The Convergence of Occupations: Evidence from Online Job Posts

Alexis Antoniades
Georgetown University Qatar

Manolis Chatzikonstantinou
Georgetown University Qatar

Mohammad Shahmeer Ahmad
Qatar Computing Research Institute

CRETE 2025

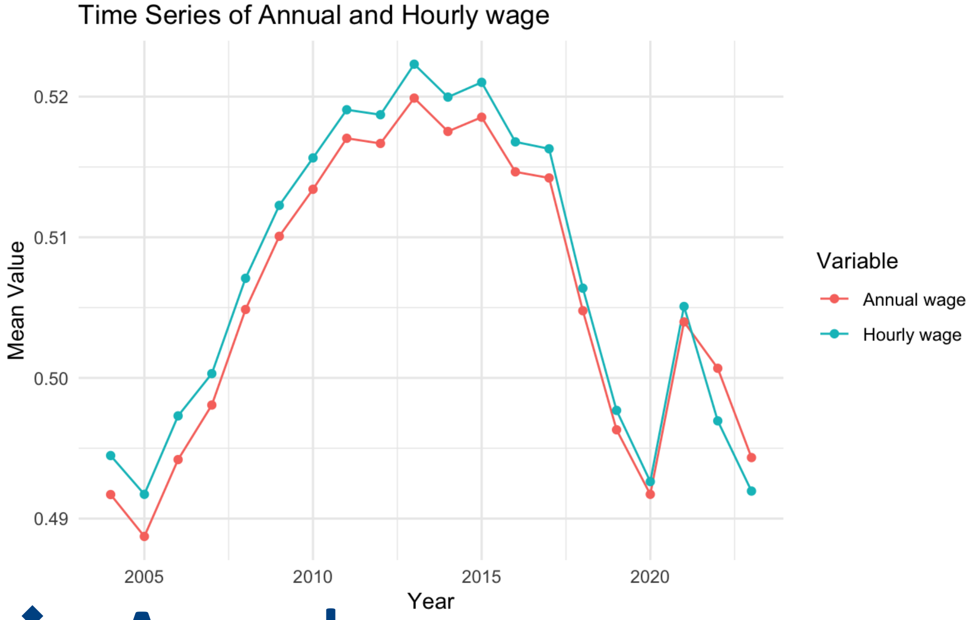
July 07, 2025

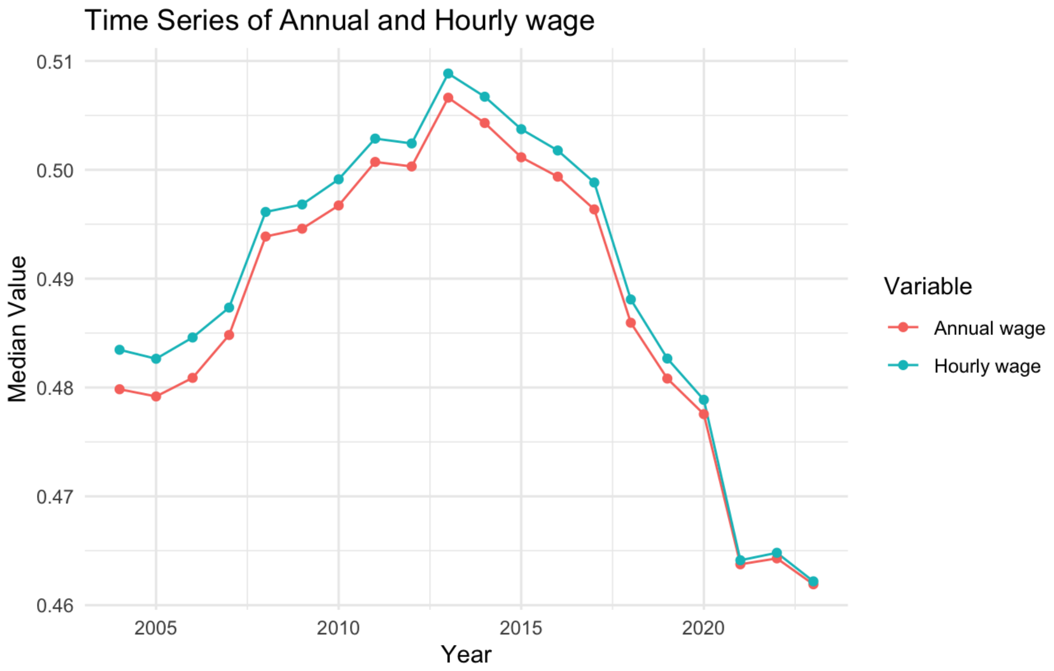
Background

- Large literature on changes in **inequality** and **wage structure**
  - Education & experience (Katz & Murphy 1992)
  - Institutions (DiNardo, Fortin, and Lemieux 1996)
  - Globalization/deindustrialization (Murphy & Welch 1991)
- **Tasks and occupations matter** in explaining inequality (Acemoglu, & Autor, 2011, Fortin and Lemieux 2016).
  - Automation (Cortes et al 2017); Mobility and tightness (Kambourov and Manovski 2009)
  - *Wage differences across occupations explained by differences in the relative price of tasks/skills (tasks within each occupation are time-invariant)* *(Lochkner et al 2025)*
  - But occupation skill/task composition changes;
  - Do changes in task composition help explain dynamics of wage inequality?

"The Unexpected Wage Compression"

❖ Relative Wages across occupations (ADH occupation classification)





❖ A puzzle:

- ❑ For the past decade, mean wages across occupations have been converging,
- ❑ while skill change has accelerated- with software replacing traditional skills

Data Sources:

- ❑ Wages from OEWS (and CPS)

“The Unexpected Wage Compression”

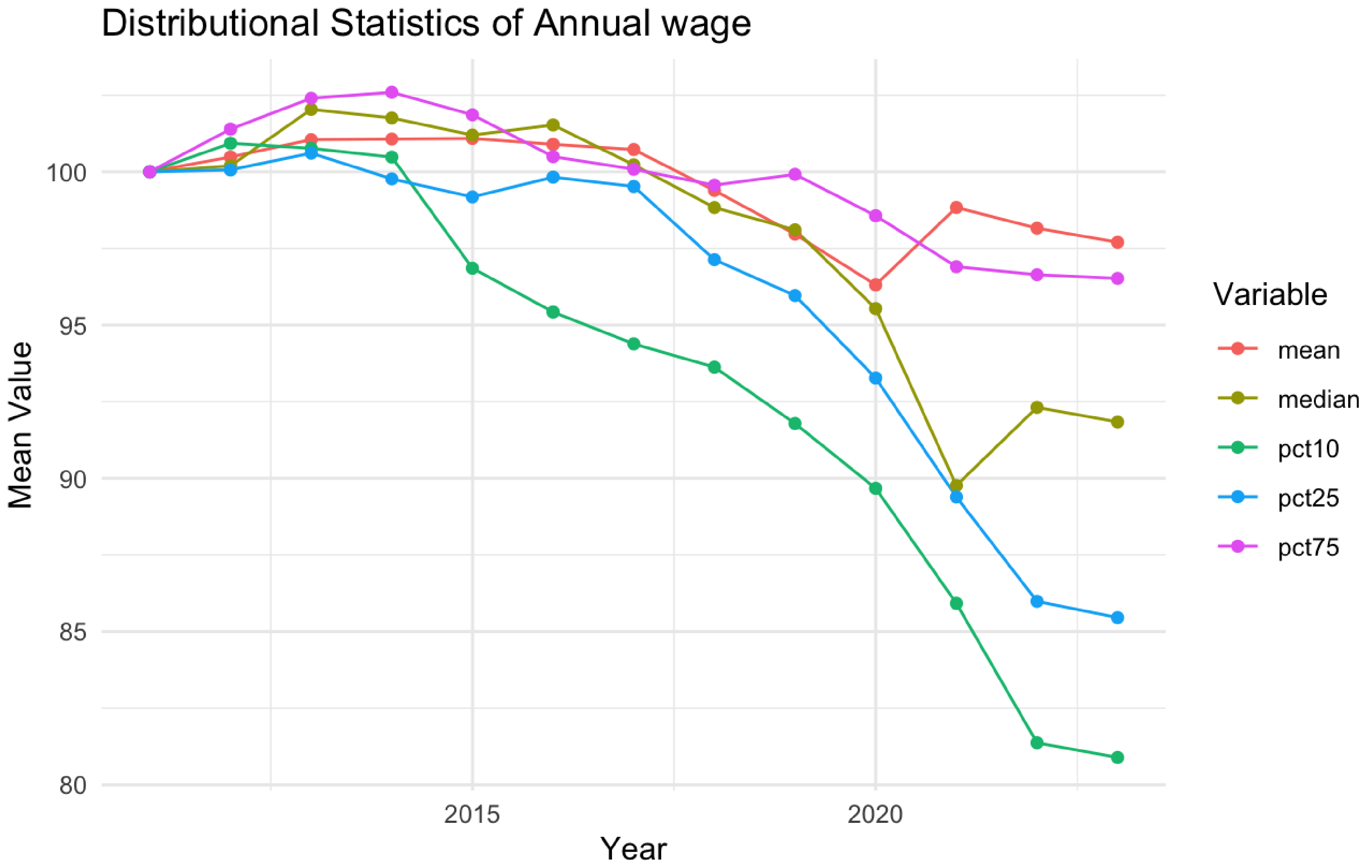


Fig 1 – Coef of variation of different stats (OCC)
Weighted by employment

“The Unexpected Mobility Increase”(?)

Occupational Mobility=Occupational Switch Rate from CPS

❖ A puzzle:

- ❑ For the past decade, occupational mobility has been rising,
- ❑ while skill change has accelerated- with software replacing traditional skills

Data Sources:

- ❑ Longitudinal mobility from CPS

Summary of paper

- **A puzzle:**
  - For the past decade wage inequality has been declining, while mobility has been rising, despite rapid technical change

“The Unexpected Wage Compression”(Aeppli and Wilmers 22):

- Increasing labor market tightness(Webber (22), and Bassier et al. (22), Autor et al(24))
- State Variation in minimum wages (Cengiz et al.,19).
- This paper – mobility and changing skill requirements (Cortes and Gallipoli(18))

- **Technological progress has not only transformed occupations, but it has also made them more similar (brought them *closer*)**
- **As occupations become more similar, wages converge and mobility rises**
- ***Caveat:*** *What determines* **similarity** *? How can it be measured?*

Skills Data

US Data

- Online job vacancy data provided by *Lightcast* (formerly known as BGT)
- Collected daily from 51,000 sources (job boards, company websites, etc).
- Includes info on skills, occupation, location, company, education requirements, experience,…

Skills/Keywords to ONET using SBERT

- Sentence-BERT (SBERT) is a powerful deep learning model.
- It generates high-quality numerical representations (embeddings) for short texts.
- These embeddings capture the *meaning* of the text, not just keywords.

Semantic Similarity: These embeddings capture the *meaning* of the text, not just keywords.

- SBERT allows us to measure how semantically similar a skill description is to all ONET descriptions.

  Even if the wording is different, similar meanings will result in close embeddings.

Semantic Skill – O*NET

1.1. 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.[1] It is a comprehensive and regularly updated database that provides detailed information on occupational characteristics and worker requirements across occupations, as listed in Table 1 below. 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 the 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. We downloaded the most recent iteration of the O*NET database (May 2025).

Table 1: O*NET Data Fields Utilized for Skill Matching

Field Name	Example	Description
Skills		Developed capacities that facilitate learning or the more rapid acquisition of knowledge.
Tasks		Specific actions or functions performed in an occupation.
Knowledge		Organized sets of principles and facts applying to a wide range of problems.
Detailed Work Activities		Specific actions performed across different types of jobs.

Keywords to O*NET

SKILL(Lightcast)	Skill definition
Accounting	Accounting is a finance skill that involves managing, analyzing, and interpreting financial data of an individual or business.
SQL	SQL (Structured Query Language) is a programming language used to manage and manipulate relational databases.
Presentation Skills	Presentation skills refer to the ability to communicate information and ideas effectively to an audience, often using tools like slideshows ...
CPR	Cardiopulmonary Resuscitation (CPR) is a life-saving emergency procedure for maintaining the flow of oxygen to the brain during a cardiac arrest .
Data Entry	Data entry involves inputting, updating and maintaining information in computer databases or systems, often utilizing software tools.

