Skill Mismatch Unemployment*

Marianna Kudlyak†
Federal Reserve Bank of San Francisco

Erin L. Wolcott‡
Middlebury College

November 25, 2019

PRELIMINARY AND INCOMPLETE

Abstract

We measure the aggregate skill match probability between the demand and supply of skills using an extended search and matching framework. In our framework, unemployed workers and vacancies meet randomly, but the probability of matching depends on how closely the unemployed worker's skillset aligns with the job requirements. We use data from the CPS, O*NET, and HWOL and find that the aggregate skill match probability declined during the 2007-09 recession but has since returned to pre-recession levels.

JEL: E24, E32, J30, J41, J63, J64

Keywords: Job Finding Rate. Unemployment. Mismatch.

*The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of San Francisco, the Federal Reserve System, or any other institution with which the authors are affiliated.

†marianna.kudlyak@sf.frb.org.

‡wolcott@middlebury.edu.

1 Introduction

What determines whether a jobseeker matches with a job opening? Many factors including: how many job openings are available, how many other jobseekers, but most importantly whether the skills possessed by the jobseeker meet the requirements of the vacancy.

We measure the aggregate skill match probability—the probability that a meeting between a jobseeker and a vacancy results in a match in an economy with heterogeneous jobseekers and vacancies.

The most intuitive way of thinking about why a jobseeker is not hired for a particular vacancy is because the jobseeker's skills do not match the vacancys requirements. Recent studies document significant secular shifts in the skill requirements due to automation and job polarization (Autor and Dorn (2013)), demand for interpersonal skills (Deming (2017)), and software substitution of skilled jobs (Aum (2017)). The skills of laid-off workers are often tied to disappearing occupations. Skills of new labor market entrants reflect current schooling curriculum that might not be up to speed with the requirements of new jobs.

We employ a search and matching function that allows for nontrivial probability of a meeting between jobseeker and vacancy with different skill characteristics to result in a match. In the search and matching framework jobseekers and vacancies meet randomly but the probability of matching depends on the distance between the skills of the jobseeker and the vacancy. We aggregate up over match probabilities of all potential heterogeneous pairs of jobseekers and vacancies to create a new measure of the aggregate skill match probability directly accounts for skill differences. This measure is close to one when jobseekers have the skills firms are looking for. This measure is close to zero when jobseekerss have skills that are very different from what firms are looking for.

To obtain data on the skillsets of jobseekers, we merge the Current Population Survey (CPS) with skill data from O*NET. To obtain data on the skill requirements of job openings, we merge The Conference Board's help-wanted online (HWOL) data with skill data from O*NET.

We find that our measure of the aggregate skill match probability declined during the 2007-09 recession but it has since recovered to pre-recession levels.

Our work relates to several papers on mismatch unemployment. Şahin, Song, Topa, and Violante (2014) examine mismatch with an island model where workers and job openings are segregated onto islands and only workers with a particular skillset (i.e. in a given occupation) can search for jobs on that island. Unlike in our framework, in their framework each period workers search only in one island. In other words, it is impossible for a worker to match with a vacancy in a different occupation. Herz and Van Rens (2018) estimate a long time series of unemployment mismatch back to 1979. They find that unemployment mismatch is quite cyclical and conclude it is largely driven by wage frictions and barriers to job mobility. We differ from these papers in that we zero-in on a specific type of mismatch: skill mismatch. Whereas the aforementioned papers study mismatch across geographic locations, industries, or occupations, we explicitly model and estimate the skill compatibility between jobseekers and job openings.

