Robots, labour-saving technologies, and occupational exposure

Fabio Montobbio[a b] **Jacopo Staccioli**[a c] M. Enrica Virgillito[c a] Marco Vivarelli[a]

[a] *Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano*

[b] *BRICK, Collegio Carlo Alberto, Torino*

[c] *Institute of Economics, Scuola Superiore Sant'Anna, Pisa*

IPR Course – University of Milan, Italy

15th November 2022

UNIVERSITÀ CATTOLICA del Sacro Cuore

Outline

Context and motivation

- the impact of automation upon employment has become a major topic of discussion both in policy and academic debate

Brynjolfsson and McAfee (2011, 2014)
: the root of current unemployment is not the Great Recession, but rather a 'Great Restructuring' characterised by an exponential growth in computers' processing power having an ever-bigger impact on jobs, skills, and the whole economy (*"This time is different"*)

Frey and Osborne (2017)
: 47% of the occupational categories are at high risk of being automated, including services and highly cognitive jobs

Context and motivation (cont'd)

> *"Automated systems, such as robotic systems, are used in a variety of industries to* ***reduce labo[u]r costs and/or increase productivity****. Additionally, the use of human operators can involve increased cost relative to automated systems."* [US20170178485A1]

> *"The use of [robotic] technology results in improved management of information, services, and data, increased efficiency, significant reduction of time,* ***decreased manpower requirements****, and substantial cost savings."* [US2010022134A1]

Our contribution

- we use natural language processing and probabilistic topic modelling techniques on the universe of 2009–2018 patent applications at USPTO, matched with ORBIS (BvD)
- we investigate the presence of explicit labour-saving heuristics among robotic patents
- we include not only patents entailing robotic artefacts as a *product* but also as *process* and complementary technology
- we analyse innovative actors engaged in robotic technology and their economic environment (identity, location, industry)
- we identify the technological fields that are particularly exposed to labour-saving innovations
- we pinpoint the technological bottlenecks underlying the search efforts inspiring robotics inventors

Outline

Original data

- universe of USPTO patent applications from 1st January 2009 to 31st December 2018
- 3,557,435 full-text applications (hereafter, patents)

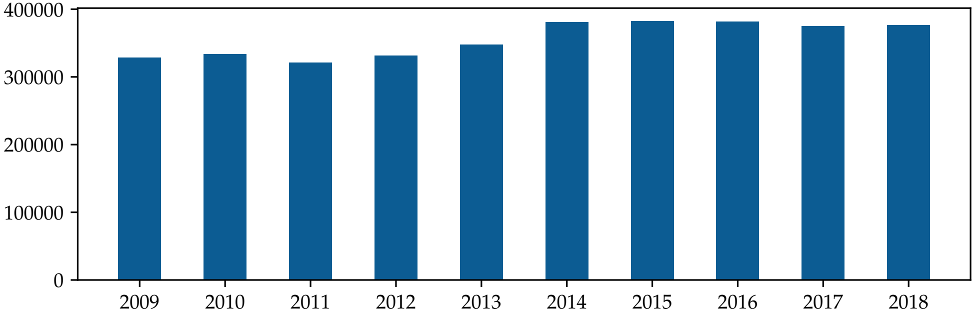


Figure: # of patents by year

Robotic patents

- identification of robotics[-related] patents
  1. via CPC codes
     - USPTO concordance table with USPC class 901
     - purely robotic technology
     - 10,929 `‘CPC’` patents
  2. via keyword search
     - multiple occurrence ($\times 10$) of morphological root *‘robot’*
     - process implementation and complementary technology
     - 18,860 `‘K10’` patents (once those already in 1 have been discarded)
- 29,789 total robotic patents

Robotic patents (cont'd)

