

**Institute for International Economic Policy Working Paper Series
Elliott School of International Affairs
The George Washington University**

Mismatch in Online Job Search

IIEP-WP-2020-1

**Tara M. Sinclair
George Washington University**

February 2020

Institute for International Economic Policy
1957 E St. NW, Suite 502
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: iiep@gwu.edu
Web: iiep.gwu.edu

Mismatch in Online Job Search

Martha E. Gimbel
Schmidt Futures

Tara M. Sinclair¹
The George Washington University and Indeed Hiring Lab

This Version: February 16, 2020

JEL Codes: E24, J11, J21, J24, J40, J62

Keywords: Job search, vacancies, employment, unemployment

Abstract

Labor market mismatch is an important measure of the health of the economy but is notoriously hard to measure since it requires information on both employer needs and job seeker characteristics. In this paper we use data from a large online job search website which has detailed information on both sides of the labor market. Mismatch is measured as the 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 and a set of English-speaking countries from January of 2014 through December of 2019. We find that title-level mismatch is substantial, with about 33% of the labor force needing to change job titles for the US to have zero mismatch in 2019, but that it declined from 40% in 2014 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 overall distribution of jobs more similar to the distribution of job seekers. We interpret this finding as evidence that mismatch between job seekers and employers eased due to jobs coming back in the slow recovery after the Great Recession.

¹Corresponding author: tsinc@gwu.edu. 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, the monetary policy conference at the Federal Reserve Board "Nontraditional Data, Machine Learning, and Natural Language Processing in Macroeconomics," and the LERA@ASSA 2020 meetings 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, 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 US and a set of English-speaking countries from January of 2014 through December 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 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 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 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 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 publicly available vacancy data are available.² 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 English-speaking countries (the US, the UK, 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 online mismatch series, where we identify a job seeker as someone who updated their resume 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

² 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. or our analysis.

employed.³ 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 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 modestly 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 in 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.

³ 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).

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 US. We find that mismatch has modestly declined as the labor market has tightened, while the distribution of jobs has changed substantially. 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 six other English speaking countries. We then conclude with a discussion of future work.

Data

The analysis 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.⁴ For comparison we also use publicly available data from the Current Population Survey (CPS) and the Job Openings and Labor Turnover Survey (JOLTS).⁵ We focus on seasonally unadjusted data from all sources. Our measure of mismatch will be in category shares of totals, which should net out any aggregate seasonal patterns, and will leave only job category seasonal patterns, which we are interested in examining.

⁴ 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 2019). Indeed has 53.6 million unique visitors per month in the US ([comScore](#), November 2019) which makes Indeed the #1 ranked job site by unique visitors in the US. Furthermore, in February of 2019, [comScore](#) estimated that 73% of US online job seekers search for jobs on Indeed (per month).

⁵ The job openings data are from the February 11, 2020, release of [JOLTS](#). The unemployed by industry data are from the [CPS](#), released on February 7, 2020. 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).

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 by job title, aggregated by Indeed from across the internet.⁶ 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, 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 Current Population Survey provided by the Bureau of Labor Statistics), or active job seekers (both employed and experienced unemployed) on Indeed, classified by their most recent job title on their resume uploaded on Indeed. Indeed has 71.7 million resumes from the US as of December of 2019. We are focusing on the subset that were active accounts during our sample from 2014 through 2019, where active is defined as having uploaded their resume on Indeed in the month we count them. Job seekers can use Indeed

⁶ 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.

without opening an account or uploading a resume, but our main sample is limited to those with accounts and resumes. Indeed only saves the latest version of resumes, so we only count each resume one time based on the date it was created. The information on the resume, if it was updated since the creation, will be later data. This could impact people who searched for a job and then updated their resume when they found a job. Therefore in the appendix we consider an alternative analysis where we place the job seeker as active in the month they last updated their resume. We aggregate to the monthly frequency, but we could look at daily or even intra-day based on the Indeed data. There are regular daily and weekly patterns in job seeker behavior, but less so for job postings data.⁷

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. 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

⁷ See <http://blog.indeed.com/2017/06/27/timing-matters-in-the-job-search/> for discussion of daily patterns in the data.

