number of studies explore peer firm identifications in specific contexts. For example, DeFranco et al. [2012] show that analysts strategically select peer firms with high valuations in order to justify their valuation of target firms. Albuquerque [2009] shows that correctly specified peer groups are critical in evaluating the use of relative performance evaluation in CEO compensation. These studies highlight that objectively selecting peer firms in the presence of potential agency conflicts are particularly difficult.

LMPs overlap with but differ from peer firms disclosed in proxy statements released by firms in a number of important ways. First, peer firms disclosed in proxy statements are used as benchmark companies in firm performance and executive compensation evaluations. From an agency theory point of view, contextual identification of peer firms could be biased due to conflicted agents (Albuquerque et al. [2012]). There is no obvious reason why peer firm identification based on LMPs is biased one way or another. Second, from a practical point view, LMPs and actual peer firms from analyst research reports and proxy statements are quite different. Take the Dow Chemical Company, for example; four of its six LMPs are in its proxy statement: 3M, Procter & Gamble, DuPont, and Monsanto. The other two LMPs, BASF and Exxon Mobil, are not its proxy statement peers.[10]

3 Data and Empirical Strategy

3.1 LinkedIn-Implied LMPs

LinkedIn analyzes members' web viewing patterns on each company's LinkedIn homepage and shows six other firms that members who have viewed a given firm also view. I classify the base firm and the six "People Also Viewed" firms as LMPs. Each firm has a maximum of six LMPs, including both private and public peer firms. I assume that professionals only visit the LinkedIn homepages of companies for which they are interested in working, and companies that they have the skills necessary to work for. Therefore, these companies are related in the labor market and share the same talent pool.

LinkedIn is the world's largest online professional networking and corporate recruiting site. Through LinkedIn's proprietary platform, members share their professional identities online, engage with their professional networks, access shared knowledge and insights, find business and career opportunities. LinkedIn first rose to popularity among technology and finance industry employees, and now covers employees across industries. LinkedIn members range from college students to senior executives. As of 2013, LinkedIn counts executives from all 2013 Fourtune 500 companies as members. Members are active on LinkedIn, executing more than 5.7 billion professionally-oriented searches on the platform in 2012. LinkedIn members represent all age groups: 46% are between ages 25 and 44, whereas 35% are between ages 45 and 64.

Companies increasingly use LinkedIn as an important recruiting channel: 90 of the Fortune 100 companies use LinkedIn's corporate talent solutions. More than three million companies use LinkedIn. LinkedIn's revenues from recruiting services such as "Talent Solutions" doubled in 2011 and 2012, while its major online competitor Monster.com's recruiting revenue only grew by 2% in 2011. A report from Bersin & Associates finds a dramatic shift in recruitment spending towards professional and social networks and away

[10]Dow Chemical's other proxy statement peers in 2009 are Alcoa, Archer Daniels, Boeing, Caterpillar, Emerson Electric, Honeywell, Johnson & Johnson, Johnson Controls, Kraft Foods, Eli Lilly& Co, Pepsi, Pfizer, PPG Industries, Tyco International, United Technologies, and GE. Another example is Accenture. Accenture's 2009 proxy statement peers are: Automatic Data Processing Inc, Cisco, Computer Sciences, EMC, HP, IBM, Lockheed Martin, Marsh & McLennan, Microsoft, Oracle, SAIC, and Xerox. Accenture's LinkedIn-implied LMPs are IBM, Deloitte, Google, Oracle, HP, and Microsoft. Deloitte and Google are Accenture's labor market but not proxy statement peers.

from traditional agencies and job boards (Leonard [2011]). A standard LinkedIn company homepage has four sections as of 2013: Home, Careers, Products and Services, and Insights. Companies use their LinkedIn company pages as a recruiting and career-related advertising platform, so most information on the pages is career oriented. For example, IBM posts articles such as "How to Build a Leadership Pipeline for Women in the Workplace," whereas Apple simply puts up new job openings as line items on their LinkedIn homepages with a link to the job description: "Apple is hiring: Consumer Apps Support Engineer." LinkedIn members need to log in to view a LinkedIn company homepage, and therefore LinkedIn can track members' viewing patterns.

The base sample of firms in this paper is the S&P 1500 universe as of December 1, 2012. I manually search for the LinkedIn company homepages for each firm in the S&P 1500 universe. A large number of S&P 1500 firms host company homepages on LinkedIn. This search results in 1,452 unique firms, and 1,476 LinkedIn company homepages (some companies have multiple homepages). Then I collect the "People Also Viewed" firms for each company as of December, 2012. I also collect the "People Also Viewed" firms for the same base firms in July, 2013. The composition of each firm's LMPs remains largely consistent over the half year time interval. I use the July sample as my final sample. This results in 5,419 unique base and peer firms. I classify firms into private and public, domestic and foreign, subsidiaries and parent firms and search for their Gvkey and Permno identifiers in CRSP/Compustat. This results in 3366 public firms, including 932 subsidiaries, 468 foreign firms, and 1275 domestic private firms. I further restrict my sample to firms listed on the NYSE, NASDAQ, or AMEX, with a CRSP share code of 10 or 11; this results in 1,360 base firms from the S&P 1500 universe, including 460 S&P 500 firms (see Table 1, panel (a)).

Existing industry classifications such as SIC and GICS impose both a restrictive transitivity property and a symmetric property.[11] My approach relaxes both, leading to more accurate and useful groupings. The transitivity property causes less accurate industry groupings because, for example, two firms that are peers to a third firm may not be peers with each other. For example, Accenture, a management consulting firm, has IBM, Deloitte, Google, Oracle, HP, and Microsoft as LMPs.[12]. Although Deloitte and Oracle are both Accenture's peers, they do not share the same talent pool and my LinkedIn-implied method does not require

[11] In mathematics, a binary relation over a set is transitive whenever an element a is related to an element b, and b is in turn related to an element c, a is also related to c. A binary relation over a set is symmetric if it holds for all a and b in the set that if a is related to b then b is related to a. TNIC measure in Hoberg and Phillips [2013] relaxes the transitivity property but not symmetric property.

[12] Accenture does not compete with IT firms such as Google, Oracle, and Microsoft in the product market. It specializes in IT consulting and does compete for talent with these technology firms in the labor market.

them to be LMPs. However, they would be classified as peer firms under a system with transitivity, because transitivity means that if firms A and B are peers, and if firms A and C are peers, then firms B and C are also peers. The symmetric property is also problematic because it gives equal weight to large and small firms, single and multi-segment firms, although they may be of unequal importance in terms of competing with each other in the labor market. For example, under the symmetric property, if Google is Telsa's peer firm, Telsa has to be Google's peer firm as well. In reality, the relationship between Google and Tesla is asymmetric because Google is larger and operates in more markets than Tesla. My approach solves this problem by identifying a unique set of firms that are the closest labor market competitors to the firm in question. For example, Tesla's closest human-capital peers include Google even though Tesla is not one of Google's six most critical labor market peers.

Table 1, panel B shows the composition of the S&P 1500 firms' peer firms. Because some parent firm and subsidiaries have separate LinkedIn company homepages, I exclude peer firms that have the same parent as the base firm in my sample. Four hundred and fifty-seven firms have at least one peer firm from the base firm's family, whereas none of 928 base firms' peer firms are from the same family. Two hundred and eighty-four firms' six peer firms are all public firms, while 1,106 firms have at least three public peer firms. One thousand forty-six firms' LMPs are all domestic firms, whereas 1,008 firms have at least one private peer firms. This suggests that LMPs likely draw a complete picture of a firm's peers, including both private and public firms.

3.2 Skill Similarity Measure

Another innovation of this study is that I develop a firm-level pair-wise skills similarity measure based on LinkedIn members' self-reported skills and expertise. Employees list their skills in the Skills & Expertise section of their LinkedIn profiles. Their skills are endorsed by people within their network, and any LinkedIn members can see how many people endorse each one of their skills.[13] LinkedIn summarizes the five most common skills of a company's current employees on each company's LinkedIn homepage. LinkedIn then uses a proprietary algorithm to calculate 10 to 40 related skills for each of the top skills. For example, the Top Skills & Expertise of Google are: Google Adwords, Python, Machine Learning, MapReduce, and Search Advertising. The related skills for Machine Learning are: Feature Selection, Text Mining, Pattern

[13]Skill endorsement likely serve as a truth-telling mechanism although it may not be the starting purpose of this function; members incur reputation costs if they list skills they do not have on their online profiles.

Recognition, and so on.

I measure firm skill similarity $s_{i,j}$ as the fraction of common skills shared by firm i and firm j. Each firm is assigned approximately 40 to 100 skills based on their current employees' self-reported skills. The skill-similarity measure is the number of common skills shared by firm i and j, divided by the number of total skills of firm i and j. Formally,

$$s_{i,j} = \frac{|C_i \cap C_j|}{|C_i \cup C_j|}, \tag{1}$$

where C_i is the set of skills and expertise of firm i, and $s_{i,j}$ is the cardinality of the intersect of firm i and firm j's skill sets divided by the cardinality of the union of firm i and firm j's skill sets. For example, Google has 85 skills and Facebook has 63 skills. They have 37 skills in common, so their skill similarity is $s_{Google,Facebook} = \frac{37}{85+63-37} = 0.33$. I show the top skills and related skills of Google and Facebook in Appendix Table 5.

