Skill Mismatch with Worker-side Learning Heterogeneity *

Bikalpa Baniya

December 1st, 2022

Abstract

Do interpersonal and cognitive skills adjust to skill mismatch? Recent papers have looked into the different characteristics of mismatch in skills. For example, Lise and Postel-Vinay (2020) find that cognitive skills adjust slowly and interpersonal skills are fixed over a lifetime. However, they and most of the literature simulate the on-the-job learning process. So, this paper revisits these results using Natural Language Processing to incorporate unique data on the heterogeneity of learning within the same occupation. I find that interpersonal skills do adjust to skill mismatch, albeit slowly.

*I would like to thank Prof. Jesse Gregory, Prof. Xiaoxia Shi, Prof. Emaad Manzoor, Prof. Paolo Martellini, Greg Kirwin, and Jingnan (Jane) Liu for their feedback at various stages of this project.

1 Introduction

Skill mismatch is a natural part of the labor market. People can be matched to unfitting jobs because of market conditions like recessions or just poor luck. Such matches are especially prevalent among younger workers, and they can be costly; switching from the worst to the best-matched decile can improve wages by at least 11% per year ((Jovanovic (1979); Guvenen et al. (2020)). Mismatches are also multi-dimensional. Heckman (1995) finds that different skills shape variations in earning and productivity, and this heterogeneity of skills among individuals along different dimensions also explains a substantial fraction of income dispersion among the college-educated (Ingram and Neumann (2006)). So, it is essential to study these skills separately to fully capture the dynamics of the labor market.

Recent papers have made essential contributions to the study of multi-dimensional mismatch and its effect on labor market outcomes (Lise and Postel-Vinay (2020);Guvenen et al. (2020);Taber and Vejlin (2020);Fredriksson et al. (2018) ; Sanders (2014)). In particular,Lise and Postel-Vinay (2020), henceforth LPV20, find that cognitive skills adjust very slowly and that interpersonal skills do not at all. The result of interpersonal skills being essentially fixed is especially surprising. Early career individuals will likely perform new tasks than their education requires, so there should be a considerable amount of on-the-job learning in the first few years. It should also be reasonable to assume that this on-the-job learning would result in the development of skills along multiple dimensions. Exploring these results is essential because task requirements have changed considerably towards social and cognitive skills and employers value these skills among younger employees (Atalay et al. (2020) ; Heller and Kessler (2022)). Social skills are particularly crucial from the worker's perspective because of their importance in wage growth in high-skill jobs (Deming (2017)). Moreover, social skills have also been shown to causally explain success in long-term labor market outcomes Heckman and Kautz (2012)).

So, this paper studies these results by relaxing a major on-the-job learning assumption used in recent skill mismatch literature. Differences in learning among workers are almost always captured using their occupational job requirements. This is usually done because data limitations make it very difficult to observe differences in learning among people in the same occupations. However, capturing this heterogeneity is important as the effect of differences in current tasks on worker productivity is well established (Heckman and Sedlacek (1985); Acemoglu and Autor (2011) ;Autor and Handel (2013) ; Manning (2003)). Here, LinkedIn is used to overcome this limitation. Every LinkedIn profile used by this paper has text descriptions of tasks completed at each job. This text is converted into numeric skill measures using Natural Language Processing to provide heterogeneity that related papers are not able to capture. This data is then integrated with occupation data from O*NET and average wage data from ACS to provide a more precise measure of skill mismatch.

I find that interpersonal skills adjust slowly among individuals with higher latent abilities. This adjustment is especially present for female workers. However, growth in interpersonal skills results in occupation switching at a lower rate than the growth in cognitive skills. So, its adjustment is harder to observe when the on-the-job learning model depends only on job characteristics. Nevertheless, initial mismatch and subsequent sorting into jobs with different skill requirements still explain a significant portion of the adjustment. The mechanisms for the adjustment are presented.

This paper is organized as follows. Section 2 provides a brief literature review. Section 3 describes the data, and section 4 describes how the heterogeneity of worker skills is extracted from the text data. Section 5 presents the reduced form models, the discussion, and the mechanisms. Section 6 concludes with avenues for future work. Finally, the appendix shows the completed modeling and tasks towards these avenues.

2 Literature Review

This paper follows a recently emerging literature on learning along several dimensions, multidimensional skill mismatch, and its effect on labor market outcomes. It is mainly related to four papers incorporating heterogeneity from both the worker and the employer sides. First is Taber and Vejlin (2020), which estimates a model incorporating Roy selection, search, compensation differentials, and on-the-job human capital accumulation. Second is Guvenen et al. (2020), which produces empirical, mostly reduced form, results to explore multidimensional skill mismatch and its impact on patterns of wages and occupational switching. Next, Sanders (2014) studies job sorting and multi-dimension learning when workers are uncertain about their skills. However, LPV20 is my basis paper. They construct a "structural model of on-the-job search [based on Cahuc et al. (2006)] in which workers differ in skills along several dimensions and sort themselves into jobs with heterogeneous skill requirements along those same dimensions." They retrieve employer- and worker-side heterogeneity from O*NET and NLSY79, respectively, and construct skill along three dimensions: interpersonal, cognitive, and mechanical. A key aspect of their model is that mismatch occurs due to random search and that they study the difference in the learning rate between skill types. They find that manual skills adjust quickly, cognitive skills slowly, and interpersonal skills not at all. They also find that the mismatch cost is much higher for cognitive skills than the other two.

LPV20 adds to the extensive literature on on-the-job learning and search. On-the-job learning models especially have had a long history in Economics from the foundational works of Becker (1964), Ben-Porath (1967), and Mincer (1974) on human capital accumulation theory. Becker (1964) outlined a learning-by-doing mode where productivity tomorrow depends on tasks completed today. Ben-Porath (1967) model looks at a person's investment in human capital as dependent on the total stock of prior knowledge, time devoted to learning, the technology used to learn, and individual ability. Mincer (1974) provides a wage model dependent on work experience and education. Similarly, the literature on search models is vast, but two are most relevant to LPV20 and my paper. First, Cahuc et al. (2006) provides an equilibrium search model with on-the-job random search and is a standard basis for any paper studying on-the-job search. Second, Burdett et al. (2011) construct another on-the-job search model based on Burdett and Mortensen (1998) with more integrated learning-by-doing effects. However, the second model does not include Bertrand's competition, so workers cannot negotiate with their current employer if they receive an outside offer.

The most significant contribution of my paper is the addition of heterogeneous on-the-job learning among workers. As discussed, almost all related papers assume that workers in the same occupation have the same exponential learning rate. So, crucial heterogeneity is lost. To my knowledge, two papers incorporate this heterogeneity to answer different questions: Autor and Handel (2013) and Stinebrickner et al. (2019). The former uses the Princeton Data Improvement Initiative (PDII) survey to capture differences in task requirements even in the same occupation. They find substantial task variations between and within occupations. Similarly, the latter uses the Berea Panel survey to measure the time Berea College graduates spend on different tasks and find that learning-by-doing effects are more significant for high-skill tasks than low-skill ones. My data differs from theirs in that I have observations from multiple locations and schools and more precise information on the type of degree[1].

Finally, there are papers (to my knowledge) that have similar resume-type data: Martellini et al. (2022) and Vafa et al. (2022). These papers have an impressive amount of observations. Still, they either answer a different question or employ a Natural Language Processing (NLP) technique unsuitable for my dataset. More specifically, Martellini et al. (2022) explores how college characteristics affect their graduates' wages, and Vafa et al. (2022) uses BERT Transformers to predict people's occupations. BERT is especially pertinent because it is considered the state-of-the-art Neural Net Language model but requires massive amounts of data. So, my sample size would not be suitable for this model. From the firm side, Atalay et al. (2020) utilize NLP to match the O*NET occupation codes with job ads.

3 Data

Three data sources have been used for the analysis: LinkedIn, O*NET, and ACS.

[1] Due to time constraints, the current version of this paper does not take advantage of this precise measure of education

Table 1: Demographic Information

Gender	
Female	44%
Non-female	56%
Education	
Post Graduate Degree	40%
Bachelors Degree	44%
Associates Degree	4%
Top Industries	
Higher Education	12%
IT and IT Services	12%
Computer Software	6%
Hospital and Health Care	5%
Financial Services	4.5%
Others	72%

3.1 Worker Characteristics

LinkedIn is a social networking website that users primarily use as an online resume. A typical user profile is the user's job history, and each profile lists previous and current job information: title, employer name, location, and dates. Many will also have job descriptions that list tasks completed during their tenure. Most people also include their education history by listing the degree name, the school name, the degree field, and enrollment dates. Finally, profiles include the total stock of skills, with endorsement from their coworkers, and a list of interests that dictates their feed.

Two thousand four hundred profiles have been scraped from 12 randomly selected cities out of the 50 most populous US cities. These profiles are chosen to have at least ten years of employment history and graduation dates between 1997 and 2012. These filters were used to ensure I captured people's "job shopping" period and focused on a relatively younger workforce. There were 125 unique industries, and the 2,400 profiles had 11,015 different skills and 17,448 different occupation titles. Out of the listed occupation titles, 15,045 had text descriptions of jobs. These are used to extract individual skills using Natural Language Processing (NLP) and Factor Modeling described in section 4.