⁸ We are 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.

indicating that they are currently employed at one of the positions listed on their resume. 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” our resume data because we are only using resumes 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 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). The same normalization rules were applied to the titles in job postings and those from resumes. 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. Including these titles does not meaningfully

affect our analysis because including a category that has zero shares for both postings and resumes does not add to or subtract from our dissimilarity metric. 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 than 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.

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:

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

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 Şahin et al. (2011, before incorporating a matching function). This index can be interpreted as the proportion

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 the assumption that it is easier for the job seeker to change occupation than employers to change their needs seems more reasonable.

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

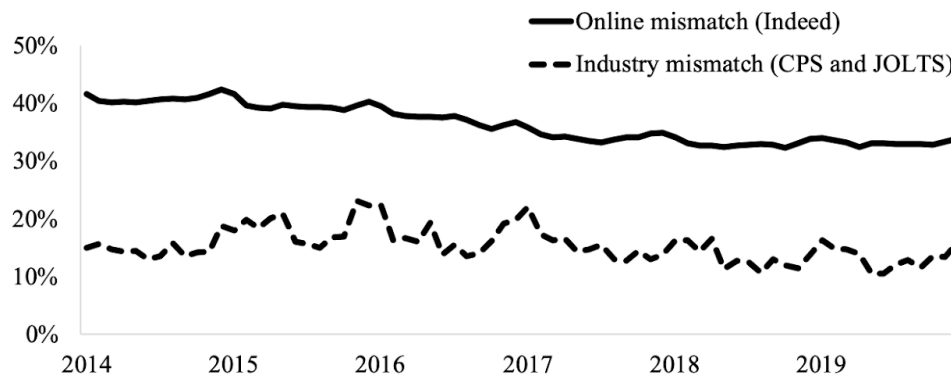
For our measure of mismatch based on online job search, we start in January of 2014 and report through December of 2019.¹⁰ One of the benefits of using the online data is more timely arrival of updated information. As soon as the first week of each month we could update our measures rather than waiting 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 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 Lazear and Spletzer (2012a and 2012b). Our measure is higher in level, as would be expected given that we are moving from 12 industry categories to over 6000 job title categories

¹⁰ The data from Indeed are available consistently over time starting in January of 2014 and analysis for this version started in January of 2019. For discussion of our initial results, see <https://www.hiringlab.org/2018/09/20/us-labor-market-changing/>.

which broadens the opportunity for mismatch. In terms of time pattern, however, they are broadly similar, although our measure is substantially smoother.

Figure 1: Comparing mismatch measures
Jan 2014 - Dec 2019



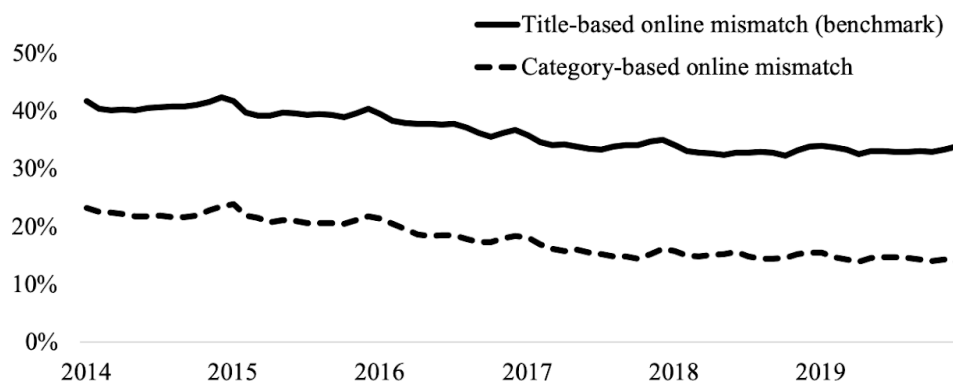
Lazear and Spletzer find much more mismatch by occupation than by industry, which is consistent with what we find for our online labor market mismatch at the normalized job title level. Job titles are much more similar to occupation than to industry. We would also expect that there would be more mismatch when there are more categories.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 modest decline over our time frame. Limiting the analysis to only titles with large numbers of

¹¹ 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.

