Senior Thesis: Mining LinkedIn Profiles to Determine Predictors of Labor Market Success

Samantha Dalal

December 1st 2019

Abstract

This thesis outlines how we can use publicly available LinkedIn profiles to generate insights on the determinants of labor market success for recently graduated MBA students. I use profile aspects, broken up into skill and work experience categories, from LinkedIn profiles of recently graduated MBA students to show that work experiences are far more effective than skills at predicting labor market success. I define labor market success in terms of salary. Among graduates who obtained a job within three months of graduation, work experience is a better predictor of labor market success with an \mathbb{R}^2 that is about five times larger than the skills model. Additionally, I find that combining both work experience and skills regressors to predict salary outcomes, results in only the maximum duration (a continuous variable measuring the maximum time in months that someone spent at a position) and same company (an indicator variable that took value 1 if the person was employed post graduation at a company where they had previously worked, and 0 if not) variables are significant, indicating that skills presented on LinkedIn profiles are of negligible importance for determining labor market outcomes. Finally, I find that everyone who attained a new job, in my sample, got that job at a company they previously worked at. LinkedIn emphasizes its skills section, encouraging users to fill it out and obtain endorsements from others in

their network, however this feature seems to have little bearing on realized labor market outcomes, inviting further exploration of what purpose the skills feature is intended to serve and who actually benefits from utilization of the skills section.

Introduction

Many young professionals use online job boards and professional social platforms as their primary environment to grow their networks, advertise themselves to employers, and conduct job searches (Peterson & Dover 2014). The most common professional social platform, LinkedIn, allows users to create a profile which showcases their work experiences, their skills, recommendations from peers, as well as other various attributes. In creating a "strong" LinkedIn profile, users must decide how to best showcase themselves using the work experiences section (structured similarly to a written resume) and the skills section (i.e language skills, programming skills, conflict resolution, etc.). LinkedIn strongly encourages its users to fill out the skills section and get endorsements (Fisher 2016), emphasizing the importance of this section in attracting the attention of recruiters and therefore making users more attractive job candidates. However, I find that amongst students who complete the same advanced degree, namely an MBA, skills and endorsements are poor predictors of labor market success and insignificant factors in determining salary. On the contrary, work experiences are a far better predictor of labor market success and the only significant determinant of salary.

According to LinkedIn's mission statement, their goal is to "Create economic opportunity for every member of the global workforce", presumably through moving the job market online and therefore making it easier for people to get access to information pertaining to jobs and to make it easier for employers to find suitable candidates. While LinkedIn's social network structure is most likely to benefit young people entering the workforce, the question remains of just how much does a having a LinkedIn profile increase one's chances of labor market success. (Skeels, 2009) For students, one of the most pressing questions they face is: What will they do upon graduation, and how do they plan on getting there? For these students, information about what has made people successful in the labor market in the past is extremely valuable as it can provide guidance as to what skills they should acquire, where they should intern, and what career paths could be attainable for them. As much of the job market has moved online, massive amounts of data about people's work experiences, skills,

and career outcomes have become available through platforms such as LinkedIn. Utilizing this massive amount of information to sketch out possible career paths based on previous work experiences and skills could potentially be insightful for students who feel lost and unsure of how to achieve their career goals. This paper aims to elucidate how specific aspects of a LinkedIn profile can be used to predict labor market outcomes, specifically salary, but does not draw a causal relationship between profile aspects and labor market outcomes.

Acknowledging that social media profiles are sometimes not the most honest portrayal of a person's characteristics, I attempt to counterbalance this by using profile aspects that are supported by endorsements from other professionals - i.e only counting skills that have endorsements.(Humphrey, 2017) Additionally, I chose only to look at people's profiles if their work experiences had some sort of detailed description attached to the title.

I will address the following questions in this paper: Are skills or work experiences better predictors of labor market success? What are strong predictors of labor market success that we can gather from LinkedIn profiles? I use data that is scraped from the LinkedIn profiles of 90 MBA students who graduated in the Spring of 2019 from top MBA programs in the US. The dataset consists of information on all employers, dates of employment, industry of employment, skills, and endorsements for each student. I use this data to determine if skills or work experiences are better predictors of labor market success and to generate insights as to what factors are tied to labor market success.

