Economic Modelling 115 (2022) 105963

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Skill requirements and labour polarisation: An association analysis based on Polish online job offers

Carlos Usabiaga [a,*], Fernando Núñez [b], Lukasz Arendt [c], Ewa Gałecka-Burdziak [d], Robert Pater [e]

[a] *Pablo de Olavide University, Seville, Spain*
[b] *University of Seville, Spain*
[c] *University of Lodz, Poland*
[d] *Warsaw School of Economics, Poland*
[e] *University of Information Technology and Management, Rzeszow, Poland*

ARTICLE INFO

JEL classification:
C38
J01
J24
J23

Keywords:
Contingency table
Cluster analysis
Labour polarisation
Online job offers
Polish labour market

ABSTRACT

This paper uses the methodological scheme of contingency tables to explore polarisation in the Polish labour market. We use a large database of online job offers published on selected Polish job portals in the period 2017–2019, whereas most of the studies on the polarisation hypothesis are based on employment data. The main advantage of our microdata is the use of information on the required skills of the vacancy. The contingency table allows us to generate clusters of vacancies whose attributes tend to appear jointly. The study reveals that office skills do not offer a particular advantage in an automated labour market, while information and computer technology skills and communication skills seem to have a shield effect in such an environment. In addition, a cluster of transversal skills –self-organisational, technical and interpersonal skills– constitutes an important requirement for most job offers. These skills should be widely developed within the educational system, at different levels.

1. Introduction

The dynamic development and implementation of new technologies, such as Information and Communication Technologies (ICT), Internet of Things, Big Data analytics, and Artificial Intelligence, have significantly influenced all spheres of social and economic life. These technologies also have a pivotal impact on the labour market. On the one hand, there are growing concerns about the "end of work" scenario in which technology replaces humans in jobs, leading to massive technological unemployment. Such fears have been fuelled, for example, by Frey and Osborne (2017), who estimated that 47% of the U.S. labour force may face a high risk of being replaced by technology. On the other hand, we can find some studies which criticise the conclusions of this last study and reveal less alarming results for OECD, European Union or G20 countries (Arntz et al., 2016; West, 2018). Likewise, there are empirical (Van Roy et al., 2018) and theoretical studies which posit that automation and employment may grow hand in hand, provided that new technologies induce the creation of new tasks in which humans have advantage over machines (Acemoglu and Restrepo, 2019, 2020). Still, this debate is an important element of mainstream literature on technical change in the labour market. Importantly, both streams of literature use a methodological framework based on tasks to analyse the impact of technical change on the labour market, where the most popular approach is a model proposed by Autor et al. (2003) –named after the authors the ALM model– which classifies the tasks performed by individuals into five task-content groups: non-routine analytical, interpersonal and manual tasks, and routine cognitive and manual tasks. Since there is still considerable ambiguity in the exact measurement of the newest technologies (such as automation or Artificial Intelligence) and their impact on the labour market, we consider that this impact can also be analysed by studying their effects on task-content groups, which may be done in line with the hypothesis of labour market polarisation, according to which the automation process in the production of goods and services leads to a growth in demand for both high-skilled and low-skilled workers, while hollowing out the middle of the skill distribution. The evident effect of labour polarisation is a shift in the employment structure among task-content groups, even if total employment remains constant (Autor et al., 2003; Autor and Dorn, 2013). This shift influences the skill-mix required on the labour market (from employees, but also from job candidates), thus influencing the

* Corresponding author. Pablo de Olavide University, Department of Economics, Carretera de Utrera, Km. 1, 41013, Seville, Spain.
E-mail address: cusaiba@upo.es (C. Usabiaga).

https://doi.org/10.1016/j.econmod.2022.105963
Received 3 November 2021; Received in revised form 7 April 2022; Accepted 10 July 2022
Available online 4 August 2022
0264-9993/

employability perspective of individuals. The fact that polarisation of the labour market in Poland differs from patterns revealed in many developed countries, as we will comment later, and that the Polish pattern may converge to these patterns in the future (Gajdos et al., 2020), make the knowledge about skills requirements across task-content groups important in order to minimise the scale of mismatch during this technological transition. This knowledge can help educational institutions to adjust the curricula to develop skills for which there will be a growing demand (as a result of deepening polarisation), and can help individuals to rationalise the acquisition of their skill-mix. Moreover, as polarisation patterns may be co-driven by other characteristics –such as industry structure or regional discrepancies–, it is reasonable to study these drivers of polarisation within one research framework.

The main goal of our study is to identify the skills most demanded in the labour market and put them in relation to the polarisation phenomenon, using online job offers from Poland. Thus, we contribute to the literature on the labour market polarisation hypothesis (Acemoglu and Autor, 2011; Autor et al., 2003; Goos et al., 2014), and skills requirements based on online job postings analysis (e.g. Marinescu and Wolthoff, 2020; Deming and Kahn, 2018; Hershbein and Kahn, 2018; Modestino et al., 2020). We use the occupation group as a typical proxy of job-related skills, but we also perform the analysis from a skills perspective. We believe that such an approach substantially complements the occupation-based technique, especially in the field of polarisation theory, where the routinisation or task-based approach is described by "what a person knows and can perform" (skills) rather than by "the occupation a person represents".

Our research aims to answer the following research questions:

1) What are the connections between the vacancy attributes (tasks, regions, occupations, industries, and skills)? Do vacancy clusters/biclusters arise from such connections?
2) Are there differences between the labour demand clusters/biclusters in the way they demand skills?
3) Which skills differentiate routine task workers from non-routine task workers?
4) Do those differences depend on whether the tasks are cognitive or manual?

Answering these research questions can help understand the phenomena of labour polarisation from the perspective of the skills and tasks requirements specified in the jobs offered in the Polish economy. Workplace automation affects, to a greater extent, those jobs dominated by routine manual tasks. However, from our results it follows that the skill composition of the vacancy is also important in this matter. Thus, those online job offers requiring only basic skills (such as Office and Availability skills) have a higher risk of being replaced by a machine than those that require a strong set of skills, whether they are transversals or job-related skills.

The available studies of the polarisation hypothesis usually focus only on the realised demand for labour, and use employment data to track polarisation patterns or the scale of the mismatch. However, to better understand the trends in the aforementioned changes, the analysis of the vacancies available in the labour market or their proxy, the job offers, can add new information to existing studies. This gives the advantage of observing the yet unmet demand, and the difficulties in the matching process presently occurring. A rich dataset of online job offers also gives the possibility of observing companies' demand for skills. This information deepens the common analysis of demand for occupations by adding the skill dimension. Finally, vacancies show prospects in labour demand and skill demand changes, as these changes were not yet accounted for in employment statistics.

We use online job offers collected with the use of the System of Online Job Offers (SOJO) for 2017–2019 in Poland. The system collects data on job offers posted on selected Polish job portals. Based on collected texts of job offers, a machine learning algorithm provides information about: (1) the required occupation (according to the International Labour Organisation ISCO-08 classification), (2) the skills (with reference to the classes of skills in the Balance of Human Capital (BHC) study) –see Appendix for the taxonomy of used skills–, (3) the activity sector (NACE industry), and (4) the region (NUTS) in which the employer operates. Having such categories linked to each job offer, we are able to characterise the online job offers by measuring the degree of association that exists between the different categories of the categorical variables used to describe them. To address this issue, those variables are cross-classified in a contingency table (CT) where rows and columns can consist of a single variable (that is, its categories) or a combination of several variables. In this latter case, we would have a new variable, a cross-variable, whose cross-categories would be given by all the possible combinations of the categories of the single variables used to create it.

The use of the CT approach may be appropriate when the information available on the object of study is fundamentally categorical (Mosteller, 1968; Agresti, 2013). A multiway CT shows the frequency of each cell corresponding to the combination of the different categories of several categorical variables; this is the starting point from which to carry out association and cluster analyses that are at the centre of the multivariate categorical analysis. The history of CTs extends back at least into the 19th century (Fienberg and Rinaldo, 2007), remaining today as an active area of research. CTs are widely used in many fields: economics, sociology, marketing, finance, biology, medicine, etc. In the economic field, two-sided matching markets are likely to be the economic topic most closely related to CTs[1] –for example, there are applications on marriage markets (Chiappori, 2020), labour matching (Álvarez de Toledo et al., 2018, 2020), college admissions, auctions (Chiappori and Salanié, 2016), allocation of loans from banks to entrepreneurs (Den Haan et al., 2003), medical residency market (Agarwal, 2015) or kidney exchanges (Roth et al., 2004). CTs are also present in other economic approaches (not based on two-sided matching), such as occupational mobility (Long and Ferrie, 2013), occupational choices (Hellerstein and Morrill, 2011), regional specialisation and concentration (Haedo and Mouchart, 2018), stock market returns (Rey et al., 2014), etc. In this article, we analyse the match or association between categorical attributes for only one of the sides of the market; the online job vacancy market.

At the methodological level, one of the most discussed issues concerns how to deal with sparse CTs. The dimensions of the CTs, measured by the number of cells, grow exponentially with the number of categorical variables and the number of different categories for each variable, giving rise to a problem of sparsity in the CT (Petitjean et al., 2013). In order to avoid the problems associated with sparsity, smoothing methods are appropriate for the analysis of large sparse CTs (Titterington and Bowman, 1985; Simonoff, 1996; Burman, 2004; among many others). We will address this problem and the smoothing solution later on.

Our working economic hypothesis in this study is that there are certain attributes of the online vacancies which tend to be strongly associated (they tend to appear together). When we refer to association, we do not mean that a particular combination of attributes has a relatively high frequency in the CT, but rather that frequency is higher than would be expected if the generation of attribute combinations were random. In other words, a certain combination of attributes may have a

[1] Observed matches in two-sided matching markets constitute a contingency table, although they may appear under other names in the literature (observed matches, match output, matching, outcomes, etc).

relatively small frequency in the CT but show instead a strong propensity to associate.

Besides the sparsity issue, high-dimension CTs present an additional problem: their high segmentation makes it difficult to find and describe association patterns in the object of study. One solution to this problem is to apply a cluster analysis on both sides of the table so that we get a panoramic view of how certain groups/clusters of row-categories tend to be associated with groups/clusters of column-categories and vice versa –the analysis is elevated to homogeneous groups of categories, rather than to individual categories. Cluster associations in both directions give rise to the formation of biclusters or local labour markets that can be explored for idiosyncratic features (in our case in terms of polarisation). Thus, we will be able to analyse how yet unmet labour demand is clustered around occupational groups, industries, regions and skill requirements. Since our analysis is rooted in the hypothesis of labour market polarisation, our approach enables clustering by task-content groups, in line with the ALM model.

The rest of the paper continues as follows. Section 2 provides a research background in a form of synthetic literature review. Section 3 presents the description of the methodology with regard to the CT and data sources. In Section 4 results are described, while Section 5 presents the discussion of those results. Finally, Section 6 is the conclusion.

2. Economic background of the research

Labour market polarisation, as well as many of the recent studies analysing the impact of technology on the labour market, is rooted in the ALM model proposed by Autor et al. (2003). The ALM model distinguishes five types of tasks (task-content groups), which we use also in our study: non-routine analytical (e.g. forming and testing hypotheses, legal writing), non-routine interpersonal (e.g. persuading and selling), non-routine manual (e.g. truck driving, janitorial services), routine cognitive (e.g. record-keeping, simple calculations) and routine manual tasks (e.g. picking and sorting, repetitive assembly). The demand for workers performing these tasks, according to the polarisation hypothesis,[2] depends on the automation process, which leads to demand growth (and wage growth) for both high- and low-skilled workers performing non-routine tasks, while hollowing out the middle of the skill distribution (this last fact applies mainly to clerical and assembly line jobs). Thus, in countries where polarisation occurs, we shall see a clustering process of labour demand around non-routine tasks and, therefore, around the skills required to perform these tasks.

The usual approach to test labour market polarisation is based on employment data (not job offers). Studies which follow such an approach reveal a polarisation pattern in several countries: the U.S. (see Autor et al., 2003; Autor and Dorn, 2013; Cortes et al., 2017; Mallick and Sousa, 2017),[3] China (Wang et al., 2021), Great Britain (Goos and Manning, 2007), Germany (Dustmann et al., 2009), Nordic countries (Asplund et al., 2011; Adermon and Gustavsson, 2015), Western EU countries (Goos et al., 2009, 2014), and several OECD countries (Michels et al., 2014). Processes of labour market polarisation have also been revealed in Canada (Green and Sand, 2015) and Portugal (Fonseca et al., 2018), however with some specific features diverging from the canonical ALM model.

There are only a few studies which analyse polarisation based on vacancy data. Hershbein and Kahn (2018) analyse online job offers in 2007 and the period 2010–2015 in the U.S. labour market. These authors show that routine-manual occupations did not face upskilling during the Great Recession, while routine-cognitive occupations did. Zilian et al. (2021) use data on vacancies provided by the Austrian public employment services and link the four-digit ISCO-08 occupational groups with the skills listed in the *European Skills, Competences, Qualifications and Occupations* (ESCO) classification. They reveal that the relative importance of medium- and high-skills increased between 2007 and 2017, as well as the number of skills demanded in the case of vacancies requiring higher skill levels; there was no evidence towards polarisation, but rather upskilling.