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.

Figure 2: Comparing online mismatch measures
Job seeker and posting shares grouped by titles (6068) or categories (57)



The trend in online mismatch 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 Online mismatch also consistently appears smoother than what we see in the industry mismatch reported in Figure 1. Since that smoothness also appears when we construct online mismatch at the category level (Figure 2) rather than the title level, this 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. The industry mismatch also uses only the (experienced) unemployed as the measure for the job seeker, and there appear to be seasonal movements in the distribution of unemployed job seekers in the CPS that are different from the

¹² We change the measure of the job seeker below to interest based on clicks and in the appendix we also use a different dating convention for identifying an active job seeker with little impact on the results.

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 less 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.

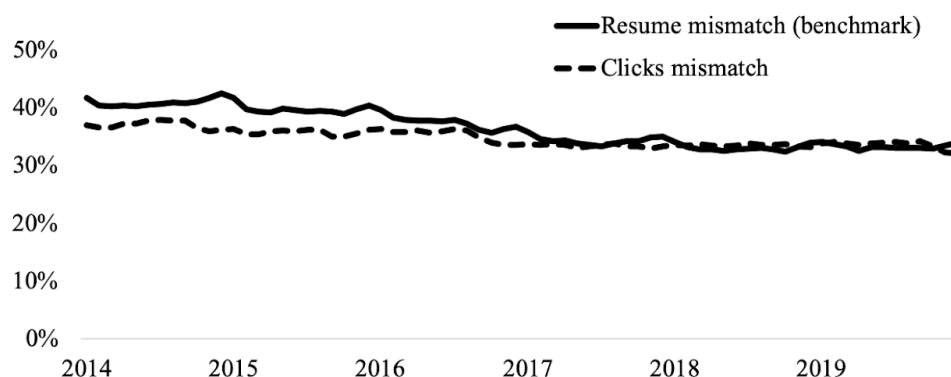
Despite the relative smoothness, we do see clear seasonality in mismatch. This might be expected because we do not use seasonally adjusted data, but it is interesting that the seasonal patterns are sufficiently different in job postings versus job seeker behavior that we see clear rises and falls each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

At least three concerns arise from our use of the latest job on job seekers' 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. In order to have the job search activity distributed across our sample we identify the time of job search based on when a job seeker first uploads or creates a resume on Indeed. That means we associate some later job titles with earlier job search. 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 to search for a job without uploading a resume. To address these concerns we consider an alternative measure of job seeker distribution based on the job titles job seekers click on (and in the appendix we also explore an alternative dating of the resumes). Clicks 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.

Figure 3: Mismatch with different job seeker measures
Click shares captures interest for next role versus experience in resume



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 titles where the resume share exceeds the posting share contribute 10.8% of mismatch, and the top ten titles where the posting share exceeds the resume share contribute 9.1% of mismatch. So, just 20 titles of the 6068 in the sample make up almost one-fifth of all mismatch. The top contributors to mismatch are also notably persistent with some seasonal patterns. For example, comparing the December 2019 list to the list for December 2018, we get slightly different ordering but remarkably similar titles. Even going back three years to December of 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 resumes and job postings in December 2019 (Indeed data)

Rank	resume share > posting share	posting share > resume share
1	customer service associate / cashier	shift manager
2	cashier	retail sales associate
3	laborer	registered nurse
4	customer service representative	babysitter/nanny
5	warehouse worker	restaurant manager
6	manager	assistant manager
7	forklift operator	delivery driver
8	administrative assistant	seasonal associate
9	office manager	licensed practical nurse
10	server	general manager