Ultimately, I find that work experiences are a far better predictor of labor market success, and every single person, in my sample, who obtained a job within three months of graduating got that job at a company they previously worked at.

Literature Review

While there have been a multitude of studies done on the effect of LinkedIn as a disruptor in the labor market, very little work has been done studying how publicly available information on LinkedIn can be gathered and utilized to better inform students of possible career paths. Most relevant to this paper are previous studies on the effect of skills on labor market success and the relationship between early career employment and adult labor market outcomes. Additionally, work on the reliability of online information is relevant to this paper, as assigning importance to potentially misleading elements of labor market success could bias results.

Preliminary analysis of the relationship between graduate students skills and labor market outcomes by Kamsuriah Ahmad, Sufian Idris, and Noor Zainal (2012)4 in their paper shows that the majority of graduates who were employed directly out of graduation possessed three of the four "employability skills" as defined by the paper. This paper defines employability skills: programming, system development, soft skills, and entrepreneur skills, as the students they studied all came from an information technology program. While it is interesting that the students' skills were strongly correlated with labor market success, it should be noted that the students self-evaluated their skills. Thus this paper really measured the relationship between graduates' perceived skills and labor market outcomes. In my paper I try to eliminate this factor of subjectivity by using the skills listed on the student's Linkedin profiles that have at least 2 or more endorsements.

In the paper "Order From Chaos? The Effects of Early Labor Market Experiences on Adult Labor Market Outcomes", Rosella Gardecki and David Neumark 5 study how early job market stability can lead to better labor market outcomes for young professionals. They conclude that adult labor market outcomes (defined as adults being in their late 20s to early/mid thirties) are not strongly tied to early labor market experiences. This paper implies that encouraging young professionals to be more strongly tied to "employers, industries, or occupations, at a younger age" does not have strong lasting effects on overall labor market outcomes, perhaps indicating that early career work experiences are not a good predictor of labor market success. However, this paper does not juxtapose experience versus skills of young professionals in their adult labor market outcomes; leaving room to explore whether

skills could serve as a better indicator of job market success than work experiences. This paper proposed a long term analysis of subjects over the span of several years, whereas my paper will only check over the span of a few months if business school graduates had attained full time positions post graduation or not. Perhaps contrary to this paper's findings, I found a strong correlation between early employment and later labor market success, as all of the subjects in my sample who obtained a job got that job at a company they had previously worked at. However, the Gardecki and Neumark paper did not explicitly address the benefits (or costs) of returning to a previous place of work.

Finally, in order to evaluate the quality of online recommendations, I refer to Lorenzo Cantoni and Chrysi Rapanta's paper, "The LinkedIn Endorsement Game: Why and How Professionals Attribute Skills to Others". Catoni and Rapanta explore how epistemic authority, source trustworthiness, and social identity must be taken into account when evaluating online endorsements as a signal for skills actually possessed by the user. They discover that a large majority of professionals who endorse others on LinkedIn do not carefully consider the weight of an endorsement and do so as an act of creating or maintaining a personal connection. This has direct implications for my paper as it highlights the fact that skills listed on user's LinkedIn may be poor representations of skills actually possessed by that user that could help them get a job. LinkedIn has been making an effort to combat the ambiguity associated with skills by offering "assessments", 10-15 minute quizzes that evaluate whether or not a person actually is competent in a specific skill and then broadcasting this information to recruiters if they pass the assessment. (Linkedin, 2019). However, as a free user I was unable to see if the skills from the that profiles I gathered had the LinkedIn certification.

Economic Theory and Background Research

While online platforms, such as LinkedIn, have played a huge role in restructuring the labor market to make jobs more accessible to people through increasing exposure to opportunities, they have also led to major congestion in the job market. Economist Alvin Roth addresses the issues that face labor markets and how their construction affects outcomes for participants in his paper "Marketplaces, Markets, and Market Design".