At the same time data on online job postings (vacancies) has become a popular source for labour market analysis, including demand for skills. These studies cover a wide range of topics: skills impact on wage for the U.S. (Deming and Kahn, 2018; Hershbein and Kahn, 2018), skills demand in selected industries for the U.S. energy sector (Lyu and Liu, 2021), demand for specific skills, as for example IT and AI-related skills (Lovaglio et al., 2018; Acemoglu et al., 2021; Samek et al., 2021; Squicciarini and Nachtigal, 2021), etc. Also the way in which skills are measured differs significantly; we move from using education level and years of experience as a proxy of skills (Modestino et al., 2020) to incorporating ESCO classification (Zilian et al., 2021).

Looking at Poland, it is noted that studies of labour market polarisation have so far been based solely on employment data. The results of these studies report an atypical pattern of polarisation; i.e., an employment increase (instead of a decline) in the middle of the skill distribution (Gajdos et al., 2020; Hardy et al., 2018) and a stronger downward pressure on wages in the case of occupations with larger degrees of routine content (Parteka, 2018), which, to a certain extent, contrasts with the existence of relatively high wage premiums in routine manual jobs compared to all non-routine task-content jobs (Arendt and Grabowski, 2019). At the same time, historically, the Polish labour market has presented a relatively high share of routine jobs (in 2004, they constituted 9.7% of employment in the case of routine cognitive jobs, and 62.2% in the case of routine manual jobs).

Technical change in the labour market, and thus labour polarisation, are not static phenomena. Autor (2015) even predicts that labour polarisation is unlikely to continue very far into the future, as even routine jobs include many non-routine tasks that human beings still perform better (more productively) than a robot/technology. This line of argument has been presented recently by Acemoglu and Restrepo (2019, 2020), who proposed a theoretical model in which automation and development of AI do not lead to “the end of work”. In the short term, the pace of change in the task-content structure of employment in Poland (taking 2010 as the year of reference) has been rather modest,[4] although this pace might be altered by the COVID-19 pandemic, which has enhanced the dynamics of implementing ICT and AI solutions and raised the discussion about creating even more accelerated automation processes across economies. These dynamic forces seem to be of great importance in Poland, a country which has been lagging behind in terms of technology uptake; for instance, it seems clear that technical change and polarisation processes shall alter the structure of jobs by shifting the demand towards non-routine jobs in the very near future.

Our study is focused on the demand side of the labour market, since we analyse job postings. Therefore, it does not deal with the phenomenon of mismatch, which is also popular in the literature (see, for example, Kracke and Rodrigues, 2020). However, we are aware that labour demand (as for levels and structure) is driven, among other

[2] Similar outcome for changes in the employment structure is predicted when globalisation or offshoring issues are taken into account (see, for example, Cavenaile, 2021, or Oldenski, 2014).

[3] However, Mallick and Sousa (2017) show that technology is biased towards skilled individuals, which leads to a growing supply of skilled labour, still reporting labour polarisation and hollowing out the middle of the skill distribution.

[4] According to the forecast from the System for Forecasting the Polish Labour Market, by 2030, routine cognitive and routine manual jobs are going to account for 7.6% and 47.5% of employment, respectively. Likewise, a growing importance of non-routine jobs is expected (cognitive analytical: 14.6%, cognitive personal: 22.2%, and manual: 8.1%).

factors, by labour supply availability. In this respect, some peculiarities of the Polish labour market should be commented on. Firstly, the massification of higher education, which started in Poland in the 1990s, led to an oversupply of individuals with a university degree. These graduates are (formally) highly-qualified and should be able to perform non-routine tasks effectively. However, some studies show that higher education institutions are not successful in equipping graduates with the skills demanded by the labour market, and that there is an overall shortage of transversal skills (Strawinski et al., 2018); at the same time, according to our results, there seems to be an unnecessary oversupply of some simple skills, for instance office skills, which could be bounded and learned during the on-the-job training. Secondly, a decline of the vocational education system in Poland (in terms of quality and number of pupils/graduates) caused a decline in the supply of individuals who may perform routine manual tasks professionally. At the same time, demand for routine-manual jobs has been quite stable, since many transnational companies located assembly lines in Poland, partly as a result of the relatively lower labour cost compared to the Western EU member states (Arendt and Grabowski, 2019). Additionally, this labour cost competitiveness stemming from low wages caused labour shortages in many industries. The scale of this shortage has been reduced to a large extent by an inflow of migrants, mainly from the eastern neighbouring countries (Ukraine, Belarus).

Another important aspect to keep in mind is that the regional heterogeneity of the Polish labour market (Rollnik-Sadowska et al., 2020) may affect important labour market processes, including that of polarisation –we will address this issue in the paper. This heterogeneity is displayed in relatively small differences across regions in terms of professional activity (and thus inactivity). For example, participation and employment rates (for individuals aged 15+) in 2020 varied between 52.6% and 51.2% respectively in the Slaske region and 60.1% and 58.0% in the Mazowieckie region. However, there are dimensions in which regional heterogeneity is much larger –for instance, in terms of employment structure, unemployment (Lewandowska-Gwarda, 2018; Tatarczak and Boichuk, 2018), wages (Adamchik and Hyclak, 2017; Rokicki et al., 2021), gender wage gap (Majchrowska and Strawiński, 2016, 2018), or wage premium across task-content groups (Arendt and Grabowski, 2019).

The employment structure by industries[5] and regions have a significant influence on labour demand and, therefore, on the structure of vacancies/job offers. The percentage of workers in the agriculture sector in 2020 varied between 6% in the Slaskie region and 36% in the Lubelskie region; in the manufacturing industry this percentage varied between 19% in the Lubelskie region and 35% in the Slaskie region, while in the service industry varied between 44% in the Podkarpackie region and 69% in the Mazowieckie region. Unemployment also shows large regional differences: the registered unemployment rate at the end of December 2021 varied between 3.0% in the Wielkopolska region to 8.6% in the Warmia and Mazury regions, while *Labour Force Survey* (LFS) unemployment rate in the 3rd quarter 2021 ranged from 2.0% in the Dolnoslaskie region to 4.7% in the Lodzkie region. As for wages, average gross wage in 2020 ranged from 4763 PLN in the Warmia and Mazury regions to 6562 PLN in the Mazowieckie region (in relative terms this translates into a 38% difference in wages between high-wage and low-wage regions).

To sum up, the atypical polarisation pattern in Poland may stem from: (i) the specific employment structure (sectoral and regional) inherited from the times of the centrally-planned economy (for instance, a relatively high share of agricultural workers in total employment); (ii) significant educational upgrading being a result of the increase in the tertiary education enrolment rate, coinciding with the stigmatisation of vocational education (Arendt and Grabowski, 2019); and (iii) globalisation, especially in terms of offshoring processes, which have been dynamic in Poland in recent years (before the COVID-19 pandemic).

[5] Significant differences are also visible in the employment structure by age groups or educational attainment. Moreover, it is argued that if employment in the industries which suffered most because of COVID-19 would not recover after the pandemic, this may translate into even bigger regional differences in the Polish labour market (OECD, 2021).

3. Methodology and data

CTs are mainly used to cross-classify categorical data. Let I denote the number of categories of a variable X and J the number of categories of a variable Y. If the categories of X are ordered in rows and the categories of Y are ordered in columns, a rectangular or wide table having I rows and J columns arises. This table displays the number of times n_{ij} that each cell ij or combination of categories is observed, and is called *contingency table* (CT_{IJ}). As we will see next, knowing the observed frequencies n_{ij} in a CT_{IJ}, its marginal frequency distributions (row and column totals), and the sample size n, it is possible to measure the similarities between the categories of variables and the association between the variables; these two measures are at the heart of multivariate statistical analysis.

Suppose that n individuals are randomly sampled from a very large population and that their characteristics are detailed in a set of descriptive variables. Further, suppose that this multivariate sample information is introduced in a CT (CT_{IJ}) where the categories of X (in rows) and Y (in columns) are obtained through the combination of the categories of the different descriptive variables, which can be categorical, binary or ordinal –continuous variables can be treated as if they were categorical, dividing its total range into a limited number of intervals. This one-to-one correspondence among rows and columns of the combined variables and their categories is illustrated in Table 1, where we combine m variables in rows (v_1 with r_1 categories, v_2 with r_2 categories, …, v_k with r_k categories, …, v_m with r_m categories) and p variables in columns (v_1' with r_1' categories, v_2' with r_2' categories, …, v_z' with r_z' categories, …, v_p' with r_p' categories). Then, the total number of X categories will be $I = \prod_{k=1}^{m} r_k$, and the total number of Y categories will be $J = \prod_{z=1}^{p} r_z'$. Each row $i = \{i_1, i_2, …, i_k, …, i_m\}$ corresponds to the combination of the particular category i_1 of v_1 with the particular category i_2 of v_2, etc. Each column $j = \{j_1, j_2, …, j_z, …, j_p\}$ corresponds to the combination of the particular category j_1 of v_1' with the particular category j_2 of v_2', etc. Each cell ij corresponds to the match of combined categories from both sides of the table: $i = \{i_1, i_2, …, i_k, …, i_m\}$ in the rows side, matches with $j = \{j_1, j_2, …, j_z, …, j_p\}$ in the columns side, with an observed frequency n_{ij} and marginal totals n_{i+} and n_{+j}.

The notion of propensity to associate between each row category and each column category in two-way CTs plays a central role in our paper; we start from the hypothesis that there are certain attributes of online vacancies which tend to be strongly associated. This propensity can be related to the notion of "departure from independence" in CTs (the difference between the observed cell frequencies and the cell frequencies expected under the independence hypothesis). However, the usual approach in CTs is mostly global (chi-squared tests of independence, etc.), whereas in our paper we make an individual analysis for

Table 1
Two-dimensional contingency table.

	Y categories							
X categories		1	2	…	j	…	J	Total
	1	n_{11}	n_{12}	…	n_{1j}	…	n_{1J}	n_{1+}
	2	n_{21}	n_{22}	…	n_{2j}	…	n_{2J}	n_{2+}
	…	.	.	…	.	…	.	.
	i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iJ}	n_{i+}
	…	.	.	…	.	…	.	.
	I	n_{I1}	n_{I2}	…	n_{Ij}	…	n_{IJ}	n_{I+}
	Total	n_{+1}	n_{+2}	…	n_{+j}	…	n_{+J}	n

Table 2
Frequency and percentage of job offers according to tasks-content groups.

Task-content groups	All job offers		Job offers with assigned skills	
	Frequency	Percentage	Frequency	Percentage
Non-routine cognitive analytical	759,787	16.86	246,199	15.64
Non-routine cognitive personal	1,161,370	25.77	443,599	28.18
Non-routine manual physical	305,282	6.78	110,789	7.04
Routine cognitive	683,879	15.18	244,826	15.55
Routine manual	1,595,501	35.41	528,709	33.59

Source: Own computations.

each cell, seeking in particular to identify which are the column categories with greater propensity to associate with each particular row category. Furthermore, in our paper we use the ratio between observed and expected cell frequencies instead of their difference.

We can measure the propensity to associate, or association factor (a_{ij}) between the row and column combinations in cell (i, j), as the ratio of the probability estimated from the observed frequencies to the random probability:

$$\widetilde{a}_{ij} = \frac{Observed\ probability\ of\ cell\ (i,j)}{Random\ probability\ of\ cell\ (i,j)} = \frac{\widetilde{p}_{ij}}{\widetilde{p}_{i+}\widetilde{p}_{+j}} = \frac{\widetilde{n}_{ij}/\widetilde{n}}{\frac{\widetilde{n}_{i+}}{n}\frac{\widetilde{n}_{+j}}{n}} = \frac{\widetilde{n}\cdot\widetilde{n}_{ij}}{\widetilde{n}_{i+}\widetilde{n}_{+j}} \tag{1}$$

Ratio values higher than one mean that the association degree is greater than in the random case, and vice versa. The symbol "~" over the variables indicates that the CT has been smoothed following the procedure proposed by Álvarez de Toledo et al. (2018, 2020). The main advantage of using smoothing techniques in the CT setting is that they provide solutions for estimating cell frequencies and their probabilities in the presence of sparsity. Sparsity may arise in finite samples when a CT is generated by the combination of multiple variables (or by the combination of few variables but with many categories) describing rows and/or columns. In a CT table like CT_{IJ}, the number of cells, *I x J*, can be so high that many cells with positive occurrence probabilities can be zero or have a very small frequency if the sample is not large enough. In this scenario of small samples with sparsity, multivariate statistical analyses (as, for example, correspondence analyses, association factors or χ^2 tests of independence) may lose the optimal properties that they have for large samples.

In a sparse high-dimensional CT with many cells with no information (zero frequencies) or limited information (small frequencies), smoothing methods can "borrow" information from neighbouring cells and obtain frequency and probability estimates that can outperform the unsmoothed estimates. Using a Cobb-Douglas (or Log-Log) functional form, Álvarez de Toledo et al. (2020) estimate the association factor (a_{ij}) between the row combined-categories and column combined-categories of a CT as a multiplicative function of the partial association factors between the row individual-categories and column individual-categories which are the result of considering the combined variables separately. In this way, they disentangle the mix of the effects of the multiple interactions among the different variables. The decomposition equation can be used to estimate the frequencies of the CT cells resulting in a smoothed estimated CT. The authors demonstrate the parsimony of their smoothing method comparing its results with the log linear and kernel-based methods.

