

Jak wnioskować o zapotrzebowaniu na kwalifikacje na podstawie ogłoszeń o pracy?

Wnioski z dotychczasowych działań

Robert Pater¹

Maciej Beręsewicz² i Herman Cherniaiev¹

¹Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie
Instytut Badań Edukacyjnych w Warszawie

²Uniwersytet Ekonomiczny w Poznaniu
Główny Urząd Statystyczny

Seminarium naukowo-informacyjne pt.:

“Edukacja, język i technologia – Zintegrowany Rejestr Kwalifikacji w perspektywie interdyscyplinarnej”

Plan prezentacji

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Postępy w badaniach rynku pracy

Metody wnioskowania o popycie na pracę z ofert pracy

Wybrane wyniki

Wnioski

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Postępy w badaniach rynku pracy

Badania reprezentacyjne (z prób losowych)

- Badanie popytu (JVS/Z05) bardziej problematyczne niż badanie podaży (LFS/BAEL)
- Coraz bardziej obciążone brakiem odpowiedzi lub niepełnym raportowaniem (*non-response, under-reporting errors*). Ważenie, jego korekta o brak odpowiedzi i kalibracja dają coraz mniej wiarygodne rezultaty.
- Główne zmienne: sekcje NACE/PKD i region NUTS2. Nie ma możliwości wprowadzenia nowych zmiennych i większej szczegółowości

Dane administracyjne (z prób nielosowych)

- Czasem obowiązek zgłoszeniowy (ZUS), czasem "powinność" (PSZ/CBOP)
- 10-20% ogółu wakatów trafia do Centralnej Bazy Ofert Pracy (CBOP), a więc ofert pracy z zasobów PSZ (Gradzewicz et al., 2016)
- Wiele zmiennych, informacja ustrukturyzowana
- Błąd pokrycia (*under-coverage error*)

Źródła (systematycznych) danych

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Big data i Internet jako źródło danych stały się ważnym zagadnieniem w statystyce, zwłaszcza w statystyce publicznej. Istnieje wiele międzynarodowych inicjatyw w tym zakresie, np. European statistical system ESSnet on Big Data; American Association for Public Opinion Research, AAPOR Task Force on Big Data; European Centre for the Development of Vocational Training, Cedefop; the Maxwell project.

Dane z sieci (z prób nielosowych)

- Uzupełniające do danych administracyjnych
- Gromadzenie wymaga ustawicznego monitorowania, ale jest relatywnie tanie
- Wiele źródeł, trudno ustalić populację ofert pracy
- Ogrom danych (stąd duże znaczenie ekonomiczne) ale nieustrukturyzowanych
- Wiele potencjalnych błędów

Badania zapotrzebowania na pracę

Badania popytu na pracę obejmują popyt zrealizowany (obsadzone miejsca pracy) i niezrealizowany (wakaty). Badania podaży pracy obejmują pracujących i zatrudnionych

- W Polsce: *Popyt na pracę* – Z05 (JVS) i *Zapotrzebowanie rynku pracy na pracowników według zawodów* – Z15
- Badanie nie obejmuje umów cywilno-prawnych (w przeciwieństwie do BAEL)
- Statystyka publiczna jest zobligowana do przekazywania do Eurostat danych o obsadzonych i wolnych miejscach pracy według sekcji PKD
- Kompletność badania *Popyt na pracę*: około 60% i spada; obniżona próba
- Badanie Z05 zaniża liczbę podmiotów z co najmniej jednym wakatem o około 10-15% (Beręsewicz, Cherniaiev and Pater, 2021). Głównymi przyczynami tej rozbieżności są błędy niezwiązane z próbkowaniem (*sampling error*) a brak odpowiedzi (*non-response error*) i niepełne raportowanie (*under-reporting error*), które można zidentyfikować, porównując dane z badania Z05 do danych administracyjnych.

Jak badać (popyt na) kwalifikacje?