Roth would categorize what LinkedIn is doing as increasing the "thickness" of the market by bringing together many job seekers and employers in the same space and time. Increasing thickness of markets leads to congestion and therefore, inefficiencies. In order for job candidates to combat marketplace congestion, Roth hypothesizes that there must be an effective signaling mechanism through which employers can tell candidates what they're looking for and candidates can broadcast their qualifications and desire to work to employers. LinkedIn's skills feature is intended to be a signaling mechanism in both directions, as employers can post desired skills on their job opening posts, and candidates can broadcast their skills to recruiters and see how their skills measure up to the desired skills on the job opening post. Additionally, LinkedIn for recruiters allows them to search through massive amounts of candidates using skills filters, further reinforcing the idea that skills serve as a signaling mechanism intended to combat inefficiencies caused by congestion. However, if this were the case - that skills serve as an important signal for recruiters, then they should have some correlation with labor market outcomes. However, in my study I find this not to be the case.

Turning to another one of Roth's theories, "unraveling", I find a potential explanation for the insignificance of skills in predicting labor market outcomes. Roth states that "Thick marketplaces that operate at a time at which transactions can be made efficiently provide a public good to the participants, by allowing them to compare many possible transactions.". (Roth, 2018) However, labor markets can experience an unraveling in time, in which the market becomes "earlier, shorter in duration, and diffuse in time", leading to inefficiencies

as participants no longer have access to full information to make decisions. Similar to the example cited on hiring at competitive law firms, a similar distorted hiring timeline occurs at top consulting firms. In this process full time offers or promotion offers are made at an earlier point in the student's education, typically during their third year of undergraduate studies, and the hiring process is then complete. By locking down candidates earlier, firms benefit from knowing that the candidate will already be familiar with their internal structure, they are a guaranteed hire, and they are a quality hire - as the internship serves as an extended interview process. The theory of unraveled time is particularly relevant to MBA graduates, as it is common practice at consulting/financial firms to either pay for an MBA, provided that the employee returns to the company upon completion of the program, or guarantee a raise contingent on the completion of an MBA and return to the company. This theory supports my findings, in that every person who obtained a job after completion of an MBA did so at a company they had previously worked at. Additionally returning to a company had a significant effect on an increase in salary. Thus it would seem that the unraveling of time in the consulting/financial labor market supersedes the use of skills to signal to goodness of fit for candidates as candidates are often not looking for a new job after completing an MBA as they are already guaranteed a job at a previous firm.

LinkedIn adds to congestion in attempt to relieve friction in the labor market by making the market thicker. It is therefore warranted to explore LinkedIn's efficiency as a marketplace for employment, and how its features, specifically the skills feature, contribute to this efficiency. LinkedIn attempts to cut through congestion by providing clear signaling mechanisms with its skills features, however this mechanism may not be effective at combating the inefficiencies caused by unraveling in time within industry specific markets.

Hypotheses

Work experience will be a more effective predictor of labor market success than skills, because many of the students from my sample will most likely have very similar skills as they all come from the same program, however their previous work experience is likely to vary widely - thus setting them apart from other prospective employees.

Data

All of my data was scraped from LinkedIn using a webscraper. I first chose from a list of top MBA programs in the US to eliminate possible variability between quality of candidates, restricting my sample to only students who had earned MBA degrees from either: UC Berkeley, HAAS, UCLA Anderson School of Business, University of Pennsylvania Wharton School of Business, and Northwestern University - Kellogg School of Management. I then opened up a premium LinkedIn account so that I could see unlimited profiles, filtered by graduate school, and graduation term being in Spring 2019 and randomly selected profiles that came up in my search results by enumerating them and using a random number generator to pick the ones I would look at. Once I chose to look at someone's profiles I verified that each of their work experiences had some kind of description attached to it to avoid possible "fake" employment records.

In order to determine if someone met the criteria for having a top 10 skill, I ensured that they had to have at least five endorsements for each skill to avoid the problem of people just listing as many skills as they could to look better for recruiters.

In order to estimate salaries for my sample, I looked on Glassdoor and Indeed for the job titles and companies and utilized this information.

Empirical Strategy

The goals of this project were to: (1) determine if skills or work experience are a better predictor of labor market success, and (2) to create a sufficient model for predicting labor market success based on a combination of skills and work experience regressors.

