

Starzenie się społeczeństwa a płace.
Wykorzystanie ujednoliconych danych o wynagrodzeniach.

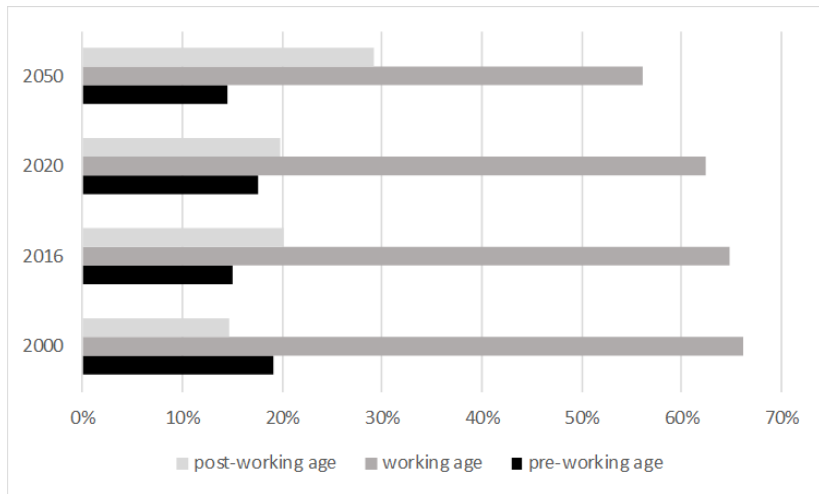
P. Broniatowska SGH M. Nasiński UW A. Majchrowska UŁ

July 9, 2018

- Analiza wpływu zmian struktury wiekowej pracujących w Polsce na przeciętny poziom plac.

- 1 Starzenie się ludności nie tylko w krajach rozwiniętych, ale też w państwach Europy Środkowo-Wschodniej.
- Mediana wieku ludności w Polsce wynosiła 38.5 roku w 2016 i ma wzrosnąć do 41.4 w 2020 oraz do 50.7 w 2050 roku.
- Old-age dependency ratio ma wzrosnąć z 22.4% w 2016 roku do 52.2% w 2050 roku.
- 2 Likwidacja wcześniejszych emerytur - dużo osób starszych zostało / pojawiło się na rynku pracy.
- 3 Wzrost odsetka osób starszych (55-64 lata) wśród pracujących: czy ma wpływ na poziom wynagrodzeń
- 4 Harmonizacja danych z ankiet Z12 - GUS dla dalszych celów badawczych

Motywacja (2)



Indywidualne dane z Badania Struktury Wynagrodzeń według zawodów
Badany okres: 2006-2014.

Zalety: bardzo duża wiarygodność danych o wynagrodzeniach.

Wady: obejmuje tylko podmioty zatrudniające co najmniej 10 osób.

Ze względu na kilkukrotne zmiany Klasyfikacji Zawodów i Specjalności konieczne było ujednoczenie grup zawodowych w czasie.

Powstała unikalna baza danych dotycząca wynagrodzeń i charakterystyk jednostek na 4-cyfrowym poziomie klasyfikacji.

Narzędzie ilościowe - regresja liniowa ważona iloczynem mnożnika GUS oraz prawdopodobieństwa występowania obserwacji.

Model wazonej regresji liniowej

$$\begin{aligned} \log(\text{hourly_wage_real}) = & \\ & \text{high_edu} + \log(\text{age}) + \log(\text{exp}) + \text{old_age_freq_g3} + \text{year} + \\ & \text{female} + \text{etat_SUM} + \text{temporary} + \log(\text{NRemp}) + \text{private} + \text{wage_negotiations1} \end{aligned}$$

Procedura łączenia danych

- Rok bazowy (kierunek łączenia danych)
- Czy występuje metoda łączenia pozwalająca na późniejsza nieobciążona analizę danych
- Jak najkorzystniej wykorzystać mnożnik obserwacji dostarczany przez GUS

Procedura Łączenia danych (1)

- Zbudowano procedurę pozwalającą na łączenie ankiet GUS Z12 publikowanych na przestrzeni lat (2006-2014).
- Każda ankieta Z12 to zbiór około 800 tysięcy obserwacji stanowiących losową próbę pracowników firm zatrudniających powyżej 10 osób.
- Częste zmiany w kodowaniu grup zawodowych stanowiły istotne utrudnienie. Ustawodawca musi korygować spis zawodów i specjalności ze względu na dopracowywane standardy międzynarodowe oraz dużą dynamikę rynku pracy. Omawiany spis nosi nazwę KZIS (Klasyfikacja zawodów i specjalności).

Procedura Łączenia danych (2)

- Procedurę zbudowano w oparciu o czterocyfrowy KZIS (można dobrać dowolny poziom).
- Dowolność w kwestii kierunku łączenia danych (rok bazowy).
- Dla uproszczenia procedura opisana zostanie dla koncepcji z 2008 jako rokiem bazowym.
- Dane replikowane są poprzez strukturę tablicy asocjacyjnej, będzie ona potrzebna także w kolejnej części procedury. Dla unikalnych czterocyfrowych kodów grup KZIS z 2010 tworzona jest mapa przejścia na odpowiadające im grupy w 2008 roku.