In this study we use data on online vacancies collected within the SOJO, developed by the Institute of Labour and Social Studies in Warsaw. SOJO retrieves data on vacancies from the following Polish web portals: pracuj.pl, gazetapraca.pl, praca.pl, careerjet.pl, gratka.pl, and olx.pl, with the use of text scraping techniques and a deduplication mechanism (this mechanism ensures that only unique job offers are downloaded –e.g. if the same job offer is posted on different web portals, the web crawler selects only one to be stored in the database). SOJO provides 5 categorical variables which describe each vacancy: industry (19 industry sectors according to NACE Rev. 2 classification at the section level); region (16 NUTS-2 regions of Poland, so called voivodeships); and three variables which are related to the skills required by the job, namely: sub-major (3-digit) occupational group (131 groups stemming from ISCO-08 classification –however, for the empirical analysis we use 38 major 2-digit occupational groups–), required skills (based on the classification used in the BHC study, PARP (2011) –see Table A1 in the Appendix–), and the task content of the job classified in line with the ALM model[6] –Table A2, in the Appendix, shows the assignment of occupation groups to task-content groups.

In our approach, we examine online job offers, covering the 2017–2019 years. Considering the period 2017–2019 jointly, our sample corresponds to n = 4,505,819 online job offers. Table 2 offers the frequency and percentage of job offers according to task-content groups (distinguishing between all job offers and offers with assigned skills[7]).

Each vacancy is described by three categorical variables that are cross-classified in a CT, where the rows represent occupation groups and the columns represent the categories of the cross-variable "region and activity". The CT has 40 rows or occupation groups (2-digit ISCO) and 304 columns or region-activity combinations (16 NUTS-2 Polish regions across 19 NACE Rev. 2 activities), resulting in a total of 12,160 cells, of which approximately 40% have zero frequency –the largest cell has 61,321 job vacancies.

Starting from the smoothed table $\widetilde{CT}_{I_1 I_2 \dots I_m}$ $(\widetilde{n}_{i_1 i_2 \dots i_m})$, which represents a sample of approximately 4.5 million online vacancies, the strategy that we followed was to analyse if certain job skill requirements tend to be associated with certain activities and regions, and vice versa. Among the three variables that can be used as a proxy for the job skill requirements, we chose the ISCO-08 occupation group to create the rows of the CT, given its higher level of disaggregation and its high correlation with the task-content groups. This analysis of statistical associations is carried out at the cluster level and at the vacancy level.

In a *I x J* CT with a large number of categories in rows and columns, the large number of arbitrarily ordered cells (*I x J*) makes it difficult to get an overall picture of the object to which the CT refers. In particular, a comprehensive analysis of the propensities to associate in each cell ("who associates with whom") may be difficult if we consider a very large number of cases. Clustering methodology enables us to address all these problems by ordering and grouping the different categories (in

[6] To categorise job offers into task-content groups and to minimise the problem of potential misclassification between task-content groups and occupational groups (ISCO-08 classification), we apply the approach followed by Hardy et al. (2018) who, in turn, follows the Acemoglu and Autor (2011) methodology. Those authors translate the US O*NET descriptions to the Polish LFS dataset by adjusting the dominant task-content of jobs to the Polish conditions. The transition matrix developed by Hardy et al. (2018) links properly the occupations at 3-digit code level of the *Polish Classification of Occupations and Specialities* to the respective task-content groups. Since each job posting in our database is assigned to the ISCO-08 (3-digit) occupational group, we are naturally able to re-classify these postings into ALM task-content groups.

[7] Job portals differ perceptibly in terms of the requirements related to the content of the job offer that can be published on-line, and there is no uniform job offer template. In some cases, the individual job offers do not contain information about demanded skills.

rows and columns) in a lower number of clusters and the *I x J* cells in a lower number of biclusters. According to Ailem et al. (2017), one of the benefits of the biclustering process is that, by merging rows and columns in larger homogeneous blocks, we are able to overcome the problem of sparsity in the context of small samples or highly segmented CTs.

Our clustering approach allows us to obtain a bicluster map (heatmap) of associations between clusters of occupation groups (row clusters) and clusters of region-activity combinations (column clusters); the structure of each bicluster within the bicluster map can be described in terms of skills and task-content groups, among other variables. The clustering methodology is based on a similarity (or dissimilarity) measure between the elements that are clustered. In an association context, we consider that row (column) categories are the more similar the more they resemble the way they associate with column (row) categories. For instance, in our application to online job offers, we argue that two occupation groups are the more similar the more they resemble the way they associate with different categories of the cross-variable region and activity, and vice versa. Consequently, we measure the similarity between each pair of rows of the CT (i_A and i_B) as the overlapping or percentage of coincidence of their row profiles (distribution of their conditional probabilities $\widetilde{p}_{ij}/\widetilde{p}_{i+}$ of matching with each of the different column categories *j*).

$$\widetilde{sim}_{i_A-i_B} = \sum_j min\left(\frac{\widetilde{p}_{i_Aj}}{\widetilde{p}_{i_A+}}, \frac{\widetilde{p}_{i_Bj}}{\widetilde{p}_{i_B+}}\right) \tag{2}$$

Its value can be between one (if the row profiles are identical) and zero (if their intersection is null). We can measure the similarity between each pair of columns (j_A and j_B) of the CT in an analogous way:

$$\widetilde{sim}_{j_A-j_B} = \sum_i min\left(\frac{\widetilde{p}_{ij_A}}{\widetilde{p}_{+j_A}}, \frac{\widetilde{p}_{ij_B}}{\widetilde{p}_{+j_B}}\right) \tag{3}$$

Based on such similarity measures, we use a hierarchical method of clustering, merging the two categories (rows or columns) with the highest similarity into a new category and, subsequently, with categories gradually fusing to form increasingly larger categories or clusters.

4. Results

Our cluster/bicluster analysis provides new evidence on the structure of the Polish labour market. Fig. 1 shows the dendrogram of the 38 (2-digit ISCO-08) occupation groups for which vacancies are observed in the sample –a dendrogram is a diagram representing a tree which illustrates the generation of clusters produced by the similarity analyses. As can be observed, some occupation groups are relatively similar in the way they associate with the region and the sector of activity of the job offer (for example, "Protective services workers" and "Market-oriented skilled agricultural workers"; "Numerical and material recording clerks" and "Mining, construction, manufacturing and transport"; or "Administrative and commercial managers" and "Business and administration professionals"), while other groups show an idiosyncratic behaviour that little resembles that of the other groups (as for example, "Food preparation assistants" or "Agricultural, forestry and fishery labourers").

It can be interesting to relate the similarities based on the association of occupation groups with regions and activities to those partial similarities based on the association of occupation groups with regions and occupation groups with activities separately. In this way, we can identify which weight each partial similarity had in the global similarity. Table 3 shows the results of the Cobb-Douglas estimation:

$$\widetilde{sim}_{occup-reg\&act} = b_0 \cdot \widetilde{sim}^{\,b_1}_{occup-reg} \cdot \widetilde{sim}^{\,b_2}_{occup-act} + \varepsilon_{occup-reg\&act} \tag{4}$$

$$\varepsilon_{occup-reg\&act} \sim N\left(0, \sigma^2_\varepsilon\right)$$

The 38 × 38 matrix of similarities between the 38 occupation groups is a symmetric matrix, where the main diagonal and the upper triangular matrix represents 741 observations –that is, 741 = 38 + (38·37/2). The estimate has a high goodness-of-fit and shows that the elasticity of $\widetilde{sim}_{occup-reg\&act}$ to $\widetilde{sim}_{occup-reg}$ is 0.59, while the elasticity of $\widetilde{sim}_{occup-reg\&act}$ to $\widetilde{sim}_{occup-act}$ accounts to 0.9; that is, when the similarity in the way of matching with the different regions of two occupation groups ($\widetilde{sim}_{occup-reg}$) increases by 1%, the similarity of those occupations in the way of matching with the different regions and activity sectors ($\widetilde{sim}_{occup-reg\&act}$) increases by 0.59%, while this last similarity increases by 0.9% when what grows by 1% is the similarity of the occupations due to the way of matching with the different sectors of activity ($\widetilde{sim}_{occup-act}$) –in other words, the variability of $\widetilde{sim}_{occup-act}$ is more determining than the variability of $\widetilde{sim}_{occup-reg}$ when explaining the variability of $\widetilde{sim}_{occup-reg\&act}$, although the estimated coefficient of $\widetilde{sim}_{occup-reg}$ is not negligible. These values led us to conclude that the clustering process of the occupation groups by association with regions and activities (dendrogram of Fig. 1) is mainly guided by the association map between occupation groups and activities, although regions also have influence on the clustering process.

When the clustering process is applied to both the rows (2-digit occupations) and the columns (regions across activities) of the CT, three things can be observed. First, the initial ordering in which rows and columns are displayed (an order that was, in principle, arbitrary) was changed to that corresponding to the base of the respective dendrograms, so that the categories with greater similarity are placed closer together. Second, as we go up through both dendrograms (the one of the rows and the one of the columns), a decreased number of row and column clusters and biclusters (combinations of a row cluster and a column cluster) is observed in the CT. Third, in our (average linkage) hierarchical cluster, a criterion to determine the optimal number of clusters has to be defined. Our criterion was based on the intra-cluster and inter-cluster distances. A well-structured cluster is one where the intra-cluster distance (the sum of the distances between each element and the centroid of its cluster) is relatively small and the inter-cluster distance (the sum of the distances between the centroids of the different clusters and the overall centroid) is relatively large. In a hierarchical cluster, it is not easy to set the optimal number of clusters because the inter-cluster distance tends to fall monotonously as the number of clusters is reduced, that is, as we climb the dendrogram (grouping elements), while the opposite happens to the intra-cluster distance. It is desirable that, as we group the units, the distance between groups is reduced as little as possible and the opposite occurs for the distance within groups.

Fig. 2 relates the inter-cluster and intra-cluster distances for each possible number of clusters (or height on the dendrogram) for both the row (occupation) cluster –Graph (a)– and the column (region and activity) cluster –Graph (b). The black line in both graphs represents the sum of their respective inter and intra-cluster distances for the different dendrogram heights. The respective optimal heights occur at the minimum of the respective sum functions –observe (more clearly in Graph b) that the growth of the respective sum functions is the signal that the intra-cluster distance is starting to increase considerably, which is not convenient. Therefore, the resulting optimal numbers are 12 clusters of rows or occupation groups (the red line in Fig. 1), and 20 clusters of columns or region-activity combinations.

We used our cluster results to order and optimally group the rows and columns of the smoothed CT. Aggregating the frequencies of each resulting row cluster, column cluster and row-column bicluster, we can generate a 12 × 20 association map (heatmap) at bicluster level, as Table 4 shows. In the table, each cell or bicluster is coloured more or less in red depending on whether the association factor between the cluster row and the cluster column is greater or lower. The biclusters with a higher propensity to associate (the redder ones) represent sets of online job offers whose required occupations tend to be associated with the same particular group of vacancy regions and activities, and vice versa. In other words, those biclusters are the most distant from a scenario of

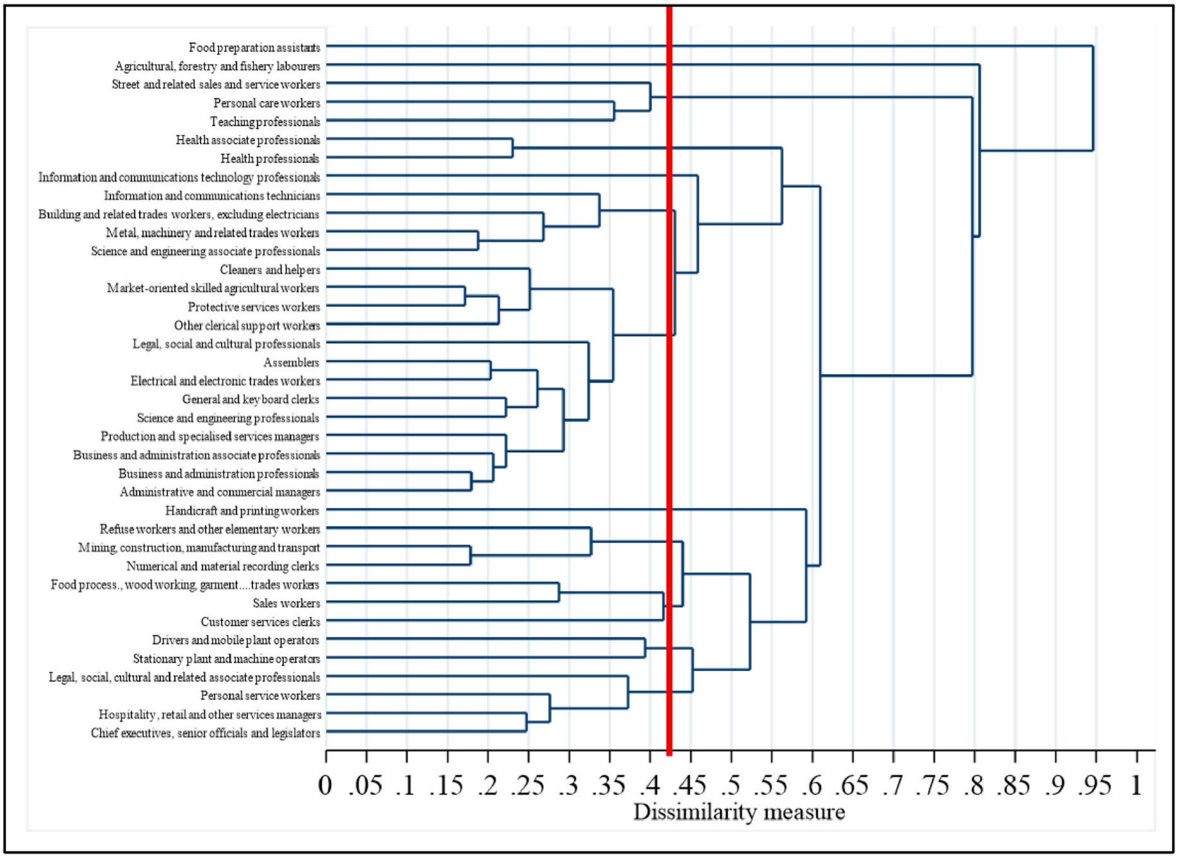


Fig. 1. Occupation group cluster (by association with regions and activities).

