X. Using Online Job Vacancy Data to Study Labor Market Dynamics

Mismatch in Online Job Searches

MARTHA E. GIMBEL
Schmidt Futures

TARA M. SINCLAIR
George Washington University and
Indeed Hiring Lab

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, based on their employment history, 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 use of online job search platforms in general and/or the market share of a particular platform.

We produce monthly mismatch measures for the United States, a set of English-speaking countries, 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, 2014) and Lazear and Spletzer (2012a, 2012b) who also quantify the level of mismatch in the economy. They use publicly available survey data from the Bureau of Labor Statistics (BLS) 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. Other research, such as Burke et al. (2019), 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 United States and particularly in the United Kingdom. Turrell et al. (2018) use data from Reed, an online recruiter in the United Kingdom, to estimate mismatch by occupation and geography in the United Kingdom. They find that regional mismatch rather than occupational mismatch 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 the Patterson et al. finding that occupational mismatch is an important contributor to weak productivity growth in the United Kingdom and the 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 publicly available vacancy data are available.[1] 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 afterward, suggesting a cyclical rather than structural pattern.

In this paper, we present a set of mismatch indexes that we compare across English-speaking countries (the United States, the United Kingdom, Australia, Canada 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 all active online job seekers, both employed and unemployed, in our benchmark series, where we identify a job seeker as someone who updated their résumé on the job search website within that month. Including employed job seekers has been challenging in previous analyses due to limited data availability on people searching while employed.[2] There is debate about how similar employed and unemployed job seekers are and what impact differences might have on economic outcomes. On the one hand, Ahn and Hamilton (2019) 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 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.

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 United States 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. We then show that the decline in mismatch is mostly driven by changes on the job posting side, suggesting that missing jobs from the recession have been returning in the recovery in a way that makes the vacancy distribution look more like the job seeker distribution.

In the following sections, we describe our data and mismatch methodology, and then we report our benchmark measure of overall online labor market mismatch for the United States. We find that mismatch has slightly declined as the labor market has tightened, while the distribution of jobs has changed substantially. The changes in the distribution of jobs and résumés 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 comparisons. We then conclude with a discussion of future work.

Data

The analysis is focused primarily on the United States, but we also include analysis for the United Kingdom, 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.[3] For comparison, we also use publicly available data from the Current Population Survey (CPS) and the Job Openings and Labor Turnover Survey (JOLTS).[4] 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 and where we can match with data available from the BLS on the industry of the unemployed, or from job postings aggregated by Indeed from across the Internet.[5] 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 that 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, Indeed collects job postings anywhere on the Internet and de-duplicates 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 (experienced) unemployed, classified by the industry of their last job (from the CPS provided by the BLS), or active job seekers (both employed and experienced unemployed) on Indeed, classified by their most recent job title on their résumé uploaded on Indeed. Indeed has 64.7 million résumés from the United States as of June 2019. We are focusing on the subset that were active accounts during our sample from 2014 through June 2019, where "active" is defined as having last updated their résumé on Indeed in that month.[6] 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.[7]

For robustness, we also use an alternative measure of job seekers based on clicks on job postings. A job seeker can click on a posting only if a job is available, and the click may not indicate the job seeker is qualified, only that they are interested in the role. We then classify the job seeker based on the titles of jobs they click on and compare the distribution of clicks to the distribution of job postings. This analysis allows us to use all job seekers on Indeed rather than being limited to account holders.

Job seekers are not just the unemployed.[8] 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 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 résumé. There is likely measurement error 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. Therefore, we do not report separate results for employed and unemployed job seekers but only combined results for all job seekers. We include only the "experienced unemployed" in our résumé data because we are only using résumés that have previous employment recorded. This is consistent with the CPS 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 6,068 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 6,068 titles were determined as the superset of English normalized titles across the countries in this study: the United Kingdom, the United States, Canada, Australia, Ireland, New Zealand, and Singapore. For some titles, the counts for both résumés and postings are zero in most or all months for one or more countries, which does not meaningfully affect our analysis. We also estimate 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 résumés and job postings and across countries, (2) titles capture skills more consistently that what is reported by job seekers in résumés, (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, and (4) titles allow us to get quite granular as compared to industries or occupations.