Description	Element Name (ONET)
Knowledge of economic and accounting principles and practices, the financial markets, banking, and the analysis and reporting of financial data.	Economics and Accounting
Compiling, coding, categorizing, calculating, tabulating, auditing, or verifying information or data.	Processing Information
Knowledge of media production, communication, and dissemination techniques and methods. This includes alternative ways to inform and entertain via written, oral, and visual media.	Communications and Media
Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.	Assisting and Caring for Others
Using computers and computer systems (including hardware and software) to program, write software, set up functions, enter data, or process information.	Working with Computers

Methodology

Using hundreds of millions of online job vacancies we represent occupations as vectors of skills, with corresponding weights

Step 1: Create corpus C_{gt}: e.g. *Purchasing Managers (SOC 11-3061)* in 2014

Step 2: Compute document freq. *df* share of each skill in the corpus.

Step 3: Express occupation as vector of skills C_{gt} ranked from most important to least

C_{gt} ={Communication skills - 0.23; Excel – 0.12; Python - 0.03; …}

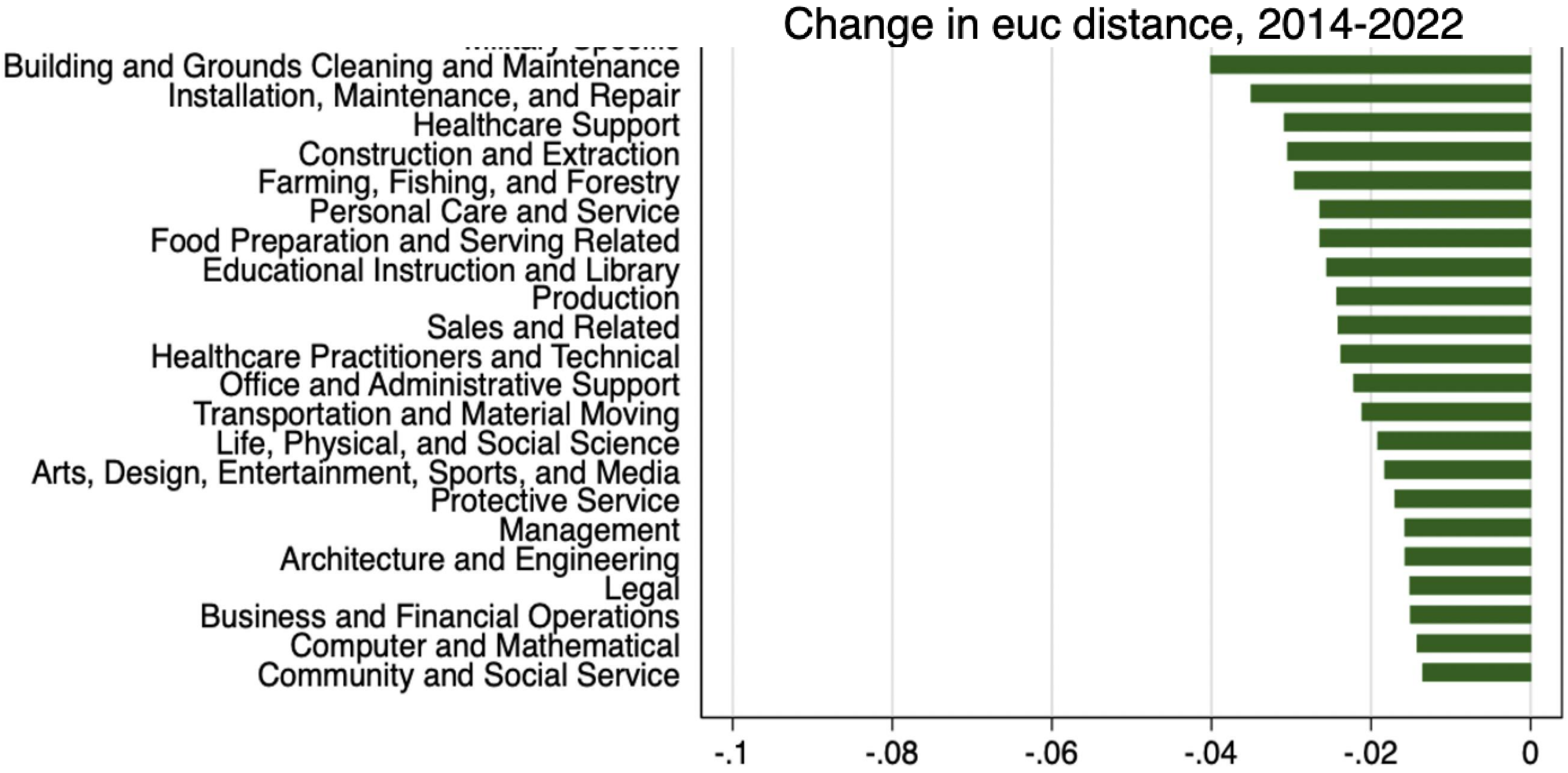
Similar vectors also based on 125 ONET groups

Then use Euclidean distance to measure similarity between two such vectors representing two different corpuses (occupations)

$$d_{ab} = \sqrt{\sum_{i}^{S} (b_i - a_i)^2}$$

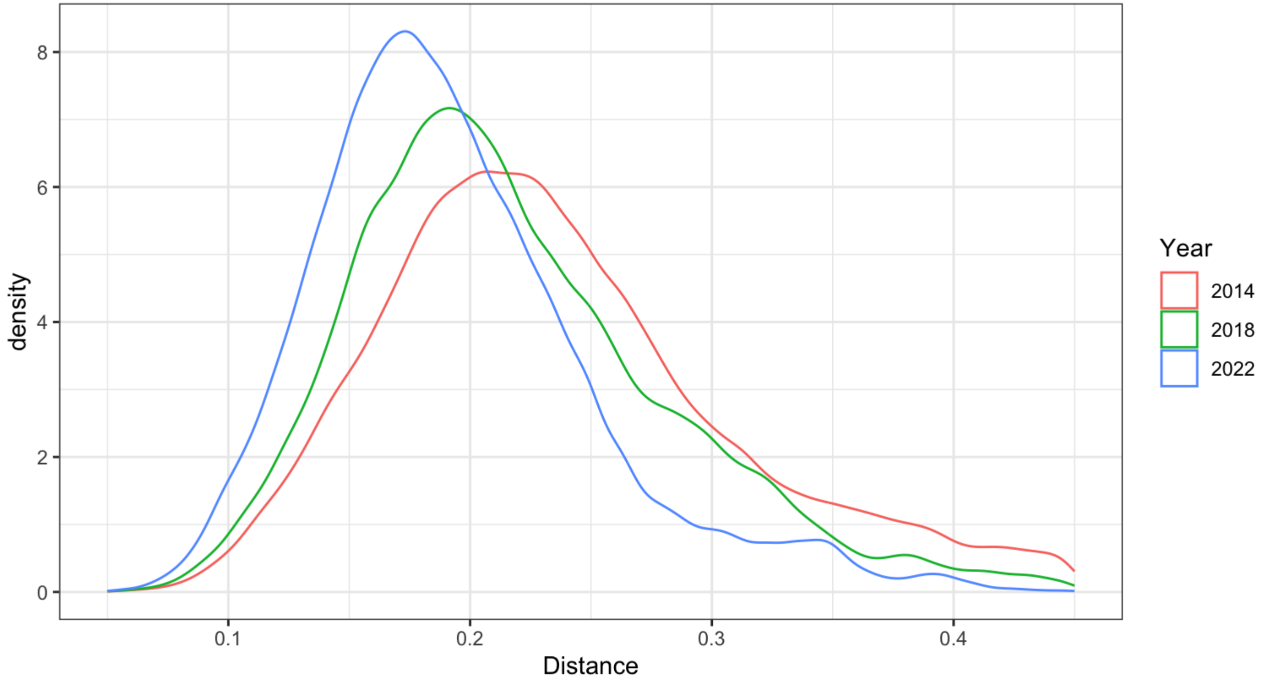
Results tend to hold using other measures of distance (Cosine,…

1.3 Estimation

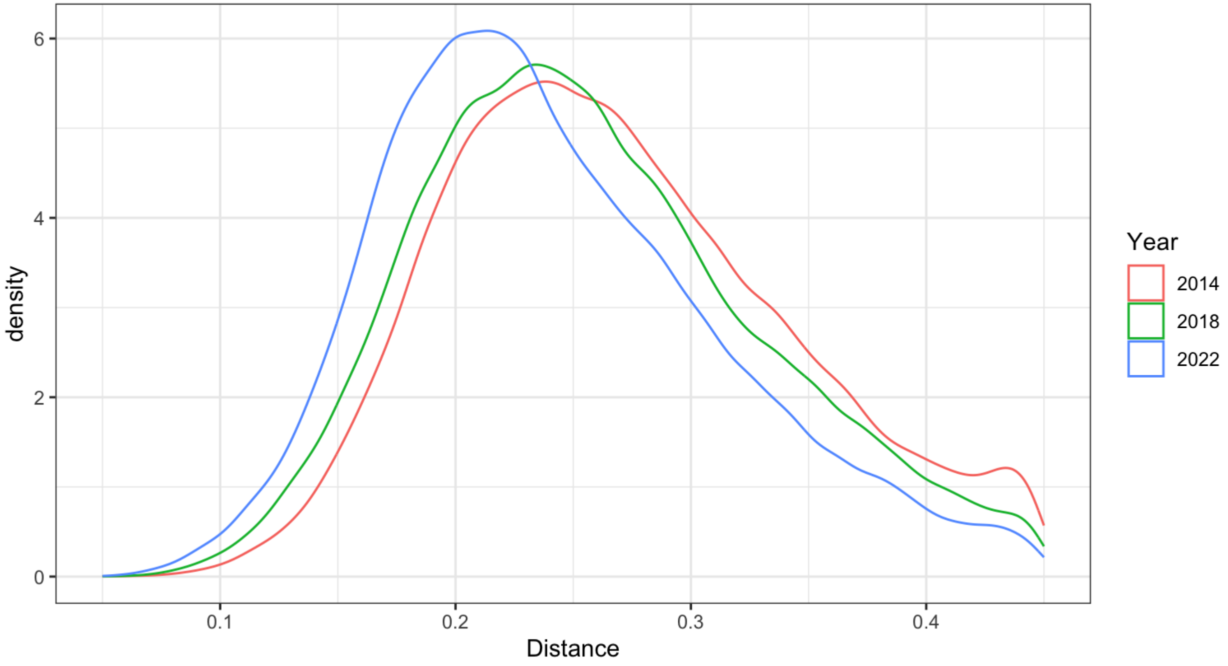


Distribution of bilateral distances for OCC pairs over time-BGT Skills/ ONET

Top 250 BGT Skills



ONET clusters



Note: A group does not have to be an occupation: We can apply our measure of distance to other sub-groups (corpuses) such as occupation-cities groups or firms. Do not find similar convergence patterns across cities.

