

Skills mismatch across EU countries using job ads and survey data at the micro-level

Francesco Trentini

University of Milano-Bicocca, CRISP and LABORatorio R.Revelli



CRISP

*Interuniversity
Research Centre
for Public Services*



LABOR
LABORATORIO RICCARDO REVELLI
CENTRE FOR EMPLOYMENT STUDIES

Overview

- Skill mismatch. Definition and used measures
- Novel approach using
 - PIAAC – Program for the International Assessment of Adult Competences (OECD)
 - Online job ads from WIH-OJA (Cedefop and Eurostat)
 - Bridging them with ESCO Skills pillar using AI
- Application of the mapping to measure skill mismatch in 17 European countries

The research has been developed in the framework of the EU funded H2020 *Pillars – Pathways towards inclusive labour markets*.

Skill mismatch

Def deviation of workers' competencies from those required by employers to perform a job.

Supply

Demand

How to identify skill mismatch?

Identification of supply and demand in empirical research suffers limitations:

- Proxy for supply is usually highest level of education attained
- Demand is measured with surveys directed to employers to state which occupations or competences are needed

Approaches to measurement:

- **Normative:** ask experts to state the needed level of education or qualification to perform a job
- **Objective:** use the distribution of realised matches and measure the distance from, e.g., the median
- **Subjective:** workers are asked to self-assess the coherence between their education/qualification and the task that perform on the job

Limitations: low granularity and high heterogeneity within education classes

Berton, F., Devicienti, F., & Grubanov-Boskovic, S. (2017). Employment Protection Legislation and Mismatch: Evidence from a Reform. IZA Discussion Papers No. 10904

Novel approach using survey and big data

Data

Supply Program for the International Assessment of Adult Competences (OECD)

Demand Online Job Ads (WIH-OJA, Eurostat and Cedefop)

Methods

- For each occupation-skill compute the degree of specialisation (Revealed Comparative Advantages) on the demand and supply side
- Calculate the skill gap and average at the occupation level.
- Insight at the occupation 2 digit and NUTS2 in relation to training and risk of automation.

Limitation

PIAAC and OJA express skills with different taxonomies. How can we bridge the two?

Skill mismatch: data: supply - PIAAC

PIAAC – Program for the International Assessment of Adult Competences (OECD)

- Cycle I, all rounds: 2012, 2014 and 2017. Cycle II, currently ongoing; data release expected in 2023.
- Representative samples of working-age individuals
- Background questionnaire + test of the performance in some activities (e.g. writing an email, search for a job online)
- **Background questionnaire:** self-declared intensity (frequency or extent) of **skill use** in different domains:
 - at work (Module F)
 - Literacy, Numeracy and ICT at work (Module G)
 - Literacy, Numeracy and ICT in everyday life (Module H).
 - And a module on «[...] about how you deal with problems and tasks you encounter.» (Module I)
- PIAAC is used in the literature to measure skill content in a global perspective (Lewandowski et al. 2022) and risk of automation (Nedelkoska and Quintini, 2018)

Lewandowski, P., Park, A., Hardy, W., Du, Y., Wu, S. (2022) Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data, *The World Bank Economic Review*, <https://doi.org/10.1093/wber/lhac005>

Nedelkoska, L. and G. Quintini (2018). Automation, Skills Use and Training. Technical report, OECD, Paris.

Skill mismatch: data: supply - PIAAC

F_START Skills used at work

Layout	Item group table
F_Q02b (JRA) (B)	How often ^DoesDid your ^JobLastjob usually involve ... instructing, training or teaching people, individually or in groups? 1 Never 2 Less than once a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day DK RF