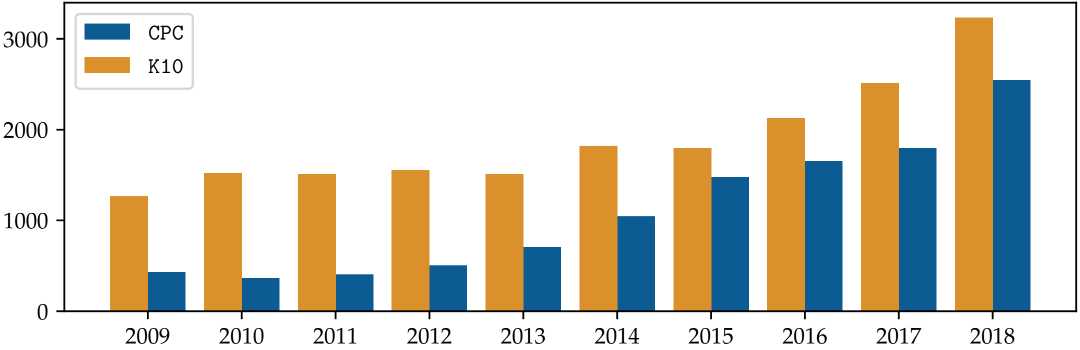


Figure: # of **robotic** patents by year

Text preprocessing

tokenisation

- each patent textual body is divided into *sentences* by means of a punctuation regexp
- patent text $\Longrightarrow$ list of sentences
- sentence $\Longrightarrow$ list of words

stemming

- each word in each sentence is reduced to its morphological root with the Porter2 stemming algorithm (an improved version of the original Porter (1980) algorithm)
- patent text $\Longrightarrow$ list of lists of stemmed words

- identification of labour-saving (LS) patents by means of a word-level text query per sentence

Labour-saving patents

$$\underbrace{\begin{bmatrix} \text{`reduc'} \\ \text{`replac'} \\ \text{`elimin'} \\ \text{`save'} \\ \text{`lower'} \\ \text{`substitut'} \\ \text{`autom'} \end{bmatrix}}_{\text{verbal predicate}} \times \underbrace{\begin{bmatrix} \text{`labor'} \\ \text{`worker'} \\ \text{`human'} \\ \text{`employe'} \\ \text{`manpow'} \\ \text{`job'} \end{bmatrix}}_{\text{direct object}} \times \underbrace{\begin{bmatrix} \text{`cost'} \\ \text{`expenditure'} \\ \text{`expens'} \\ \text{`hour'} \\ \text{`intens'} \\ \text{`task'} \\ \text{`time'} \\ \text{`skill'} \end{bmatrix}}_{\text{object attribute}}$$

- 336 combinations of triplets (**not** *trigrams*, as we do not require adjacency)
- a patent is flagged as *potentially* LS if contains at least one triplet within a sentence
- 1,666 potentially LS patents

Labour-saving patents (cont'd)

- all matched sentences are manually examined and flagged as *explicitly* LS if appropriate
- 1,276 explicitly LS patents ($\approx$ 77% of potentially LS; $\approx$ 4.3% of robotic patents)
- of which 461 ($\approx$ 36.1%) are CPC and 815 ($\approx$ 63.9%) are K10

