The Convergence of Occupations: Evidence from Online Job Posts *

Alexis Antoniades
Georgetown University Qatar

Manolis Chatzikonstantinou
Georgetown University Qatar

Mohammed Shahmeer Ahmad
Qatar Computational Research Institute

January 29, 2026

Abstract

We document, in the past decade, a reversal in the trend of increasing wage inequality and declining labor mobility patterns across occupations. This is despite the rapid increase in technological progress and increase in technical skills demanded, as evident in online job advertisements. We argue that rapid technological progress over the past decade is transforming occupations and driving them closer together, leading to wage convergence and increased mobility. To study this, we introduce a novel measure of occupational distance based on job postings that allow us to represent each occupation as a vector of skills with corresponding weights that evolve over time. We compute bilateral distances between occupations using vector distance measures and find a steady decline in these distances, indicating occupational convergence. We show that this convergence has led to a decline in wage differences across occupations and increased bilateral labor mobility. Our results are not driven by labor market tightness or changes in minimum wages showcasing a new force that may affect wage inequality in the long term. Using microdata from the CPS, we show that the effects are larger for low-wage workers.

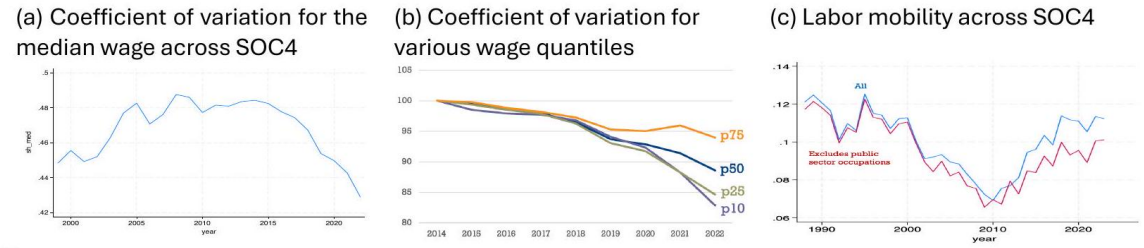
*Contact author: Alexis Antoniades, Georgetown University Qatar, aa658@georgetown.edu; Emmanouil Chatzikonstantinou, Georgetown University Qatar, ec1269@georgetown.edu; These views are our own and not those of any affiliated institutions.

1. Introduction

Skill-biased technical change, driven by technological progress favoring certain types of workers and occupations, has been a driving force behind job polarization, declining labor mobility, and rising inequality since the 1980s. However, after 2010, amid further technological advancements driven by innovations related to computing and information technology, such as big data analytics and artificial intelligence, we observe a convergence in wages across occupations along with a rise in labor mobility.

This trend is illustrated in Figure 1a, which plots the coefficient of variation for median wages across occupations, (defined at the standard occupation classification 4-digit level), from 1999 to 2022. Wages diverged up until 2008, plateaued thereafter, and have been converging since 2014. A similar pattern emerges when examining the coefficient of variation across occupations at different wage quantiles (Figure 1b). Finally, labor mobility across occupations, measured using the longitudinal responses of workers in CPS declined until 2010 but has increased since then, as shown in Figure 1c.

Figure 1: Figure 1 - Wage convergence and mobility across occupations in the U.S.



We argue that technological progress does not only transform tasks and skills performed in each occupation but may also bring occupations closer together, leading to wage convergence and increased occupational mobility. As occupations evolve, tasks and skills become obsolete and disappear (e.g., typewriting and teller transactions), while new skills (e.g., data analytics and Python) that are highly transferable emerge. If a significant set of highly transferable and in demand skills exists across occupations, we would expect wages to converge and labor mobility to increase as occupations become more homogeneous-effectively reducing the distance between occupations. Drawing on the seminal work of Lucas and Prescott (1974), which depicts occupations as islands with workers distributed across them, our hypothesis is that over the past decade these islands have moved closer together, effectively reducing the distance between them.

To test this hypothesis, we introduce a novel measure of occupational distance

based on online job vacancy data. This measure enables us to assess how similar or dissimilar two occupations are and how this distance evolves over time and across space, using high frequency and current data compared to occupation surveys that tend to be updated irregularly. Specifically, by analyzing hundreds of millions of job vacancies posted in the US between 2014 and 2022 and taking stock of the skills requirements provided in the job description, we represent each occupation as a vector of skills with corresponding weights. We then apply similarity analysis to calculate the distance between these vectors, effectively quantifying the similarity between occupations. Analyzing these data over time, we can determine whether occupations have converged.

We begin by applying our proposed methodology to measure bilateral distances between US occupations for the years 2014-2022.[1] Using the detailed standard occupation classification in the United States(6-digit SOC), which comprises 1,326 groups, we calculate all bilateral distances between occupation pairs (e.g., Purchasing Managers, 11-3061 and Library Technicians, 25-4031) for each year. This allows us to quantify and observe how these distances evolve over time. Our first finding is that occupations are indeed converging. The distribution of bilateral distances between 2014 and 2022 has shifted to the left, reflecting a reduction in distances among occupations. Furthermore, both the average and median bilateral distances have consistently decreased each year within this period.

Before proceeding to examine the impact of declining bilateral distances on wages and labor mobility, we provide some evidence supporting the validity of the skill-based distance measure. First, as expected, we find that the computed distances between detailed(SOC 6-digit) occupations within the same broad occupation groups- SOC 2-digit (e.g., Managers, 11) are smaller on average than distances across two digit SOCs(broad occupation groups). Second, the smallest distances observed between occupations from different broad occupation groups involve occupations with nearly identical titles (e.g., Purchasing Managers with 11-3061 and Wholesale and Retail Buyers or 13-1022; Industrial Engineering Technologists and Technicians and 17-3026 and Industrial Machinery Mechanics, 49-9041). Third, we demonstrate a negative correlation between the computed distances and labor mobility across occupations: labor mobility is lower for occupations further apart. These three tests confirm that our measure of distance is intuitive and robust and captures relevant information for occupation classification and labor market dynamics.

Between 2014 and 2022, we demonstrate that distances between occupations decreased, wages converged, and labor mobility increased. To move beyond simple correlations, we leverage the richness of the data to test whether occupation pairs that

[1]Our data start from 2010, and for robustness we replicate our analysis for our full sample. But online job vacancy data are more reliable post 2014.

experienced greater convergence also exhibited higher wage convergence and greater increases in labor mobility compared to pairs that did not. Our regression analysis confirms these relationships. We find a positive and statistically significant coefficient of the estimated bilateral skill-based distance on wage inequality between occupation pairs, indicating that smaller distances are associated with reduced wage disparities. Similarly, we observe a negative and statistically significant coefficient of distance on labor mobility, suggesting that labor mobility increases as occupations become more similar.

Our distance measure based on the raw Lightcast data is based on free-form keywords from the parsed text of online job vacancies. This means that our measure of distance can potentially be affected by additions and deletions from this vocabulary of skills. We introduce as a result a new methodology to standardize occupational skills within the O*NET database. Using the definitions of occupational characteristics from the comprehensive O*NET database, we replace the vector of raw skills in Lightcast for every job post with a vector of O*NET characteristics. The core of the approach relies on the power of Sentence Transformers to generate semantically rich embeddings. In particular, we first source for each keyword its definition from Wikipedia, and then we move to find the nearest neighbor in terms of meaning to an O*NET characteristic. Then, we recalculate the distance measure based on vectors of O*NET keywords, and confirm our findings with both raw keywords and O*NET characteristics.

Our work makes three key contributions: Theoretically, we provide a mechanism that helps reconcile the empirical evidence of declining across-occupation inequality and rising labor mobility with the predictions of existing occupational models. Methodologically, we introduce a novel skill-based measure of distance, which can also be applied to other groupings, such as industries, firms, and cities and create an algorithm to standardize raw keywords to skills, tasks ,knowledge and work activities in the O*NET database. Empirically, we provide evidence of a reduction in bilateral distance measures across occupations since 2014 and confirm that as occupations have converged in their skills requirements, this has led to wage convergence across occupations and an increase in occupational mobility.

The remainder of the paper is organized as follows. In Section 2 we present a short review of the literature. The data and methodology are presented in Section 3 and 4 respectively. Section 5 presents empirical findings on the distance between occupations, and Section 6 tests the implications of convergence on relative wages and occupational mobility. Section 7 concludes.

2. Literature Review

Our work contributes to several strands of the literature. Firstly, our paper builds on a growing body of literature that leverages online job postings as a comprehensive snapshot of the labor demand landscape-a strategy pioneered by researchers over the past two decades (Kuhn & Skuterud, 2004; Marinescu, 2017; Deming & Kahn, 2018). In this work, we provide a novel application of online job vacancy data by measuring the similarity between occupations and examining its implications for wage inequality and labor mobility.

Our study explores a new dimension of online job vacancy data, complementing the work of Modestino, Shoag, and Ballance (2020) and Deming and Kahn (2018), which highlights the importance of such data in understanding trends like upskilling, the evolving task content of occupations, and the growing complementarity between cognitive and computer skills in professional roles.