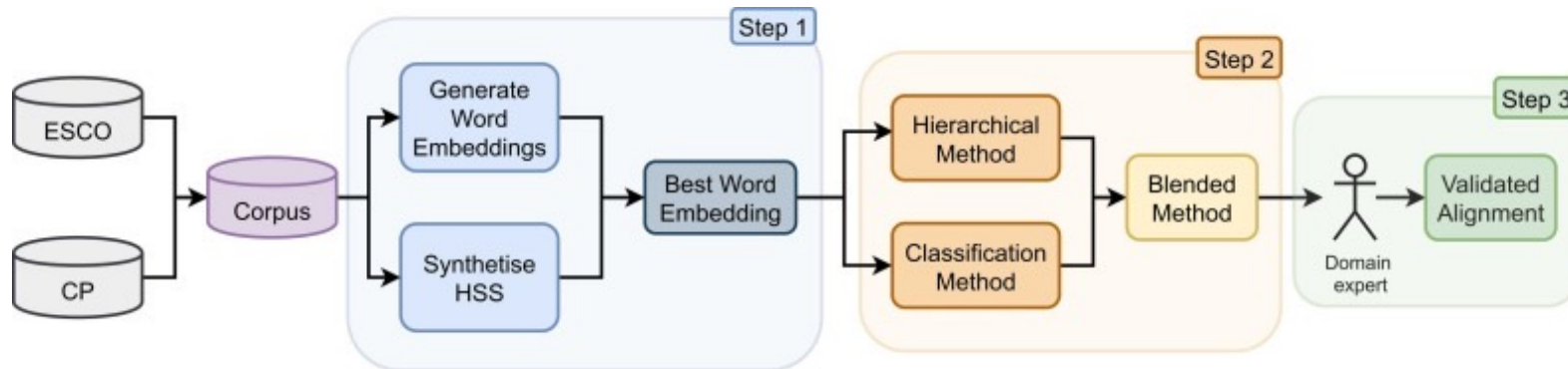
Skill mismatch: data: demand - OJA

Online Job Ads (WIH-OJA, Eurostat and Cedefop)

- Collection of online job ads from 27 European countries + UK and EFTA countries. Since 2018Q4
- Data on occupations and related skills as they emerge from online job postings
- Skills are extracted based on the ESCO Skill Pillar

How to link PIAAC and OJA? Use ESCO and AI

- The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation
- PIAAC questions are processed to tag the most similar ESCO Skills.
- The embedding is trained on OJA UK data and the matching is done on the English language.



Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2022). WETA: Automatic taxonomy alignment via word embeddings. *Computers in Industry*, 138, 103626.

PIAAC2ESCO mapping

PIAAC Question Id	PIAAC Question Description	Label	Group	Type of variable
F_Q02b	instructing training or teaching people individually or in groups?	Teaching people	General	Frequency (time units)
F_Q02d	selling a product or selling a service?	Selling	General	Frequency (time units)
F_Q04a	persuading or influencing people?	Influencing people	General	Frequency (time units)
F_Q05a	[...] «Problem solving» [...]. How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?	Simple problems	Problem solving	Frequency (time units)
G_Q01b	read letters memos or e-mails?	Read letters memos or mails	Literacy	Frequency (time units)
G_Q01g	read bills invoices bank statements or other financial statements?	Read financial statements	Literacy	Frequency (time units)
G_Q01h	read diagrams maps or schematics?	Read diagrams maps or schematics	Literacy	Frequency (time units)
G_Q02a	write letters memos or e-mails?	Write letters memos or mails	Literacy	Frequency (time units)
G_Q03b	calculate prices costs or budgets?	Calculating costs or budgets	Numeracy	Frequency (time units)
G_Q03c	use or calculate fractions decimals or percentages?	Use or calculate fractions or percentages	Numeracy	Frequency (time units)
G_Q03d	use a calculator - either hand-held or computer based?	Use a calculator	Numeracy	Frequency (time units)
G_Q03g	use simple algebra or formulas?	Use simple algebra or formulas	Numeracy	Frequency (time units)
G_Q03h	use more advanced math or statistics such as calculus complex algebra trigonometry or use of regression techniques?	Use advanced math or statistics	Numeracy	Frequency (time units)
G_Q04	you use a computer in your job?	Experience with computer in job	ICT	Yes (1) / No (2)
G_Q05a	use email?	For mail	ICT - Internet	Frequency (time units)
G_Q05d	conduct transactions on the internet for example buying or selling products or services or banking?	Conduct transactions	ICT - Internet	Frequency (time units)
G_Q05e	use spreadsheet software for example Excel?	Spreadsheets	ICT - Computer	Frequency (time units)
G_Q05f	use a word processor for example Word?	Word	ICT - Computer	Frequency (time units)
G_Q05g	use a programming language to program or write computer code?	Programming language	ICT - Computer	Frequency (time units)
I_Q04d	I like learning new things	Like learning new things	Learning strategies	Extents
I_Q04I	I like to figure out how different ideas fit together	Figure out how different ideas fit together	Learning strategies	Extents

