Mismatch in Online Job Search

Martha E. Gimbel
Indeed Hiring Lab

Tara M. Sinclair[1]
The George Washington University and Indeed Hiring Lab
tsinc@gwu.edu

PRELIMINARY: DO NOT DISTRIBUTE, QUOTE, OR CITE WITHOUT PERMISSION
Public Blog Posts Here:
https://www.hiringlab.org/2018/09/20/us-labor-market-changing/
https://www.hiringlab.org/2019/02/28/tech-smaller-skills-gap/
https://www.hiringlab.org/2019/03/20/healthcare-skills-gap/
https://www.hiringlab.org/2019/05/30/changing-labor-markets-around-world/

This Draft: November 12, 2019
JEL Codes: E24, J11, J21, J24, J40, J62
Keywords: Job search, vacancies, employment, unemployment

Abstract

Much attention has been paid to the potential of new private sector data sources as well as the challenges in using these new sources. In this paper we use data from a large online job search website in a way that can particularly complement existing data on the labor market. We focus on labor market mismatch which is an important measure of the health of the economy but is notoriously hard to measure since it requires detailed information on both employer needs and job seeker characteristics. Mismatch is measured as dissimilarity between the distribution of job seekers across a set of predefined categories and the distribution of job vacancies across the same categories. We produce time series measures of mismatch for the US, a set of English-speaking countries, US states, and select US sectors from January of 2014 through June of 2019. We find that title-level mismatch is substantial, hovering at about 33% for the US in the first half of 2019, but that it has slightly declined as the labor market has tightened. Furthermore, over the same time period the mix of job opportunities has shifted substantially but in a way that has made the distribution of jobs more similar to the distribution of job seekers.

[1] Corresponding author. Adhi Rajaprabhakaran provided excellent research assistance and Indeed's Data Insights Platform team provided essential data support. The authors thank Joelle Abramowitz, Pawel Adrjan, Jim Albrecht, Brendon Bernard, Steven Davis, Jason Faberman, Kyu-Heong Kim, Jed Kolko, Peter Kuhn, Claudio Labanca, Joe McCarthy, Donal McMahon, Christina Patterson, Alicia Sasser-Modestino, Callam Pickering, Bryan Stuart, Arthur Turrell, Christiaan Visser, Larry Warren, Abigail Wozniak, and participants at the 2018 Chicago Fed-Upjohn Institute Workshop on Job Search and Vacancies, the IDSC of IZA Workshop: Matching Workers and Jobs Online - New Developments and Opportunities for Social Science and Practice, the Rand seminar, the Department of Economics, Monash Business School Applied Microeconomics Seminar, the First DC Search and Matching Workshop, the Western Economic Association meetings, the 4th Workshop on Spatial Dimensions of the Labour Market, and the monetary policy conference at the Federal Reserve Board "Nontraditional Data, Machine Learning, and Natural Language Processing in Macroeconomics." for helpful comments and suggestions. Both authors of this paper had an employment relationship with Indeed when involved in the research.

Introduction

Public debate keeps returning to the issue of whether or not there are structural problems in the labor market in terms of a mismatch between the background, skills, and/or interests of job seekers as compared to the needs perceived by employers. The "skills gap" or "talent shortage" conversation often relies on anecdotes because it can be hard to collect data at a sufficiently detailed level to appropriately quantify mismatch. Previous research has provided measures based on connecting data from a variety of different sources with varying levels of detail. Online labor market data provides the potential for new insights based on a single source of rich data on both vacancies and job seekers.

The mismatch index is designed to measure the level of mismatch, or dissimilarity, in the economy. It compares the number of job seekers in a job category to the number of vacancies in the same category. Mismatch can arise because there are too few or too many job seekers in a particular category relative to the number of job opportunities. Importantly, our measure of mismatch is relative to the overall availability of job seekers and vacancies. Thus we are focused here on the mismatch across categories rather than movements in the aggregate job seeker to vacancy ratio which might be affected by changes in the aggregate use of online job search platforms in general and/or the market share of a particular platform.[2]

We produce monthly mismatch measures for the US, a set of English-speaking countries, US states, and select US sectors from January of 2014 through June of 2019. Our main finding is that mismatch has declined as the economy has improved. This decline has been driven primarily by a return of jobs to bring the distribution of jobs more in line with the distribution of job seekers.

Our analysis is closely related to Şahin et al. (2011 and 2014) and Lazear and Spletzer (2012a, 2012b) who also quantify the level of mismatch in the economy. They use publicly available data from BLS (JOLTS and CPS) and measure mismatch based on industry categories. They also use vacancy data from the Conference Board's Help Wanted Online (HWOL) Index to construct mismatch measures for a set of occupation categories.[3] Other research, such as Burke et al. (2018), uses job postings data aggregated by Burning Glass Technologies for vacancy information. Marinescu and Rathelot (2018) use data from job board CareerBuilder.com to estimate the role of geographic mismatch and find that it plays a minor role in explaining aggregate unemployment.

There has also been substantial research on mismatch outside the US and particularly in the UK. Turrell et al. (2018) use data from Reed, an online recruiter in the UK, to estimate mismatch by occupation and geography in the UK. They find that it is regional mismatch rather than

[2] In the appendix, however, we discuss relative tightness for a set of sectors to explore mismatch across broad sectors versus within sectors.

[3] Modestino (2010) also uses HWOL data for vacancies and BLS data for workers for her study focusing on mismatch of educational requirements.

occupational mismatch that affects UK productivity. Patterson et al. (2016) and Smith (2012) use data from the UK government employment agency JobCentre Plus to construct estimates of mismatch with Patterson et al. finding that occupational mismatch is an important contributor to weak productivity growth in the UK and Smith finding that occupational mismatch has had a substantial impact on UK unemployment rates.

Şahin et al. (2014) focus on measuring "mismatch unemployment", i.e. the share of unemployment due to sectoral mismatch. For their occupation-level analysis they report results using 22 of the 23 major (two-digit) SOC groups and 36 of 96 minor (three-digit) SOC groups. In the working paper version, Şahin et al. (2011) use the same mismatch formula we use here for a benchmark measure with no heterogeneity across markets. They consider all 17 industries where JOLTS vacancy data are available.[4] They conclude that mismatch explains up to one third of the increase in the unemployment rate during the Great Recession.

Lazear and Spletzer (2012a, 2012b) used a measure of mismatch as part of a broader set of indicators on the recent performance of the US labor market. In terms of mismatch they focused on their finding that mismatch rose in the recession and then declined afterwards suggesting a cyclical rather than structural pattern.

In this paper we present a set of mismatch indexes that we compare across regions including US states and English-speaking countries (the US, the UK, Australia, Canadam Ireland, New Zealand, and Singapore). Similar to Lazear and Spletzer, we are particularly interested in what the patterns in our mismatch measures over time tell us about how different types of mismatch are related to changes in economic conditions. With our unique dataset we can focus on a range of different levels of disaggregation to create different measures of mismatch in terms of geography, sector, and job seeker characteristics.

For example, we include both employed and unemployed job seekers in our benchmark series. Including employed job seekers which has been challenging in previous analyses due to limited data availability on people searching on the job.[5] There is debate about how similar employed and unemployed job seekers are and what impact that might have on economic outcomes. On the one hand, Ahn and Hamilton (2016) argue that the unemployed differ in terms of relevant unobservables for job finding that vary over time and Longhi and Taylor (2014), using UK data, find that the unemployed and employed are quite different and that the differences vary over the

---

[4] The 17 industries used by Şahin et al. are: arts, construction, mining, accommodations, retail, professional business services, real estate, wholesale, other, transportation and utilities, manufacturing - nondurables, education, health, government, manufacturing - durables, finance, and information. The 12 industries we use in our analysis are: construction, durable goods manufacturing, nondurable goods manufacturing, wholesale and retail trade, transportation and utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality, other services, and government. Lazear and Spletzer use 12 industries but differ from ours by including mining but grouping together durable and nondurable goods manufacturing. We exclude mining due to different definitions between JOLTS and CPS. Results are little changed between the different choices of Lazear and Spletzer, Şahin et al. or our analysis.

[5] Şahin et al. (2014) did provide an estimate of their measure including on-the-job search. They used the American Time Use Survey to identify employed job seekers. This survey likely underestimates the number of employed job seekers as discussed in Faberman et al. (2017).

business cycle. On the other hand, Kroft et al. (2016) find that "shifts in observable characteristics of the unemployed do not go very far in accounting for the rise in long-term unemployment." Most related to our analysis, Şahin et al. (2014) see little difference when adding in employed job seekers based on time use surveys into their measure of mismatch.[6]

In addition to mismatch, we also produce measures of vacancy dissimilarity over time as well as job seeker dissimilarity over time. Comparing the distribution of job opportunities today to what was available in the past and doing the same for job seekers gives us a measure of how much the labor market has shifted over time from both the labor supply and labor demand dimensions. This is particularly important given one of our key findings for the US is that mismatch is declining somewhat over our sample period. At the same time, we find substantial change in the distribution of both vacancies and job seekers over this period, so the slightly declining mismatch suggests that jobs and job seekers are becoming more similar to each other as the economy has improved.

In the following sections we describe our data and mismatch methodology, then we report our benchmark measure of overall online labor market mismatch for the US. We find that mismatch has not increased as the labor market has tightened and also show that the distribution of jobs has changed substantially over this time period. The changes in the distribution of jobs and resumes have overall drawn job seekers and employers closer together over the sample. We also provide results for a set of sectors as well as cross-country an cross-state comparisons. We then conclude with a discussion of future work.

Data

This draft is focused primarily on the US, but we also include analysis for the UK, Ireland, Australia, New Zealand, Singapore, and Canada. Our main data source is online job postings and job seekers from Indeed, the largest job site in the world based on unique visitors according to ComScore, an independent analytics firm.[7] For comparison we also use publicly available data from the Current Population Survey (CPS) and the Job Openings and Labor Turnover Survey

---

[6] In preliminary analysis where we infer whether or not a job seeker was currently employed, we find the same trend whether we limit to just unemployed or also include the employed.