The advantage of the skill-similarity measure is that it enables within and across labor-market grouping comparisons. I measure the skill similarity between the base firm and every firm in my sample.

4 Results

In this section, I find that LMPs exhibit strong skill-similarity with their base firms. I show that LMPs have significant incremental power over traditional industry groupings to explain cross-sectional variation of returns, valuation multiples, profitability and expense ratios. Consistent with the idea that labor market skills reflect an important dimension of production, whenever a cluster of firms exhibits stronger commonality of skills, LMPs better explain cross-sectional variation of returns.

4.1 Summary of LMPs and Skill Similarity

Table 2 summarizes the degree of concordance between each firm's labor market and product market peers. Specifically, I report the proportion of the LinkedIn-implied LMPs that belong to the same two-digit GICS (GICS2), four-digit GICS (GICS4), two-digit SIC (SIC2), three-digit SIC (SIC3), three-digit NAICS (NAICS3), and six-digit NAICS (NAICS6) as the base firm.

The degree of overlap between input and output markets differs across industries. At one end of the spectrum, energy and healthcare firms have more than 74% of their LMPs from the same two-digit GICS industries, suggesting that the necessary skills for employment in these industries are relatively industry-specific. Financial and information technology firms also have more than 74% of LMPs from the same two-digit GICS industries as the base firm.

At the other end of the spectrum, materials, consumer staples, and industrial firms have the fewest (approximately 60%) LMPs from the same two-digit GICS industry codes. When firms employ labor with skills that are portable across related industries, they have fewer LMPs from the same product market. One example is Dow Chemical Company, which belongs to the materials industry. Its LinkedIn-implied LMPs cover five two-digit GICS industries: BASF (chemicals, materials), Dupont (chemicals, materials), 3M (mechanical and industrial engineering, industrials), Exxon & Mobil (energy), Monsanto (biotechnology, healthcare), and Procter & Gamble (consumer staples). Dow is connected with its LinkedIn-implied LMPs because they all use chemicals in the production processes and they all hire people with chemistry expertise to conduct R&D. This linkage is identified plausibly through potential employees with chemistry expertise's viewing behaviors on LinkedIn company homepages, when they are searching for jobs.

Table 3 summarizes skill-similarity between the base firm and LMPs, and shows that LMPs capture labor market linkages. Under Hypothesis 1, if LMPs do not capture firms' labor market relatedness, I expect the average skill similarity to be the average of any random firm pairs. To impose a higher standard than random firm pairs, I calculate skill similarity between the base firm and any other firms in my sample that share at least one skill with the base firm, and use this as the benchmark. Table 3 shows that S&P 1500 firms' share on average 11% of skills with their LMPs, whereas share only 6.5% of skills with the benchmark groups. Firms in the information technology industry share the fewest skills with their peer firms (6%), whereas energy, utilities, and financial firms share 18%, 15%, and 14% of skills with their LMPs, respectively.[14] In general, the more skills the base firm shares with its peer firms, the more related they are in the labor market. This results validates Hypothesis 1 that LMPs capture labor market relatedness.[15]

[14]It is worth noting that we need to be cautious when comparing the skill similarity measure across industries because the fineness of skill definition may vary across industries. For example, IT firms have very specific delineation of skills, each programing language, such as C, JAVA, Perl are listed as different skills. Financial industries have broader definitions of skills, such as corporate finance, mortgage lending, and so on.

[15]It is difficult to validate my results externally because there are no data available elsewhere that keeps track of a firm's skill set. Donangelo [2013] provides a measure of labor market mobility but the focus of the type of labor in that paper is less on human capital with specific skills and in scarce supply.

4.2 Return Comovements of LMPs

An important aspect of firms' economic linkages is degree of contemporaneous correlations in their stock returns. I test Hypothesis 2a in this section. I investigate the extent to which LMPs help explain stock return comovements over standard industry classifications of peer firms. I exclude private and foreign peer firms because of stock return data availability, and keep base firms with at least three public peer firms in my sample (about 80% of the S&P 1500 firms). My results are robust to using the whole sample.

My first test compares the ability of LMPs with peer firms based on existing industry classification schemes to explain the base firm's stock return variation.[16] I first estimate the regression specification

$$R_{i,t} = \alpha_t + \beta_t R_{peer,t} + \varepsilon_{i,t} \tag{2}$$

where $R_{i,t}$ is the monthly stock return for each base firm i drawn from CRSP monthly files, $R_{peer,t}$ is the equally weighted average monthly portfolio return based on either the LMP, or four-digit GICS (GICS4), or four-digit SIC (SIC4), or six-digit NAICS (NAICS6) or TNIC (Hoberg and Phillips [2013]). I estimate equation (2) across firms for every month from 2003 to 2012 and obtain an average adjusted R^2 based on the 120 regressions. This regression specification is analogous to that in Bhojraj et al. [2003]. I show that my results are robust to using value weighted instead of equal weighted monthly returns in section 5.

If LMPs measure firms' connections beyond the output market, it should have incremental explanatory power over standard industry classifications to explain the base firm's stock return variation. When firms are connected by important economic linkages, they are exposed to common economic shocks. I test the regression specification

$$R_{i,t} = \alpha_t + \beta_{LMP,t} R_{LMP,t} + \sum_{output-based} \beta_{peer,t} R_{peer,t} + \varepsilon_{i,t}, \tag{3}$$

where $R_{LMP,t}$ is the equally weighted average monthly portfolio return based on LMPs, and $R_{peer,t}$ is the equally weighted average monthly portfolio return based on GICS4, SIC4, NAICS6 or TNIC.

Table 4 reports monthly regression results for equations (2) and (3). I focus on R^2 as the main measure for explanatory power. Table 4, columns (1) to (5) report estimation results for equation (2). I find that

[16] A firm's LMPs have at most six public peer firms. In my sample, there are about 25 four-digit GICS industries, and firms have on average 285 peer firms with the same four-digit GICS code, with a maximum of 711 peer firms. There are about 449 four digit SIC industries and firms in my sample have on average have 79 peer firms with the same four-digit SIC and a maximum number of 465 peer firms. A firm has 285 TNIC peers on average.

LMPs significantly outperform GICS4, SIC4, NAICS6 and TNIC peer portfolios. LMPs explain 10.4% of the cross-sectional variation in realized returns on average. Output-market-based or investment-oriented industry groupings explain about 6.9% (NAICS6), 7.4% (SIC4), 8.7% (TNIC), 9.2% (GICS4) of the cross-sectional variation in realized returns on average. Pairwise t-tests show differences of the average R^2 between LMPs and existing industry measures are statistically significant at the 1% level.

Table 4, columns (6) to (7) show the incremental power of LMPs to explain cross sectional returns over existing industry groupings. Bhojraj et al. [2003] shows that GICS outperforms SIC, NAICS, Fama and French 49 industries peer portfolios to explain the base firms' return variation. In column (6), GICS4, SIC4, NAICS6, TNIC altogether explain on average 13.0% of cross-sectional variation in realized returns. In column (7), LMPs increase the explanatory power to 15.6%. My results are robust to controlling for market returns (see robustness tests).

A firm's LMPs include at most six public peers due to data availability from LinkedIn, whereas a firm's alternative industry peers include a larger number of peers. Taking average of a larger number of peer firms might either bias for or against alternative industry groupings. Taking the average of more firms help cancel out the idiosyncratic component in stock returns, whereas a smaller set of more relevant firms should perform better than less relevant firms. It is an empirical question. To make a fair comparison, I randomly select the same number of firms as the number of LMPs out of the base firm's GICS4 group and calculate the equal-weighted return (Random GICS4). Table 5, column 2 shows that random GICS4 only explain 2.3% of cross-sectional variation in returns on average and LMPs increases the explanatory power to 11.4%. Traditional industry groupings provide a group of output-market peers without a ranking of relevance. The LinkedIn-implied LMPs potentially show a firm's most relevant peers in the labor market.

It is also possible that LMPs select a firm's most relevant product market peers without additional information about the input market. I construct a "best" 4-digit SIC peer firms to mimic a group of peer firms comparable to LMPs that have the most relevant capital market performance with the base firm within the same product market group. Specifically, for each base firm, I select the same number of firms as the number of LMPs within the base firm's 4-digit SIC industry that have the highest return correlation with the base firm. For each base firm, a pooled time-series cross-sectional regression is estimated with the monthly stock returns of each firm in the base firm's 4-digit SIC code as the independent variable. I rank firms based on the regression R^2 and select the same number of firms with the highest R^2 as the "best" 4-digit SIC peer firms. Then I calculate the equal-weighted "best" performance return based on the "best" 4-digit SIC peer portfolios. If LMPs select peer firms based only on the output market, I expect to see no incremental power of LMPs over the "best" 4-digit SIC peers. LMPs also explain return comovement over the "best" SIC4 at the 1% significance level, from 14.2% to 17.4%. The results suggest that LMPs is not a mere selection based on the product market.

In sum, LMPs increase the explanatory power of existing industry peer portfolios in terms of their ability to explain cross-sectional variation in base firm returns. If GICS4, SIC4, NAICS6 and TNIC peer portfolios together do a good job of capturing common output market shocks, the results in Table 4, 5 and 5 illustrate that that LMPs measure important common input market economic shocks. My results achieve an improvement of 20% in R^2 for the equally weighted peer portfolios relative to the aggregate power of GICS4, SIC4, NAICS6 and TNIC in S&P 1500.