Distance based on ONET clusters

Figure 2 – Average and median Euclidean distance across all occupations by year

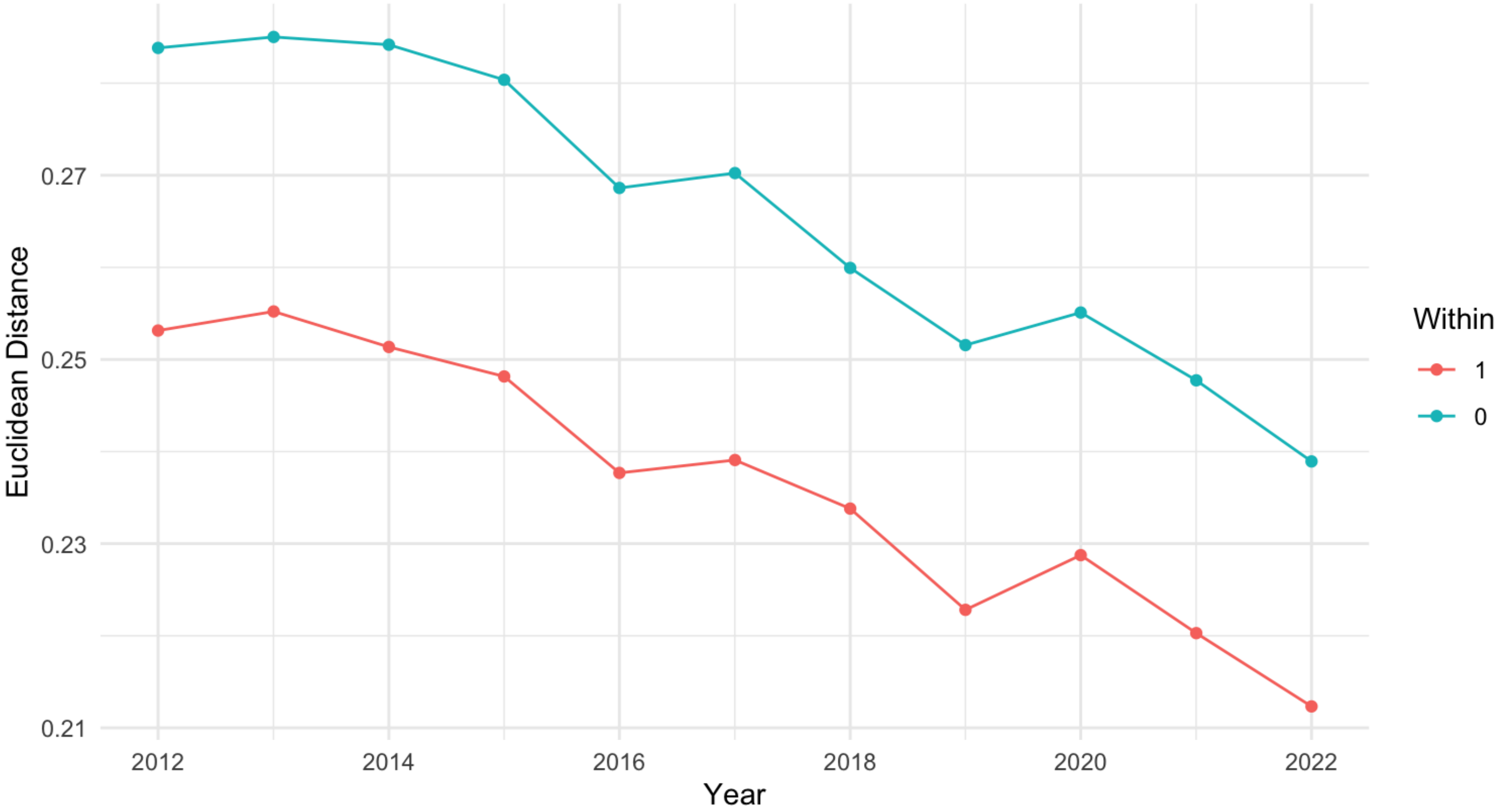
1.4 Robustness/Validity

- **Robustness**
  1. Same dynamics for other distance measures
  2. Not driven by increase in size of sample of ads (choose random samples of jobs every year, e.g. 1m each year.
- **Is our measured distance a good measure of …distance (i.e similarity)?**
  1. Distance within vs across occupation groups
  2. Occupations with the highest similarity
  3. Mobility
     1. Measured using LinkedIn employee histories
     2. Measured from CPS-ASEC sample (vom Lehn et al 22)- retrospective measure

Validity

1. Within group occupations are closer than across group

Figure 3 – Median Euclidean distance across and within SOC2 groups



2. Closest across-group occupations have similar names

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

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	683,102	683,102	656,910	275,292	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

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

Occupations with largest decline in distance

occ1990dd_1	Occupation 1 Name	occ19990dd_2	Occupation 2 Name	year	Δeucdis
577	Electric power installers and repairers	853	Excavating and loading machine operators	2022	-0.348
228	Broadcast equipment operators	617	Other mining occupations	2022	-0.336
669	Shoemakers, other prec. apparel and fabric workers	853	Excavating and loading machine operators	2022	-0.319
228	Broadcast equipment operators	467	Motion picture projectionists	2022	-0.31
467	Motion picture projectionists	617	Other mining occupations	2022	-0.309
228	228 Broadcast equipment operators	853	Excavating and loading machine operators	2022	-0.306
577	Electric power installers and repairers	669	Shoemakers, other prec. apparel and fabric workers	2022	-0.305
518	Industrial machinery repairers	853	Excavating and loading machine operators	2022	-0.303
577	Electric power installers and repairers	637	Machinists	2022	-0.298
256	Advertising and related sales jobs	583	Paperhangers	2022	-0.298

2.1 Convergence Implications: Wages

Hypotheses 1:

As occupations become more similar, do wages converge?

Approach: Combine skilled-based measures of distance with wages at OCC level

- ***Δdist***$_{ij}$: Change in Euclidean distance between 2014 and 2022 for occupations *i* and *j.*
- ***Δrel.wage***$_{ij}$: Change in relative wage between 2014 and 2022 for occupations *i* and *j*.

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

- ***Same***: 1 if *i* & *j* belong to same SOC2 group, 0 otherwise

2.1 Using Online Job Posts(2012-2022)

Model:	(1)	(2)	(3)	(4)
Variables				
$\Delta Dist$	0.0656**	0.0527**	0.0339**	0.0364**
	(0.0282)	(0.0207)	(0.0136)	(0.0156)
In. 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)
$\Delta Dist$ × Same		-0.0223	-0.0224**	-0.0278*
		(0.0180)	(0.0110)	(0.0150)
In. 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*

2 Robustness

❖ Wage measurement:

1. Wages are measured using BLS OEWS data
2. We confirm our results using
    1. CPS measurements (at occsoc level)
    2. Posted Wages (at soc6 level)
    3. Residualized wages (CPS sample)

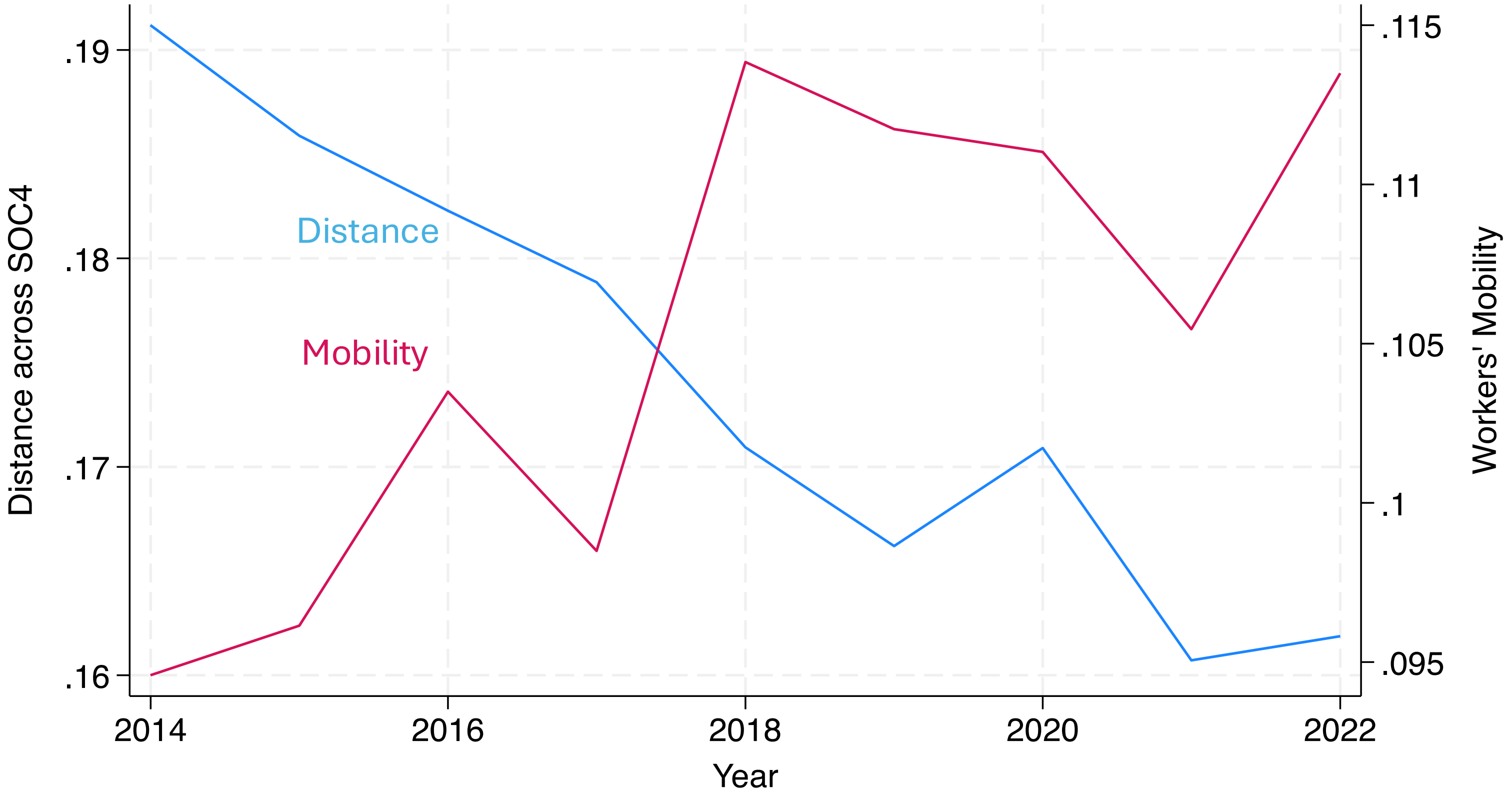
❖ Time periods:

1. Regressions presented for 2014-2022 (2010 OJV data , 2011-2013 online job ads are less reliable)
2. We confirm our results using
    1. 2011-2022 sample period
    2. 2011-2019 pre-covid period
    3. 2014-2019 pre-covid period

2.2 Implications: Mobility

Hypotheses 2:

As occupations become more similar, does mobility rise?



2.2 Implications: Long Difference Mobility

Hypotheses 2:

As occupations become more similar, does mobility rise?

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*

2.2 Implications: YoY Mobility

جامعة جورجتاون قطر
GEORGETOWN UNIVERSITY QATAR

Hypotheses 2:

As occupations become more similar, does mobility rise?

	Dependent Variable: Change in Transition shares between Occ I and Occ j		
	(1)	(2)	(3)
Change in Distance	-0.1504***	-0.1213**	-0.1172**
	(0.0575)	(0.0579)	(0.0556)
Same		-0.0127	-0.0111
		(0.0137)	(0.0132)
Change in Distance* Same		-0.0464	-0.0430
		(0.0338)	(0.0330)
Occupation 1*Year FE	Yes	Yes	Yes
Occupation 2*Year FE	Yes	Yes	Yes
Occupation 1*Occupation 2 FE	Yes	Yes	Yes
Weighted by log employment	No	No	Yes
Observations	24,159	24,159	24,159
R2	0.78044	0.78052	0.77131

Source CPS. Robust standard errors clustered by occupation 1 in parenthesis, *** p<0.01, ** p<0.05, * p<0.1

2.3 Alternative Explanations

Spurious association of increasing similarity and wage compression.