[7] Globally Indeed has 250 million unique visitors per month, (Google Analytics, Unique Visitors, September 2018) and is the #1 job site worldwide according to comScore total visits (March 2018). Indeed has 55.4 million unique visitors per month in the US (comScore, November 2018) which makes Indeed the #1 ranked job site by unique visitors in the US. Furthermore, in July of 2018, comScore estimated that 75% of US online job seekers search for jobs on Indeed (per month). Indeed Canada has 8.1 million unique visitors per month (comScore, November 2018) which makes Indeed the #1 ranked job site by unique visitors in Canada. As of December, 2018, Indeed has 150 million resumes worldwide. As of December, 2018, Indeed has 58 million resumes in Canada with 290,000 added or updated every month. According to SimilarWeb in December 2018 the Indeed UK site had 35.9 million total visits making it the #1 ranked job site in the UK. As of December, 2018, Indeed has 12.5 million resumes ("CVs") in the UK with 720,000 added or updated every month. Indeed Australia is the #2 ranked job site in Australia with 8.7 million total visits in December of 2018 per SimilarWeb. As of December, 2018, Indeed has 2.7 million total resumes for Australia, with 100,000 added or updated every month.

(JOLTS).[8] We focus on seasonally unadjusted data from all sources. Our measure of mismatch will be in shares of totals which should net out any common seasonal patterns and will leave only job category seasonal patterns which we are interested in examining.

Our measure of job openings will either be from JOLTS by industry, where we focus on the 12 industries where we can match with data available from the Bureau of Labor Statistics on the industry of the unemployed, or from job postings aggregated by Indeed from across the internet. [9] The Indeed postings number for each month is the average daily postings visible on Indeed in that category for that month. We also considered job postings visible on the last business day of the month to line up with the definition from JOLTS, but found that it was typically similar to average daily postings and using the average daily posting number smoothed out any single-day effects. We also compared all visible postings to only those from employer websites (excluding job boards whose visibility on Indeed has varied over time) and found the results to be similar.

It is important to note that we are not restricted to advertisers on Indeed. Instead they collect job postings anywhere on the internet and de-duplicate them as part of their business. Indeed is a generalist site in the sense that they focus on providing "all jobs" not a niche market.

Our measure of job seekers will either be the unemployed from the CPS or active job seekers on Indeed. In our analysis we are focused on the job seekers who have accounts and have uploaded resumes to provide further detailed background information. Indeed has 64.7 million resumes from the US as of June of 2019. We are focusing on the subset that were active accounts during our sample from 2014 through June of 2019, where active is defined as having last updated their resume on Indeed in that month.[10] We aggregate to the monthly frequency, but we could look at daily or even intra-day based on the Indeed data. Higher frequency is interesting when looking at the job seeker data (there are interesting daily and weekly patterns in the job search data), but less so for job postings data.

Job seekers are not just the unemployed.[11] In fact, it appears that the majority of job seekers on Indeed are employed based on reported employment status by account holders as well as

[8] The job openings data are from the September 10, 2019, release of JOLTS. The unemployed by industry data are from the CPS. The data are not seasonally adjusted, and using the 12 industries available from both CPS and JOLTS: construction, durable goods manufacturing, nondurable goods manufacturing, wholesale and retail trade, transportation and utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality, other services, and government. Note that we exclude mining due to different definitions between JOLTS and CPS (although including it does not give noticeably different results).

[9] Şahin et al. (2011, 2014) and Lazear and Spletzer (2012a and b) also each produce measures of *occupational* mismatch using Help Wanted Online Index (HWOL) data as their measure of vacancies for a subset of standard occupation categories (since only industry groupings are available from JOLTS). The HWOL data by occupation is not publicly available and thus we focus on the industry mismatch as our comparison. Canon et al (2013) provide a review of mismatch indexes using HWOL job vacancy data.

[10] Indeed only saves the latest version of resumes, so we only count each resume one time based on latest update date, since the last job title from the resume is key to our analysis. We recognize this might cause a bias in the analysis if there is a systematic pattern in who updates resumes frequently and/or who was a job seeker on Indeed early in our sample and again later in our sample. We address this further in the robustness checks section.

[11] We're only looking at active job seekers, so they are either employed or unemployed, there is no "out of the labor force" group in our analysis.

reported in internal surveys. This is consistent with the finding by Faberman et al. (2017) that employed job seeking is "pervasive." We identify labor market status in the Indeed data based on information reported by the user. Users opt-in to being counted as employed by checking a box indicating that they are currently employed at one of the positions listed on their resume. There is likely measurement error in this as some employed workers may not select the box and others may try to hide that they are unemployed by selecting the box or by not updating that information if they leave their employer but continue searching for a job on Indeed. We include only the "experienced unemployed" for our resume data because we are only using resumes that have previous employment recorded. This is consistent with the BLS data where an industry is only available for people who were previously employed. For our clicks analysis, however, the clicks can come from any job seeker and we do not observe their current employment status.

In the online labor market data we have much finer job type groupings than what is available in the data used in previous research: for our benchmark measure we include 6068 normalized title pairs per month in our analysis as compared to the 9 to 36 categories used by Lazear and Spletzer (2012b) and Şahin et al. (2014). For example, "registered nurse" is a normalized title that contains: Registered Nurse, RN, RN Staff Nurse, Registered Nurse (RN), Registered Nurse - RN, Registered Nurse Traveler, etc. "Economist" is a normalized title that contains: economist, health economist, principal economist, chief economist, associate economist, lead economist, and so on. The 6068 titles were determined as the superset of English normalized titles across the countries in this study: the UK, the US, Canada, Australia, Ireland, New Zealand, and Singapore. For some titles the counts for both resumes and postings are zero in most or all months for one or more countries which does not meaningfully affect our analysis. We also estimated a version excluding low observation categories with no meaningful impact on the estimates. We organize our analysis around job titles for a number of reasons: 1) titles are relatively easy to standardize across resumes and job postings and across countries 2) titles capture skills more consistently that what is reported by job seekers in resumes 3) employment background provides a blend of interest and skills to better connect with where a job seeker will likely go than just a narrow classification of job seekers by skills alone 4) titles allow us to get quite granular as compared to industries or occupations.

For robustness, we also use an alternative measure of job seekers based on clicks on job postings. A job seeker can only click on a posting if one is available and the click may not indicate the job seeker is qualified, only interested in the role.

Methodology

The mismatch measure is the Duncan and Duncan (1955) dissimilarity index. With this measure we assume that only the job seekers can change occupation whereas job vacancies are fixed in their category.[12] The Duncan and Duncan measure is:

[12] The Duncan and Duncan measure has come under criticism when applied to occupational gender segregation (Watts 1992, 1994, 1998). An alternative measure, the IP index of Karmel and MacLachlan (1988) is the preferred

$$\frac{1}{2}\sum_i \left| \frac{S_i}{S} - \frac{V_i}{V} \right|, \tag{1}$$

where S_i is the job seekers in category i, S is the total number of job seekers, V_i is the number of vacancies in category i, V is the total number of vacancies.

This is the same measure used by Lazear and Spletzer (2012a and 2012b) and Sahin et al. (2011, before incorporating a matching function). This index can be interpreted as the proportion of job seekers who would need to be moved to make the job seeker to posting ratio the same for all job categories, where a job category in our analysis will either be industry or normalized job title. Other measures of mismatch, notably Şahin et al. (2014), are reported as a fraction of hires lost per period due to job seeker misallocation. Thus our index will likely be much higher in magnitude as a share of job seekers as compared to a share of monthly hires.

Benchmark Results

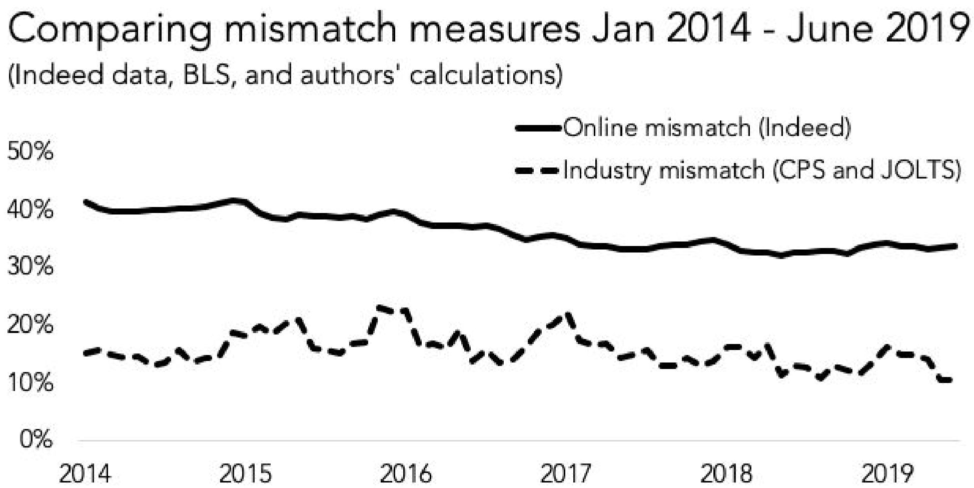
For our measure of mismatch based on online job search we start in January of 2014[13] and currently report through June of 2019. One of the benefits of using the online data is more timely arrival of updated information. As soon as the first week of each month we could update our measures rather than waiting for JOLTS vacancy data which arrives over a month later and then is revised further in the following months when later surveys come in. JOLTS vacancies are further revised annually all the way back to the beginning of the series in December of 2000 to incorporate updates to the Current Employment Statistics employment estimates. Seasonally adjusted data are also revised with updated seasonal factors, but we are using seasonally unadjusted data throughout.