3.2 Job Requirement and Wages

O*NET is used to retrieve Job requirements. It is an online database organized jointly by the US Department of Labor and the Employment and Training Administration. O*NET contains hundreds of job definitions and skill requirements for each job, and each job contains 277 descriptors of skills and a numeric value for relevance for each descriptor. Each job also contains text descriptions of the tasks involved; these (along with labeled LinkedIn job descriptions) will be used to train the NLP model. Two hundred sixteen of these skill descriptors are selected for analysis, and they fall into five categories: skills, abilities, knowledge, work activity, and work context. Moreover, only 875 occupations had data on all five categories, so the analysis is restricted to these occupations.

Similarly, the average wage data was retrieved using ACS because neither LinkedIn nor O*NET has this information. More specifically, the Public Use Micro Samples (PUMS) from 2004 to 2021 are used to find the average wage for a given occupation, year, and state. The Census Bureau collects PUMS data, and the aggregate files are published for either 5-year or 1-year intervals. I am using the 1-year State level data to ensure that all the states have wage data for most of the occupation titles.

3.3 Linking LinkedIn, O*NET and ACS

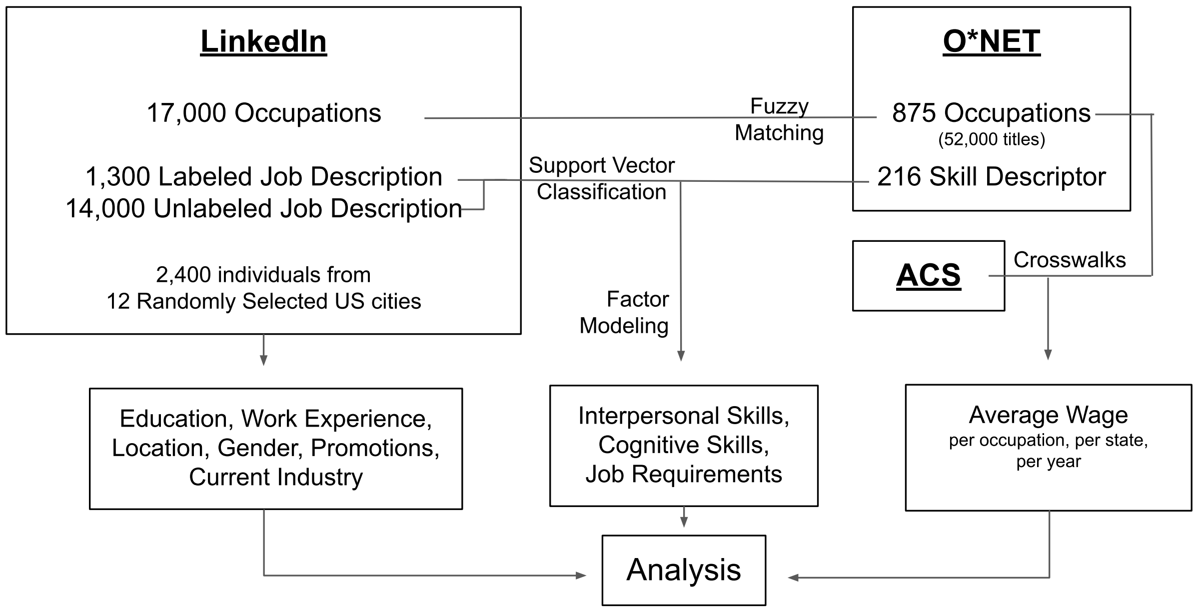
The 2,400 scraped LinkedIn profiles have 17,482 occupation titles, while O*NET has 923 occupations with 52,000 alternate titles. 67% of the 17,482 LinkedIn titles are linked to the O*NET titles using Fuzzy Matching [2] [3]. Similarly, LinkedIn users have labeled 9% of the job description with the different skills involved in their jobs. Six hundred twenty-six of these labeled skills are present in at least 20 profiles, so these are manually linked to the 216 O*NET skill descriptors[4]

[2] These are record-linking methods that account for the semantic variation of words that appear across separate datasets. Jaro-Winkler [Jaro (1989); Winkler (2006)] and Levenshtein [Yujian and Bo (2007)] are primarily used.

[3] This version of the paper only uses the linked titles, so there is a considerable decrease in sample size

[4] It is important to note that a skill might be associated with multiple skill descriptors.

Figure 1: Outline of Data and Data Processing



Correspondingly, using multiple crosswalks, I connect O*NET occupation titles to the ACS's PUMS occupation codes. More specifically, the O*NET website provides crosswalks between their occupation code and the Bureau of Labor Statistics' 2018 Standard Occupation Code (SOC). This 2018 SOC is linked to each of the classification systems present in ACS PUMS from 2004 to 2021 [5]. Autor and Handel (2013) and Deming and Kahn (2018) also use similar strategies.

4 Capturing heterogeneity of skill

Four data sources are used to predict skills in each job for each individual: unlabeled LinkedIn job descriptions, labeled LinkedIn job descriptions, O*NET job descriptions, and associated O*NET skills. The first is treated as the prediction set, and the remaining as the train-

[5] There were multiple occupations that were split or merged between years. These cases are handled individually, and multiple titles are merged whenever required.

ing/testing sets.

4.1 Estimating O*NET Skill Descriptors Values on LinkedIn

First, all text data is vectorized using a Bi-gram Bag of Words model. These models produce matrices with the job identifier as rows, pairs of words as columns, and the relevancy score of a given pair for a given job as the cell value[6]. Bi-grams (pairs of words) are used because they have more information than singleton words when describing skills. It is also essential to weigh the values of the Bi-grams. In the case of O*NET, some tasks are more relevant for a job, so the text vectors for these tasks are assigned higher weights [7]. Similarly, for labeled LinkedIn text, some skills descriptions might be exaggerated, so the number of skill endorsements an individual gains from their peers is used as the weight for its outcome variable.

Then, Support Vectors Classification (SVC) is used for skill estimation. SVC is one of the most robust classification methods in NLP (Vapnik (1982, 1995) and Chervonenkis (1974)). It uses labeled training data to estimate a hyperplane between categories. These hyperplanes are optimized to maximize the gap between the hyperplane and the classified groups. Prediction is then based on which side of the plane new data is observed. In my case, take a skill descriptor S and discretize it into four quartiles. The weighted vectorized training data creates linear SVC hyperplanes for each quartile of S. Unlabeled LinkedIn text is plotted in this "graph" and the O*NET skill value S are predicted based on the location of the text relative to the SVCs. This process is repeated for all 216 skills descriptors. A 5-fold cross-validation on 1/3 of the Labeled LinkedIn data resulted in 72% prediction accuracy [8].

[6]Here, the relevancy measure is the term frequency-inverse document frequency (TF-IDF) score. It is simply a multiple of the count a word pair appears in a job description and the inverse of how often it appears across all text. TF-IDFs are often better than counts because certain words might appear often but do not carry much information

[7]Importance of tasks to a given job is given in O*NET, and these are used as the weights

[8]Naive Bayes and Logistic regressions were also used, but SVC produced the highest cross-validation accuracy

4.2 Constructing Interpersonal and Cognitive Skills

A given skill descriptor might be relevant to multiple skill dimensions, so latent factors are considered. Moreover, a variation on Factor Modeling is used to produce Interpersonal and Cognitive skills to improve interpretability. The strategy for O*NET is identical to LPV20, and the following describes a similar process for LinkedIn. However, it is essential to note that LPV20 produces three skills, but I only produce two because mechanical jobs and skills are less relevant to my sample.

There are 2,400 LinkedIn profiles containing 15,428 jobs with text descriptions. SVCs produce estimations for these jobs along the 216 skill descriptors used by O*NET. Now let these jobs be denoted by a matrix $\mathbf{M}$ with dimension $N \times P$ matrix where N is the number of jobs and P are the number of skill descriptors. Rearrange the columns such that the first two columns are descriptors for *social perceptiveness* and *mathematics knowledge*, respectively. This rearrangement will ensure that the first two factor loadings are relevant to interpersonal and cognitive skills. These were also the descriptors used by LPV20.

Now, Principal Component Analysis (PCA) decomposes $\mathbf{M}$ into $\mathbf{M} = \mathbf{FL}$. Here, $\mathbf{F}$ is $N \times P$ matrix of orthogonal eigenvectors of the matrix $\mathbf{M}^T\mathbf{M}$. L is a $P \times P$ matrix of factor loadings. Now $\mathbf{M}$ can be reformulated as $\mathbf{M} = (\mathbf{F}_2\mathbf{T})(\mathbf{T}^{-1}\mathbf{L}_2) + \mathbf{U}$, where F_2 is first two columns of F, L_2 is the first two rows of L and $\mathbf{T}$ is the is the first two columns of L_2. Under this new formulation, $(\mathbf{F}_2\mathbf{T})$ gives two principal components for the $\mathbf{M}$. Here, the highest factor loading for the first principal components were *Speaking*, *Speech Clarity* and *Oral Expression*, and those for the second were *Inductive Reasoning*, *Number Facility* and *Complex Problem Solving*. Thus, the two PC do describe interpersonal and cognitive skills, respectively.

5 Reduced Form Models and Discussion

This section provides reduced-form models and the subsequent discussion. It is organized as follows. Section 5.1 presents a wage regression to introduce the data and present some comparisons with LPV20. Section 5.2 presents the primary regression model to observe the change in skill match, and section 5.3 discusses a possible mechanism for the observations.