4.3 Accounting Ratios and Valuation Multiples

Another measure of firms' connections is the degree of contemporaneous correlation in their accounting-based performance ratios and valuation multiples. I test Hypothesis 2b in this subsection and gather annual data from Compustat on a range of accounting ratios and valuation multiples, including expense ratios, profitability ratios, and employment ratios. If LMPs do a better job than product-market-based peer portfolios of measuring common input market shocks that affect firm value, I expect LMPs to have incremental power over GICS4 and SIC4 to explain cross-sectional variation in employment growth rates and expense ratios that are directly related to the labor market. Specifically, I look at the one-year-ahead realized number of employees growth rate (empgrowth), R&D expenses scaled by net sales (rdpersales), SG&A expenses scaled by net sales (sgapersales), and the sum of R&D and SG&A expenses scaled by net sales (rdsgapersales).

Brown and Ball [1967] show that industry earnings explain a substantial amount of the earnings of an individual firm. LMPs should have incremental explanatory power over GICS and SIC to explain profitability ratios and valuation multiples if LMPs measure input market shocks that affect firm value. Following Bhojraj et al. [2003], Lee et al. [2012], I include valuation multiples such as the price to book ratio (pb), enterprise value to sales ratio (evs), price to earnings ratio (pe), and performance ratios such as the return on asset (roa), return on equity (roe), leverage (lev), and one-year-ahead realized sales growth (salesgrowth) to proxy for firms' profits in my stylized model. The calculation and definition of accounting-based performance metrics are detailed in the Appendix Table 4.

I first estimate the cross-sectional baseline regression of the form for each variable

$$Ratio_{i,t} = \alpha_t + \beta_{1,t} Ratio_{GICS,t} + \beta_{2,t} Ratio_{SIC,t} + \varepsilon_{i,t} \tag{4}$$

and then I add the ratio of LMPs to the above regression

$$Ratio_{i,t} = \alpha_t + \beta_{LMP,t} Ratio_{LMP,t} + \beta_{1,t} Ratio_{GICS,t} + \beta_{2,t} Ratio_{SIC,t} + \varepsilon_{i,t} \tag{5}$$

where $Ratio_{i,t}$ is the variable of interest for each base firm i, and the independent variable $Ratio_{LMP,t}$, $Ratio_{GICS,t}$, $Ratio_{SIC,t}$ is the average yearly peer portfolio ratios, based on either LMPs, or GICS4, or SIC4. I estimate these regressions on a yearly basis for each calendar year from 2003 to 2012. I use yearly ratios because LMPs cover different industries, and different industries are more comparable at yearly levels due to seasonality issues. For the entire sample, I drop observations with missing total assets, long term debt, net income before extraordinary items, or operating income after depreciation. I also drop observations with negative common or total equity, share price smaller than \$3 at the end of the fiscal year, and net sales smaller than \$100 million (Bhojraj et al. [2003]). My results are robust to including the dropped observations. To mitigate the effect of outliers, I winsorize observations at the 1st and 99th percentiles for each variable in each regression.

Table 6 presents the regression results. Column (1) shows the number of observations for each estimation. Columns (2) and (3) report the the average value of adjusted R^2 on yearly cross-sections of the above regression with different dependent variables for equation (6) and (7), respectively. Column (4) reports the estimate of $\beta_{LMP,t}$ and its significance level.

Consistent with the idea that LMPs provide an input-oriented industry measure, LMPs exhibit significant incremental explanatory power to the aggregate power of GICS and SIC to explain cross-sectional variation in expense ratios, including rdpersales, sgapersales, rdsgapersales. For example, the baseline regression produces an adjusted R^2 of 0.580 for rdsgapersales. When I take into LMPs, the adjusted R^2 increases by 9.1% to 0.633. The coefficient estimate on the portfolio ratio of LMPs report in column (4) is 0.221 and statistically significant at the 1% level. This result suggests that one standard deviation increase in LMPs' rdsgapersales implies 4.0% increase in that of the base firm.

The coefficient estimate on the portfolio accounting-based performance ratios of LMPs and all statistically significant at the 1% level, except pe, which is statistically significant at the 5% level. Together, Table 6 shows that LMPs measure common input market shocks that affect firms' valuation multiples, profits, and expenses.

4.4 Cross-Sectional Variation of Return Comovements

In this section, I test Hypothesis 3 and show that LMPs do an increasingly better job when the base firm share more skill with its labor market peers. Specifically, I examine situations where the base firm and its LMPs are more likely to share common input market shocks, and investigate whether LMPs outperform traditional industry groupings when the base firm shares more labor market skills with its peers, or when the base firm is less R&D intensive.

I divide the sample into three quantiles based on the base firm's characteristics. First, for each quantile I examine the explanatory power of LMPs by estimating

$$R_{i,t} = \alpha_t + \beta_t R_{LMP,t} + \varepsilon_{i,t}, \tag{6}$$

where $R_{i,t}$ is the monthly stock return of the base firm and $R_{LMP,t}$ is the monthly average return of LMPs. Second, for each quantile I examine the explanatory power of GICS4 analogously and estimate

$$R_{i,t} = \alpha_t + \beta_t R_{GICS4,t} + \varepsilon_{i,t}, \tag{7}$$

where $R_{GICS4,t}$ is the average monthly return of GICS4 peer firms. I estimate cross-sectional regressions of the above equations for every month from 2003 to 2012 and obtain an average adjusted R^2 based on the 120 regressions. I compare the explanatory power of each peer measure and test when LMPs outperform traditional industry measures in measuring common shocks that affect firm value. The results are similar if I replace GICS4 with SIC4, NAICS6 or TNIC.

4.4.1 Skill-Similarity and Return Comovements

I group firms based on the skill similarity between the base firm and peer firms into three quantiles. Skill similarity is defined in equation (1) and measures the proportion of the number of common skills shared by firm i and firm j, divided by the total skills of firm i and firm j. Untabulated results show that small, growth and R&D intensive firms share more skills with their LMPs. For the following estimation, I take the average skill similarity between the base firm and its public LMPs, excluding its subsidiaries or affiliated firms. Therefore, each base firm is assigned with a within labor market peer group skill similarity score. Then I sort the S&P 1500 sample into three groups based on the base firm's skill similarity score. Firms in the low skill similarity group on average share 1.2% of skills with their LMPs. Firms in the median skill similarity group share on average 8.4% of skills, and the high skill similarity groups have 21.3% skills in common. According to equation (**??**) and Hypothesis 3a, I expect LMPs to have higher explanatory power of the base firm's return variation when the base firm shares more labor market skills with its LMPs.

For each skill similarity group, I estimate equation (6) in columns (1) to (3) in Table 7 , test equation (7) for each skill similarity group in columns (4) to (6) in Table 7. For each group, I run the regression every month from 2003 to 2012 and obtain an average coefficient and adjusted R^2.

Table 7 shows the regression results for different skill similarity groups. I show that labor market relatedness is positively related with the amount of common shocks that affect firm value. The R^2 for LMPs increases monotonically and significantly when within group firms share more labor market skills, from 0.035 to 0.101 to 0.190. LMPs have about five times more explanatory power in the high skill similarity group than in the low skill similarity group. R^2 for GICS4 are also positively correlated with the ratio of shared labor market skills. R^2 for SIC4 increases monotonically and significantly when within group firms share more labor market skills, from 0.041 to 0.096 to 0.139.

Next, I examine when LMPs outperform peer groups based on GICS4. The R^2 of columns (1) and (4) is not significantly different, showing that LMPs and GICS4 both explain about 4% of return variation when firms share the least skills. In columns (3) and (6), LMPs explain 19.0% of cross-sectional return variation, significantly higher than the 14.0% by GICS when the base firm and its peers have more skills in common. This suggests that when firms share more labor market commonalities, LMPs do a better job of measuring common shocks.

Taken together, these results supports Hypothesis 3a and show that LMPs do an increasingly good job when the base firm share more labor market commonalities with its within group peers. In addition, Second, LMPs outperform traditional measures based on the output market to explain stock return variation when firms are more closely connected in the labor market.

4.4.2 R&D, Skill Similarity and Return Comovements

I test Hypothesis 3b and compare the explanatory power of LMPs in different R&D intensity groups. In Table 8, columns (1) to (3), I find the explanatory power of LMPs decreases monotonically from 14.2% to 8.7%, and then to 4.0% when R&D ratio increases. Then I estimate the performance of GICS4 using equation (7) in columns (4) to (6). LMP explains 14.2% of cross sectional variation when the base firm has zero R&D expense, significantly outperforming that of GICS (11.4%) and SIC (11.0%). LMPs are insignificantly different from GICS4 peer portfolios and explain about 4% of cross-sectional return variation when R&D ratios are high. The results is consistent with the idea that LMPs outperform existing peer portfolios when the base firm and its peer firms share more labor market skills.

One possible explanation for the above results is that R&D intensive firms share less skills with their peers and have a smaller wage variance term in equation (**??**) in the stylized model. Hoberg and Phillips [2013] find that R&D can create unique products that appeal to quality-sensitive consumers, making it more expensive to enter. They find that a higher ex-ante R&D level is correlated with a lower ex-post product market similarity. To create unique products, firms need to hire labor with unique skills to conduct R&D. Consequently, R&D intensive firms need to differentiate and thus share fewer skills with their peer firms. This hypothesis is challenging to test but my skill similarity measure enables me to test this hypothesis. Consistent with this idea, I find a significant negative correlation (-0.11) between R&D intensity and skill similarity. The base firm shares 13.2% of skills with peer firms when the base firm has R&D expense of zero. The base firm share 7.4% of skills with peer firms when R&D expense is low, and 6.4% of skills when its R&D expense is high.