Table 3
Estimation of global similarity from partial similarities.

Explanatory variables	Coef.	Std. Err.	*t*	$P > t$	[95% Conf. Interval]	
Constant	0.996	0.003	397.4	0.000	0.991	1.000
Similarity occupation-region	0.587	0.021	28.1	0.000	0.546	0.628
Similarity occupation-activity	0.902	0.004	210.2	0.000	0.894	0.911

Number of obs = 741; R-squared = 0.9987; Adj R-squared = 0.9987; Root MSE = 0.0182.

random assignment of occupation groups to regions and activities; a random assignment that would correspond to a unitary association factor.

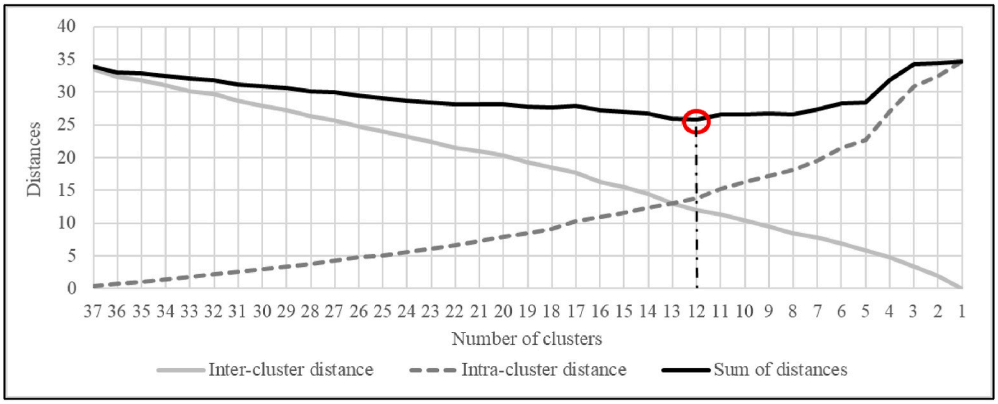
Our analysis reveals that the biclusters appear to be quite concentrated in terms of occupation groups and regions/activities combinations. Some biclusters could be considered obvious or expected, like those observed in health, education or food services, but even these biclusters are interesting in terms of the rest of attributes which describe them. An example of such obvious association is the occupational cluster presented in row 9, which contains only health professionals and associate health professionals. Those occupations are visibly associated with column 12, and very weakly with others. Column 12 represents a cluster containing the whole “Human health and social work activities” sector in the most-developed Polish regions. This association explains most of the assignment of health professional occupations to sectors and regions in Poland.

As we already mentioned, the structure of each bicluster within the bicluster map can be described in terms of skills and task-content groups, among other variables –note that the description of variables that do not belong to those used to create the CT (as is the case with skills or tasks) must be made on the original CT (not smoothed) as these characteristics are not observed for the “virtual” observations that the smoothing procedure created. Table 5 describes those biclusters that show an association factor greater than 1.5 and a volume of vacancies greater than 10,000 (to avoid excessive table size, only the categories of each variable that exceeded 10% are shown).

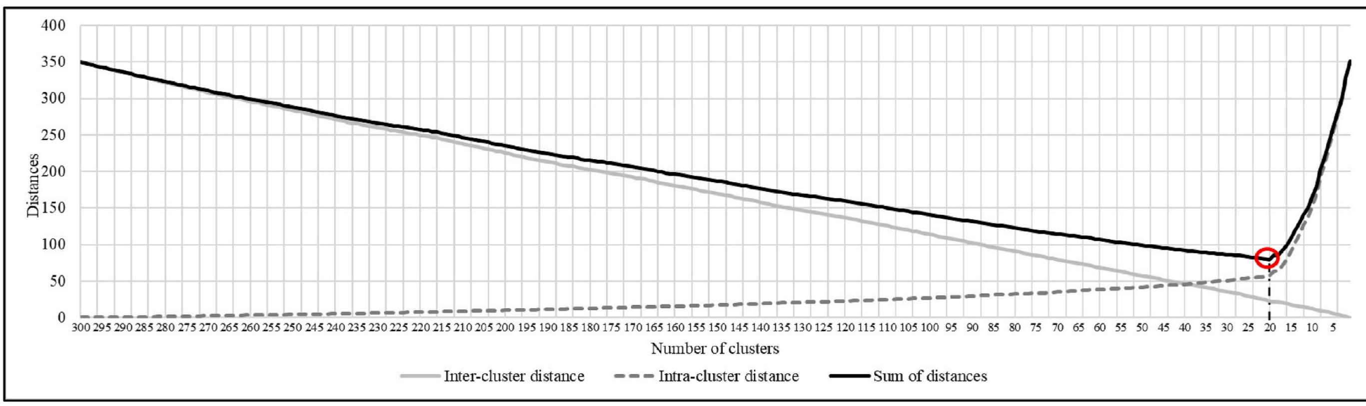
As a single vacancy may require several skills, the skills description in Table 5 has been obtained by replacing each observation or vacancy i in the database (in long format) with m_i copies of that observation which only differ in the required skills, that is, each copy only contains one of the skills required by the copied vacancy; m_i is therefore the number of skills required by the vacancy i. This expanded database contains 4,643,058 rows, which are the result of expanding the 1,574,175 job offers for which required skills were known –note that the biclusters of Table 5 have been ordered by partially altering the dendrograms of rows (occupation groups) and columns (regions and activities) in order to place at the beginning of the table those biclusters where routine tasks represent an important weight.

5. Discussion

In this section, we discuss some of the results observed in our bicluster/heatmap. Most of the biclusters in Table 4 show expected results; this is the case, for example, of the bicluster {row cluster 12,



Graph (a). Row clusters



Graph (b). Column clusters

Fig. 2. Inter-cluster and intra-cluster distances for each possible number of clusters.

Table 4
Bicluster heatmap of occupation groups, regions and activities.

		20 Region & Activity clusters																			
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
12 (2-digit) occupation clusters	**1**	0.27	0.10	0.15	0.28	0.09	2.45	1.03	0.45	0.83	1.78	0.32	0.58	0.22	0.17	0.07	0.13	3.21	1.98	5.26	0.32
	2	0.08	0.99	1.68	0.02	0.76	3.48	0.32	0.09	0.37	0.08	0.18	0.18	3.77	0.14	7.41	0.04	0.40	0.03	0.35	0.56
	3	0.12	0.63	1.71	0.67	0.68	1.28	1.91	0.14	0.39	0.19	0.21	0.88	0.10	0.05	0.07	0.05	0.30	5.83	1.52	0.12
	4	1.03	0.57	1.38	0.30	0.68	1.36	2.87	1.18	0.15	0.64	0.23	0.28	0.71	0.07	4.57	0.24	0.85	0.16	1.74	0.13
	5	0.08	0.10	0.06	0.09	1.57	0.26	1.36	0.09	0.41	4.23	0.10	0.01	0.10	0.11	0.11	11.3	0.10	0.19	0.60	0.01
	6	0.84	1.38	0.85	1.84	0.95	0.75	0.76	1.70	1.44	1.24	1.10	0.20	0.61	1.38	0.48	0.35	0.23	0.17	0.48	0.34
	7	0.70	1.54	1.44	0.06	3.23	0.63	0.39	0.29	1.03	1.25	0.91	0.19	4.27	3.04	0.59	0.49	0.15	0.11	0.37	0.03
	8	0.14	0.33	0.39	0.40	0.05	0.32	0.35	0.61	1.27	1.45	0.46	0.62	0.07	0.13	0.08	11	0.02	0.17	0.09	0.02
	9	0.13	0.25	0.56	0.42	0.42	0.21	0.49	0.09	0.83	0.14	8.78	20.4	0.17	0.17	0.01	0.02	0.13	1.21	2.44	0.48
	10	0.09	0.08	0.05	0.07	0.52	0.27	0.45	0.36	0.25	0.20	2.87	5.34	0.04	0.02	0.05	0.25	0.06	0.18	0.34	31.43
	11	1674	0.10	0.02	0.10	0.03	5.21	0.02	0.08	0.10	0.08	0.10	5.96	0.09	0.10	0.03	0.09	0.18	0.03	4.47	0.10
	12	0.09	0.16	0.29	0.11	0.35	0.01	0.04	0.09	0.00	0.67	0.10	0.02	0.10	0.11	0.03	0.00	54.1	0.03	0.37	0.02

column cluster 17}, which shows that vacancies of the occupation group "Food preparation assistants" have a probability of being assigned to the activity "Accommodation and food service activities" which is well above the probability corresponding to a random assignment. In addition, from Table 5 we know that this cluster contains 100% routine manual jobs that mainly require Interpersonal (31%) and Office skills (28%). On the other hand, there are some biclusters in which the relationship in terms of occupations and activities is not easily interpreted or intuitive. For example, vacancies in occupations "Numerical and material recording clerks" (67%) and "Labourers in mining, construction, manufacturing and transport" (26%) tend to associate with the activity sector "Transporting and storage" {row cluster 4, column cluster 15, in Table 4}. The vacancies of this bicluster mainly contain Interpersonal (36%), Technical (24%) and Self-organisational (21%) skills for performance of Routine cognitive (67%) and Routine manual (33%) tasks. The association of occupation "Numerical and material recording

Table 5
Large biclusters with a high degree of association.

Assoc. Factor	Cluster occup.	Cluster Reg-Act	Online job offers	Competence	Comp. (%)	Task Group	Task (%)	(2 digit) ISCO occupation groups	ISCO (%)	NACE activity sector	NACE (%)	Region	Reg. (%)
3.2	1	17	14,315	Interpersonal skills	28%	Routine manual	87%	Personal service workers	87%	Accommodation and food service activities	100%	Mazowieckie	20%
				Technical skills	21%							Malopolskie	15%
				Self-organisational skills	21%								
				Availability	11%								
2.0	1	18	19,570	Office skills	21%	Routine manual	86%	Personal service workers	86%	Other services activities	100%	Mazowieckie	21%
				Interpersonal skills	20%	Non-routine manual physical	11%	Legal, social, cultural and related associate professionals	11%			Pomorskie	15%
				Self-organisational skills	19%							Slaskie	11%
				Technical skills	19%								
3.5	2	6	81,918	Technical skills	38%	Routine manual	100%	Drivers and mobile plant operators	90%	Wholesale and retail trade; repair of motor vehicles and motorcycles	100%	Mazowieckie	16%
				Interpersonal skills	22%			Stationary plant and machine operators	10%				
				Office skills	17%								
				Self-organisational skills	15%								
7.4	2	15	41,005	Interpersonal skills	38%	Routine manual	100%	Drivers and mobile plant operators	99%	Transporting and storage	100%	Dolnoslaskie	15%
				Technical skills	29%							Mazowieckie	13%
				Self-organisational skills	19%							Wielkopolskie	12%
												Lodzkie	10%
1.9	3	7	187,575	Interpersonal skills	28%	Routine manual	83%	Sales workers	73%	Administrative and support service activities	100%	Mazowieckie	16%
				Technical skills	26%	Routine cognitive	17%	Customer services clerks	17%			Slaskie	11%
				Self-organisational skills	22%			Food processing, wood working, garment and other craft and related trades workers	10%				
5.8	3	18	129,099	Interpersonal skills	30%	Routine manual	98%	Sales workers	96%	Other services activities	100%	Mazowieckie	14%
				Self-organisational skills	24%							Slaskie	12%
				Technical skills	20%							Pomorskie	11%
				Office skills	13%								
2.9	4	7	199,218	Interpersonal skills	32%	Routine manual	68%	Refuse workers and other elementary workers	53%	Administrative and support service activities	100%	Mazowieckie	15%
				Technical skills	19%	Routine cognitive	32%	Numerical and material recording clerks	32%			Wielkopolskie	14%
				Self-organisational skills	19%			Labourers in mining, construction, manufacturing and transport	15%			Dolnoslaskie	12%
				Office skills	16%							Slaskie	11%
												Lodzkie	11%