Secondly, our paper also contributes to the literature examining the dynamics of wages and inequality. Hurst and Kahn (2023) explore the stagnation of median real wages and its implications for income inequality, focusing on challenges in wage measurement, wage cyclicality, and the role of monopsony power in labor markets. Complementing their work, we place greater emphasis on the measurement of occupations and how their evolving nature impacts labor markets and wage dynamics.

By introducing a novel perspective on wage determination, our analysis also complements studies on occupational polarization, such as Autor and Dorn (2013), the effects of occupational mobility (Kampourov and Manowski, 2008), the decline of labor unions (Western and Rosenfeld, 2011), and the changing boundaries of firms and their impact on wages (Bloom et al., 2018). Furthermore, our findings on wage inequality align with recent evidence by Aeppli and Wilmers (2022), who document the stagnation of earnings inequality

Thirdly, our focus on the analysis of the skill content of occupations and especially the transformation of occupations over provide further insights into labor market shifts shaped by technological advancements, particularly through the lens of Skill-Biased Technical Change (SBTC). The SBTC hypothesis suggests that technological progress, especially in information and communication technology, has disproportionately benefited high-skilled workers over low-skilled workers, contributing to wage disparities and shifts in employment patterns.

Early work by Autor, Katz, and Krueger (1998) highlighted the role of computerization in increasing demand for skilled workers, while Acemoglu (2002) provided a theoretical framework to explain how innovations can disproportionately enhance the productivity of skilled workers, thereby driving wage inequality. In contrast, our paper demonstrates that while these dynamics are present, they may be mitigated when technical change transforms occupations in ways that increase occupational mobility.

Similarly, Card and DiNardo (2002) questioned the primacy of the SBTC hypothesis as a driver of rising wage inequality, emphasizing the role of institutional factors. Our findings contribute to this ongoing debate by highlighting the interplay between technological change, occupational transformation, and labor market outcomes.

3. Data Sources

Our primary dataset consists of hundreds of millions of online job vacancies (OJVs) posted in the US between January 2011 and March 2022 [2]. We consider data for the period 2014-2022 due to its representativeness in our main specification, but we also confirm our results for the period 2011-2022. The online job vacancies data, provided by Lightcast (formerly Burning Glass Technologies), has been widely used by scholars over the past decade. Each day, Lightcast collects OJVs from 40,000 to 50,000 sources, including web boards, company websites, forums, and other directories. These postings typically include details about the job (e.g., title, description, requirements), the ideal candidate (e.g., qualifications, experience, education), the company (e.g., name, industry), and the location (e.g., address). Lightcast processes this unstructured text to create a structured dataset with over 70 parameters for each post, covering aspects such as skill content, company, location, industry, and occupation.

Of particular importance to our analysis are the data on location, occupation, and skill composition. Location information allows us to assign each job posting to a specific geographical area using FIPS codes (county-level identifiers), which can then be aggregated to metropolitan statistical areas (MSAs), commuting zones, states, or other geographic units of interest. Occupation data is provided at the detailed Standard Occupational Classification (SOC 6) level, encompassing 1,326 distinct occupational groups (see Table 1, columns 1-6).

Lightcast processes each job posting into a list of keywords and phrases, referred to as skills. Examples of these skills include Communication Skills, Python, Bitcoin, 3D Printing, Movement Disorders, Plumbing Repairs, and Copper Plumbing Installation. In the first quarter of 2014, the dataset contained 13,989 unique skills, which increased to 14,483 unique skills by the first quarter of 2022. To illustrate how these skills data are useful to analyze changes in occupations over time, we conduct two preliminary exercises. A more formal analysis of skills' distance calculation follows in Section 4.

In the first exercise, we compile a list of skills related to Al and a list related to Data Analytics. These skills are selected based on a combination of consulting reports and academic papers, with the complete lists provided in Appendix A. For each broad

[2]Data exist also for year 2007 and 2010 but they were irregularly collected with large differences in the number of vacancies collected in the two years with the post 2011 period. As a result, we exclude these two years from our sample period.

Table 1: Summary Statistics

Table 1 – Online Job Vacancies Descriptive Statistics

Occupations				Online Job Vacancies (in thousands, 000s)		Occupational Characteristics				
							AI Intensity		Data Analytics	
SOC 2	SOC2 Name	SOC4	SOC6	2014 Q1-Q4	2022 Q1	High Nonroutine Cognitive	2014 Q1-Q4	2022 Q1	2014 Q1-Q4	2022 Q1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	23 Total occupation groups (SOC 2)	98	1,326	24,261	12,595	0.45	0.5%	4.6%	2.3%	8.5%
SOC2 occupation group										
11	Management	4	68	2,694	1,522	0.90	0.4%	1.9%	5.1%	10.6%
13	Business and Financial Operations	2	58	1,506	778	0.78	0.4%	1.4%	8.9%	16.3%
15	Computer and Mathematical	2	33	2,528	1,170	1.00	3.3%	15.8%	19.5%	38.7%
17	Architecture and Engineering	3	57	752	333	0.86	0.8%	4.8%	4.6%	10.2%
19	Life, Physical, and Social Science	5	73	253	134	0.95	0.6%	2.3%	9.1%	16.1%
21	Community and Social Service	2	24	283	169	0.79	0.1%	0.2%	2.9%	3.8%
23	Legal	2	12	203	85	0.71	0.1%	0.6%	1.6%	4.5%
25	Educational Instruction and Library	5	94	610	324	0.93	0.1%	0.4%	1.8%	3.0%
27	Arts, Design, Entertainment, Sports, and Media	4	57	619	277	0.72	0.2%	0.7%	1.7%	3.8%
29	Healthcare Practitioners and Technical	3	99	2,381	1,400	0.73	0.1%	0.1%	1.2%	1.7%
31	Healthcare Support	3	24	511	330	0.13	0.0%	0.1%	0.4%	0.7%
33	Protective Service	4	38	265	160	0.53	0.1%	0.6%	1.4%	2.0%
35	Food Preparation and Serving Related	4	28	1,049	622	0.00	0.0%	0.0%	0.2%	0.4%
37	Building and Grounds Cleaning and Maintenance	3	14	323	264	0.00	0.0%	0.0%	0.3%	0.5%
39	Personal Care and Service	8	52	592	294	0.10	0.0%	0.0%	0.2%	0.5%
41	Sales and Related	5	37	3,031	1,248	0.38	0.1%	1.0%	1.4%	2.3%
43	Office and Administrative Support	7	101	2,567	1,360	0.14	0.1%	0.2%	3.7%	7.7%
45	Farming, Fishing, and Forestry	4	22	21	11	0.00	0.0%	0.2%	0.7%	2.6%
47	Construction and Extraction	5	98	282	159	0.08	0.1%	0.4%	0.5%	0.8%
49	Installation, Maintenance, and Repair	4	70	785	430	0.12	0.4%	0.8%	1.1%	1.9%
51	Production	9	159	755	374	0.10	0.1%	0.3%	0.9%	1.9%
53	Transportation and Material Moving	7	86	1,308	634	0.11	0.0%	0.1%	0.2%	0.7%
55	Military Specific	3	22	29	3	-	0.3%	1.7%	0.7%	5.5%

Source: Lightcast, Authors' calculations

Note: This table is included as an image. The text discusses its content, but the data to reproduce it in LaTeX is not available.

group of occupation(SOC 2), we count the number of job vacancies that mention each specific skill. This process is repeated for all skills available during the analyzed period, generating a skill vector for each SOC 2 group along with a corresponding measure of frequency, referred to as document frequency (df). To calculate the AI intensity of job vacancies in a particular SOC 2 group, we sum the df of all Al-related skills identified on our list and divide it by the total df of all skills (AI and non-AI). We apply the same methodology to compute a Data Analytics intensity.

In Table 1, Columns 8 and 9 report the AI intensity for 2014 and 2022Q1, respectively, while Columns 10 and 11 present the Data Analytics intensity. Additionally, we include the share of SOC6 occupations within each SOC2 group classified as High Nonroutine Cognitive (High NR-C). This classification is determined by comparing the Nonroutine Cognitive score of an occupation, based on Acemoglu and Autor (2021), to the median score for all occupations.

The first row of Table 1 presents aggregated data across all occupations. The dataset includes job vacancies allocated to 1,324 occupations at the SOC6 level, which can be aggregated into 98 SOC3 or 23 SOC2 groups. In 2014, the data set contained 24,261 observations, while in the first quarter of 2022, it included 12,595. Among these 1,324 occupations, 45% are classified as High Nonroutine Cognitive (High NR-C). Our calculations show that the AI intensity increased from 0.5% in 2014 to 4.5% in 2022, while the data analytics intensity increased from 2.3% to 8.5% over the same period.