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.[9] The Duncan and Duncan measure is

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

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

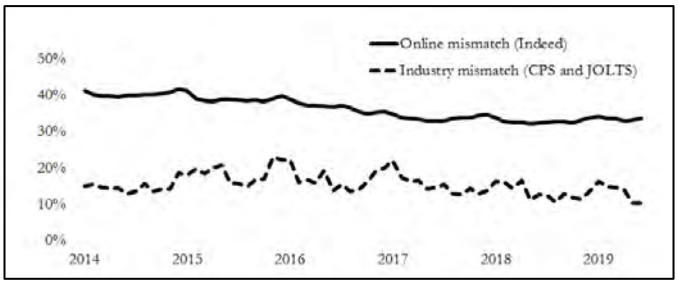
This is the same measure used by Lazear and Spletzer (2012a, 2012b) and Şahin et al. (2011, before incorporating a matching function). This index can be interpreted as the proportion of job seekers who would need to be moved to make the job seeker to posting ratio the same for all job categories, where 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 2014 and report through June 2019.[10] 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 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 CPS and vacancies from the JOLTS following a similar methodology to that used by Lazear and Spletzer (2012a, 2012b). Our measure is higher in level, as would be expected given that we are moving from 12 industry categories to over 6,000 job title categories. In terms of time pattern, however, they are broadly similar, although our measure is substantially smoother.

FIGURE 1
Comparing Mismatch Measures, January 2014–June 2019

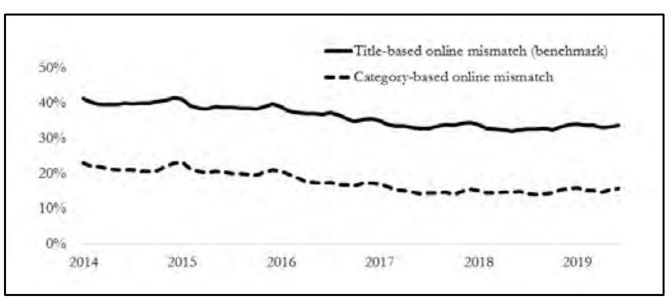


Source: Indeed data, BLS, and authors' calculations.

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.[11]

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 résumés (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.

FIGURE 2
Comparing Online Mismatch Measures
[Job seeker and posting shares grouped by titles (6,068) or categories (57)]



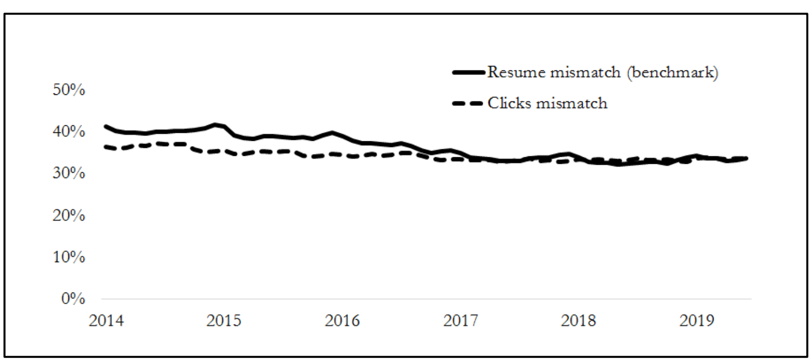
The smoothness of online mismatch may be due to the consistency of the data since the online data source is a common labor market with as much as possible the same definitions applied to both groups. It does not appear to be sensitive to changes in aggregation level, the particular dissimilarity metric used, or changes in our definition of an active job seeker.[12] There appear to be seasonal movements in the distribution of unemployed job seekers in the CPS that are different from the JOLTS job openings numbers, which results in seasonal fluctuations appearing in the industry mismatch series. One interpretation of the smoothness in the online mismatch series is that employed job seekers may have fewer seasonal differences from openings as compared to the unemployed, but further analysis beyond the scope of this paper would be needed to confirm that interpretation.