Figure 1 presents our online labor market mismatch estimate along with industry mismatch based on unemployment from the Current Population Survey (CPS) and vacancies from the Job Openings and Labor Turnover Survey (JOLTS) following a similar methodology to that used by Lazeer and Spletzer(2012a and 2012b). Appendix A reports further details on the industry mismatch analysis. Our measure is higher in level, as would be expected given that we're going from 12 industry categories to over 6000. In terms of time pattern, however, they're broadly similar, although our measure is substantially smoother.

measure in that literature. In the gender segregation case, however, both men and women could change occupations, whereas in our analysis we assume only the job seeker can change occupations.

[13] The data from Indeed are only available consistently over time starting in January of 2014.

Figure 1: Online Mismatch and Industry Mismatch

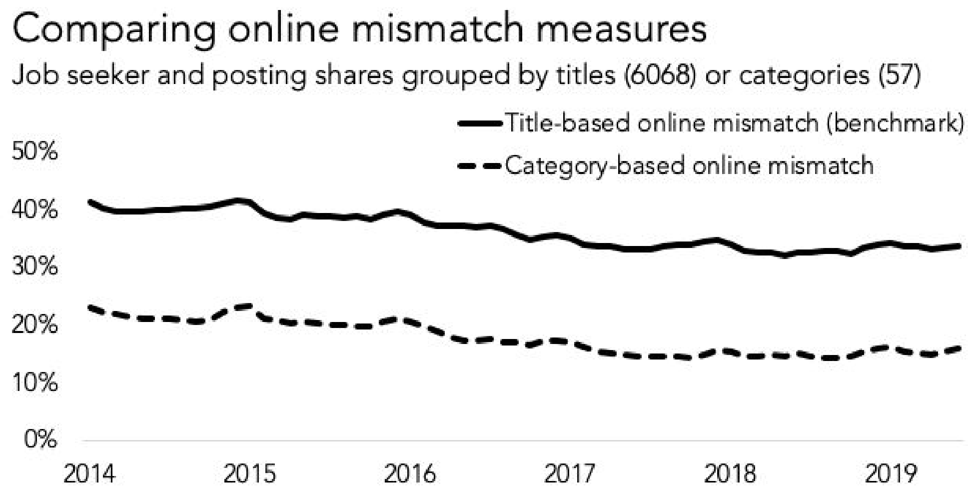


Lazear and Spletzer find much more mismatch by occupation than by industry, which is consistent with what we find for our online labor market mismatch at the normalized job title level. Job titles are much more similar to occupation than to industry. We would also expect that there would be more mismatch at lower levels of aggregation.[14]

We have explored a number of different groupings and our results are consistent with what is expected: grouping the job titles into broader categories (Indeed's proprietary categories) results in a lower level of mismatch overall as seen below in Figure 2, but a similar pattern of slight decline over our time frame. Limiting the analysis to only titles with large numbers of postings and resumes (e.g. the top 700) gives very similar results in both level and slope, which is consistent with how mismatch is measured because it is driven by large categories. It is also similar in terms of smoothness, which suggests it is not the large number of titles that is driving the smoothness of online mismatch as compared to industry mismatch based on publicly available data.

[14] According to Şahin et al. (2014) "...every statement about the role of mismatch should be qualified with respect to the degree of sectoral disaggregation used" (pg. 3538). Comparing across different aggregation approaches (occupation versus industry for example) and/or across different data sets can also shift the level of mismatch. We are focused less on the level of mismatch and more on the pattern in mismatch over time.

Figure 2: Category Mismatch and Title Mismatch



The smoothness of online mismatch may be due to the consistency of the data since our source is a common labor market with as much as possible the same definitions applied to both groups. It does not appear to be sensitive to changes in aggregation level, the particular dissimilarity metric used, or changes in our definition of an active job seeker.[15] It also appears that some of the smoothness comes from including employed job seekers.

Despite the smoothness, we do see clear seasonality in mismatch. This might be expected because we do not use seasonally adjusted data, but it is interesting that the seasonal patterns are sufficiently different in job postings versus job seeker behavior that we see clear rises and falls each year in our mismatch measure.

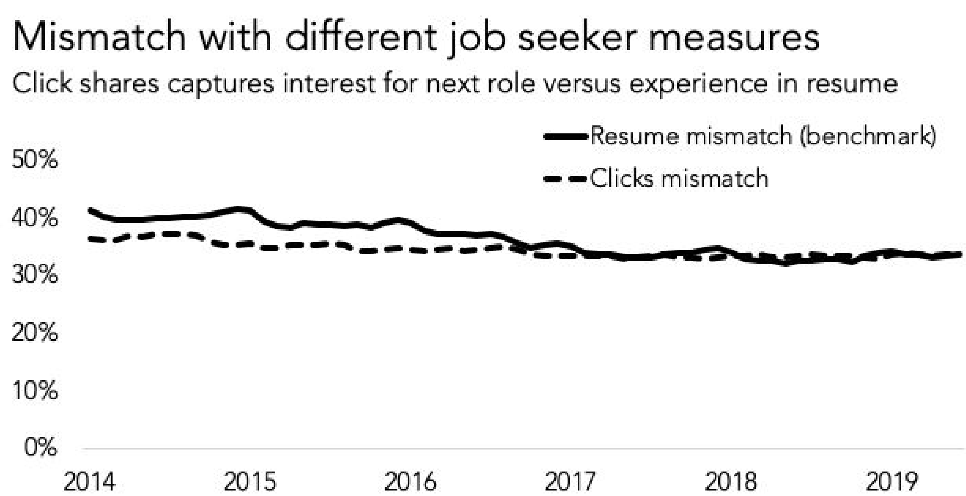
At least three concerns arise from our use of latest job on job seekers' resumes in order to classify them. The first is that job seekers may be aware of the changing landscape of job opportunities and they may be looking for roles different from their current or most recent job title. The second is a concern about the way the resume data is stored that may be affecting our results. Per the terms of Indeed's user agreement, only the latest resume a job seeker has uploaded is kept. That means we lose some of the earlier job seekers in our sample since we count an active job seeker based on the month the resume was last updated. Third, using resume data means we limit the sample to job seekers who have uploaded a resume on Indeed, but many people use the website without uploading a resume. To address these concerns we consider an alternative measure of job seeker distribution based on the job titles job seekers

[15] We change the measure of the job seeker below to interest based on clicks. We also considered an alternative measure of dissimilarity, the Kullback-Leibler (KL) divergence measure (using Bayesian Dirichlet priors, see the recent survey by Yang, 2018, for more details on the KL divergence measure) and find broadly similar results in terms of trend some decline early but broadly flat since 2017 (see appendix for graph).

click on. This allows us to focus on the jobs a job seeker is looking for rather than their experience. The job seekers may not always be qualified for the roles they look at, so the clicks-based measure is more about interest whereas the resume title captures work experience. Another caveat of this measure is that job seekers cannot click on a job if they are not shown the role so the clicks are affected by both job posting availability and the Indeed search algorithm.

Despite the caveats and substantial differences between our two different job seeker measures, the mismatch series created by using the same job posting shares as before and measuring job seeker shares in the two different ways are surprisingly similar. As shown in Figure 3, clicks mismatch is lower than resume mismatch early in the sample, but by 2017 the two measures are very similar. Both show some decline over time, but it is more muted for the clicks measure. This leads us to emphasize "not increasing" rather than clearly declining in interpreting our US results.

Figure 3: Mismatch Counting Job Seekers by their Clicks on Job Postings



Looking into the normalized titles that are the largest contributors to mismatch presented in Table 1, a few features stand out. First, these titles are large categories. This is important to keep in mind for the dissimilarity measure we use - it is based on differences between the shares in the postings and the resumes, so even a large percentage difference in a small category would not result in a large move in overall mismatch. The top ten where the resume share exceeds the posting share contribute 10.9% of mismatch, and the top ten where the posting share exceeds the resume share contribute 10.2% of mismatch. The top contributors to mismatch are also notably persistent with some seasonal patterns. For example, comparing this list to the list for December 2018, we get slightly different ordering but remarkably similar titles with the exception of seasonal associate appearing prominently in the December list for posting share exceeding resume share. Comparing June 2019 mismatch contributors with June 2016 results in substantial overlap with over 50% of the same titles showing up on both the 2019 and 2016 lists.

Table 1: Top Contributors to Online Mismatch

Top contributors to online mismatch

Comparing job seeker resumes and job postings in June 2019 (Indeed data)

Rank	resume share > posting share	posting share > resume share
1	customer service representative	retail sales associate
2	cashier	shift manager
3	customer service associate / cashier	registered nurse
4	server	restaurant manager
5	receptionist	babysitter/nanny
6	warehouse worker	assistant manager
7	laborer	shift leader
8	forklift operator	store manager
9	manager	restaurant staff
10	nursing assistant	general manager

Changing job postings and changing resumes

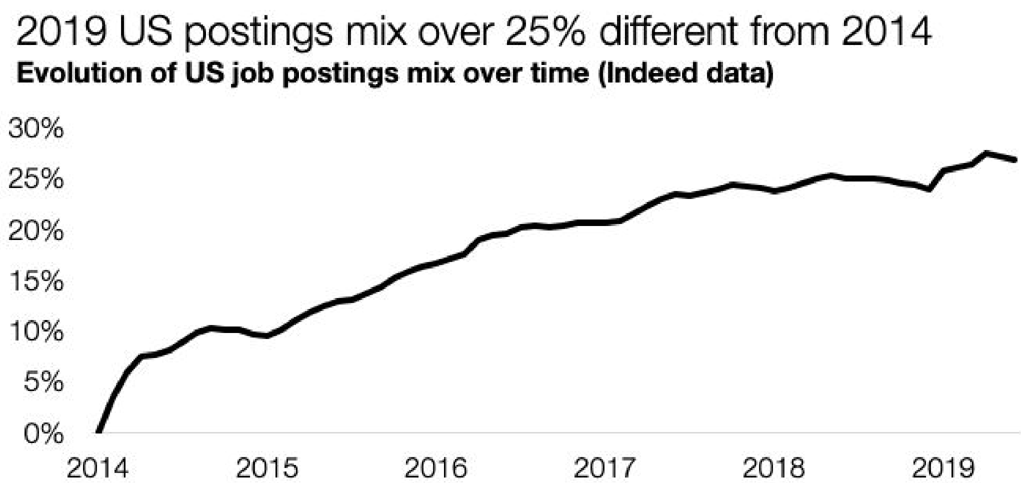
Mismatch could be flat to declining for two reasons: either little is changing underneath or job seekers and jobs opportunities are seeing their distribution across titles change in similar ways over the last several years. To examine this we used the same dissimilarity index but applied it to jobs and resumes separately over time to see how different jobs and resumes are today from what they were in 2014. Thus for each time period *t*, from January 2014 through June of 2019, we constructed the following dissimilarity metric:

$$\frac{1}{2}\sum_i \left| \frac{V_{i,t}}{V_t} - \frac{V_{i,2014m1}}{V_{2014m1}} \right|. \tag{2}$$

We find that the jobs mix has changed substantially over the last few years. The job seeker mix has also changed, although not as dramatically. Overall, as we show below, it is the change of job postings towards job seekers that has brought about the small decline in mismatch over the sample.