In our second exercise, we move beyond a curated list of AI and Data Analytics skills to investigate broader trends. Specifically, we examine which skills, out of over 15,000 in the dataset, experienced significant changes between 2014 and 2022. We classify these skills into three categories:

1. Common Skills: Skills that existed in both 2014 and 2022 and experienced the highest growth.

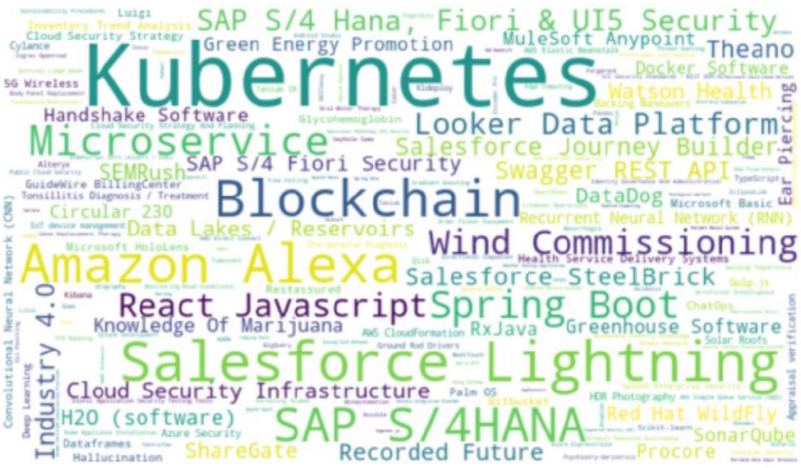
2. New Skills: Skills that did not exist in 2014 but gained significant prominence in 2022 (based on high document frequency).

3. Disappearing Skills: Skills that were important in 2014 (based on a document frequency threshold) but saw a decrease of 85% or more by 2022.

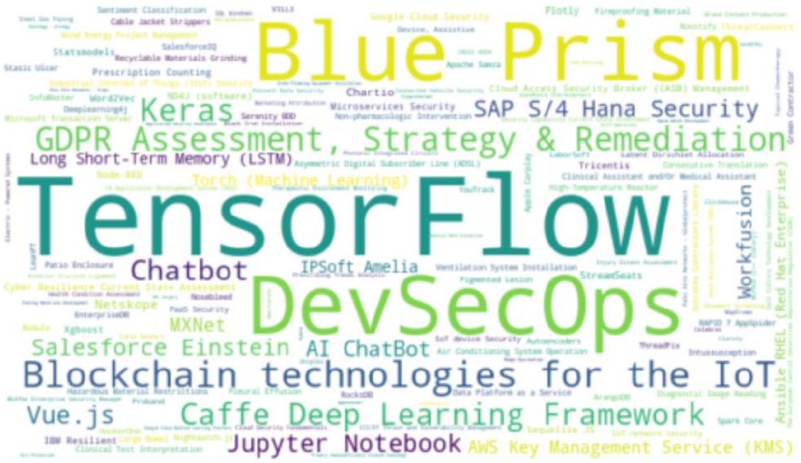
To visualize these trends, we present WordClouds for each of these categories in Figure 2a.[3] To further underscore our paper's motivation, we repeat this analysis for three specific occupational groups: Business Operation Specialists (SOC4 13-10),

[3]Source: Lightcast, Authors' calculations. Each word cloud visually summarizes the key terms and technologies that are prevalent in the Lightcast data. The size of each term corresponds to its

Figure 2a - Growing Skills and Fading Skills, 2014-2022



(a) Common: Skills with the highest growth in demand



(b) New: Skills in high demand, 2022 (not present in 2014)



(c) Disappearing: Skills with 85% or more drop in demand, sorted by 2014 importance

Note: Source: Lightcast, Authors' calculations. Each word cloud visually summarizes the key terms and technologies that are prevalent in the Lightcast data. The size of each term corresponds to its frequency or importance, with the first panel using only skills important in both 2014 and in 2022, the second panel concentrating in new skills, and the last panel using disappearing skills: those skills with 85% or more drop in demand.

Engineers (SOC4 17-20), and Lawyers, Judges, and Related Workers (SOC4 23-10). The results for these groups are displayed in Figure 2b.

Two key findings emerge from our analysis: First, within just five years, numerous new skills have emerged, most of which are technology-related. This trend is clear when examining the fastest-growing existing skills between 2014 and 2019, as well as the new skills that ranked highest in employer demand in 2019. These patterns hold true whether we analyze data across all occupations (Figure 2a) or focus on specific occupational groups (Figure 2b).

Second, Figure 2 reveals that the skills with the highest growth in demand tend to be common across the three distinct occupational groups. In contrast, the skills that are disappearing vary significantly between these groups. This indicates that while occupations evolve over time by shedding older, occupation-specific skills, they increasingly adopt new, occupation-agnostic skills.

This dynamic of skill entry and exit, as demonstrated in the word cloud analysis, suggests that occupations are converging. To formalize this observation, we introduce a measure of distance in the next section.

Lastly, the vocabulary of keywords from the parsed text of online job vacancies is not necessarily the vocabulary of skills describing occupations(or more specifically the vo skills, tasks ,knowledge and work activities of occupations). The O*NET framework is a long-running effort to provide such a standardized set of skills, with the intensity that each occupation is associated with every one of those. However, the O*NET database is not regularly updated and does not provide the optimal framework for calculating the occupational distance based on skills over time.

To make sure that our analysis is not affected by the fact that there are additions and deletions from the skills vocabulary in Lightcast, we introduce a new methodology to standardize occupational skills in Lightcast within the O*NET database.

frequency or importance, with the first panel using only skills important in both 2014 and in 2022, the second panel concentrating in new skills, and the last panel using disappearing skills: those skills with 85% or more drop in demand.

Figure 2b - Growing Skills and Fading Skills for Selected SOC4 Groups, 2014-2022



3.1 Matching Skill-to-O*NET

This subsection details the algorithmic approach developed for semantically matching Lightcast Online Job Posts "Skills" (keywords) to standardized occupational characteristics within the O*NET framework. The methodology leverages advanced natural language processing (NLP) techniques, specifically sentence embeddings and a two-stage retrieval-and-re-ranking architecture, to achieve high-precision and scalable skill-O*NET matching.

O*NET Database The O*NET (Occupational Information Network) program, maintained by the U.S. Department of Labor, serves as the source for this methodology. It is the main database that provides detailed information on occupational characteristics and worker requirements across occupations in the United States. The updating process of these occupational characteristics involves ongoing surveys of workers and expert input to create relevant task, skill and worker abilities definitions and provide the importance of each of these characteristics for every occupation. We use the descriptions of skills, tasks, knowledge and work activities to provide a robust and semantically rich target for the matching task on hand.[4] For example, some of the more than 150 data fields in O*NET that we use are; 2A1d: Speaking, 2C1e: Customer and Personal Service, 4A4b1: Coordinating the Work and Activities of Others.

We then use a natural language processing technique to transform the textual data into a vector numerical format that effectively preserves and represents semantic meaning. At the heart of this methodology lies the concept of embeddings, which are dense numerical representations (vectors) of textual units such as words, sentences, or entire documents. These vectors are designed to capture the semantic meaning of the information they represent.[5] This is in contrast to traditional text representation methods, such as bag-of-words models, often ignore syntax and deeper contextual relationships.

This capture of meaning is important for nuanced skill matching, where 16000 skill definitions necessarily implies subtle differences in phrasing, conveying significant semantic meanings. The use of transformer-based models, notably BERT (Bidirectional Encoder Representations from Transformers), has allowed semantic embedding generation. These models analyze words in context from both preceding and succeeding words. For the task of matching Lightcast keyword definitions with O*NET definitions, where terminology can be highly specialized, industry-specific, or where

[4] We downloaded the most recent iteration of the O*NET database (May 2025).

[5] For example, the embedding of the sentence "Today is a sunny day" will be numerically very similar to that of "The weather is nice at the moment," enabling quantitative comparison of their underlying meaning.

multiple synonyms might exist (e.g., "coding," "programming," "software development"), our semantic vector based algorithm enables the system to accurately match skill definitions like "PC Literacy" to an O*NET skill description such as "Working with computers".

In simple terms, here's how we're matching job skills: First, we're using a specific tool called SBERT to understand the meaning of skill descriptions, i.e. for the initial embedding generation phase. The sentence-transformers open-source library (SBERT) is chosen for its efficiency and ease of use in generating high-quality sentence embeddings from a wide array of pre-trained models. We chose a special type of model called a cross-encoder for the initial matching process. Think of this model as a smart matchmaker. Unlike other models that look at pairs of two skills separately and then compare them, a cross-encoder looks at all skills at the same time. This model gives us a relevance score for each pair of skills, telling us how well they match.[6]

For example, it would give a high score to "Microsoft Excel" and "Working with Computers" because they are very similar in meaning. After getting all the relevance scores, we sort the skills from highest score to lowest. After retrieving the best match for every skill in Lightcast, instead of associating every job post with a list of keywords from the parsed text, e.g. PC literacy,Detail-Oriented, and Microsoft Excel, we associate each job post with a list of O*NET occupation characteristics, in this example: Working with computers, Attention to Detail and Processing Information. This gives us a more consistent and organized way to describe the skills needed for a job. For example, "Microsoft Excel" from one list gets matched with "Working with Computers" from the other. This helps to standardize how we categorize skills across different job posts. A list of examples of the matches between skills in Lightcast and O*NET are provided in table 7 in the Appendix.