5.1 Current Wage and Skills

Table 2 provides the summary statistics of most variables used in the regressions. Notice that the mean and maximum log wage is higher in the current job. This increase points to wage trends and that current jobs require more ability along these dimensions. Also, the sample has an average of 16 years of work experience, so people on average should be settled into jobs rights after their early career "shopping period".

$$\ln(W_{ik}) = \beta_0 + \sum_{j=I,C}\Big[\beta_{1j}R_{ij} + \beta_{2j}S_{ij} + \beta_{3j}R_{ij} \times S_{ij}\Big] + \mathbf{M}_i + \epsilon_i \tag{1}$$

Regression (1) presents a wage regression. Here, W_{iK} is the average wage of an individual i's current occupation K. R_{iI} refers to the interpersonal job requirement at K whereas R_{iC} refers to the cognitive job requirement at K. Similarly, S_{iI} is the interpersonal skill level of i at K while S_{iC} is their cognitive skill level at K. M_i is a vector of i's Mincerian variables like current work experience and education dummies.

Table 3 presents the results of the regression above. Higher job requirements would result in higher wages, as expected. However, note that my sample has a higher average education level than LPV20, so they could sort into jobs with different skill requirements. This sorting effect can be seen in the magnitude of the coefficient of job requirement in reg (1.3) and (1.4). Although returns to cognitive skills are higher than interpersonal skills, the difference is lower than in LPV20. This observation is explored further in section 5.3.

Next, the coefficient of work experience and education dummies in Table 3 is not prac-

Table 2: Summary Statistics

	Count	Mean	S.D.	Min	Max
First Job					
Interpersonal Skill	1306	-.1341	6.696	-18.435	17.892
Cognitive Skill	1306	-0.211	5.250	-13.720	13.015
Interpersonal Job Requirement	1306	4.743	6.737	-17.238	17.744
Cognitive Job Requirement	1306	4.601	10.547	-27.083	26.198
Promotions	2400	0.320	0.978	0	11
Tenure (Years)	2391	3.385	3.341	0.083	24
ln(Wage)	637	10.598	0.581	7.101	11.904
Current Job					
Interpersonal Skill	1047	-0.290	7.820	-18.435	38.823
Cognitive Skill	1047	-0.117	6.078	-13.454	30.114
Interpersonal Job Requirement	1047	6.606	6.183	-18.744	17.744
Cognitive Job Requirement	1047	8.315	9.220	-22.240	28.941
Promotions	2400	0.386	0.937	0	8
Tenure (Years)	2361	2.660	3.140	0.083	28.5
ln(Wage)	733	10.836	0.486	8.869	12.170
Female	2400	0.435	0.495	0	1
Work Experience (Years)	2400	16.440	5.381	9	27
Associate Degrees	2400	0.058	0.233	0	1
Bachelors Degree	2400	0.442	0.496	0	1
Postgraduate Degree	2400	0.344	0.475	0	1

Table 3: Log Wage of Current Job

	(1.1)	(1.2)	(1.3)	(1.4)	(1.5)	(1.6)
Interpersonal Skill in Current Job						
Job Requirement	2.220***	2.039*	2.259***	1.960*		
	(0.542)	(1.100)	(0.480)	(1.044)		
Skill	-1.020	0.920			-5.645***	-4.259
	(1.485)	(2.714)			(1.758)	(3.562)
Job Requirement × Skill	3.328	4.475				
	(3.046)	(6.749)				
Cognitive Skill in Current Job						
Job Requirement	2.550***	2.803***	2.567***	2.853***		
	(0.290)	(0.702)	(0.256)	(0.668)		
Skill	1.764	-0.003			7.915***	6.271
	(1.808)	(3.250)			(2.253)	(4.370)
Job Requirement × Skill	-1.491	-8.506				
	(2.870)	(6.473)				
ln (Wage), 1st Job		0.145***		0.143***		0.253***
		(0.048)		(0.049)		(0.070)
Tenure, Current Job	-0.005	-0.005	-0.004	-0.001	-0.012**	-0.013
	(0.004)	(0.007)	(0.004)	(0.007)	(0.006)	(0.009)
Work Experience	0.002	-0.005	0.004	0.003	0.008	-0.008
	(0.009)	(0.022)	(0.009)	(0.022)	(0.011)	(0.031)
(Work Experience)^2	0.000	0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Associates Degree	-0.095	-0.033	-0.098*	-0.073	-0.136*	0.031
	(0.058)	(0.104)	(0.059)	(0.101)	(0.078)	(0.142)
Bachelors Degree	-0.020	0.009	-0.019	-0.002	0.041	0.121
	(0.046)	(0.081)	(0.046)	(0.078)	(0.058)	(0.115)
Postgraduate Degree	-0.076	0.015	-0.076	-0.010	0.054	0.193
	(0.047)	(0.086)	(0.047)	(0.086)	(0.060)	(0.126)
Constant	10.569***	9.037***	10.548***	8.961***	10.708***	8.020***
	(0.092)	(0.533)	(0.092)	(0.536)	(0.120)	(0.760)
R^2	0.454	0.484	0.447	0.467	0.049	0.107
Observations	727	217	727	217	727	217

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

tically significant. This observation reflects those seen in LPV20. Their interpersonal and cognitive variables did not capture the wage growth trend, so they introduced "general skills" into their structural model to capture this growth [9]

5.2 Mismatch and Skill Adjustment

The following regressions show the factors that might affect the current mismatch in skills, including mismatch in their first job. This model and the subsequent mechanisms are used to show that interpersonal and cognitive skills do adjust.

$$\begin{aligned}\triangle I_{iK} = \beta_0 + \beta_1 \triangle I_{iF} + &\sum_{j\in\{F,K\}}\Big[\beta_{2j}P_{ij} + \beta_{3j}T_{ij} + \beta_{4j}P_{ij}\times T_{ij}\Big] \\ &+ \beta_5 \triangle Y_i + \beta_6(\triangle Y_i)^2 + \mathbf{X}_i + \epsilon_i\end{aligned} \tag{2}$$

$$\begin{aligned}\triangle C_{iK} = \gamma_0 + \gamma_1 \triangle C_{iF} + &\sum_{j\in\{F,K\}}\Big[\gamma_{2j}P_{ji} + \gamma_{3j}T_{ji} + \gamma_{4j}P_{ji}\times T_{ji}\Big] \\ &+ \gamma_5 \triangle Y_i + \gamma_6(\triangle Y_i)^2 + \mathbf{X}_i + \epsilon_i\end{aligned} \tag{3}$$

Regression (2) and (3) presents the mismatch in interpersonal and cognitive skill, respectively. Here, $\triangle I_{iK} = R_{iK} - S_{iK}$ is the interpersonal mismatch in the current job, and $\triangle I_{iF} = R_{iF} - S_{iF}$ is the interpersonal mismatch in the first job. $\triangle C_{iK}$ and $\triangle C_{iK}$ is defined analogously for cognitive skills. Now change in skill mismatch can occur either because skill level changes or because people switch to a job with different skill requirements. The remaining variables are used to condition on these scenarios.

The number of promotions within a company P_{ij} and occupation tenure T_{ij} is used to condition on skill level change. More specifically, promotion is a proxy for i's latent ability on the task. It is assumed that individuals with higher abilities will have a higher rate of on-the-

[9] Note that introduction of wage in the first job decreases the sample size substantially. This decrease is caused by only 67% of the LinkedIn occupation titles being connected to O*NET. Future iterations of this paper will link more titles to increase the sample size when wages are observed at multiple periods.

job learning. Tenure could have a similar effect as better-matched individuals can hold their jobs for longer. Concurrently, the difference in job requirements $\triangle Y_i = R_{ijK} - R_{ijC}$ is used to condition on the possibility that i might switch to jobs with different skill requirements. Finally, $\mathbf{X}_i$ is a vector of relevant controls such as gender, work experience, and education.

Table 4 presents the results of regression (2) and (3), and there are several things to note. First, a mismatch in the first job positively impacts the mismatch in the current job or, in other words, initial match quality matters. This result can be observed throughout the literature. For example, Fredriksson et al. (2018) argue that employers have lower information about the match quality with an inexperienced worker and the initial misalignment decreases the amount of information available throughout the job "shopping" period of workers.

Second and most importantly, promotions in the current job do have a significant effect. Note that the number average number of promotions from Table 2 is 0.32 and 0.38 in the first and current jobs, respectively. Likewise, the average tenure is 3.38 years and 2.66 years, respectively. Given the coefficients, the impact of promotion in the first job for an average worker is $(-1.022) * 0.32 + (0.33) * 3.38 * 0.32 = 0.031$, which in close to zero[10]. However, estimated *average lower bound* of the effect of promotion in the second job is $(-1.269) * 0.38 + (0.31) * 2.66 * 0.386 = -0.168$[11]. Since promotions have been used as a proxy for ability, being good in the current job seems to be related to decreased mismatch. This effect might not necessarily be causal because mismatch can affect promotions, and the effect seems small. Nevertheless, the correlation is meaningful because it signals that LPV20's result about interpersonal skills not adjusting might not always hold.