PIAAC2ESCO validated dataset

- The validated dataset covers **21 PIAAC questions** and the **mapped ESCO skills, enriched** with alternative labels

F_Q02b: instructing training or teaching people individually or in groups?

- | | |
|--------------------------------------|-------------------------|
| • coach young people | • instruct young people |
| • coach youngsters | • instructing others |
| • educate others | • teach others |
| • educate young people | • teach young people |
| • facilitate young peoples education | • train others |
| • facilitate young peoples mentoring | • train young people |
| • instruct colleagues | • train youngsters |
| • instruct others | • tutoring |

PIAAC2ESCO – Open data

<https://crisp-unimib.github.io/PIAAC2ESCO/>

Open access to:

- Dataset – enriched mapping
- Methodological annex

PIAAC2ESCO - An AI-driven classification of the PIAAC Background questionnaire onto the ESCO Skills Pillar

[View on GitHub](#)

PIAAC2ESCO - An AI-driven classification of the PIAAC Background questionnaire onto the ESCO Skills Pillar

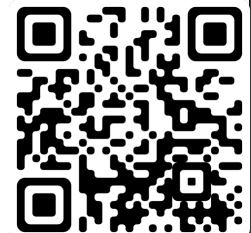
What is PIAAC2ESCO?

PIAAC2ESCO provides a characterisation of the [PIAAC background questionnaire](#) on the base of the [ESCO Skills Pillar](#). In practice it associates a list of ESCO skills (v1) to questions of the PIAAC background questionnaire (version 2010), based on their similarity. We use the section F to I of the PIAAC background questionnaire, from which we select the relevant questions (73 questions out of 84) and all the ESCO skills (13600 items). The validated dataset covers 21 PIAAC questions and the mapped ESCO skills, which are enriched using alternative labels.

How does PIAAC2ESCO work?

The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation. A description of the adopted methodology is available in the [Technical Annex](#).

The training dataset of the embedding is the representative sample of the job ads collected by Eurostat and Cedefop as part of the [Web Intelligence Hub - Online Job Advertisements \(WIH-OJA\)](#)



Publication

Yuchen Guo & Christina Langer & Fabio Mercorio & Francesco Trentini, 2022. "[Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads](#)," [EconPol Forum](#), CESifo, vol. 23(05), pages 11-15, September.

econPOL
FORUM

5
2022
September
Vol. 23

ECONOMIC POLICY AND ITS IMPACT

What Works? Regional Effects of Universities
Nicolas Bunde, Nina Czernich, Oliver Falck and Gabriel Felbermayr

INSTITUTIONS ACROSS THE WORLD

Institutions – Moving to the Global?
Harold James

BIG-DATA-BASED ECONOMIC INSIGHTS

How Do Taxation and Regulation Affect the Real Estate Market?
Mathias Dolls, David Gstrein, Carla Krolage and Florian Neumeier

POLICY DEBATE OF THE HOUR

European Labor Markets: How Can We Effectively Manage Technological and Structural Change?

POLICY DEBATE OF THE HOUR

CONTENT

Yuchen Guo, Christina Langer, Fabio Mercorio and Francesco Trentini

Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads^{*}

Labor markets face challenges due to globalization, structural transformation, and advancing technological change. This can lead to skills gaps and skills mismatch between firms' skill demand and employees' skill supply, which can go in two directions: workers having a skill surplus, where skill supply exceeds demand, or workers experiencing skill shortage, where firms' skill demand is greater than the skills workers actually possess. In light of this, the EU's Agenda for New Skills and Jobs (European Commission 2020) states that creating a more skilled workforce "is a considerable challenge given the rapidly-changing skills needed, and the persistent skills mismatches in the EU labor market." In this report, the Commission also "established the anticipation and matching of labor market and skills needs as a top priority for the EU."