4 Methodology

In this section, we introduce a method to measure the distance between groups of workers based on the skill content of jobs, as derived from online job vacancies. Unlike

[6]For the initial embedding generation phase, the cross-encoder *stsb-distilroberta-base* model is selected, which is a cross-encoder that encodes the query and each candidate skill simultaneously in a single forward pass. This joint encoding allows the model to leverage attention mechanisms to analyze the relationships between vectors across the query-skill pair, leading to a more accurate assessment of relevance. Unlike bi-encoders which produce independent embeddings, cross-encoders typically output a single classification score (e.g., a relevance score) for every pair. In particular, for each of the candidate O*NET skills definitions, the cross-encoder model computes a relevance score for every query, i.e. every Lightcast skill definition. This is achieved by feeding the full list of encoded O*NET vectors and the candidate skill descriptions as a concatenated input to the cross-encoder.

the work done using data that do not change frequently as they rely on surveys– to measure similarity, our distance metrics can be time varying and can be also customized to any particular group (corpus) of interest, e.g. industries or cities.

For each group of interest g, we compile all online job postings related to that group at time t. A group can represent an occupation, city, industry, firm, or any other defined set. This generates a corpus C_{gt} of online job postings for group g at time t. Next, for each skill s[7] within the total set of skills available during that period, we calculate the number of documents in the corpus C_{gt} that mention the skill. This count is referred to as the document frequency (df_{st}). We then compute the intensity of each skill as the share of df_{st} relative to the sum of df_{st} values for all skills in the corpus. This process reduces the corpus into a skill vector with corresponding shares: $(c_1, c_2, \ldots, c_S)$.

To calculate the distance between two corpora, we represent them as vectors of corresponding shares $\mathrm{A} = (\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{s})$ and $B = (b_1, b_2, \ldots, b_S)$. We then compute the Euclidean distance between these vectors, defined as:

$$\text{Distance}(A, B) = \sqrt{\sum_{i}^{S} (a_s - b_s)^2}$$

In case we use Lightcast skills [8], given that there are over 14,000 skills, most of which have values of zero, we limit our analysis to the top 200 skills to reduce computational complexity. Specifically, every vector is sorted from the most to least frequent skills, and the weights of the top 200 skills are compared across the two vectors. In the next section, we demonstrate that the results are robust to varying this cutoff of 200 skills. For further robustness, we replicate the analysis for both Lightcast and O*NET characteristics using Cosine similarity as an alternative metric and results are very similar.

When the two corpora correspond to different occupations at time t, the Euclidean distance measures the difference in the skill mix of the two occupations. By repeating this calculation for a later period t + T, we can assess how the distance between the two occupations has changed over time, thereby determining whether convergence across occupations has occurred.[9]

[7] In our case, a skill can be 1. a Lightcast skill for the distance measures that rely on the vocabulary provided by Lightcast and 2. an O*NET occupational characteristic for the distance measures that are calculated after we restate the Lightcast online job posts to a prespecified list of occupational characteristics are described above.

[8] We have two types of vectors in general for every occupation, one based of document frequencies of raw keywords and one based of document frequency of O*NET characteristics after using our matching algorithm. As show in Appendix C, both lead to similar time series patterns and regression results.

[9] Additionally, by comparing the distance of an occupation across two points in time (i.e., corpus

This methodology is flexible and can be applied beyond occupations. It enables us to compute skill-based distance measures for various groups, such as cities, counties, industries, firms, or commuting zones, providing a versatile framework for analyzing transformations across different entities.

5. Analysis

To explore how relationships between occupations evolved the past decade, we analyze changes in occupational distances over time. This dynamic perspective reveals patterns of skill convergence among occupation, that we then try to understand its implication.

Following the computation of these distances, we first present validity checks to ensure the robustness and reliability of the results, and then in the next section, we study the implication of occupational characteristics convergence.

5.1 Bilateral distances over time across SOC6

We begin the analysis by computing bilateral measures of distance between occupations at the SOC6 level over time. For each year between 2014 and 2022, we use the OJV data to summarize the skill landscape of each occupation during the first quarter (Q1). Our focus on Q1 is driven by data availability, as 2022 data exist only for that quarter. Using our proposed similarity measure, we calculate the bilateral distances between all SOC6 occupations for each year.

To illustrate how occupational distances evolve, Figure 3 presents kernel density plots of the computed bilateral distances for the years 2014, 2018, and 2022. The results show a clear leftward shift in the distribution over time, while the shape of the distribution remains largely unchanged. This indicates that occupations are converging-distances between them are decreasing, and they are becoming more similar in terms of skill content.

We further demonstrate convergence in Figure 4, which reports the average and median bilateral distances across all SOC6 occupation pairs for each year. Both measures consistently decline over time, reinforcing the finding that occupational distances are narrowing. Importantly, this result is robust to the choice of distance metric: when we use O*NET occupational classifications or original raw Lightcast data and when the Cosine Similarity method is applied instead of the Euclidean

C_{gt} and $C_{g(t+T)}$), we can measure the extent of transformation within that occupation over time. A larger distance indicates greater transformation. To distinguish this self-comparison from the bilateral distance between two groups, we refer to this metric as selfdistance.

measure, the conclusions remain unchanged (see Appendix C). These findings suggest a clear trend of skill convergence across occupations.

We further explore whether convergence depends on the nature of tasks performed within an SOC6 occupation. To do this, we classify occupations depending on their nature of tasks into Manual and Low Manual groups, following the framework of Acemoglu and Autor (2021). We repeat by 2. High Manual or not, 3. High Abstract or not and 4. Highly offshorable or not. The results, presented in Figure 5, reveal that high manual occupations and less abstract occupations tend to be more distance, but there is a broad trend reflecting greater skill similarity within all groups.

5.2 Validity of Computed Distances and Robustness

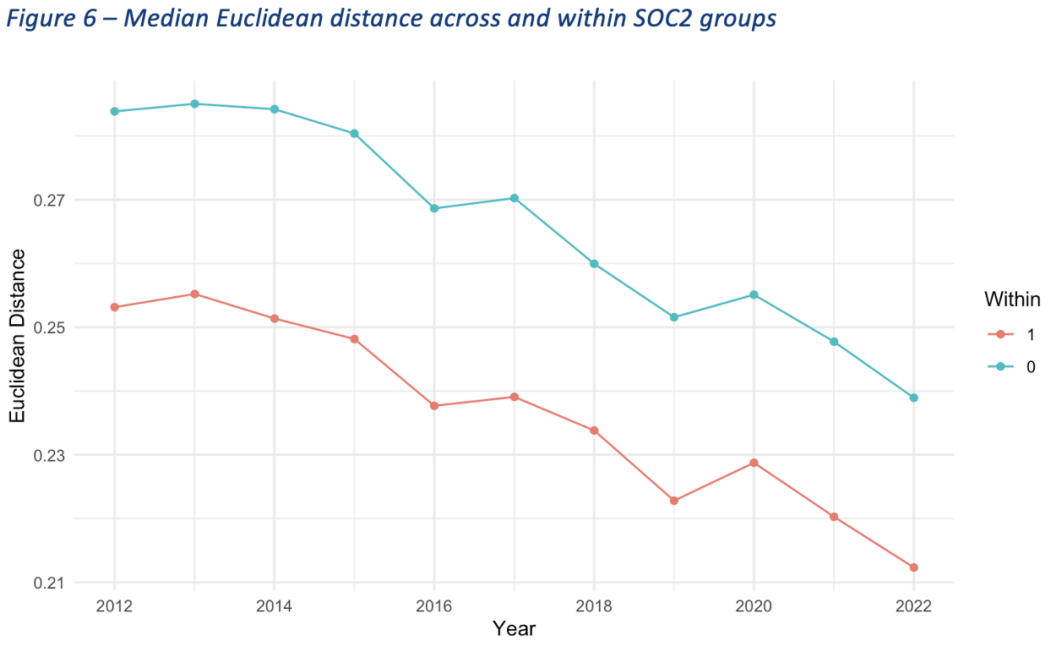
Before we proceed any further, we want to ensure that this newly introduced skill-based measure of distance does indeed capture similarity. We show that it does using three different approaches. First, we expect that occupations within the same SOC2 group (e.g. Management) should be more similar (shorter distance) compared to occupations across SOC2 groups. We show this in Figure 6 below, where we plot the average bilateral distance d_{ij} for SOC6 occupations i, j within the same SOC2 group (Within) and the average distance when i, j come from different SOC2 groups (Across). The figure demonstrates that computed Within distances are always smaller than Across, as expected, and that both distances decrease over time.