First, looking at the distribution of job postings over time: Figure 4 shows that there has been a substantial change in the distribution across titles in job postings over recent years. Comparing January of 2019 with January of 2014 (comparing January to January to exclude potential seasonal differences), 25.8% of job postings in 2019 would need to change in order to have the same distribution as five years before.

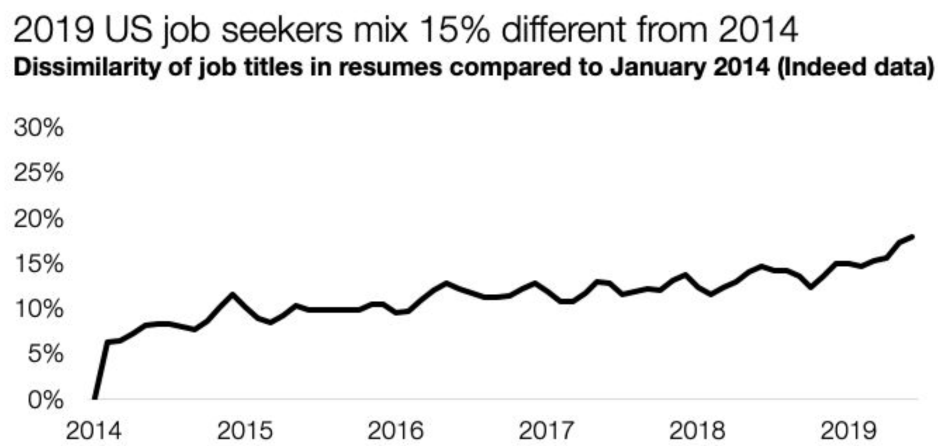
Figure 4: Changing Mix of Job Postings Over Time[16]



Resumes have, however, changed less over the sample than job postings have. Again comparing January of 2019 with January of 2014, resumes are 15.0% different than they were five years before (Figure 5). One data note: because of the nature of Indeed's data, where only the latest resume a job seeker has uploaded is kept, resumes today are less comparable with resumes five years ago than job postings over the same time period.

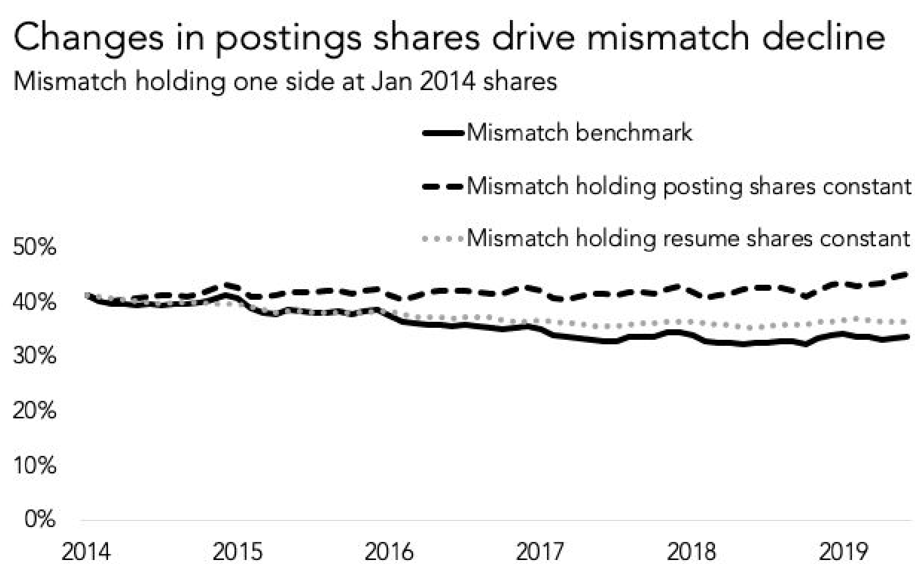
[16] We also considered our alternative dissimilarity measure, KL divergence. The results are consistent across the two measures, with January of 2018 compared to January of 2014 having a KL statistic of 0.23 and a similar trend over time (see appendix).

Figure 5: Changing Mix of Online Resumes over Time



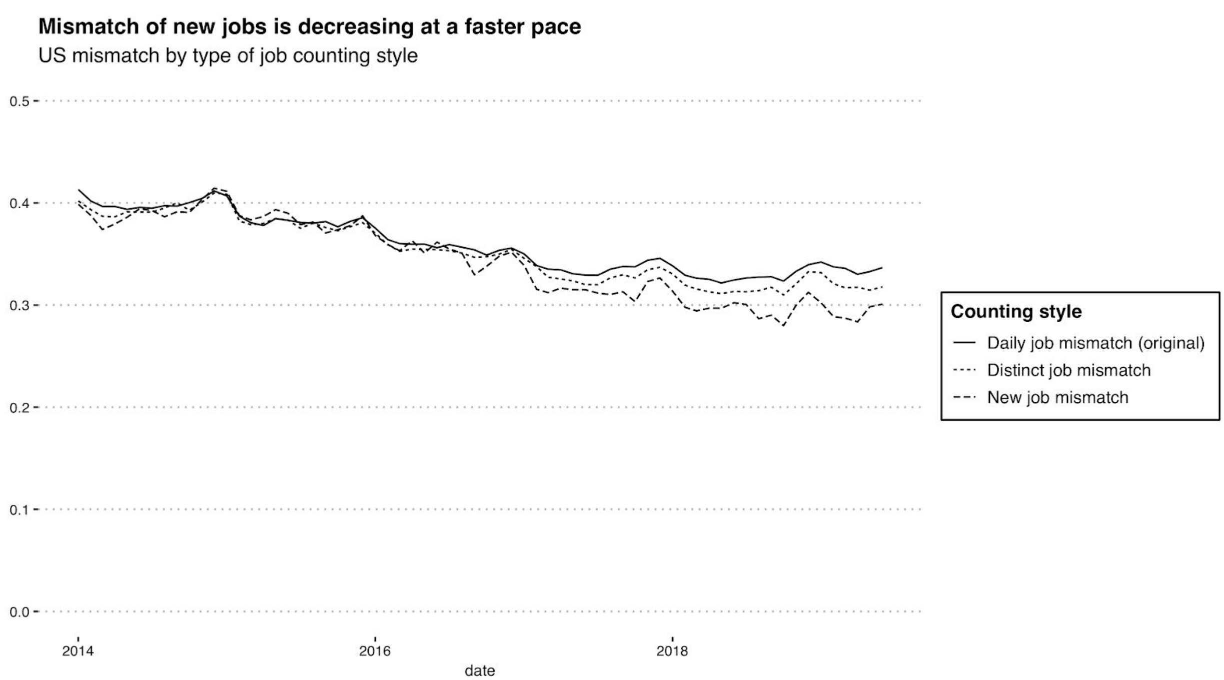
In order to explore the role of the changes in postings and resumes and their contribution to mismatch, we constructed counterfactual mismatch measures where we held the labor supply (resumes) or the labor demand (postings) distribution constant at the shares of the beginning of the sample (January 2014). Figure 6 shows that mismatch would have been a bit higher in 2019 if the resume distribution had not changed, but it's much more dramatic when we hold the postings distribution constant: in that case mismatch would have risen rather than declined over the sample.

Figure 6: Analysis Holding One Side of Mismatch Constant



Since postings shares are particularly important for this analysis, we also considered some different ways to measure them. Our benchmark is the average daily visible jobs in the month since open vacancies on a day follows the official vacancy numbers from JOLTS. Instead of focusing on the number of open vacancies on a specific day (like the last business day of the month which is used for JOLTS), we average over the month to smooth out potential data anomalies such as a major employer website being down on a specific day. But there are other ways to count the number of vacancies and we considered two alternatives: total number of distinct jobs that appeared within a month and the total number of jobs that appeared for the first time in that month. The distinct count might best match up with our definition of active job seeker since we count them once in the month they last updated their resume. Figure 7 below shows our mismatch measure for the different cases. Overall the level and trend are broadly similar across the three measures, but it is interesting to see that mismatch appears to be decreasing more over the sample if we use either of the two alternative measures, and particularly so if we focus only on new job postings.