Our work also relates to a large literature studying the decline in matching efficiency during and after the 2007-09 recession. Hall and Schulhofer-Wohl (2018) measure matching efficiency when the population of jobseekers is heterogeneous. They find that overall matching efficiency declined smoothly over the period from 2001 through 2013 and that measures of matching efficiency neglecting heterogeneity among the unemployed and neglecting jobseekers other than the unemployed predict large declines in efficiency between 2007 and 2009. Davis, Faberman, and Haltiwanger (2013) study how much of the decline in the aggregate matching efficiency over the 2007-09 recession can be attributed to the changes in recruiting intensity. Barnichon and Figura (2015) find that matching efficiency is procyclical. They show that matching efficiency declines substantially when, as in the Great Recession, the average characteristics of the unemployed deteriorate substantially, or when some labor markets fare worse than others. Sedláček (2016) finds that almost half of the rise in U.S. unemployment during the 2007-09 recession is explained by a drop in match efficiency of the unemployed. In line with these papers, we find that the skill mismatch increased during the 2007-09 recession, which contributed to the decline in the aggregate matching efficiency during the 2007-09 recession.

The rest of the paper is organized as follows. Section 2 describes a theoretical framework for a matching function with skill heterogeneity. Section 3 describes the data. Section 4 provides descriptive statistics about the supply and demand of skills. Section 6 describes the main results. Section 5 concludes.

2 Matching Function with Skill Heterogeneity

2.1 Aggregate Matching Function

Consider an economy with a heterogeneous pool of jobseekers and vacancies. Let j index a vacancy type characterized by an S-dimensional vector of requirements $V_j = [v_j^1, v_j^2 ..., v_j^S]$. Let i index an unemployed jobseeker's type characterized by an S-dimensional vector of characteristics that correspond to the vacancy requirements $U_i = [u_i^1, u_i^2 ..., u_i^S]$. These requirements can encompass a vector of skills, tasks or an amount of experience.

The total number of jobseekers in the economy in period t is $U_t = \sum_{i=1}^{I} U_{it}$ and the total number of vacancies is $V_t = \sum_{j=1}^{J} V_{jt}$.

Search is random. The number of meetings between jobseekers and vacancies is described by the standard Cobb-Douglas function

$$M_t = \xi_t V_t^{\alpha} U_t^{1-\alpha}, \tag{1}$$

where $\alpha \in (0, 1)$ is the meeting function elasticity with respect to the number of vacancies and ξ_t is the meeting efficiency.

A meeting between jobseeker i and vacancy j results in a match with probability P_{ij}, $P_{ij} \leq 1$, which we describe below.

The probability that a jobseeker meets a vacancy (Petrongolo and Pissarides (2001)) is

$$\frac{M_t}{U_t} = \xi_t \left(\frac{V_t}{U_t} \right)^{\alpha}. \tag{2}$$

The probability that jobseeker i meets vacancy j and the meeting results in a match is

$$\lambda_{ijt} = \xi_t \left(\frac{V_t}{U_t} \right)^{\alpha} \frac{V_{jt}}{V_t} P_{ij}. \tag{3}$$

Note that probability P_{ij} is time-invariant but the framework can be easily extended to the time-varying one.

The job finding rate of jobseeker i is

$$\lambda_{it} = \sum_{j=1}^{J} \lambda_{ijt} = \xi_t \left(\frac{V_t}{U_t}\right)^{\alpha} \sum_{j=1}^{J} \frac{V_{jt}}{V_t} P_{ij}. \tag{4}$$

The average job finding rate in the economy is

$$\lambda_t = \sum_{i=1}^{I} \frac{U_{it}}{U_t} \lambda_{it} = \xi_t \left(\frac{V_t}{U_t}\right)^{\alpha} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{U_{it}}{U_t} \frac{V_{jt}}{V_t} P_{ij}. \tag{5}$$

or

$$\lambda_t = \mu_t \left(\frac{V_t}{U_t}\right)^{\alpha}, \tag{6}$$

where

$$\mu_t \equiv \xi_t \times \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{u_{it}}{u_t} \frac{v_{jt}}{v_t} P_{ij}}_{\phi_t}$$

is the aggregate matching efficiency. The aggregate matching efficiency is the product of the aggregate meeting efficiency, ξ_t, and ϕ_t.