Next, we rank the bilateral distances d_{ij} across SOC2 groups from smallest (most similar) to largest (most dissimilar) and present the top 25 occupational pairs in Table 2. The results reveal that the closest pairs often have nearly identical names or highly overlapping job roles. For example, the smallest distance is observed between *Purchasing Managers* (`11-3061`; SOC2 = Management) and *Wholesale and Retail Buyers* (`13-1022`; SOC2 = Business and Financial Operations), highlighting their strong skill alignment. Similarly, *Industrial Engineering Technologists and Technicians* (`17-3026`; SOC2 = Architecture and Engineering) are most closely related to *Industrial Machinery Mechanics* (`49-9041`; SOC2 = Installation, Maintenance, and Repair). Another notable pair is *Library Technicians* (`25-4031`; SOC2 = Educational Instruction and Library) and *Library Assistants, Clerical* (`43-4121`; SOC2 = Office and Administrative Support), which reflect highly similar roles despite belonging to different SOC2 categories.

The ability of our methodology to identify and link such closely related occupations-based solely on skill information extracted from job advertisements-demonstrates the robustness and accuracy of our approach. The fact that these relationships emerge naturally from the data further validates the reasonableness and practical utility of the computed bilateral distances.

For the final validity test we explore whether pairs of occupations that we mea-

Figure 4: Figure 6: Average Bilateral Distance Within and Across SOC2 Groups



sure to be closer together tend to have higher bilateral labor mobility. To do that, we regress the computed distance d_{ij} with mobility data from Schubert et al (2024) and from the CPS. Schubert et al (2024) use online worker profiles to estimate labor transitions for pairs of occupations and provide estimates of bilateral transition shares for each of 840 SOC 6-digit origin occupation to each 6 digit destination occupation. These estimates are averages for the period 2002-2016 at the national level. We also confirm these results using data of transition shares for each occupation pair in CPS for the same time period, mostly preceding our sample of distance metrics. The results from CPS are reported in column 5 and because CPS reports occupations at a more aggregated level (305 levels), the number of observations falls. In CPS, transition shares are calculated using the current occupation reported by each worker and the occupation of the previous year as reported by each worker the day of the survey (longitudinal measurement). The results are reported in Table 3. Bilateral distance is negatively correlated with measures of transitional shares across the different sample, and at the more granular level (columns 1-4) this negative correlation with occupational distance is larger, compared to the more aggregated level(column 5). In addition, when pairs of occupations with zero transitions are excluded(column 4), the estimated coefficient becomes even larger.

The three validity tests presented above demonstrate that the proposed measure

Table 2: Top 25 Closest Occupation Pairs Across SOC2 Groups

Table 2 - Top 25 SOC6 pairs with the lowest distance, across SOC2 groups, 2014

Rank	Occ 1	Occ 2	Occupation 1 Name	Occupation 2 Name
1	11-3061	13-1022	Purchasing Managers	Wholesale and Retail Buyers, Except Farm Products
2	17-3026	49-9041	Industrial Engineering Technologists and Technicians	Industrial Machinery Mechanics
3	25-4031	43-4121	Library Technicians	Library Assistants, Clerical
4	13-1151	11-3131	Training and Development Specialists	Training and Development Managers
5	43-4041	13-2041	Credit Authorizers, Checkers, and Clerks	Credit Analysts
6	11-9033	21-1012	Education Administrators, Postsecondary	Educational, Guidance, and Career Counselors and Advisors
7	11-9151	21-1093	Social and Community Service Managers	Social and Human Service Assistants
8	13-1141	11-3111	Compensation, Benefits, and Job Analysis Specialists	Compensation and Benefits Managers
9	11-1011	19-3041	Chief Executives	Sociologists
10	15-2031	19-1042	Operations Research Analysts	Medical Scientists, Except Epidemiologists
11	11-9033	19-3041	Education Administrators, Postsecondary	Sociologists
12	13-1199	11-9199	Business Operations Specialists, All Other	Managers, All Other
13	13-1023	11-3061	Purchasing Agents, Except Wholesale, Retail, and Farm Products	Purchasing Managers
14	49-9043	17-3026	Maintenance Workers, Machinery	Industrial Engineering Technologists and Technicians
15	19-3099	15-2031	Social Scientists and Related Workers, All Other	Operations Research Analysts
16	21-1012	43-4111	Educational, Guidance, and Career Counselors and Advisors	Interviewers, Except Eligibility and Loan
17	11-9161	55-1019	Emergency Management Directors	Military Officer Special and Tactical Operations Leaders, All Other
18	11-1021	43-1011	General and Operations Managers	First-Line Supervisors of Office and Administrative Support Workers
19	49-9012	17-3023	Control and Valve Installers and Repairers, Except Mechanical Door	Electrical and Electronic Engineering Technologists and Technicians
20	11-9033	21-1099	Education Administrators, Postsecondary	Community and Social Service Specialists, All Other
21	11-9161	13-1199	Emergency Management Directors	Business Operations Specialists, All Other
22	19-1042	11-9121	Medical Scientists, Except Epidemiologists	Natural Sciences Managers
23	51-9151	17-3029	Photographic Process Workers and Processing Machine Operators	Engineering Technologists and Technicians, Except Drafters, All Other
24	47-1011	11-9021	First-Line Supervisors of Construction Trades and Extraction Workers	Construction Managers
25	43-9041	13-1031	Insurance Claims and Policy Processing Clerks	Claims Adjusters, Examiners, and Investigators

Source: Lightcast, Authors' calculations
Note: This Table enumerates the top 25 pairs of SOC6 occupations that exhibit the closest relationship in terms of their skills but differ in the broader SOC2 classification for the year 2014. The low Euclidean distances suggest high similarity, which could indicate overlapping skills or comparable work activities even if they are classified in separate occupational groups.

Note: This table is included as an image. The text discusses its content, but the data to reproduce it in LaTeX is not available.

of distance produces robust and reasonable estimates. These results help address valid concerns about the quality and consistency of the OJV data over time.

One potential concern is that the description of tasks in job postings may become more detailed over time, or that certain skills may be replaced by newer equivalents while the underlying tasks remain unchanged (e.g., replacing TensorFlow with PyTorch). In both scenarios, the computed distances between occupations could artificially increase, making it more difficult to detect convergence, which is not what is currently observed. In addition, this concern is dealt with since we also repeat our exercises and results do not change with the O*NET characteristics that remain unchanged over the sample period. We also check robustness of the results by replicating the analysis using Cosine Similarity instead of Euclidean Distance to assess similarity. The results are identical and omitted for brevity.

Another measure of robustness is to test the effect of using alternative vector sizes

Table 3: Closer Occupations Have Higher Mobility

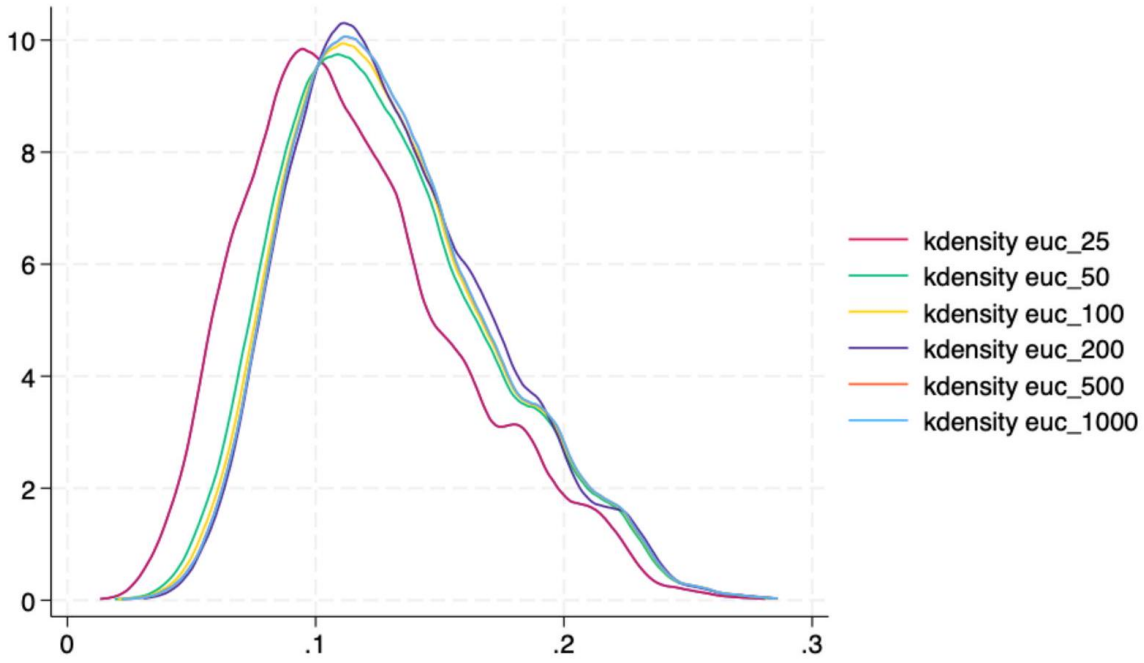
Dependent Variables:	Transition Share				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Distance	−0.0226***	−0.0195***	−0.0216***	−0.0370***	−0.0849***
	(0.0007)	(0.0006)	(0.0007)	(0.0012)	(0.0054)
Same		0.0030***	0.0033***	0.0052***	0.0257***
		(0.0002)	(0.0002)	(0.0004)	(0.0023)
Distance × Same		−0.0082***	−0.0090***	−0.0152***	−0.0437***
		(0.0008)	(0.0009)	(0.0016)	(0.0080)
Fixed-effects					
Occupation 1	Yes	Yes	Yes	Yes	Yes
Occupation 2	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	683102	683102	656910	275292	93.721
R^2	0.32585	0.32922	0.34051	0.37590	0.05630
Within R^2	0.03279	0.03762	0.04120	0.06680	0.03725