The independent variables include consist of the following:

- 1. Ten or more Endorsements
 - (a) Whether or not they had at least 10 endorsements among the top 10 skills
- 2. At_Least_2_of_Top_10
 - (a) Whether or not they at least 2 of the top 10 skills, where the top 10 skills were the 10 skills that were the most common within my sample
- 3. At Least 1 Slevel
 - (a) Whether or not they had at least one prior senior level job
- 4. At Least 7 WE
 - (a) Whether or not they had at least 7 prior work experiences
- 5. At Least 30 months MaxDur
 - (a) Whether or not their maximum duration of employment was at least 30 consecutive months
- 6. SameCompany
 - (a) Whether or not their post-MBA employment was at the same company as before
- 7. MaxDur
 - (a) The length of the maximum duration of consecutive months spent at one position

8. TotalWe

(a) The total number of work experiences

The response (dependent) variables consist of the following:

1. Estimated Salary

(a) The estimated salary based on title and company, from Glassdoor

2. Recoded NewJob

(a) An ordinal variable where 0 corresponds to no job, 1 corresponds to obtaining a job, and 2 corresponds to obtaining a senior level job

The summary indicator variables were constructed to reduce the total number of regressors by capturing all relevant information pertaining to skills and work experience into those five variables. I constructed the skill summary indicator variables by first determining what the top ten skills were (the skills that showed up the most frequently within my sample) in my sample and then determining for each person if they had those skills, and if so did they have at least ten endorsements across all of those skills. The work experience summary indicator variables were constructed by first determining that the average number of work experience for MBA graduates is ranges from five to seven, and then comparing each person with this average. Then I looked at the average length of employment for MBA graduates, which was around 30 months and compared each person to this average. The thinking was that longer stints of employment conveyed information about the quality of worker and depth of experience generated in that role. Finally, it seemed as though having a senior level job would probably increase your chances of finding a job after completing an MBA program, so I included this variable as several people in my sample had previous senior level job titles. Senior level job titles are defined as being: senior manager, president, vice president, or director.

I included Same Company because I knew that many MBA students return to their prior places of work after finishing a degree as their employers typically pay for their degree. I included maximum duration as working in a job for a longer period of time can show that the person is gaining valuable skills and knowledge related to that one type of work, making them stronger candidates for similar jobs post graduation. Total number of work experiences are included because having high friction in the job market could either indicate a poor candidate as they cannot "hold down a job", or a well rounded candidate that is flexible enough to succeed in many lines of work and therefore will be able to get a high paying job after completing an MBA degree.

I chose both estimated salary and type of labor market success as response variables as initially I thought that I would do an ordinal logistic regression model, and therefore I needed an ordinal response variable. However, due to my small sample size ordinal logistic regression was a poor choice and I ended up doing an OLS regression with a continuous response variable, estimated salary.

I began my analysis by running a two separate OLS regressions in order to compare skills to work experience as predictors of labor market success, in this case salary. I compared the R^2 values for my skills and my work experiences regressions to determine which set of predictors was better for determining labor market success.

Then I regressed estimated salary on all predictors to determine among both skills and work experience, which are significant predictors of labor market success and to what magnitude do they affect salary post-graduation.

The first set of regressions I run is the following:

$$Estimated_Salary = \beta_0 + \beta_1 * Ten_or_more_Endorsements + \beta_2 * At_Least_2_of_Top_10 + \epsilon$$

$$\tag{1}$$

and

$$Estimated_Salary = \beta_0 + \beta_1 * At_Least_1_Slevel + \beta_2 * At_Least_7_WE + \beta_3 * At_Least_30_months_MaxDur + \epsilon$$
 (2)

These regressions are intended to compare how good of predictors of salary skills and work experiences are. I determine the quality of predictor by looking at the R^2 value for each regression.

Seeing as neither skills or work experiences alone would be sufficient to explain labor market outcomes, I then create a new regression, combining all of my summary indicator variables as well as my other indicator and continuous predictors. Equation (3) is intended to show what aspects of a LinkedIn profile are significantly linked to labor market success and by what magnitude do they affect salary outcomes.

$$Estimated_Salary = \beta_0 + \beta_1 * Ten_or_more_Endorsements + \beta_2 \\ * At_Least_2_of_Top_10 + \beta_3 * At_Least_1_Slevel \\ + \beta_4 * At_Least_7_WE + \beta_5 * At_Least_30_months_MaxDur \\ + \beta_6 * SameCompany + \beta_7 * MaxDur + \beta_8 * TotalWe + \epsilon \end{aligned}$$

$$(3)$$

Seeing that only SameCompany and MaxDur were the significant predictors from Equation (3), I create a separate regression with only these two predictors and estimated salary as the response to verify their unique significance, as seen below in Equation (4)

$$Estimated_Salary = \beta_0 + \beta_1 * SameCompany + \beta_2 * MaxDur + \epsilon$$
 (4)