Hypotheses 1:

Labor Market Tightness -> competition between firms -> similar posted ads

Hypotheses 2:

State minimum wage changes-> automation-> similar posted ads for the low skilled workforce

Using MSA-Occupation variation:

1. Results hold in states with no change in min. wage
2. Results hold in states with low and high labor market tightness

2.3 Alternative Explanations

Model:	(No Min. Wage)	(Low Tightness)	(High Tightness)
Variables			
$\Delta Dist$	0.0471***	0.0530***	0.0526***
	(0.0084)	(0.0082)	(0.0083)
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}

2.3 Alternative Explanations

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Constant	-0.0271***	-0.0286***				
	(6.29×10^{-5})	(0.0001)				
$\Delta Dist$	0.0075***	-0.0010	0.0540***	0.0530***	0.0528***	0.0616***
	(0.0010)	(0.0021)	(0.0072)	(0.0071)	(0.0062)	(0.0067)
Tigh		0.0004***				
		(3.41×10^{-5})				
$\Delta Dist \times$ Tigh		0.0024***				
		(0.0005)				
Same						-0.0040***
						(0.0006)
$\Delta Dist \times$ Same						-0.0310***
						(0.0060)
Fixed-effects						
MSA			Yes	Yes		
occ1			Yes	Yes		
occ2			Yes	Yes		
MSA:occ1				Yes	Yes	Yes
MSA:occ2				Yes	Yes	Yes
Fit statistics						
Observations	4,030,816	4,030,816	4,030,816	4,030,816	4,030,816	4,030,816
R^2	1.48×10^{-5}	5.9×10^{-5}	0.07839	0.46175	0.46616	0.46624
Within R^2			5.97×10^{-5}	9.83×10^{-5}	9.83×10^{-5}	0.00023

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

Conclusion

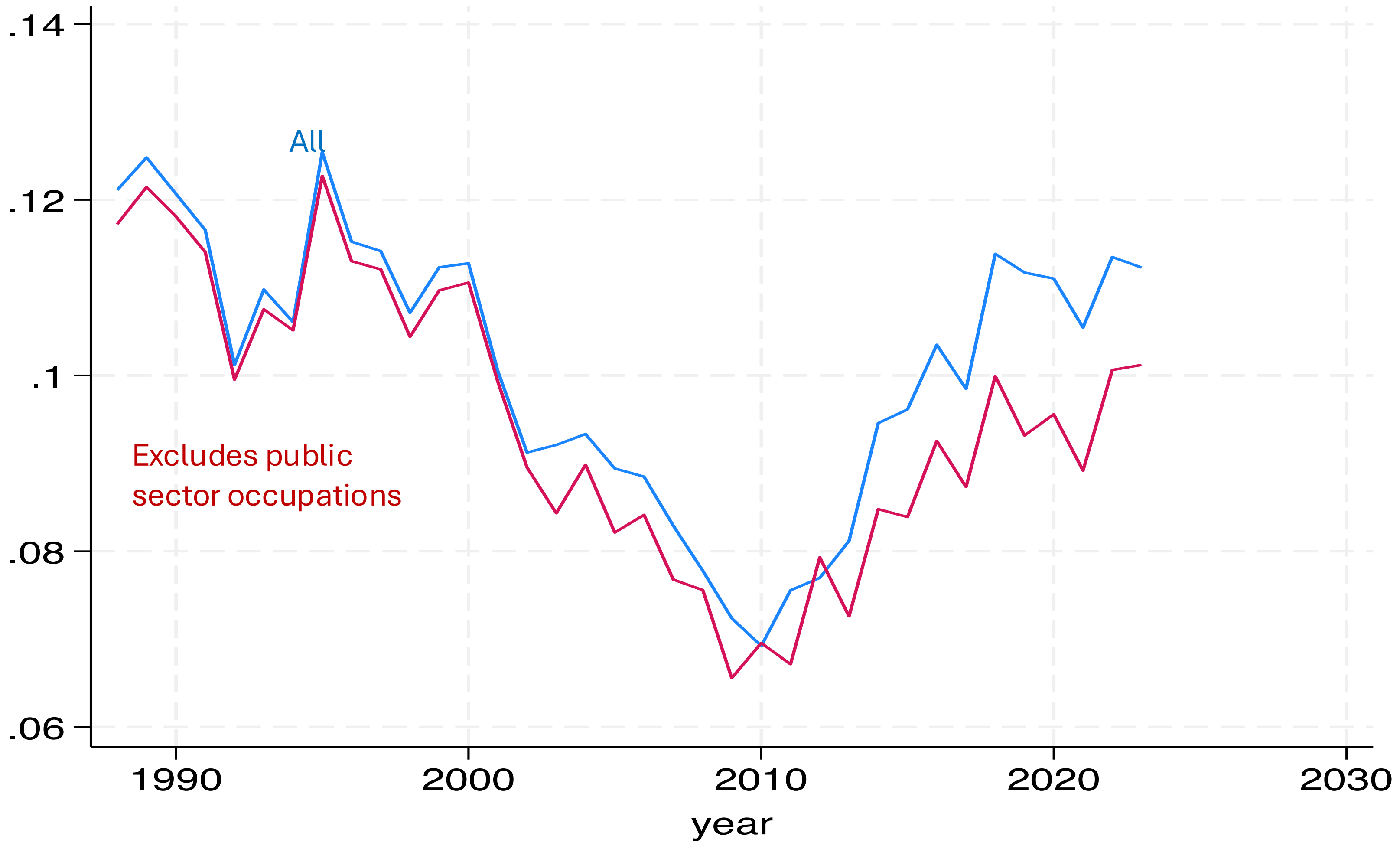
- Changes in task composition can help explain recent dynamics of wage inequality:
  - Convergence in skill requirements-> Increase in mobility and reduction in cross-occupation wage inequality
- Next Steps: We have built a Burdett Mortensen type model with two occupations
  - Similarity/Distance determines the arrival rate of offers from “other” occupation
  - As Similarity/arrival rate increases: low wage earners in each occupation move to higher paying jobs: Mobility increases and relative wages(below the mean) converge
  - In the process of testing model implications: wage increases for movers and patterns of mobility as distance changes

The Convergence of Occupations: Evidence from Online Job Posts

Appendix

Excluding Public sector

❖ Mobility across occupations



1.2 Methodology

Comments:

1. **Similarity:** The distance between two occupations at time t
2. **Convergence**: The change in similarity between two occupation across two points in time t and $t+T$
3. **Self-distance/transformation**: The distance between an occupation at time t and itself at $t+T$
4. **A group does not have to be an occupation**: We can apply our measure of distance to other sub-groups (corpuses) such as occupation-cities groups or firms. Do not find similar convergence patterns across cities.

1.1 Data

GEORGETOWN UNIVERSITY QATAR

Keyword search: “Plumbing”

Skill	2014	2019	Growth
Plumbing Design	605	1,971	2.3
Mechanical, Electrical, and Plumbing (MEP) Design	1,525	4,357	1.9
Commercial Plumbing	2,382	6,691	1.8
Plumbing Installation	418	1,134	1.7
Plumbing Systems	11,787	30,005	1.6
Plumbing Tools	429	1,082	1.5
Plumbing Fixture Installation	225	536	1.4
Plumbing Maintenance	4,544	10,663	1.4
Plumbing	199,046	445,228	1.2
Plumbing Repairs	11,132	24,334	1.2
Copper Plumbing Installation	1	2	1.0
Industrial Plumbing	377	660	0.8
Residential Plumbing	1,848	3,201	0.7

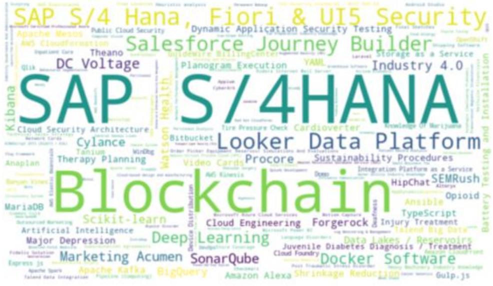
Occupations				Online Job Vacancies (in thousands, 000s)		Occupational Characteristics				
							AI Intensity		Data Analytics	
				2014	2022	High Nonroutine	2014	2022	2014	2022
SOC 2	SOC2 Name	SOC4	SOC6	Q1-Q4	Q1	Cognitive	Q1-Q4	Q1	Q1-Q4	Q1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	23 Total occupation 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%

1.1 Data

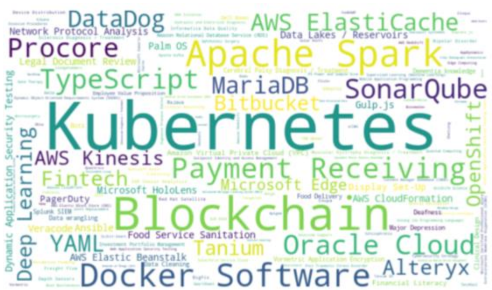
Sample: OJVs for selected occupations (SOC3), 2014-2019

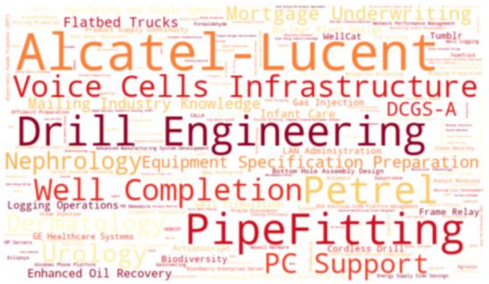
(a) Common (b) New (c) Disappearing

Business Operations Specialists (SOC2 131)



Engineers (SOC2 172)



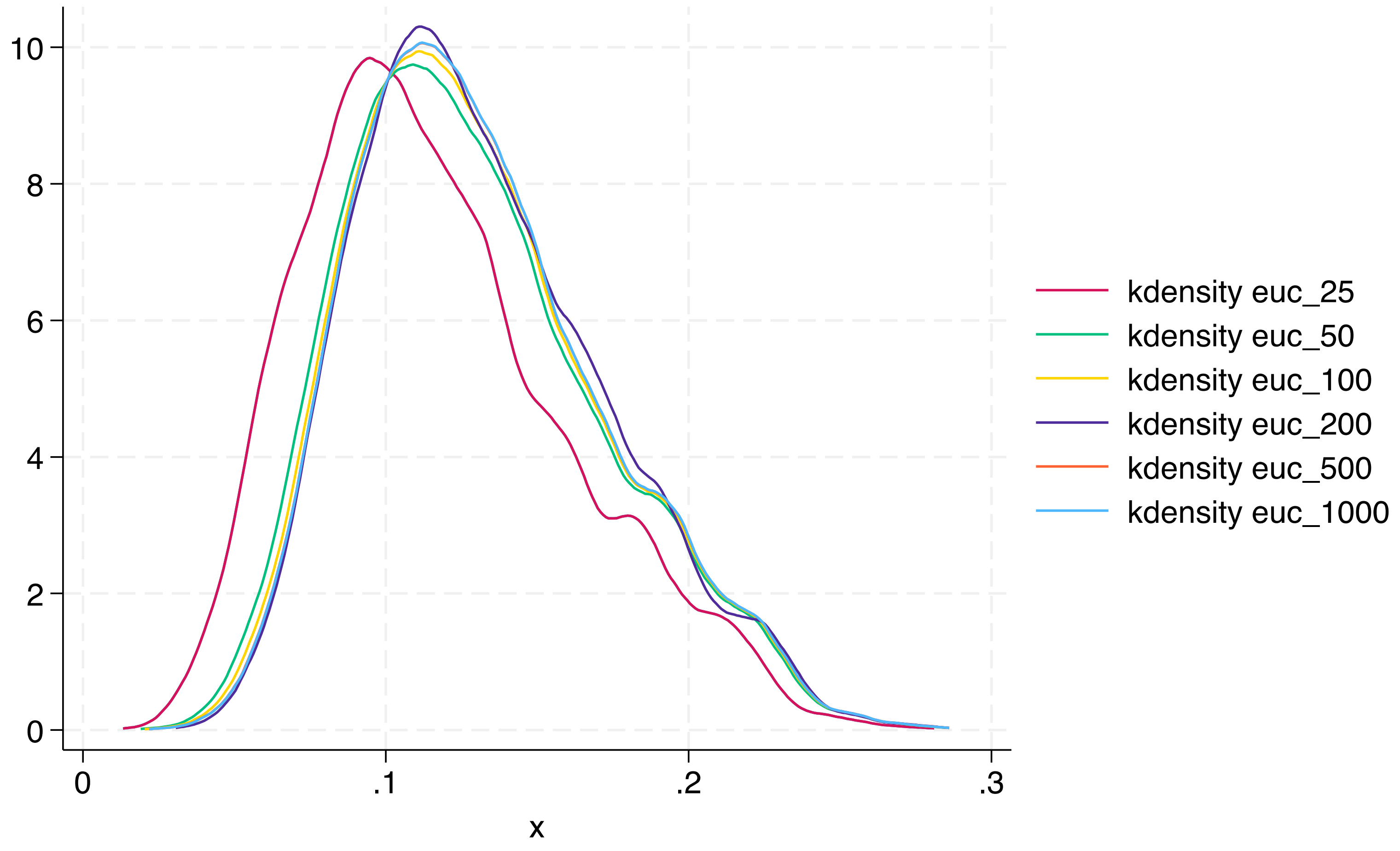


Judges, Lawyers, and Related Workers (SOC2 231)

1.3 Estimation

❖ A note on the choice of vector size:

Kernel density functions of bilateral occupation distances (Year = 2014) for various skill-vector sizes.