(*continued on next page*)

Table 5 (*continued*)

Assoc. Factor	Cluster occup.	Cluster Reg-Act	Online job offers	Competence	Comp. (%)	Task Group	Task (%)	(2 digit) ISCO occupation groups	ISCO (%)	NACE activity sector	NACE (%)	Region	Reg. (%)
3.2	7	5	188,951	Interpersonal skills	29%	Routine manual	52%	Science and engineering associate professionals	40%	Manufacturing	100%	Dolnoslaskie	13%
				Technical skills	26%	Non-routine manual physical	48%	Metal, machinery and related trades workers	28%			Slaskie	12%
				Self-organisational skills	21%			Building and related trades workers, excluding electricians	24%			Mazowieckie	12%
												Wielkopolskie	10%
3.0	7	14	10,866	Office skills	35%	Routine manual	99%	Building and related trades workers, excluding electricians	99%	Real estate activities	100%	Slaskie	16%
				Interpersonal skills	28%							Mazowieckie	11%
				Technical skills	15%								
				Self-organisational skills	12%								
54.1	12	17	45,189	Interpersonal skills	31%	Routine manual	100%	Food preparation assistants	100%	Accommodation and food service activities	100%	Mazowieckie	18%
				Office skills	28%							Pomorskie	12%
				Self-organisational skills	17%							Slaskie	11%
				Technical skills	14%							Malopolskie	11%
5.3	10	12	12,557	Technical skills	66%	Routine cognitive	74%	Personal care workers	77%	Human health and social work activities	100%	Mazowieckie	28%
				Self-organisational skills	13%	Non-routine cognitive personal	23%	Teaching professionals	23%			Malopolskie	11%
				Interpersonal skills	13%							Pomorskie	10%
4.6	4	15	55,781	Interpersonal skills	36%	Routine cognitive	67%	Numerical and material recording clerks	67%	Transporting and storage	100%	Dolnoslaskie	21%
				Technical skills	24%	Routine manual	33%	Labourers in mining, construction, manufacturing and transport	26%			Mazowieckie	15%
				Self-organisational skills	21%							Wielkopolskie	15%
												Lodzkie	12%
												Slaskie	11%
2.4	1	6	82,130	Interpersonal skills	22.5%	Non-routine cognitive personal	49%	Hospitality, retail and other services managers	43%	Wholesale and retail trade; repair of motor vehicles and motorcycles	100%	Mazowieckie	16%
				Technical skills	22.2%	Routine manual	43%	Personal service workers	43%			Dolnoslaskie	10%
				Self-organisational skills	21.8%							Slaskie	10%
31.4	10	20	66,124	Interpersonal skills	27%	Non-routine cognitive personal	67%	Teaching professionals	67%	Education	100%	Mazowieckie	24%
				Technical skills	24%	Routine cognitive	31%	Personal care workers	31%			Pomorskie	14%
				Self-organisational skills	19%							Wielkopolskie	10%
				Office skills	10%								

(*continued on next page*)

Table 5 (*continued*)

Assoc. Factor	Cluster occup.	Cluster Reg-Act	Online job offers	Competence	Comp. (%)	Task Group	Task (%)	(2 digit) ISCO occupation groups	ISCO (%)	NACE activity sector	NACE (%)	Region	Reg. (%)
20.4	9	12	53,718	Self-organisational skills	30%	Non-routine cognitive personal	62%	Health professionals	73%	Human health and social work activities	100%	Mazowieckie	24%
				Interpersonal skills	25%	Non-routine manual physical	27%	Health associate professionals	27%			Slaskie	14%
				Technical skills	19%	Non-routine cognitive analytical	11%					Malopolskie	12%
				Office skills	15%								
11.0	8	16	105,021	Technical skills	30%	Non-routine cognitive analytical	100%	Information and communications technology professionals	100%	Information and communication	100%	Mazowieckie	26%
				Interpersonal skills	29%							Malopolskie	15%
				Self-organisational skills	20%							Slaskie	12%
												Dolnoslaskie	11%
1.8	6	4	66,357	Interpersonal skills	30%	Non-routine cognitive analytical	59%	Business and administration professionals	63%	Financial and insurance activities	100%	Mazowieckie	18%
				Technical skills	25%	Non-routine cognitive personal	31%	Business and administration associate professionals	23%			Slaskie	12%
				Self-organisational skills	18%	Routine cognitive	11%						

clerks" with the "Transporting and storage" activity may be explained by the characteristics of this sector and the dynamic growth of logistics centres in Poland in recent decades[8]; since logistics focuses more and more on optimisation of routes and thus costs, companies look for individuals able to provide calculations and record-keeping types of tasks, which were routine, by the token of the ALM model. As for the sector "Labourers in mining, construction, manufacturing and transport", it seems that most of the vacancies are concentrated in the transportation sector, with less demand for workers in construction and manufacturing branches, as well as in the mining industry.

The highest association (1673.8) is recorded in the bicluster {row cluster 11, column cluster 1} in Table 4. This is a small cluster −124 online offers; not described in Table 5– containing specialised occupations in the sector "Agriculture, forestry and fishing". The third among the highest associated is the bicluster {row cluster 10, column cluster 20} in Table 4. It associates teaching professionals and personal care workers with the education sector. These occupations are connected to intense non-routine cognitive personal tasks and routine cognitive personal tasks, and require strong interpersonal, but also technical skills. The connection of personal care workers to the education sector shows that these kinds of professionals may be required to participate in the vocational (or even tertiary) education system. More advanced medical skills are the source of another strong association {row cluster 9, column cluster 12} in Table 4. This bicluster includes health professionals working in health care. Job offers for them require the use of specific technical skills for non-routine tasks, for which the demand-supply ratio has always been high, but also important transversal skills, especially self-organisational and interpersonal ones.

A detailed analysis of the results presented in Table 5 leads to interesting conclusions. Firstly, in a prevailing number of cases, the Mazowieckie region records the highest volume of online vacancies, which only confirms a well-known characteristic of the Polish labour market –labour demand is distributed unevenly across regions, with a leading role of Mazowieckie (especially Warsaw, capital of Poland) in the employment structure. Secondly, task-content groups reveal perceptible internal differences in terms of required skills, depending on the occupational group in which the recruitment takes place. It may sound pretty obvious; however, this issue has not been discussed within the labour market polarisation framework. Even within the routine manual task-content group, where interpersonal skills usually dominate, there are examples of biclusters in which technical skills were in high demand (e.g. Drivers and mobile plant operators, and Stationary plant and machine operators). It should also be noted that, in general, the required skill-mix varies within the task-content group. Thirdly, by analysing the task-content groups and NACE industries dominating in the biclusters presented in Table 5, we are able to identify those industries that are most susceptible to automation in the Polish economy –sectors such as "Accommodation and food service activities", "Transportation and storage", "Administrative and support activities", or "Education" (see Frey and Osborne, 2017; Fossen and Sorgner, 2022). Unsurprisingly, the "Information and communication" sector seems to be least prone to automation. These are important insights from the point of view of the polarisation hypothesis within the labour heatmap.

An advantage of our methodology is that it allows us to zoom in on any bicluster and analyse its structure of rows (occupations) and columns (regions and activities). As an example, Table 6 amplifies the bicluster {row cluster 6, column cluster 4} from Table 4 which belongs to the activity sector "Financial and insurance activities". Although the association factor of the overall bicluster is 1.8, there are rows and columns within it whose propensity to associate is much greater. For example, the occupation group "Business and administration professionals" shows a propensity to associate with all regions of the bicluster that exceeded 4 in all cases, being particularly large with the regions of Lodzkie, Opolskie, Swietokrzyskie and Slaskie. This is not the case for other occupation groups, such as "General and keyboard clerks" or "Other clerical support workers", which tend to be strongly associated only with certain regions –Slaskie and Kujawskopomorskie in the first group and Mazowieckie in the second one. Returning to the "Business and administration professionals" group, we observe that non-routine cognitive analytical and personal job offers dominate this row of the table with percentages across regions always close to or greater than 90% of the online job offers. Furthermore, interpersonal and technical skills are the most frequently required in this occupation group, always appearing across regions in more than 20% of the vacancies, which is in line with the common perception of the job profile in this sector. This observation points to the fact that, in some occupation groups (it does not have to occur in all occupation groups), there are certain task-content groups and skills that are in high demand irrespective of the regional dimension. This, in turn, could point to the existence of a potential imbalance stemming from insufficient labour supply of those task-content groups and skills. At the same time, an uneven regional distribution of labour demand in the case of other occupations/jobs may be a symptom of socio-economic regional disparities, which may be strengthened by inter-regional disparities in the occupational structure of labour (Krupowicz, 2020).

Our cluster analysis based on CTs can be complemented with an econometric model that estimates the probability of demanding certain skills from non-routine vs. routine task-content groups, both at general and bicluster level. For this purpose, we estimate three logistic regressions (models 1–3, Table 7).

The logistic model 1 explains the extent to which skills are interrelated with routine or non-routine tasks. This informs us about what skills generate a lower risk of "hollowing out" the labour market. The dependent variable takes value 1 if the job offer encompasses non-routine tasks and 0 if it covers routine tasks. The results show that three skills are required in routine jobs more often than in non-routine ones; these are: Office, Availability and Cognitive skills. Office and Availability skills are basic skills. Cognitive skills appear in jobs with routine and non-routine task-groups, but the former were in high demand in Poland, hence their slightly stronger connection to routine tasks. In the case of non-routine tasks, the strongest relation is with ICT and Communication skills. Moreover, non-routine jobs require a wider range of skills than routine ones.

The following two estimates allow us to analyse the specificity of those biclusters of occupations, regions and activities with the strongest association (see heatmap of Table 4). We choose six biclusters with association factor $a_{ij} > 10$ for a case study –in this way, our sample is reduced to 273,629 online vacancies. We find that most of those biclusters were homogenous in terms of task groups; that is, they are dominated by a single task group –note that each individual vacancy in the sample is assigned to a single task group. The exception is bicluster "10.20", which contains jobs with both routine and non-routine cognitive tasks and routine manual tasks; routine tasks are mostly carried out in this group by personal care workers, while non-routine ones by teaching professionals. With the logistic model 2, we analyse which skills are associated with tasks within this bicluster compared to all vacancies in the estimation sample that do not belong to it. We also include routinisation-skill interactions in the estimation. We find that, in the case of both types of task-groups altogether (routine and non-routine), Communication, Managing, and Technical skills are the most important for jobs from this bicluster. Interactions between skills and task-content groups point out that to be in a non-routine job in bicluster "10.20" increases (with respect to the rest of the vacancies in the sample) the probability of demanding Interpersonal, Availability, Self-organisational, and Office skills. These results are interesting because

[8] Bentyn (2016) showed that in the global logistic performance ranking Poland moved from 40th place in 2007 to 31st place in 2014. In recent years, the logistic performance index kept increasing –in 2018 Poland was in 28th place (https://lpi.worldbank.org/international/scorecard/radar/254/C/POL/2018).

Table 6
Bicluster {row cluster 6, column cluster 4} from Table 4.

Activity →	Financial and insurance activities															
Regions → (2-digit) Occupation groups ↓	Dolnoslaskie	Pomorskie	Malopolskie	Podlaskie	Kujawskopomorskie	Lubelskie	Warminsk-omaz-urskie	Lubuskie	Wielkopolskie	Podkarpackie	Zachodniopomorskie	Lodzkie	Opolskie	Swiätokrzyskie	Slaskie	Mazowieckie
Administrative and commercial managers	0.83	1.06	1.09	0.74	0.62	1.27	0.84	1.02	1.02	1.40	0.59	0.60	0.64	0.68	0.86	0.94
Business and administration professionals	4.84	4.57	4.88	4.45	5.27	5.30	5.12	5.29	5.27	5.20	5.27	5.64	6.03	5.77	5.74	4.34
Business and administration associate professionals	1.69	1.99	1.87	2.46	1.41	1.36	1.31	1.63	1.62	1.70	1.95	1.27	1.39	1.50	1.40	1.51
Production and specialised services managers	0.10	0.10	0.09	0.27	0.16	0.13	0.19	0.19	0.10	0.12	0.13	0.11	0.12	0.13	0.10	0.18
Science and engineering professionals	0.07	0.05	0.06	0.05	0.06	0.05	0.05	0.08	0.05	0.06	0.05	0.06	0.08	0.07	0.09	0.14
General and keyboard clerks	0.33	0.49	0.31	0.37	1.00	0.32	0.40	0.32	0.33	0.39	0.27	0.35	0.41	0.40	1.05	0.57
Electrical and electronic trades workers	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
Assemblers	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Legal, social and cultural professionals	2.14	1.19	0.95	1.62	1.15	1.97	1.95	1.47	1.28	1.10	1.20	1.12	1.28	1.64	0.85	1.29
Other clerical support workers	0.84	0.88	0.69	0.65	0.83	0.74	0.85	0.52	0.71	0.73	0.73	0.81	0.38	0.39	0.62	1.59
Protective services workers	0.06	0.14	0.06	0.07	0.06	0.07	0.10	0.05	0.07	0.05	0.09	0.08	0.08	0.10	0.08	0.23
Market-oriented skilled agricultural workers	0.09	0.13	0.09	0.09	0.18	0.08	0.11	0.08	0.15	0.08	0.12	0.09	0.11	0.13	0.09	0.08
Cleaners and helpers	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01

they show that some biclusters (or local labour markets in terms of occupations, regions and activities) can show an idiosyncratic behaviour in terms of the skills required by their job offers, which may have consequences in terms of polarisation. For example, the fact that the non-routine vacancy jobs of bicluster “10.20” (vs. the rest of the sample) more likely demand Office and Availability skills might be indicative that the risk of polarisation/automation of this particular bicluster (or, at least, of some job offers within it) could be relatively larger compared to other local labour markets with non-routine job offers. Non-routine jobs from this bicluster contain teaching professionals. The risk of polarisation for this occupation may be justified by a strong worldwide tendency to use online courses. Poland is lagging in their use, but the demand for them grows strongly, and potentially will be much higher in the future at a cost to traditional teaching. It may also be connected to the prospects of decreasing demand for traditional teaching services due to Polish demographic trends.