Results

From equations (1) and (2), I find that the R^2 for the work experience regression ($R^2 = .040$) is significantly larger in magnitude than that of the skills regression ($R^2 = .006$), indicating that the work experience predictors are explain a larger percentage of the variability in the estimated salaries. Table 1: Estimated Effects of Skills on Salary and Table 2: Estimated

Effects of Work Experience on Salary provide more detail on the regressions themselves. It should be noted that both work experiences and skills alone are poor predictors of estimated salary, indicating that one cannot look at solely one type of predictor to explain labor market outcomes.

Equation (3) and its accompanying Table 3: Estimated Effects of Work Experience and Skills on Salary show that the only statistically significant predictors of salary outcomes were whether or not the person was hired at a company they had previously worked at, and the longest duration a person spent at their job. Additionally, the R^2 value for this regression is reasonably high, with around 78% of the variability of estimated salary being explained by this model.

In order to verify the unique significance of SameCompany and MaxDur, I run a regression with only SameCompany and MaxDur as predictors and Estimated Salary as the response - shown in Equation (4). I find that when you remove all predictors except for SameCompany and MaxDur from the regression, the R^2 remains approximately the same and the significance level of these two variables remain unchanged - as can be seen in Table 4: Estimated Effects of Employment at Same Company and Maximum Job Duration on Salary. This indicates that these SameCompany and MaxDur alone explain the variability of salary outcomes, and the other predictors were superfluous.

Tables

Table 1: Estimated Effects of Skills on Salary

	$Dependent\ variable:$	
	Estimated_Salary	
At_Least_2_of_Top_10	14,527.370	
	(26,760.020)	
Ten or more Endorsements	-17,555.000	
	(26,628.190)	
Constant	90,392.070***	
	(12,973.870)	
Observations	70	
\mathbb{R}^2	0.006	
Adjusted R^2	-0.023	
Residual Std. Error	$70,796.210 \; (\mathrm{df}=67)$	
F Statistic	$0.219 \; (\mathrm{df} = 2; 67)$	

Note:

*p<0.1; **p<0.05; ***p<0.01

 $At_Least_2_of_Top_10:$ $Ten_or_more_Endorsements:$

Having at least 2 of the top 10 skills in the sample Having at least 10 endorsements among top 10 skills

Table 2: Estimated Effects of Work Experience on Salary

_	$Dependent\ variable:$
	Estimated_Salary
At Least 1 Slevel	-4,204.582
	(17,551.860)
At Least 7 WE	$-25{,}101.370$
	(17,493.830)
At Least 30 months MaxDur	8,122.243
	(16,996.520)
Constant	101,125.200***
	(17,080.840)
Observations	70
\mathbb{R}^2	0.040
Adjusted R^2	-0.003
Residual Std. Error	$70{,}109.860 \; (\mathrm{df} = 66)$
F Statistic	0.921 (df = 3; 66)
Note:	*p<0.1; **p<0.05; ***p<0.01
$At_Least_1_Slevel:$	Having at least 1 prior senior level job

Note: p<0.1; **p<0.05; ***p<0.01At_Least_1_Slevel: Having at least 1 prior senior level job

At_Least_7_WE: Having at least 17 prior work experiences

At_Least_30_months_MaxDur: Longest work experience at least 30 months

Table 3: Estimated Effects of Work Experience and Skills on Salary

	Dependent variable:	
	Estimated_Salary	
At_Least_2_of_Top_10	5,612.486	
	(13,282.840)	
Ten_or_more_Endorsements	-10,191.350	
	(13,428.740)	
SameCompany	131,511.300***	
	(9,146.541)	
At_Least_1_Slevel	-593.604	
	(8,825.742)	
At_Least_7_WE	3,976.671	
	(13,307.710)	
At_Least_30_months_MaxDur	-6,053.556	
	(10,608.260)	
MaxDur	762.929***	
	(284.791)	
TotalWe	969.333	
	(2,506.233)	
Constant	-32,757.040	
	(20, 252.660)	
Observations	70	
\mathbb{R}^2	0.788	
Adjusted R^2	0.761	
Residual Std. Error	$34,245.060 \; (\mathrm{df} = 61)$	
F Statistic	$28.403^{***} (df = 8; 61)$	