Semantic Skill – O*NET: Steps

- **Steps:**
  - Obtain definitions from Skills(sourced from Wikipedia or GPT generated if missing)
  - Obtain definitions for 125 O*NET clusters (following Atalay et al.)
  - Embed using a sentence-transformers open-source library (SBERT) (cross-encoder) on a pre-trained LLM model
  - For each of the 16000 skills find top-1 matches to O*NET clusters

Element Name	value	SKILL_US
Speaking	2A1d	Communication Skills
Customer and Personal Service	2C1e	Customer Service
Coordinating the Work and Activities of Others	4A4b1	Teamwork / Collaboration
Sales and Marketing	2C1d	Sales
Administration and Management	2C1a	Organizational Skills
Complex Problem Solving	2B2i	Problem Solving
Processing Information	4A2a2	Microsoft Excel
Scheduling Work and Activities	4A2b5	Scheduling
Organizing, Planning, and Prioritizing Work	4A2b6	Planning
Performing General Physical Activities	4A3a1	Physical Abilities
Working with Computers	4A3b1	Computer Literacy
Attention to Detail	1C5b	Detail-Oriented
Reading Comprehension	2A1a	Writing
Administrative	2C1b	Microsoft Office

1.1 Data (1950-2000)

US Data

- Job vacancy data provided by *Atalay et al.(2020)*
- Collected from New York Times (from 1940 to 2000), Wall Street Journal (from 1940 to 1998), and Boston Globe (from 1960 to 1983)
- NLP techniques to extract keywords from every post. Categorize these occupational characteristics(keywords) into groups: O*NET's work styles, skills, knowledge requirements,and work activities
- Includes info on all O*NET's work styles, skills, knowledge requirements,and work activities and occupation

Keywords to O*NET

Job Characteristic	CBOW Model
O*NET Work Activity 4A1a1 Getting Information	'details', 'info', 'gelling', 'formation', 'facts', 'informant', 'gaffing', 'qet', 'fads', 'finding'
O*NET Work Activity 4A1a2 Monitor Processes, Materials, or Surroundings	'material', 'etc', 'procedures', 'methods', 'monitoring', 'process', 'documents', 'forms', 'activities', 'records'
O*NET Work Activity 4A1b1 Identifying Objects, Actions, and Events	'activities', 'issues', 'action', 'measures', 'changes', 'occurrences', 'processes', 'strategies', 'activity', 'risks'
O*NET Work Activity 4A1b3 Estimating the Quantifiable Characteristics of Products, Events	'analysis', 'materials', 'studies', 'components', 'data', 'product', 'activities', 'experience', 'software', 'design'

1.2 Methodology

We repeat the same methodology

Step 1: Create corpus C_{gt}: e.g. *Purchasing Managers (SOC 11-3061)* in 1985

Step 2: Compute freq. measure of each (O*NET) skill in the corpus.

Step 3: Express occupation as vector of skills C_{gt} ranked from most important to least

C_{gt} ={4A1a1 Getting Information ; Monitor Processes, Materials; …}

Then use Euclidean distance(or cosine) to measure similarity between two such vectors representing two different corpuses (occupations):

