Big Data for Labour Market Intelligence

Capacity development programme 2022

Module 2: dissemination and analysis

Session 8

OJV and labour market statistics: key elements for a methodological approach

Speaker: Mario Mezzanzanica, Emilio Colombo, Francesco Trentini

06/12/2022

Overview

- What do OJA represent?

- How are they related to labour market variables?

- Can they help to predict labour market variables?

- Linking Big data and traditional survey data using AI to study skills mismatch

- An application: PIAAC2ESCO and analysis of skill gaps in Europe

- Conclusions

OJA and representativeness

Are OJA representative of the LM structure?

- We need to define what we observe

- We need to select a benchmark

- We need to compare OJA with the benchmark

ETF
Working together
Learning for life

European Training Foundation

OJA and the population

We observe some (large) number of OJA

Are they representative of the universe of OJA?

We do not know as the true population of OJA is unobservable

We need to estimate it

- Capture recapture models: need ads to be posted in different sites
- Multilevel modelling: individual response is a function of different covariates

OJA finding the right benchmark

Which labour market variable and statistics is closest to the concept of OJA?

1. Job Vacancy Statistics
Advantage: JVS is the closest definition to OJA. Allows for sectoral dimension
Problem: JVS are constructed from a survey of firms which has a relatively small sample size and has not information about occupations nor it allows a regional division. Moreover JVS measure the stock of vacancies OJA the flow

2: LFS
Advantage: by far the most detailed and granular survey on Labour Market in Europe
Problem: LFS is a supply side variable. Measures the stock of employment, (not vacancies)
We can restrict the analysis to job changes <3months but different concept wrt OJA
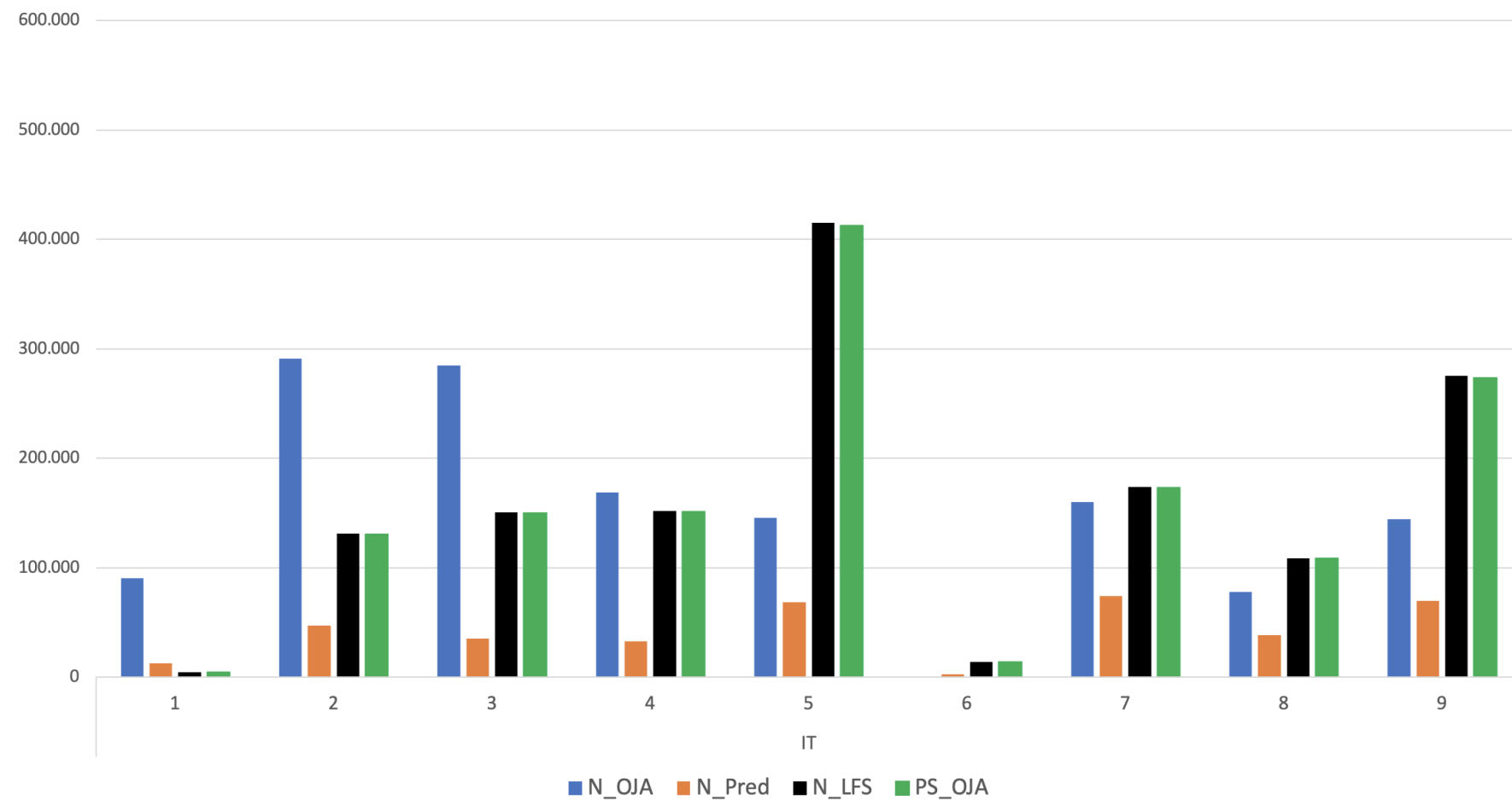
3. Administrative data: very promising not available across countries

