Economics Bulletin

Volume 43, Issue 4

Skill preferences in job postings

Andrei Ternikov
HSE University

Abstract

This paper investigates the order of skills mentioned in job ads, their frequency, and whether there is a relation between skill groups and salary offered. A novel methodology was used across three job board datasets to demonstrate existing skill preferences in job ads. By identifying skill preferences empirically, the methodology yields valuable insights into the job market.

The author would like to thank the reviewer for his/her insightful comments and suggestions, which the author found indeed valuable for the final version of the manuscript.
Citation: Andrei Ternikov, (2023) "Skill preferences in job postings", *Economics Bulletin*, Volume 43, Issue 4, pages 1928-1943
Contact: Andrei Ternikov - aternikov@hse.ru.
Submitted: July 03, 2023. **Published:** December 30, 2023.

1. Introduction

A growing body of academic literature explores the problem of analyzing the demand for skills in the labor market (Napierala and Kvetan, 2023). Recent studies focus on the rapid growth of information technology (IT) and artificial intelligence (AI), which has implications for the occupational structure of the labor market (Acemoglu et al., 2022; Alekseeva et al., 2021). Using the advances in job board data and focusing on the phenomenon of job polarization and the hypothesis of skill-biased technical change (Autor et al., 2003; Acemoglu and Autor, 2011), the authors investigate issues related to the relationship between technical and non-technical skills, their complementarity, and their association with the offered salary (Ao et al., 2023; Banfi and Villena-Roldán, 2019; Deming and Kahn, 2018; Deming and Noray, 2020). However, the authors primarily assess wage premiums based on whether a particular skill is present or absent in a job vacancy. To the best of my knowledge, a major limitation of previous studies is the lack of attention paid to the relationship between skill preferences (both in terms of ranking and frequencies) highlighted in job advertisements and their impact on disclosed salaries by companies.

This paper suggests that job ads reveal skill preferences through the order and frequency of mention, which in turn are linked to the salary offered. This study presents a novel approach for identifying skill preferences by supplementing standard regression analysis with graph analysis using data from three job boards.

2. Data

The paper uses three job posting datasets covering the Commonwealth of Independent States (CIS) region (HeadHunter job portal), the United Kingdom (Reed job board), and Singapore (MyCareersFuture platform). Each dataset contains job postings that include the following key fields: the date the job was posted, the salary offered, the job description, and the company identifier. Categorical variables representing occupation and experience are also included in some datasets.

Following the skill classification introduced by Deming and Noray (2020) and Alekseeva et al. (2021), 14 skill groups were considered. Skill names were extracted from job descriptions and provided as skill fields, preserving their order in the text within the SkillSpan deep learning framework (Zhang et al., 2022). Then, skill groups were assigned using the keyword and regular expression based skill mapping approach of Deming and Noray (2020). Finally, vacancies with disclosed salary were used. Table 1 shows the main characteristics of the proposed datasets used in the following analysis (the percentages of job postings by skill groups and their frequency are provided in Appendix A).

Table 1: Characteristics of the data sets.

Source	Region	Date Range	Currency	Pay freq.	Obs.	Experience	Occupation
HeadHunter[a]	the CIS	05/2015–03/2021	RUB	Monthly	988,187	yes	yes
Reed[b]	the UK	01/2018–03/2018	GBR	Hourly	30,940	no	yes
MyCareersFuture[c]	Singapore	05/2019–06/2019	SGD	Monthly	13,161	yes	no

Notes: [a] Job postings collected using HeadHunter API (https://dev.hh.ru/); [b] "50000 job board records from Reed UK" available at https://www.kaggle.com/datasets/jobspikr/50000-job-board-record-from-reed-uk/; [c] The dataset from Bhola et al. (2020).

3. Methodology

The research design is based on detecting skill preferences, estimating wage premiums using regression analysis, and graphs to represent preference relationships. To formalize the characteristics of job boards and the preference relations towards skills, let X denote the set of existing skill groups introduced *ex ante*. $S_i^n : (s_k \in X)_{k=1}^n$ — the n-tuple (an ordered sequence with repetitions) consisting of skill groups in the i-th job posting.

3.1. Preliminaries: Two definitions of skill preferences

Let S_i^n be a sequence of skill groups in the i-th job posting; a and b are two distinct skill groups such that $a, b \in X$. Then, I formalize two definitions of skill preferences that are used in the following analysis. The rationale for using two definitions relates to the varying structures of job postings. The initial definition pertains to the order in which a specific skill is mentioned in a job advertisement. For instance, skills mentioned first may hold greater value in the recruitment process, and subsequent skills may be considered supplementary. The second definition pertains to the comparison of skill frequency, where skills mentioned more frequently may be attributed to the primary tasks of a given vacancy.

Definition 1 (Strict preference in ranking)**.** Denote the strict preference in ranking between two different skill groups a and b in the i-th vacancy as $a \overset{r}{\succ}_i b$. A skill group a is strictly preferred over (succeeds) b in the i-th vacancy if $a, b \in S_i^n$ and the maximum position of a in S_i^n is lower than the minimum position of b in S_i^n.

According to definition 1, the binary indicator variables are constructed as follows:

$$a \overset{r}{\succ}_i b = \begin{cases} 1, & a, b \in S_i^n \text{ and } \max\{k \in \overline{1,n} \mid s_k = a\} < \min\{k \in \overline{1,n} \mid s_k = b\} \\ 0, & \text{otherwise} \end{cases},$$

where $\max\{k \in \overline{1,n} \mid s_k = a\}$ — maximum position (rank) of a in a tuple S_i, $\min\{k \in \overline{1,n} \mid s_k = b\}$ — minimum position of b in a tuple S_i.

Definition 2 (Strict preference in frequency)**.** Denote the strict preference in frequency between two different skill groups a and b in the i-th vacancy as $a \overset{f}{\succ}_i b$. A skill group a is strictly preferred over (succeeds) b in the i-th vacancy if $a, b \in S_i^n$ and the frequency of a in S_i^n is higher than the frequency of b in S_i^n.