Term ϕ_t is the average probability of a meeting resulting in a match given the composition of the aggregate search pool. It takes values between 0 and 1. When $\phi_t = 0$, no meetings result in a match. When $\phi_t = 1$, all meetings result in a match. The higher the ϕ_t, the greater the number of matches in the economy, given the aggregate meeting efficiency, and the aggregate number of vacancies and jobseekers. More generally, ϕ_t captures any heterogeneity that influences pairwise matching probability P_{ij}, not only the one related to skill. In what follows, we refer to ϕ_t as the aggregate skill match probability of the aggregate matching efficiency.

In an economy without heterogeneity, i.e., when any meeting between jobseeker i and vacancy j is equally likely to turn into hiring, i.e., $P_{ij} = 1 \forall i, j$, the aggregate matching efficiency equals the aggregate meeting efficiency, $\mu_t = \xi_t$, and the number of meetings equals the number of matches. However, since vacancies are heterogeneous in their requirements and jobseekers are heterogeneous in their characteristics, not all random meetings between a vacancy and a jobseeker result in a match. In the economy with heterogeneity, the number of matches is smaller than the number of meetings because the aggregate skill match probability is less than 1.

Since the aggregate skill match probability being less than 1 is a defining characteristic of the economy with heterogeneity, unity is not a useful benchmark for ϕ_t. Instead, we envision a long-run average level of the aggregate skill match probability. The deviations of the aggregate skill match probability from this long-run average constitute a skill mismatch in the matching efficiency.

The aggregate skill match probability changes with the changes in the probability of a successful match between jobseeker i and vacancy j across all (i, j) pairs and with the changes in the composition of different types of jobseekers and vacancies in the aggregate search pool. Thus, the aggregate skill match probability can vary over time even if the probability of a pairwise match P_{ij} is constant.

In what follows, we analyze how much of the change in the aggregate matching efficiency μ_t can be attributed to the change in the aggregate skill match probability ϕ_t and how that has changed over time.

2.2 Measuring Match Probability

We focus on the heterogeneity in terms of occupational skills, which is an important dimension for a successful match (Autor et al. (2003)). Each occupation is characterized by a vector of skills. We assume that jobseekers retain the vector of skills from their most recent occupation.

The probability of a successful match between a jobseeker with most recent occupation i and a vacancy in occupation j depends inversely on the distance between skill vectors characterizing occupations i and j.

We construct a distance between occupations i and j as a simple average of the absolute differences between skill level requirements for i and j across S possible skills,

$$distance_{ij} = \frac{1}{S} \sum_{s=1}^{S} |s_i - s_j|, \tag{7}$$

where s_i is normalized such that $s_i \in [0, 1]$.

The probability of a match between jobseeker i and vacancy j is the inverse function of the distance between i and j:

$$P_{ij} = 1 - distance_{ij}. \tag{8}$$

As the distance between the skill set possessed by jobseeker i and the skill set required by vacancy j increases, the probability that a match forms from their meeting decreases. Note that we scale $distance_{ijt}$ to lie between zero and one so that the probability lies between zero and one.

There are alternative concepts of the distance between multidimensional vectors of skills. Gathmann and Schönberg (2010) use the angular separation or un-centered correlation of the observable vectors to measure the distance between two multidimensional-skill vectors. Our measure of the distance between skill sets is the same as the one used by Macaluso (2017). Macaluso (2017) constructs a measure of skill remoteness, which measures the distance between worker's current occupation and the rest of the occupations in the city in which others are employed.

Alternatively, we can construct empirical probabilities of a successful match between jobseeker i and vacancy j from the data on jobseekers' transitions between occupations. The empirical method will produce $P_{ij} > P*_{ij}$, where $P*_{ij}$ is a latent level of the match probability below which no match takes place.

3 Data

3.1 Datasets

We use data on occupational skills from the U.S. Department of Labor's O*NET 22.0 database. Our jobseeker data are from the Current Population Survey (CPS). Our vacancy data are from the Conference Board's Help-Wanted Online dataset (HWOL) from HAVER. In the current version, we use monthly data from May 2005 to August 2018.