Figure 7: Alternative measures of mismatch using different vacancy definitions

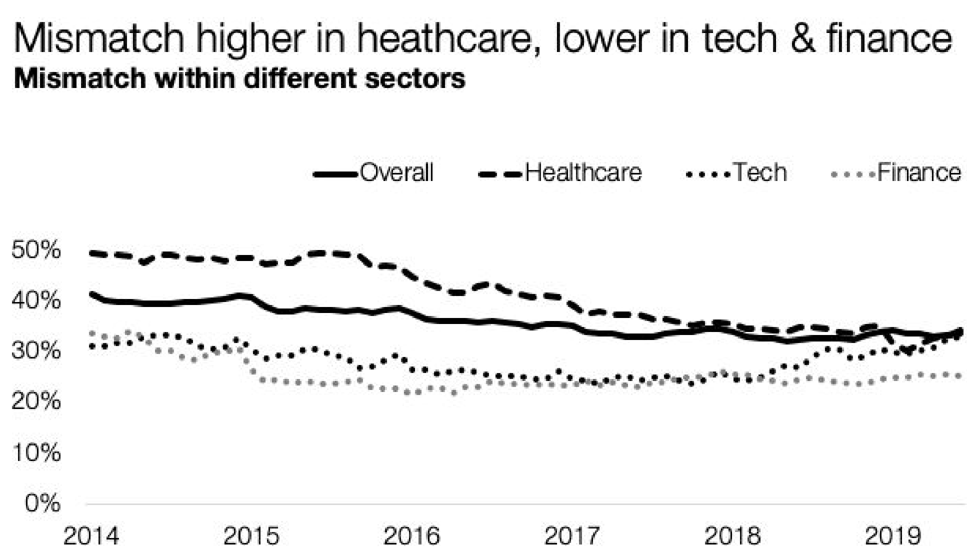


Sector Analysis

We can also explore the question of how well matched the job seekers are within the sector, which we might think of as sort of intensive margin mismatch.[17] For the within sector mismatch we return to our dissimilarity measure and calculate mismatch based on resumes of job seekers currently or most recently employed in that sector and job postings in that sector. Each sector is defined by a set of normalized titles that can clearly be mapped to that sector . Our three sectors are tech (550 titles), healthcare (289 titles), and finance (571 titles). In June of 2019, healthcare was the largest sector with approximately 14% of all US job postings in this sector. Finance had less than 2% and tech had almost 6% of all US job postings.

In Figure 8 we show that for most of the sample healthcare shows greater mismatch than our benchmark overall results for the US and tech and finance are both below. Interestingly in recent months healthcare mismatch has declined and tech mismatch has risen to converge close to the overall national level of mismatch. Finance, however, has stayed flat and well below the national level.

Figure 8: Mismatch for Tech, Healthcare, and Finance Sectors



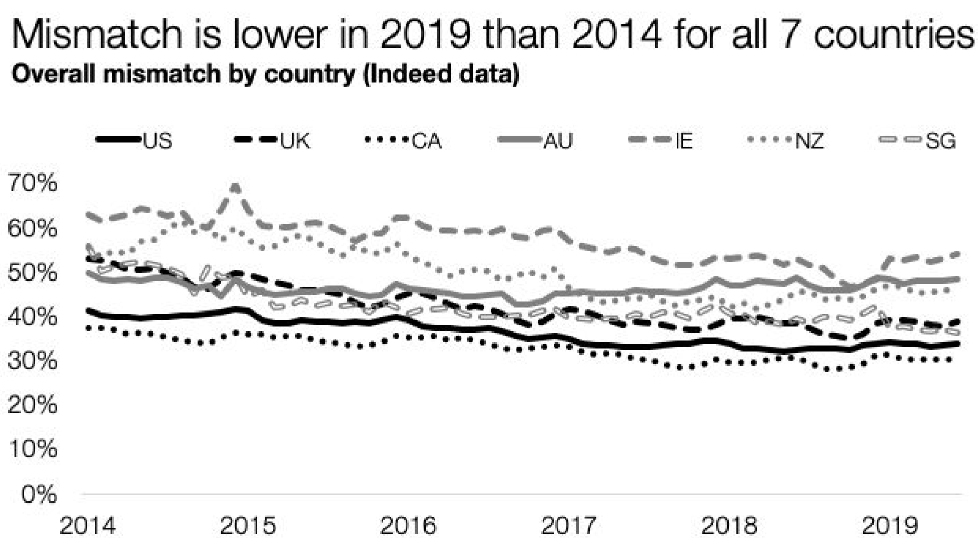
Cross Country Comparisons

For the same set of 6068 normalized titles (selected as the superset of normalized titles across the countries) we construct comparable mismatch measures again monthly from 2014 through

[17] See the appendix for discussion of relative tightness for these sectors. Also see these two blog posts for further discussion of the healthcare and tech results: https://www.hiringlab.org/2019/03/20/healthcare-skills-gap/, https://www.hiringlab.org/2019/02/28/tech-smaller-skills-gap/.

June of 2019 (Figure 9). The countries have slightly different levels and seasonal patterns, but perhaps the most interesting pattern is the trends: all seven of the countries studied: US, UK, Canada, Australia, Ireland, New Zealand, and Singapore. Canada and the US have very similar levels and patterns, with Canada just slightly below the US throughout the sample. Australia is at almost the same level of mismatch at the end of the sample as in 2014. The similarity of the US, US, and Canada is consistent with other labor market indicators for these countries over this time period.[18]

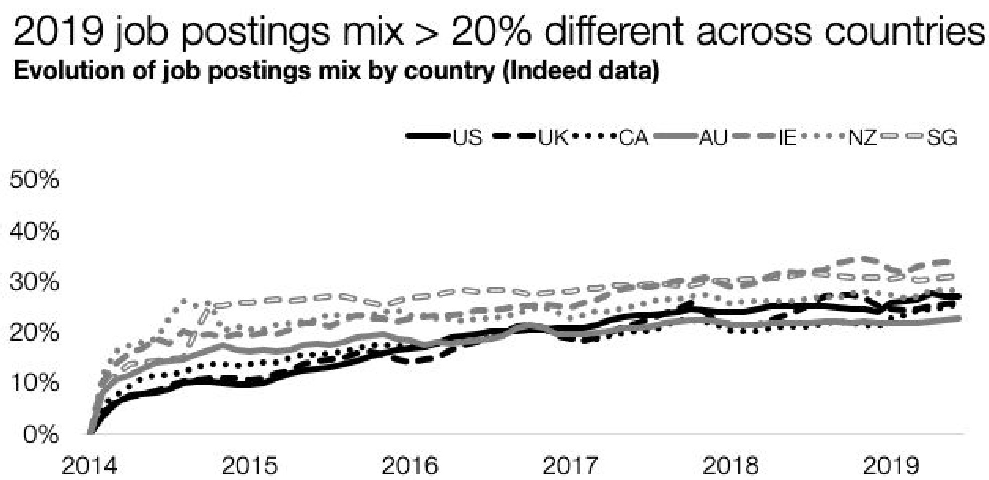
Figure 9: Within Country Mismatch Comparisons



We also constructed the dissimilarity index for job postings over time for each of the countries in our dataset and report the comparison of the results in Figure 10. We see that all the countries have seen a substantial change in the distribution of their mix of job postings between 2014 and 2019, ranging from Australia's 21.7% change to Ireland's 32.3% change (comparing January to January to avoid seasonal differences).

[18] For more analysis of the Canadian and Australian data, see the following blog posts:
https://www.hiringlab.org/en-ca/2019/05/16/labour-market-mismatch-canada/
https://www.hiringlab.org/au/blog/2019/06/28/australias-mismatched-labour-market/

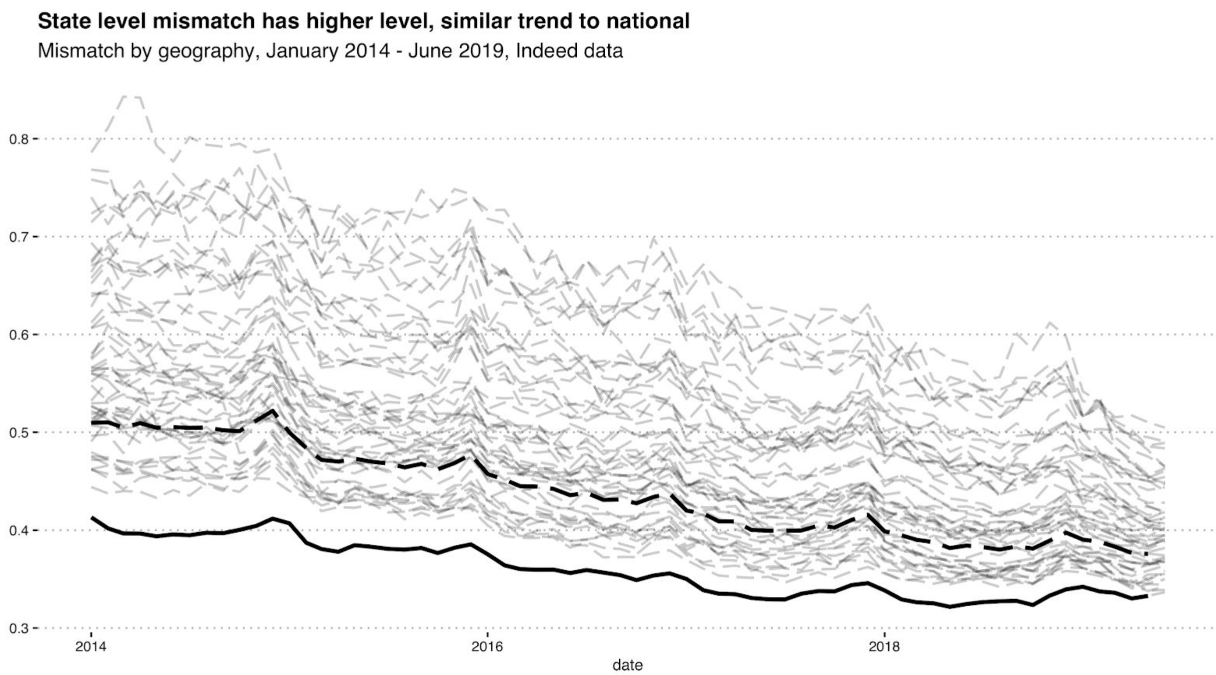
Figure 10: Postings Shares Changes over Time for 7 Countries



Mismatch for US States

Our data also allow us to look at more granular slices, for example by state. Mismatch for the states is higher than national because we are adding a second requirement that the job seekers be in the same state as the job posting (Figure 11). In general the smaller states have higher mismatch and the largest states have lower mismatch, but the pattern doesn't apply strictly. There is movement in the rankings between the beginning and end of the sample. We also see broadly similar patterns in terms of the changes in the distribution of job postings and resumes respectively over time for the states.