According to definition 2, the binary indicator variables are constructed as follows:

$$a \overset{f}{\succ}_i b = \begin{cases} 1, & a, b \in S_i^n \text{ and } |\{k \in \overline{1,n} \mid s_k = a\}| > |\{k \in \overline{1,n} \mid s_k = b\}| \\ 0, & \text{otherwise} \end{cases},$$

where $|\{k \in \overline{1,n} \mid s_k = a\}|$ — the frequency of a in a tuple S_i, $|\{k \in \overline{1,n} \mid s_k = b\}|$ — the frequency of b in a tuple S_i.

An example of how skill groups are mapped with ranking and frequency preference relationships is shown in the Figure 1.

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ \text{Skill set :} & \text{Data Mining} & \text{NLP} & \text{Computer Vision} & \text{Teamwork} & \text{ML} \\ \downarrow \text{maps to Skill groups} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ (s_1, s_2, \dots, s_5) & \text{Data Analysis} & \text{AI} & \text{AI} & \text{Social} & \text{AI} \end{array}$$

Figure 1: An illustration of the mapping of skill sets to skill groups.

Notes: An example above refers to a sequence of five skills ("Data Mining", "NLP", "Computer Vision", "Teamwork", "ML") resulting in a sequence of skill groups ("Data Analysis", "AI", "AI", "Social", "AI") according to the classification of Deming and Noray (2020) and Alekseeva et al. (2021). The particular example shows only two strict preference relations in ranking ("Data Analysis" $\overset{r}{\succ}$ "AI" and "Data Analysis" $\overset{r}{\succ}$ "Social") and two strict preference relations in frequency ("AI" $\overset{f}{\succ}$ "Data Analysis" and "AI" $\overset{f}{\succ}$ "Social").

3.2. Wage and skill preferences: An empirical strategy

The primary empirical approach is based on the analysis of skill premiums using skill preferences transformed into indicator variables (e.g. Alekseeva et al., 2021; Deming and Noray, 2020). The following baseline regression model is estimated:

$$\ln W_{i,j,t} = \alpha + K_{i,j,t}^{'} \beta + \gamma_j + \delta_t + \varepsilon_{i,j,t}, \quad (1)$$

where $\ln W_{i,j,t}$ is the logarithm of the suggested salary in job posting i in the firm j, and year-month t. The vector $K_i = (\max\{a_k \succ_i b \mid a_k, b \in X \text{ and } a_k \neq b\})_{k=1}^{14}$ for the i-th job posting denotes a preference relation of 14 binary indicators representing skill groups, following the classification introduced by Deming and Noray (2020) and Alekseeva et al. (2021); γ_j and δ_t are firm and time fixed effects; $\varepsilon_{i,j,t}$ is an error term.[1]

The strategy facilitates the identification of skill-based premiums in a job posting. To illustrate, utilizing preference definitions can ascertain if a specific skill has a higher ranking than others (strict preference in ranking) or is mentioned more frequently (strict preference in frequency) in a vacancy.

3.3. Wage and skill preferences: A graph-based representation

The baseline empirical strategy eliminates the wage effect generated by pairwise preference relations between skills. Therefore, to extract knowledge from skill pairs, an alternative approach is also used.

First, a baseline regression model is estimated by including pairwise preferences among all skills as follows:

$$\ln W_{i,j,t} = \alpha + G^{'}_{i,j,t}\,\beta + \gamma_j + \delta_t + \varepsilon_{i,j,t}, \tag{2}$$

where $\ln W_{i,j,t}$ is the logarithm of the suggested salary in job posting i in the firm j, and year-month t; a vector of binary indicators for the i-th job posting for each pairwise permutation of possible skill groups is denoted as $G_i = (a \succ_i b \mid a, b \in X \text{ and } a \neq b)$; γ_j and δ_t are firm and time fixed effects; $\varepsilon_{i,j,t}$ is an error term.

Second, using post-estimation of the model (2), let ξ be a given significance level for the obtained coefficients. For a binary indicator describing the preference relation $a \succ b$, denote the significant (at ξ-level) coefficient as $\hat{\beta}^*_{a\succ b}$, and the insignificant (at ξ-level) coefficient as $\hat{\beta}_{a\succ b}$. Denote the sign of the coefficient with sgn. Then, the strict and weak skill preferences within the wage are defined as follows.

Definition 3 (Strict preference within wage)**.** A strict preference $a \overset{w}{\succ} b$ exists if $\operatorname{sgn} \hat{\beta}^*_{a\succ b} > \operatorname{sgn} \hat{\beta}^*_{b\succ a}$.

Definition 4 (Weak preference within wage)**.** A weak preference $a \overset{w}{\succsim} b$ exists if $\operatorname{sgn} \hat{\beta}_{a\succ b} > \operatorname{sgn} \hat{\beta}_{b\succ a}$ and there exists at least $\hat{\beta}^*_{a\succ b}$ or $\hat{\beta}^*_{b\succ a}$.

Definition 5 (Wage preference indifferent to skill order / frequency)**.** A permutation indifferent preference $a \overset{w}{\sim} b$ exists if $\operatorname{sgn} \hat{\beta}^*_{a\succ b} = \operatorname{sgn} \hat{\beta}^*_{b\succ a}$.

Finally, a graph-based approach to identifying pairs of skills based on strict and weak preferences can be performed as described below. This approach relies on the definitions provided in 3 and 4. Let $H_s = (V_s, E_s)$ be a directed graph for the set of strict preferences, where $V_s = \{x \in X \mid \exists\, y \in X : x \overset{w}{\succ} y \text{ or } y \overset{w}{\succ} x\}$ is a set of vertices containing skill groups,

[1] The following analysis uses two baseline regression models based on skill preference definitions, where separate K_i vectors are calculated for $\overset{r}{\succ}_i$ and $\overset{f}{\succ}_i$ relations.