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 the latest job on job seekers' résumés 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 résumé data are stored that may be affecting our results. Per the terms of Indeed's user agreement, only the latest résumé 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 résumé was last updated.[13] Third, using résumé data means we limit the sample to job seekers who have uploaded a résumé on Indeed, but many people use the website without uploading a résumé. To address these concerns, we consider an alternative measure of job seeker distribution based on the job titles job seekers 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 résumé 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 résumé 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 with Different Job Seeker Measures
(Click shares captures interest for next role vs. experience in résumé)



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 résumés, 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 résumé share exceeds the posting share contributes 10.9% of mismatch, and the top ten where the posting share exceeds the résumé share contributes 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 résumé 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
[Comparing job seeker résumés and job postings in June 2019 (Indeed data)]

Rank	Résumé share > Posting share	Posting share > Résumé 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 Résumés

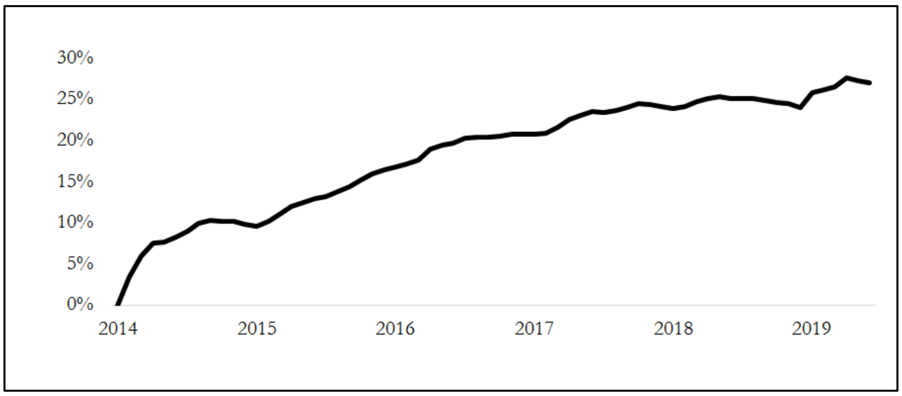
Mismatch could be flat to declining for two reasons: either little is changing underneath or job seekers and jobs opportunities are seeing their distribution across titles change in similar ways over the past several years. To examine this, we used the same dissimilarity index but applied it to jobs and résumés separately over time to see how different jobs and résumés 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 past 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 toward job seekers that has brought about the small decline in mismatch over the sample.

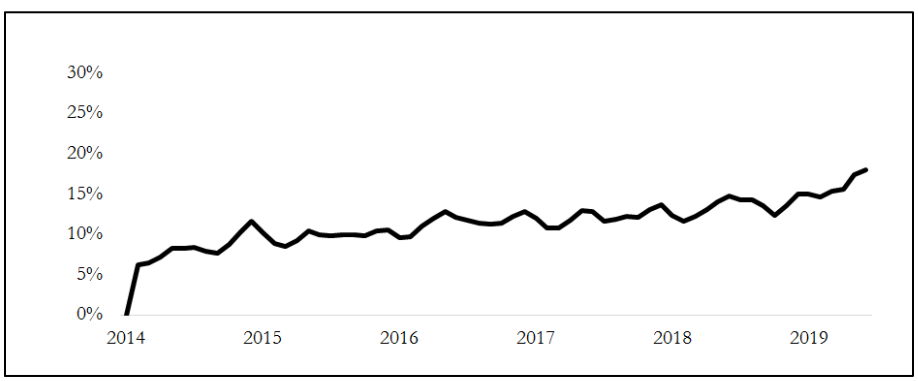
First, looking at the distribution of job postings over time: Figure 4 shows that there has been a substantial change in the distribution across titles in job postings over recent years. Comparing January 2019 with January 2014 (comparing January to January to exclude potential seasonal differences), 25.8% of job postings in 2019 would need to change in order to have the same distribution as five years before.[14]

FIGURE 4
Changing Mix of Job Postings over Time
(Evolution of US job postings mix over time)