Third, change in job requirements still considerably contributes to the change in mismatch, and people seem to find a better fit over time. This result aligns with Cahuc et al. (2006) and similar on-the-job search models. They look at a similar population and explain that while job search and subsequent switching decrease with experience, the effect is still

[10] Average Effect of promotion = (Coefficient of Promotion) * Average Number of Promotion + (Coefficient of Interaction term) * Average Number of Promotion * Average Tenure

[11] Here, the coefficient of the interaction is not statistically significant, so the 95% CI is taken using the standard error. This should estimate the lower bound of the average effect

Table 4: Mismatch in Current Job

	Interpersonal	Cognitive
Interpersonal Mismatch, 1st Job	0.439***	
	(0.046)	
Cognitive Mismatch, 1st Job		0.784***
		(0.031)
Tenure, Current Job	-0.176*	-0.098
	(0.097)	(0.083)
Promotions, Current Job	-1.269**	-0.970**
	(0.501)	(0.427)
Tenure × Promotions, Current Job	0.026	-0.019
	(0.146)	(0.129)
Tenure, 1st Job	0.085	0.093
	(0.100)	(0.084)
Promotions, 1st Job	-1.022*	-1.247***
	(0.531)	(0.393)
Tenure × Promotions, 1st Job	0.331**	0.321***
	(0.160)	(0.103)
△ Job Requirement, Interpersonal	0.675***	
	(0.047)	
(△ Job Requirement, Interpersonal)^2	-0.010***	
	(0.003)	
△ Job Requirement, Cognitive		0.847***
		(0.028)
(△ Job Requirement, Cognitive)^2		-0.000
		(0.001)
Female	-0.549	-1.747***
	(0.669)	(0.583)
Work Experience	-0.178***	-0.095*
	(0.067)	(0.057)
Associates Degree	3.850**	3.244**
	(1.720)	(1.390)
Bachelors Degree	0.882	0.403
	(1.088)	(0.932)
Postgraduate Degree	1.627	1.194
	(1.172)	(1.022)
Constant	6.446***	3.386***
	(1.506)	(1.239)
R^2	0.311	0.651
Observations	580	580

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

meaningful in the first ten years. Furthermore, the coefficient on change in cognitive requirements points out that this change effects present cognitive mismatch more than that for the interpersonal requirement.

Finally, the coefficient on Associates degree is positive in both cases, so having an Associates degree interestingly makes individuals worse off than those with lower education. The possible mechanism for this and prior observations are explored further next.

5.3 Mechanisms

As noted above, a mismatch can change either because an individual switched jobs or their skills improved. So, it is essential to examine these effects separately.

Mechanism 1: Employer Characteristics

How does prior experience affect sorting into the current job? The following regression examines the effect of characteristics of the first job and other relevant variables in sorting into jobs with different interpersonal and cognitive requirements.

$$R_{K_{iI}} = \beta_0 + \sum_{j \in \{I,C\}} \left[\beta_{1j} S_{F_{ij}} + \beta_{2j} R_{F_{ij}} \right] + \beta_3 P_{iF} + \beta_4 T_{iF} + \beta_5 P_{iF} \times T_{iF} + \mathbf{X}_i + \epsilon_i \tag{4}$$

Regression (4) presents the model for interpersonal job requirements, and the model for the cognitive requirement is analogous. Here, $R_{K_{iI}}$ is the interpersonal job requirement in the current job. $S_{F_{ij}}$ and $R_{F_{ij}}$ are skills and job requirements of i at the first job respectively. Similarly, P_{iF} and T_{iF} respectively define the number of promotions and tenure at the first job. Finally, $\mathbf{X}_i$ is a vector of relevant controls, including gender and education.

Table 5 presents the result for the regression. First, note the high value of the coefficient for interpersonal and cognitive skills for the current cognitive job requirement. In other words, higher interpersonal and cognitive skills in the first job would result in sorting into a current job with higher cognitive skill requirements. As Autor et al. (2003) and Autor et al.

Table 5: Skill Requirement of Current Job

	Interpersonal	Interpersonal	Cognitive	Cognitive
Interpersonal Skill, 1st Job	0.195	0.181	1.261***	1.258***
	(0.328)	(0.323)	(0.471)	(0.464)
Cognitive Skill, 1st Job	-0.175	-0.160	1.572***	1.568***
	(0.420)	(0.413)	(0.603)	(0.592)
Interpersonal Requirement, 1st Job	0.305***	0.286***	0.206***	0.247***
	(0.054)	(0.053)	(0.076)	(0.075)
Cognitive Requirement, 1st Job	-0.048	-0.054*	0.275***	0.272***
	(0.031)	(0.031)	(0.047)	(0.047)
Promotions, 1st Job	0.686*	0.690*	0.391	0.377
	(0.381)	(0.401)	(0.469)	(0.484)
Tenure, 1st Job	-0.085	-0.071	-0.210*	-0.174
	(0.074)	(0.075)	(0.115)	(0.115)
Promotions× Tenure, 1st Job	-0.066	-0.080	-0.107	-0.143
	(0.134)	(0.131)	(0.117)	(0.116)
Female	0.888*	0.901*	-2.168***	-2.185***
	(0.483)	(0.478)	(0.771)	(0.766)
Associates Degree		0.817		3.761**
		(1.122)		(1.747)
Bachelors Degree		1.672*		3.725***
		(0.870)		(1.222)
Postgraduate Degree		2.811***		5.593***
		(0.887)		(1.306)
Constant	4.919***	3.170***	9.337***	5.465***
	(0.469)	(0.882)	(0.745)	(1.200)
R^2	0.117	0.148	0.137	0.177
Observations	588	588	588	588

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(2008) explain, the returns to cognitive skills have increased in the past few decades because computers complement cognitively complex and abstract tasks. So, signaling this ability in the first job would explain positive sorting into similar tasks.

There is also extensive literature studying the effect of interpersonal ability and tasks on cognitive job sorting (Deming (2017); Deming and Kahn (2018); Weinberger (2014); Beaudry et al. (2016)). This literature suggests a high complementary between the two skills among high skills jobs (the majority of jobs in my sample). For example, Deming (2017) notes that while automation has substituted a significant number of tasks in the labor market, the tasks that cannot be substituted away are usually complemented by interpersonal skills. This complementarity is especially pronounced by the difficulty of automating social skills Autor (2015). The coefficient of the interpersonal requirement for the current cognitive requirement regression provides further evidence for this complementarity.

Second, gender seems to play an essential role in job sorting, and the effect differs between skills. Being female would result in positive sorting into jobs with higher interpersonal requirements. As Autor and Wasserman (2013) note, researchers have only recently examined the varied effect of change in technology for women. They argue that women are disproportionately endowed with valuable interpersonal skills in information and technology-rich high-skill environments.

On the other hand, the challenges for females in the labor market, especially in high-skill jobs, is also well known (Brenøe and Zölitz (2020); Beede et al. (2011)) and can be seen in the negative sorting into cognitive jobs. The magnitude of the coefficient is particularly revealing. Kahn and Ginther (2017) explain the myriad of factors that affect women: for example, stereotyping, cultural roles, competition, risk aversion, and beliefs that are not directly tied to gender. Finally, education has the expected effect. Higher education would result in sorting into jobs with higher requirements, and this magnitude is greater for cognitive skills and higher degrees.

Mechanism 2: Worker Characteristics

Most of the literature on skill mismatch estimates the variation among workers' skills using the characteristics of their job. So, the majority of literature is relevant to sorting through Mechanism 1. This paper can provide a unique contribution to the literature through Mechanism 2 because it can capture the heterogeneity of workers without relying on employer characteristics. I also provide a possible explanation as to why LPV20 does not clearly observe interpersonal skill adjustment.

$$\begin{aligned} S_{K_{iI}} =&\beta_0 + \beta_1 P_{iK} + \beta_2 T_{iK} + \beta_3 P_{iK} \times T_{iK} \\ &+ \beta_4 T_{iF} + \sum_{J \in \{I,C\}} \Big[\beta_{5j} \triangle J_{iF} + \beta_{6j} \triangle J_{iF} \times T_{iF} \Big] + \mathbf{X}_i + \epsilon_i \end{aligned} \tag{5}$$

Equation (5) presents a regression of current interpersonal skills on promotion and tenure at current job and past job characteristics. The regression for cognitive skills is analogous. Unlike Mechanism 1, *current job* characteristics, promotions P_{iK} and tenure T_{iK}, are added because current employment depends only on prior skills. but the current skill might be affected by the current job too. $\triangle J_{iF}$ is the skill mismatch in the first job along dimension $J \in \{I, C\}$, T_{iF} is the tenure in first job and $\mathbf{X}_i$ is a vector of relevant controls.

Table 6 presents the result for the regression above, and the results point to the importance of ability. There is a positive correlation between promotion (a proxy for ability in current job requirements) and current skills, even after considering the coefficient and standard error of the interaction terms. This suggests further evidence of the adjustment of skills. Moreover, note that the coefficient of promotion for cognitive skills is smaller than that for interpersonal. A combination of these two effects might cause this difference in coefficients. First, Stinebrickner et al. (2019) find that there is strong evidence of on-the-job learning for high-skill tasks and learning interpersonal skills might be easier than cognitive ones for the same time frame. Second, the on-the-job learning rate might be higher if an