ETF
Working together
Learning for life

European Training Foundation

OJA representativeness

If OJA represents a labour market phenomena and they are a biased description of it they can be "corrected" to match the distribution of some benchmark variable
Post-stratification delivers just that, i.e. it reweights the population of OJA with weights obtained from the reference sample.

ETF
Working together
Learning for life

European Training Foundation

OJA representativeness: PS by ISCO 1D

Legend: N_OJA, N_Pred, N_LFS, PS_OJA

Job Postings Source: Lightcast

Do OJA predict LM variables?

Even if OJA are a biased description of some LM phenomena/variable they may be a good predictor of the same phenomena

For a good fit both in sample and out of sample representativeness is not needed, what is needed is predictive ability (fit or MSE)
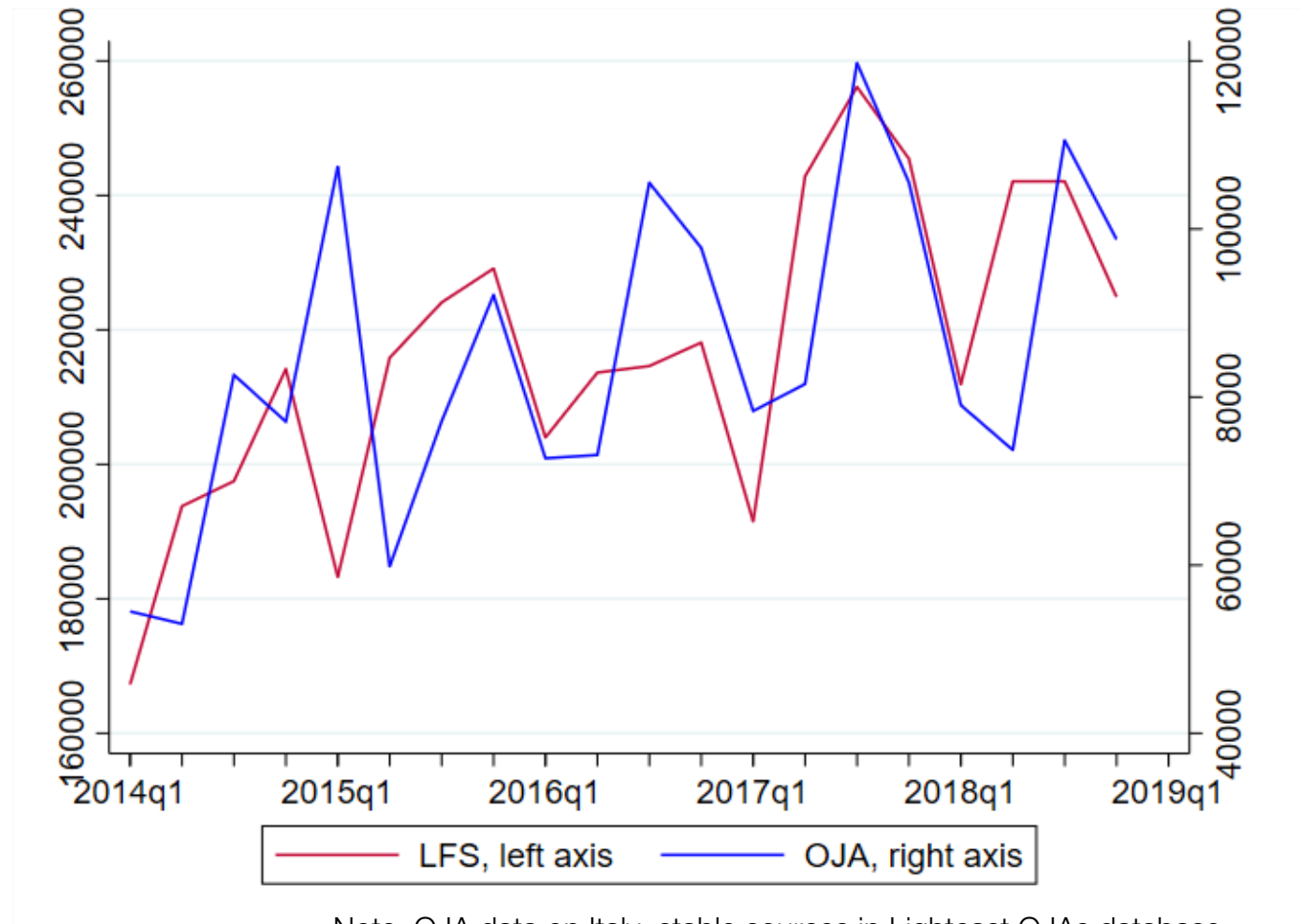
This is true both at cross sectional and at time series analysis

Predictive ability: OJA vs JVS

Note: OJA data on Italy, stable sources in Lightcast OJAs database.

Predictive ability: OJA vs LFS

Note: OJA data on Italy, stable sources in Lightcast OJAs database.