Finally, we estimate a multinomial logistic regression (model 3) with the six aforementioned high-association biclusters in order to compare skill requirements. The reference bicluster is again the “10.20”. Biclusters “5.16” (Handicraft and printing workers), “11.1” (Agricultural, forestry and fishery labourers), and “12.17” (Accommodation and food service activities) contain only jobs with routine manual tasks (they might be at risk of automation). This fact results in lower Communication, Managing, Technical, and Cognitive skill requirements for those three biclusters compared to “10.20”. Likewise, job offers from 100% non-routine bicluster “9.12” (Human health and social work activities) require Managing, Self-organisational, and Technical skills more often than “10.20” vacancies do. Finally, bicluster “8.16” stands out for being the most specialised of the six chosen. It requires ICT specialists for non-routine cognitive analytical tasks. From such workers, companies required strong ICT and Technical skills, with low requirements in other skill dimensions, which indicates that this bicluster may be considered as one with low risk of polarisation.

Our study concludes by deepening the analysis of the skills required by online job offers. One of the results observed in the description of the bicluster heatmap (Table 5) is that Interpersonal, Self-organisation, Office and Technical skills are the ones most required in high association biclusters with a large size. There is no doubt that the first two types of skills are categorised as transversal, and thus required across many occupations or task-content groups. However, if we look carefully at the definition of technical skills as proposed in the BHC study (see Table A1 in the Appendix), we may come to the conclusion that Office and Technical skills may also be treated as semi-transversal. To get a complete picture of the relationship between all the skills and the remaining vacancy requirements, we have clustered them taking into account their degree of association with the cross-variable occupation-activity –these calculations have been obtained with the aforementioned expanded database. Fig. 3 shows the resulting dendrogram. Basically, five well-differentiated clusters of skills are observed (ordered from highest to lowest number of vacancies): (1) Self-organisational, Technical, Interpersonal and Availability skills, (2) Cognitive and Managing skills, (3) Office skills, (4) ICT skills, and (5) Communication skills. Note that this dendrogram provides additional and complementary information to the previous logistic estimates, since it allows us to describe neighbouring skills that are close by the way they associate with occupations and industries.

Table 8 shows the description of these clusters –only the categories of each variable that exceeded 5% are shown. Cluster 1, the largest one, represents 77.7% of the expanded database of job offers. The assignment to task-content groups is less concentrated in this skill cluster, which seems to indicate that its transversal (or semi-transversal) skills are required in almost all types of vacancies. In clusters 2 (Cognitive and Managing skills, 12.5% of the offers), 4 (ICT skills, 3.3%) and 5 (Communication skills, 1.1%), non-routine cognitive (personal and analytical) tasks represent more than 60% of the cluster. This would indicate that the acquisition of these skills should mitigate the probability of being hollowed out of the labour market due to polarisation processes. On the other hand, since more than half of the cluster 3 (Office skills, 5.4%) corresponds to routine manual tasks, it seems that possessing mainly this type of skills puts a worker in relatively high risk of automation of the job place, so this skill does not provide a particular advantage in the labour market in the light of polarisation trends. From

Table 7
Skill-task connections. Regression results.

	Dependent variable		
	Non-routine tasks	Bicluster 10.20	Six chosen biclusters[a]
	logistic	logistic	multinomial logistic
	(1)	(2)	(3)
regional dummies	YES	NO	NO
2018	−0.165*** (0.013)		
2019	0.218*** (0.013)		
2020	−0.291*** (0.036)		
nonroutine		−0.547*** (0.011)	
communication	0.981*** (0.309)	3.160*** (0.856)	
managerial	0.725*** (0.058)	2.396*** (0.149)	
availability	−0.352*** (0.045)	−0.578*** (0.140)	
office	−1.048*** (0.019)	−0.656*** (0.037)	
interpersonal	0.030 (0.019)	−3.309*** (0.089)	
self-organisation	0.849*** (0.022)	−1.742*** (0.091)	
technical	0.793*** (0.023)	2.422*** (0.075)	
ICT	3.996*** (0.334)	1.455* (0.860)	
cognitive	−0.187*** (0.034)	1.002*** (0.120)	
nonroutine:communication		−1.256 (0.864)	
nonroutine:managerial		−2.293*** (0.154)	
nonroutine:availability		2.251*** (0.144)	
nonroutine:office		1.075*** (0.045)	
nonroutine:interpersonal		3.768*** (0.092)	
nonroutine:self-organisation		1.172*** (0.094)	
nonroutine:technical		−3.088*** (0.079)	
nonroutine:ICT		−2.970*** (0.864)	
nonroutine:cognitive		−0.764*** (0.124)	
(Intercept):11.1			−5.531*** (0.115)
(Intercept):12.17			0.615*** (0.009)
(Intercept):5.16			−1.971*** (0.021)
(Intercept):8:16			−28.466 (8825.804)
(Intercept):9.12			−27.993 (6455.563)
nrcp:11.1			−25.546 (6003.875)
nrcp:12.17			−30.649 (7197.291)
nrcp:5.16			−28.78 (6590.781)
nrcp:8:16			−0.618 (11,568.430)
nrcp:9.12			27.669 (6455.563)
nrmp:11.1			3.489 (59,188.750)
nrmp:12.17			−2.292 (63,494.270)
nrmp:5.16			−0.585 (59,608.600)
nrmp:8:16			27.892 (65,734.690)
nrmp:9.12			59.121 (46,360.180)
nrca:11.1			3.468 (18,851.660)
nrca:12.17			−2.269 (20,097.650)
nrca:5.16			−0.588 (19,330.910)
nrca:8:16			59.208 (17,354.860)
nrca:9.12			55.854 (16,277.920)
communication:11.1			−0.652 (142,905.600)
communication:12.17			−3.744*** (1.116)
communication:5.16			−1.887 (1.221)
communication:8:16			−2.663*** (0.809)
communication:9.12			−2.609*** (0.347)
managerial:11.1			−20.659 (24,037.040)
managerial:12.17			−2.287*** (0.151)
managerial:5.16			−3.412*** (0.271)
managerial:8:16			0.016 (0.104)
managerial:9.12			0.645*** (0.057)
availability:11.1			0.204 (1.033)
availability:12.17			0.358** (0.143)
availability:5.16			2.008*** (0.169)
availability:8:16			−0.373* (0.199)
availability:9.12			−1.961*** (0.073)
office:11.1			−22.241 (14,396.660)
office:12.17			0.726*** (0.037)
office:5.16			−1.547*** (0.137)
office:8:16			−1.820*** (0.070)
office:9.12			−0.050 (0.033)
interpersonal:11.1			5.831*** (0.212)
interpersonal:12.17			3.239*** (0.090)
interpersonal:5.16			3.901*** (0.106)
interpersonal:8:16			−0.202*** (0.066)
interpersonal:9.12			−0.238*** (0.029)
self-organisation:11.1			−0.696 (0.738)
self-organisation:12.17			1.715*** (0.091)

(continued on next page)

Table 7 (*continued*)

	Dependent variable		
	Non-routine tasks	Bicluster 10.20	Six chosen biclusters[a]
	logistic	logistic	multinomial logistic
	(1)	(2)	(3)
self-organisation:5.16			2.064*** (0.122)
self-organisation:8:16			−0.456*** (0.064)
self-organisation:9.12			1.235*** (0.032)
technical:11.1			−3.220*** (0.487)
technical:12.17			−2.402*** (0.075)
technical:5.16			−2.666*** (0.116)
technical:8:16			2.197*** (0.080)
technical:9.12			−0.293*** (0.035)
ICT:11.1			−15.380 (47,868.550)
ICT:12.17			−24.976 (60,819.490)
ICT:5.16			1.320 (0.898)
ICT:8:16			2.818*** (0.437)
ICT:9.12			0.109 (0.139)
cognitive:11.1			−21.106 (19,322.190)
cognitive:12.17			−0.932*** (0.121)
cognitive:5.16			−1.557*** (0.166)
cognitive:8:16			−0.089 (0.122)
cognitive:9.12			−1.365*** (0.053)
Constant	0.611*** (0.033)	−0.689*** (0.009)	
Observations	273,629	273,629	273,629
R^2			0.67
Log Likelihood	−147,990.8	−145,302.8	−125,355.5
Akaike Inf. Crit.	296,037.5	290,645.6	
LR Test			508,030.100*** (df = 65)

Notes: standard errors are in parentheses; *p < 0.1; **p < 0.05; ***p < 0.01.
[a] reference bicluster 10.20; nrca: non-routine cognitive analytical tasks, nrcp: non-routine cognitive physical tasks, nrmp: non-routine manual physical tasks.

the regional point of view, we can observe that the better developed Polish regions (with the Mazowieckie region being the first) provide most job offers for every skill cluster. Paying attention to sectors, the highest labour demand comes from Professional, scientific and technical activities in all clusters; followed by sectors such as Administration, Trade and Manufacturing –all these sectors have a significant weight in the Polish economy and seem to require a wide range of skills. Finally, regarding the occupation groups, we observe that the most frequent vacancies correspond to the groups of Business professionals, Clerical and administration workers, Sales workers and ICT professionals; an exception to this predominance is found in the relatively high weight of Refuse workers and other elementary workers in the case of Office skills. The predominance of those occupations is due to the fact that they are relatively frequent in the population of online vacancies and, in turn, tend to demand a wide range of skills –this does not happen with the occupations involving manual and physical tasks, where the range of required skills is usually small; the companies demanding these occupations do not usually demand transversal skills, but job-related skills.

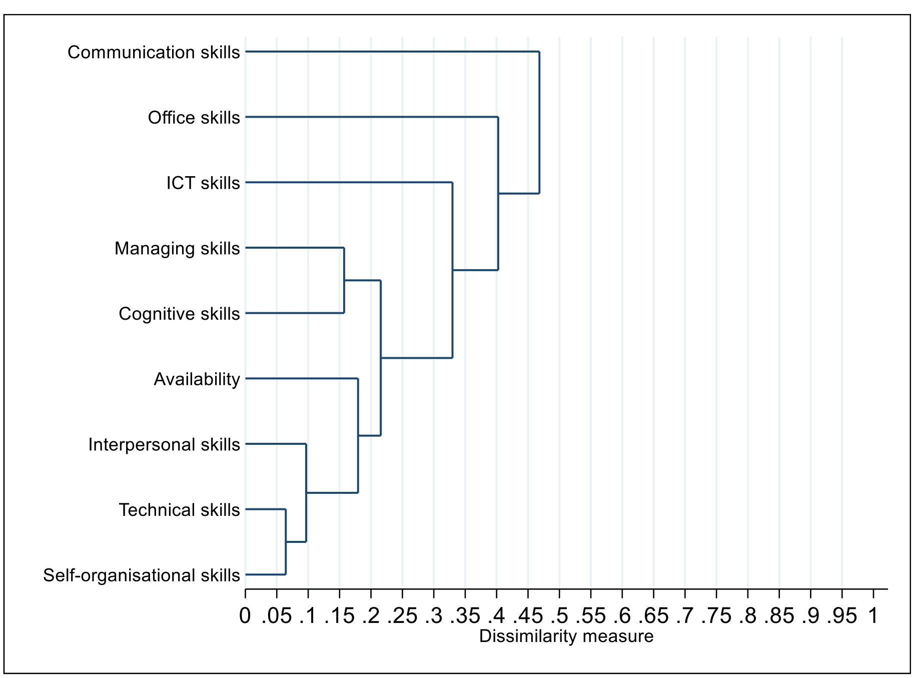


Fig. 3. Clusters of skills.

Table 8
Clusters of skills. Description.