Significance Codes: ***: 0.01, **: 0.05, *: 0.1

when computing bilateral distances. As explained in the Methodology section, the skill data provided by Lightcast include approximately 16,000 keywords or phrases. However, since most of these skills are not relevant to most occupations, we represent each occupation at time t as a vector of its most important skills. Specifically, we use the top 200 skills (ranked by importance) for comparison across occupations. To assess the robustness of this choice, Figure 7 presents the kernel density functions for bilateral distances in 2014 using alternative vector sizes (e.g., 25, 50, 100, ..., 1,000 skills). The results demonstrate that the density functions remain nearly indistinguishable when using 50 or more observations from the skills vector. This suggests that the choice of vector size does not significantly impact the computed distances.

To ensure that the results are not driven by outliers, we drop the top 10 or top 25 skills by importance and repeat the analysis. Again, the results remain unchanged and are available upon request. Lastly, one may be concerned with the increasing number of job ads over time driving the results. First, our vectors of skill intensity based on the frequency of appearances are normalized by the total number of skill mentions. As a result, the number of advertisements should not affect the measure of distance, but to ensure the robustness of the results, we also repeat the analysis by

Figure 5: Figure 7 - Kernel density functions of bilateral distances, 2014, various skill-vector sizes



using the following strategy. We first create a subsample of random job ads per year (e.g. 1 million, or 2 million every year for each robustness exercise), we recalculate our distance and repeat the analysis in this and the next section and results remain unchanged.

5.3 What Drives the Convergence?

Having established that occupational distances are declining, we next investigate why this convergence is occurring. We decompose this question into two parts: 1. Is the convergence driven by the content of jobs changing (e.g., technology adoption) or by a change in the mix of jobs in the economy (e.g., growth in "generalist" jobs)? 2. If job content is the driver, which specific skill groups are pulling occupations closer together?

To address the first question, we conduct a Kitagawa–Oaxaca–Blinder decomposition. This method allows us to parse the total change in Euclidean distance into a "Within Effect" (Job Content), which captures the change caused by O*NET skill vectors changing within occupations, and a "Between Effect" (Job Mix), which captures the change caused by employment shares shifting between occupations. In

particular, Let $\bar{D}$ be the aggregate average skill distance in the economy, which is the weighted average of all individual bilateral distances d_{ij} (the Euclidean distance between occupation i and occupation j). The weight $w_{ij} = s_i \cdot s_j$ for each pair is the product of the two occupations share of total employment.

The average distance in the initial period (t) is equal to the weighted mean of each occupation's average distance , letting :

$$\bar{D}_t = \sum_{i,j} w_{ij,t} \cdot d_{ij,t} = \sum_{i} s_{i,t} \cdot D_{i,t}$$

The average distance in the final period ($t+1$) is:

$$\bar{D}_{t+1} = \sum_{i,j} w_{ij,t+1} \cdot d_{ij,t+1}$$

Then the total change we want to explain is $\Delta\bar{D} = \bar{D}_{t+1} - \bar{D}_t$. We decompose this total change as:

$$\Delta\bar{D} = \underbrace{\sum_{i,j} w_{ij,t} \cdot (d_{ij,t+1} - d_{ij,t})}_{\text{Within Effect (Job Content)}} + \underbrace{\sum_{i,j} (w_{ij,t+1} - w_{ij,t}) \cdot d_{ij,t+1}}_{\text{Between Effect (Job Mix)}}$$

The Within Effect $= \sum_{i,j} w_{ij,t} \cdot (d_{ij,t+1} - d_{ij,t})$ measures the change in average distance due to the changing content of jobs. It calculates the change in distance for each occupation pair $(d_{ij,t+1} - d_{ij,t})$ and weights that change by the occupation pair's initial employment share. The Between Effect $= \sum_{i,j} (w_{ij,t+1} - w_{ij,t}) \cdot d_{ij,t+1}$ calculates the change in distance due to employment shares of different occupations changing.

The results are presented in Table 4. The decomposition is striking: the "Job Content" (Within) effect is the overwhelming driver of the change, accounting for 89.5% to 95.6% of the total convergence across all time periods. The "Job Mix" (Between) effect is minimal. This provides strong evidence that the shrinking average skill distance is driven by the changing content of jobs themselves, not by a structural shift in the occupational composition of the economy.

Given that the convergence is driven by what is in a job, we next identify which skill groups are responsible. We conduct a counterfactual analysis that asks, "What would the change in similarity have been if a specific skill group didn't change?". We classify certain skills into four groups—Cognitive (Hi), IT Skills, Routine (Hi), and Work Styles—and present the ratio of the counterfactual change to the actual average change in Table 5.

In this framework, a positive ratio means the skill group contributed to convergence (made jobs more similar), while a negative ratio indicates it was a diverging

Table 4: Decomposition of Total Change in Euc. Distance

Period	Total Change (Euc. Distance)	Job Content	Job Mix
2011–2022	-0.0325	-0.0291 (89.5%)	-0.0034 (10.5%)
2014–2022	-0.0344	-0.0314 (91.3%)	-0.0030 (8.7%)
2011–2019	-0.0207	-0.0191 (92.3%)	-0.0015 (7.7%)
2014–2019	-0.0225	-0.0215 (95.6%)	-0.0010 (4.4%)

Note: Cosine distance results show the exact same pattern, with the 'Within' effect explaining 91-98% of the decline.

Table 5: Ratio of Counterfactual Avg. Change to Avg. Change by Skill Group

Skill Group	2011–2019 (Pre-Pandemic)	2014–2019 (Pre-Pandemic)	2011–2022 (Full Period)	2014–2022 (Main Sample)
Cognitive (Hi)	1.605	0.524	1.606	1.080
IT Skills	0.196	−0.104	0.183	−0.030
Routine (Hi)	−0.344	0.500	−0.434	0.086
Work Styles	0.127	0.220	0.016	0.001

force (made jobs less similar). The results show that the Cognitive (Hi) skill group is the primary driver of convergence. In the 2014-2022 main sample, its ratio is 1.080, and in the full 2011-2022 period, it is 1.606. A ratio greater than 1.0 implies this group was a primary driver that also amplified the effects of other changes. Conversely, IT Skills show a mixed and often negative contribution, such as -0.104 in the 2014-2019 period and -0.030 in the main 2014-2022 sample. This suggests that while IT skills are proliferating, they are not the primary force pulling all occupations together. In summary, the story of occupational convergence since 2011 is one of within-job transformation, overwhelmingly driven by the changing nature and diffusion of high-level cognitive tasks.

6. Implications of Convergence on Wages and Mobility

So far, we have introduced a novel methodology for measuring skill-based distance across occupations (or any other groups of interest). Our findings demonstrate that, over time, occupations are converging as the skill distances between them decrease. In this section, we empirically demonstrate that occupational convergence has resulted in

wage convergence across occupations and higher mobility. Drawing on the Lucas and Prescott island model of occupations, we hypothesize that as islands (occupations) move closer together, the cost of transitioning from one island to another decreases. As a result, we expect to observe two key outcomes: As a pair of occupations becomes more similar (i) wages in these two occupations converge (wage inequality across occupations declines), and (ii) bilateral mobility increases as workers find it easier to transition between occupations.

6.1 Wage convergence

We begin by investigating whether the convergence of occupations over time has led to convergence in wages. First, we plot in Figure 8 the average computed bilateral distances between occupational pairs at the SOC4 level alongside the coefficient of variation of the median wage over time. The results clearly illustrate that since 2014, as occupational distances decrease-indicating greater skill convergence-wages have also converged. The declining coefficient of variation suggests a convergence of wages within SOC6 occupation groups, while the decline in median Euclidean distance trend suggests an increase in occupation homogeneity.[10]. Why would skill convergence lead to lower across occupation wage inequality. Our hypothesis rests on the increasing occupational mobility, since workers now obtain experience in one occupation that is relevant for a broader set of occupations effectively increasing their outside option. The panel B of figure 8 suggests that this may be part of what is the underlying mechanism. In this section, we try to formally test this hypothesis.