- Najbliższa zmienna w statystyce oficjalnej to *zawód*, ale to nie główna zmienna dla statystyki oficjalnej
- Z punktu widzenia szkolnictwa branżowego *zawód* to *kwalifikacja pełna*. Czy pozostałe zawody też można tak traktować?
- Klasyfikacje zawodów: ISCO, ESCO, O*NET, KZiS i KZSB
- Szczegółowość Klasyfikacji zawodów i specjalności (KZiS) oraz Klasyfikacji zawodów szkolnictwa branżowego (KZSB)
- Czy jest możliwość badania kwalifikacji cząstkowych, uregulowanych i rynkowych?
- Zadania zawodowe i umiejętności (kompetencje), a także wiele innych informacji o stanowisku pracy
- Język Instytucji rynku pracy – Edukacji – Przedsiębiorstw

Minusy

- Jedna z metod poszukiwań pracowników (*measurement, under-coverage errors*)
- Nie można zbadać wszystkich portali w Internecie (*under-coverage, selection errors*)
- Niestabilność części źródeł danych (*selection error*, zob. badania Cedefop, Colombo et al. (2018))
- Jakie ogłoszenia można uznać za oferty pracy – staże/praktyki, oferty pracy za granicą, umowy cywilno-prawne, umowy typu B2B? (*over-coverage error*)

Plusy

- Rozwój portali dedykowanych sektorowi
- Szczegółowość danych jest czasem bardzo duża, ale zróżnicowana w zależności od portalu i poziomu stanowiska
- Analiza historyczna i z wysoką częstotliwością (tygodniową)
- Reprezentacja nowych podmiotów (w przeciwieństwie do innych źródeł)

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Metody wnioskowania o popycie na pracę z ofert pracy

Aby oszacować liczbę wolnych miejsc pracy kluczowy jest dobór odpowiednich źródeł, które charakteryzują się wysoką jakością danych oraz umożliwiają łączenie danych między źródłami. W związku z tym dobór źródeł jest związany z następującymi elementami

- weryfikacją ogłoszeń,
- występowaniem identyfikatorów ogłoszeń oraz nazw pracodawców,
- występowaniem identyfikatorów podmiotów takich jak NIP oraz REGON,
- możliwością ustalenia kodu zawodu,
- możliwością ustalenia faktycznego miejsca pracy.

Etapy przed analizą

- przygotowanie danych, czyszczenie itp.,
- łączenie danych po identyfikatorach: deterministyczne lub probabilistyczne,
- ew. imputacja danych nt. liczby wakatów w Internecie na podstawie CBOP,
- uwzględnienie ujemnej lub dodatniej korelacji pomiędzy źródłami danych.

Główne kierunki dotychczasowych prac

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

- Przygotowanie, opracowanie i dopracowywanie odpowiednich metod uczenia maszynowego
- Integracja danych ze źródeł nielosowych (administracyjnych i online) z tymi ze źródeł losowych (badań reprezentacyjnych)
- Szacowanie populacji ofert pracy przy użyciu źródeł nielosowych

Metoda integracji danych oparta na modelu regresji LASSO

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Celem integracji danych jest oszacowanie pewnej wielkości (np. średniej, sumy) zmiennych docelowych, które występują tylko w danych online. Tutaj zajęliśmy się oszacowaniem zapotrzebowania na umiejętności (istniejących tylko w danych online) wg klasyfikacji BKL i integracją tych danych z tymi z badania reprezentacyjnego pn. *Popyt na pracę*.

Wykorzystane zmienne pomocnicze (z danych reprezentacyjnych): duże grupy zawodowe, sekcja PKD i województwo (Beręsewicz, Białkowska, Marcinkowski, Maślak, Opiela, Katarzyna and Pater, 2021).