Competences	Online job offers	Task Group	Task (%)	(2 digit) ISCO occupation groups	ISCO (%)	NACE activity sector	NACE (%)	Region	Reg. (%)
Self-organisational skills	4,359,354 (77.7%)	Non-routine cognitive personal	33.1%	Business and administration associate professionals	14.1%	Professional, scientific and technical activities	42.3%	Mazowieckie	21.5%
Technical skills Interpersonal skills		Routine manual	27.6%	Business and administration professionals	12.2%	Administrative and support service activities	17.5%	Dolnoslaskie	10.1%
Availability skills		Non-routine cognitive analytical	16.7%	Other clerical support workers	7.3%	Manufacturing	13.4%	Slaskie	10.1%
		Routine cognitive	15.6%	Sales workers	7.2%	Wholesale and retail trade; repair of motor vehicles and motorcycles	12.2%	Wielkopolskie	9.8%
		Non-routine manual physical	6.9%	Administrative and commercial managers	5.7%			Malopolskie	8.1%
				Information and communications technology professionals	5.5%			Pomorskie	6.8%
								Lodzkie	6.6%
Cognitive skills Managing skills	702,577 (12.5%)	Non-routine cognitive personal	40.5%	Business and administration professionals	14.6%	Professional, scientific and technical activities	49.1%	Mazowieckie	25.2%
		Non-routine cognitive analytical	20.9%	Business and administration associate professionals	14.5%	Administrative and support service activities	16.1%	Slaskie	9.9%
		Routine cognitive	17.5%	Other clerical support workers	9.6%	Wholesale and retail trade; repair of motor vehicles and motorcycles	12.1%	Dolnoslaskie	9.8%
		Routine manual	15.4%	Administrative and commercial managers	8.9%	Manufacturing	10.9%	Wielkopolskie	9.3%
		Non-routine manual physical	5.7%	Information and communications technology professionals	7.0%			Malopolskie	8.4%
				Science and engineering professionals	5.8%			Pomorskie	6.9%
				Sales workers	5.3%			Lodzkie	6.3%
Office skills	302,143 (5.4%)	Routine manual	51.3%	Refuse workers and other elementary workers	11.6%	Professional, scientific and technical activities	34.2%	Mazowieckie	21.6%
		Non-routine cognitive personal	19.9%	Business and administration associate professionals	9.3%	Administrative and support service activities	21.9%	Slaskie	11.5%
		Routine cognitive	13.7%	Sales workers	8.7%	Wholesale and retail trade; repair of motor vehicles and motorcycles	16.1%	Dolnoslaskie	9.9%
		Non-routine cognitive analytical	9.4%	Business and administration professionals	7.4%	Manufacturing	9.9%	Wielkopolskie	9.2%
		Non-routine manual physical	5.7%	Drivers and mobile plant operators	7.4%	Other services activities	5.5%	Malopolskie	9.1%
				Personal service workers	5.8%			Pomorskie	6.6%
								Lodzkie	5.9%
ICT skills	184,723 (3.3%)	Non-routine cognitive analytical	30.4%	Business and administration professionals	18.5%	Professional, scientific and technical activities	60.6%	Mazowieckie	30.8%
		Non-routine cognitive personal	29.8%	Business and administration associate professionals	15.8%	Administrative and support service activities	14.8%	Slaskie	10.1%
		Routine cognitive	28.9%	Other clerical support workers	15.3%	Manufacturing	8.1%	Dolnoslaskie	9.7%
		Routine manual	6.3%	Science and engineering professionals	8.8%	Wholesale and retail trade; repair of motor vehicles and motorcycles	7.2%	Wielkopolskie	9.1%
				Information and communications technology professionals	8.7%			Malopolskie	7.5%
				Administrative and commercial managers	7.1%			Pomorskie	6.5%
								Lodzkie	6.4%
Communication skills	62,402 (1.1%)	Non-routine cognitive personal	73.5%	Business and administration associate professionals	34.0%	Professional, scientific and technical activities	41.6%	Mazowieckie	20.0%
			9.4%		23.9%	Manufacturing	21.8%	Wielkopolskie	9.8%

(continued on next page)

Table 8 (*continued*)

Competences	Online job offers	Task Group	Task (%)	(2 digit) ISCO occupation groups	ISCO (%)	NACE activity sector	NACE (%)	Region	Reg. (%)
		Non-routine cognitive analytical		Business and administration professionals					
		Routine cognitive	8.7%	Administrative and commercial managers	13.8%	Wholesale and retail trade; repair of motor vehicles and motorcycles	15.7%	Slaskie	9.4%
		Routine manual	7.6%	Sales workers	5.8%	Administrative and support service activities	10.0%	Dolnoslaskie	8.4%
				Other clerical support workers	5.6%			Malopolskie	7.6%
								Pomorskie	6.9%
								Lodzkie	6.2%

6. Conclusions

The aforementioned literature analyses labour market polarisation mainly according to changes in employment. In this study, we use job offers as proxies of vacancies, which enable us to observe unmet demand for labour –a hard to observe and less known fraction of the labour demand– and also to analyse the skill demand. Observation of job vacancies provides an opportunity to identify direct and timely demand for workers and detailed characteristics of new workplaces. On the basis of job vacancies one can infer the possible directions in the employment flows, and the barriers to job creation in routine and non-routine task-content jobs. Skills included in job offers enable us to look at the polarisation patterns not only from the common perspective of occupations (proxies of job-related skills), but also from the perspective of transversal skills. This gives us a valuable addition to the analysis, because non-routine tasks may require not only technical knowledge, but also certain soft skills. Thus, we report the skills (and the skill-mix) which provide better prospects for individuals in the labour market undergoing automation and polarisation processes.

To analyse clustering patterns of online job offers in the Polish labour market, we take advantage of the categorical approach provided by the CT methodology. On the basis of this approach, we were able to analyse in detail the connections between tasks, regions, occupations, sectors, and skills. Specifically, we generated different biclusters (or local labour markets) by identifying those occupations, regions and sectors that tend to appear together in the vacancies offered. These biclusters can then be analysed in terms, not only of the variables used to generate them, but also of the skills and tasks required by the job offers that belong to them. Our CT methodology can be seen as a useful information system for policy makers (and labour economists), since it allows them to monitor (in a static or dynamic way) the demands for skills in the different local labour markets that coexist in the economy. In this way, those clusters/biclusters that evolve towards higher levels of automation could be identified, which can help the transition of workers at risk of polarisation towards employment-generating sectors.

Some structural changes have been observed in various economic sectors in Poland. Demographic changes lead to a dynamic increase in medium-level and high-level healthcare system workers, as well as in the instructors of the required skills. Such changes were also visible in logistics, where routine tasks started to demand numeracy and material recording job-related skills, as well as interpersonal and self-organisational transversal skills. These structural changes may affect the identified biclusters differently. Thus, those biclusters dominated by routine activities are the ones that would have to be followed over time to see if they are diluted as a result of the robotisation of the economy. The opposite would apply to those biclusters focused on analytical and interpersonal non-routine job offers; they should maintain, or even increase, their presence in the online job offers. According to our data, there are some biclusters whose sectors are at high risk of automation (given their skills and task requirements), such as Accommodation and food service activities, Transportation and storage, Administrative and support activities, and Education.

We analyse the probability of demanding certain skills for routine and non-routine tasks. Of the nine skills that we analysed, the most important for non-routine tasks are ICT and Communication skills, while the least important are Office skills and Availability skills. Those skills differentiate between the typical routine and non-routine task workers. The question of whether differences between routine and non-routine jobs depend on whether the tasks are cognitive or manual can be studied at general and bicluster level. At both levels, we observe that manual non-routine tasks (versus manual routine ones) demand more skills that are more difficult to automate, especially ICT and communication skills, and that those differences are not so significant when the tasks are cognitive.

Our results may shed light on which skills should be developed at the vocational and tertiary education levels in Poland in order to adjust the skills supply to the changes in labour demand due to the automation and polarisation processes –our study complements recent discussion on this topic (see for example Antczak et al., 2019; Gajdos et al., 2020). It seems that there will still be a high demand for routine jobs in Poland, as the labour market lacks such workers. At the same time, given that technical change and polarisation are dynamic phenomena, we can foresee a gain in the share of non-routine jobs (mainly cognitive analytical, but also manual) in labour demand and employment structure; this fact will make the development of transversal skills within the life-long learning scheme even more important than it is today. Thus, within the framework of an adequate educational policy, vocational education should provide, not only technical skills, but also certain key transversal skills. The supply of Communication, Managing and ICT skills should especially be increased to ensure the transition of the economy to non-routine manual tasks. We also find that, at least in the Polish vacancy market, Office skills do not offer a particular advantage in the light of the polarisation trends, because they are perceived by employers as basic skills that every employee should have. The extensive supply of such simple skills is not needed, as they can be learned during on-the-job training. On the contrary, ICT and Communication skills can increase the chances of employment for an individual in the upper levels of the skill distribution (reducing the routinisation hypothesis risk). Tertiary education should be more focused on teaching and developing such skills.

A complementary policy recommendation, in the face of structural imbalances between the supply and demand for workers across task-content groups, would be to use the immigration policy, mainly from Eastern countries, to try to balance those gaps by means of the right incentives.

Declaration of competing interest

None.

Acknowledgments

We are grateful to the Research Group PAIDI SEJ-513 (Andalusian Board), the Project ECO2017-86780-R (Spanish Ministry of Science, Innovation and Universities), the I + D + i Project P20-00808 (Andalusian Board), and the National Science Centre of Poland (Project "The polarisation of the Polish labour market in the context of technical change", contract number 2016/23/B/HS4/00334), for the funding provided, and the Institute of Labour and Social Studies in Warsaw for providing data on online job offers. Funding for open access publishing: Universidad Pablo de Olavide/CBUA. We would also like to thank Pablo Álvarez de Toledo, the participants in *EcoMod 2021* conference, *2021 RSAI World Congress*, *XV Labour Economics Meeting* (Albacete, 2022), and the anonymous referees and the Editor of this journal for their useful comments and suggestions. All the remaining errors are our sole responsibility."

Appendix

Table A1
Classes of skills based on the Balance of Human Capital[a] study (and SOJO).

Skills	Behavioural dimension	Behavioural sub-dimensions
Cognitive	Retrieval and analysis of information; drawing conclusions	Quick summarising large amounts of text Logical thinking, analysis of facts Constant learning of new things
Computer	Computer and Internet use	Basic knowledge of office software Knowledge of specialised software, ability to develop software or creating websites Internet use: searching websites, e-mail operations
Technical	Technical imagination and the use of devices	Use of devices Ability to repair the devices
Self-organisation	Self-organisation of work and taking initiative (planning and implementation of tasks on time, achieving goals)	Making decision independently Entrepreneurship and taking initiative Creativity (being innovative, coming up with new solutions) Resistance to stress Implementation of planned activities on time
Interpersonal	Developing relations with colleagues, clients or subordinates	Team-working Ease in entering into contact with co-workers and customers Solving conflicts between people
Communication	Communication with other people	Being communicative and ability to communicate thoughts clearly
Office	Organising and performing office work	–
Managerial	Managerial skill and work organisation of others	Assigning tasks to the employees Work coordination Providing discipline at work
Availability	Availability	Readiness for frequent business trips Flexible working time

Source: own elaboration based on PARP (2011, p. 32).

[a] The original classification of skills from BHC has been adjusted to the data characteristics of online job offers retrieved from the SOJO system. Communication skills do not appear in the BHC study as a separate category. However, as we have identified this skill group in a large number of job offers, we have decided to separate it from Interpersonal skills, since the Interpersonal skills range is very broad in our microdata. We understand the term "communication" as debating, interrogating, persuading, and negotiating. Interpersonal skills include the use of language, spoken production, and non-verbal communication. Moreover artistic, physical, and mathematical skills (which are listed in the BHC study) do not appear in the final set of job offers that we analysed. We considered only job offers which were fully described by occupation, sector, region, task-content group and skill. These three types of skills might have appeared in advertisements of simple jobs –those which do not have a full description.

Table A2
Correspondence between occupation groups and task-content groups

ISCO groups (2 digits)		Non-routine cognitive personal	Non-routine cognitive analytical	Non-routine manual physical	Routine cognitive	Routine manual
12	Administrative and commercial managers	121, 122				
11	Chief executives, senior officials and legislators	111, 112				
13	Production and specialised services managers	131 to 134				
14	Hospitality, retail and other services managers	141 to 143				
23	Teaching professionals	231 to 235				
24	Business and administration professionals	243, 244	241, 242			
22	Health professionals	221 to 226, 229	227, 228			
21	Science and engineering professionals		211 to 216			
26	Legal, social and cultural professionals		261 to 265			
25	Information and communications technology professionals		251, 252			
32	Health associate professionals			321 to 325		
31	Science and engineering associate professionals			311 to 315		
35	Information and communications technicians			351, 352		
33	Business and administration associate professionals	332 to 335			331	
34	Legal, social, cultural and related associate professionals			342, 343	341	
41	General and keyboard clerks				411 to 413	
42	Customer services clerks				421 to 422	
43	Numerical and material recording clerks				431 to 432	
44	Other clerical support workers				441	

(*continued on next page*)

Table A2 (*continued*)

ISCO groups (2 digits)		Non-routine cognitive personal	Non-routine cognitive analytical	Non-routine manual physical	Routine cognitive	Routine manual
53	Personal care workers				531	532
51	Personal service workers					511 to 516
52	Sales workers					521 to 524
54	Protective services workers					541
61	Market-oriented skilled agricultural workers					611 to 613
62	Market-oriented skilled forestry, fishery and hunting workers					621, 622
63	Subsistence farmers, fishers, hunters and gatherers					631 to 634
71	Building and related trades workers, excluding electricians					711 to 713
72	Metal, machinery and related trades workers					721 to 723
73	Handicraft and printing workers					731, 732
74	Electrical and electronic trades workers					741, 742
75	Food processing, wood working, garment and other craft and related trades workers					751 to 754
81	Stationary plant and machine operators					811 to 818
82	Assemblers					821
83	Drivers and mobile plant operators					831 to 835
91	Cleaners and helpers					911, 912
92	Agricultural, forestry and fishery labourers					921
93	Labourers in mining, construction, manufacturing and transport					931 to 933
94	Food preparation assistants					941
95	Street and related sales and service workers					951 to 952
96	Refuse workers and other elementary workers					961 to 962

Source: Own elaboration based on Acemoglu and Autor (2011) and Hardy et al. (2018).
Note: Numbers in the table refer to the International Labour Organisation classification of occupations (ISCO). The rows indicate 2-digit occupation groups, while the content in each cell indicates 3-digit occupation groups assigned to task-content groups. Non-routine cognitive personal and analytical jobs are concentrated on the first and second major occupational groups (managers, professionals), while non-routine manual physical jobs fall into the third major group (technicians). Routine cognitive and routine manual jobs are more diversified along the major occupational groups: cognitive ones encompass technicians, clerical and services workers; and manual ones expand across five major groups –services, agricultural, craft, elementary workers and operators. This implies, from the point of view of the labour market polarisation hypothesis, that the hollowing out effect affects a wide range of occupations across many major occupational groups according to ISCO-08 classification.