5 Robustness Tests

In this section, I show that LMPs is not a mere reflection of firm performance, and it reflect firms' fundamental linkages that are persistent over time. I show that my results are robust to controlling for market returns, different test specifications and sampling criteria.

It is possible that LinkedIn members' viewing behavior is driven by similar transient firm performance, and therefore LMPs do not reflect firms' fundamental linkages. I expect LMPs to have stronger explanatory power to explain returns in more recent periods but not in past periods under this story.

Appendix Table 1 shows that LMPs do not exhibit better performance in the most recent period. I divide my sample into three year groups: 2003 to 2007, 2008 to 2010, 2011 to 2012, and estimate equation (4) separately. I compare the explanatory power of LMPs for the above sample periods in columns (1) to (3). LMPs explain 10.1% of cross-sectional variation of returns from 2003 to 2007, 12.5% from 2008 to 2010, and 9.6% from 2011 to 2012, and LMPs' explanatory power in 2003 to 2007 is not significantly different from that in 2011 to 2012.

Next I test the incremental power of LMPs over existing industry groupings to explain return variation in different sample periods. Appendix Table 1, columns (4) and (5) show that the explanatory power of LMPs increases from 13.0% to 15.0% when LMPs are included in 2003 to 2007. R^2 increases from 14.8% to 17.8% from 2008 to 2010 in columns (6) and (7) and from 10.5% to 13.6% from 2011 to 2012. There is no particular trend in the incremental power of LMPs over time. These findings suggest that LMPs is not a pure reflection of firm performance and reflect fundamental connections that are persistent over time.

I also examine the explanatory power of LMPs after controlling for firm characteristics in Fama and French [1992], Jegadeesh and Titman [1993]. I include log of book to market $Log(B/M)_{i,t}$, log of market cap $Log(Size)_{i,t}$, and $Momentum_{i,t}$, which is the average of the stock's own return from time $t-12$ to $t-1$, to equation (2) and equation (3), and estimate the regression specifications

$$R_{i,t} = \alpha_t + \beta_{1,t} R_{peer,t} + \beta_{2,t} Log(B/M)_{i,t} + \beta_{3,t} Log(Size)_{i,t} + \beta_{4,t} Momentum_{i,t} + \varepsilon_{i,t} \tag{8}$$

and

$$R_{i,t} = \alpha_t + \beta_{LMP,t} R_{LMP,t} + \beta_{1,t} R_{peer,t} + \beta_{2,t} Log(B/M)_{i,t} + \beta_{3,t} Log(Size)_{i,t} + \beta_{4,t} Momentum_{i,t} + \varepsilon_{i,t} \tag{9}$$

The main measure of interest is still the average R^2. Table 2 shows the monthly regression results for equations (8) and (9). LMPs explain about 13.3% of the cross-sectional variation in realized returns on average in Table 2, column (1), controlling for growth, size and momentum. Columns (2) to (5) show that the existing industry peer portfolios explain about 10.1% to 12.7% of the cross-sectional variation in realized returns on average. Columns (6) and (7) show that LMP significantly increases the explanatory power of existing industry groupings (SIC4, GICS4, NAICS6, TNIC altogether) from 16.0% to 18.1%.

I next show that my results are robust to controlling for market returns and different sampling criteria.

It is likely that most of the return variation come from the market. King [1966] shows that about half of a stock's variance is explained by an element of price change that affects the whole market, and industry effects account for perhaps an additional 10%. I run pooled regressions in Appendix Table 3, cluster standard errors at the firm level and include year fixed effects (Petersen [2009]), and use value-weighted returns for all variables in Appendix Table 3.[17] I first estimate

$$R_{i,t} = \alpha_t + \beta_1 R_{SIC4,t} + \beta_2 R_{GICS4,t} + \beta_3 R_{NAICS6,t} + \beta_4 R_{TNIC,t} + \beta_5 Market_t + \varepsilon_{i,t} \tag{10}$$

and then estimate the above equation with LMPs,

$$R_{i,t} = \alpha_t + \beta_1 R_{LMP,t} + \beta_2 R_{SIC4,t} + \beta_3 R_{GICS4,t} + \beta_4 R_{NAICS6,t} + \beta_5 R_{TNIC,t} + \beta_5 Market_t + \varepsilon_{i,t} \tag{11}$$

Appendix Table 3 of column (1) includes the average stock returns of existing industry classifications. The adjusted R^2 in column (1) is 0.335. Then I add the average monthly returns of my LMPs in column (2). The adjusted R^2 increases to 0.341.

Following Rauh and Sufi [2012], I use F-test of a single linear restriction to test whether the increase in R^2 from column (1) to column (2) of Appendix Table 3 is statistically significantly different from zero at a reasonable confidence level. The F statistic when examining the increase in R^2 from column 1 to column 2 is 670, which implies an easy rejection of the null hypothesis that the return of LMPs does not lead to an increase in explanatory power of the regression.

Then I show that the performance of LMPs is robust to the number of public firms in the LMPs and robust over time. In Appendix Table 3, columns (3) and (4), I require that each base firm have at least four public LMPs and restrict my sample period from 2003 to 2010. The adjusted R^2 increases from 0.346 to 0.351. The F-statistic when examining the increase in R^2 from column 1 to column 2 is 276, which implies an easy rejection of the null hypothesis that the return of LMPs does not lead to an increase in explanatory power of the regression. In Appendix Table 3, columns (5) and (6), I show the performance of LMPs for the more recent years. I restrict my sample to firms with at least four public LMPs and from 2010 to 2012. LMPs increase the adjusted R^2 from 0.369 to 0.374. The F-statistic is 82, and rejects the null hypothesis that the return of LMPs does not lead to an increase in the explanatory power of the regression.

[17] I assume the residuals of a given firm's stock returns are correlated over time and each month's residuals are correlated over different firms. In my main tests, I report the average of monthly cross-sectional regressions and assume the standard errors are correlated over time.

6 Conclusion

This paper shows that labor market peer firms (LMPs) measure firms' shocks to value in both input and output markets. Standard industry groupings based on product market similarities fall short of identifying input market relatedness, an important dimension of a firm's business environment. I show that grouping firms based on their labor market similarities is well suited to measure both input and output markets shocks, as reflected in earnings and stock return comovements.

Identifying LMPs is inherently difficult because no data are available to track firms hiring similar labor. I propose an innovative way to identify LMPs using LinkedIn members' viewing behaviors on LinkedIn company homepages. I also create a measure of employee skill-similarity across firms to directly measure firms' relatedness on the labor market. I find that LMPs provide significant incremental power over output-market-based industry groupings to explain firms' cross-sectional returns and accounting-based performance metrics variation. LMPs explain 19.0% of cross-sectional variation in returns, whereas output-market-based industry groupings explain at most 14.7% when firms exhibit a stronger skill set commonality. I use LinkedIn members' disclosure about their work history to create a labor turnover measure, and find that LMPs do a better job to explain returns when firms hire more specialized labor, holding the level of skills constant.

This paper uses labor market participants' revealed preferences and shows the latent intelligence of search behaviors. LinkedIn members select a firm's most relevant LMPs. This is another advantage of the LinkedIn-implied measure – it measures the intensity of the relation between companies and can select groupings on the basis of closeness. In addition, the LinkedIn-member sample likely represents economically important human capital, and therefore, my methodology likely identify firms' crucial linkages.

This paper joins the growing literature on how social media helps shape firms' information environments. My results highlight the usefulness of employee disclosures. Traditional disclosure studies focus on the information content of corporate disclosures, however research on employee disclosures is scant. This paper is a first step in research on employee disclosures, and I look forward to future research projects in this area.

In contrast to other research on peer firms and industry groupings, the LinkedIn setting offers a unique labor market perspective. Industry groupings are useful in accounting, economics, and finance. I believe that the concept of LMP will be valuable for peer firm selections in executive compensation, analyst reports, antitrust litigation, risk management, controlling for fundamentals in academic research, policy recommendations and other related areas.

Appendix: A Stylized Model of LMPs and Earning Comovements

I propose a stylized model of LMPs and earning comovements. When LMPs hire common types of labor, labor market shocks are reflected in earning comovements of LMPs. The earning comovements are stronger when LMPs share more common types of labor. I also examine the relation between LMPs and labor specificity.

The stylized model represents two firms, 1 and 2. They share same types of specific labor in the production process. Specific labor is in limited supply in the short run. To keep the focus on the dynamics of labor, capital is assumed away.[18] Outputs of firm 1 and 2 are given by:

$$Y_1 = \sum_{s \in S_1} \beta_{1s} L_{1s}^{\alpha_s},$$

$$Y_2 = \sum_{s \in S_2} \beta_{2s} L_{2s}^{\alpha_s},$$

where S_1 and S_2 are the labor skill sets used by firm 1 and 2 respectively, L_{1s} and L_{2s} denotes the mass of labor with skill s employed, $0 < \alpha_s < 1$ so that firms' production functions are decreasing returns to scale, and β_{1s} and β_{2s} are parameters that denote production technologies.