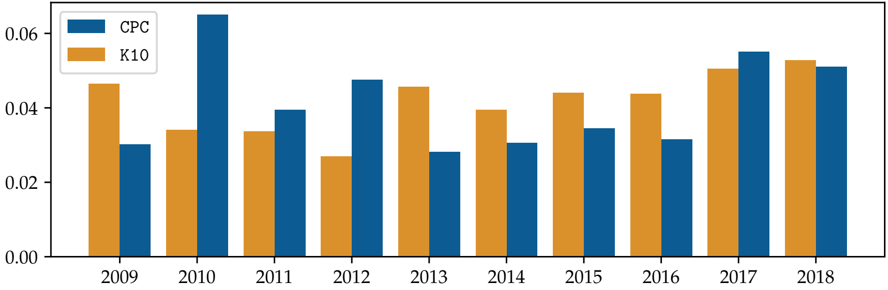


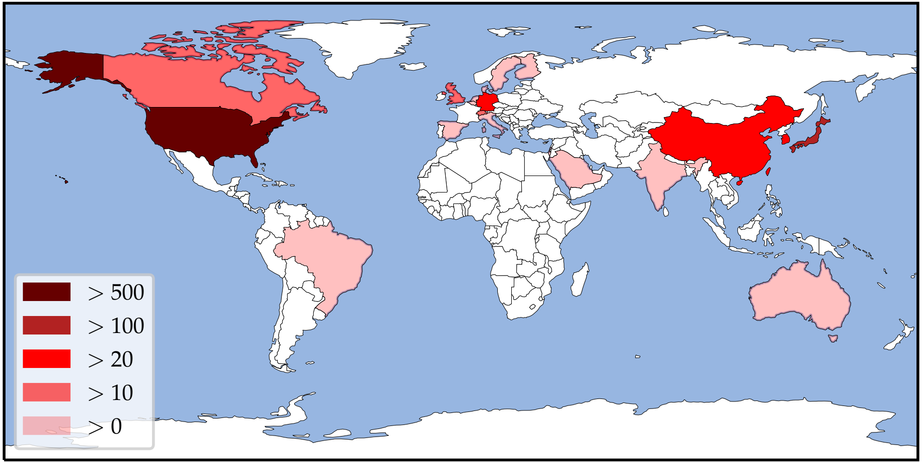
Figure: Fraction of **explicitly** LS patents over **robotic** patents by year

Firm level match

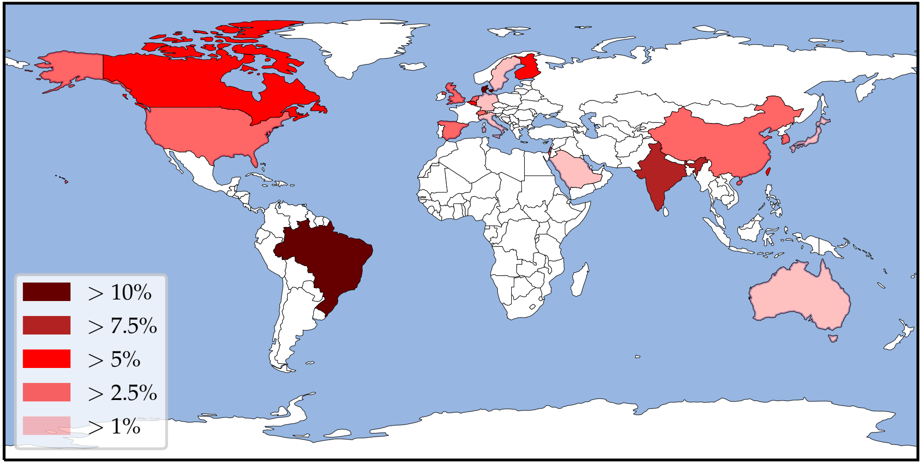
- LS patents are matched to their assignee via ORBIS (BvD)
- number reduces to 1,136 ($\approx$ 89%) due to truncation on 31st July 2018 (140 discarded)
- of these, 903 ($\approx$ 79%) are matched to at least one firm (233 find no match)
- there are 408 LS firms in total