References

Acemoglu, D., Autor, D., 2011. In: Card, D., Ashenfelter, O. (Eds.), Skills, Tasks and Technologies: Implications for Employment and Earnings, ume 4B. Handbook of Labor Economics, North-Holland, Amsterdam, pp. 1043–1171.

Acemoglu, D., Restrepo, P., 2019. Automation and new tasks: how technology displaces and reinstates labor. J. Econ. Perspect. 33 (2), 3–30.

Acemoglu, D., Restrepo, P., 2020. The wrong kind of AI? Artificial intelligence and the future of labour demand. Camb. J. Reg. Econ. Soc. 13 (1), 25–35.

Acemoglu, D., Autor, D., Hazell, J., Restrepo, P., 2021. AI and Jobs: Evidence from Online Vacancies. NBER Working Paper, 28257.

Adamchik, V.A., Hyclak, T.J., 2017. Economic transition and regional wages: the evidence from Poland. Transit. Stud. Rev. 24 (1), 47–69.

Adermon, A., Gustavsson, M., 2015. Job polarisation and task-biased technological change: evidence from Sweden, 1975-2005. Scand. J. Econ. 117 (3), 878–917.

Agarwal, N., 2015. An empirical model of the medical match. Am. Econ. Rev. 105 (7), 1939–1978.

Agresti, A., 2013. Categorical Data Analysis, third ed. Wiley, New York.

Ailem, M., Role, F., Nadif, M., 2017. Sparse Poisson latent block model for document clustering. IEEE Trans. Knowl. Data Eng. 29 (7), 1563–1576.

Álvarez de Toledo, P., Núñez, F., Usabiaga, C., 2018. Matching and clustering in square contingency tables. Who matches with whom in the Spanish labour market. Comput. Stat. Data Anal. 127 (C), 135–159.

Álvarez de Toledo, P., Núñez, F. y, Usabiaga, C., 2020. Matching in segmented labor markets: an analytical proposal based on high-dimensional contingency tables. Econ. Modell. 93 (C), 175–186.

Antczak, E., Gałecka-Burdziak, E., Pater, R., 2019. What affects efficiency in labour market matching at different territorial aggregation levels in Poland? Bull. Econ. Res. 71 (2), 160–179.

Arendt, Ł., Grabowski, W., 2019. Technical change and wage premium shifts among task-content groups in Poland. Econ. Res.-Ekon. Istraživanja 32 (1), 3392–3410.

Arntz, M., Gregory, T., Zierahn, U., 2016. The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. OECD, Social, Employment and Migration Working Papers, p. 189.

Asplund, R., Barth, E., Lundborg, P., Nilsen, K.M., 2011. Polarisation of the Nordic labour markets. Fin. Econ. Pap. 24 (2), 87–110.

Autor, D., 2015. Why are there still so many jobs? The history and future of workplace automation. J. Econ. Perspect. 29 (3), 3–30.

Autor, D., Dorn, D., 2013. The growth of low-skill service jobs and the polarisation of the US labor market. Am. Econ. Rev. 103 (5), 1553–1597.

Autor, D., Levy, F., Murnane, R.J., 2003. The skill content of recent technological change: an empirical exploration. Q. J. Econ. 118 (4), 1279–1333.

Bentyn, Z., 2016. Poland as a regional logistic hub serving the development of Northern corridor of the new silk route. J. Manag. Mark. Logist. 3 (2), 135–144.

Burman, P., 2004. On some testing problems for sparse contingency tables. J. Multivariate Anal. 88 (1), 1–18.

Cavenaile, L., 2021. Offshoring, computerization, labor market polarization and top income inequality. J. Macroecon. 69 (C), 103317.

Chiappori, P.A., 2020. The theory and empirics of the marriage market. Annu. Rev. Econ. 12 (1), 547–578.

Chiappori, P.A., Salanié, B., 2016. The econometrics of matching models. J. Econ. Lit. 54 (3), 832–861.

Cortes, G.M., Jaimovich, N., Siu, H.E., 2017. Disappearing routine jobs: who, how, and why? J. Monetary Econ. 91 (C), 69–87.

Deming, D., Kahn, L.B., 2018. Skill requirements across firms and labor markets: evidence from job postings for professionals. J. Labor Econ. 36 (S1), 337–369.

Den Haan, W.J., Ramey, G., Watson, J., 2003. Liquidity flows and fragility of business enterprises. J. Monetary Econ. 50 (6), 1215–1241.

Dustmann, C., Ludsteck, J., Schonberg, U., 2009. Revisiting the German wage structure. Q. J. Econ. 124 (2), 843–881.

Fienberg, S.E., Rinaldo, A., 2007. Three centuries of categorical data analysis: log-linear models and maximum likelihood estimation. J. Stat. Plann. Inference 137 (11), 3430–3445.

Fonseca, T., Lima, F., Pereira, S.C., 2018. Job polarisation, technological change and routinization: evidence for Portugal. Lab. Econ. 51 (C), 317–339.

Fossen, F., Sorgner, A., 2022. New digital technologies and heterogeneous wage and employment dynamics in the United States: evidence from individual-level data. Technol. Forecast. Soc. Change 175, 121381.

Frey, C.B., Osborne, M.A., 2017. The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Change 114 (C), 254–280.

Gajdos, A., Arendt, Ł., Balcerzak, A.P., Pietrzak, M.B., 2020. Future trends of labour market polarisation in Poland. The perspective of 2025. Transform. Bus. Econ. 19 (3), 114–135, 51.

Goos, M., Manning, A., 2007. Lousy and lovely jobs: the rising polarisation of work in Britain. Rev. Econ. Stat. 89 (1), 118–133.

Goos, M., Manning, A., Salomons, A., 2009. Job polarisation in Europe. Am. Econ. Rev. 99 (2), 58–63.

Goos, M., Manning, A., Salomons, A., 2014. Explaining job polarisation: routine-biased technological change and offshoring. Am. Econ. Rev. 104 (8), 2509–2526.

Green, D., Sand, B., 2015. Has the Canadian labour market polarized? Can. J. Econ. 48 (2), 612–646.

Haedo, C., Mouchart, M., 2018. A stochastic independence approach for measuring regional specialization and concentration. Pap. Reg. Sci. 97 (4), 1151–1168.

Hardy, W., Keister, R., Lewandowski, P., 2018. Educational upgrading, structural change and the task composition of Jobs in Europe. Econ. Transit. 26 (2), 201–231.

Hellerstein, J.K., Morrill, M.S., 2011. Dads and daughters. The changing impact of fathers on women's occupational choices. J. Hum. Resour. 46 (2), 333–372.

Hershbein, B., Kahn, L.B., 2018. Do recessions accelerate routine-biased technological change? Evidence from vacancy posting. Am. Econ. Rev. 108 (7), 1737–1772.

Kracke, N., Rodrigues, M., 2020. A task-based indicator for labour market mismatch. Soc. Indicat. Res. 149, 399–421.

Krupowicz, J., 2020. The convergence or divergence of labour resources in Poland? Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu 64 (4), 75–100.

Lewandowska-Gwarda, K., 2018. Geographically weighted regression in the analysis of unemployment in Poland. Int. J. Geo-Inf. 7 (17), 7010017.

Long, J., Ferrie, J., 2013. Intergenerational occupational mobility in great Britain and the United States since 1850. Am. Econ. Rev. 103 (4), 1109–1137.

Lovaglio, P.G., Cesarini, M., Mercorio, F., Mezzanzanica, M., 2018. Skills in demand for ICT and statistical occupations: evidence from web-based job vacancies. Stat. Anal. Data Min. 11 (2), 78–91.

Lyu, W., Liu, J., 2021. Soft skills, hard skills: what matters most? Evidence from job postings. Appl. Energy 300 (C), 117307.

Majchrowska, A., Strawiński, P., 2016. Regional differences in gender wage gaps in Poland: new estimates based on harmonized data for wages. Cent. Eur. J. Econ. Modell. Econ. 8, 115–141.

Majchrowska, A., Strawiński, P., 2018. Impact of minimum wage increase on gender wage gap: case of Poland. Econ. Modell. 70 (C), 174–185.

Mallick, S.K., Sousa, R.M., 2017. The skill premium effect of technological change: new evidence from United States manufacturing. Int. Lab. Rev. 156 (1), 113–131.

Marinescu, I., Wolthoff, R., 2020. Opening the black box of the matching function: the power of words. J. Labor Econ. 38 (2), 535–568.

Michels, G., Natraj, A., van Reenen, J.V., 2014. Has ICT polarized skill demand? Evidence from eleven countries over 25 years. Rev. Econ. Stat. 96 (1), 60–77.

Modestino, A.S., Shoag, D., Ballance, J., 2020. Upskilling: do employers demand greater skill when workers are plentiful? Rev. Econ. Stat. 102 (4), 793–805.

Mosteller, F., 1968. Association and estimation in contingency tables. J. Am. Stat. Assoc. 63 (321), 1–28.

OECD, 2021. Regional Economic Inactivity Trends in Poland. OECD Reviews on Local Job Creation, Paris.

Oldenski, L., 2014. Offshoring and the polarisation of the U.S. labor market. Ind. Labor Relat. Rev. 67 (3), 734–761.

PARP, 2011. *Bilans Kapitału Ludzkiego W Polsce* (Balance of Human Capital in Poland), Report. Polish Agency for Enterprise Development, Warsaw.

Parteka, A., 2018. Import intensity of production, tasks and wages: micro-level evidence for Poland. Entrep. Bus. Econ. Rev. 6 (2), 71–89.

Petitjean, F., Webb, G.I., Nicholson, A.E., 2013. Scaling log-linear analysis to high-dimensional data. In: Proceedings of the 13th IEEE International Conference on Data Mining, pp. 597–606.

Rey, C., Rey, S., Viala, J.R., 2014. Detection of high and low states in stock market returns with MCMC method in a Markov switching model. Econ. Modell. 41 (C), 145–155.

Rokicki, B., Blien, U., Hewings, G.J.D., thi Hong Van, P., 2021. Is there a wage curve with regional real wages? An analysis for the US and Poland. Econ. Modell. 102 (C), 105582.

Rollnik-Sadowska, E., Jarocka, M., Dąbrowska, E., 2020. Diversity of regional labour markets in Poland. Eur. Res. Stud. J. 23 (4), 33–51.

Roth, A.E., Sönmez, T., Ünver, M.U., 2004. Kidney exchange. Q. J. Econ. 119 (2), 457–488.

Samek, L., Squicciarini, M., Cammeraat, E., 2021. The Human Capital behind AI: Jobs and Skills Demand from Online Job Postings. OECD Science, Technology and Industry Policy Papers, p. 120.

Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer Series in Statistics, Heidelberg.

Squicciarini, M., Nachtigal, H., 2021. Demand for AI skills in jobs: evidence from online job postings. OECD Sci. Technol. Ind. Pol. Pap. 3.

Strawinski, P., Majchrowska, A., Broniatowska, P., 2018. Wage returns to different education levels. Evidence from Poland. Ekonomista 1, 25–49.

Tatarczak, A., Boichuk, O., 2018. The multivariate techniques in evaluation of unemployment analysis of Polish regions. Oecon. Copernic. 9 (3), 361–380.

Titterington, D.M., Bowman, A.W., 1985. A comparative study of smoothing procedures for ordered categorical data. J. Stat. Comput. Simulat. 21 (3–4), 291–312.

Van Roy, V., Vertesy, D., Vivarelli, M., 2018. Technology and employment: mass unemployment or job creation? Empirical evidence from European patenting firms. Res. Pol. 47 (9), 1762–1776.

Wang, J., Hu, Y., Zhang, Z., 2021. Skill-biased technological change and labor market polarization in China. Econ. Modell. 100 (C), 105507.

West, D.M., 2018. The Future of Work. Robots, AI, and Automation. Brookings Institution Press, Washington.

Zilian, L.S., Zilian, S.S., Jager, G., 2021. Labour market polarisation revisited: evidence from Austrian vacancy data. J. Labour Mark. Res. 55 (7) s12651-021-00290-4.