For each occupation, O*NET assigns ratings to a set of skills. O*NET rates the importance of 35 skills on a relative scale. Examples of the skills include reading comprehension, writing, critical thinking, coordination, etc. We merge skills data to the jobseeker and vacancy datasets through occupations. The HWOL occupational data are available to us on 2-digit occupational level. We convert the 8-digit SOC codes in O*NET to 2-digit occupation codes, taking the simple average, in order to merge the skills data with HWOL data and

CPS data.

3.2 Skills

There are two rating measures in O*NET. One measure rates the "level" used of each skill to a specific occupation on a scale of 0 to 7 and another measure rates the "importance" of each skill on a scale of 1 to 5. We use the "level" measure and rescale it so that it lies on the unit interval, but results are robust to the "importance" measure. Using equation (7), we calculate the distance between all occupation pairs based on their 35-element skill vectors. We construct probabilities P_{ij} by subtracting theses distances from one so that low distances correspond to high probabilities and high distances correspond to low probabilities.

3.3 Abilities and Knowledge

O*NET includes data on dimensions of occupations other than skills, such as "Abilities" and "Knowledge." The abilities dimension lists 52 elements including oral comprehension, spatial orientation, finger dexterity, and reaction time. The knowledge dimension lists 33 elements including design, physics, law and government, and transportation. We test robustness to using these alternative dimensions when calculating the distance between all occupational pairs in equation (7).

3.4 Jobseeker Occupations

We consider two definitions of jobseekers: (1) the unemployed, which is the narrow definition because two thirds of new hires from nonemployment are from out of labor force (Hornstein, Kudlyak, and Lange (2015)), (2) the nonemployed which is a broader definition because it includes the unemployed and those out of the labor force.

We assign skills to jobseekers based on their most recent occupation. We calculate shares of jobseekers (unemployed and nonemployed) in each occupation using the CPS. We restrict attention to prime-age respondents, who are between 25 and 54 years of age. Of the unemployed, 84 percent of our sample has a valid occupation entry. Only 2.4 percent of the sample labeled out of the labor force has a valid occupation entry. These missing occupations take place for two reasons: (1) we do not have information on previous occupation for new labor market entrants (students etc.), and (2) we do not have recent employment information for individuals who were employed a while ago. To address these missing occupations, we use the matched structure of the CPS. We look across the 8 months individuals are surveyed. First, we carry occupations forward: if a respondent has a valid occupation entry in month $t-1$, but not in t, we use the occupation in month $t-1$ for t. Second, we carry occupations backwards: if a respondent has a valid occupation in $t+1$, but invalid in t, we use $t+1$ for t. This procedure results in 89 percent of the nonemployed having valid occupations between 2005 and 2018 compared to 80 percent before the adjustment. The vast majority of respondents who do not have a valid occupation entry (and we therefore drop), are out of the labor force for all of the eight survey months. A handful are unemployed for all of the eight survey months, or some combination of unemployed and out of the labor force. We do not worry much about these dropped observations because these individuals are very marginally attached the labor force.

4 Skill Supply and Demand

Figure 1 plots the average vacancy skill profile (black) and average unemployed skill profile (blue) for the entire sample period, May 2005 - August 2018. Skills are ordered according to decreasing tightness (i.e. vacancy-unemployment ratios). That is, we order skills based on the ratio between the black and blue bars which is not plotted. On the top, programing has the most demand relative to supply (i.e its tightness ratio is large) and on the bottom, repairing has the least demand relative to supply (i.e. its tightness ratio is small). This figure averages across all occupations using vacancy shares and unemployment shares as weights. Put differently, this is how the average skill profile of an unemployed workers compares to the average skill profile of a vacancy.

Perhaps unsurprisingly, the three tightest skills are programming, science, and technology design. The three slackest skills are repairing, equipment maintenance, and operation monitoring.