The firm's profit is $\pi_i = p_i Y_i - \sum_{s \in S_i} w_s L_{is}$, where $i = 1,2$. Each firm chooses labor to maximize the profit. The first order condition is $p_i \beta_{is} \alpha_s L_{is}^{\alpha_S - 1} = w_s$. Thus, the demand for each labor from each firm is $L_{is} = (\frac{w_s}{p_i \beta_{is} \alpha_s})^{\frac{1}{\alpha_s - 1}}$. Perfect competition in the labor market drives firms to equate the marginal profitability of skills employed to wages for the particular type of skill. Hence firms' profits are:

$$\pi_1^* = \sum_{s \in S_1} (\alpha_s p_1 \beta_{1s})^{\frac{1}{1-\alpha_s}} ((\alpha_s p_1)^{-1} - 1) w_s^{\frac{\alpha_s}{\alpha_s - 1}},$$

$$\pi_2^* = \sum_{s \in S_2} (\alpha_s p_2 \beta_{2s})^{\frac{1}{1-\alpha_s}} ((\alpha_s p_2)^{-1} - 1) w_s^{\frac{\alpha_s}{\alpha_s - 1}},$$

where p_1 and p_2 are product prices for firm 1 and 2, w_s is the wage for labor skill s, and demand for labor skill s are $L_{1s} = (\frac{w_s}{\alpha_s p_1 \beta_{1s}})^{\frac{1}{\alpha_s - 1}}$ and $L_{2s} = (\frac{w_s}{\alpha_s p_2 \beta_{2s}})^{\frac{1}{\alpha_s - 1}}$. I assume that there are barriers to enter the product market in the short run, and thus firms have nonzero profits.

The supply of labor with skill s is assumed to be $L^S = w_s^{b_s}$, where the supply increases with wages

[18] A simple extension of the model is to add capital as another input. This does not affect the results of the stylized model.

and labor supply elasticity ($b_s > 0$). The supply of workers with skill s is more elastic when b_s is higher, suggesting a small increase in the wage leads to a large increase in labor supply. It takes time for workers to acquire new skills and therefore I assume that labor supply is inelastic and labor markets for different labor skills are segmented in the short run.

Labor markets are in equilibrium when the demand for labor skill s equals its supply, $L_{1s} + L_{2s} = L^S \varepsilon_s$, where ε_s is either a demand or supply $i.i.d.$ shock to the labor market. For example, technology improvements might change the demand for certain types of labor and induce labor demand shocks (Bresnahan et al. [2002]), whereas changes in the number of work permit given to foreign workers, changes in certain major's college graduation rates, changes in industry qualification requirements might induce labor supply shocks. When the labor market clears ($L^S = L^D$), wages per unit of labor skill s are endogenously determined in equilibrium:

$$w_s = [\alpha_s^{\frac{1}{\alpha_s - 1}} ((p_1\beta_{1s})^{\frac{1}{1-\alpha_s}} + (p_2\beta_{2s})^{\frac{1}{1-\alpha_s}})^{-1} \varepsilon_s]^{\frac{\alpha_s - 1}{1+(1-\alpha_s)b_s}}.$$

Profits (earnings) of firm 1 and 2 comove due to shocks to the shared labor markets. The comovement is represented by the covariance of firm 1 and 2's earnings :

$$cov(\pi_1^*, \pi_2^*) = \sum_{s \in S_1 \cap S_2} \lambda_s var(w_s^{\frac{\alpha_s}{\alpha_s - 1}}) = \sum_{s \in S_1 \cap S_2} \lambda_s var[(\gamma_s \varepsilon_s)^{\frac{\alpha_s}{1+(1-\alpha_s)b_s}}], \tag{12}$$

where $\lambda_s = (\alpha_s^2 p_1 p_2 \beta_{1s} \beta_{2s})^{\frac{1}{1-\alpha_s}} ((\alpha_s p_1)^{-1} - 1)((\alpha_s p_2)^{-1} - 1)$ and $\gamma_s = \alpha_s^{\frac{1}{\alpha_s - 1}} ((p_1\beta_{1s})^{\frac{1}{1-\alpha_s}} + (p_2\beta_{2s})^{\frac{1}{1-\alpha_s}})^{-1}$, which are both positive parameters, and $S_1 \cap S_2$ is the common labor skill sets that firm 1 and 2 uses in production. Therefore, the covariance of firm 1 and 2's earnings is the summation of the wage variance in the shared labor markets. In equilibrium, the wage variance is represented by the variance of the labor market shocks times a positive parameter.

Proposition: (Comovements and common labor sets): Earning comovements are stronger when firms share more labor skills.

Proof: When firms share more labor skills, $S_1 \cap S_2$ is larger, the summation of more positive terms leads to larger $cov(\pi_1^*, \pi_2^*)$, and thus stronger comovements.

The relationship between stock prices and the labor market is more complicated than the relationship between earnings and labor market because of hedging components in stock prices, and because labor supply and demand adjust to shocks in the long run. In the long run, a firm's stock prices are positively related to its earnings and I rely on this stylized model to infer the relation between stock returns and the labor market.

References

David Aboody and Baruch Lev. The value relevance of intangibles: the case of software capitalization. *Journal of Accounting Research*, 36:161–191, 1998.

Ana Albuquerque. Peer firms in relative performance evaluation. *Journal of Accounting and Economics*, 48 (1):69–89, 2009.

Ana M Albuquerque, Gus DeFranco, and Rodrigo S Verdi. Peer choice in ceo compensation. *Journal of Financial Economics*, 2012.

Paul Almeida and Bruce Kogut. Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917, 1999.

Nicholas Barberis, Andrei Shleifer, and Jeffrey Wurgler. Comovement. *Journal of Financial Economics*, 75 (2):283–317, 2005.

Santiago Bazdrech, Frederico Belo, and Xiaoji Lin. Labor hiring, investment and stock return predictability in the cross section. *Available at SSRN 1267462*, 2013.

Gary S Becker. Investment in human capital: A theoretical analysis. *The Journal of Political Economy*, 70 (5):9–49, 1962.

Philip G Berger. Challenges and opportunities in disclosure research, a discussion of "the financial reporting environment: Review of the recent literature". *Journal of Accounting and Economics*, 51(1):204–218, 2011.

Anne Beyer, Daniel A Cohen, Thomas Z Lys, and Beverly R Walther. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, 50(2):296–343, 2010.

Sanjeev Bhojraj and Charles Lee. Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, 40(2):407–439, 2002.

Sanjeev Bhojraj, Charles Lee, and Derek K Oler. What's my line? a comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, 41(5):745–774, 2003.

Elizabeth Blankespoor, Gregory S Miller, and Hal D White. The role of dissemination in market liquidity: Evidence from firms' use of twitter. *The Accounting Review*, 2013.

Timothy F Bresnahan, Erik Brynjolfsson, and Lorin M Hitt. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117 (1):339–376, 2002.

Philip Brown and Ray Ball. Some preliminary findings on the association between the earnings of a firm, its industry, and the economy. *Journal of Accounting Research*, pages 55–77, 1967.

Lauren Cohen and Andrea Frazzini. Economic links and predictable returns. *The Journal of Finance*, 63 (4):1977–2011, 2008.

Lauren Cohen, Andrea Frazzini, and Christopher Malloy. Sell-side school ties. *The Journal of Finance*, 65 (4):1409–1437, 2010.

Zhi Da, Joseph Engelberg, and Pengjie Gao. In search of attention. *The Journal of Finance*, 66(5):1461–1499, 2011.

Gus DeFranco, SP Kothari, and Rodrigo S Verdi. The benefits of financial statement comparability. *Journal of Accounting Research*, 49(4):895–931, 2011.

Gus DeFranco, Ole-Kristian Hope, and Stephannie Larocque. Analysts' choice of peer companies. *Available at SSRN 2046396*, 2012.

Andres Donangelo. Labor mobility: Implications for asset pricing. *forthcoming, The Journal of Finance*, 2013.

Esther Eiling. Industry-specific human capital, idiosyncratic risk, and the cross-section of expected stock returns. *The Journal of Finance*, 68(1):43–84, 2013.

Andrea L Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *The Journal of Finance*, 2013.

Joseph Engelberg, Pengjie Gao, and Christopher A Parsons. The price of a ceo's rolodex. *Review of Financial Studies*, 26(1):79–114, 2013.

Eugene Fama and Kenneth French. The cross-section of expected stock returns. *Journal of Finance*, 47: 427–465, 1992.

Eugene Fama and Kenneth French. Industry costs of equity. *Journal of Financial Economics*, 43:153–194, 1997.

Joseph PH Fan and Larry HP Lang. The measurement of relatedness: An application to corporate diversification*. *The Journal of Business*, 73(4):629–660, 2000.

George Foster. Intra-industry information transfers associated with earnings releases. *Journal of Accounting and Economics*, 3(3):201–232, 1981.

Paul M Healy and Krishna G Palepu. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1):405–440, 2001.

Gerard Hoberg and Gordon Phillips. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, 23(10):3773–3811, 2010.

Gerard Hoberg and Gordon M Phillips. Text-based network industries and endogenous product differentiation. *National Bureau of Economic Research*, 2013.

Ravi Jaganathan and Zhenyu Wang. The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, 51:3–54, 1996.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winner and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48:65–91, 1993.

Chandra Kanodia, Haresh Sapra, and Raghu Venugopalan. Should intangibles be measured: What are the economic trade-offs? *Journal of Accounting Research*, 42(1):89–120, 2004.