Procedura łączenia danych (3)

Fragment tabeli przejść.

```
> head(to_2008)
$`1111` [1] "1111"
$`1112` [1] "1112"
$`1113` [1] "1112"
$`1114` [1] "1121" "1122" "1123"
$`1120` [1] "1211" "1212"
$`1311` [1] "1221" "1311"
```

Fragment tabeli przejść z prawdopodobieństwami.

```
#summ_bsw_2008 - dane zagregowane z 2008 -> podstawa dla 2010
> head(kzis_apply_freq(to_2008, summ_bsw_2008))
$`1111` [1] 1
$`1112` [1] 1
$`1113` [1] 1
$`1114` [1] 0.3333333 0.3333333 0.3333333
$`1120` [1] 0.8343979 0.1656021
$`1311` [1] 0.95471934 0.04528066
```

Procedura Łączenia danych (5)

- Dane z konkretnego roku następującego po roku 2008 replikujemy zgodnie z kluczem KZIS, konkretne obserwacje replikowane są tyle razy ile odpowiadających grup z kodowania w 2008 roku przypada na konkretna badana grupę.
- Prawdopodobieństwa pochodzą z relatywnej częstotliwości danej grupy z poprzedniego roku.
- Dla przejść jeden do jednego przypisujemy wszędzie prawdopodobieństwo 1.
- Jeżeli zer jest więcej niż jedno a żadna z grup nie wystąpiła w roku poprzednim to prawdopodobieństwo rozłożone jest jednostajnie.
- W przypadku grup posiadających kilka niewystępujących grup a chociaż jedna występująca to całe prawdopodobieństwo przypisane jest takowej grupie.

Podsumowanie Procedury Łączenia danych (1)

- Procedura pozwala na otrzymanie niezagregowanych danych z 3 dodatkowymi zmiennymi oraz większą ilością obserwacji w wyniku replikacji. Dodatkowe zmienne to odpowiadająca grupa zawodowa w stosunku do kodowania z 2008 roku (rok bazowy), prawdopodobieństwo obserwacji oraz liczba replikacji danej obserwacji (dla każdej z zreplikowanych ta sama).
- Istotnym jest ze statystyki dla zagregowanych zmiennych innych niż liczebność NIE są obojętne na implementowaną procedurę, należy wyważyć je poprzez zmienna prawdopodobieństwa.
- Przez fakt wykonywania obliczeń na czterocyfrowym KZIS, do danej grupy mogła trafić obserwacja z kilku różnych grup. W grupach mogą znajdować się osoby o różnej liczbie replikacji jak i prawdopodobieństwach.
- Nieobciążone parametry przy obciążonych błędach standardowych (z powodu sztucznego zwiększenia liczebności próbki) - konieczna korekta

Podsumowanie Procedury Łaczenia danych (2)

WYNIKI – cała próba (2006-2016)

	Wartość parametru	Stand. błąd szacunku	Statystyka t	p-value
highedu	0.667	0.000	1448.057	0.000
log(age)	0.099	0.001	82.145	0.000
log(ex)	0.084	0.000	257.728	0.000
freq_oldage	-0.749	0.004	-211.153	0.000
gender	-0.194	0.000	-450.580	0.000
etat	0.002	0.000	20.654	0.000
temp	-0.228	0.000	-433.404	0.000
log(size)	0.054	0.000	484.154	0.000
sector	-0.106	0.000	-225.731	0.000
nego	0.082	0.001	79.144	0.000
2008	0.301	0.001	465.006	0.000
2010	0.325	0.001	496.124	0.000
2012	0.323	0.001	480.216	0.000
2014	0.391	0.001	558.291	0.000
R ²	0.45			
Liczba obserwacji	5202362			

Figure 2: WYNIKI – cała próba

Podsumowanie Procedury Łaczenia danych (3)

WYNIKI dla poszczególnych wielkich grup zawodowych (2006-2016)

	Nazwa grupy	Wartość parametru	Stand. błąd szacunku	Statystyka F-Snedecora	p-value	R ²	Liczba obserwacji
1	Przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy	3.307	0.019	178.079	0.000	0.319	512,650
2	Specjaliści	-2.553	0.009	-280.750	0.000	0.343	590,353
3	Technicy i inny średni personel	-0.120	0.010	-12.146	0.000	0.282	759,579
4	Pracownicy biurowi	0.520	0.030	17.245	0.000	0.297	348,174
5	Pracownicy usług i sprzedawcy	-0.811	0.009	-90.945	0.000	0.354	300,080
7	Robotnicy przemysłowi i rzemieślnicy	0.467	0.011	42.613	0.000	0.518	645,688
8	Operatorzy i monterzy maszyn i urządzeń	0.345	0.010	34.575	0.000	0.428	758,457
9	Pracownicy przy pracach prostych	-0.934	0.005	-173.313	0.000	0.344	1,287,381

Figure 3: WYNIKI dla poszczególnych wielkich grup zawodowych

- 1 Ujemna i istotna statystycznie zależność pomiędzy odsetkiem osób starszych a poziomem płac dla całej próby.
- 2 Silnie zróżnicowana jednak pomiędzy grupami zawodowymi.
- 3 Wyniki zgodne z teorią ekonomii. Potwierdzają też wyniki uzyskane dla innych gospodarek.