Zastosowanie estymatora Least Angle Shrinkage and Selection Operator (LASSO) – metody regularyzacji. Adaptacyjna metoda LASSO została zaproponowana przez Zou (2006) i dla k zmiennych docelowych polega na znalezieniu współczynników

$$\hat{\beta}_k = \underset{k}{\operatorname{argmin}} \left(\sum_{i=1}^{n^A} [-y_{ki}(k_i k) + \log(1 + \exp(T_{ki} k))] + \lambda_{n^A k} \sum_{j=1}^p \alpha_{kj}^{\gamma_k} |\beta_{kj}| \right), \quad (1)$$

gdzie $\alpha_{kj}^{\gamma_k}$ jest wagą do ustalenia, a γ_{n^A} jest "karą" używaną do optymalizacji miary dopasowania modelu. Mając $\lambda_{n^A k}$ i γ_k można oszacować \hat{k} w procedurze iteracyjnej. Jeżeli $\alpha_{kj}^{\gamma_k} = 1$ otrzymujemy standardowy model LASSO. Parametr γ_k jest stałą większą niż 0 wchodzącym w interakcje z kj i kontroluje prawdopodobieństwo odrzucenia parametrów.

Metoda capture-recapture

Posiadamy tablicę kontyngencji Table 1. Dla dwóch źródeł danych A i B, możliwa jest sytuacja, w której mamy dane dla źródła A i nie mamy dla źródła B, co oznaczamy jako n_{10} lub mamy dla źródła B, a nie dla źródła A oznaczone jako n_{01} lub dane pojawiają się w źródle A i B, co oznaczamy jako n_{11} . Celem analizy jest zatem oszacowanie liczby n_{00} , tj. liczby jednostek nieuwzględnionych w żadnym ze źródeł. Ostateczną szacowaną wielkość populacji uzyskuje się przez dodanie wszystkich wartości z Table 1 po uprzednim oszacowaniu wielkości n_{00} .

Table 1: Przykład dwóch źródeł danych – tablica kontyngencji 2×2

		Lista 2		Σ
		Yes (1)	No (0)	
Lista 1	Yes (1)	n_{11} (p_{11})	n_{10} (p_{10})	$n_{1.}$ ($p_{1.}$)
	No (0)	n_{01} (p_{01})	n_{00} (p_{00})	$n_{0.}$ ($p_{0.}$)
Σ		$n_{.1}$ ($p_{.1}$)	$n_{.0}$ ($p_{.0}$)	n ($p_{.}$)

Metoda capture-recapture c.d.

W celu oszacowania całkowitej wielkości populacji możemy skorzystać z estymatora Lincolna-Petersena (LP)

$$\hat{N}_{LP} = \frac{n_{10}n_{01}}{n_{11}}. \quad (2)$$

Ale warunek braku korelacji pomiędzy źródłami danych zostanie naruszony gdy wystąpi (i) zależność dodatnia – istnieje większe prawdopodobieństwo zaobserwowania jednostek w drugim źródle/czasie; oraz (ii) zależność ujemna – prawdopodobieństwo zaobserwowania jednostek w drugim źródle/czasie jest mniejsze. Na podstawie pracy Nour (1982) zaproponowano estymatory

$$\hat{N}_L = n_{11} + n_{10} + n_{01} + 2 \frac{n_{11}n_{10}n_{01}}{n_{11}^2 + n_{10}n_{01}} \quad (3)$$

$$\hat{N}_M = n_{11} + n_{10} + n_{01} + \frac{1}{2} \sqrt{n_{10}n_{01}} + \frac{n_{11}n_{10}n_{01}}{n_{11}^2 + n_{10}n_{01}} \quad (4)$$