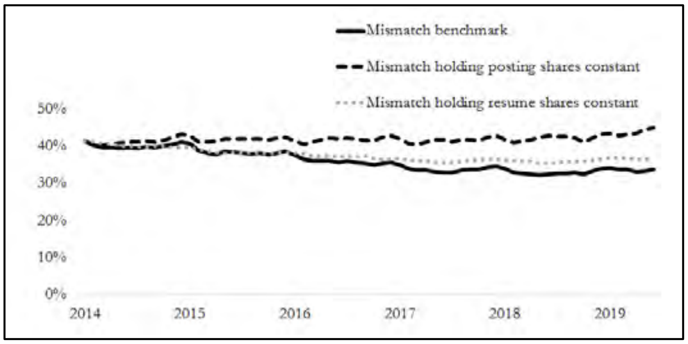
Résumés have changed less over the sample than job postings have. Again comparing January 2019 with January 2014, résumés are 15.0% different than they were five years before (Figure 5). One data note: because of the nature of Indeed's data, where only the latest résumé a job seeker has uploaded is kept,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

FIGURE 5
Changing Mix of Online Résumés over Time
(Dissimilarity of job titles in résumés compared to January 2014)



In order to explore the role of the changes in postings and résumés and their contribution to mismatch, we constructed counterfactual mismatch measures where we held the labor supply (résumés) 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 résumé distribution had not changed, but it is 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
(Mismatch holding one side at January 2014 shares)

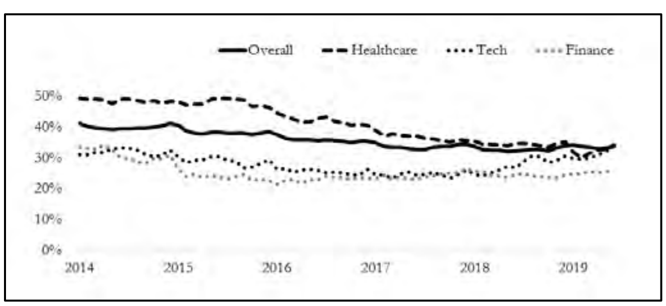


Sector Analysis

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.[15] For the within-sector mismatch we return to our dissimilarity measure and calculate mismatch based on résumés 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 2019, healthcare was the largest sector, with approximately 14% of all US job postings. Finance had less than 2% and tech had almost 6% of all US job postings.

In Figure 7, we show that for most of the sample, healthcare shows greater mismatch than our benchmark overall results for the United States, and tech and finance are both below. Interestingly, at the end of the sample, healthcare mismatch declines and tech mismatch rises to converge close to the overall national level of mismatch. Finance, however, stays flat and well below the national level.

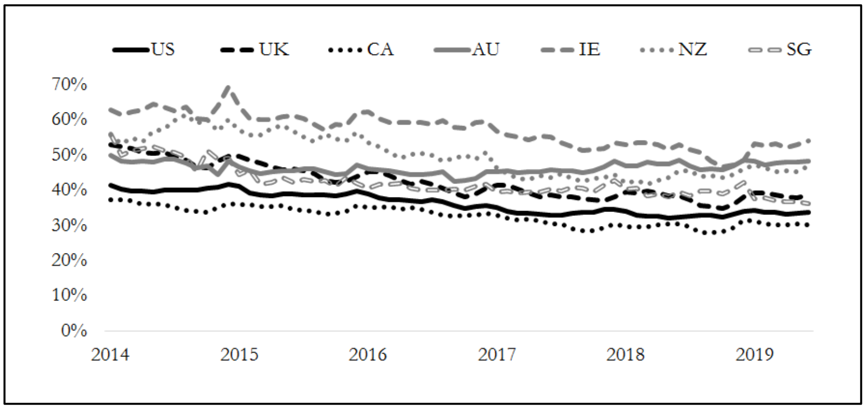
FIGURE 7
Mismatch for Tech, Healthcare, and Finance Sectors
(Mismatch within different sectors)