$E_s \subseteq \{(x, y) \mid (x, y) \in V_s^2 \text{ and } x \overset{w}{\succ} y\}$ is a set of edges. Let $H_w = (V_w, E_w)$ be a directed graph for the set of weak preferences, where $V_w = \{x \in X \mid \exists y \in X : x \overset{w}{\succsim} y \text{ or } y \overset{w}{\succsim} x\}$ is a set of vertices containing skill groups, $E_w \subseteq \{(x, y) \mid (x, y) \in V_w^2 \text{ and } x \overset{w}{\succsim} y\}$ is a set of edges. An illustrative example of the mapping of skill preferences to the graph representation is shown in Figure 2.

$$
\begin{array}{c}
\begin{array}{cc}
 & \begin{array}{cccc} A & B & C & D \end{array} \\
\begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} \cdot & 1 & 0 & -1 \\ -1 & \cdot & 0 & -1 \\ 1 & 0 & \cdot & 0 \\ -1 & 0 & 0 & \cdot \end{pmatrix}
\end{array}
\end{array}
\longrightarrow
\begin{tikzcd}
A \arrow[r] & B \\
C \arrow[u, dashed] & D \arrow[u, dashed]
\end{tikzcd}
$$

Figure 2: An illustration of skill preferences within wages in matrix form and its mapping to the graph representation.

Notes: An example above refers to the notation of a regression post-estimation for model (2) over 4 skill groups denoted as A, B, C, and D. The matrix cell indicates the preference relation "the skill in a row *relates to* the skill in a column". It contains the signs of significant coefficients and zeros otherwise. The graph is based on the matrix values. Strict and weak preferences within wage are represented by straight and dashed lines, respectively. The particular example shows the following set of preferences: $A \overset{w}{\succ} B$, $C \overset{w}{\succsim} A$, $D \overset{w}{\succsim} B$, and $A \overset{w}{\sim} D$.

4. Empirical results

4.1. Wage regression: Ranking versus frequency

Table 2 shows estimates from a baseline regression model (1) for ranking and frequency preferences in three datasets.[2] The estimations for the HeadHunter dataset are presented in the first two columns, followed by estimations for Reed in the next two columns, and for MyCareersFuture in the last two columns. The results reveal a common pattern of skill preferences across all datasets in terms of both ranking and frequency preferences relation. For instance, having preferences in AI, Cognitive, Character, and Management skills is associated with wage premiums, while preferences in Office skills generally result in wage penalties. Additionally, platform-specific preferences in skills such as Social, Creativity, and Writing are observed in job datasets. These skills are associated with salary penalties in HeadHunter, whereas in Reed and MyCareersFuture, they are connected to premiums or have no corre-

[2]Following the research methodology, indicator variables in regressions have the same names but calculated differently for ranking and frequency preferences.

lation with salary. Another example is Customer Service skills, which are associated with penalties in Reed, but premiums in the other two datasets.

Table 2: Wage premium for skills estimations

	Dependent variable: log (suggested salary)					
	HeadHunter		Reed		MyCareersFuture	
	Rank.	Freq.	Rank.	Freq.	Rank.	Freq.
AI	0.242***	0.338***	0.088***	0.183***	0.097***	0.033
	(0.009)	(0.028)	(0.032)	(0.040)	(0.034)	(0.050)
Social	−0.005***	−0.012***	0.037***	0.043***	0.075***	0.107***
	(0.001)	(0.001)	(0.006)	(0.007)	(0.013)	(0.016)
Cognitive	0.068***	0.091***	0.053***	0.043***	0.044***	0.019
	(0.002)	(0.005)	(0.008)	(0.010)	(0.012)	(0.015)
Character	0.031***	0.028***	0.018*	0.014	0.061***	0.127***
	(0.002)	(0.003)	(0.009)	(0.020)	(0.014)	(0.033)
Creativity	−0.028***	−0.097***	0.039**	0.003	0.026	0.009
	(0.005)	(0.023)	(0.020)	(0.034)	(0.032)	(0.061)
Writing	−0.021***	−0.081***	0.014	0.039	0.004	0.061
	(0.002)	(0.006)	(0.017)	(0.035)	(0.039)	(0.105)
Management	0.184***	0.258***	0.085***	0.156***	0.044***	0.068***
	(0.002)	(0.002)	(0.007)	(0.009)	(0.012)	(0.017)
Finance	−0.021***	0.066***	0.095***	0.152***	0.089***	0.108***
	(0.003)	(0.004)	(0.009)	(0.011)	(0.017)	(0.025)
Bus.Systems	0.049***	0.071***	0.074***	0.020	0.009	0.037
	(0.002)	(0.002)	(0.010)	(0.017)	(0.019)	(0.040)
Cust.Service	0.039***	0.192***	−0.041***	−0.041***	0.017	0.051***
	(0.002)	(0.001)	(0.006)	(0.007)	(0.012)	(0.013)
Office	−0.084***	−0.080***	−0.061***	−0.110***	−0.027	0.008
	(0.001)	(0.002)	(0.010)	(0.012)	(0.033)	(0.090)
Tech.Support	0.013***	−0.035***	0.024***	0.050***	−0.002	0.019*
	(0.003)	(0.003)	(0.006)	(0.007)	(0.010)	(0.011)
Data.Analysis	0.099***	0.147***	0.053***	0.009	−0.001	0.014
	(0.011)	(0.027)	(0.017)	(0.029)	(0.025)	(0.040)
Spec.Software	0.072***	0.219***	−0.017***	0.065***	0.034***	0.053***
	(0.002)	(0.003)	(0.006)	(0.005)	(0.010)	(0.009)
Observations	988,187	988,187	30,940	30,940	13,161	13,161
Adjusted R^2	0.508	0.522	0.436	0.447	0.598	0.599

Notes: Year-month and firm fixed effects are included in all specifications. Models are conducted over three datasets: HeadHunter, Reed, and MyCareersFuture. For each dataset two approaches for skill preferences identification are used: Rank. — the order of appearance of a skill in a job posting (the skill that is unambigously mentioned before the other skill), Freq. — the relative frequency of a skill in a job posting (the most unambigously frequent skill is identified). Standard errors in parentheses. Significance levels are denoted as: *p<0.1; **p<0.05; ***p<0.01.

A notable aspect of this research, however, is the different estimates produced by the two methods of defining preference relationships. In particular, these are preferences for Finance and Technical Support skills in HeadHunter dataset, and preferences for Specialized Software in Reed dataset. This phenomenon can be described as follows: If the skill is ranked negatively and its frequency positively, the market is willing to pay more for tasks that involve that skill (even if the skill group is not specific to a particular occupation). On the other hand, if the skill is positively ranked, the market values that skill within a specific occupation (and the skill's prevalence in job postings may be related to prior job tasks). To ensure the robustness of the results and validate their interpretation, an occupational decomposition has been included.