However, empirical evidence on the prevalence of skills mismatch between the skills requested by employers and the skills provided by employees across Europe is scarce. We contribute to the understanding of skills mismatch in the EU by presenting novel evidence on skills gaps across countries, occupations, and skill domains using innovative job ad data and survey data for 16 EU countries and the UK.¹ In particular, we leverage two different data sources: online job vacancy data on skills requested by employers and survey data on skills supplied by workers. We document four key findings: first, skill gaps in the European Union exist, but the extent and direction vary across occupation types: workers in cognitive intensive occupations provide more skills than are demanded (skill surplus), whereas workers in manual intensive occupations face higher skill demand compared to the skills they have (skill shortage). Second, this pattern is consistent across almost all 17 countries that are part of our analysis. This suggests that overall patterns of skills mismatch do not reflect country-specific factors but are rather a European-wide phenomenon. Delving deeper into different skill domains (i.e., digital, numeracy, literacy, and social skills), we document similar skills gaps for different occupation types. Thus, the ob-

KEY MESSAGES

- Linking survey data and online job ads offers new insights into skills gaps in the EU
- Matching labor market needs and skill supply remains a Europe-wide challenge
- Manual workers have skill supply shortages, cognitive workers have skill supply surpluses
- Workers at higher risk of automation experience higher skill shortages, potentially because their job tasks are changing more rapidly
- On-the-job training might be a potential measure to meet future skills needs

LINKING SURVEY DATA AND ONLINE JOB ADS OFFERS NEW INSIGHTS INTO SKILLS GAPS IN THE EU

We propose a novel measure of the gap between the skills demanded by employers and the skills supplied by workers. On the demand side, we rely on online job vacancies (OJV) data from the European Center for the Development of Vocational Training (CEDEFOP), collected in 2015, to capture skills de-

served skills gaps are not driven by a lack of specific skill domains such as digital skills. Finally, we investigate potential mechanisms – i.e., an occupation's automation probability and workers' propensity to participate in on-the-job training – that might underlie the observed patterns of skills mismatch.

Yuchen Guo

is a Ph.D. candidate in Economics at KU Eichstätt-Ingolstadt and guest researcher at the Ifo Institute.

Christina Langer

is a Ph.D. candidate in Economics at KU Eichstätt-Ingolstadt, guest researcher at the Ifo Institute, and Visiting Research Fellow at the Project on Workforce at Harvard University.

^{*} This article is based on a larger research project within the scope of POLAR, which has greatly benefited from contributions by Mario Mazzancora, Filippo Palucchi, and Dennis Roeschke.

¹ These countries are Belgium, Cyprus, Czech Republic, Denmark, France, Germany, Greece, Ireland, Italy, Lithuania, Netherlands, Poland, Slovakia, Slovenia, Spain, Sweden, and the UK.

CESifo ifo INSTITUTE

EconPol Forum 5/2022 September Volume 23 11

13

Skill mismatch across Europe

Descriptive insights on skills mismatch 17 European countries in 2019. Relation with automation and training.

Samples

- PIAAC comprises 250,000 observations (4,000 - 8,000 per country). 2012 and 2014. Projected to 2019 using changes in US (observed in 2014 and 2017) as inflation parameters.
- WIH-OJA includes 17,966,812 observations. 2019.

Skill mismatch - measure

We follow Alabdulkareem et al. (2018) and Giabelli et al. (2022) and calculate the relevance of each skill for each occupation.

Skill frequency

Occupations $\bar{O} = \{O_k, k=1, \dots, m\}$

Skills $\bar{S} = \{S_j, j=1, \dots, p\}$

$$sf(o_k, s_j) = \frac{\sum_{i=1}^n I(o_i = o_k) \cdot I(s_i = s_j)}{\sum_{i=1}^n I(o_i = o_k)}$$

PIAAC use number of respondents - by occupation-question.
From Likert scale to Indicator of use

OJA use number of ads observed - by occupation-skill

Revealed Comparative Advantage

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^p sf(o_i, s_j)}{\sum_{k=1}^m sf(o_k, s_l) / \sum_{k=1}^m \sum_{j=1}^p sf(o_k, s_j)}$$

Alabdulkareem, A., M. R. Frank, L. Sun, B. Al Shebli, C. Hidalgo and I. Rahwan (2018), "Unpacking the Polarization of Workplace Skills", Science Advances 4(7), DOI: [10.1126/sciadv.aao6030](https://doi.org/10.1126/sciadv.aao6030).