Cross-Country Comparisons

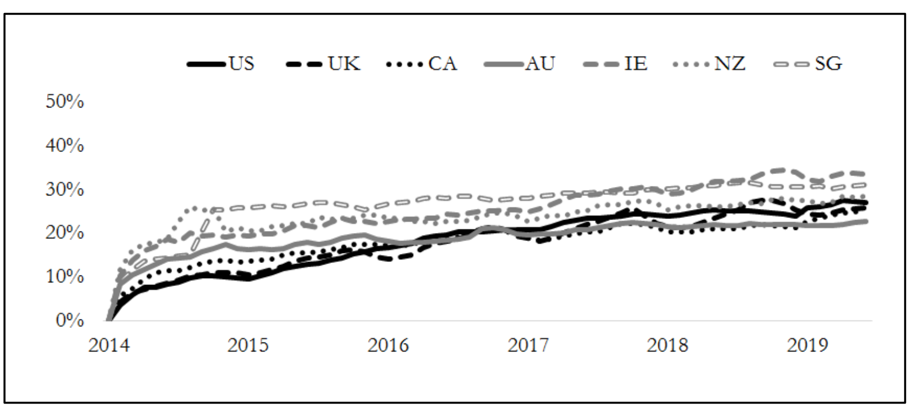
For the same set of 6,068 normalized titles (selected as the superset of normalized titles across the countries), we construct comparable mismatch measures, again monthly from 2014 through June 2019 (Figure 8). The countries have slightly different levels and seasonal patterns, but perhaps the most interesting pattern is the trends: all seven of the countries studied: the United States, the United Kingdom, Canada, Australia, Ireland, New Zealand, and Singapore. Canada and the United States have very similar levels and patterns, with Canada just slightly below the United States 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 United States, the United Kingdom, and Canada is consistent with other labor market indicators for these countries over this time period.[16]

FIGURE 8
Within-Country Mismatch Comparisons
(Overall mismatch in online job search by country)



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 9. 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).

FIGURE 9
Postings Shares Changes over Time for Seven Countries
(Evolution of job postings mix by country)



Source: Indeed data.

Conclusion

This paper shows that even though the distribution of job vacancies has changed substantially since 2014, we see a robust trend of slight decline in mismatch between the distribution of online job vacancies as compared to the distribution of online job seekers over the past several years for the United States and across a range of English-speaking countries. The decline in mismatch appears to be driven by the change in the distribution of jobs toward 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 past several years.

This analysis opens up several directions for future work. In particular, this analysis, consistent with Lazear and Spletzer (2012b), suggests there is a cyclical component to mismatch, which means if we knew the trend or natural rate of mismatch, we could potentially use mismatch as an additional indicator of slack. 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 we expect there to be more information along these lines as we update the series over time.

Furthermore, modeling and weighting for potential career changers may provide additional insights. Although we consider a variety of different aggregation levels with robust results, for each set of categories, our analysis is binary: same category or not same category. 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 in “adjacent” job titles by occupation grouping. A related approach was used by Şahin et al. 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” (2014: 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).

Finally, we can produce estimates of mismatch for more types of job seekers and more regions and countries. In preliminary work, we have estimated mismatch for all the US states and found a decline in mismatch across all states over our time frame to suggest that the national decline is broad based rather than driven by a subset of states. In future work, further analysis of the state-level patterns may provide additional insights. 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,[17] 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 will have to address how to get consistent job titles across languages.

Acknowledgments

Adhi Rajaprabhakaran provided excellent research assistance and Indeed's Data Insights Platform team provided essential data support. Both authors of this paper had an employment relationship with Indeed when involved in the research.

Endnotes

[1] 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 applied to vacancies and job seekers in the publicly available data. Results are little changed between the different choices of Lazear and Spletzer, Şahin et al., and our analysis.

[2] 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).

[3] 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 United States (comScore, November 2018), which makes Indeed the #1 ranked job site by unique visitors in the United States. Furthermore, in July 2018, comScore estimated that 75% of US online job seekers search for jobs on Indeed (per month).

[4] 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).

[5] Şahin et al. (2011, 2014) and Lazear and Spletzer (2012a, 2012b) 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 are 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.