Combined use of official statistics and LMI

Web data can enrich and complement existing datasets.

Enrichment

AI-methods using web data 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

Complementarity

Web data can provide detailed information about phenomena otherwise not observed

- **Pros**: uniqueness (frequency, detail)
- **Caveat:** representativeness, find robust measures

ETF Working together Learning for life

European Training Foundation

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.

ETF Working together Learning for life

European Training Foundation

PIAAC background questionnaire

F_START Skills used at work

Layout	Item group table
	How often ^DoesDid your ^JobLastjob usually involve …
F_Q02b (JRA) (B)	**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

ETF
Working together
Learning for life

European Training Foundation

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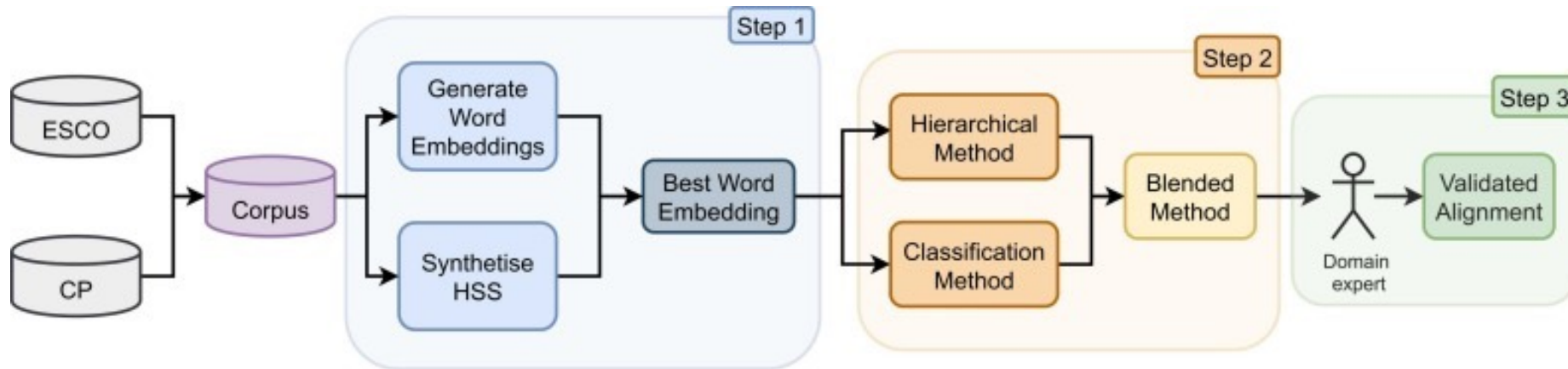
Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2022). WETA: Automatic taxonomy alignment via word embeddings. Computers in Industry, 138, 103626.