$$\hat{N}_U = n_{11} + n_{10} + n_{01} + \sqrt{n_{10}n_{01}} \quad (5)$$

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Wybrane wyniki

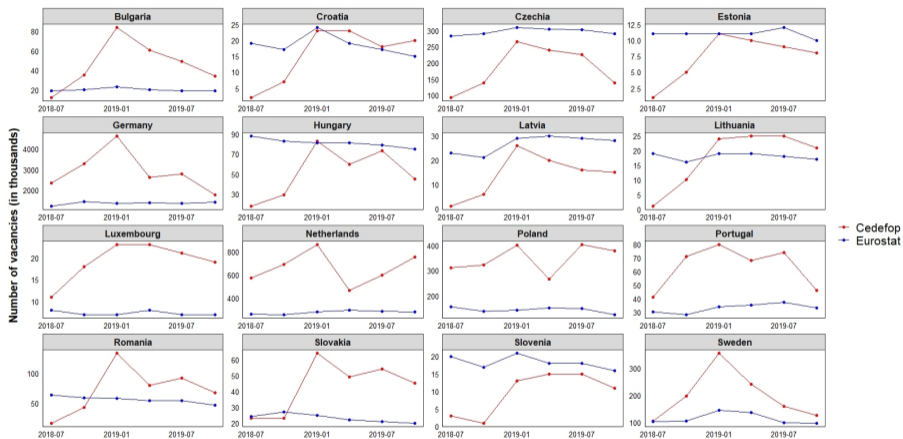


Fig. 1: Liczba ofert pracy z Cedefop (OJA) i wakatów z Eurostat (JVS)

Korelacje zmiennej docelowej ze zmiennymi pomocniczymi

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Table 2: Współczynnik V Cramera pomiędzy umiejętnościami a średnią grupą zawodową, sekcją PKD i województwem

Skill	Occupation (2 digits)	NACE	Region
Artistic	0.22	0.11	0.05
Availability	0.15	0.14	0.05
Cognitive	0.21	0.06	0.06
Computer	0.45	0.23	0.10
Interpersonal	0.42	0.23	0.06
Managerial	0.34	0.15	0.04
Mathematical	0.05	0.02	0.03
Office	0.11	0.06	0.03
Physical	0.17	0.09	0.04
Self-organization	0.34	0.19	0.04
Technical	0.31	0.11	0.07

Oszacowany udział umiejętności w wakatach w Polsce

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Table 3: Udział umiejętności w ofertach pracy i (oszacowany) udział w wakatach w Polsce

SKILLS	JO	LASSO
Artistic	15.8	12.5
Availability	20.9	19.5
Cognitive	20.9	14.6
Computer	33.0	22.6
Interpersonal	53.8	34.9
Managerial	26.2	16.8
Mathematical	0.4	0.4
Office	3.9	3.2
Physical	5.4	7.6
Self-organization	58.6	43.8
Technical	4.3	7.7

Liczba podmiotów gospodarczych według źródła i rozmiaru

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Table 4: Liczba podmiotów gospodarczych wg wielkości i źródła pochodzenia ofert pracy na koniec kwartałów 2018 roku

Size	CBOP	Pracuj	Q1	Q2	Q3	Q4
Small & medium-sized	Yes	Yes	100	129	107	76
	Yes	No	8,900	8,571	8,199	4,019
	No	Yes	3,641	3,543	3,116	2,795
	Σ		12,641	12,243	11,422	6,890
	\hat{c}		0.0111	0.0148	0.0129	0.0186
Large	Yes	Yes	534	582	552	303
	Yes	No	2,584	2,705	2,657	1,528
	No	Yes	3,780	3,608	3,506	3,202
	Σ		6,898	6,895	6,715	5,033
	\hat{c}		0.1713	0.1771	0.1720	0.1655

Oszacowana populacja podmiotów publikujących oferty pracy

Table 5: Szacunkowa liczba podmiotów posiadających wolne miejsca pracy w 2018

Parameter	Q1	Q2	Q3	Q4
Point and standard errors estimates				
\hat{N}_{DL}	54,655 (2,174)	56,076 (2,179)	50,805 (2,031)	44,108 (1,869)
\hat{N}_{LP}	153,960 (–)	132,548 (–)	127,224 (–)	99,694 (–)
\hat{N}_{dep}	73,536 (3,358)	63,364 (2,893)	63,724 (2,969)	49,147 (3,278)
$\hat{N}_{DL,A}$	47,393 (2,118)	46,979 (2,037)	42,297 (1,922)	36,210 (1,759)
$\hat{N}_{LP,A}$	336,690 (–)	247,647 (–)	250,189 (–)	154,694 (–)
$\hat{N}_{dep,A}$	54,620 (2,626)	46,004 (2,224)	45,971 (2,275)	33,967 (2,413)
$\hat{N}_{DL,B}$	7,262 (500)	9,097 (783)	8,508 (669)	7,898 (641)
$\hat{N}_{LP,B}$	25,189 (–)	23,664 (–)	23,591 (–)	21,180 (–)
$\hat{N}_{dep,B}$	18,916 (777)	17,360 (707)	17,753 (734)	15,181 (909)