Outline

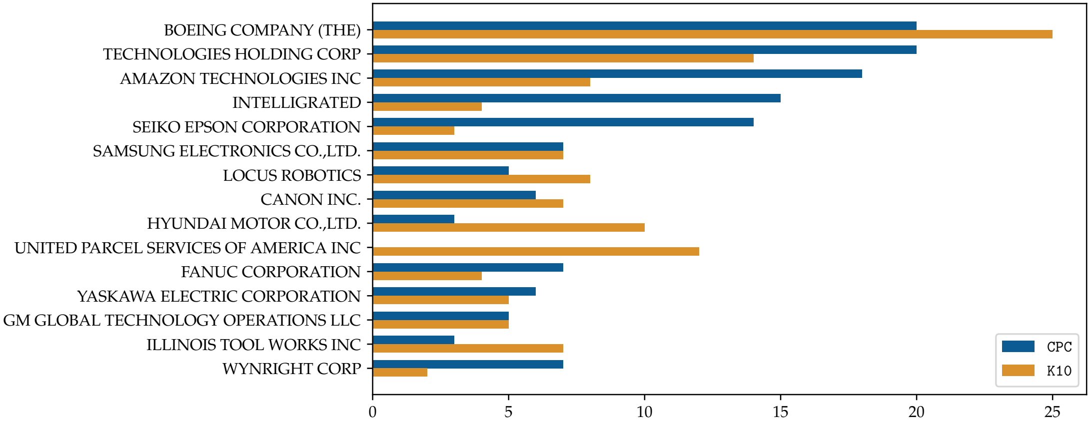
LS patents by country – absolute value



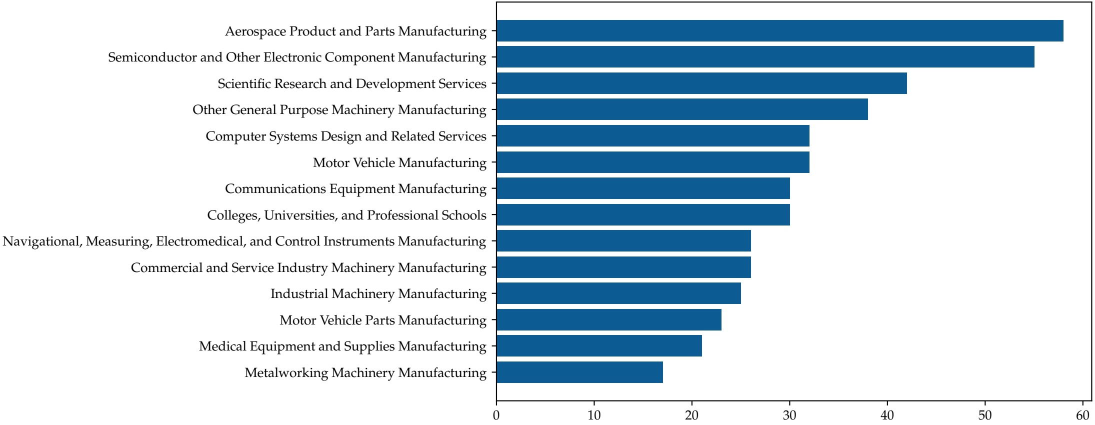
LS patents by country – as % of robotic patents



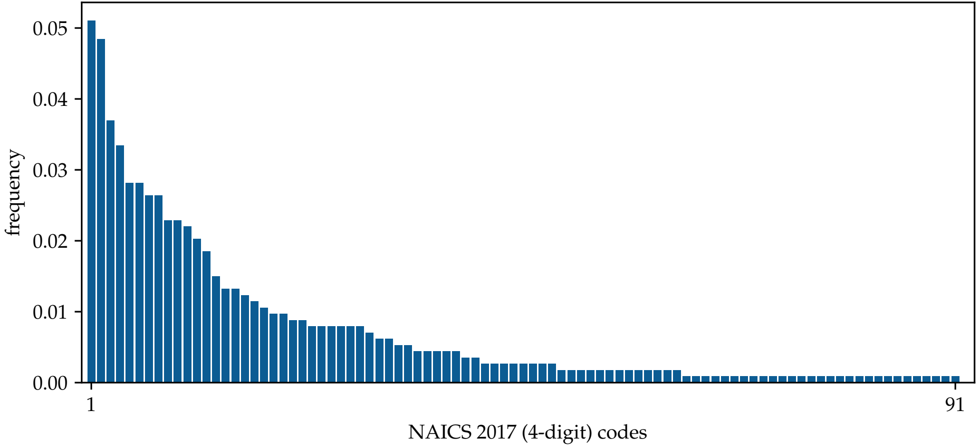
LS patents by assignee



LS patents by industry (cont'd)