Wage regression results within occupations are reported in Appendix B. Table 5 performs results for HeadHunter, Table 6 — for Reed. The study reveals certain skill preferences across various occupations. Taking into account the HeadHunter job board, it is evident that the occupational sample shows different skill preferences than those estimated based on the entire dataset. For instance, ranking preferences alter the value of certain skills in IT occupations (Cognitive, Character, Customer Service, and Technical Support skills have a negative sign) and Sales occupations (AI skills have a negative sign, while Social skills have a positive sign). Both ranking and frequency preferences are altered in the Healthcare and Services occupations. Creativity and Specialized Software have negative signs while Office skills have positive signs in Healthcare. Meanwhile, Technical Support skills have positive signs in Services. Additionally, it can be assumed that Customer Service skills in Healthcare and IT and Technical Support skills in Sales are not job-specific, as their rank is negative while frequency is positive.

There are distinct patterns in occupational skill preferences for the Reed dataset. In Service occupations, there is a negative ranking preference for Writing skills, but a positive preference for Business Systems skills in terms of frequency. In Sales occupations, Management skills are insignificant in terms of wage, whereas Technical Support skills can be considered as non-specific skills for the occupation. Moreover, the frequency preferences of Social skills for Healthcare occupations is negative, while the frequency of Writing skills in IT is positive. This difference is also attributed to Technical Support preferences in IT, which have a negative impact on both types of preferences.

Therefore, the shared skill preference relations within occupations between the two datasets are predominantly related to IT occupations, specifically Management, Office, and Technical Support skills. Moreover, the preference relations for Management skills are similar in Healthcare and Services occupations. Lastly, Technical Support skills have a nonspecific role in Sales occupations.

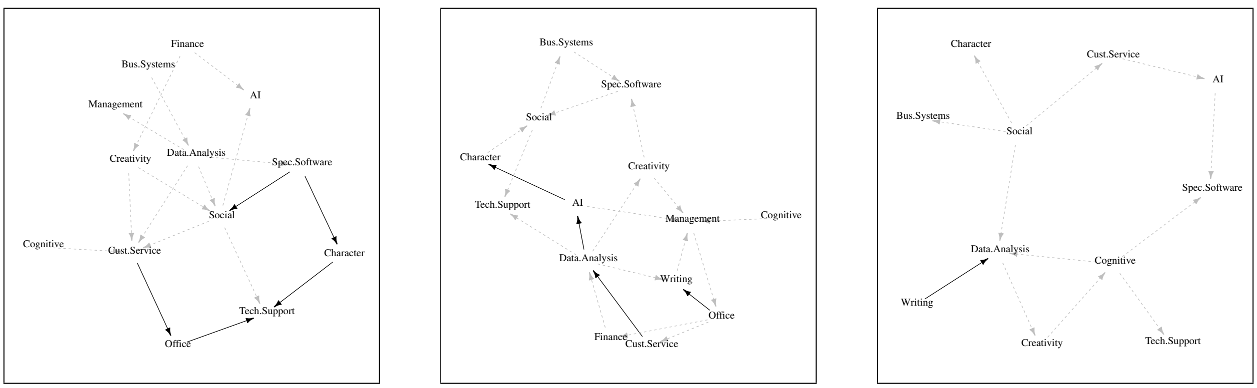
Interestingly, the dynamics of skill preferences can also be estimated using the provided methodology. Preliminary estimates for the HeadHunter dataset, covering the entire job board and IT occupations, are presented in Appendix C. However, the estimation results should be interpreted with caution. Based on the results of the baseline regression model estimations, it is possible to determine the order of coefficient estimations for skill group indicators in relation to their impact on wages. As a result, skills can be ranked both in terms of their absolute impact on wages (based on regression coefficients) and their relative ranking. It is important to interpret the latter proposition with precision, taking into consideration the significance, sign, and confidence intervals of each coefficient.

For instance, consider the rankings of all HeadHunter occupations (shown in Figure 4a). The preferences for Social and Finance skills underwent changes between 2016 and 2021, as evidenced by the signs of the coefficients. Prioritizing Social skills in a job vacancy led to wage premiums in 2016, but in 2021, this can result in wage penalties due to the reversed dynamics of Finance skills. Additionally, some skills, such as AI, Management, and Cognitive skills, bring wage premiums during the entire period if they are ranked higher than at least one other skill in the vacancy. When comparing skills in terms of suggested salary, it is important to consider confidence intervals of the estimated coefficients. In 2016, both AI and Management skills were valued higher than other skills, however, it is difficult to determine a clear ranking between these two skills. As of 2021, the same uncertainty remains. AI, Management, Cognitive, and Finance skills have resulted in wage premiums, while Social and Office skills have brought wage penalties and can both be ranked below the previously mentioned skill groups. This logic can be applied to specific occupations. Figure 5a illustrates the ranking of skill preferences for IT occupations. Interestingly, Social and Cognitive skills had no significant impact on wages and both resulted in wage penalties until the start of 2021.

Considering the frequency preferences of all HeadHunter occupations (Figure 4b) or solely IT occupations (Figure 5b), the ranking of skills may vary. The impact of the skill on wage during the investigated period may also change based on the preference relation for the indicator variable, resulting in its more frequent appearance in a job ad compared to other skills. Thus, it can be observed that Finance and Social skills may alter their impacts on wage. Furthermore, it can be observed that there is a more consistent ranking of skills, with Finance skills being prioritized over Social skills from 2016 to 2021. However, the frequency and importance of AI skills are not clearly apparent, particularly at the beginning of the examined time period.

4.2. Graph-based representation results

Following the presented baseline framework that provides a little information about pairwise skill preferences, a graph-based approach was implemented to address this issue. Figure 3 shows the results of the implemented methodology for model (2) on three datasets, with the ξ-level fixed at 0.1 (robustness checks are presented in Appendix D). To interpret the results, let us observe a particular example. Considering the HeadHunter dataset, the strict relationship between "Specialized Software" and "Social" skills could be technically described as follows: we can expect an increase in the suggested salary in a vacancy if "Specialized Software" skills are mentioned earlier than "Social" skills. However, the results should be interpreted with caution in order to capture both evidence and robustness insights. Accordingly, in order to simplify the interpretation, the term "prefers to" will be used in the following to describe both the strict and the weak preference relationship within the skill order and the salary increase.