Note:

*p<0.1; **p<0.05; ***p<0.01

 $Same Company: \\ MaxDur:$

Hired post-MBA at a company previously employed at Longest consecutive duration in months in a position

 $\hbox{ Table 4: Estimated Effects of Employment at Same Company and Maximum Job Duration on Salary } \\$

	Dependent variable:
	Estimated_Salary
SameCompany	129,667.400***
	(8,530.119)
MaxDur	601.205***
	(208.011)
Constant	-21,943.970**
	(10,369.390)
Observations ————————————————————————————————————	70
\mathbb{R}^2	0.782
Adjusted R ²	0.776
Residual Std. Error	33,126.320 (df = 67)
F Statistic	$120.509^{***} (df = 2; 67)$
Note:	*p<0.1; **p<0.05; ***p<0

Conclusion

My models confirmed my hypothesis that work experiences are a better predictor of labor market outcomes, as they alone explain more of the variability in salaries than skills do and within the larger regression, the only significant predictors of salary are related to work experience. This is most likely due to the fact that within the financial and consulting industries, firms will pay for their employees to obtain an MBA contingent on them returning to the company, largely eliminating uncertainty surrounding labor market outcomes. The tendencies of the financial and consulting industries to "unravel" time in the context of the labor market most likely biased the outcomes of this studies and restricts these findings to only people from these industries. In other words, because financial and consulting firms bypass the typical recruitment process of MBA graduates by offering incentives to current employees to complete a funded MBA program, provided they return to the company upon completion, the labor market for MBA graduate students experiences a distorted hiring timeline. This predetermined hiring process skews results for MBA students as many of them are guaranteed a job at a prior company upon graduation, making them not ideal candidates for studying the effect of skills or work experiences, as showcased on LinkedIn profiles, in determining labor market success. In the future, it would be of interest to extrapolate this study a larger selection of advanced degree graduates from a variety of industries to determine if work experience still function as strong predictors of labor market outcomes outside of the financial and consulting industries. Additionally, it would be interesting to study the efficacy of LinkedIn in reducing congestion through effective signaling mechanisms as their claim is that they bring economic opportunities to all strata of society via online connections and access to relevant job information. Possible ways to do so would be through ethnographic studies of recruiting professionals to understand how they utilize LinkedIn to find candidates.

Sources

Ahmad, Kamsuriah, Noor Faridatul Ainun Zainal, and Masura Rahmat. "Relationship Between Employability and Graduates' Skills." International Business Management 6.4 (2012): 440-45. Web.

Fisher, Catherine. (2016, October 26). Expert Tips for Making Skills on Your LinkedIn Profile Stand Out [Blog post]. Retrieved from https://engineering.linkedin.com/blog/2019/the-building-blocks-of-linkedin-skill-assessments

Gardecki, R., Neumark, D., & National Bureau of Economic Research. (1997). Order from chaos? the effects of early labor market experiences on adult labor market outcomes (Working paper series (National Bureau of Economic Research); working paper no. 5899). Cambridge, MA: National Bureau of Economic Research.

Humphrey, Aaron. "User Personas and Social Media Profiles." Persona Studies 3.2 (2017): 13-20. Web.

"LinkedIn Skill Assessments." LinkedIn Skill Assessments | LinkedIn Help, Nov. 2019, www.linkedin.com/help/linkedin/answer/94427/linkedin-skill-assessments?lang=en.

Peterson, Robert M, and Howard F Dover. "Building Student Networks with LinkedIn: The Potential for Connections, Internships, and Jobs." Marketing Education Review 24.1 (2014): 15-20. Web.

Roth, Alvin. "Marketplaces, Markets, and Market Design." The American Economic Review 108.7 (2018): 1609-658. Web.

Skeels, M. M., & Grudin, J. (2009). When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn. Proceeding GROUP '09 Proceedings of the ACM 2009 international conference on Supporting group work. New York, NY, USA: GROUP '09.