LS patents by industry



Outline

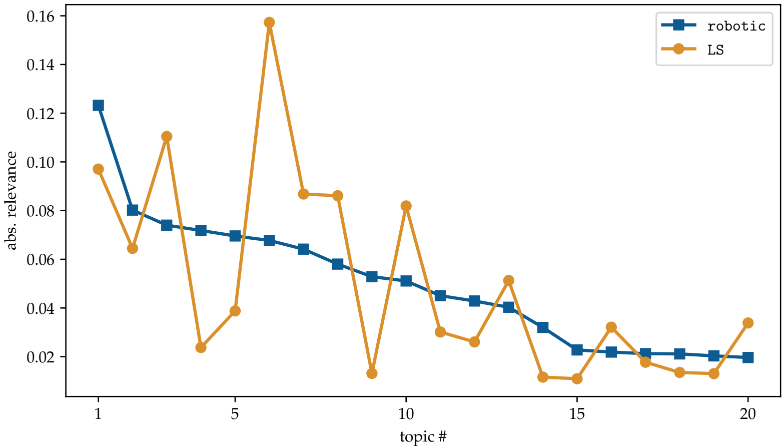
Probabilistic topic model

1. we estimate a topic model with $K = 20$ topics on the whole collection of robotic patents D
   - each topic $k \in K$ is identified as a list of keywords ranked by frequency
   - each patent $d \in D$ is assigned a distribution $\theta_{d,k}$ over the K topics
2. we assign a significance measure of CPC codes ($c \in C$) originally attributed to patents to each topic k by leveraging on the *latent semantic structure* of the whole collection of patents, through relevance distributions $\theta_{d,k}$ obtained in 1

$$\varphi_{c,k} = \sum_{d \in D} \mathbf{1}_{\{c \in d\}} \cdot \theta_{d,k} \qquad \forall\, k = 1, \ldots, K; \quad \forall\, c \in C$$

   - this brings useful information for labelling the topics
3. we compare the relevance of the K topics for robotic patents and the subset of LS patents

Topic relevance for robotic and LS patents



Technological taxonomy

Topic #	LS relev.	Words	CPC	Weight	Description
6	+132.2%	carrier	B65	24.4%	Conveying; packing; storing; handling thin or filamentary material
		conveyor			
		item	H01	6.8%	Basic electric elements
		gripper	G11	6.0%	Information storage
		tape	Y02	4.6%	Technologies or applications for mitigation or adaptation against climate change
			B23	4.3%	Machine tools; metal-working not otherwise provided for
...	...	...	...	...	...
...	...	...	...	...	...
9	−75.2%	heater	H01	8.6%	Basic electric elements
		hydrocarbon	E21	6.6%	Earth drilling; mining
		pipe	B23	5.5%	Machine tools; metal-working not otherwise provided for
		drill	Y10T29	4.4%	Metal working
		gas	Y02	4.4%	Technologies or applications for mitigation or adaptation against climate change

Outline

Occupations and labour-saving patents

- we compute a text similarity measure between technological codes and tasks

CPC corpus

- technological definitions from CPC `v.2019.08`
- 671 4-digit CPC codes

Task–Occupation corpus

- tasks description from O*NET `v.25.1`
- 19,231 tasks mapped to 923 8-digit SOC2018 occupations

preprocessing: every piece of text is *tokenised*, *stemmed*, and *stop words* are removed

Document-term matrix

1. construct the *document-term matrix* $\mathcal{D}_{CPC}$ of the corpus D of CPC definitions
   - each cell contains the frequency of term t in definition d
   - tf-idf: term frequency–inverse document frequency

$$\text{tf-idf}(t, d, D) := \text{tf}(t, d) \cdot \text{idf}(t, D)$$

$$\text{tf}(t, d) := \mathbf{1}_d(t) = \begin{cases} 1 & \text{if } t \in d \\ 0 & \text{otherwise} \end{cases}$$

$$\text{idf}(t, D) := \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right)$$