(a) HeadHunter (b) Reed (c) MyCareersFuture

Figure 3: Graphs of skill preferences in the baseline model (2) over three job board datasets.

Notes: The main indicator variables are based on ranking preferences (the graphs for frequency preferences are not substantially different from depicted in the figure).

Taking into account the advanced computer skills such as "AI", "Data Analysis", and "Specialized Software", we cannot make a definite statement about their superiority over the other skills. For example, both "AI" and "Data Analysis" are preferable to "Management" and inferior to "Finance" skills. Moreover, some interesting insights could be gained regarding the relationship with non-technical skills that are widely discussed in the academic literature (e.g. Deming and Kahn, 2018; Deming and Noray, 2020). In general, "Social" and "Character" skills are inferior to "Specialized Software" (except for IT-related occupations), but both might be preferred to "AI". In addition, "Data Analysis" is preferred over "Creativity". However, this is not the case for "Writing" skills. Accordingly, the MyCareersFuture data shows that "Writing" skills are superior to "Data Analysis" except for IT occupations, but the Reed data indicates the opposite relationship. Another interesting effect relates to the superiority of “Office Software” over “Writing” that is captured in two data sources, but the effect is reversed when controlling for specification within occupations. The latter findings concern the attribution to “Customer Service” and “Technical Support”. For IT occupations, “Customer Service” is inferior to “Specialized Software” skills in all databases. However, “Customer Service” prefers “Data Analysis” and inferior to “Office Software” in Reed, and the relationships are reversed in HeadHunter. Interestingly, “Technical Support” is inferior to “Social” and “Character” skills (except for IT occupations), and also inferior to “Office Software” (HeadHunter) and “Data Analysis” (Reed).

5. Conclusion

This paper proposes a methodology for formalizing and identifying skill preferences within job postings in relation to suggested salaries. Skill preferences were found to be prevalent and to have a direct impact on suggested salaries through analysis of various job board datasets. The research provides unique insights into studying skill preferences in posting. Depending on job board data origin, occupation, and time period, skill ranking and frequencies may offer complementary information. Therefore, the subordination of skills in job ads cannot be ignored.

The findings present new insights into the preferences of technical versus non-technical skills. Specifically, advanced computer skills are typically favored over other job-specific skills, including management and service skills. Non-technical skills are more highly valued in IT occupations than technical skills. However, there may be slight variations in the effects across different job boards. To address these issues in future research, enriching the data coverage is possible.

The methodology and results presented in this paper could also provide valuable insights for human resource professionals and educational institutions in monitoring labor market demands and achieving better workforce matching. For example, the alignment of educational curriculum to meet the needs of the labor market could provide a more targeted means of communication between companies and educational institutions. Additionally, the methodology presented has significant potential to be incorporated into existing occupation and skill classification systems, such as ESCO, ISCO, and O*NET. Accordingly, policymakers can broaden labor market interventions in specific countries, industries, and occupations by ranking skills based on their relative importance to employers.

References

Acemoglu, D. and D. Autor (2011) "Skills, tasks and technologies: Implications for employment and earnings" in *Handbook of Labor Economics* **4**, Elsevier, 1043–1171.

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022) "Artificial intelligence and jobs: Evidence from online vacancies" *Journal of Labor Economics* **40(S1)**, S293–S340.

Alekseeva, L., J. Azar, M. Gine, S. Samila, and B. Taska (2021) "The demand for AI skills in the labor market" *Labour Economics* **71**, 102002.

Ao, Z., G. Horváth, C. Sheng, Y. Song, and Y. Sun (2023) "Skill requirements in job advertisements: A comparison of skill-categorization methods based on wage regressions" *Information Processing & Management* **60(2)**, 103185.

Autor, D. H., F. Levy, and R. J. Murnane (2003) "The skill content of recent technological change: An empirical exploration" *The Quarterly Journal of Economics* **118(4)**, 1279–1333.

Banfi, S. and B. Villena-Roldán (2019) "Do high-wage jobs attract more applicants? Directed search evidence from the online labor market" *Journal of Labor Economics* **37(3)**, 715–746.

Bhola, A., K. Halder, A. Prasad, and M.-Y. Kan (2020, December). "Retrieving skills from job descriptions: A language model based extreme multi-label classification framework" in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online), 5832–5842.

Deming, D. and L. B. Kahn (2018) "Skill requirements across firms and labor markets: Evidence from job postings for professionals" *Journal of Labor Economics* **36(S1)**, S337–S369.

Deming, D. and K. Noray (2020) "Earnings dynamics, changing job skills, and STEM careers" *The Quarterly Journal of Economics* **135(4)**, 1965–2005.

Napierala, J. and V. Kvetan (2023) "Changing job skills in a changing world" in *Handbook of Computational Social Science for Policy*, Springer International Publishing Cham, 243–259.

Zhang, M., K. N. Jensen, S. Sonniks, and B. Plank (2022, July) "SkillSpan: Hard and soft skill extraction from English job postings" In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, United States, Association for Computational Linguistics, 4962–4984.

Appendices

A. Descriptive statistics

Table 3: Percentage of jobs posted by skill group (in %).

Skill group	HeadHunter	Reed	MyCareersFuture
AI	0.5	1.3	2.4
Social	44.5	42.9	26.5
Cognitive	9.0	18.7	26.4
Character	18.6	12.0	12.9
Creativity	1.6	2.6	3.0
Writing	10.0	2.9	2.1
Management	16.3	23.6	23.1
Finance	5.6	14.3	13.4
Business Systems	25.3	10.1	6.2
Customer Service	32.6	37.3	39.9
Office Software	29.9	13.3	2.2
Technical Support	8.8	45.6	50.5
Data Analysis	0.3	2.7	4.1
Specialized Software	11.6	68.9	71.1

Table 4: Frequency of unique skill groups in job ads (in %) across job boards.

