Using online vacancy and applicants’ data to study skills dynamics

Fidel Bennett, Verónica Escudero, Hannah Liepmann & Ana Podjanin

Date: 3 / February / 2022
Motivation

- Rich literature on skills dynamics in Europe and the US thanks to rich data (O*NET)
  - Computer technology replaces routine work (e.g., Autor et al. 2003; Goos et al. 2014)
  - AI and robots are predicted to replace also non-routine analytical labour (e.g., Frey and Osborne, 2017)
  - Yet, interactive skills have been found to be less replicable (Arntz et al., 2016)

- Less knowledge outside of high-income countries, despite importance of the topic
  - Imputation of US data requires strong assumptions (e.g., Almeida et al. 2017; Bhorat et al. 2020)
  - Survey data exist as single cross-sections (e.g., Lewandowski et al. 2019, 2020)

- Task approach using online labour platform data is an alternative
  - Growing literature in the US using vacancy data (e.g., Deming and Kahn, 2018; Hershbein and Kahn, 2018)
  - How skills change within occupations & overtime (granularity and panel structure of the data)
Our study

- Explores whether skills dynamics can be studied using online vacancies and job applicants’ data outside Europe and the US
- Develops a taxonomy based on the social sciences literature applicable to developing country contexts
- Implements the taxonomy using data from the Uruguayan portal *BuscoJobs*
- Creates ISCO-08 occupations at the 2-digit level to assess how skills change within occupations
**Taxonomy**

- The aim is for the taxonomy to be comprehensive and succinct, suitable for emerging and developing economies, and adapted to online vacancy and applicants’ data.

- Based on social science literature (in particular, labour economics & psychology)
  - Skill-biased technological change literature (e.g., Acemoglu and Autor, 2011)
  - Psychology literature (e.g., Almlund et al. 2011).

- 3 broad categories – cognitive, socioemotional, manual skills – & 14 sub-categories

- Sub-categories are defined based on keywords and expressions, also called “unique skills”.
  - When a keyword appears in the data, we assume that the respective sub-category is relevant for the vacancy or the applicant job spell.
BuscoJobs data

- Private portal in Uruguay, supported by the Ministry of Labor and Social Security

- For 2010-2020, rich information on (i) job vacancies (N=87,000), (ii) applicants’ work biographies (N=1.23 million), (iii) search behavior

- Data are not fully representative: e.g., applicants tend to be younger, more educated, and more likely to live in the capital, compared to labor force

- Might require weighting and/or focus on selected labour force segments
Occupations: BuscoJobs applicants vs average worker in Uruguay

(a) Shares in 2020

(b) Yet, no big bias across time, pp difference

Advancing social justice, promoting decent work
Machine learning and text mining approach

Methodological implementation of the skills taxonomy into the data
- Using open text descriptions of vacancies and applicants’ job spells in the data AND the keywords and expressions (i.e., unique skills) that characterize each of the 14 sub-categories
- NLP approach to store keywords and their synonyms to define and create skills sub-categories in the data
- Coded proxy variables of skill intensity (i.e., number of times a unique skill appears)
- We classify over 73% of applicants’ observations and 91% of vacancies.

Creation of the 2-digit ISCO-08 occupations for both the vacancy and job applicants’ data
- Allow assessing how skills change within occupations
- Using similar NLP approach, plus a machine learning predictive model to improve precision
Skills sub-categories: keywords/expressions plus synonyms

(a) Applicants’ job spells

Advancing social justice, promoting decent work

(b) Vacancies
Wrapping up

- Online data on vacancies and applications to a job platform are a suitable source for studying skills dynamics outside of high-income countries.

- Our approach relies on data that is currently available in many countries across the world.
  - It allows country-specific analysis that does not need to assume that occupational skills are the same across countries.

- Another distinguishable feature is the ability to study detailed skills dynamics across time, representative of both labour demand and supply.
  - Panel nature and granularity of the data.

- Issues of representativeness need to be taken into account.

- To the best of our knowledge, we are the first to explore this approach in the context of an emerging economy.
Thank you.
Verónica Escudero: escudero@ilo.org
Hannah Liepmann: liepmann@ilo.org