   - 671×2309 matrix

Cosine similarity

2. construct the *document-term matrix* $\mathcal{D}_{ONET}$ of the corpus of task descriptions
   - projected on the *vocabulary* of the CPC matrix $\mathcal{D}_{CPC}$
   - 19231×2309 matrix
3. construct the cosine similarity (CS) measure between the two corpora
   - for each couple of row vectors $X \in \mathcal{D}_{CPC}, Y \in \mathcal{D}_{ONET} \qquad (X, Y \in \mathbb{R}_{+}^{2309})$

$$\cos(X, Y) := \frac{X \cdot Y}{\|X\| \, \|Y\|} = \frac{\sum_t x_t y_t}{\sqrt{\sum_t x_t^2} \sqrt{\sum_t y_t^2}}$$

   - $\cos(X, Y) \in [0, 1]$ since vectors X and Y are non-negative valued
   - w.r.t. Euclidean distance, cosine similarity normalises for varying lengths of documents
   - 671×19231 cosine similarity matrix[1]
   - each task obtains a similarity score to each CPC code (12,904,001 pairs)

[1] under tf-idf, it is possible to show that $\cos(\mathcal{D}_{CPC}, \mathcal{D}_{ONET}) \equiv \mathcal{D}_{CPC} \cdot \mathcal{D}'_{ONET}$

Cosine similarity (cont'd)

OCCUPATION	11-1011.00			...		53-7121.00	
TASK	8823	8824	...	...	...	12809	12810
CPC							
A01B	cos(A01B,8823)	cos(A01B,8824)	...	...	...	cos(A01B,12809)	cos(A01B,12810)
A01D	cos(A01D,8823)	cos(A01D,8824)	...	...	...	cos(A01D,12809)	cos(A01D,12810)
...	...	...	...	...	...	...	...
H05H	cos(H05H,8823)	cos(H05H,8824)	...	...	...	cos(H05H,12809)	cos(H05H,12810)
H05K	cos(H05K,8823)	cos(H05K,8824)	...	...	...	cos(H05K,12809)	cos(H05K,12810)

4. weight by CPC frequency in LS patents[2], sum across CPCs, and rescale between [0,1]

[2] codes B25*, G01*, G05*, G06*, and Y* are excluded because too general

From tasks to occupations

- each O*NET occupation consists of a number of *core* and *supplementary* tasks
- we attribute task CS to occupations with weights

$$\text{core}: \quad \frac{2/3}{\#\text{ tasks in the occupation}}$$

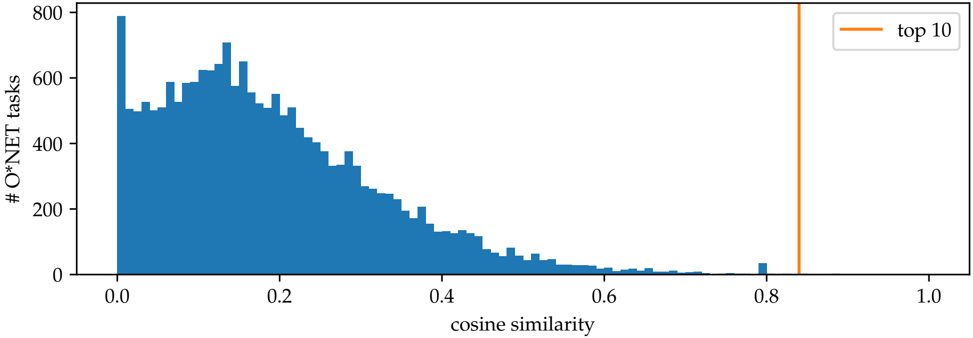
$$\text{supplementary}: \quad \frac{1/3}{\#\text{ tasks in the occupation}}$$

  - this weighting scheme reflects O*NET cutoff between core and supplementary tasks (based on a blend of *frequency*, *importance*, and *relevance* to underlying occupation)