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

Wnioski

Wnioski i kierunki dalszych badań

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

- Nie ma wystarczająco dobrych metod szacowania populacji ofert pracy przy wszystkich ich założeniach; a badania reprezentacyjne tracą zastosowanie

Wnioski i kierunki dalszych badań

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

- Nie ma wystarczająco dobrych metod szacowania populacji ofert pracy przy wszystkich ich założeniach; a badania reprezentacyjne tracą zastosowanie
- Metody badań nielosowych źródeł danych statystycznych dynamicznie się rozwijają. Chcemy mieć drobny wkład w rozwój metod do szacowania popytu na pracę :)

Wnioski i kierunki dalszych badań

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

- Nie ma wystarczająco dobrych metod szacowania populacji ofert pracy przy wszystkich ich założeniach; a badania reprezentacyjne tracą zastosowanie
- Metody badań nielosowych źródeł danych statystycznych dynamicznie się rozwijają. Chcemy mieć drobny wkład w rozwój metod do szacowania popytu na pracę :)
- Trwają prace podstawowe (wyprowadzanie odpowiednich estymatorów) i stosowane (zastosowania dla danych nt. wakatów) zmierzające do oszacowania wielkości i struktury popytu na pracę przede wszystkich w przekroju zawodów; etapów prac (i potencjalnych obciążeń wyników :/) jest wiele

Wnioski i kierunki dalszych badań

Postępy w
badaniach rynku
pracy

Metody
wnioskowania o
popycie na pracę
z ofert pracy

Wybrane wyniki

Wnioski

References

- Nie ma wystarczająco dobrych metod szacowania populacji ofert pracy przy wszystkich ich założeniach; a badania reprezentacyjne tracą zastosowanie
- Metody badań nielosowych źródeł danych statystycznych dynamicznie się rozwijają. Chcemy mieć drobny wkład w rozwój metod do szacowania popytu na pracę :)
- Trwają prace podstawowe (wyprowadzanie odpowiednich estymatorów) i stosowane (zastosowania dla danych nt. wakatów) zmierzające do oszacowania wielkości i struktury popytu na pracę przede wszystkich w przekroju zawodów; etapów prac (i potencjalnych obciążeń wyników :/) jest wiele
- Dalsze prace będą ewoluowały w kierunku pełnej charakterystyki miejsc pracy, tj. ze względu na inne zmienne, takie jak umiejętności, kwalifikacje *sensu stricto*, płace, typy umów itd.

Beręsewicz, M., Cherniaiev, H. and Pater, R. (2021), 'Estimating the number of entities with vacancies using administrative and online data', *arXiv preprint arXiv:2106.03263* .

Beręsewicz, M. E., Białkowska, G., Marcinkowski, K., Maślak, M., Opiela, P., Katarzyna, P. and Pater, R. (2021), Enhancing the demand for labour survey by including skills from online job advertisements using model-assisted calibration, *in* 'Survey Research Methods', Vol. 15, pp. 147–167.

Beręsewicz, M. and Pater, R. (2021), *Inferring job vacancies from online job advertisements*, Publications Office of the European Union.

Colombo, E., Mercurio, F. and Mezzanzanica, M. (2018), 'Applying machine learning tools on web vacancies for labour market and skill analysis', *Terminator or the Jetsons* .

Gradzewicz, M., Saczuk, K., Strzelecki, P., Tyrowicz, J. and Wyszyński, R. (2016), 'Badanie ankietowe rynku pracy', *Warszawa: Instytut Ekonomiczny NBP* .

Nour, E.-S. (1982), 'On the estimation of the total number of vital events with data from dual collection systems', *Journal of the Royal Statistical Society Series A: Statistics in Society* **145**(1), 106–116.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**(476), 1418–1429.