Changing job postings and changing resumes

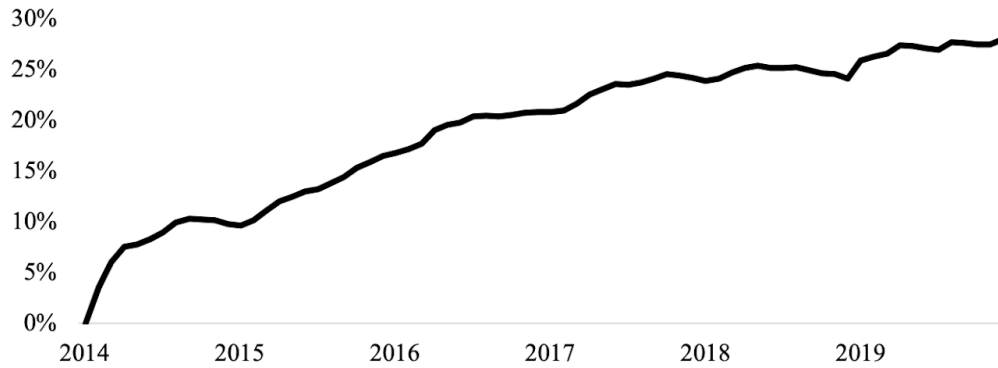
Mismatch could be modestly declining for two reasons: either relatively little is changing underneath or job seekers and jobs opportunities are seeing their distribution across titles change in similar ways over the last several years with one or both drawing somewhat closer to the other. To examine 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 December of 2019, we constructed the following dissimilarity metric:

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

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.

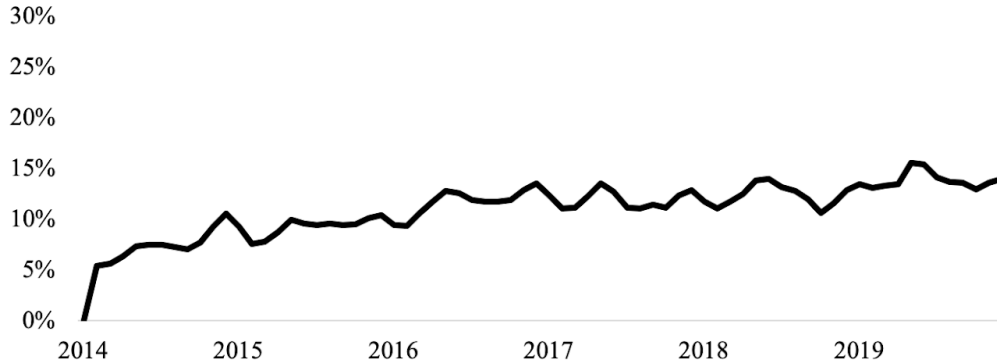
First, looking at the distribution of job postings over time: Figure 4 shows that there has been a substantial change in the distribution across titles in job postings over recent years. Comparing January of 2019 with January of 2014 (comparing January to January to exclude potential seasonal differences), 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
Evolution of US job postings mix over time