Benjamin F King. Market and industry factors in stock price behavior. *the Journal of Business*, 39(1): 139–190, 1966.

Edward P Lazear. Balanced skills and entrepreneurship. *American Economic Review*, pages 208–211, 2004.

Edward P Lazear. Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, 117(5):914–940, 2009.

Mark Leary and Michael Roberts. Do peer firms affect corporate financial policy? *forthcoming, The Journal of Finance*, 2013.

Charles Lee, Paul Ma, and Charles Wang. Identifying peer firms: Evidence from edgar search traffic. *Available at SSRN 2171497*, 2012.

Karen Leonard. Talent acquisition factbook 2011: Benchmarks and trends of spending, staffing and key talent metrics, 2011.

Alvin Leung, Ashish Agarwal, Prabhudev Konana, and Alok Kumar. Online search: Identifying new investment habitats. 2012.

Baruch Lev and Paul Zarowin. The boundaries of financial reporting and how to extend them. *Journal of Accounting Research*, 37(2):353–385, 1999.

Hanno Lustig, Chad Syverson, and Stijn Van Nieuwerburgh. Technological change and the growing inequality in managerial compensation. *Journal of Financial Economics*, 99(3):601–627, 2011.

David Mayers. Nonmarketable assets and the determination of capital asset prices in the absence of a riskless asset. *The Journal of Business*, 46(2):258–267, 1973.

Derek Neal. Industry-specific human capital: Evidence from displaced workers. *Journal of labor Economics*, pages 653–677, 1995.

Mitchell Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22:435–480, 2009.

Sundaresh Ramnath. Investor and analyst reactions to earnings announcements of related firms: An empirical analysis. *Journal of Accounting Research*, 40(5):1351–1376, 2002.

Joshua D Rauh and Amir Sufi. Explaining corporate capital structure: Product markets, leases, and asset similarity. *Review of Finance*, 16(1):115–155, 2012.

John Sutton. *Sunk Costs and Market Structure: price competition, advertising and the evolution of concentration*. The MIT press, 1991.

Prasanna Tambe. Big data investment, skills, and firm value. *Management Science*, 60(6):1452–1469, 2014.

Bruce A Weinberg. Long-term wage fluctuations with industry-specific human capital. *Journal of Labor Economics*, 19(1):231–264, 2001.

Table 1: Summary Statistics of LMPs

This Table provides summary statistics on the S&P 1500 firms' LinkedIn-implied labor market peer firms (LMPs). I search for the LinkedIn company homepages of S&P 1500 firms (base firms), and collect the "People Also Viewed" firms (LMPs) for each S&P 1500 firm on its LinkedIn company webpage between July 15th-19th, 2013. I also search for the LMPs' Gvkey if there is one. If the LMP is a subsidiary, I record the Gvkey of its parent firm. Panel (A) reports the coverage of S&P 1500 firms on LinkedIn. The first row reports the number of base firms with a LinkedIn company homepage; the second row reports the total number of LinkedIn homepages for firms because some firms have multiple LinkedIn company homepages. The third and fourth rows restrict the S&P 1500 sample to firms listed on either NYSE, NASDAQ, or AMEX, with a CRSP share code of 10 or 11. Panel (B) reports the composition of LMPs. Each firm has six LMPs. Column (1) shows firms with different number of LMPs that are subsidiaries or affiliates in the S&P 1500 sample. For example, 5 firms' all six LMPs are their affiliated firms, and 928 firms have zero LMPs that are affiliated with the base firm. Column (2) shows firms with different number of LMPs that are public firms. For example, 86 base firms have only one public peer firms among their LMPs. Column (3) shows firms with different number of LMPs that are foreign firms and Column (4) shows firms with different number of LMPs that are private firms.

Panel A: Coverage of S&P 1500 Firms On LinkedIn Company Homepage

	LinkedIn Representation of S&P 1500 Universe	S&P 1500 Firms	S&P 500 Firms
		(1)	(2)
(1)	Number of Firms with a LinkedIn Company Homepage	1452	492
(2)	Total Number of LinkedIn Company Homepages	1476	493
(3)	Number of Firms with a LinkedIn Company Homepage and have Common Shares Listed on NYSE, NASDAQ or AMEX	1360	460
(4)	Total Number of LinkedIn Company Homepages	1385	476

Panel B: Composition of S&P 1500 Firms' LMPs on LinkedIn

	Composition of the Six LMPs for S&P 1500 Firms			
Number of Firms	Affiliates of the Base Firm	Public Firms	Foreign Firms	Private Firms
	(1)	(2)	(3)	(4)
0	928	32	1046	633
1	261	86	260	375
2	118	161	58	217
3	42	252	16	108
4	25	277	5	34
5	6	293	0	16
6	5	284	0	2
Total	1385	1385	1385	1385

Table 2: Correspondence of LMP with Alternative Industry Classifications

This Table provides summary statistics on the degree of correspondence between LMPs and other alternative industry classification schemes by each GICS2 industry sector. The base firm is grouped by its GICS2 and the fraction of its publicly traded peer firms that have the same industry classification (for GICS2, GICS4, Fama and French 49, SIC2, SIC3, SIC4, NAICS3, NAICS6) are reported.

GICS2 Groupings	Number of firms	Same GICS2	Same GICS4	Fama and French 49	Same SIC2	Same SIC3	Same SIC4	Same NAICS3	Same NAICS6
Financial	183	0.79	0.69	0.68	0.65	0.51	0.49	0.67	0.47
Health Care	146	0.78	0.72	0.59	0.58	0.55	0.37	0.49	0.34
Consumer Discretionary	244	0.78	0.67	0.64	0.50	0.38	0.37	0.50	0.30
Energy	82	0.74	0.74	0.60	0.53	0.49	0.45	0.51	0.45
Information Technology	268	0.74	0.59	0.49	0.51	0.45	0.31	0.42	0.29
Utilities	61	0.72	0.72	0.71	0.73	0.32	0.31	0.71	0.27
Industrials	200	0.62	0.58	0.44	0.44	0.31	0.25	0.42	0.23
Consumer Staples	76	0.61	0.54	0.43	0.47	0.22	0.14	0.44	0.13
Telecommunication Services	18	0.61	0.61	0.61	0.61	0.44	0.38	0.60	0.41
Materials	95	0.57	0.57	0.35	0.43	0.24	0.21	0.42	0.18
S&P1500	1385	0.72	0.64	0.55	0.53	0.41	0.34	0.50	0.31
S&P500	474	0.69	0.61	0.53	0.50	0.39	0.33	0.48	0.31

Table 3: Summary Statistics of the Skill Similarity Measure

This Table provides summary statistics of the average skill similarity scores between the base firm and its LMPs by each GICS2 industry sector. Skill similarity $s_{i,j}$ is defined in equation (1). $s_{i,j}$ is measured as the fraction of common skills shared by firm i and firm j, divided by the number of total skills of firm i and firm j. Higher skill similarity implies stronger connections in the labor market. Column (1) reports the number of firms in each GICS2 sector; Column (2) reports the average skill similarity between the base firm and its LMPs; Column (3) reports the average skill similarity between the base firm and any firm that shares at least one common skill with the base firm in my sample. Column (4) reports the difference between column (2) and (3), and *** indicates significant at the 1% level.

GICS2 Groupings	Number of Firms	LMP (Mean)	Non-LMP (Mean)	(2)-(3)
	(1)	(2)	(3)	(4)
Energy	82	0.18	0.09	0.097***
Utilities	61	0.15	0.10	0.052***
Financial	183	0.14	0.10	0.044***
Consumer Discretionary	244	0.11	0.06	0.045***
Health Care	146	0.10	0.05	0.046***
Materials	95	0.10	0.04	0.049***
Consumer Staples	76	0.10	0.065	0.042***
Telecommunication Services	18	0.10	0.06	0.036***
Industrials	200	0.08	0.04	0.034***
Information Technology	268	0.06	0.04	0.020***
S&P1500	1384	0.11	0.06	0.045***
S&P500	466	0.12	0.07	0.055***

Table 4: Return Comovements Tests based on LMPs vs. Alternative Industry Groupings

This Table reports the average of 120 monthly cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings. The dependent variables are base firms' stock-level monthly returns. In columns (1) to (5), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs, 4-digit SIC peer firms, 6-digit NAICS peers, 4-digit GICS peers, or TNIC (Hoberg and Phillips [2013]) peers, respectively. Column (6) and (7) include both returns from peer firms from all alternative industry groupings, and column (7) include returns from LMPs. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012.