[6] It is possible to use Indeed for job searching without opening an account or uploading a résumé, but our main sample is limited to those with accounts and résumés. Indeed saves only the latest version of résumés, so we count each résumé only one time based on the latest update date because the last job title from the résumé is key to our analysis. We recognize this might cause a bias in the analysis if there is a systematic pattern in who updates résumés frequently and/or who was a job seeker on Indeed early in our sample and again later in our sample. We also estimated a version with the latest résumé attached to all accounts, but activity was determined by the date that the résumé was created. This could also cause a bias because a job seeker could have been in one role in 2014 and searching for a different role, gotten that role, updated their résumé, and searched again in 2016. Interestingly, however, the results were nearly identical in the two models, so there does not appear to be much bias from the updating of the résumés.

[7] There are interesting daily and weekly patterns in the job search data—but less so for job postings data. See https://indeedhi.re/2U55hso for discussion of daily patterns in the data.

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

[9] 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 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.

[10] The data from Indeed are only available consistently over time starting in January 2014, and analysis for this version started in June 2019. For discussion of our initial results, see https://bit.ly/3eAujcy.

[11] According to Şahin et al., "Every statement about the role of mismatch should be qualified with respect to the degree of sectoral disaggregation used" (2014: 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.

[12] We change the measure of the job seeker below to interest based on clicks and, in additional unreported robustness checks, we also used different dating conventions for identifying an active job seeker with little impact on the results. 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.

[13] We also estimated mismatch identifying job seekers using their latest résumé and the date they first uploaded it to Indeed as an alternative and found extremely similar results for mismatch suggesting the updating is not causing much bias.

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

[15] See these two blog posts for further discussion of the healthcare and tech results: https://bit.ly/2Ibyy22 and https://bit.ly/3l6IzMC.

[16] See also https://bit.ly/2U0VPq3. For more analysis of the Canadian and Australian data, see the following blog posts: https://bit.ly/3k53vlT and https://bit.ly/2U0W4RZ.

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

References

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

Ahn, Hie Joo, and James D. Hamilton. 2019. "Heterogeneity and unemployment dynamics." Journal of Business & Economic Statistics, pp. 1-26.

Burke, Mary A., Alicia Sasser Modestino, Shahriar Sadighi, Rachel Sederberg, and Bledi Taska. 2019. "No Longer Qualified? Changes in the Supply and Demand for Skills within Occupations." https://bit.ly/2I8Zi3o

Cappelli, Peter H. 2015. "Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the United States." ILR Review Vol. 68, no. 2, pp. 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." Federal Reserve Bank of St. Louis Review, Vol. 95 (May/June), pp. 237-272.

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A methodological analysis of segregation indexes." American Sociological Review, Vol. 20, no. 2, pp. 210-217.

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

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2017. "Job search behavior among the employed and non-employed." National Bureau of Economic Research Working Paper No. 23731.

Flowers, Andrew. 2018. "The Top Jobs People Don't Leave." https://bit.ly/3k1kUvz

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

Hobijn, Bart. 2012. "The Industry-Occupation Mix of U.S. Job Openings and Hires." Federal Reserve Bank of San Francisco Working Paper No. 2012-09.

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

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

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

Lazear, Edward P., and James R. Spletzer. 2012a "The United States labor market: Status quo or a new normal?" National Bureau of Economic Research Working Paper No. 18386.

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

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

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

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

Rim, Nayoung. 2018. "The Effect of Title IX on Gender Disparity in Graduate Education." AccessLex Institute Research Paper No. 18-05.

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

Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. 2000. "The earth mover's distance as a metric for image retrieval." International Journal of Computer Vision Vol. 40, no. 2, pp. 99-121.

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Giovanni L. Violante. 2011. "Measuring mismatch in the US labor market." https://nyfed.org/3k6bMWG

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

Sinclair, Tara M. 2014. "What Job Seekers Want: Occupation Satisfaction & Desirability Report." https://indeedhi.re/3k3SiSv

Smith, Jennifer C. 2012. "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.

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

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

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

Watts, Martin. J. 1998. “Occupational gender segregation: Index measurement and econometric modeling.” Demography Vol. 35, no. 4, pp. 489-496.

Wiczer, David. 2015. “Long-Term Unemployment: Attached and Mismatched?” Federal Reserve Bank of St. Louis Working Paper No. 2015-042.

Yang, Jangho. 2018. “Information theoretic approaches in economics.” Journal of Economic Surveys, Vol. 32, no. 3, pp. 940-960.