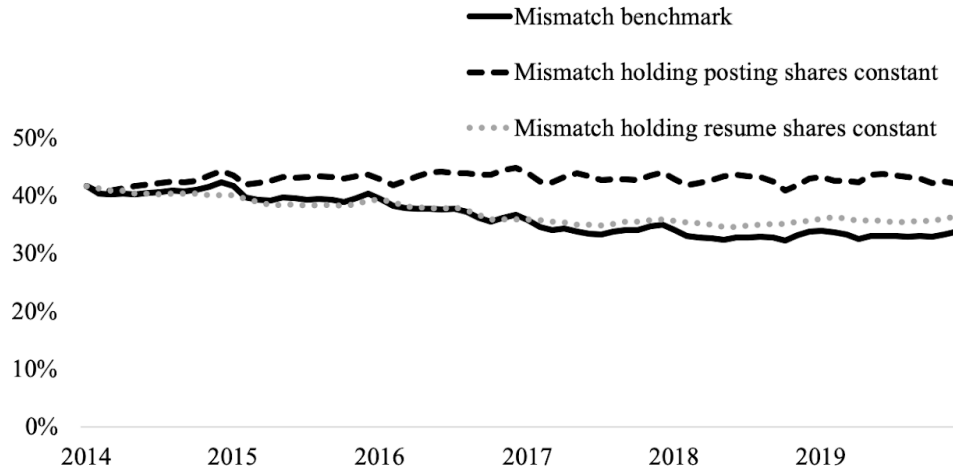
Resumes have changed less over the sample than job postings have. Again comparing January of 2019 with January of 2014, resumes are 13.5% different than they were five years before (Figure 5).

Figure 5: Changing mix of online resumes over time
Dissimilarity of job titles in resumes compared to January 2014



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 that the same downward trend is still evident. When we hold the postings distribution constant, however, mismatch rises rather than declines over the sample. This analysis suggests that the key driver of the decline in mismatch over our sample is an adjustment of the distribution of job postings towards the distribution of job seekers. This suggests that the Great Recession affected not just the quantity of jobs but also the distribution of job opportunities and roles have returned as the economy has recovered.

Figure 6: Analysis holding one side of mismatch constant
Mismatch holding one side at Jan 2014 shares

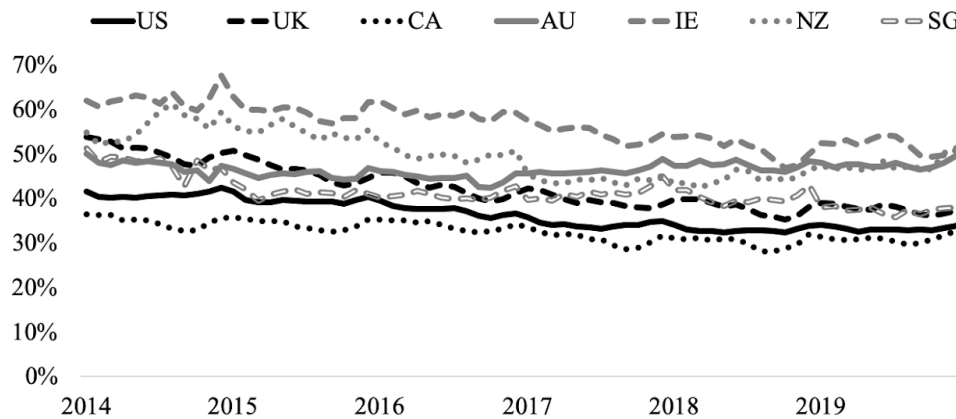


Cross Country Comparisons

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 December of 2019 (Figure 7). The countries have slightly different levels and seasonal patterns, but perhaps the most interesting pattern is the trends: six of the seven countries studied, the US, the UK, Canada, Ireland, New Zealand, and Singapore, have seen declines over the sample. The seventh country, Australia, is at almost the same level of mismatch at the end of the sample as in 2014, after having seen a decline over the first half of the sample and then a rise over the second half back to 2014 values. Canada and the US have very similar levels and patterns, with Canada just slightly below the US throughout the sample. The similarity of the US,

UK, and Canada is consistent with other labor market indicators for these countries over this time period.¹³