VARIABLES	(1) 2003-2012 ewret	(2) 2003-2012 ewret	(3) 2003-2012 ewret	(4) 2003-2012 ewret	(5) 2003-2012 ewret	(6) 2003-2012 ewret	(7) 2003-2012 ewret
LMP	0.507***						0.290***
	(45.41)						(29.14)
SIC4		0.455***				0.135***	0.0900***
		(32.00)				(10.24)	(7.162)
GICS4			0.916***			0.465***	0.369***
			(43.14)			(20.22)	(18.14)
NAICS6				0.390***		0.0880***	0.0623***
				(27.24)		(7.672)	(5.835)
TNIC					0.574***	0.272***	0.177***
					(39.12)	(23.50)	(17.08)
Observations	73,568	73,568	73,568	73,568	73,568	73,568	73,568
R-squared	0.107	0.079	0.093	0.069	0.091	0.130	0.156
Number of groups	120	120	120	120	120	120	120

Table 5: Return Comovements Tests Based on LMPs v.s. Random and Best Alternative Industry Groupings

This Table reports the average of 120 monthly cross-sectional regressions of return comovement tests based on LMPs alternative industry groupings. The dependent variables are the base firm's stock-level monthly returns. In columns (1), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs. The LinkedIn implied LMPs show at most six firms, among which x firms are publicly traded with stock return data available. I create a comparable GICS and SIC portfolio with the same number of peers. In column (2), the dependent variable is contemporaneous equal-weighted average returns of the "Random" 4-digit GICS peer firms. For each base firm, I randomly select x peer firms with the same GICS code as the base firm and compute the average stock returns. In column (4), the dependent variable is contemporaneous equal-weighted average returns of the "Best" 4-digit SIC peer firms. For each base firm, a pooled time-series cross-sectional regression is run with the monthly return of each firm in the same 4-digit SIC code as the independent variable, and x firms with the highest R-squared are selected as the "best" 4-digit SIC peer firms. Column (3) and (5) includes both returns from LMPs and peer firms from "Random" 4-digit GICS and "Best" 4-digit SIC respectively. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012.

VARIABLES	(1) 2003-2012 ewret	(2) 2003-2012 ewret	(3) 2003-2012 ewret	(4) 2003-2012 ewret	(5) 2003-2012 ewret
LMP	0.513***		0.489***		0.313***
	(43.87)		(45.78)		(30.26)
Random GICS		0.168***	0.0871***		
		(14.92)	(12.06)		
Best SIC4				0.485***	0.370***
				(45.88)	(36.70)
Observations	65,687	65,687	65,687	65,687	65,687
R-squared	0.107	0.023	0.114	0.142	0.174
Number of groups	120	120	120	120	120

Table 6: Accounting Ratios Comovements Tests Based on LMPs v.s. Alternative Industry Groupings

This Table reports the average of 10 annual cross-sectional regressions of accounting-based performance metrics comovement tests based on LMPs and alternative industry groupings. The dependent variables are base firm's annual accounting-based performance metrics. Time-series average of annual cross-sectional R-squared are reported in columns (2) to (3). In column (2), the independent variable is contemporaneous average performance metrics of 4-digit SIC peer firms and 4-digit GICS peer firms, $Ratio_{i,t} = \alpha_t + \beta_{1,t} Ratio_{GICS,t} + \beta_{2,t} Ratio_{SIC,t} + \varepsilon_{i,t}$. In column (3), average performance metrics from LMPs are add to the estimation specification in column (3), $Ratio_{i,t} = \alpha_t + \beta_{LMP,t} Ratio_{LMP,t} + \beta_{1,t} Ratio_{GICS,t} + \beta_{2,t} Ratio_{SIC,t} + \varepsilon_{i,t}$. In column (4), Time-series average of yearly cross-sectional regression coefficient estimates $\beta_{LMP,t}$ are reported, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012. Observations with missing total assets, long term debt, net income before extraordinary items, or operating income after depreciation, or with negative common or total equity, share price smaller than $3 at the end of the fiscal year, net sales smaller than $100 million are dropped. The calculation and definition of accounting-based performance metrics are detailed in the Appendix Table **4.**

	Num. Panel Obs	GICS+SIC	GICS+SIC +LMP	LMP
		R-squared	R-squared	Coefficient
	(1)	(2)	(3)	(4)
Expense Ratios				
rdpersales	6,324	0.677	0.757	0.627***
sgapersales	5,321	0.556	0.623	0.565***
rdsgapersales	5,321	0.597	0.676	0.623***
Employment Ratios				
empgrowth	5,333	0.041	0.057	0.180***
Valuation Multiples				
pb	5,940	0.081	0.111	0.284***
evs	5,940	0.358	0.430	0.463***
pe	5,940	0.007	0.012	0.0872**
Profitability Ratios				
roa	6,324	0.103	0.1659	0.430***
roe	6,324	0.032	0.052	0.211***
leverage	6,324	0.064	0.075	0.129***
salesgrowth	5,537	0.151	0.187	0.315***

Table 7: Return Comovements Tests Based on LMPs v.s. Alternative Industry Groupings By Skill Similarity

This table reports the average of 120 monthly cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings by skill similarity. The sample is sorted into terciles based on the average skill similarity between the base firm and its LMPs. Skill similarity $s_{i,j}$ is defined in equation (1). $s_{i,j}$ is measured as the fraction of common skills shared by firm i and firm j, divided by the number of total skills of firm i and firm j. Higher skill similarity implies stronger connections in the labor market. The dependent variables are base firm's stock-level monthly returns. In columns (1) to (3), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs. In columns (4) to (6), the independent variable is equally-weighted contemporaneous average return of 4-digit GICS peer firms. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012. Base firms in column (1) and (4) share the fewest skills with their LMPs, and base firms in column (3) and (6) share the most skills with their LMPs.

	(1)	(2)	(3)	(4)	(5)	(6)
Sample Period	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012
Skill Quantile	Low	Medium	High	Low	Medium	High
LMP	0.306***	0.474***	0.658***			
	(19.23)	(32.49)	(49.41)			
GICS4				0.717***	0.888***	1.015***
				(19.76)	(33.40)	(35.89)
Total Num. of						
panel observations	21,122	25,936	28,965	21,122	25,936	28,965
R-squared	0.039	0.100	0.197	0.041	0.096	0.139
Number of months	120	120	120	120	120	120

Table 8: Return Comovements Tests Based on LMPs v.s. Alternative Industry Groupings By R&D Intensity

This table reports the average of 120 monthly cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings by R&D intensity. The sample is sorted into terciles based on the average R&D intensity of the base firm every year. Firms have zero R&D expense are in the "None" group. Firms with positive R&D expense are sorted into high R&D and low R&D groups, respectively. The dependent variables are base firm's stock-level monthly returns. In columns (1) to (3), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs. In columns (4) to (6), the independent variable is equally-weighted contemporaneous average return of 4-digit GICS peer firms. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012. Base firms in column(1) and (4) report zero R&D expense, base firms in column (2) and (5) have relatively low R&D expense, and base firms in column (3) and (6) have relatively high R&D expense.

	(1)	(2)	(3)	(4)	(5)	(6)
Sample Period	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012
R&D Quantile	Zero	Low	High	Zero	Low	High
LMP	0.540***	0.431***	0.317***			
	(43.19)	(18.60)	(15.28)			
GICS4				0.948***	0.758***	0.721***
				(33.68)	(16.40)	(12.11)
Total Num. of panel observations	50,686	11,740	19,784	50,686	11,740	19,784
R-squared	0.142	0.087	0.040	0.114	0.066	0.044
Number of months	120	120	120	120	120	120

Appendix Table 1: Return Comovement Tests Based on LMPs v.s. Alternative Industry Groupings with Out of Sample Tests

This table reports the average of several monthly cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings with different time periods. The sample period is from 2003 to 2007 in columns (1), (4), (5); from 2008 to 2010 in columns (2), (6), (7); from 2011 to 2012 in column (3), (8), (9). The dependent variables are base firm's stock-level monthly returns. In columns (1) to (4), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs. In columns (5) to (9), the independent variables include returns from 4-digit SIC peer firms, 4-digit GICS peer firms, or TNIC (Hoberg and Phillips [2013]) peer firms as controls. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, from 2003 to 2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2003-2007	2008-2011	2012	2003-2007	2003-2007	2008-2011	2008-2011	2012	2012
VARIABLES	ewret	ewret	ewret	ewret	ewret	ewret	ewret	ewret	ewret
LMP	0.505***	0.525***	0.487***		0.272***		0.303***		0.315***
	(31.72)	(23.74)	(22.93)		(19.50)		(16.06)		(15.23)
GICS4				0.536***	0.431***	0.399***	0.320***	0.386***	0.286***
				(16.16)	(14.30)	(10.09)	(9.630)	(8.759)	(7.471)
SIC4				0.136***	0.0915***	0.151***	0.104***	0.109***	0.0652**
				(7.279)	(5.313)	(5.722)	(3.985)	(4.343)	(2.690)
NAICS6				0.0551***	0.0416***	0.114***	0.0771***	0.131***	0.0918***
				(3.714)	(2.922)	(5.474)	(3.798)	(4.763)	(3.720)
TNIC				0.267***	0.174***	0.279***	0.181***	0.272***	0.179***
				(16.31)	(12.12)	(12.26)	(8.771)	(11.64)	(8.138)
Observations	34,896	23,026	15,646	34,896	34,896	23,026	23,026	15,646	15,646
R-squared	0.101	0.125	0.096	0.130	0.150	0.148	0.178	0.105	0.136
Number of groups	60	36	24	60	60	36	36	24	24

Appendix Table 2: Return Comovements Tests Based on LMPs v.s. Alternative Industry Groupings, with Firm Characteristic Controls