$$d_{ab} = \sqrt{\sum_{i}^{S} (b_i - a_i)^2}$$

1.3 Estimation

Insight 1: Bilateral distances across occ (SOC6) pairs are falling over time

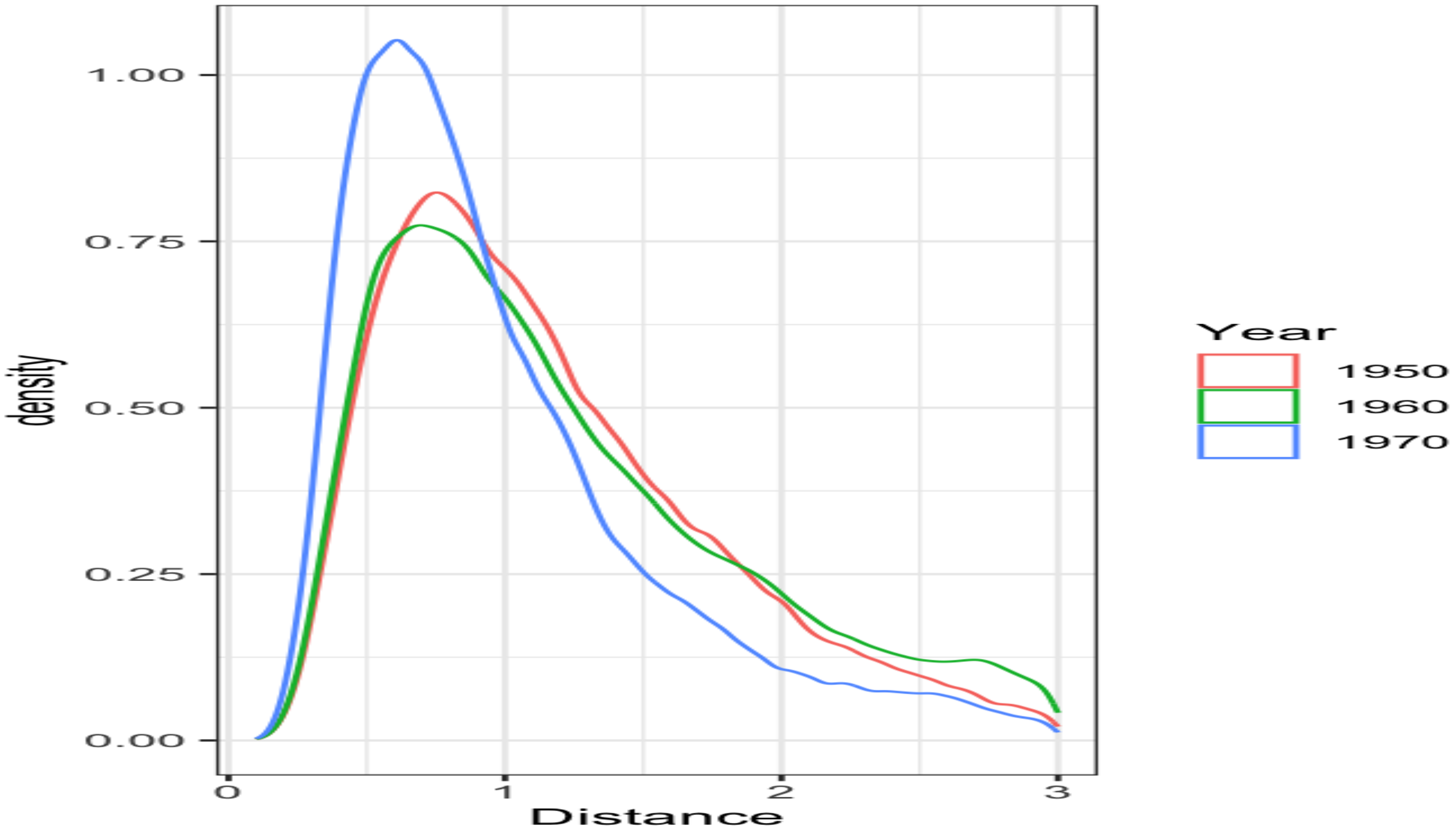


Figure 1 – Distribution of bilateral distances for SOC6 pairs over time

1.3 Estimation

GEORGETOWN UNIVERSITY QATAR

Insight 1: Bilateral distances across occ (SOC6) pairs are falling over time

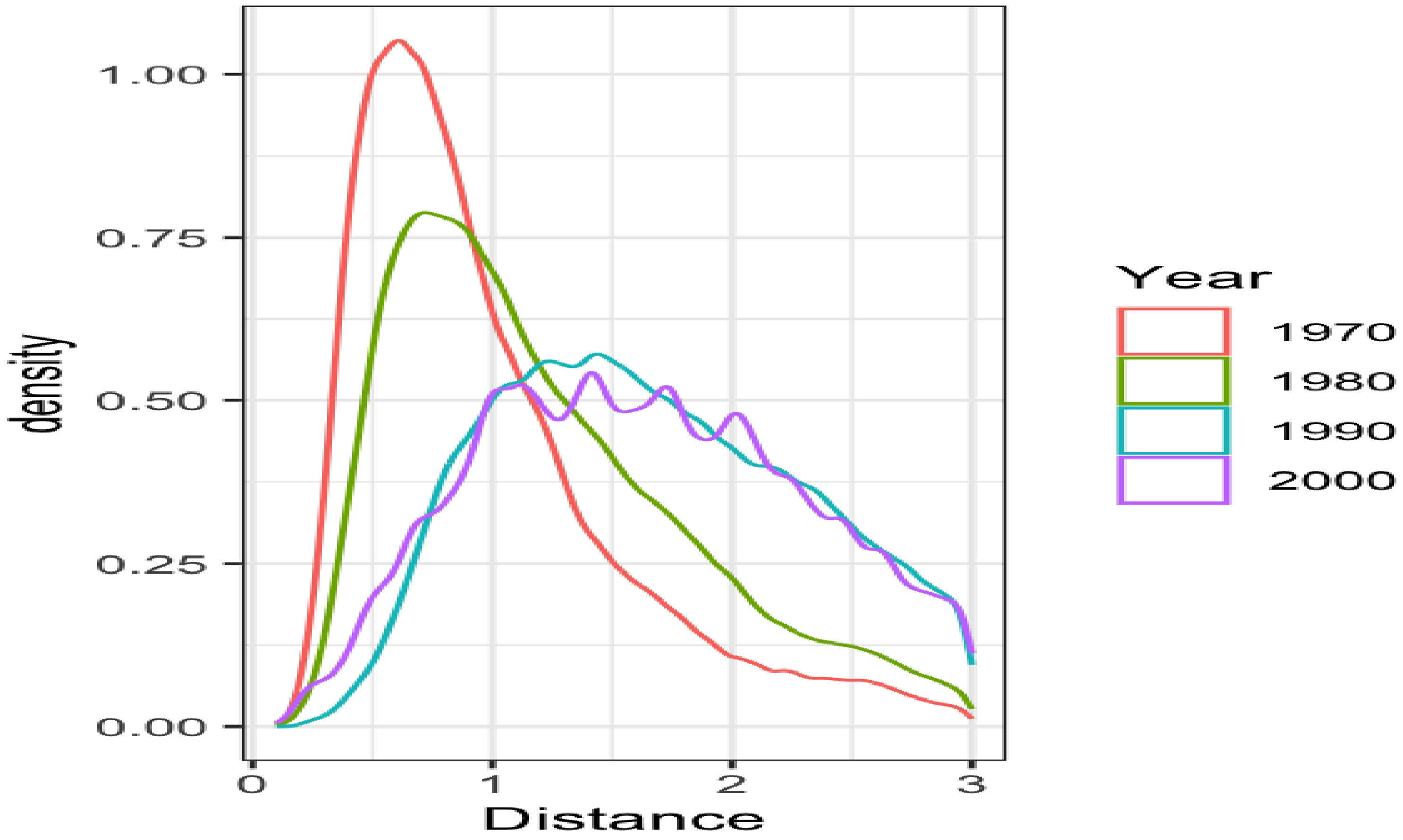


Figure 1 – Distribution of bilateral distances for SOC6 pairs over time

1.3 1950-2000

Insight 1: Bilateral distances across occ (SOC6) pairs is increasing over time

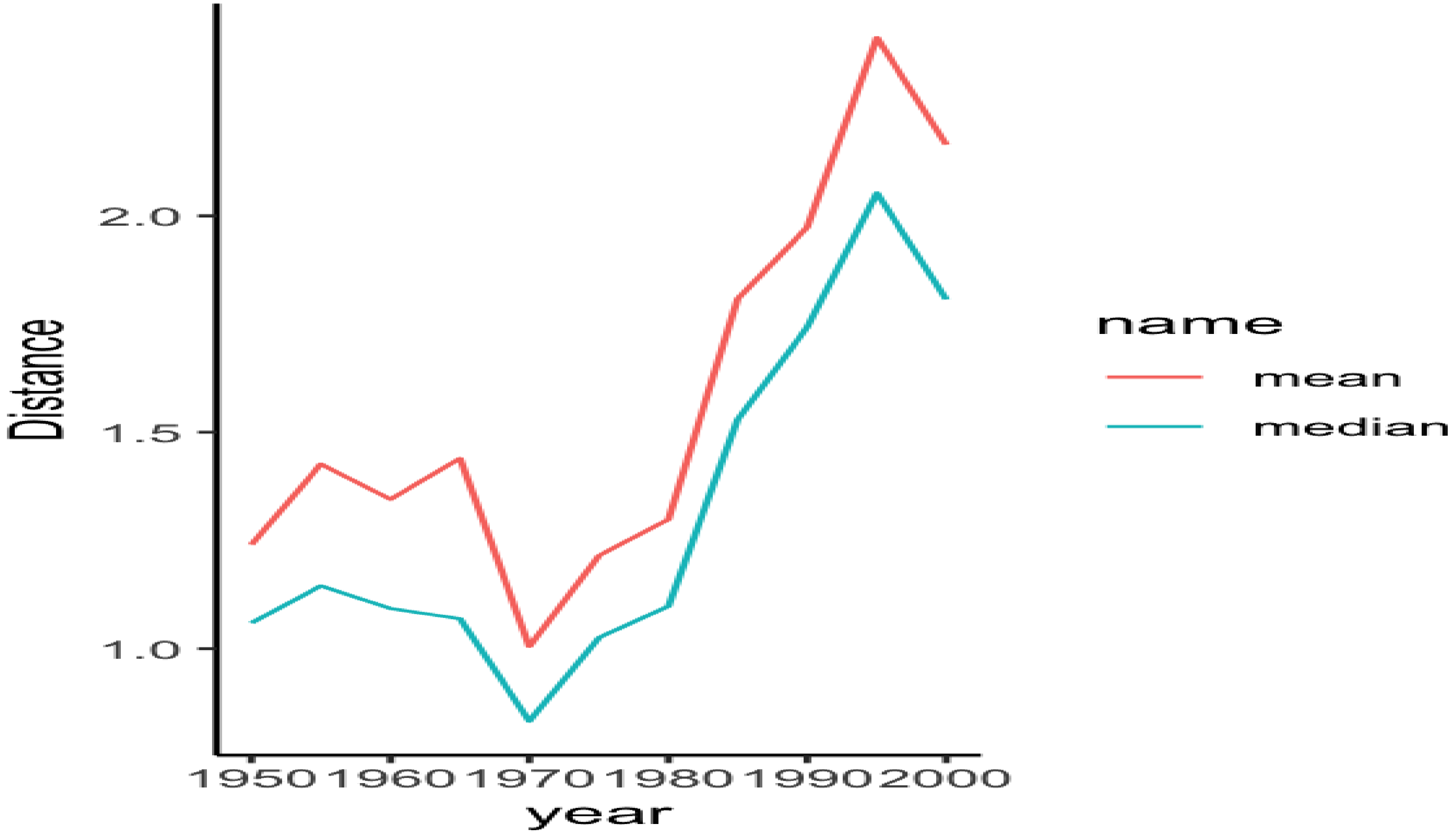


Figure 2 – Average and median Euclidean distance across all occupations by year

2.4 Additional results

GEORGETOWN UNIVERSITY QATAR

❖ **Self-distance:**

1. Occupations that have transformed the most have experienced bigger declines in *within* occupation wage inequality between 2014 and 2022.
2. Occupations that have transformed the most experienced a higher increase in wage growth between 2014 and 2022

3 Extensions

GEORGETOWN UNIVERSITY QATAR

Back

Finding 1: The higher the transformation, the higher the fall in within inequality

	(1)	(2)
Ineq_2014	-0.190***	-0.192***
	(0.0241)	(0.0376)
Low NR-C		-0.0420
		(0.0301)
Low NR-C * Ineq_2014		-0.0687
		(0.0505)
Self Distance	-0.147**	-0.333**
	(0.0605)	(0.132)
Low NR-C *Self Distance		0.290*
		(0.161)
Constant	0.0578***	0.102***
	(0.0143)	(0.0231)
Observations	753	637
R-squared	0.080	0.170

Standard errors in parenthesis, *** p<0.01, ** p<0.05, * p<0.1

3 Application:

❖ Are cities converging?

- ❑ United States

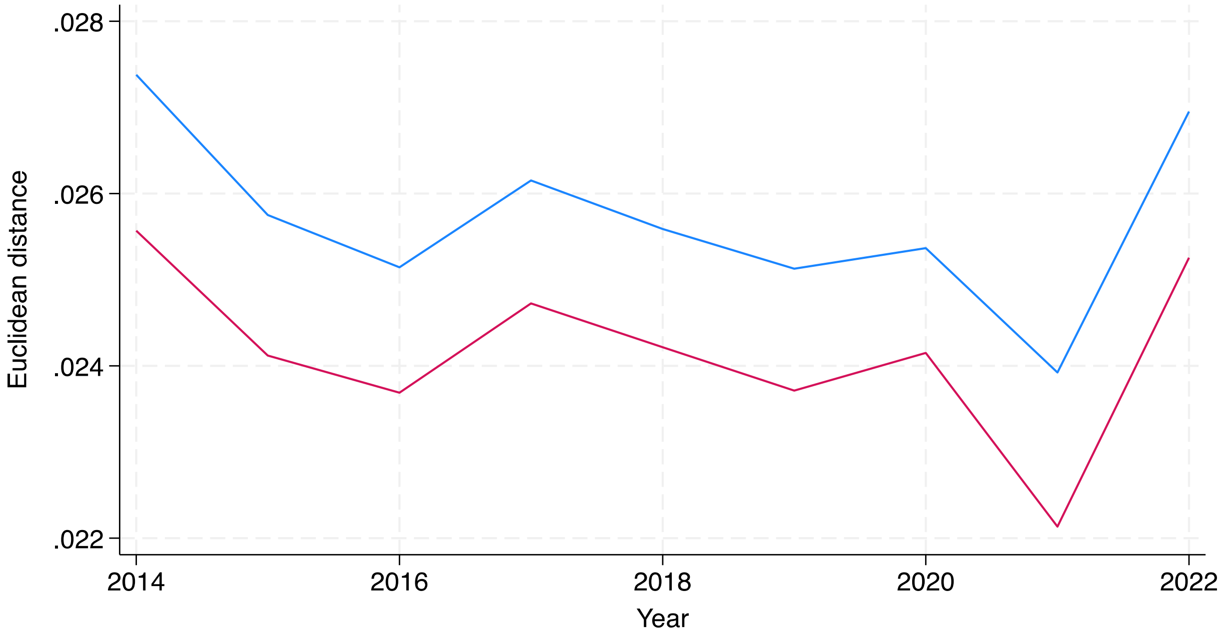


Fig 1 - Bilateral distanced across US cities

- ❑ Compare distance between US cities (MSAs) and Chinese cities

Beijing: San Jose-Sunnyvale-Santa Clara; San Francisco-Oakland-Hayward

Median

Wuhan: Seattle-Tacoma-Bellevue; Chicago-Naperville-Elgin

Shanghai: San Francisco-Oakland-Hayward; Boston-Cambridge-Newton

Foreign: Reno; Las Vegas-Henderson-Paradise

“The Unexpected Wage Compression”

❖ Wages across occupations (Weighted)

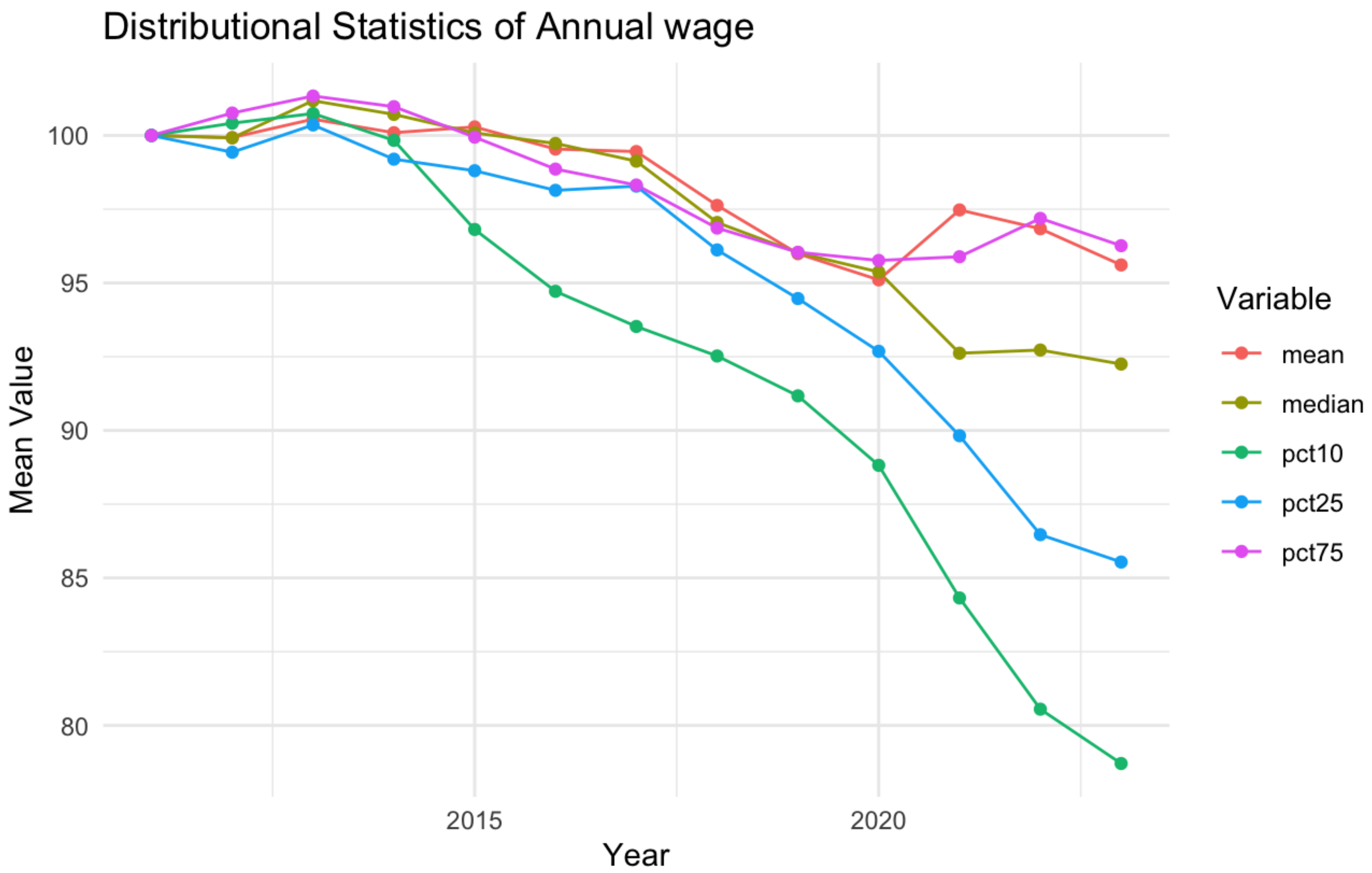
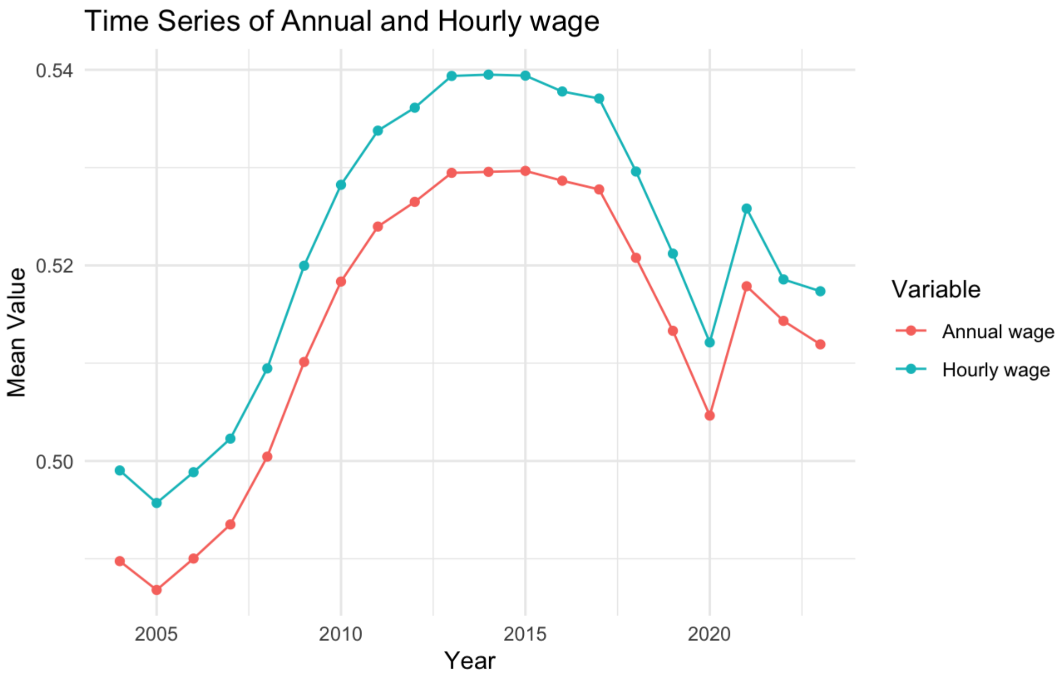
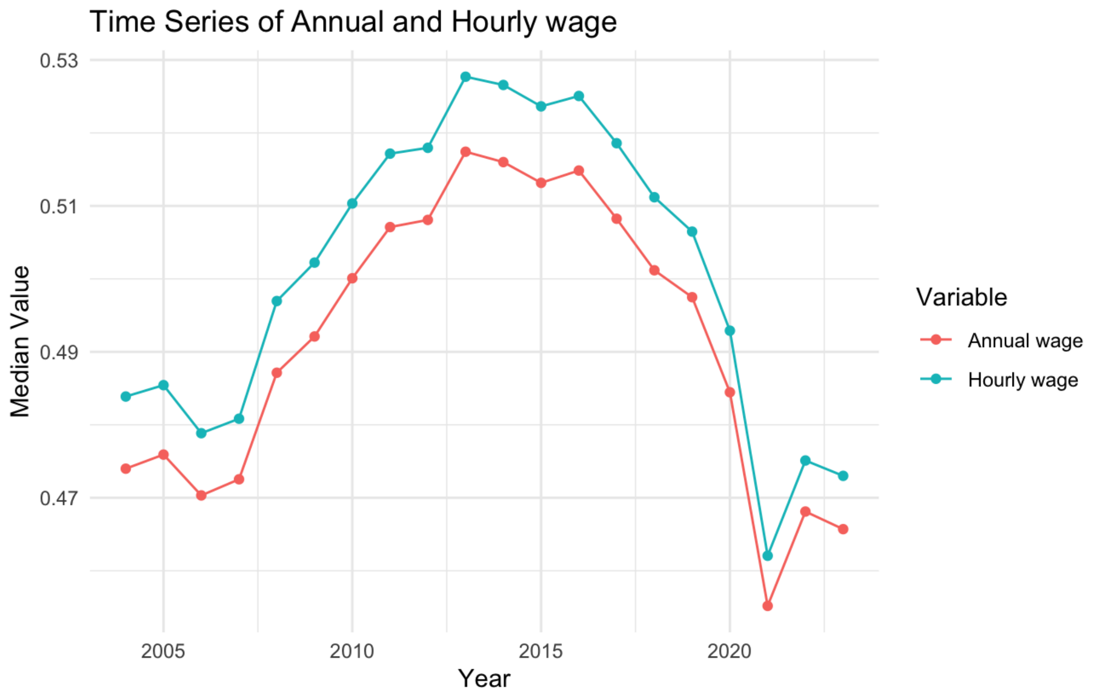