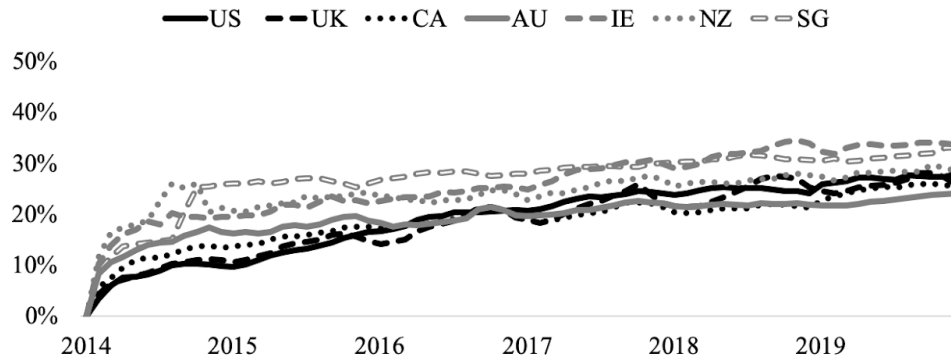
Figure 7: Within country mismatch comparisons
Overall mismatch in online job search by country



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 8. 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).

¹³ See also <https://www.hiringlab.org/2019/05/30/changing-labor-markets-around-world/>. 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 8: Postings shares changes over time for 7 countries
Evolution of job postings mix by country (Indeed data)



Conclusion

This paper shows that even though the distribution of titles in job vacancies has changed substantially since 2014, we see a robust, although modest, decline in mismatch between the distribution of online job vacancies as compared to the distribution of online job seekers over the last several years for the US and also for a set of English-speaking countries. The decline in mismatch 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.

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. Another is that skills may be transferable across categories or that 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 amounts 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. Recent analysis by Bunker (2020) of the Job-to-Job Flows (J2J) data from the US Census Bureau’s Longitudinal Employer-Household Dynamics program shows that for most industries it is true that the most common job move is within the same industry, but it is still less than 50% of the time. This suggests that cross-industries moves are important, but may not end up impacting aggregate analyses such as mismatch. 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,¹⁴ 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.

¹⁴ 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.

Bunker, Nick. 2020. "The State of Job Switching and Hiring."
<https://www.hiringlab.org/2020/02/11/job-switching-and-hiring/> [February 13, 2020].

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". http://conference.iza.org/conference_files/DATA_2019/modestino_a28319.pdf [January 14, 2020].

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://www.hiringlab.org/2018/03/29/top-jobs-people-dont-leave/> [January 14, 2020].

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." Manuscript available at https://www.newyorkfed.org/medialibrary/media/research/economists/topa/USmismatch_v14.pdf [January 14, 2020].

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." Available at: <https://www.indeed.com/hire/hiring-lab/March-2014.pdf> [January 14, 2020].

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.

Appendix

An ideal set of data on the job seeker side would be the job title the job seeker had at the time of each job search spell. However, per the terms of Indeed's user agreement, only the latest resume a job seeker has uploaded is kept. This means we have at most one resume per job seeker, and this resume is the latest version. We know when they first created their resume, but if they later updated it we only know the latest job title on their latest resume.

We thus have two options for dating a job search. The one we use in the main body of the paper is to date the user's activity from when they first uploaded their resume. This approach prevents the loss of job seeker counts early in the sample by job seekers who created a resume and then later updated it. But, this approach also associates later information for earlier job searches. For example, a job seeker could have been in one role in 2014, searched at that time for a different role, gotten that role and then updated their resume on Indeed for a future job search. The job seeker, however, was qualified to be hired for the role they moved into, so in some sense using the later title captures the job seeker's true qualifications and may even control for some easily transferrable skills that might be missed when using a large number of titles to categorize the job seekers.

To examine the impact of our dating choice, we present here the mismatch estimates using the last update date as the timing for the job search. This approach 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, since we will only count them on their last search. This approach also results in a much larger number of observations later in the sample as compared to the first approach because lots of job seekers update their

resume. Figure A1 presents the two approaches for resume dating as well as the clicks version for identifying job searches. All three approaches are reassuringly similar and show a modest downward trend.

Figure A1: Mismatch with different job seeker measures
Benchmark: job seeker active based on last resume update date
Alternative: job seeker active based on resume create date