Figure 2 plots tightness ratios for the three tightest and three slackest skills over time. The horizontal line represents where the tightness ratio is one, meaning average supply and demand of these skills is equivalent. The higher type skills are all above one, indicating demand for these skills exceed supply for the entire sample period. Conversely, the lower type skills are all below one, indicating the supply of these skills exceed demand.

Figure 2 also shows movements in these six selected skills over the business cycle. Leading up to and during the recession, demand of the high skills increasingly exceed supply, while supply of the low skills increasingly exceeded demand. This flipped following the recession: the high skills became less tight and the low skills became tighter. We use the narrow definition of jobseeker, but results are robust to calculating vacancy-*nonempoyment* ratios instead of vacancy-*unemployment* ratios.

5 Results

5.1 Aggregate Skill Match Probability

Figure 3 plots the aggregate skill match probability using O*NET data to calculate the likelihood a meeting turns into a match. The three lines illustrate alternative dimensions of occupational heterogeneity, namely skills, knowledge, and abilities. Here, we plot the measure constructed using only unemployed jobseekers but things look similar when the jobseeker pool is extended to include all nonemployed jobseekers (unemployed and OLF). As expected with the heterogeneity, Figure 3 shows that the aggregate skill match probability, regardless of the dimension, is below unity. That is, the aggregate matching efficiency μ_t is lower than the aggregate meeting efficiency ξ_t. The skill match probability using O*NET data hovers between 85 an 91 percent during 2005-2018.

The data on vacancies by occupations covers only one recession and thus discerning a cyclical pattern is challenging. Figure 3 seems to suggest that the aggregate skill match probability declines during recessions and then recovers during recoveries. After 2015, the aggregate skill match probability has tapered off.

Figure 4 uses empirical transition probabilities between occupations to measure whether a meeting turns into a match. Here, the average skill match probability is about 14 percent which is much lower than in Figure 3. This is because the likelihood a farmer becomes a lawyer after stumbling across an opening for a lawyer is essentially zero in the data, while according to O*NET there is some skill overlap. Nevertheless, the procyclicality persists, suggesting there was a robust decline in aggregate skill match probability during the Great Recession and a robust recovery.

5.2 Contribution to Aggregate Matching Efficiency

We can now decompose changes in the aggregate job finding rate into changes due to the aggregate skill match probability, the aggregate labor market tightness, and the residual term - the aggregate meeting efficiency using (5):

$$\ln \lambda_t = \ln \xi + \ln \phi_t + \alpha \ln \left(\frac{V_t}{U_t} \right). \tag{9}$$

We have data on the aggregate job finding rate from the CPS, the V-U ratio from HWOL and FRED and the aggregate skill mismatch component from Section 5.1. We set the elasticity parameter $\alpha = 0.5$. With this data we solve for the residual component of the matching function, $\ln \xi_t$. The solid line in Figure 5 plots the log of aggregate matching efficiency μ_t and the dashed lines plot its two components: aggregate skill match probability (red) and the residual (dashed black).

Figure 5 uses the skills dimension of O*NET to measure occupational heterogenity when calculating the distance between all occupational pairs, and the resulting probability a meeting between a job opening and job seeker results in a match P_{ij}. Figure 6 plots similar objects but uses empirical transitions to measure occupational heterogenity and calculate P_{ij}. Both figures show that the residual term closely mimics movements in aggregate matching efficiency. Aggregate skill match probability is responsible for the difference between the two black lines. Although aggregate skill match probability can account for a sizable share of the level of aggregate matching efficiency, it does not account for much of the cycilcial fluctuations.

6 Conclusion

We propose a theoretical framework to unpack (some of) the black box of the the aggregate matching efficiency, namely the part responsible for skill (in)compatibility which we term aggregate skill match probability. We find that aggregate skill match probability is procycical regardless if it is measured from a theoretical concept of distance between occupations or from empirical transition rates. However, we find that deviations from its long-run average are small compared to the changes in aggregate matching efficiency.