Table 6: Current Skills

	Interpersonal	Interpersonal	Cognitive	Cognitive
Promotions, Current Job	1.457**	1.420**	1.209***	1.175***
	(0.576)	(0.573)	(0.439)	(0.444)
Tenure, Current Job	0.111	0.104	0.102	0.092
	(0.087)	(0.088)	(0.073)	(0.073)
Promotions × Tenure, Current Job	-0.097	-0.080	-0.078	-0.064
	(0.174)	(0.175)	(0.123)	(0.125)
Tenure, 1st Job	0.014	0.024	-0.040	-0.030
	(0.102)	(0.103)	(0.081)	(0.082)
Interpersonal Mismatch, 1st Job	-0.024	-0.027		
	(0.050)	(0.050)		
Interpersonal Mismatch × Tenure, 1st Job	-0.029***	-0.028***		
	(0.008)	(0.008)		
Cognitive Mismatch, 1st Job			-0.040	-0.042
			(0.029)	(0.029)
Cognitive Mismatch × Tenure, 1st Job			-0.011**	-0.011**
			(0.005)	(0.005)
Female	1.411**	1.504**	0.367	0.461
	(0.637)	(0.643)	(0.500)	(0.505)
Work Experience	0.017	-0.049	0.075	0.043
	(0.300)	(0.307)	(0.227)	(0.235)
(Work Experience $)^2$	0.004	0.005	0.002	0.003
	(0.007)	(0.007)	(0.006)	(0.006)
Associates Degree		-3.237**		-2.713**
		(1.614)		(1.244)
Bachelors Degree		-0.785		-0.518
		(1.010)		(0.797)
Postgraduate Degree		0.185		0.115
		(1.071)		(0.861)
Constant	-2.774	-1.586	-2.566	-1.914
	(2.810)	(3.007)	(2.123)	(2.279)
R^2	0.071	0.082	0.061	0.073
Observations	580	580	580	580

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

individual is especially good at current tasks. We can put this together with the result of positive sorting between cognitive skills and requirements in Table 5. If individual is good at their job and gains some cognitive skills, they are more likely to switch to a job with a higher cognitive requirement because of its high return. LPV20 can observe this switching. However, the same is not necessarily valid for growth in interpersonal skills, as occupational switching might occur at lower rates, and LPV20 will not necessarily observe this. So, even though interpersonal skills might be adjusting their model would not detect it.

Gender also seems to have an essential effect on interpersonal ability, but similar to mechanism 1, this effect might be attributed to the difference in interpersonal ability between gender. Finally, having an Associate's degree makes an individual worse off than those with less education. However, only 4% of the sample has this degree so the effect might not be as important.

6 Conclusion and Future Work

This paper builds on the recent and emerging work on skill mismatch by incorporating heterogeneity of learning among workers even within the same occupation. The dynamics of mismatch-change from both worker and employer sides have been explored since mismatch could adjust from either path. I find counter-evidence to LPV20's result that interpersonal skills do not adjust to skill mismatch. Interpersonal skills adjust slowly among individuals with higher latent abilities, and this adjustment is especially present for female workers.

Future work will build on this paper in two regards. First, a direct comparison is needed to properly compare the results with those of LPV20. So, the data will be modified to fit their structural model and to carry out the simulations. It is important to note that LPV20's model assumes exponential growth. This assumption is restrictive because everyone faces time and resource constraints. It is reasonable to consider that a change in the learning rate would follow an S-shaped logistic curve. In other words, people learn new skills slowly, and the rate increase with time. However, they eventually face constraints, and the rate decreases with time. Appendix A presents the completed extension of the model to include this new growth model, and the simulation will be carried out in the future version of the paper.

Moreover, two kinds of skills could also be extracted from the text. One is more specific to an occupation group, like teacher training, and the other is more general, like leadership. SVCs are good at extracting the first kind, but a latent variable model is needed for the second kind. Appendix B presents work towards this skill extraction.

References

D. Acemoglu and D. Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier, 2011.

E. Atalay, P. Phongthiengtham, S. Sotelo, and D. Tannenbaum. The evolution of work in the united states. *American Economic Journal: Applied Economics*, 12(2):1–34, 2020.

D. Autor. Why are there still so many jobs? the history and future of workplace automation. *Journal of economic perspectives*, 29(3):3–30, 2015.

D. H. Autor and M. J. Handel. Putting tasks to the test: Human capital, job tasks, and wages. *Journal of labor Economics*, 31(S1):S59–S96, 2013.

D. H. Autor and M. Wasserman. Wayward sons: The emerging gender gap in labor markets and education. *Third Way Report*, 20013, 2013.

D. H. Autor, F. Levy, and R. J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333, 2003.

D. H. Autor, L. F. Katz, and M. S. Kearney. Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2):300–323, 2008.

P. Beaudry, D. A. Green, and B. M. Sand. The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):S199–S247, 2016.

G. S. Becker. Human capital theory. *Columbia, New York*, 1964, 1964.

D. N. Beede, T. A. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms. Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11), 2011.

Y. Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of political economy*, 75(4, Part 1):352–365, 1967.

A. A. Brenøe and U. Zölitz. Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4):1009–1054, 2020.

K. Burdett and D. T. Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273, 1998.

K. Burdett, C. Carrillo-Tudela, and M. G. Coles. Human capital accumulation and labor market equilibrium. *International economic review*, 52(3):657–677, 2011.

P. Cahuc, F. Postel-Vinay, and J.-M. Robin. Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, 74(2):323–364, 2006.

D. Deming and L. B. Kahn. Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369, 2018.

D. J. Deming. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4):1593–1640, 2017.

P. Fredriksson, L. Hensvik, and O. N. Skans. Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review*, 108(11):3303–38, 2018.

F. Guvenen, B. Kuruscu, S. Tanaka, and D. Wiczer. Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–44, 2020.

J. J. Heckman. Lessons from the bell curve. *Journal of Political Economy*, 103(5):1091–1120, 1995.

J. J. Heckman and T. Kautz. Hard evidence on soft skills. *Labour economics*, 19(4):451–464, 2012.

J. J. Heckman and G. Sedlacek. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *Journal of political Economy*, 93(6): 1077–1125, 1985.

S. B. Heller and J. B. Kessler. Soft skills in the youth labor market. In *AEA Papers and Proceedings*, volume 112, pages 121–25, 2022.

B. F. Ingram and G. R. Neumann. The returns to skill. *Labour economics*, 13(1):35–59, 2006.

M. A. Jaro. Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. *Journal of the American Statistical Association*, 84(406):414–420, 1989.

B. Jovanovic. Job matching and the theory of turnover. *Journal of political economy*, 87(5, Part 1):972–990, 1979.

S. Kahn and D. Ginther. Women and stem. Technical report, National Bureau of Economic Research, 2017.

J. Lise and F. Postel-Vinay. Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–76, 2020.

A. Manning. The real thin theory: monopsony in modern labour markets. *Labour economics*, 10(2):105–131, 2003.

P. Martellini, T. Schoellman, and J. Sockin. The global distribution of college graduate quality. *Federal Reserve Bank of Minneapolis Working Paper*, 791, 2022.

J. Mincer. Schooling, experience, and earnings. human behavior & social institutions no. 2. 1974.

C. Sanders. Skill accumulation, skill uncertainty, and occupational choice. *Working Paper*, 2014.

R. Stinebrickner, T. Stinebrickner, and P. Sullivan. Job tasks, time allocation, and wages. *Journal of Labor Economics*, 37(2):399–433, 2019.

C. Taber and R. Vejlin. Estimation of a roy/search/compensating differential model of the labor market. *Econometrica*, 88(3):1031–1069, 2020.

K. Vafa, E. Palikot, T. Du, A. Kanodia, S. Athey, and D. M. Blei. Learning transferrable representations of career trajectories for economic prediction. *arXiv preprint arXiv:2202.08370*, 2022.

C. J. Weinberger. The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics*, 96(5):849–861, 2014.

W. E. Winkler. Overview of record linkage and current research directions. In *Bureau of the Census*. Citeseer, 2006.

K. Yu, S. Yu, and V. Tresp. Multi-label informed latent semantic indexing. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 258–265, 2005.

L. Yujian and L. Bo. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1091–1095, 2007.

Appendix A: Structural Model

The following model is based on LPV20(2020) with three major exceptions. First, they consider workers to differ across four skill dimensions: Interpersonal, Cognitive, Mechanical, and General [12]. Second, they assume that two workers working in the same occupation have the same learning rate. The LinkedIn data allows me to relax this assumption so that direct observation of learning is incorporated into the model. The scraped LinkedIn data contains very few occupations and skills where mechanical skills are relevant, so only the remaining three are considered.

Third, occupations have different codes or titles across the three datasets [13], and multiple crosswalks were created or used for this linkage. However, the current version of this paper has only been able to link 67% of the LinkedIn occupations. Restricting the analysis to these linked occupations would result in a considerable reduction in sample size. So, I am using all profiles with at most one missing (unlinked) occupation. This occupation is then simulated to follow logistics growth from the previous job. It is important to note that Lise and Postel-Vinay's (2020) model assumes exponential growth. This assumption is restrictive because everyone faces time and resource constraints. It is reasonable to consider that a change in the learning rate would follow an S-shaped logistic curve. In other words, people learn new skills slowly, and the rate increase with time. However, they eventually face constraints, and the rate decreases with time.

Mismatch Characterization, On-the-job Learning

Consider a worker i and their occupation j. Worker i is described by their skill in three skill groups: General $s(0)$, Interpersonal $s(1)$, and Cognitive $s(2)$. Let $x_{i_s}(t=0)$ denote their skill at job-market entry with $x_{i_s} \in \mathbb{R}$. Similarly, occupation j is described by its skill requirement across these two dimensions, Interpersonal $y_{j_{s(1)}}$ and cognitive $y_{j_{s(2)}}$, with $y_{j_s} \in \mathbb{R}$[14]. The

[12]Here, general skills are those that are relevant to all occupations, such as basic literacy skills
[13]This is discussed in section 4
[14]These requirements are fixed and assumes that the firm's technology does not change

growth of general skills is assumed not to affect the worker's market productivity. In other words, an occupation only cares about $s(0)$ and $s(1)$. Mismatch between i and j at time t in skill dimension s is defined as $m_s(i,j,t) = x_{i_s}(t) - y_{j_s}$. This is the difference between the occupation's skill requirements and the worker's skills at time t.