Figure 11: Mismatch within US States



Note: Solid black line is national mismatch, dashed black line is weighted average of the states, the dashed gray lines are for the individual states.

Conclusions and Future Work

This version introduced several different measures of online mismatch and showed a robust trend of slight decline over the last several years for the US, across a range of English-speaking countries, and across US states. This decline appears to be driven by the change in the distribution of jobs towards the distribution of job seekers. One interpretation is that jobs came back that were a better fit for job seekers as the global economy continued to improve over the last several years.

In future work we plan to explore several topics including:

Estimating a Natural Rate of Mismatch: With our estimates only available for a recovery period, we have little business cycle variation to estimate what is trend and what is cycle, but based on connecting our results to those of Lazear and Spletzer (2012b) we have a few initial thoughts. We see a slight downward trend in our mismatch which is consistent with the Lazear and Spletzer (2012b) interpretation that mismatch goes down as labor markets improve. We expect there to be more information along these lines as we continue to update the series over time.

The Role of Job Switchers: Our analysis is currently binary: same or not same. One concern about grouping job seekers into categories is that job seekers may not stay in the same category and that skills may be transferable across categories and/or job seekers may develop new skills over time that might lead them to change categories. This may be particularly true of the finer categories we use at the normalized job title level. Furthermore, people may have the skills for jobs, but be uninterested in doing them (interest mismatch as compared to skills mismatch). Hobijn (2012) combined data from the CPS, JOLTS, and state-level job vacancy surveys and found that the "majority of job openings in all industries and occupations are filled with persons who previously did not work in the same industry or occupation." Sinclair (2014) and Flowers (2018) have both examined the behavior of job seekers using Indeed to search for jobs in categories other than their most recent employment and find substantial amount of searching across even very broad categories. They also each document that specialization and pay are both positively related to retention by job type. This analysis suggests we may want to weight by some measure of skills and/or interest overlap for our mismatch index. In that case we may be able to think about the distance between normalized job titles and estimate a smaller amount of mismatch if in "adjacent" job titles by occupation grouping. A related approach was used by Şahin et al. (2014) to allow their unemployed job seekers to search in a new industry/occupation, but they find that the "bulk of unemployed workers keep searching in their previous employment sector" (page 3559) so their estimate of mismatch unemployment is little affected. We can also rank order the normalized titles by estimated average salary to construct a weighted variant of the dissimilarity index called the Earth Mover's Distance (Rubner et. al, 2000; for an application to the labor market see Rim, 2018), or use a measure of occupational distance such as Robinson (2018).

Mismatch for more types of job seekers and more countries: Besides overlapping categories, it may be interesting to zoom in not just on narrower geographies and sectors, but also on mismatch by other features of the job seeker. For example we can look at employment status, long term versus short term unemployed,[19] and age categories. Indeed also has data for over 60 countries with broadly similar data collection and structure, so we would like to build indexes that are comparable across countries, although we'll have to address how to get consistent job titles across languages.

[19] Wiczer (2015) argues that occupation-specific shocks are important for understanding the pattern of unemployment duration over the business cycle.

References

Abraham, Katharine G. "Is skill mismatch impeding US economic recovery?." ILR Review 68, no. 2 (2015): 291-313.

Ahn, Hie Joo, and James D. Hamilton. Heterogeneity and unemployment dynamics. No. w22451. National Bureau of Economic Research, 2016.

Burke, Mary A., Alicia Sasser Modestino, Shahriar Sadighi, Rachel Sederberg, and Beldi Taska. "No Longer Qualified? Changes in the Supply and Demand for Skills within Occupations" Manuscript 2018.

Cappelli, Peter H. "Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States." ILR Review 68, no. 2 (2015): 251-290.

Canon, Maria, Chen, Mingyu and Marifian, Elise, (2013), Labor mismatch in the Great Recession: a review of indexes using recent U.S. data, Review, issue May, p. 237-272, https://EconPapers.repec.org/RePEc:fip:fedlrv:y:2013:i:may:p:237-272:n:v.95no.3.

Duncan, Otis Dudley, and Beverly Duncan. "A methodological analysis of segregation indexes." American sociological review 20, no. 2 (1955): 210-217.

Faberman, Jason, and Marianna Kudlyak. "What does online job search tell us about the labor market?." FRB Chicago Economic Perspectives 40, no. 1 (2016).

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. Job search behavior among the employed and non-employed. No. w23731. National Bureau of Economic Research, 2017.

Flowers (2018) https://www.hiringlab.org/2018/03/29/top-jobs-people-dont-leave/

Galaasen, Sigurd Mølster and Andreas Kostøl. "Mismatch and the Consequence of Job Loss" Presentation at the 2018 Symposium for Nonlinear Dynamics and Econometrics

Hobijn: FRBSF Working Paper
http://www.frbsf.org/economic-research/publications/working-papers/2012/wp12-09bk.pdf

Karmel, Tom, and Maureen MacLachlan. "Occupational sex segregation—increasing or decreasing?." Economic Record 64, no. 3 (1988): 187-195.

Kroft, Kory, Fabian Lange, Matthew J. Notowidigdo, and Lawrence F. Katz. "Long-term unemployment and the Great Recession: the role of composition, duration dependence, and nonparticipation." Journal of Labor Economics 34, no. S1 (2016): S7-S54.

Kuhn, Peter, and Hani Mansour. "Is Internet job search still ineffective?." The Economic Journal 124, no. 581 (2014): 1213-1233.

Lazear, Edward P., and James R. Spletzer. The United States labor market: Status quo or a new normal?. No. w18386. National Bureau of Economic Research, 2012a.

Lazear, Edward P., and James R. Spletzer. "The United States labor market: status quo or a new normal?" In Proceedings-Economic Policy Symposium-Jackson Hole, pp. 405-451. Federal Reserve Bank of Kansas City, 2012b.

Longhi, Simonetta, and Mark Taylor. "Employed and unemployed job seekers and the business cycle." Oxford Bulletin of Economics and Statistics 76, no. 4 (2014): 463-483.

Marinescu, Ioana, and Roland Rathelot. "Mismatch unemployment and the geography of job search." *American Economic Journal: Macroeconomics* 10, no. 3 (2018): 42-70.

Modestino, Alicia. "Mismatch in the labor market: Measuring the supply of and demand for skilled labor in New England." *FRB of Boston Public Policy Discussion Paper* 10-2 (2010).

Patterson, Christina, Ayşegül Şahin, Giorgio Topa, and Giovanni L. Violante. "Working hard in the wrong place: A mismatch-based explanation to the UK productivity puzzle." European Economic Review 84 (2016): 42-56.

Rim, Nayoung. The Effect of Title IX on Gender Disparity in Graduate Education. February 2018 Manuscript.

Robinson, Chris. "Occupational Mobility, Occupation Distance, and Specific Human Capital." Journal of Human Resources 53, no. 2 (2018): 513-551.

Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. "The earth mover's distance as a metric for image retrieval." International journal of computer vision 40, no. 2 (2000): 99-121.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. "Measuring mismatch in the US labor market." Manuscript, revised March (2011).

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. "Mismatch unemployment." American Economic Review 104, no. 11 (2014): 3529-64.

Sinclair (2014) https://www.indeed.com/hire/hiring-lab/March-2014.pdf

Smith, Jennifer C. "Unemployment and Mismatch in the UK." In Bank of England/Institute of Macroeconomics conference on ‘Unemployment, productivity and potential output: the aftermath of the crisis. 2012.

Turrell, Arthur, Bradley Speigner, Jyldyz Djumalieva, David Copple, and James Thurgood. Using job vacancies to understand the effects of labour market mismatch on UK output and productivity. No. 737. Bank of England, 2018.

Watts, Martin. "How should occupational sex segregation be measured?." Work, Employment and Society 6, no. 3 (1992): 475-487.

Watts, Martin J. "A critique of marginal matching." Work, Employment and Society 8, no. 3 (1994): 421-431.

Watts, Martin. "Occupational gender segregation: Index measurement and econometric modeling." Demography 35, no. 4 (1998): 489-496.

Wiczer, David. "Long-Term Unemployment: Attached and Mismatched?." Federal Reserve Bank of St. Louis Working Paper Series 2015-042 (2015).

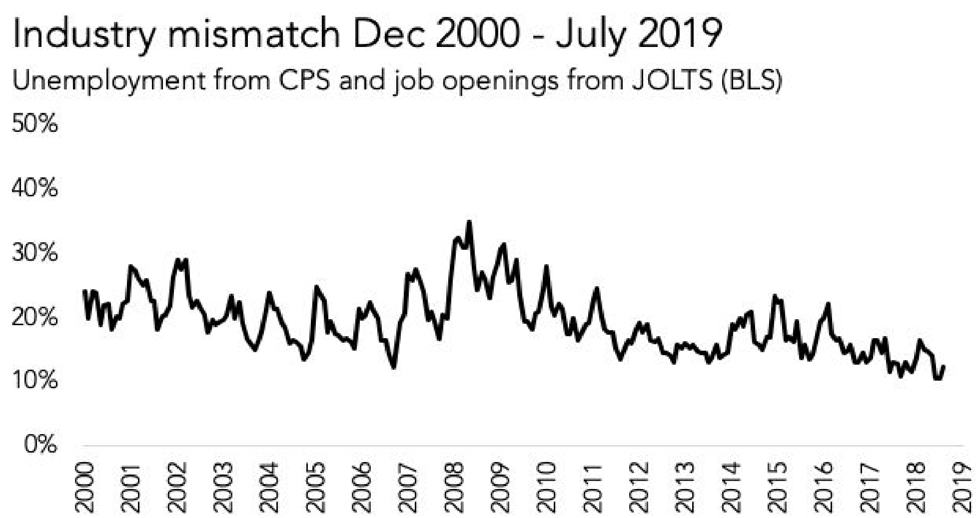
Yang, Jangho. "Information theoretic approaches in economics." Journal of Economic Surveys 32, no. 3 (2018): 940-960.