This Table reports the average of 120 monthly cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings. The dependent variables are the base firm's stock-level monthly returns. In all specifications, log of book to market $Log(B/M)_{i,t}$, log of market cap $Log(Size)_{i,t}$, and the base firm's own past 12 months returns, skipping the most recent one month, $Momentumn_{i,t}$ are included as additional controls. In columns (1) to (5), the independent variable is equally-weighted contemporaneous average return of LinkedIn-implied LMPs, 4-digit SIC peer firms, 6-digit NAICS peers, 4-digit GICS peers, or TNIC (Hoberg and Phillips [2013]) peers, respectively. Column (7) includes both returns from LMP and peer firms from all alternative industry groupings. Time-series average of monthly cross-sectional regression coefficients, and R-squared are reported. T-statistics of coefficient estimates are reported in parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes S&P 1500 firms that have LinkedIn company homepages and have at least three publicly traded LMPs firms, between 2003 and 2012.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012	2003-2012
VARIABLES	ewret	ewret	ewret	ewret	ewret	ewret	ewret
LMP	0.480***						0.265***
	(42.62)						(27.23)
SIC4		0.437***				0.135***	0.0936***
		(31.48)				(10.44)	(7.612)
GICS4			0.896***			0.476***	0.386***
			(41.62)			(20.58)	(18.40)
NAICS6				0.371***		0.0767***	0.0546***
				(26.78)		(6.956)	(5.265)
TNIC					0.550***	0.252***	0.167***
					(37.27)	(21.54)	(15.51)
Characteristic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	71,976	71,976	71,976	71,976	71,976	71,976	71,976
R-squared	0.133	0.111	0.127	0.101	0.121	0.160	0.181
Number of groups	120	120	120	120	120	120	120

Appendix Table 3: Return Comovement Tests based on LMP vs. Alternative Industry Groupings, Value Weighted Average Returns and Different Number of Publicly Trade LMPs

This table reports pooled time-series cross-sectional regressions of return comovement tests based on LMPs and alternative industry groupings. In all specification, year fixed effects are included. The dependent variables are base firm's stock-level monthly returns. In columns (1), (3) and (5), the independent variables include contemporaneous value-weighted average returns of 4-digit SIC peer firms, 4-digit GICS peer firms, and TNIC peer firms. In columns (2), (4) and (6), the independent variables include both contemporaneous value-weighted average returns of LinkedIn-implied LMP firms, and 4-digit SIC peer firms, 4-digit GICS peer firms, as well as TNIC peer firms. Specifications in columns (1) to (6) include market excess return as an additional control. The sample period in columns (1) and (2) is from 2003 to 2012, in columns (3) and (4) is from 2003 to 2010, and in columns (4) and (5) is from 2011 to 2012. The sample in columns (1) and (2) require at least three publicly trade LMP firms, while the rest requires at least four publicly traded LMP firms. T-statistics of coefficient estimates are calculated based on the standard errors clustered at firm level, and they are reported in the parenthesis, where *, **, and *** denote significance at 10, 5, and 1% significance level, respectively. The sample includes all S&P1500 firms that have LinkedIn profiles and have at least three publicly traded LMPs.

Appendix Table 3: Return Comovement Tests based on LMP vs. Alternative Industry Groupings, Value Weighted Average Returns and Different Number of Publicly Trade LMPs (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	2003-2012	2003-2012	2003-2010	2003-2010	2011-2012	2011-2012
VARIABLES	vwret	vwret	vwret	vwret	vwret	vwret
VW-LMP		0.182***		0.203***		0.178***
		(11.63)		(7.998)		(6.321)
VW-TNIC	0.244***	0.180***	0.253***	0.181***	0.231***	0.183***
	(10.49)	(8.410)	(7.413)	(5.781)	(6.270)	(4.868)
VW-SIC4	0.191***	0.151***	0.179***	0.132***	0.251***	0.201***
	(8.367)	(7.037)	(5.537)	(4.237)	(7.761)	(6.482)
VW-GICS4	0.387***	0.316***	0.397***	0.319***	0.285***	0.207***
	(15.12)	(12.92)	(11.08)	(8.981)	(7.161)	(5.230)
VW-NAICS6	0.182***	0.177***	0.232***	0.224***	0.164***	0.159***
	(9.737)	(9.844)	(8.415)	(8.355)	(6.330)	(6.324)
Market	0.140***	0.135***	0.0506*	0.0549*	0.266***	0.272***
	(6.027)	(5.956)	(1.656)	(1.804)	(6.029)	(6.153)
Total Num. of panel observations	73,566	73,566	35,864	35,864	10,242	10,242
Adjusted R-squared	0.335	0.341	0.346	0.351	0.369	0.374
Numb. Public Peers	At Least 3	At Least 3	At Least 4	At Least 4	At Least 4	At Least 4
Cluster	Firm	Firm	Firm	Firm	Firm	Firm
Fixed Effect	Year	Year	Year	Year	Year	Year

Appendix Table 4: Definitions of Accounting-Based Performance Variables

This table provides detailed definitions for the accounting-based performance variables used in the paper, including the corresponding Compustat item numbers.

Variable Names	Variable Definitions
Expense Ratios	
rdpersales	R&D expense (xrd)/net sales (sale)
sgapersales	SG&A expense (xsga)/net sales (sale)
rdsgapersales	(R&D expense (xrd)+ SG&A expense (xsga))/net sales (sale)
Employment Ratios	
empgrowth	(one year ahead number of employees (emp_{t+1})-emp_t)/emp_t
Valuation Multiples	
pb	market cap/total common equity (ceq)
evs	(market cap+long-term debt(dltt))/net sales(sale)
pe	market cap/net income before extraordinary items (ib)
Profitability Ratios	
roa	net operating income after depreciation (oiadp)/total assets (at)
roe	net income before extraordinary items(ib)/total common equity (ceq)
leverage	long term debt (dltt)/total stockholder's equity (seq)
salesgrowth	(one year ahead realized sales ($sale_{t+1}$)-$sale_t$)/$sale_t$

Appendix Table 5: LinkedIn-Implied Labor Market Skill Set and Skill Similarity

This table provides the proxy for Google's and Facebook's labor market skill sets in Panel A and B respectively. For each firm, I include the top five skills & expertise listed on its LinkedIn company homepage and the related skills for each of the top five skills. Google has 85 skills and Facebook has 63 skills. They have 37 skills in common, so their skill similarity score is $s_{Google,Facebook} = \frac{37}{85+63-37} = 0.33$.

Panel A: Google's LinkedIn-Implied Labor Market Skill Set

Top Skills & Expertise	Related Skills
Google Adwords	MSN AdCenter, Yahoo Search Marketing, Search Advertising Organic Search, Adsense, Paid Search Strategy, Google Merchant Center Google Ad Planner, Conversion Optimization, Keyword Research, Paid Search Campaigns, Marin Software, Google Website Optimizer, Kenshoo, Google Adwords Professional, Adgooroo Landing Page Optimization, Google Webmaster Tools, Search Analysis,
Python	NumPy, Django, SciPy, SQLAlchemy, PyQt, Matplotlib, wxPython, Celery, NLTK, WSGI, Flask, CherryPy, Web2py, TurboGears, Pygame, Pylons, PyGTK, SCons, Zope, PyUnit
Machine Learning	Feature Selection, Semi-supervised Learning, Classifiers, Dimensionality Reduction, Graphical Models, Reinforcement Learning, Unsupervised Learning, Text Classification, Pattern Recognition, Recommender Systems, Natural Language Processing, Text Mining, Object Detection, Collaborative Filtering, SVM, Statistical Machine Translation, Mahout, Bayesian networks, NLTK, Natural Language
MapReduce	Oozie, Sqoop, Flume, BigTable, Mahout, HBase, Amazon Elastic MapReduce, Cascading, Hive, Avro, Voldemort, Katta, Cascalog, Collaborative Filtering, Apache Pig, CDH, Relevance, Nutch, Stream processing, Recommender Systems
Search Advertising	Yahoo Search Marketing, Search Syndication, Account Optimization, Paid Search Strategy, MSN AdCenter, Adgooroo, Kenshoo, Marin Software, Text Ads, Keyword Advertising, Paid Search Campaigns, DART Search, Google Adwords Professional, Conversion Tracking, Mobile Search, SearchCenter, Local Search, Keyword Generation, Google Website Optimizer, Landing Page Optimization

Appendix Table 5: LinkedIn-Implied Labor Market Skill Set and Skill Similarity (Continued)

Panel B: Facebook's LinkedIn-Implied Labor Market Skill Set	
Top Skills & Expertise	Related Skills
Hive	Sqoop, Oozie, Flume, Amazon Elastic MapReduce, Mahout, HBase, Avro, Cascading, CDH, Apache Pig, Cascalog, MapReduce, Voldemort, Google Ad Planner, Conversion Optimization, Keyword Research, Paid Search Campaigns, Marin Software, Google Website Optimizer, Kenshoo, Google Adwords Professional, Adgooroo Landing Page Optimization, Google Webmaster Tools, Search Analysis,
Facebook API	LinkedIn API, YouTube API, FQL, Google API, Paypal Integration, OAuth, OpenSocial, Custom Facebook Pages, Social Engine, eBay API, Google Checkout, OpenSceneGraph, Papervision3D, Twilio, Away3D, Authorize.net, Kohana, Chrome Extensions, Social Graph, Socket.io
Machine Learning	Same as those of Google
MapReduce	Same as those of Google
Hadoop	HBase, Oozie, Sqoop, Flume, Mahout, MapReduce, Hive, Cascading, Amazon Elastic MapReduce, Cascalog, Nutch, Voldemort, Apache Pig, Katta, Avro, Cassandra, Greenplum, Vertica, Collaborative Filtering, CDH