Whenever on-the-job learning needs to be modeled, it follows the following standard logistic growth model. If worker i has started working with j and time $t-1, t_1 < t$, their on-the-job learning for skill s follows from

$$x_{i_s}(t) = y_{j_s} - \frac{m_s(i,j,t_1)\cdot A_s \cdot e^{(t-t_1)\cdot\pi_s(i,t)}}{(A_s - m_s(i,j,t_1)) + m_s(i,j,t_1)\cdot e^{(t-t_1)\cdot\pi_s(i,t)}} \text{ for } s \in \{s(1), s(2)\} \tag{6}$$

Here, A_s denotes the upper asymptote of learning in skill s or the maximum value of x anyone could achieve in this skill. $\pi_s(i,t)$ is the learning growth rate of person i in skills s in time t.

Worker and Firm Interaction

Frictions: Consider four exogenous frictions: λ_1, λ_0, δ and μ. λ_1 is the sampling rate of alternate jobs for employed workers, and λ_0 is the sampling rate of alternate jobs for unemployed workers. An employed worker is forced into unemployment at the rate δ; similarly, a worker chooses to exit the market with the rate μ. Finally, let r be the discount rate for the future.

Private Values: Let $\mathbf{x}$ and $\mathbf{y}$ denote the vectors of workers' skills and the skills requirements of a firm, respectively. The worker with $\mathbf{x}$ values unemployment with $U(\mathbf{x}) = b\cdot x_{i_{s(0)}}/(r+\mu-g)$, where b is a positive constant and g is the constant growth rate of general skills. In other words, the worker's unemployment income is only dependent on their general skills, discounted for the possibility their might exit anyway μ, the future discount rate r, and the loss in learning g[15].

The private value of working for type-$\mathbf{y}$ firm is $P(\mathbf{x}, \mathbf{y})$. If a worker does work with a firm, we know that the worker's value of the contract W is $P(\mathbf{x}, \mathbf{y}) \geq W \geq U(\mathbf{x})$. Similarly, the worker's share of total surplus generated with the match W is $(W - U(\mathbf{x}))/(P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x}))$.

Bertrand Competition: Consider the case where an employed worker with skill bundle $\mathbf{x}$ receives an offer from an outside firm with requirements $\mathbf{y'}$. The worker will switch to the new job if the private value for working in the new job is higher than the old job, $P(\mathbf{x}, \mathbf{y}) < P(\mathbf{x}, \mathbf{y'})$. However, when that is not the case, $W < P(\mathbf{x}, \mathbf{y'}) \leq P(\mathbf{x}, \mathbf{y})$, the worker can negotiate their current wage, and the renegotiated share of the matched surplus is

$$\sigma(\mathbf{x}, \mathbf{y}, \mathbf{y'}) = \frac{P(\mathbf{x}, \mathbf{y'}) - U(\mathbf{x})}{P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})} \in [0, 1] \tag{7}$$

$$\Rightarrow W'(\mathbf{x}, \mathbf{y})P(\mathbf{x}, \mathbf{y'}) = [1 - \sigma(\mathbf{x}, \mathbf{y}, \mathbf{y'})]U(\mathbf{x}) + \sigma(\mathbf{x}, \mathbf{y}, \mathbf{y'})P(\mathbf{x}, \mathbf{y}) \tag{8}$$

where W' is the value of their new contract.

Production function and Flow Disutility

Whenever there is a skill mismatch, a worker might be underqualified or overqualified across either interpersonal or cognitive skills. The following production function specification accounts for both cases

$$f(\mathbf{x}, \mathbf{y}) = x_{i_{s(0)}} \times \left(\sum_{l \in \{s(1), s(2)\}} \alpha_l y_{j_l} + k_l^u \min\{m_l(i, j, t), 0\} \right) \tag{9}$$

Here, α_l denotes the return to the firm for completing tasks that are required by y_{j_l}. $k_l^u > 0$ is used to account for output loss when a worker is underqualified. In other words, in such case the mismatch $m_l(i, j, t) = x_{i_s}(t) - y_{j_s} < 0$ so the term $k_l^u \min\{m_l(i, j, t), 0\}$ decreases the

[15]This carries an implicit assumption that general skills do not grow logistically as the other skills. Rather it grows exponentially $x_{i_{s(0)}}(t) = x_{i_{s(0)}}(0) \times e^{gt}$ This is done to capture wage/experience trend

output generated by the first term $\alpha_l y_{j_l}$. On the other hand, when a worker is overqualified in skill dimension l, $m_l(i,j,t) = x_{i_s}(t) - y_{j_s} > 0$, so the second term becomes zero and there is no production loss[16]. The general skill term $x_{i_{s(0)}}$ is used to scale the output generated.

Similarly, the flow disutility of working is specified in the following specification.

$$c(\mathbf{x},\mathbf{y}) = \sum_{l\in\{s(1),s(2)\}} k_i^o \max\{m_l(i,j,t),0\} \tag{10}$$

Here, if the worker is overqualified, $m_l(i,j,t) > 0$, so the amount of disutility from a skill is proportional to the amount of mismatch scaled by $k_i^o > 0$. It is important to note that being underqualified does not cause disutility.

Value functions and Wage Equation

To tie the private values, learning, production function, disutility and wages, consider the following sets of specifications.

Total Private Value of Match and Unemployment:

$$(r+\mu+\delta)P(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x}) + g(\mathbf{x},\mathbf{y}).\nabla_{\mathbf{x}} P(\mathbf{x},\mathbf{y}) \tag{11}$$

$$(r+\mu)U(\mathbf{x}) = b(\mathbf{x}) + g(\mathbf{x},0).\nabla U(\mathbf{x}) \tag{12}$$

Equation (6) implies that friction-adjusted total private value is equal to the production minus the disutility for the worker, in addition to the possibility of being forced into unemployment $\delta U(\mathbf{x})$ and the learning that occurs. However, during unemployment, equation (7), the worker not only gets their unemployment income $b(x)$ but also sustains a loss of their human capital.

[16] Notice here that overqualified does not mean that a worker will produce additional output

Wage Equation: The wage equation is the solution to

$$
\begin{aligned}
(r+\sigma+\mu)W(\mathbf{x},\mathbf{y},\sigma) = {} & w(\mathbf{x},\mathbf{y},\sigma) - c(\mathbf{x},\mathbf{y}) + \delta U(\mathbf{x},\mathbf{y}) \\
& + \lambda_1 \mathbb{E}\max[0, \min[P(\mathbf{x},\mathbf{y'}) - P(\mathbf{x},\mathbf{y})] - W(\mathbf{x},\mathbf{y},\sigma)] \\
& + g(\boldsymbol{\pi},\mathbf{x},\mathbf{y}) \cdot \nabla_x W(\mathbf{x},\mathbf{y},\sigma)
\end{aligned}
\tag{13}
$$

Equation (8) implies that workers value their wage contract using five things: actual wage, the disutility of work, the possibility of being forced into unemployment, the value and rate of outside offers they will receive, and their on-the-job learning. Now, combining (3), (5), (6), (7), and (8), we get the following solution to the wage equation.

$$
\begin{aligned}
w(\mathbf{x},\mathbf{y},\sigma) = {} & \sigma f(\mathbf{x},\mathbf{y}) + [1-\sigma](b(\mathbf{x}) + c(\mathbf{x},\mathbf{y})) \\
& - \lambda_1 \mathbb{E}\max\Big\{0, \min[0, P(\mathbf{x},\mathbf{y'}) - P(\mathbf{x},\mathbf{y})] + (1-\sigma)(P(\mathbf{x},\mathbf{y}) - U(\mathbf{x}))\Big\} \\
& - (1-\sigma)(g(\mathbf{x},\mathbf{y}) - g(\mathbf{x},0)) \cdot \nabla U(\mathbf{x})
\end{aligned}
\tag{14}
$$

Here, the first term of two terms with σ denotes the static sharing of surplus based on the worker's prior outside offers. The third term with the expectation reflects the possibility of future outside offers. Finally, the last term of the on-the-job learning the worker will accrue while working for a firm $\mathbf{y}$.

Total Matched Surplus

We can now use equations (1)-(9) to form the following solution for the surplus. The derivation is relegated the next section. It is important for several reasons, including counterfactual analysis, the most important of which is identified as explained below.

$$P(\mathbf{x}(t), \mathbf{y}) - U(\mathbf{x}(t)) = x_{i_{s(0)}} \times \sum_{l \in \{s(1), s(2)\}} \Bigg(\frac{\alpha_i y_l - b x_{i_l}}{r + \mu + \delta} - \Bigg[k_l^u \min\Big\{0, A_l \sum_{n=0}^{\infty} \frac{D_n}{(r + \mu + \delta)^{n+1}} \Big\} + k_l^o \max\Big\{0, A_l \sum_{n=0}^{\infty} \frac{D_n}{(r + \mu + \delta)^{n+1}} \Big\} \Bigg] \Bigg) \tag{15}$$

where $D_n = \sum_{L=0}^{n} (-1)^{n+L} \cdot L! \cdot S_{n,L} \cdot (\pi^*)^n \cdot \frac{C^L}{(C+1)^{L+1}}$, $S_{n,L}$ is the Sterling number of the second kind and $C = \frac{A_i - m_{it_1}}{m_{it_1}}$.