Giabelli, A., L. Malandri, F. Mercorio and M. Mezzanzanica (2022), "GraphLMI: A Data Driven System for Exploring Labor Market Information through Graph Databases", Multimedia Tools and Applications 81, 3061–3090.

Skill mismatch - measure

For each skill in demand and supply, the RCA is ranked **among all occupation** and mapped to the percentile of belonging.

Our **mismatch measure** at the occupation level is **the mean RCAs-percentile-rank gap between demand and supply**. Negative values indicate skill surplus, vice versa positive values indicate skill shortage.

An example:

F_Q02b: “[...] teaching people individually or in groups?” → ESCO skill: “Teaching others”

RCA_{oja} percentile rank: 0.95

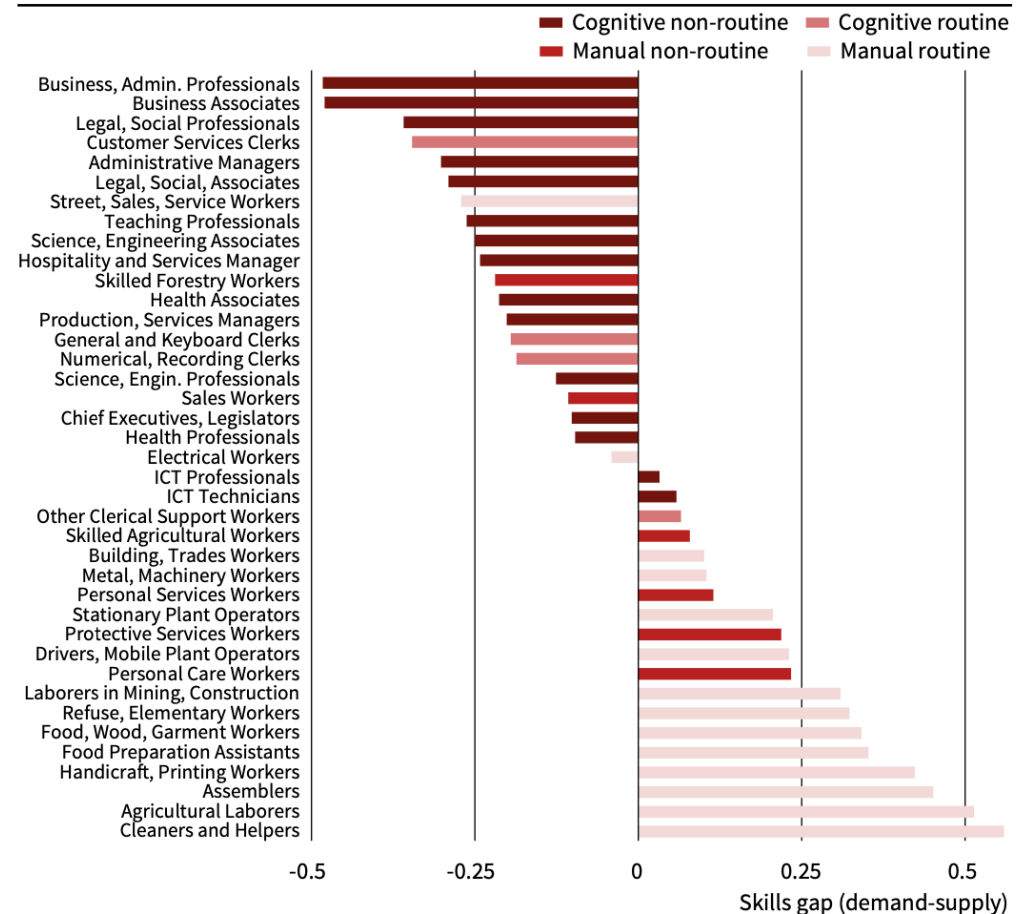
RCA_{piaac} percentile rank : 0.97

Skill mismatch: $pRCA_{oja} - pRCA_{piaac} = -0.02$ **<0 Skill surplus**

Findings

Skill shortage impacts manual workers more than cognitive workers

Average Skill gap by occupation ISCO08 II digit. 2019 pooled 17 countries.