Appendix A: Industry Mismatch using Publicly Available Data

In this section we provide an analysis similar to Şahin et al (2011) and Lazear and Spletzer (2012a & b) to produce a measure of industry mismatch using publicly available data. We produce an updated estimate of mismatch based on 12 industry categories that are available for vacancies from JOLTS and for the unemployed from the CPS data reported by the Bureau of Labor Statistics.[20] We estimate this industry mismatch for the full sample where JOLTS vacancy data are available as of the writing of this paper: December 2000 through July 2019.

Our estimates are reported in Figure 1. Similar to what was noted by Lazear and Spletzer (2012a and 2012b), we find that industry mismatch rose during the recession from the end of 2007 through mid=2009 and fell during the recovery. Comparing with our online mismatch time sample of 2014-June 2019, we see a similar slight decline in industry mismatch.

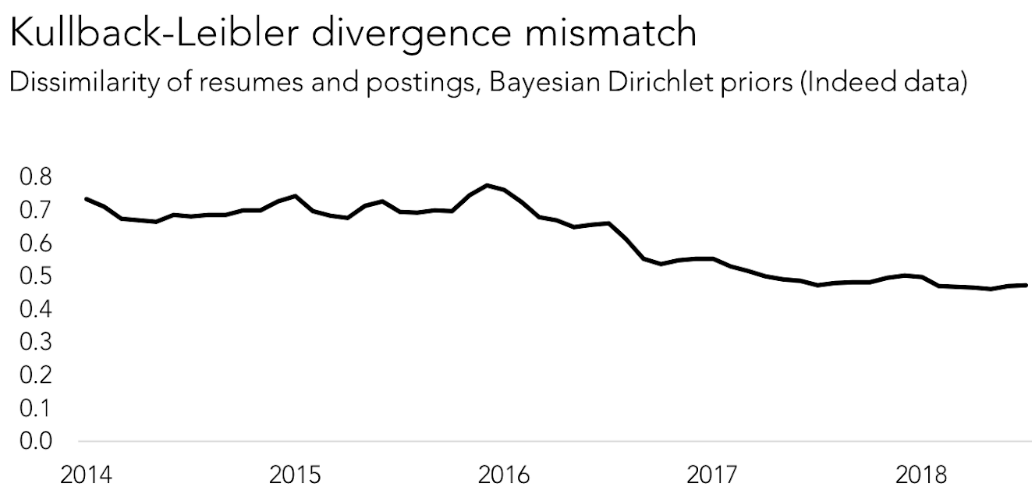
Figure A.1: US labor market mismatch based on publicly available data



[20] The job openings data are from the February 12, 2019, release of JOLTS. The unemployed by industry data are from the CPS. The data are not seasonally adjusted, and using the 12 industries available from both CPS and JOLTS: construction, durable goods manufacturing, nondurable goods manufacturing, wholesale and retail trade, transportation and utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality, other services, and government. Note that we exclude mining due to different definitions between JOLTS and CPS (although including it does not give noticeably different results). Lazear and Spletzer use 12 industries by grouping together durable and nondurable goods manufacturing and including mining. Şahin et al. use CPS microdata to include all 17 industries available in JOLTS. Results are little changed between the different choices of Lazear and Spletzer, Şahin et al. or our analysis. The largest difference is due to our choice of using seasonally unadjusted data.

Appendix B: Results and Robustness Checks Based on Different Samples

Robustness of Mismatch based on Kullback-Leibler Divergence Jan 2014-July 2018



Top Contributors[21] of Changes in the Mix of Job Postings

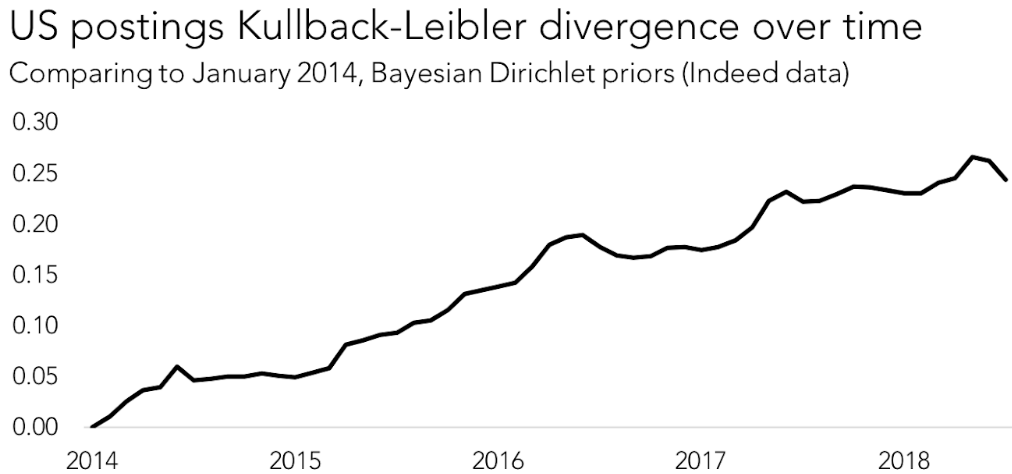
Top contributors to jobs mix changes

Comparing job postings in Dec 2018 to Dec 2014 (Indeed data)

Rank	2018 share > 2014 share	2014 share > 2018 share
1	assistant manager	occupational therapist
2	retail sales associate	physical therapist
3	team member	speech language pathologist
4	cook	owner operator driver
5	shift leader	store manager
6	general manager	stylist
7	seasonal associate	tax professional
8	hair stylist	babysitter/nanny
9	lead associate	cashier/clerk
10	delivery driver	pharmacy assistant

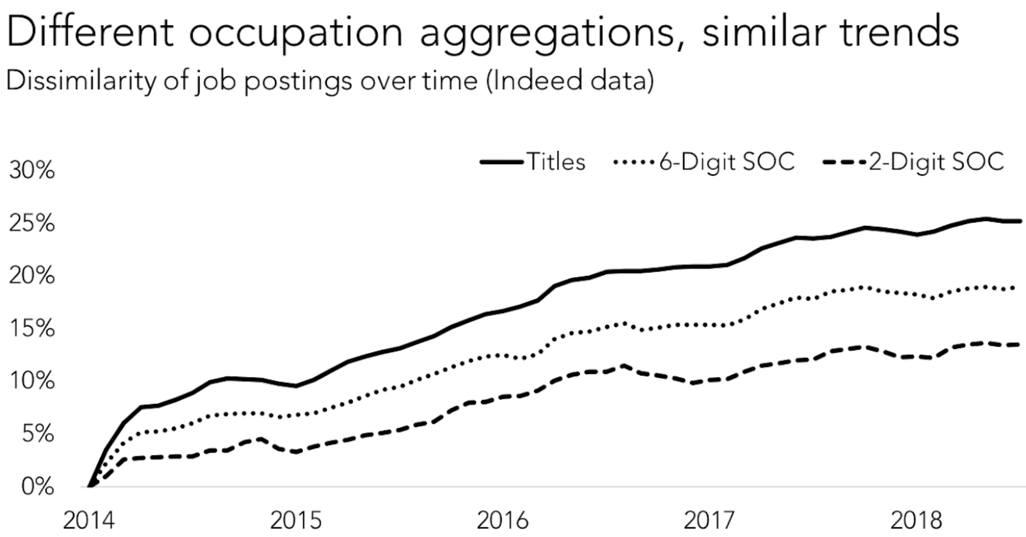
[21] For the top contributors to jobs mix changes, the top ten in growing share contribute 10.6% of the dissimilarity when comparing January of 2018 with January of 2014, while the top ten in shrinking share contribute 9.1%.

Robustness of Jobs Mix Changes using Kullback-Leibler Divergence (1/14-7/18)



Given the sheer number of categories we have, one might think the trend in dissimilarity could be due to the amount of disaggregation. Therefore we looked at different numbers categories based on standard occupation codes (SOC) of between 23 (based on 2-digit SOC code mappings) and 821 (based on 6-digit SOC code mappings) and found a similar trend, although the levels of dissimilarity compared to January of 2014 were lower which is consistent with the higher aggregation.

Changing Mix of Job Postings for Different Aggregation Levels (1/14-7/18)



Top Contributors[22] of Changes in the Mix of Resumes

Top contributors to resume mix changes

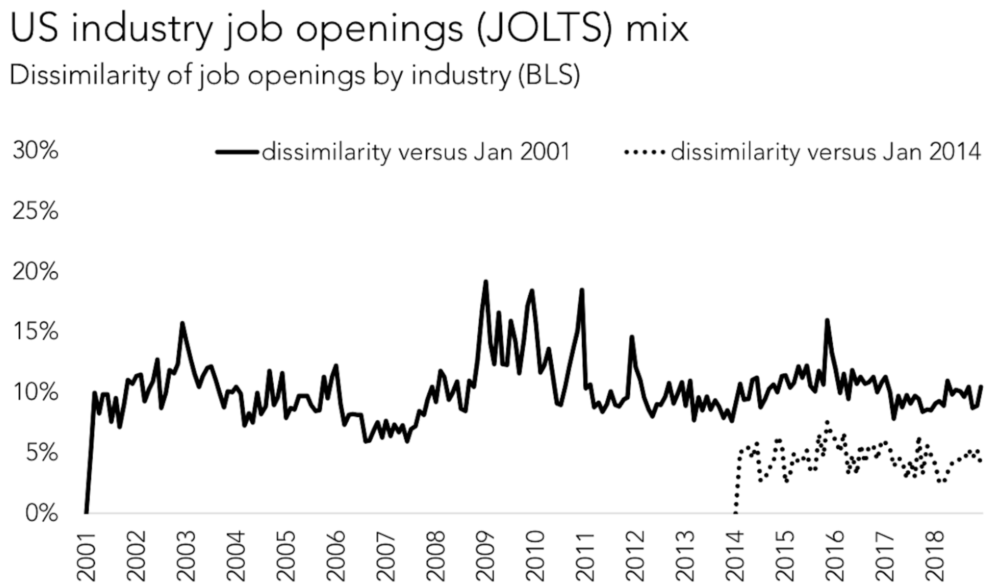
Comparing resumes in Dec 2018 to Dec 2014 (Indeed data)

Rank	2018 share > 2014 share	2014 share > 2018 share
1	server	owner
2	crew member	recruiter
3	customer service associate / cashier	president
4	warehouse worker	office manager
5	delivery driver	administrative assistant
6	nursing assistant	sales representative
7	housekeeper	manager
8	package handler	registered nurse
9	bartender	consultant
10	prep cook	account executive

[22] For the top contributors to resume mix changes, the top ten in growing share contribute 10.4% of the dissimilarity when comparing January of 2018 with January of 2014, while the top ten in shrinking share contribute 6.7%.