The left-hand side of the equation indicates the total surplus gained when a worker $\mathbf{x}$ is matched to a firm $\mathbf{y}$. The first term on the right-hand side, $x_{i_{s(0)}}$, refers to a multiplier for general skills. The first term of the summation denotes the friction-discounted surplus gained when a worker is perfectly matched to an employer. The next term is the value lost when a worker is underqualified, and finally, the last term is the value lost when they are overqualified.

Note that when a worker comes out of unemployment, the only $\mathbf{y}$ they accept is in $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) \geq U(\mathbf{x})\}$ and the boundary set is $\{\mathbf{y} : P(\mathbf{x}, \mathbf{y}) = U(\mathbf{x})\}$. We can then use this boundary set to identify many of the parameters in the surplus solution because on the left-hand side, the value for over-qualification and under-qualification becomes zero. The full identification is given in at the end of appendix A.

Closed Form Solution of Surplus Function

Take the expression of $P(\mathbf{x}, \mathbf{y})$ and parameterize P and x as a function of worker's tenure t in a specific job y. First order partial different equations of (6) can be used to defined the following system of $K+1$ characteristic equations:

$$\frac{dx_k}{dt} = g_k(\mathbf{x}(t), \mathbf{y}) \text{ with } k = 1, ..., K \tag{16}$$
$$\frac{dz}{dt} = (r+\mu+\delta)z - [f(\mathbf{x}(t), \mathbf{y}) - c(\mathbf{x}(t), \mathbf{y})] - \delta U(\mathbf{x(t)}) \tag{17}$$

Under this formulation, $P(\mathbf{x}, \mathbf{y}) = z(t)$ so it can be defined as the following.

$$P(x(t), y) = \int_t^\infty [f(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y}) - c(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y}) + \delta U(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y}))]e^{-(r+\mu+\delta)(s-t)}ds \tag{18}$$

We can similarly find the following

$$U(\mathbf{x}(t)) = \int_t^\infty b(\mathbf{X}(s; \mathbf{x}(t), 0))e^{-(r+\mu+\delta)(s-t)}ds \tag{19}$$

Using (15) and (16) we can get the surplus function as

$$P(x(t), y) - U(x(t)) \tag{20}$$
$$= \int_t^\infty [f(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y}) - c(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y})) - b(\mathbf{X}(s; \mathbf{x}(t), \mathbf{y}))]e^{-(r+\mu+\delta)(s-t)}ds \tag{21}$$
$$= \int_t^\infty \Bigg[\sum_{l \in \{s(1), s(2)\}} (\alpha_l y_l - k_l^u \min\{m_{lt}, 0\} - k_l^o \max\{m_{lt}, 0\} - bx_l) \Bigg] e^{-(r+\mu+\delta)(s-t)}ds \tag{22}$$
$$= \int_t^\infty \Bigg[\sum_{l \in \{s(1), s(2)\}} (\alpha_l y_l - bx_l) - k_l^{u/o}(m_{lt}) \Bigg] e^{-(r+\mu+\delta)(s-t)}ds \tag{23}$$

We can now look at both elements of (18) individually

$$
\begin{aligned}
M_1 &:= \int_t^{\infty} \sum_{l \in \{s(1), s(2)\}} (\alpha_l y_l - b x_l) e^{-(r+\mu+\delta)(s-t)} ds \\
&= \sum_{l \in \{s(1), s(2)\}} (\alpha_l y_l - b x_l) \int_t^{\infty} e^{-(r+\mu+\delta)(s-t)} ds \\
&= \frac{1}{r+\mu+\delta} \sum_{l \in \{s(1), s(2)\}} (\alpha_l y_l - b x_l)
\end{aligned}
$$

$$
\begin{aligned}
M_2 &:= \int_t^{\infty} \sum_{l \in \{s(1), s(2)\}} k_l^{u/o} (m_{lt}) e^{-(r+\mu+\delta)(s-t)} ds \\
&= \sum_{l \in \{s(1), s(2)\})} k_l^{u/o} \int_t^{\infty} m_{lt} e^{-(r+\mu+\delta)(s-t)} ds
\end{aligned}
$$

Here, we know the following by the definition of mismatch. If a worker is overqualified $m_{lt} > 0$ and when they are underqualified $m_{lt} < 0$.

$$
\begin{aligned}
m_{lt} = y_l - x_{lt} = \frac{m_{lt} A_l e^{(s-t)\pi}}{(A_l - m_{lt}) + m_{lt} e^{(s-t)\pi}} &= A_l u(x) \\
u(x) := \frac{m_{lt} e^{(s-t)\pi}}{(A_l - m_{lt}) + m_{lt} e^{(s-t)\pi}} &= \frac{1}{C e^(- x\pi) + 1} \\
C := \frac{A_l - m_{lt}}{m_{lt}} \text{ and } x := s - t &
\end{aligned}
$$

Now let $u_d(t) = Ce^{-x\pi} + 1$ and applying Foa' di Bruno's formula ,

$$\begin{aligned}
\frac{d^n}{dx^n}u(x) &= \frac{d^n}{dx^n}\frac{1}{u_d(t)} \\
&= \sum_{k=0}^{n}(-1)^k \cdot k! \cdot u_d(t)_{n,k}^{-(k+1)}(u_d'(t), u_d''(t), ..., u_d^{(n-k+1)}(t)) \\
&= \sum_{k=0}^{n}(-1)^k \cdot k! \cdot u_d(t)^{-(k+1)} \cdot (-1)^n C^k S_{n,k}\pi^n e^{-k\pi x} \\
&= \sum_{k=0}^{n}(-1)^k \cdot k! \cdot S_{n,k}(\pi)^n(1 + Ce^{-x\pi})^{-k-1}C^k e^{-kx\pi} \\
&= \sum_{k=0}^{n}(-1)^k \cdot k! \cdot S_{n,k}(\pi)^n u(x)(1 - u(x))^k
\end{aligned}$$

Here, $B_{n,k}(.)$ is the Bell Polynomial and $S_{n,k}$ is the sterling number of the second kind. Now we can approximate $u(x)$ at $x = 0$ using Taylor Polynomial to get the following

$$\begin{aligned}
u(x) &= \frac{1}{1 + e^{-x\pi}} = \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n = \sum_{n=0}^{\infty}\frac{D_n}{n!}x^n \\
D_n &:= \sum_{k=0}^{n}(-1)^k \cdot k! \cdot S_{n,k}(\pi)^n u(0)(1 - u(0))^k = \sum_{k=0}^{n}(-1)^{n+k}k!S_{n,k}(\pi)^n\frac{C^k}{(C+1)^{k+1}}
\end{aligned}$$

Putting this together, we can finally solve M_2,

$$\begin{aligned}
M_2 &= \sum_{l\in\{s(1),s(2)\})} k_l^{u/o}\int_t^{\infty} m_{lt}e^{-(r+\mu+\delta)(s-t)}ds \\
&= \sum_{l\in\{s(1),s(2)\})} k_l^{u/o}A_l\sum_{n=0}^{\infty}\frac{D_n}{n!}\int_0^{\infty}\left[\frac{y}{(r+\mu+\delta)}\right]^n\frac{e^y}{(r+\mu+\delta)}dy \\
&= \sum_{l\in\{s(1),s(2)\})} k_l^{u/o}A_l\sum_{n=0}^{\infty}\frac{D_n}{n!}\frac{1}{(r+\mu+\delta)^{n+1}}\int_0^{\infty}y^n e^{-y}dy \\
&= \sum_{l\in\{s(1),s(2)\})} k_l^{u/o}A_l\sum_{n=0}^{\infty}\frac{D_n}{n!}\frac{1}{(r+\mu+\delta)^{n+1}}\Gamma(n+1) \\
&= \sum_{l\in\{s(1),s(2)\})} k_l^{u/o}A_l\sum_{n=0}^{\infty}\frac{D_n}{(r+\mu+\delta)^{n+1}}
\end{aligned}$$

Plugging M_1 and M_2 in (18), we get

$$
\begin{aligned}
&P(x(t),y) - U(x(t)) \\
=&\frac{1}{r+\mu+\delta} \sum_{l\in\{s(1),s(2)\}} (\alpha_l y_l - b x_l) - \sum_{il\in\{s(1),s(2)\}} k_l^{u/o} A_l \sum_{n=0}^{\infty} \frac{D_n}{(r+\mu+\delta)^(n+1)} \\
=&\sum_{l\in\{s(1),s(2)\}} \Bigg(\frac{\alpha_i y_l - b x_{i_l}}{r+\mu+\delta} - \Bigg[k_l^u \min\Big\{0, A_l \sum_{n=0}^{\infty} \frac{D_n}{(r+\mu+\delta)^{n+1}} \Big\} \\
&+ k_l^o \max\Big\{0, A_l \sum_{n=0}^{\infty} \frac{D_n}{(r+\mu+\delta)^{n+1}} \Big\} \Bigg] \Bigg)
\end{aligned}
$$

Parameter Summary

	Parameter	Identification Strategy
Unobserved Skill Group	S	Factor Model on O*NET skills
Workers Skill	$\mathbf{x}$	NLP on LinkedIn
Production Tech	$\mathbf{y}$	O*NET
Production Function (Productivity)	$\alpha_i, .., \alpha_S$	Wage and O*NET
Production Function (Mismatch cost)	$k_1, .., k_S$	$P(x,y) = U$
Joint Dist of Initial Worker Skill		Obs of Initial Sample
Sampling Dist of Job Attributes	$\Upsilon(\mathbf{y})$	Gaussian Copula
Job Destruction Rate	δ	Sample E2U Rate
Offer Arrival Rate	$\lambda_0, \lambda_1 \sim \Upsilon$	Random from Υ
Attrition, Discount	μ, r	Calibrated

Appendix B. Extracting Skills from text

Prediction of each job for each individual is completed with four data sources: Unlabeled LinkedIn job description, labeled job description, and O*NET job description and O*NET skills. The first is treated as the prediction set, and the former two as the training/testing datasets.