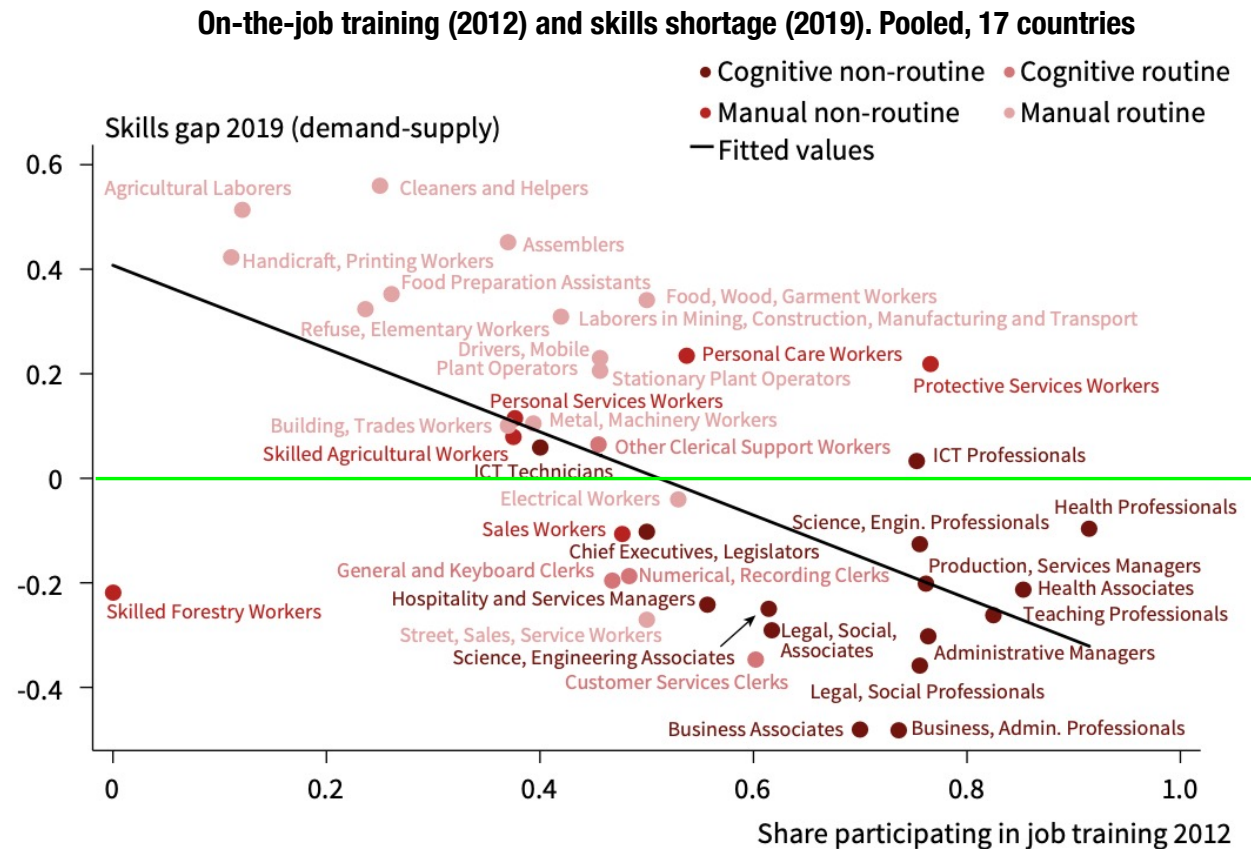
Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. The Quarterly Journal of Economics 118 (4), 1279–1333.

Note: Pooled for 17 European countries.
Source: CEDEFOP; PIAAC.