Number of unique skill groups	HeadHunter	Reed	MyCareersFuture
0	10.5	10.4	3.4
1	25.2	16.0	20.2
2	29.9	18.5	22.6
3	18.5	18.4	22.1
4	9.2	14.8	16.2
5	4.1	10.6	8.7
6	1.7	6.3	4.7
7+	0.9	5.0	2.2

B. Wage premium and skills per occupation

Table 5: Wage premium for skills estimations across occupations: HeadHunter data

	Dependent variable: log (suggested salary)							
	IT		Healthcare		Sales		Services	
	Order	Freq.	Order	Freq.	Order	Freq.	Order	Freq.
AI	0.164***	0.241***	0.405***	0.308	−0.204***	−0.421	0.493***	−0.248
	(0.010)	(0.030)	(0.139)	(0.369)	(0.073)	(0.376)	(0.073)	(0.395)
Social	−0.067***	−0.114***	−0.016***	−0.071***	0.013***	−0.00001	−0.034***	−0.047***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.002)	(0.002)	(0.004)	(0.004)
Cognitive	−0.034***	−0.004	0.097***	0.078***	0.117***	0.082***	0.048***	0.205***
	(0.007)	(0.011)	(0.012)	(0.026)	(0.004)	(0.014)	(0.010)	(0.024)
Character	−0.020***	0.014	−0.005	0.058***	0.027***	0.002	0.051***	0.044***
	(0.007)	(0.012)	(0.008)	(0.014)	(0.003)	(0.004)	(0.005)	(0.008)
Creativity	−0.051***	−0.004	0.316***	0.487*	−0.055***	−0.077*	0.011	−0.100*
	(0.017)	(0.084)	(0.016)	(0.277)	(0.010)	(0.046)	(0.015)	(0.060)
Writing	−0.121***	−0.256***	−0.002	−0.002	−0.011***	−0.057***	−0.004	−0.062**
	(0.007)	(0.015)	(0.014)	(0.050)	(0.003)	(0.017)	(0.008)	(0.029)
Management	0.178***	0.238***	0.094***	0.250***	0.175***	0.292***	0.187***	0.280***
	(0.006)	(0.007)	(0.009)	(0.014)	(0.003)	(0.003)	(0.005)	(0.006)
Finance	0.075***	0.061**	0.165***	0.105	0.011	0.138***	0.199***	0.156***
	(0.016)	(0.026)	(0.038)	(0.070)	(0.009)	(0.019)	(0.017)	(0.027)
Bus.Systems	0.067***	0.133***	−0.011	0.077***	0.042***	0.057***	0.109***	0.163***
	(0.005)	(0.006)	(0.009)	(0.015)	(0.003)	(0.004)	(0.007)	(0.011)
Cust.Service	−0.071***	0.089***	−0.038***	0.061***	0.006***	0.144***	0.015***	0.110***
	(0.006)	(0.005)	(0.007)	(0.008)	(0.002)	(0.002)	(0.005)	(0.005)
Office	−0.137***	−0.160***	0.048***	0.063***	−0.099***	−0.070***	−0.081***	−0.079***
	(0.005)	(0.008)	(0.005)	(0.011)	(0.002)	(0.005)	(0.005)	(0.010)
Tech.Support	−0.020***	−0.072***	0.009	0.050	−0.028***	0.048***	0.022**	0.088***
	(0.005)	(0.005)	(0.025)	(0.045)	(0.010)	(0.016)	(0.011)	(0.026)
Data.Analysis	0.084***	0.109***	0.068	−0.227	0.190***	0.342***	0.076*	0.033
	(0.021)	(0.037)	(0.044)	(0.396)	(0.027)	(0.101)	(0.045)	(0.076)
Spec.Software	0.026***	0.173***	−0.106***	−0.295***	0.043***	0.004	0.116***	0.083***
	(0.005)	(0.004)	(0.013)	(0.019)	(0.010)	(0.018)	(0.013)	(0.024)
Observations	109,733	109,733	52,530	52,530	296,124	296,124	83,881	83,881
Adjusted R^2	0.558	0.567	0.610	0.611	0.593	0.607	0.628	0.633

Notes: Year-month and firm fixed effects are included in all specifications. Models are conducted across four occupations (each vacancy may attribute to several occupations following the HeadHunter's classification): IT, Healthcare, Sales, Services. For each subsample two approaches for skill preferences identification are used: Rank. — the order of appearance of a skill in a job posting, Freq. — the relative frequency of a skill in a job posting. Standard errors in parentheses. Significance levels are denoted as: *p<0.1; **p<0.05; ***p<0.01.

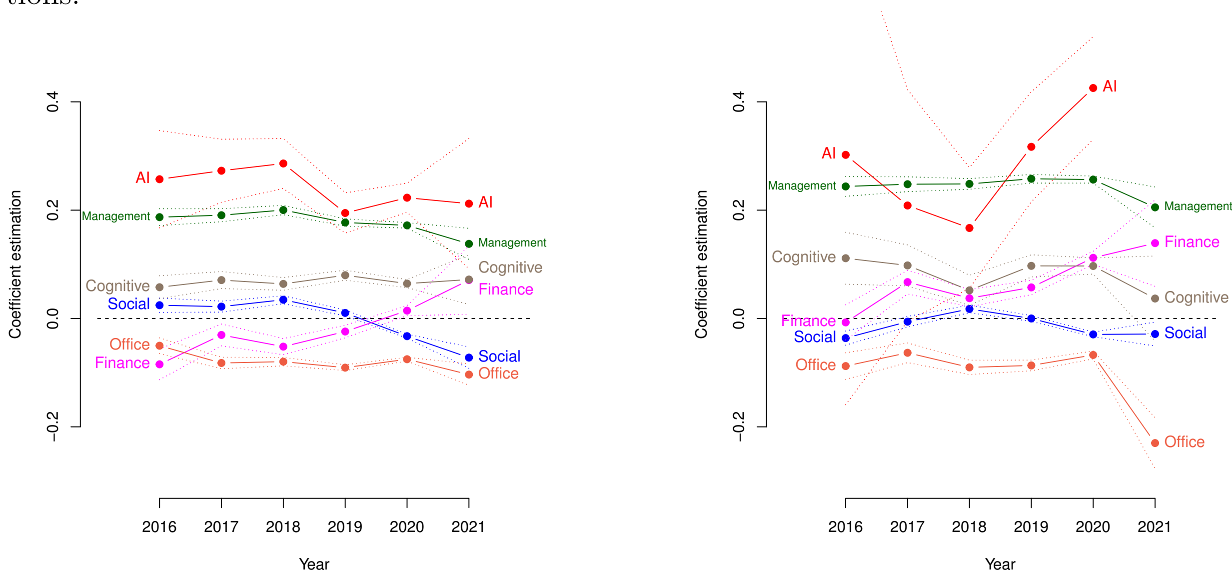
Table 6: Wage premium for skills estimations across occupations: Reed data