Another point of context is comparing our data to data available from the Bureau of Labor Statistics (BLS). For categories we use the 12 industry categories for which we can get monthly data on unemployed and on job vacancies. With a much smaller number of categories (12 as compared to over 6000) we expect the dissimilarity to be smaller, but we might still expect an upward trend.

Changing Industry Mix of Job Openings over Time (Jan 2014 - Dec 2018)

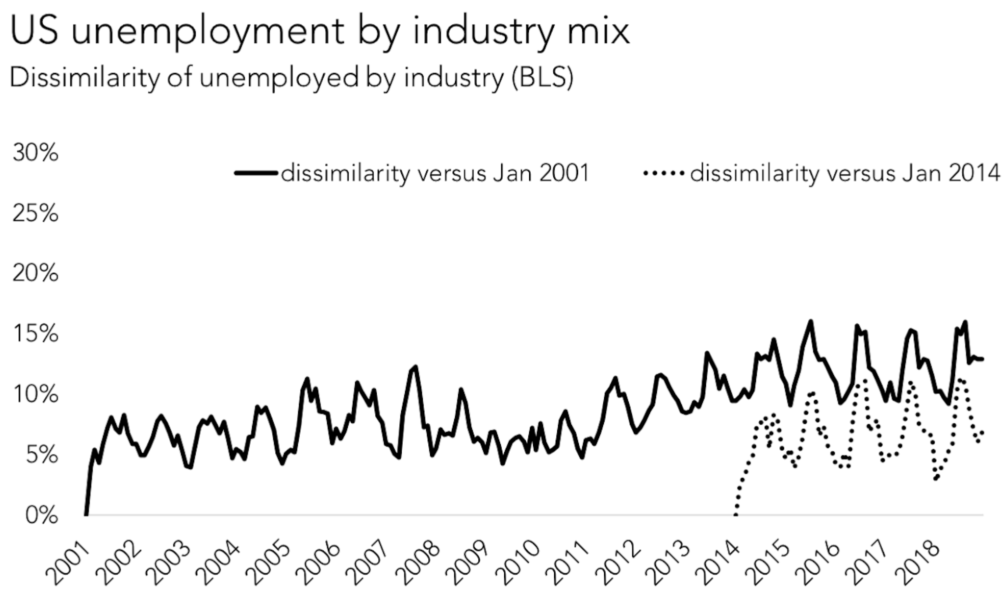


However, JOLTS is showing lower dissimilarity over time and no clear upward trend when compared to either January of 2001 or January of 2014 (and other months are similar so it is not a seasonal issue). The fact that the dissimilarity is lower is likely related to the number of categories: with a smaller number of categories, we would expect the overall level of dissimilarity to be lower (e.g. in the measure above nurses vs. doctors shows up, whereas in this measure they are both under healthcare).

The fact that the trend in industry is holding steady suggests that this isn't about changes in industries, but rather about changes in who those industries are employing (the shift away from occupational therapist to pharmacy technician in healthcare for example).

Shifting to the job seeker side, similar to what we see for job postings, we see little trend by industry in the CPS data for the unemployed, but we do see an upward trend in our resume data.

Changing Industry Mix of Unemployed Job Seekers over Time (Jan 2014-Jan 2018)

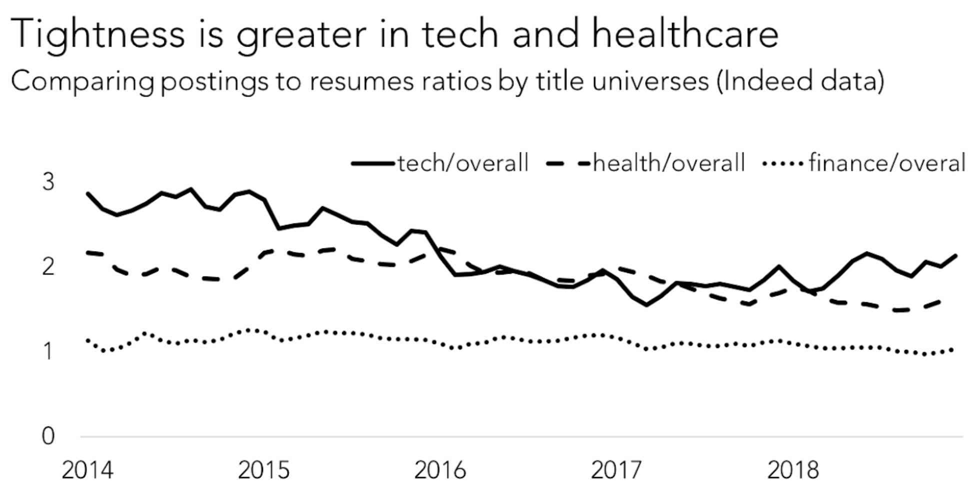


Additional Sector Analysis

We cannot directly compare resumes for a job title with postings for a job title for a number of reasons. Most notably is the use of online platforms for job search and in particular the use of the Indeed website has grown over time. Thus it would not be surprising to see an increasing number of resumes per job posting even as the labor market has tightened. We can, however, compare the ratios of resumes and postings from one sector to that of the economy overall to get a sense of relative tightness over time.

The figure below presents our measures of relative tightness for three sectors, where each sector is defined by a set of normalized titles that can clearly be mapped to that sector (for examples see the tables of top contributors below). Our three sectors are tech, healthcare, and finance. In December of 2018, healthcare was the largest sector with approximately 16% of all US job postings in this sector. Finance had less than 2% and tech had just over 6% of all US job postings.

Relative Tightness for Tech, Healthcare, and Finance Sectors



One interpretation of the variation in sector tightness is that it is relatively harder to attract job seekers to tech and healthcare roles than to other roles, which we might think of as extensive margin mismatch. In fact, if we construct our same dissimilarity measure just comparing shares in sector versus rest we get a similar pattern. Notably for finance since its tightness is so similar to overall, the mismatch between finance and the rest of the economy is basically zero.

Top Contributors to Jobs Mix by Sector

Top contributors to tech jobs mix changes

Comparing job postings in Dec 2018 to Dec 2014 (Indeed data)

Rank	2018 share > 2014 share	2014 share > 2018 share
1	full stack developer	.net developer
2	software engineer	java developer
3	senior software engineer	database administrator
4	data scientist	web developer
5	development operations engineer	developer
6	data engineer	systems analyst
7	cloud engineer	senior system engineer
8	product owner	front end developer
9	machine learning engineer	senior java developer
10	it security specialist	senior systems analyst

Top contributors to healthcare jobs mix changes

Comparing job postings in Dec 2018 to Dec 2014 (Indeed data)

Rank	2018 share > 2014 share	2014 share > 2018 share
1	licensed practical nurse	occupational therapist
2	pharmacy technician	physical therapist
3	nursing assistant	pharmacy assistant
4	caregiver	travel nurse
5	medical assistant	nurse practitioner
6	home health aide	registered nurse - icu
7	massage therapist	family medicine physician
8	certified medical assistant	psychiatrist
9	x-ray technician	physician
10	behavior technician	labor and delivery nurse

Top contributors to finance jobs mix changes

Comparing job postings in Dec 2018 to Dec 2014 (Indeed data)

Rank	**2018 share > 2014 share**	**2014 share > 2018 share**
1	insurance agent	senior operations analyst
2	lending officer	personal banker
3	relationship banker	financial analyst
4	insurance account position	financial representative
5	revenue cycle specialist	operations analyst
6	client associate	finance associate
7	client services associate	investment consultant
8	loan officer	teller
9	adjuster	risk manager
10	closer	banker

Top Contributors to Mismatch by Sector

Top contributors to tech mismatch

Comparing job seeker resumes and job postings in December 2018 (Indeed data)

Rank	resume share > posting share	posting share > resume share
1	technical support	software engineer
2	help desk analyst	senior software engineer
3	IT technician	software architect
4	data analyst	system engineer
5	technical support representative	data engineer
6	IT support	java developer
7	computer technician	senior system engineer
8	technical support supervisor	data scientist
9	desktop support technician	application developer
10	web developer	product manager

Top contributors to healthcare mismatch

Comparing job seeker resumes and job postings in December 2018 (Indeed data)

Rank	resume share > posting share	posting share > resume share
1	nursing assistant	registered nurse
2	caregiver	physical therapist
3	medical assistant	speech language pathologist
4	home health aide	registered nurse - operating room
5	dental assistant	occupational therapist
6	medical receptionist	registered nurse - icu
7	charge nurse	licensed practical nurse
8	dietary aide	registered nurse - emergency room
9	medical biller	nurse practitioner
10	certified medical assistant	registered nurse - home health

Top contributors to finance mismatch

Comparing job seeker resumes and job postings in December 2018 (Indeed data)

Rank	resume share > posting share	posting share > resume share
1	teller	financial analyst
2	loan processor	senior financial analyst
3	collection representative	actuary
4	chief financial officer	loan officer
5	debt collector	mortgage loan originator
6	claims adjuster	risk manager
7	fraud analyst	lending officer
8	claims specialist	relationship banker
9	claims processor	insurance account position
10	insurance specialist	financial planning analyst