© ifo Institute

Findings

On aggregate, on-the-job training is negatively related to skill shortage



Note: Correlation between on-the-job training (measured in 2012) and the skills gap (measured in 2019), pooled for 17 European countries.
Source: CEDEFOP; PIAAC.

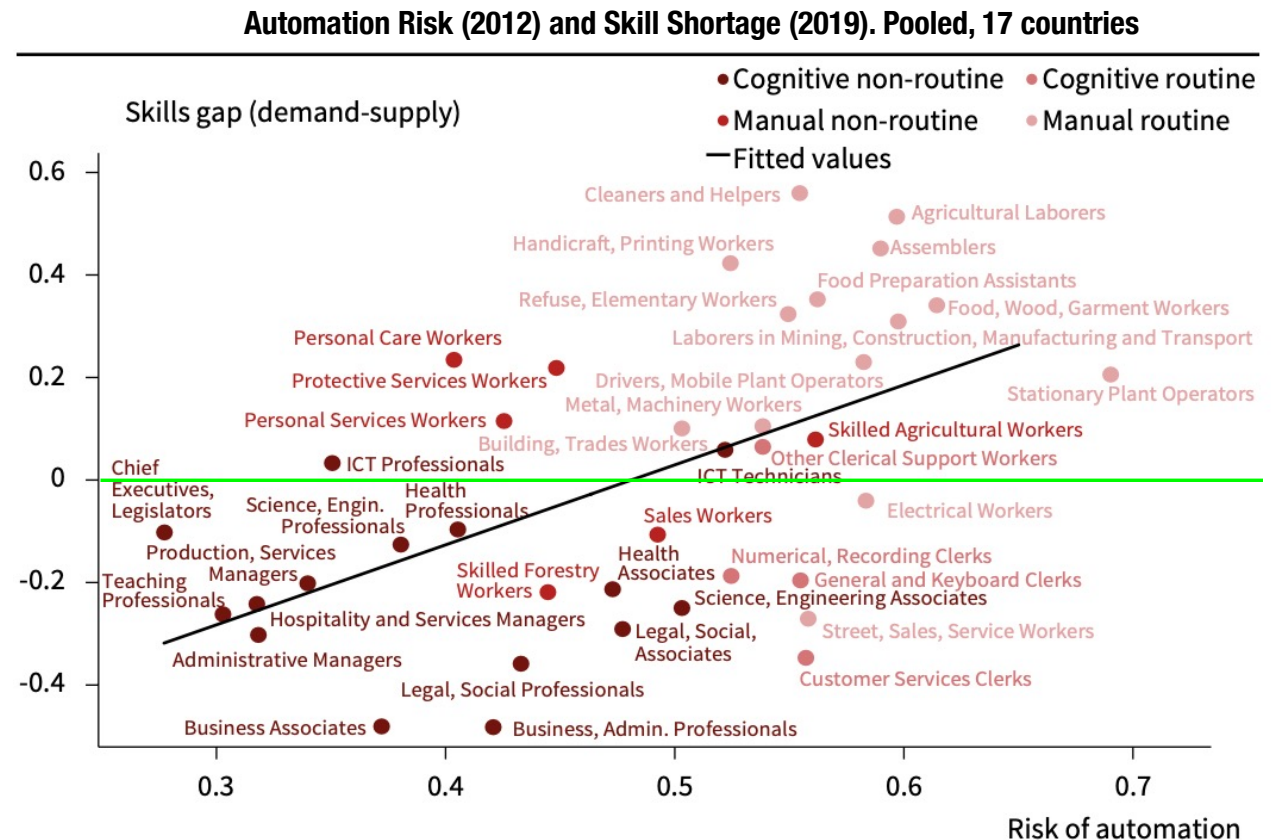
© ifo Institute

Findings

Risk of automation from Nedelkoska and Quintini (2018)

Positive relation between skill shortage and risk of automation.

Nedelkoska, L. and G. Quintini (2018). Automation, Skills Use and Training. Technical report, OECD, Paris.



Note: Correlation between automation risk and the skills gap, pooled for 17 European countries. Our measure of automation risk stems from Nedelkoska and Quintini (2018), who constructed the automation probability for all occupations and countries in our sample using PIAAC data.