Fig 1 – Coef of variation of different stats (OCC)

“The Unexpected Wage Compression”

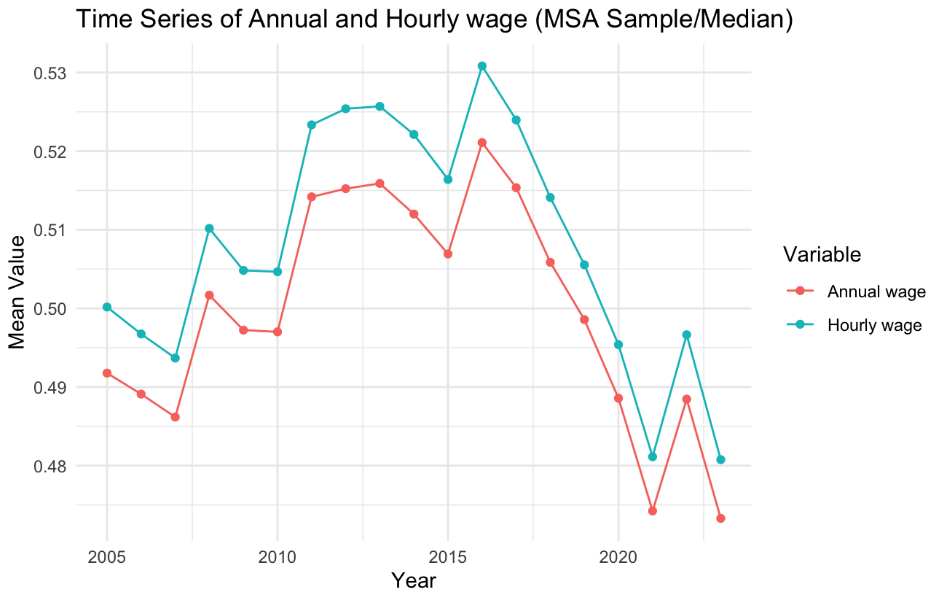
❖ Weighted by employment of each occupation(Weighted)

