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
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
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.
OJA representativeness: PS by ISCO 1D

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

- 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)
  

PIAAC background questionnaire

F_START Skills used at work

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

13
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.

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
## PIAAC2ESCO mapping

<table>
<thead>
<tr>
<th>PIAAC Question Id</th>
<th>PIAAC Question Description</th>
<th>Label</th>
<th>Group</th>
<th>Type of variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_Q02b</td>
<td>instructing training or teaching people individually or in groups?</td>
<td>Teaching people</td>
<td>General</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>F_Q02d</td>
<td>selling a product or selling a service?</td>
<td>Selling</td>
<td>General</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>F_Q04a</td>
<td>persuading or influencing people?</td>
<td>Influencing people</td>
<td>General</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>F_Q05a</td>
<td>[...] «Problem solving» [...] How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?</td>
<td>Simple problems</td>
<td>Problem solving</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q01b</td>
<td>read letters memos or e-mails?</td>
<td>Read letters memos or mails</td>
<td>Literacy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q01g</td>
<td>read bills invoices bank statements or other financial statements?</td>
<td>Read financial statements</td>
<td>Literacy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q01h</td>
<td>read diagrams maps or schematics?</td>
<td>Read diagrams maps or schematics</td>
<td>Literacy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q02a</td>
<td>write letters memos or e-mails?</td>
<td>Write letters memos or mails</td>
<td>Literacy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q03b</td>
<td>calculate prices costs or budgets?</td>
<td>Calculating costs or budgets</td>
<td>Numeracy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q03c</td>
<td>use or calculate fractions decimals or percentages?</td>
<td>Use or calculate fractions or percentages</td>
<td>Numeracy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q03d</td>
<td>use a calculator - either hand-held or computer based?</td>
<td>Use a calculator</td>
<td>Numeracy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q03g</td>
<td>use simple algebra or formulas?</td>
<td>Use simple algebra or formulas</td>
<td>Numeracy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q03h</td>
<td>use more advanced math or statistics such as calculus complex algebra trigonometry or use of regression techniques?</td>
<td>Use advanced math or statistics</td>
<td>Numeracy</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q04</td>
<td>you use a computer in your job?</td>
<td>Experience with computer in job</td>
<td>ICT</td>
<td>Yes (1) / No (2)</td>
</tr>
<tr>
<td>G_Q05a</td>
<td>use email?</td>
<td>For mail</td>
<td>ICT - Internet</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q05d</td>
<td>conduct transactions on the internet for example buying or selling products or services or banking?</td>
<td>Conduct transactions</td>
<td>ICT - Internet</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q05e</td>
<td>use spreadsheet software for example Excel?</td>
<td>Spreadsheets</td>
<td>ICT - Computer</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q05f</td>
<td>use a word processor for example Word?</td>
<td>Word</td>
<td>ICT - Computer</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>G_Q05g</td>
<td>use a programming language to program or write computer code?</td>
<td>Programming language</td>
<td>ICT - Computer</td>
<td>Frequency (time units)</td>
</tr>
<tr>
<td>I_Q04d</td>
<td>I like learning new things</td>
<td>Like learning new things</td>
<td>Learning strategies Extents</td>
<td></td>
</tr>
<tr>
<td>I_Q04f</td>
<td>I like to figure out how different ideas fit together</td>
<td>Figure out how different ideas fit together</td>
<td>Learning strategies Extents</td>
<td></td>
</tr>
</tbody>
</table>
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.

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.
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”

<table>
<thead>
<tr>
<th>RCA\textsubscript{oja} percentile rank: 0.95</th>
<th>Skill gap: pRCA\textsubscript{oja} – pRCA\textsubscript{piaac} = -0.02</th>
<th>&lt;0 Over-skilling</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCA\textsubscript{piaac} percentile rank: 0.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Findings

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

Automated 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.

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.
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.