Source: CEDEFOP; PIAAC.

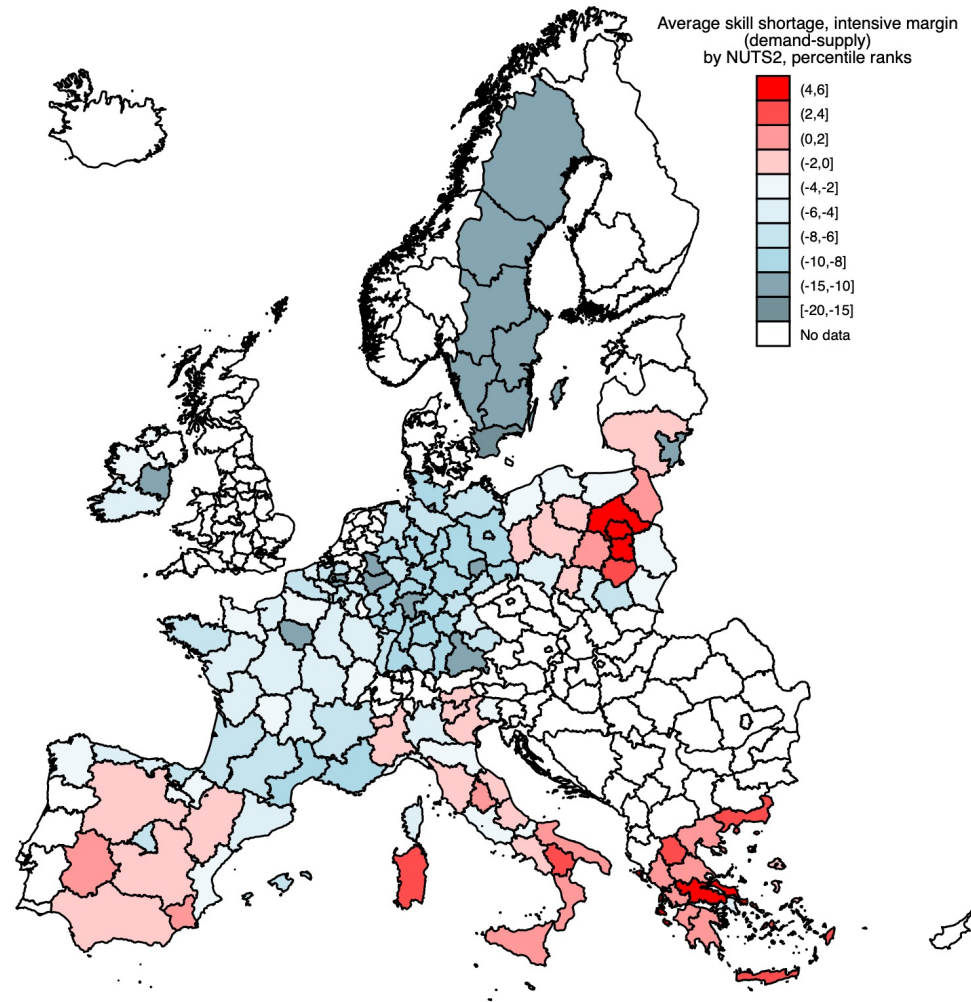
© ifo Institute

Findings

Descriptive figures show high heterogeneity among European regions.

Within-country heterogeneity is also relevant

- Surplus: BE, DE, FR, IE, SE
- Shortage: EL
- Mix: ES, IT, LT, PL



Conclusions

Web data can enrich and complement existing datasets.

AI-methods can simplify complex tasks and support human experts

- **Pros:** cost reduction and bounded risk of non-systematic errors
- **Caveat:** need of transparent design and account of the process

Jobs are changing in terms of tasks and the skill composition of jobs changes. Skill shortages are negatively related to training and positively related to automation risk.

High heterogeneity at the country and regional level calls for more depth in understanding the role of institutional features of the countries labour markets.

Thank you for your attention! Questions?

Francesco Trentini

francesco.trentini@unimib.it