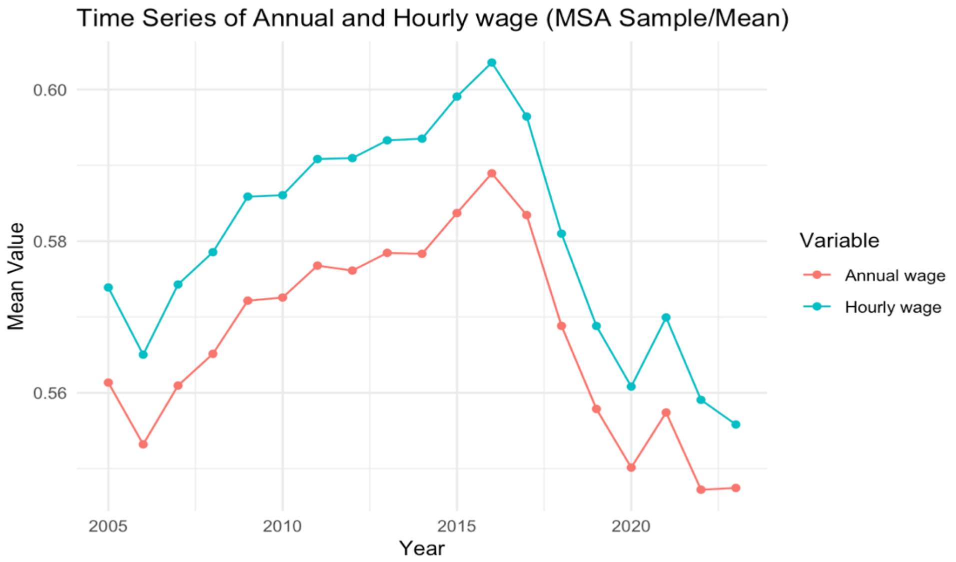




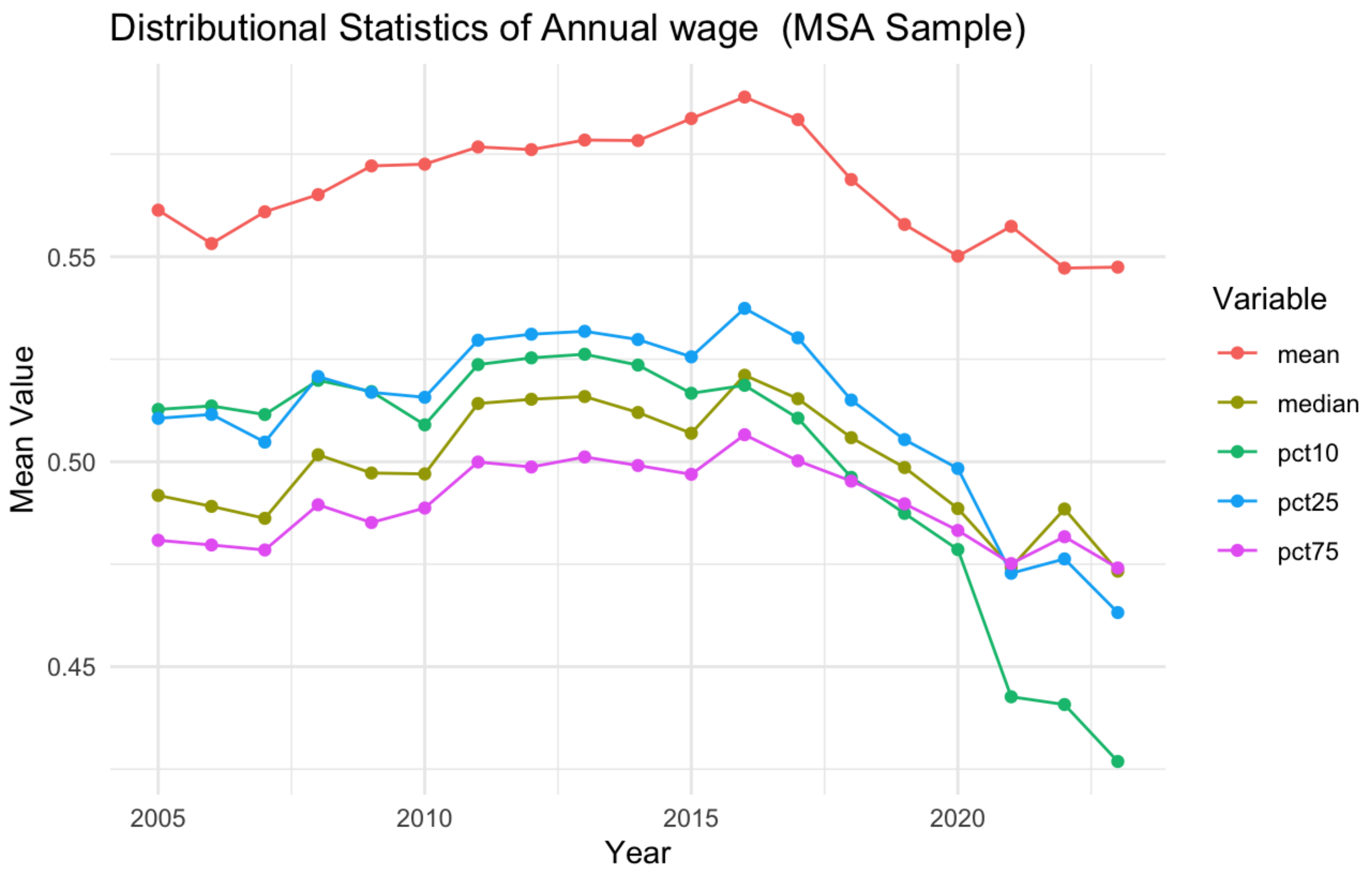
A New Stylized Fact

❖Coef. Variation Median and Mean over time (Within MSA variation_





Replicating using MSA level occupation wages

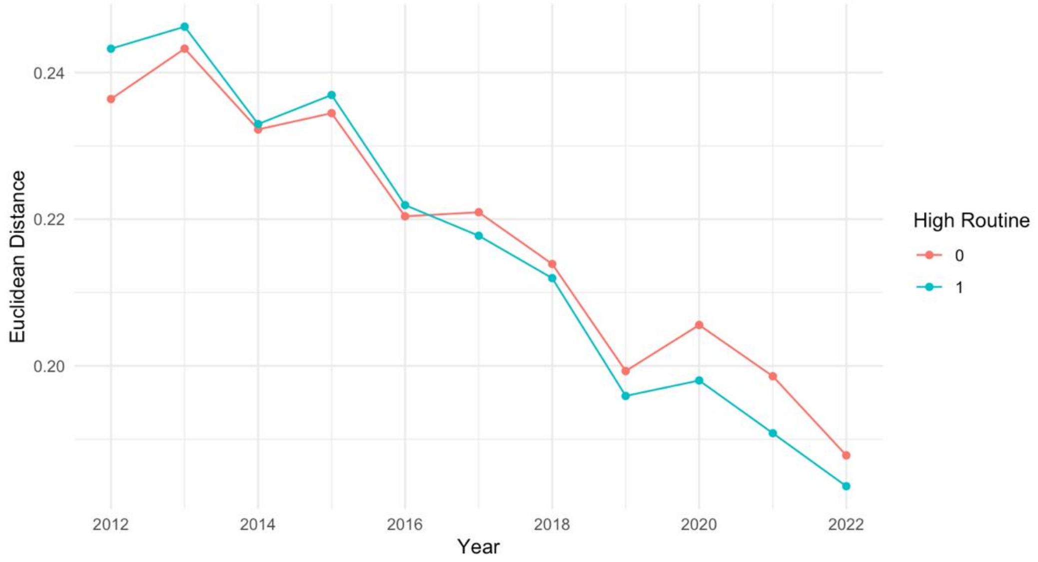


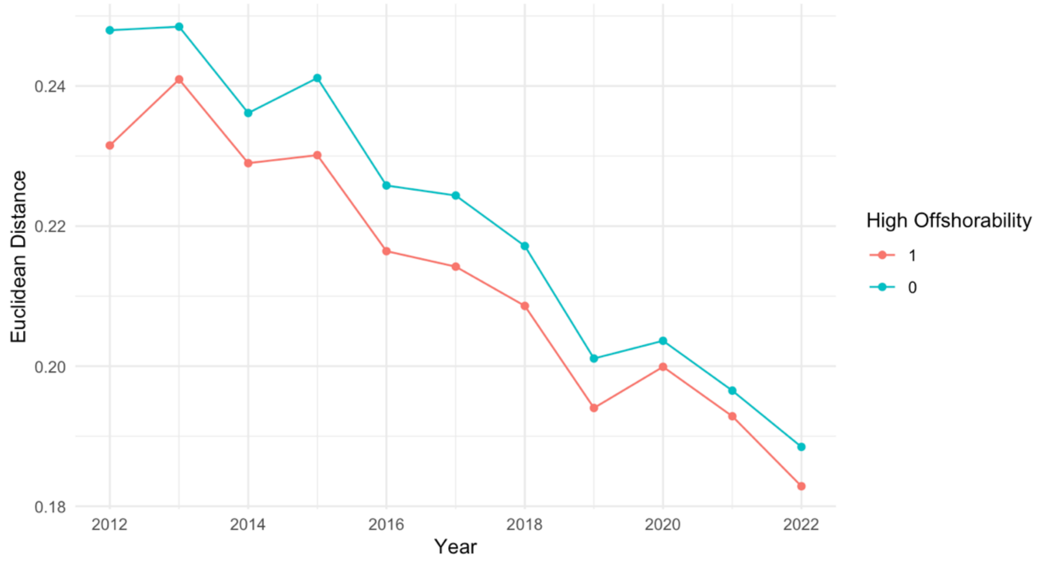
OEWS data are also published fpr major cities(we keep these that report positive employment for at least XXX occupations.

For each SOC6-pair, we have annual and hourly earnings for the following stats(mean, median, p10,p25,p75) – calculate coef. of variation at the MSA level and then average to calculate the mean across MSAs – as a counterpart to what we do with the national OEWS sample

Distance Dynamics over time across groups

GEORGETOWN UNIVERSITY QATAR





BGT Skills-OCC occupations

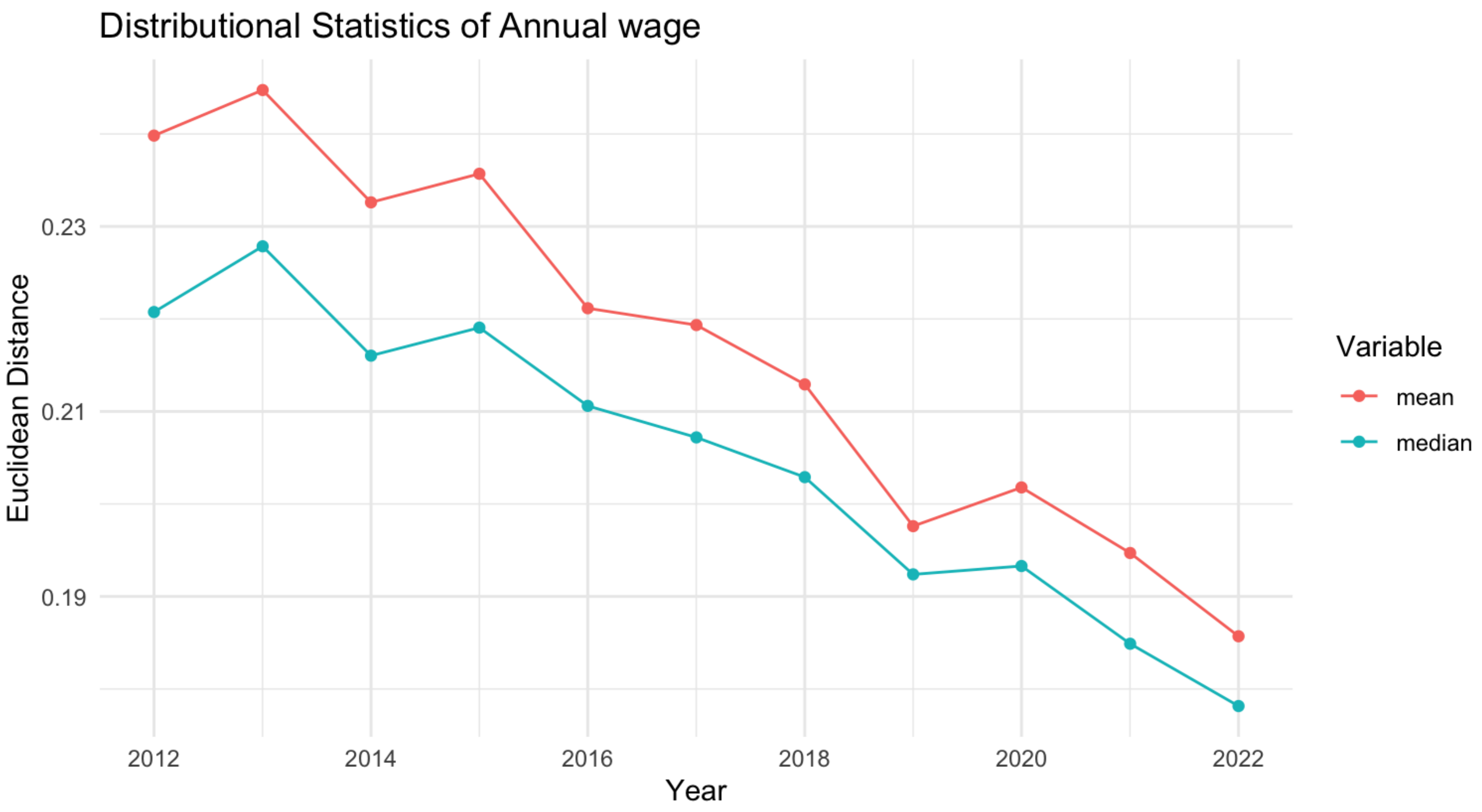
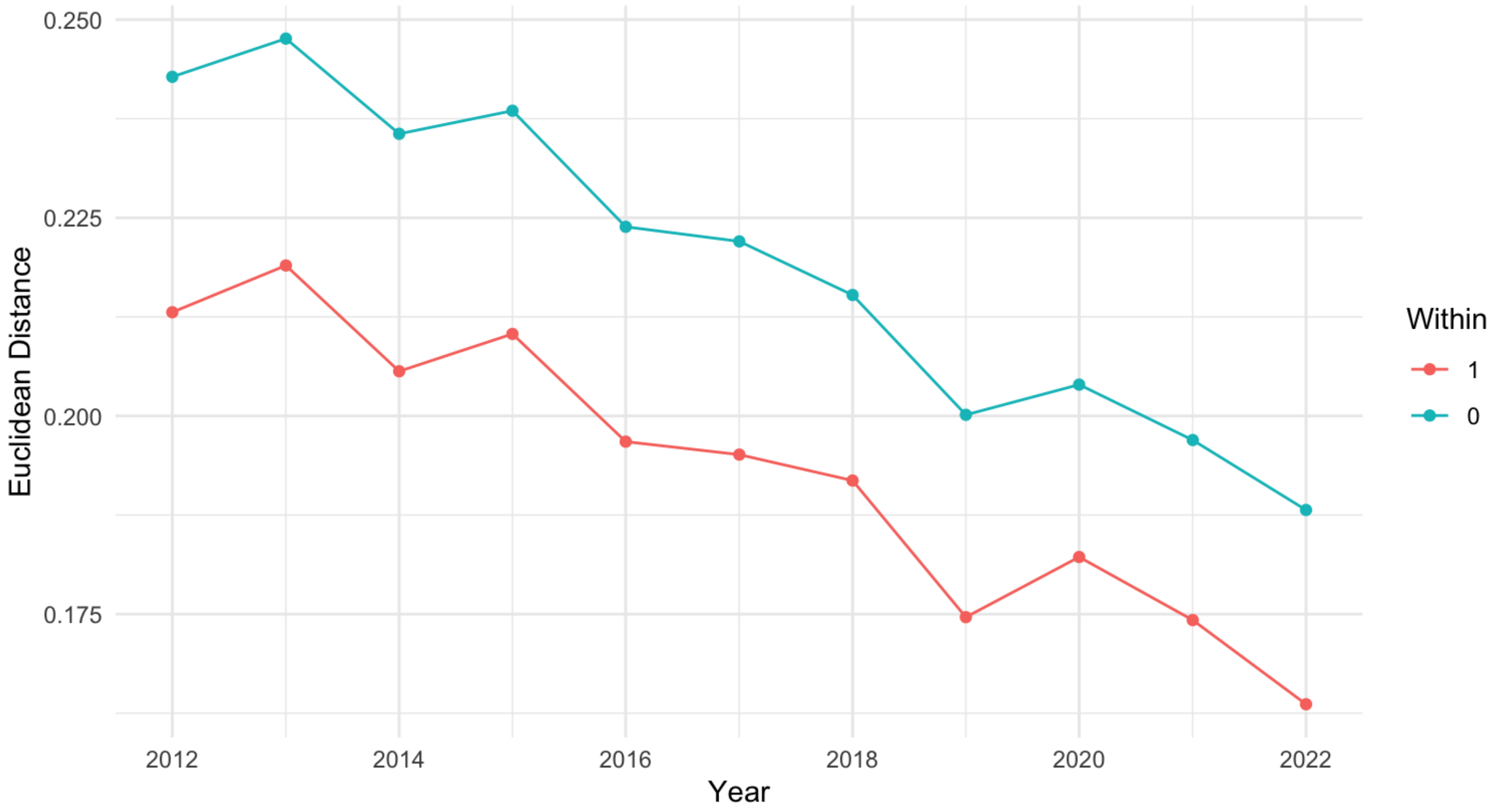


Figure 2 – Average and median Euclidean distance across all occupations by year

1.4 Top 250

1. Within group occupations are closer than across group

Figure 3 – Median Euclidean distance across and within SOC2 groups



1.4 Validity 250

3. Closer occupations have higher mobility

Dependent Variables:	Transition Share				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Distance	-0.0607***	-0.0533***	-0.0595***	-0.1055***	-0.1670***
	(0.0016)	(0.0015)	(0.0016)	(0.0034)	(0.0140)
Same		0.0033***	0.0036***	0.0053***	0.0205***
		(0.0002)	(0.0003)	(0.0004)	(0.0021)
Distance × Same		-0.0151***	-0.0168***	-0.0266***	-0.0310***
		(0.0012)	(0.0014)	(0.0029)	(0.0111)
Fixed-effects					
Occupation 1	Yes	Yes	Yes	Yes	Yes
Occupation 2	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	696,390	696,390	666,672	275,733	93,721
R^2	0.32601	0.32931	0.34109	0.37861	0.05271
Within R^2	0.03567	0.04038	0.04392	0.07064	0.03358