To formalize this finding and address concerns about omitted variable bias, we conduct a regression analysis. Specifically, we regress the change in relative wages between occupations i and j on the change in bilateral distances d_{ij} for the years 2014 and 2022. This approach strengthens our test of the hypothesis by moving beyond aggregate trends over time. Instead, it leverages the richness of the bilateral data to determine whether wage convergence is greater for occupations that become more similar in terms of skills.

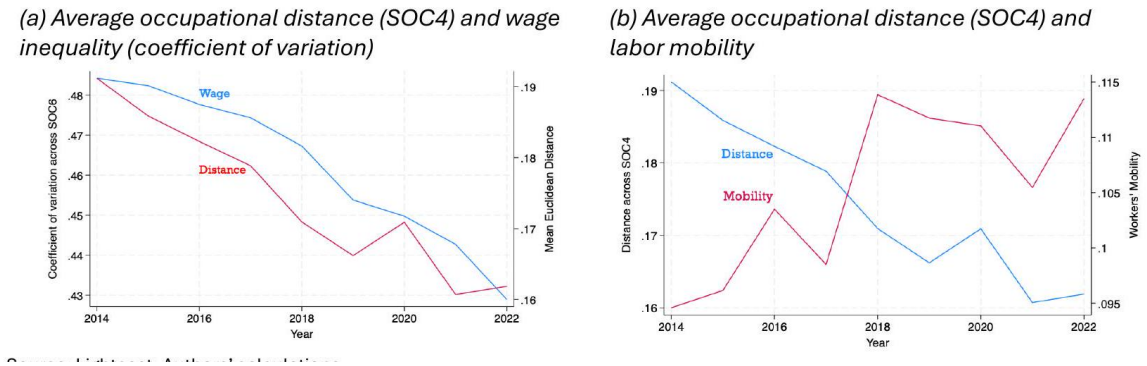
We compute relative wages between occupations i and j as follows:

$$\Delta rw_{ij} = \lceil rw_{ij}^{2022} - w_{ij}^{2014} \rceil = \left| \left(\frac{w_i^{2022} - w_j^{2022}}{w_i^{2022} + w_j^{2022}} \right) - \left(\frac{w_i^{2014} - w_j^{2014}}{w_i^{2014} + w_j^{2014}} \right) \right|$$

The regression results are presented in Table 4. In Column 1, we do not control other determinants other than the change in bilateral distance (Δeuc) between occupations from 2014 to 2022. Here, a higher value of Δ Dist indicates that the two

[10]Notably, these trends observed in recent years between wage inequality and occupational distance is not present when wage variation and distance are calculated by MSAs

Figure 6: Figure 8 - CV and average distance over time



Source: Lightcast, Authors' calculations

occupations have moved further apart. The results reveal a key finding: occupations that have become closer in terms of skill similarity exhibit significantly higher wage convergence.

In Column 2, we extend the analysis by controlling for the initial gap in wages, an indicator for occupation belonging in the same broad group of occupationa and all the interactions. The results remain robust: occupations that converge in skills also experience greater wage convergence. Additionally, in Columns 3 and 4, we extend the analysis by considering the robustness of rerunning the analysis for a different set of first and final year. In Column 3, we take the difference between 2022 and 2011 and in Column 4 between 2011 and 2019.

6.2 Labor mobility

We also investigate the impact of occupational convergence on labor mobility. The regression results show a clear pattern: for a pair of occupations as occupational distance for the period 2014-2022 or 2011-2022 (table 5, Columns 1 and 2) decreases, mobility increases. In particular in table 5 below, the regression results presented provide strong evidence for this pattern. This finding is consistent and statistically significant across all models and all time periods considered, with the coefficient for "Change in Distance" being negative and highly significant ($p<0.05$). This negative relationship implies that a decrease in the "distance" (i.e., an increase in similarity) between two occupations is associated with a greater change in the share of workers transitioning between them. The inclusion of extensive fixed effects, such as Occupation 1, Occupation 2 , and Occupation 1*Occupation 2 pair FE, is crucial as it controls for unobserved factors specific to each occupation pair, strengthening the

Table 6: Table 4: Wage Convergence Implications (2014-2022)

Model:	(1)	(2)	(3)	(4)
Variables				
$\Delta Dist$	0.0656**	0.0527**	0.0339**	0.0364**
	(0.0282)	(0.0207)	(0.0136)	(0.0156)
ln Rel. Wage		−0.1535***	−0.0563***	−0.1471***
		(0.0077)	(0.0041)	(0.0080)
Same		−0.0031**	−0.0020**	−0.0021
		(0.0015)	(0.0010)	(0.0016)
$\Delta Dist\times$ Same		−0.0223	−0.0224**	−0.0278**
		(0.0180)	(0.0110)	(0.0150)
ln Rel. Wage × Same		−0.0325***	−0.0003	−0.0396***
		(0.0072)	(0.0025)	(0.0075)
Fixed-effects				
Occupation 1	Yes	Yes	Yes	Yes
Occupation 2	Yes	Yes	Yes	Yes
Fit statistics				
Observations	93.942	93.942	93.942	93.942
R^2	0.34340	0.51443	0.37530	0.50459
Within R^2	0.00058	0.26091	0.10475	0.23869

Clustered (Occupation 1) standard-errors in parentheses
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

conclusion that the observed effect is due to changes in occupational similarity.

The robustness of this core result is further demonstrated by its persistence in Table 6 included in the Appendix, which includes regressions on YoY mobility rates instead of long difference changes.In particular, in table 6 an increase in the change in distance in a particular year is associated with a decline in the change in mobility bilaterally. In table 6, the change in distance is calculated for each pair relative to the year before and similarly for mobility patterns.

7. Robustness and alternative mechanisms

The analysis of wage trends and occupational similarity is corroborated by a series of robustness checks to ensure the findings are not a product of specific data choices, time

Table 7: Table 5: Long Difference Mobility Implications

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
$\Delta Dist$	−0.0389*	−0.0377**	−0.0488***	−0.0346***	−0.0398**	−0.1627***
	(0.0211)	(0.0165)	(0.0186)	(0.0126)	(0.0193)	(0.0611)
$\Delta Dist \times$ Same			0.0328	0.0602*	0.0074	0.0227
			(0.0576)	(0.0355)	(0.0366)	(0.1540)
Fixed-effects						
Occupation12(2dig)	Yes	Yes	Yes	Yes	Yes	Yes
Occupation1(2dig)	Yes	Yes	Yes	Yes	Yes	Yes
Occupation2(2dig)	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	6.109	6.461	6.109	9.087	6.461	3.183
R^2	0.04555	0.04060	0.04562	0.02998	0.04060	0.08321
Within R^2	0.00051	0.00039	0.00059	0.00064	0.00040	0.00229

Clustered (o12) standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

periods, or spurious correlations. First, we ensure that the primary regression results, which show a relationship between increasing skill similarity across occupations and wage compression are not driven by the data used to construct relative wages.

The main results are based on BLS OEWS data, a standard and reliable source for occupational wage estimates. To confirm these findings, the analysis is replicated using two alternative wage measures: 1. Using the Current Population Survey (CPS) provides an independent measure of wages [11], 2. Posted Wages (at the SOC6 level) This measure is derived directly from online job vacancies and 3. Residualized wages (CPS sample) which involves regressing wages on a set of characteristics (e.g., education, experience, gender) and using the residuals as the wage measure. This method helps to isolate the portion of wages that is not explained by these standard factors, providing a cleaner look at wage changes driven by skill similarity. Across all measures, our results, that are provided in the online Appendix, consistently show that an increase in occupational similarity is associated with a decline in bilateral relative wages.

In addition, the main analysis is conducted over the period 2014-2022. However,

[11]The CPS is a large, nationally representative survey of households, and its wage data, when aggregated to the census-specific occupational classification have low levels of measurement error.

to ensure the results aren't sensitive to this specific time frame, the regressions are also run for alternative periods: 1. 2011-2022 sample period: While the early data introduce some noise, we obtain consistent results indicating the robustness of the core findings. We also replicate the analysis excluding the COVID-119 pandemic years: in particular we use: 2. the 2011-2019 pre-COVID period: and 3. 2014-2019 pre-COVID period. The COVID-19 pandemic significantly disrupted labor markets, potentially affecting wage dynamics. Our results for the period before the pandemic, confirm that the relationship between skill similarity and wage compression existed independent of the unique economic shocks of 2020 and beyond.

A key concern is that the observed relationship is not causal but rather a spurious association driven by other factors. Two main alternative hypotheses are considered and tested using the availability of wage data at the MSA-Occupation level from the BLS OEWS statistics. Hypothesis 1: It's plausible that a tight labor market (low unemployment, many job openings) forces firms to compete more aggressively for talent. This competition could lead them to post job ads that are more similar in their skill requirements, resulting in wage compression as firms are forced to pay similar wages for similar skills. To test this, the analysis is conducted separately for states with low and high labor market tightness. If the results hold in both, it suggests that tightness is not the primary driver of the observed relationship. Results in Table 7, Columns 2 and 3 confirm that tightness is not the main driver. Hypothesis 2: Increases in the state minimum wage could force employers to automate certain tasks or restructure roles, leading to a shift in skill requirements for low-skilled jobs. This could increase the similarity of job postings for this group and, in turn, affect wage compression. To rule this out, the regressions are run for states that experienced no changes in their minimum wage. Results in Table 7, Columns 1 confirm that changes in state mininum wage do not drive the results.