References

Aum, S. (2017). The rise of software and skill demand reversal. Technical report.

Autor, D. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–97.

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

Barnichon, R., M. Elsby, B. Hobijn, and A. Şahin (2012). Which industries are shifting the beveridge curve. *Monthly Lab. Rev. 135*, 25.

Barnichon, R. and A. Figura (2015). Labor market heterogeneity and the aggregate matching function. *American Economic Journal: Macroeconomics 7*(4), 222–49.

Davis, S. J., R. J. Faberman, and J. C. Haltiwanger (2013). The establishment-level behavior of vacancies and hiring. *The Quarterly Journal of Economics 581*, 622.

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

Gathmann, C. and U. Schönberg (2010). How general is human capital? a task-based approach. *Journal of Labor Economics 28*(1), 1–49.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2015). Multidimensional skill mismatch. Technical report, National Bureau of Economic Research.

Hall, R. E. and S. Schulhofer-Wohl (2018). Measuring job-finding rates and matching efficiency with heterogeneous job-seekers. *American Economic Journal: Macroeconomics 10*(1), 1–32.

Herz, B. and T. Van Rens (2018). Accounting for mismatch unemployment.

Hornstein, A. and M. Kudlyak (2016). Estimating matching efficiency with variable search effort.

Hornstein, A. and M. Kudlyak (2017). Generalized matching functions and resource utilization indices for the labor market.

Hornstein, A., M. Kudlyak, and F. Lange (2015). Measuring resource utilization in the labor market.

Korinek, A., J. A. Mistiaen, and M. Ravallion (2007). An econometric method of correcting for unit nonresponse bias in surveys. *Journal of Econometrics 136*(1), 213–235.

Macaluso, C. (2017). Skill remoteness and post-layoff labor market outcomes.

Martellini, P. and G. Menzio (2018). Declining search frictions, unemployment and growth. Technical report, National Bureau of Economic Research.

Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature 39*(2), 390–431.

Pissarides, C. A. (2000). *Equilibrium unemployment theory.* MIT press.

Şahin, A., J. Song, G. Topa, and G. L. Violante (2014). Mismatch unemployment. *American Economic Review 104*(11), 3529–64.

Sedláček, P. (2016). The aggregate matching function and job search from employment and out of the labor force. *Review of Economic Dynamics 21*, 16–28.