Top tasks by similarity

#	Code	Description	CS
1	14587	Load materials and products into machines and equipment, or onto conveyors, using hand tools and moving devices	1.0
2	3202	Move levers or controls that operate lifting devices, such as forklifts, lift beams with swivel-hooks, hoists, or elevating platforms, to load, unload, transport, or stack material	0.96
3	3203	Position lifting devices under, over, or around loaded pallets, skids, or boxes and secure material or products for transport to designated areas	0.9
4	17928	Lift and move loads, using cranes, hoists, and rigging, to install or repair hydroelectric system equipment or infrastructure	0.89
5	15266	Manually or mechanically load or unload materials from pallets, skids, platforms, cars, lifting devices, or other transport vehicles	0.88
6	14584	Remove materials and products from machines and equipment, and place them in boxes, trucks or conveyors, using hand tools and moving devices	0.86
7	11839	Transport machine parts, tools, equipment, and other material between work areas and storage, using cranes, hoists, or dollies	0.85
8	3217	Load materials and products into package processing equipment	0.85
9	12805	Operate conveyors and equipment to transfer grain or other materials from transportation vehicles	0.85
10	12323	Communicate with systems operators to regulate and coordinate line voltages and transmission loads and frequencies	0.84

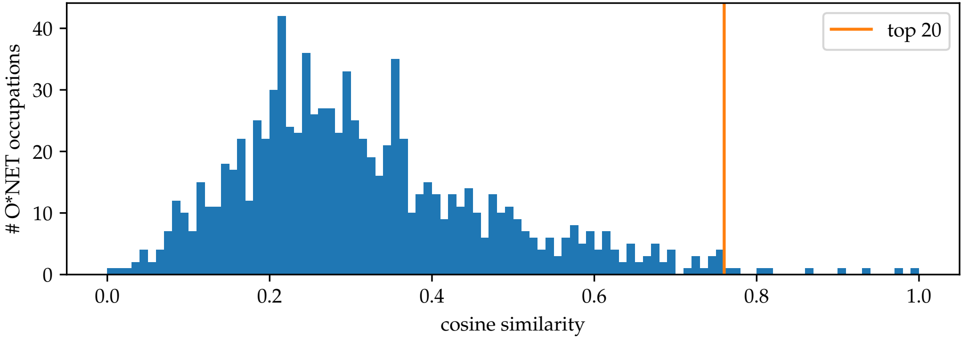
Tasks by similarity



Top occupations by similarity

#	Code	Title	CS
1	53-7051.00	Industrial Truck and Tractor Operators	1.0
2	49-9043.00	Maintenance Workers, Machinery	0.97
3	53-7063.00	Machine Feeders and Offbearers	0.94
4	53-7064.00	Packers and Packagers, Hand	0.91
5	49-2091.00	Avionics Technicians	0.87
6	51-9111.00	Packaging and Filling Machine Operators and Tenders	0.81
7	49-3041.00	Farm Equipment Mechanics and Service Technicians	0.81
8	49-3092.00	Recreational Vehicle Service Technicians	0.78
9	49-3042.00	Mobile Heavy Equipment Mechanics, Except Engines	0.77
10	47-2111.00	Electricians	0.76
11	49-9098.00	Helpers–Installation, Maintenance, and Repair Workers	0.75
12	49-9041.00	Industrial Machinery Mechanics	0.75
13	51-9082.00	Medical Appliance Technicians	0.75
14	47-3011.00	Helpers–Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters	0.75
15	51-9191.00	Adhesive Bonding Machine Operators and Tenders	0.75
16	51-9023.00	Mixing and Blending Machine Setters, Operators, and Tenders	0.74
17	13-1032.00	Insurance Appraisers, Auto Damage	0.73
18	51-4111.00	Tool and Die Makers	0.73
19	49-9081.00	Wind Turbine Service Technicians	0.72
20	51-8013.04	Hydroelectric Plant Technicians	0.72