PIAAC2ESCO 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
- coach youngsters
- educate others
- educate young people
- facilitate young peoples education
- facilitate young peoples mentoring
- instruct colleagues
- instruct others

- instruct young people
- instructing others
- teach others
- teach young people
- train others
- train young people
- train youngsters
- tutoring

ETF
Working together
Learning for life

European Training Foundation

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_Q04l	I like to figure out how different ideas fit together	Figure out how different ideas fit together	Learning strategies	Extents

ETF
Working together
Learning for life

European Training Foundation

PIAAC2ESCO – Open data

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

Open access to:
- Dataset – enriched mapping
- Methodological annex

Skill mismatch across Europe

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

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.

Data

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

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 in 2019.

ETF
Working together
Learning for life

European Training Foundation

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 over-skilling, vice versa positive values indicate under-skilling.

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 gap: pRCA$_{oja}$ – pRCA$_{piaac}$ = -0.02   <0 Over-skilling

---
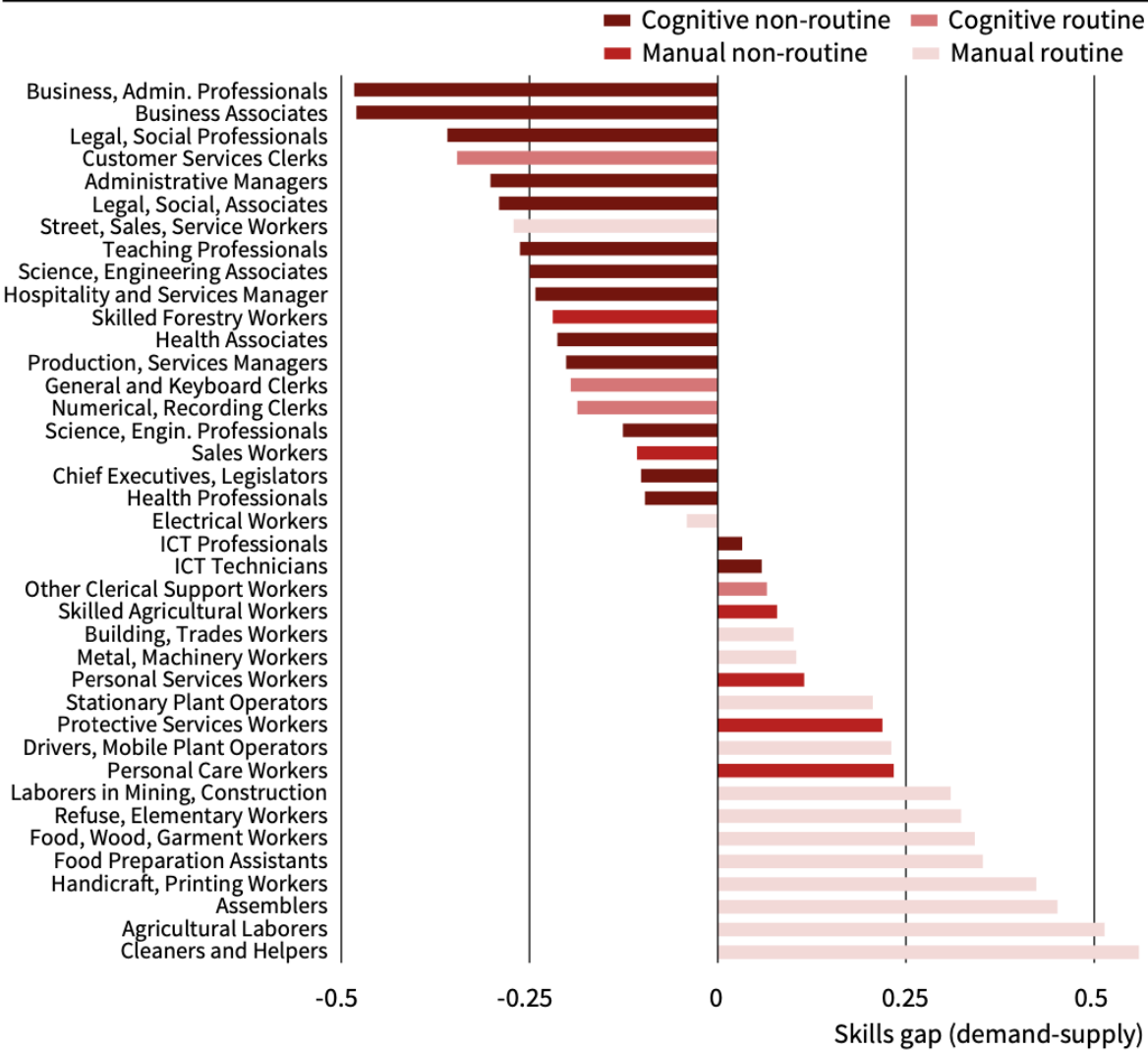
ETF
Working together
Learning for life