The two training/testing sets are linked using brute force. Each job listed by a person on LinkedIn has a position title and a job description. The total stock of skills (in all jobs) is posted at the end of the profile. The position title matches the ones listed on O*NET using fuzzy matching. Then, the O*NET skills for that title are linked with the LinkedIn title if the skills are listed among the total stock.

Appendix B1. Support Vector Classification

About 8% of the LinkedIn text job descriptions are already labeled with skills. O*NET job titles also have text job descriptions. A Linear Support Vector Machine is trained on these data and used to predict labels for the unlabeled LinkedIn job descriptions. Naive Bayes and Logistic regression were also used, but SVC produced the best results with cross-validation on 1/3 of the Labeled LinkedIn data resulting in 72% prediction accuracy.

Appendix B2. Semi-supervised Latent Semantic Indexing

There might be skills whose descriptions are context-dependent. For example, an Economist's use of the word "data analysis" could differ significantly from that of an Anthropologist. So, the paper tried identifying these skills using a version of Latent Semantic Indexing based on Yu et al. (2005). The latent representation has already been produced, and the only task is cross-validation to find the optimum tuning parameters.

This following subsection explains the use of the algorithm below

Algorithm Summary

Input:

The inputs are Unlabeled LinkedIn Document,$\mathbf{X} \in \mathbb{R}^{N\times M}$ Labeled Document $\mathbf{Y} \in \mathbb{R}^{N\times L}$, Tuning parameter for bias towards input vector $0 \leq \beta \leq 1$, Tuning parameter for stability $\gamma \geq 0$, Dimension of output latent space $K > 0$

Steps:

1. Specify the following inner products

   - $(\mathbf{K}_x)_{i,j} = k_x(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$
   - $(\mathbf{K}_y)_{i,j} = k_y(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{y}_i^T\mathbf{y}_j$
   - $\mathbf{C} = (1-\beta)\mathbf{K}_x + \beta\mathbf{K}_y$

2. Solve the generalized eigenvalue problem:

$$\mathbf{K}_x^2\boldsymbol{\alpha} = \lambda[\mathbf{K}_x\mathbf{C}^{-1}\mathbf{K}_x + \gamma\mathbf{K}_x]\boldsymbol{\alpha}$$

3. Obtain eigenvectors $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_K$ with largest eigenvalues $\lambda_1 \geq ... \geq \lambda_K$

Output :

The output should be the matrix created by the indexing function

$$\psi_j(\mathbf{x}) = \sum_{i=1}^{N}(\boldsymbol{\alpha}_j)_i k_x(\mathbf{x}_i, \mathbf{x}, j = 1, ..., K$$

General Latent Semantic Indexing (LSI) Problem

The general problem minimizes the following *reconstruction error*

$$\min_{\mathbf{A},\mathbf{V}} ||\mathbf{X} - \mathbf{V}\mathbf{A}||^2$$

$$\text{subjected to: } \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$$

Here $\mathbf{X} \in \mathbb{R}^{N \times M}$ is the vectorization of the input text, N is the number of document and M is the total number of words. $\mathbf{V} \in \mathbb{R}^{N \times K}$ is a low-dimensional latent space $\mathbf{A} \in \mathbb{R}^{K \times M}$ with $K \leq M$ is the factor loading. The constraint ensures each columns of $\mathbf{V}$, denoted as the latent variables, to be uncorrelated and has unit variance.

Semi-supervised LSI Problem

I now introduce manually labeled document, $\mathbf{Y} \in \mathbb{R}^{N \times L}$, and factor loading of the manual labels $\mathbf{B} \in \mathbb{R}^{K \times L}$

$$\min_{\mathbf{A},\mathbf{V},\mathbf{B}} (1-\beta)||\mathbf{X} - \mathbf{VA}||^2 + \beta||\mathbf{Y} - \mathbf{VB}||^2$$
$$\text{subjected to: } \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$$

Here, $0 \leq \beta \leq 1$ is the tuning parameter that determined how much the indexing should be biased by the output. The second part of the objective function forces the latent semantic to incorporate manual labels

Appendix B2.4. Semi-supervised LSI Problem: Eigenvector form

Here, instead of solving for $\mathbf{A}, \mathbf{V}, \mathbf{B}$, we can simplifying the optimization by introducing $\mathbf{C}$. It solves the eigenvalue problem $\mathbf{Cv} = \lambda\mathbf{v}$ for K leading eigenvectors of $\mathbf{C} = (1-\beta)\mathbf{X}\mathbf{X}^T + \beta\mathbf{Y}\mathbf{Y}^T$, which is equivalent to

$$\max_{\mathbf{v} \in \mathbb{R}^N} \mathbf{v}^T\mathbf{C}\mathbf{v}$$
$$\text{subjected to: } \mathbf{v}^{\mathrm{T}}\mathbf{v} = 1$$

Here, $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_K]$, $\mathbf{A} = \mathbf{V}^T\mathbf{X}$, $\mathbf{B} = \mathbf{V}^T\mathbf{Y}$. Proof that this produces the same optimum is given in Yu et.al. (2006).

Semi-supervised LSI Considerations

There are at least three thing to consider

1. *Linear Constraint :* The optimization above should be able to work even in the presence of non-labeled data (i.e. it should still be unsupervised). As currently constructed, it will only be to handle texts with a portion of data already labeled.

   If ψ is the linear mapping from the input documents to the latent space, we want the latent space to be reduced from $\mathbf{X}$, not a combination of $\mathbf{X}$ and $\mathbf{Y}$. So, restricting the latent variables (the final labels) as *linear mapping* of $\mathbf{X}$, i.e, $\mathbf{V} = \mathbf{XW}$. Here, $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_k] \in \mathbb{R}^{M \times K}$ is the new restricted latent space. This will ensure indexing does not rely too much on the labels $\mathbf{Y}$. The new optimization problem is

$$\max_{\mathbf{w} \in \mathbb{R}^M} \mathbf{w}^T \mathbf{X}^T \mathbf{C} \mathbf{X} \mathbf{w}$$
$$\text{subjected to: } \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} = 1$$

2. *Over-fitting :*As with other ML methods, we do not want the algorithm to over-fit to the already label data. Since the training set in this case will be small, the algorithm might also not produce stable results

   So, we can introduce the penalty term $\|\mathbf{w}\|^2$ with γ being the tuning parameter

$$\min_{\mathbf{w} \in \mathbb{R}^M} \mathbf{w}^T X^T \mathbf{C}^{-1} \mathbf{X} \mathbf{w} + \gamma \|\mathbf{w}\|^2$$
$$\text{subjected to: } \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} = 1$$

3. *Computational Tractability:* High dimensionality is an important issue as minimization might not be computationally tractable. However, note that the solution is $\mathbf{w} \in \mathbb{R}^M$, where M is the number of words. M could be a very big number and cause

computational issues. We could instead find the dual solution $\boldsymbol{\alpha} \in \mathbb{R}^N$, where $N << M$ is the number of documents

Now consider the *kernel* function k_x to be the inner product in X. Then

- $\mathbf{v} = \mathbf{X}\mathbf{w} = \boldsymbol{\alpha}^T\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T\mathbf{K}_x\boldsymbol{\alpha}$ where $\mathbf{K}_x = \mathbf{X}\mathbf{X}^T$
- $\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w} = \boldsymbol{\alpha}^T\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha} = \boldsymbol{\alpha}^T\mathbf{K}_x\boldsymbol{\alpha}$
- $\mathbf{C} = (1-\beta)\mathbf{K}_x + \beta\mathbf{K}_y$ where $\mathbf{K}_y = \mathbf{Y}\mathbf{Y}^T$

Here $\mathbf{K}_x$ is the $N \times N$ kernel matrix satisfying $(\mathbf{K}_x)_{i,j} = k_x(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$ and $\mathbf{K}_y$ is the $K \times K$ kernel matrix satisfying $(\mathbf{K}_y)_{i,j} = k_x(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{y}_i^T\mathbf{y}_j$. Finally, plugging the values in the optimization problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^N} \boldsymbol{\alpha}^T\mathbf{K}_x\mathbf{C}^{-1}\mathbf{K}_x\boldsymbol{\alpha} + \gamma\boldsymbol{\alpha}^T\mathbf{K}_x\boldsymbol{\alpha}$$
$$\text{subjected to: } \boldsymbol{\alpha}^T\mathbf{K}_x^2\boldsymbol{\alpha} = 1$$

Semi-supervised LSI: Solution

Finally, taking the Lagrangian of the problem and setting the derivative with respect to w to zero and rearranging, we get

$$\mathbf{K}_x^2\boldsymbol{\alpha} = \lambda[\mathbf{K}_x^2\mathbf{C}^{-1}\mathbf{K}_x + \gamma\mathbf{K}_x]\boldsymbol{\alpha}$$

Solving this we can get $\alpha_1, ..., \alpha_K$ eigenvectors with eigenvalues $\lambda_1 \geq ... \geq \lambda_K$ The j-th mapping function ,i.e. the j-th index, is given by

$$\psi_j(\mathbf{x}) = \sum_{i=1}^{N} (\boldsymbol{\alpha}_j)_i k_x(\mathbf{x}_i, \mathbf{x})$$