Figure 1: Average Vacancy-Unemployment Ratios by Skill: 2005m5-2018m8

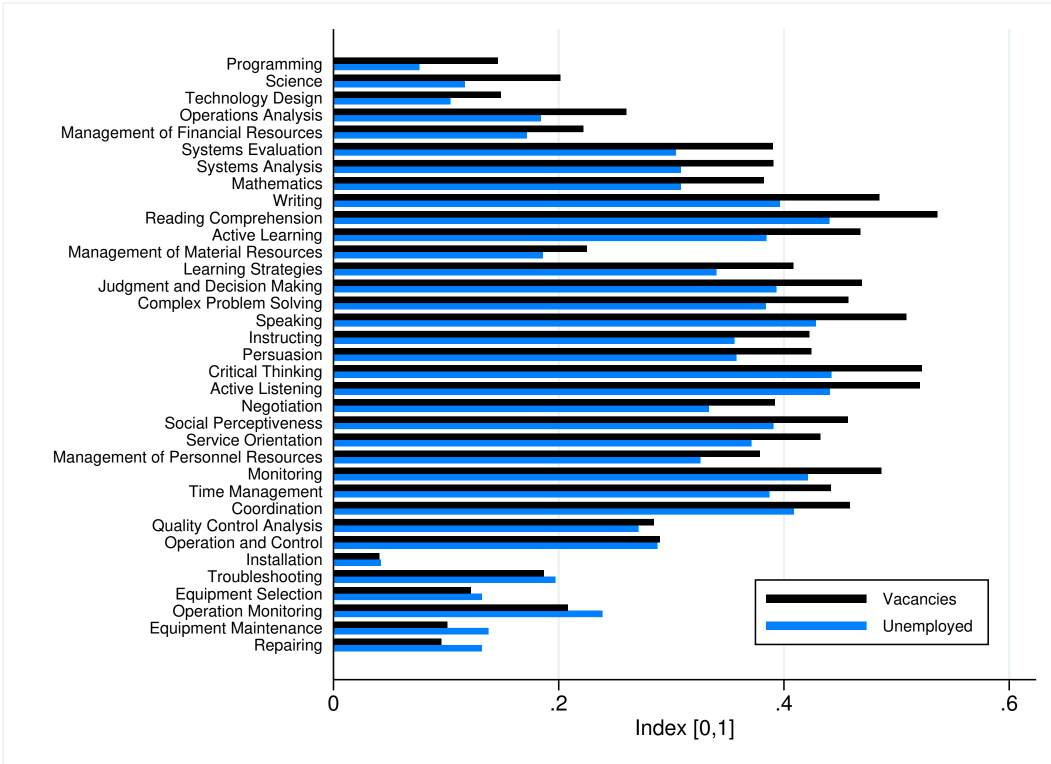

Figure 2: Average Vacancy-Unemployment Ratios by Skill Over Time

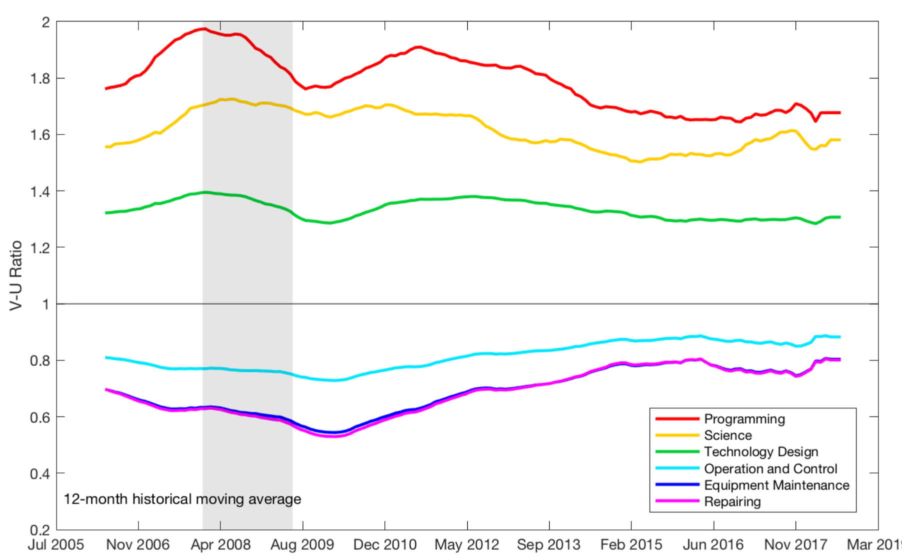

Figure 3: Aggregate Skill Match Probability with Different Dimensions

P_{ij} Constructed from O*NET Distance

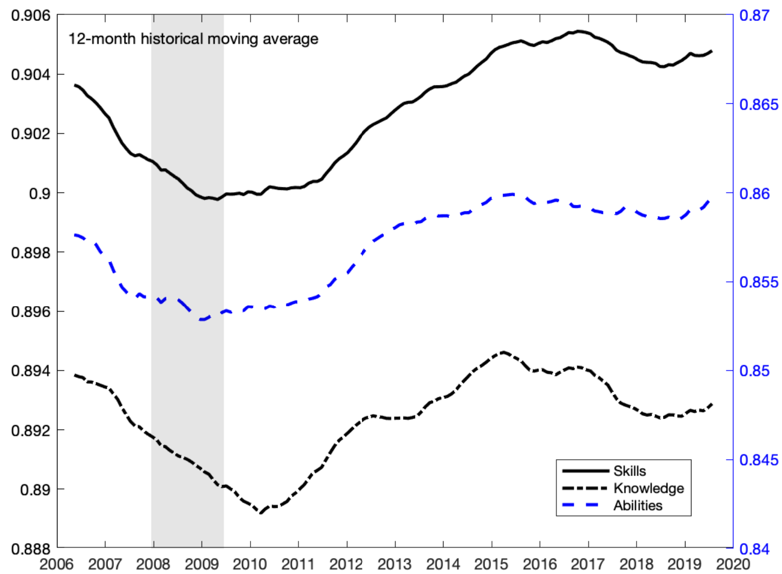