European Training Foundation

Findings

Skill gaps indicates under-skilling is more pervasive among manual workers than cognitive workers.

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.

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

Legend:
- Cognitive non-routine
- Cognitive routine
- Manual non-routine
- Manual routine

Occupations (top to bottom):
Business, Admin. Professionals; Business Associates; Legal, Social Professionals; Customer Services Clerks; Administrative Managers; Legal, Social, Associates; Street, Sales, Service Workers; Teaching Professionals; Science, Engineering Associates; Hospitality and Services Manager; Skilled Forestry Workers; Health Associates; Production, Services Managers; General and Keyboard Clerks; Numerical, Recording Clerks; Science, Engin. Professionals; Sales Workers; Chief Executives, Legislators; Health Professionals; Electrical Workers; ICT Professionals; ICT Technicians; Other Clerical Support Workers; Skilled Agricultural Workers; Building, Trades Workers; Metal, Machinery Workers; Personal Services Workers; Stationary Plant Operators; Protective Services Workers; Drivers, Mobile Plant Operators; Personal Care Workers; Laborers in Mining, Construction; Refuse, Elementary Workers; Food, Wood, Garment Workers; Food Preparation Assistants; Handicraft, Printing Workers; Assemblers; Agricultural Laborers; Cleaners and Helpers

X-axis: -0.5, -0.25, 0, 0.25, 0.5 — Skills gap (demand-supply)

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

© ifo Institute

Working together Learning for life

European Training Foundation

Automation changes job content

Risk of automation from Nedelkoska and Quintini (2018)

Positive relation between under-skilling and risk of automation.

Jobs are changing in terms of tasks and the skill composition of jobs changes.