	Dependent variable: log (suggested salary)							
	IT		Healthcare		Sales		Services	
	Order	Freq.	Order	Freq.	Order	Freq.	Order	Freq.
AI	−0.053	0.047	0.184					
	(0.074)	(0.079)	(0.344)					
Social	0.025	−0.012	−0.005	-0.135^{***}	0.034	0.053	0.048^{**}	0.025
	(0.037)	(0.045)	(0.034)	(0.037)	(0.030)	(0.037)	(0.024)	(0.026)
Cognitive	−0.049	−0.029	0.097	0.043	0.010	−0.005	0.030	0.027
	(0.042)	(0.052)	(0.066)	(0.100)	(0.049)	(0.073)	(0.031)	(0.051)
Character	−0.028	−0.116	−0.016	−0.060	−0.034	0.299^{***}	0.016	0.090
	(0.065)	(0.148)	(0.058)	(0.145)	(0.048)	(0.109)	(0.033)	(0.089)
Creativity	−0.224		0.199		0.008	0.200	−0.041	
	(0.142)		(0.209)		(0.082)	(0.272)	(0.216)	
Writing	−0.066	0.350^{*}	−0.087		−0.182	−0.215	-0.199^{***}	
	(0.084)	(0.180)	(0.157)		(0.178)	(0.302)	(0.073)	
Management	0.094^{**}	0.164^{***}	0.156^{***}	0.395^{***}	0.025	0.044	0.138^{***}	0.173^{***}
	(0.044)	(0.052)	(0.051)	(0.062)	(0.037)	(0.055)	(0.034)	(0.048)
Finance	−0.022	0.012	0.048	−0.039	0.073	0.041	0.003	0.174
	(0.067)	(0.095)	(0.138)	(0.187)	(0.072)	(0.172)	(0.059)	(0.106)
Bus.Systems	0.065	0.054	−0.110	−0.043	−0.067	−0.034	0.016	0.144^{*}
	(0.055)	(0.078)	(0.069)	(0.290)	(0.045)	(0.066)	(0.043)	(0.080)
Cust.Service	-0.167^{***}	-0.147^{***}	0.013	−0.039	0.024	−0.016	0.008	0.003
	(0.042)	(0.056)	(0.042)	(0.069)	(0.026)	(0.026)	(0.021)	(0.020)
Office	-0.318^{***}	-0.354^{***}	−0.039	−0.118	−0.060	−0.036	−0.046	−0.042
	(0.048)	(0.064)	(0.089)	(0.215)	(0.053)	(0.075)	(0.034)	(0.040)
Tech.Support	-0.060^{*}	-0.152^{***}	−0.037	−0.013	-0.065^{**}	0.064^{**}	0.054^{**}	0.029
	(0.035)	(0.032)	(0.033)	(0.049)	(0.028)	(0.031)	(0.022)	(0.027)
Data.Analysis	0.062	0.003	0.103	0.675	0.349		0.103^{*}	
	(0.082)	(0.110)	(0.139)	(0.442)	(0.220)		(0.062)	
Spec.Software	−0.017	0.087^{**}	-0.077^{**}	−0.014	0.019	0.059^{**}	-0.049^{**}	0.045^{**}
	(0.037)	(0.035)	(0.030)	(0.029)	(0.026)	(0.028)	(0.022)	(0.021)
Observations	1,235	1,235	1,099	1,099	1,170	1,170	952	952
Adjusted R^2	0.370	0.373	0.306	0.331	0.627	0.632	0.505	0.510

Notes: Year-month and firm fixed effects are included in all specifications. Models are conducted across four occupations (each vacancy may attribute to one occupation only following the Reed's classification): IT, Healthcare, Sales, Services. For each subsample two approaches for skill preferences identification are used: Rank. — the order of appearance of a skill in a job posting, Freq. — the relative frequency of a skill in a job posting. Several specifications have blank coefficient estimates due to not detected preference relation for a particular skill. Standard errors in parentheses. Significance levels are denoted as: *p<0.1; **p<0.05; ***p<0.01.

C. Skill preferences dynamics

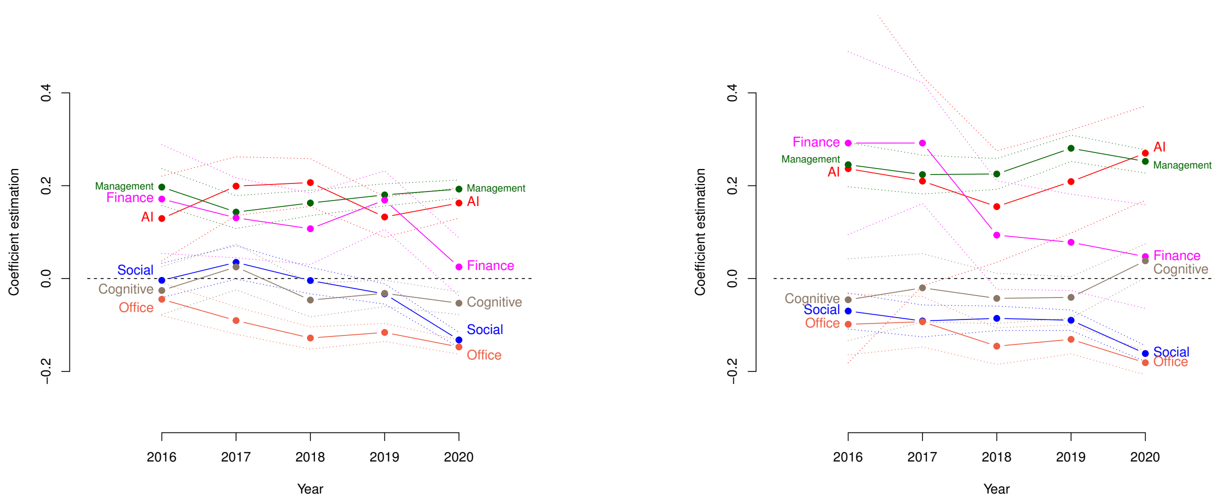
The figures below show obtained estimations (with 95% confidence intervals) from the baseline regression model (1) in dynamics for several skill groups in HeadHunter dataset. Both preference in ranking and in frequency were estimated separately. Each year estimates correspond to a separate regression model conducted across the subsample of particular year. Figure 4 shows the skill dynamics across the whole dataset; Figure 5 — across IT occupations.