Figure 4: Aggregate Skill Match Probability

P_{ij} Constructed from Empirical Transition Rates

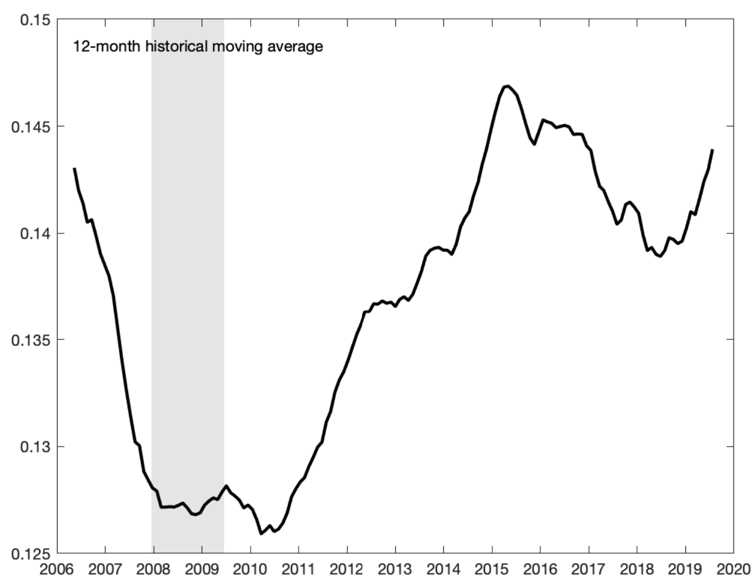

Figure 5: Decomposition of the Aggregate Matching Efficiency

P_{ij} Constructed from O*NET Distance

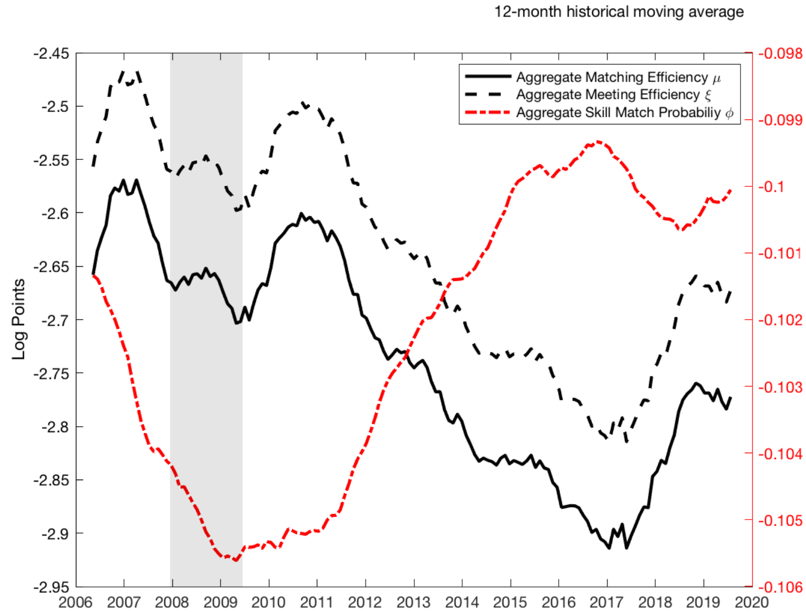

Figure 6: Decomposition of the Aggregate Matching Efficiency

P_{ij} Constructed from Empirical Transition Rates