Occupations by similarity



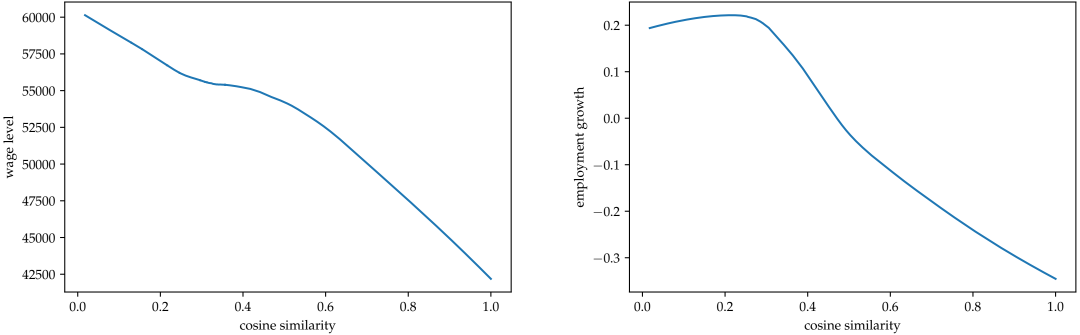
Occupational exposure and employment

- match with Occupational Employment Statistics (OES) from US Bureau of Labor Statistics
- employment (excluding self-employed) and median wage data for 6-digit SOC occupations

  2019 for levels

  1999 for 20-year growth rates

Wage levels and employment growth



- robust LOWESS estimates of the underlying scatter plots (bandwidth = 0.8)

Outline

Main findings

- LS firms are not only robots producers, but mainly adopters (archetypical cases are Boeing, Amazon, and UPS)
- the overall number of robotic patents is rapidly expanding (3-fold increase in a decade)
- conversely, LS patents do not exhibit a clear trend, supporting the idea that labour-saving is a rather established heuristic
- LS robotic patents emerge along the entire supply chain, signalling pervasiveness
- LS patents are concentrated in labour intensive industries (e.g. logistics, healthcare)
- technological bottlenecks identified by Frey and Osborne (2017) (occupations requiring social and cognitive intelligence, finger dexterity and manipulation) are under active research efforts by innovative firms

Main findings (cont'd)

- the cosine similarity matrix is overall very *sparse*
  - skewed distributions in both tasks and occupations
  - high similarity is a **rare event** (low probability of false positives)
- considering the top quartile of the similarity distribution, around 6.6% of employees (≈10m) are exposed to substitution
- we do not know how many workers a single machine is able to substitute

Main findings (cont'd)

- exposure to substitution is monotonically decreasing in wage
  - no U-shaped pattern but rather a negative declining relationship
- most affected occupations (2-digit) include "transportation and material moving" (logistics), "installation, maintenance, and repair" (automotive), "food preparation and serving"
- exposure to substitution is decreasing in employment growth
  - innovative efforts towards the weakest and cheapest segment of the labour market

Outline

References

- Brynjolfsson, E. and A. McAfee (2012) *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Digital Frontier Press.
- Brynjolfsson, E. and A. McAfee (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Frey, C.B. and M.A. Osborne (2017) The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114, pp. 254–280. DOI: 10.1016/j.techfore.2016.08.019
- Montobbio, F., J. Staccioli, M.E Virgillito, and M. Vivarelli (2022) "Robots and the origin of their labour-saving impact". *Technological Forecasting and Social Change* 174, 121122. DOI: 10.1016/j.techfore.2021.121122
- Montobbio, F., J. Staccioli, M.E Virgillito, and M. Vivarelli (2021) "Labour-saving automation: a direct measure of occupational exposure". *World Economy* [forthcoming]
- Squicciarini, M. and J. Staccioli (2022) *Labour-saving technologies and employment levels: are robots really making workers redundant?* OECD Science, Technology and Industry Policy Paper n. 124