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

Automation Risk (2012) and Skill gaps (2019). Pooled, 17 countries
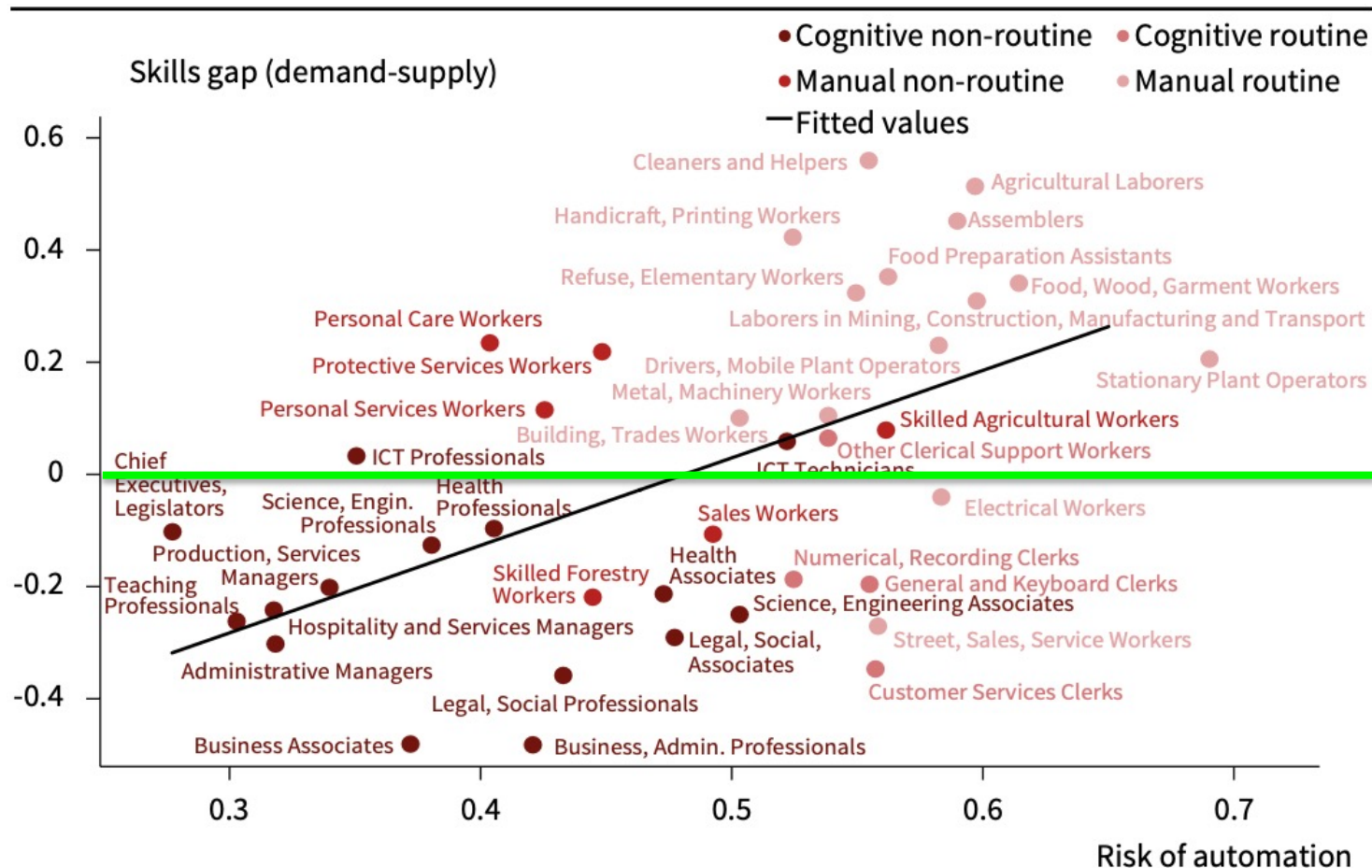
Note: Correlation between automation risk and the skills gap, pooled for 17 European countries. Our measure of automation risk stems from Nedelkoska and Quintini (2018), who constructed the automation probability for all occupations and countries in our sample using PIAAC data.
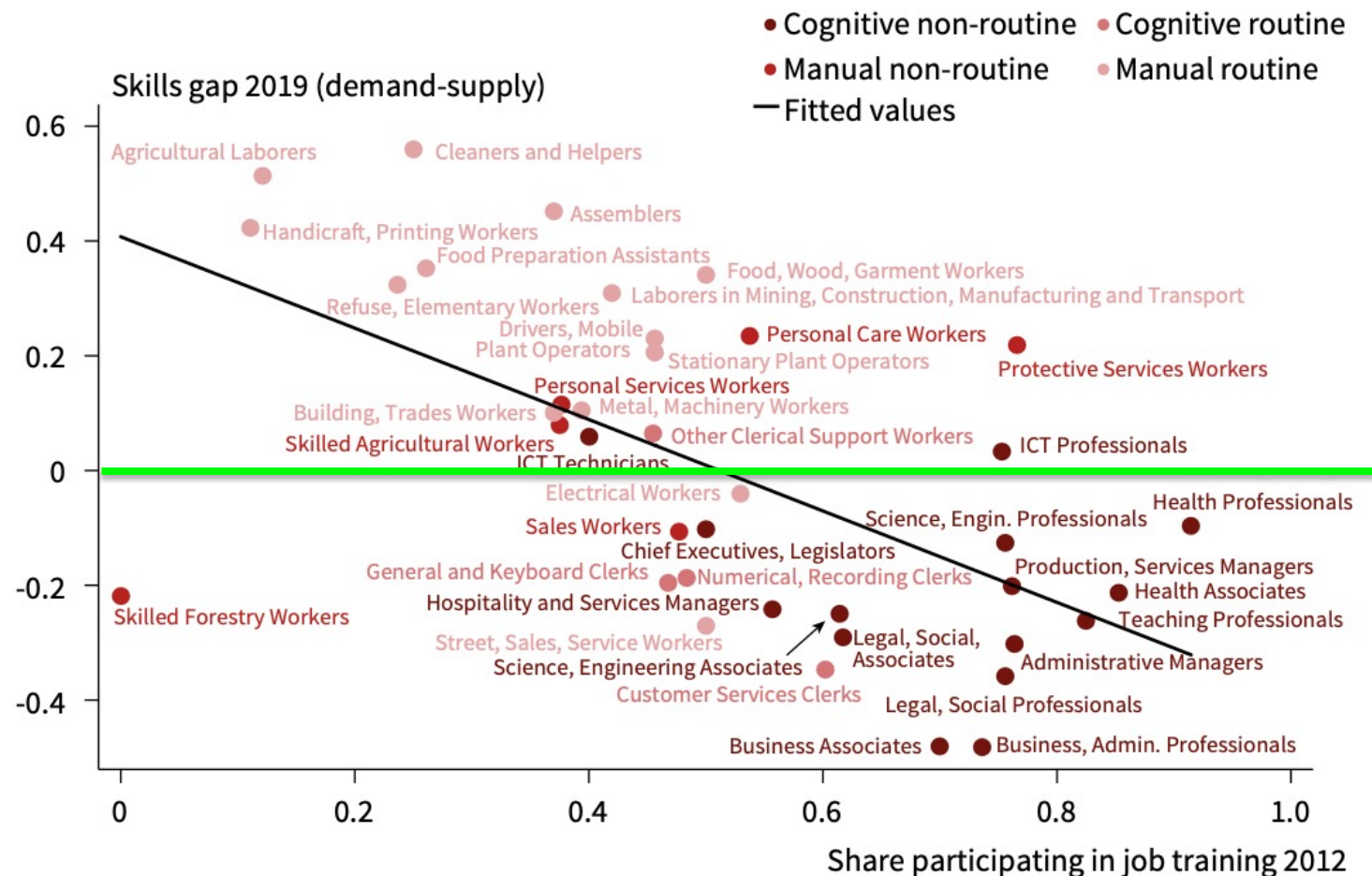Source: CEDEFOP; PIAAC.

© ifo Institute

together for life

European Training Foundation

Training

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

Occupation with high levels of under-skilling also present low participation in on-the-job training.

On-the-job training (2012) and skills gaps (2019). Pooled, 17 countries

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

European Training Foundation

Findings

Descriptive figures show high heterogeneity among European regions.

Within-country heterogeneity is also relevant
- Over-skilled (on average): BE, DE, FR, IE, SE
- Mix: EL, ES, IT, LT, PL

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

Average skill shortage, intensive margin (demand-supply) by NUTS2, percentile ranks

- (4,6]
- (2,4]
- (0,2]
- (-2,0]
- (-4,-2]
- (-6,-4]
- (-8,-6]
- (-10,-8]
- (-15,-10]
- (-20,-15]
- No data

:TF Working together Learning for life

European Training Foundation