(a) Skill preferences in ranking

(b) Skill preferences in frequency

Figure 4: Dynamics of skill preferences in HeadHunter dataset.

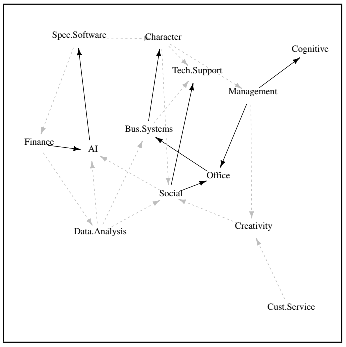


(a) Skill preferences in ranking

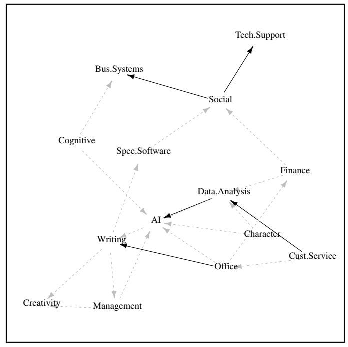
(b) Skill preferences in frequency

Figure 5: Dynamics of skill preferences across IT occupations in HeadHunter dataset.

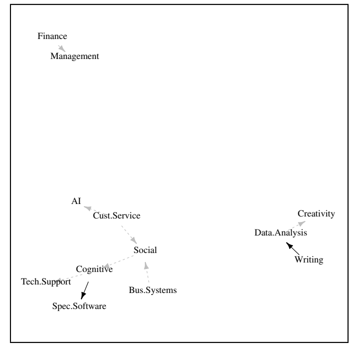
D. Graph-based representation: Robustness checks



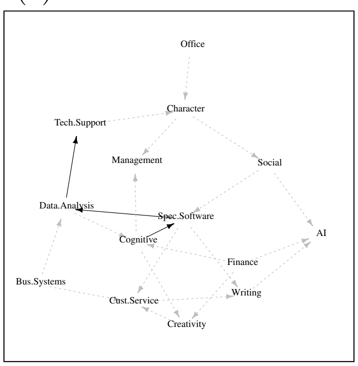
(a) HeadHunter with skills



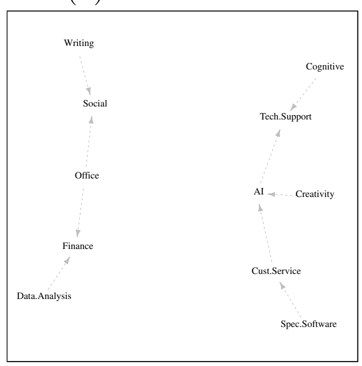
(b) Reed with skills



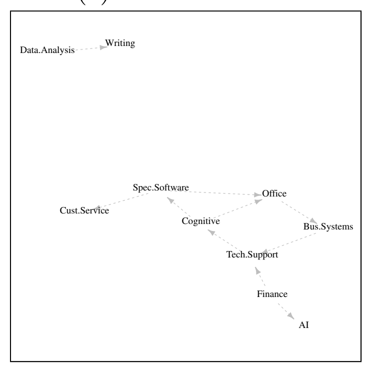
(c) MCF with skills



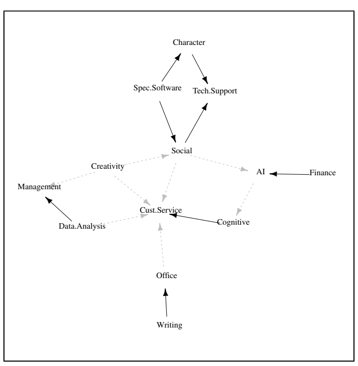
(d) HeadHunter: IT vacancies



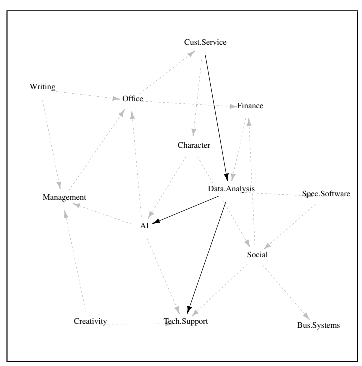
(e) Reed: IT vacancies



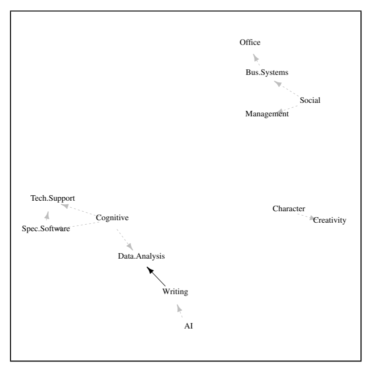
(f) MCF: IT vacancies



(g) HeadHunter with controls



(h) Reed with controls



(i) MCF with controls

Figure 6: Graphs of skill preferences within different model specifications across job boards.

Notes: The grid of graphs is organized by dataset in columns (first column — HeadHunter, second column — Reed, third column — MyCareersFuture (MCF)). The rows indicate different specifications of model (2) as follows. The first row covers model specifications with the inclusion of skill binaries (14 binary variables indicating the presence of a particular skill group in a job posting) in the baseline model. The second row provides baseline specifications for the subsample of vacanies related to information technology (IT) occupations (the share of IT-related jobs is 8% from the HeadHunter dataset, 4% — Reed, 26% — MyCareersFuture). The third row shows specifications with additional controls for a baseline model: experience and occupation for HeadHunter; experience for Reed; occupation for MyCareersFuture. The main indicator variables are based on ranking preferences (the graphs for frequency preferences are not substantially different from depicted in the figure).