8. Conclusion

In this study, we introduced a methodology for measuring skill-based distances between occupations and demonstrated its wide-ranging applications. By leveraging detailed skill data from job advertisements, 1. we created an algorithm to standardise a vocabulary of raw keywords from parsed text to a standardized vocabulary of occupational characteristics and then 2.we computed bilateral distances between 1,326 SOC6 occupations and observed a clear trend of convergence over time for the period 2014-2022, meaning that occupations have become more similar in their skill requirements.

We empirically showed that this occupational convergence has significant implications for labor market outcomes. First, wage convergence is correlated occupational

Table 8: Table 7: Alternative Explanations: Effects of Minimum Wage and Labor Market Tightness

Model:	**(No Min. Wage)**	**(Low Tightness)**	**(High Tightness)**
Variables			
$\Delta Dist$	0.0471^{***}	0.0530^{***}	0.0526^{***}
	(00084)	(00082)	(00083)
Fixed-effects			
MSA:occ1	Yes	Yes	Yes
MSA:occ2	Yes	Yes	Yes
Fit statistics			
Observations	1,824,712	1,961,853	2,068,963
R^2	0.46916	0.46348	0.46865
Within R^2	7.96×10^{-5}	9.92×10^{-5}	9.74×10^{-5}

convergence. Initial gaps in wages narrow as skill distances decrease, and this relationship holds even after controlling for initial wage disparities, self-distances, and other factors. Notably, occupations classified as Low Non-Routine Cognitive (Low NR-C) experienced the largest reduction in wage inequality, reversing earlier trends of rising inequality in these roles.

Second, we demonstrated that occupational convergence leads to higher labor mobility. As skill-based distances between occupations decrease, workers find it easier to transition between roles, increasing overall mobility within the labor market. Regression analysis confirmed this relationship: the reduction in bilateral distances is significantly associated with rising transitional shares between occupations, even after accounting for fixed effects and employment-based weights.

Beyond these core findings, our methodology enabled us to construct occupational networks using skill-based distances. By employing tools such as minimum spanning trees and centrality measures, we visualized the interdependencies among occupations and identified key nodes, such as Business and Management Specialists, which occupy central roles within the occupational landscape. This network-based approach provided further validation of our results and highlighted the practical utility of our distance measures for understanding workforce dynamics.

In summary, the convergence of occupations over time has led to wage convergence and increased labor mobility, signaling a more integrated and dynamic labor market. These findings provide valuable insights into the changing structure of occupations and underscore the importance of skill-based analysis for informing policy decisions related to workforce development, reskilling initiatives, and wage inequality.

Bibliography

1. Acemoglu, Daron. "Directed technical change." The review of economic studies 69.4 (2002): 781-809.

2. Acemoglu, Daron, et al. "Artificial intelligence and jobs: Evidence from online vacancies." Journal of Labor Economics 40.S1 (2022): S293-S340.

3. Aeppli, Clem, and Nathan Wilmers. "Rapid wage growth at the bottom has offset rising US inequality." Proceedings of the National Academy of Sciences 119.42 (2022): e2204305119.

4. Autor, David H., Lawrence F. Katz, and Alan B. Krueger. "Computing inequality: have computers changed the labor market?." The Quarterly journal of economics 113.4 (1998): 1169-1213.

5. Autor, David H., Frank Levy, and Richard J. Murnane. "The skill content of recent technological change: An empirical exploration." The Quarterly journal of economics 118.4 (2003): 1279-1333.

6. Autor, David H. "Why are there still so many jobs? The history and future of workplace automation." Journal of economic perspectives 29.3 (2015): 3-30.

7. Autor, David H., and David Dorn. "The growth of low-skill service jobs and the polarization of the US labor market." American economic review 103.5 (2013): 1553-1597.

8. Song, Jae, et al. "Firming up inequality." The Quarterly journal of economics 134.1 (2019): 1-50.

9. Cammeraat, Emile, and Mariagrazia Squicciarini. "Burning Glass Technologies' data use in policy-relevant analysis: An occupation-level assessment." (2021).

10. Card, David, and John E. DiNardo. "Skill-biased technological change and rising wage inequality: Some problems and puzzles." Journal of labor economics 20.4 (2002): 733-783.

11. Carrillo-Tudela, Carlos, and Ludo Visschers. "Unemployment and endogenous reallocation over the business cycle." Econometrica 91.3 (2023): 1119-1153..

12. Chodorow-Reich, Gabriel, and Johannes Wieland. "Secular labor reallocation and business cycles." Journal of Political Economy 128.6 (2020): 2245-2287.

13. Deming, David, and Lisa B. Kahn. "Skill requirements across firms and labor markets: Evidence from job postings for professionals." Journal of Labor Economics 36.S1 (2018): S337-S369.

14. Kambourov, Gueorgui, and lourii Manovskii. "Occupational mobility and wage inequality." The Review of Economic Studies 76.2 (2009): 731-759.

15. Katz, Lawrence F., and Kevin M. Murphy. "Changes in relative wages, 1963-1987: supply and demand factors." The quarterly journal of economics 107.1 (1992): 35-78.

16. Kuhn, Peter, and Mikal Skuterud. "Internet job search and unemployment durations." American Economic Review 94.1 (2004): 218-232.

17. Lilien, David M. "Sectoral shifts and cyclical unemployment." Journal of political economy 90.4 (1982): 777-793.

18. Lucas Jr, Robert, and Edward C. Prescott. "Equilibrium search and unemployment." Journal of Economic Theory 7.2 (1974): 188-209.

19. Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. "Upskilling: Do employers demand greater skill when workers are plentiful?." Review of Economics and Statistics 102.4 (2020): 793-805.

20. Moscarini, Giuseppe, and Kaj Thomsson. "Occupational and Job Mobility in the US." The Scandinavian Journal of Economics 109.4 (2007): 807-836.

21. Western, Bruce, and Jake Rosenfeld. "Unions, norms, and the rise in US wage inequality." American Sociological Review 76.4 (2011): 513-537.

Tables and Figures

Appendix A: Artificial Intelligence and Data Analytics

ARTIFICIAL INTELLIGENCE (AI) SKILLS	DATA ANALYTICS (DA) SKILLS
Apache Ant	Apache Hadoop
Apache Spark	Apache Hive
Artificial Intelligence	Apache Kafka
Automated Testing	Bayesian Inference
Automation	Bayesian Methods
Automation Consulting	Bayesian Modeling
Automation Systems	Bayesian Networks
Automation Techniques	Big Data
Automation Tools	Big Data Analytics
AWS Elastic MapReduce (EMR)	Data Analysis
BigQuery	Data Analytics
Boosting (Machine Learning)	Data Conversion
Caffe Deep Learning Framework	Data Engineering
Cluster Analysis	Data Management
Clustering	Data Modeling
Clustering Algorithms	Data Science
Computer Vision	Data Visualization
Convolutional Neural Network (CNN)	Extraction Transformation and Loading (ETL)
Decision Trees	Microsoft Power BI
Deep Learning	Microsoft SQL
Kubernetes	MongoDB
Machine Learning	MySQL
Machine Vision	NoSQL
MapReduce	Oracle PL/SQL
ML/AI	PostgreSQL
Natural Language Processing	Predictive Analytics
Natural Language Toolkit (NLTK)	Predictive Models
Neural Networks	Prepare Spreadsheets
Splunk	Spreadsheets
Supervised Learning (Machine Learning)	SQL
Support Vector Machines (SVM)	SQLite
TensorFlow	SQL Plus
Torch (Machine Learning)	SQL Server

Figure 7: BGT to O*NET examples

BGT Skill	O*NET Element	Description
Microsoft Excel is a job skill that involves being proficient with the Excel spreadsheet software, used for creating, editing, and managing data through formulas and functions, generating reports, and performing statistical analysis.	Processing Information	Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.
Computer literacy refers to the ability to use and understand various types of computer software, hardware and applications effectively and efficiently.	Working with Computers	Using computers and computer systems (including hardware and software) to program, write software, set up functions, enter data, or process information.
SQL (Structured Query Language) is a programming language used to manage and manipulate relational databases to store, track, and retrieve specific data efficiently.	Processing Information	Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.
Leadership is the ability to inspire, motivate, and guide individuals or teams to achieve goals, demonstrating qualities like decision-making,...	Developing and Building Teams	Encouraging and building mutual trust, respect, and cooperation among team members.
Quality Assurance and Control refers to the systematic process of ensuring that a product or service meets specified standards of quality and reliability	Quality Control Analysis	Conducting tests and inspections of products, services, or processes to evaluate quality or performance.