

# Krzyżowe porównanie danych o wynagrodzeniach z polskich badań przekrojowych

Paweł Strawiński\*

Nadesłany: 27 lipca 2015 r. Zaakceptowany: 28 września 2015 r.

---

## Streszczenie

W artykule zaprezentowano analizę porównawczą informacji o charakterystykach osób zatrudnionych i wysokości ich wynagrodzeń. Dane pochodzą z trzech głównych przekrojowych badań społeczno-ekonomicznych prowadzonych w Polsce. Podstawowym celem jest identyfikacja ewentualnych rozbieżności między badaniami oraz ich wyjaśnienie. Cel dodatkowy to określenie, w jaki sposób różnice między reprezentatywnymi zbiorami danych zawierającymi informacje o rynku pracy wpływają na wartości parametrów modeli ekonomicznych. Oszacowania parametrów przykładowego modelu stopy zwrotu z wykształcenia pokazują, że w przypadku zmiennych objaśniających o rozkładzie ciągłym niewielkie różnice między danymi nie powodują zróżnicowania ocen parametrów. W przypadku zmiennych o rozkładzie dyskretnym różnice między wartościami oszacowań mogą być jednak znaczne.

---

**Słowa kluczowe:** dane jednostkowe, BAEL, BBGD, BSW, stopa zwrotu z wykształcenia

**JEL:** C81, I26, J31

## 1. Wstęp

Celem niniejszego artykułu jest przeprowadzenie analizy porównawczej charakterystyk osób zatrudnionych oraz informacji o wysokości ich wynagrodzeń. Wykorzystano w tym celu dane z trzech polskich badań o charakterze przekrojowym: badania aktywności ekonomicznej ludności (BAEL), badania struktury wynagrodzeń według zawodów (BSW) oraz badania budżetów gospodarstw domowych (BBGD). Są to zbiory publicznie dostępnych danych jednostkowych. Opracowanie ma charakter statystyczny, jednak wnioski są ważne dla osób prowadzących badania ekonomiczne. Autor odwołuje się do analizy graficznej i podstawowych metod analizy statystycznej. Zgodnie z najlepszą wiedzą autora jest to pierwsze tego typu badanie w Polsce. Ograniczenie do powszechnie stosowanych metod analizy i podstawowych charakterystyk badanej populacji służy rozszerzeniu kręgu odbiorców artykułu. Elementem składowym wielu badań o charakterze wtórnym, prowadzonych na poziomie jednostkowym, jest szacowanie równania płacy typu Mincera dla Polski. Jako przykłady z ostatnich lat można wskazać prace Puhaniego (2000), Newella i Sochy (2005), Strawińskiego (2006; 2008), Rogut i Roszkowskiej (2007), Mycka, Morawskiego i Mycielskiego (2007), Mycka, Nicińskiej i Morawskiego (2009), Gajderowicza, Grotkowskiej i Wincenciaka (2012) oraz Majchrowskiej i Roszkowskiej (2013). Warto zaznaczyć, że każdy zespół badawczy uzyskał inne rezultaty dotyczące parametrów równania płacy. Źródłem obserwowanych rozbieżności może być fakt, że oszacowania tych parametrów opierają się na danych pochodzących z różnych badań.

Przeprowadzone w tym artykule porównanie informacji o rozkładzie uzyskiwanych płac i innych cech respondentów pozwoli na zidentyfikowanie rozbieżności między badaniami, a także ich wyjaśnienie. Dodatkowo, analizowane będą oszacowania stopy zwrotu z wykształcenia jako przykład ważnego parametru ekonomicznego. Pokaże to, w jakim stopniu niewielkie różnice między zbiorami danych wpływają na zróżnicowanie punktowych ocen parametrów. Alternatywnym wytłumaczeniem obserwowanych rozbieżności między wynikami badań empirycznych są różnice metodologiczne pomiędzy badaniami. Przeprowadzona analiza pozwoli stwierdzić, czy przyczyną odmiennych wyników są rozbieżności metodologiczne czy zróżnicowanie wykorzystanego materiału empirycznego.

Wszystkie trzy badania mają dużą część wspólną. Analizowane są w nich między innymi wynagrodzenia osób pracujących w podmiotach gospodarki narodowej zatrudniających 10 pracowników i więcej. Porównanie informacji statystycznej dotyczącej wynagrodzeń oraz innych informacji na temat struktury pracowników może być ciekawe. Dodatkowo, możliwe jest porównanie względnej jakości zbieranych danych. Wyniki analizy opisowej i statystycznej zaprezentowane w drugiej części opracowania pozwolą ukazać różnice i podobieństwa pomiędzy omawianymi zbiorami danych. W celu porównania informacji statystycznej wykorzystano uspołnione definicje analizowanych cech stosowane w badaniach. W końcowej części rozdziału porównano rozkłady wynagrodzeń netto w trzech badaniach.

W ostatniej części zaprezentowano przykład empiryczny. Ilustruje on, jakie skutki dla uzyskiwanych ocen parametrów modeli ekonomicznych ma zróżnicowanie zbiorów danych.

## 2. Porównanie danych

Bezpośrednie porównanie informacji statystycznej dostarczonej przez trzy reprezentatywne badania jest niemożliwe, gdyż różnią się one celami i zakresem. W tej części opracowania zostanie zaprezentowany sposób wyznaczenia porównywalnych podzbiorów reprezentatywnych danych.

Wszystkie wymienione badania podają informacje o wysokości wynagrodzeń i pewnych cechach uzyskujących je osób. Są one reprezentatywne dla Polski na poziomie badania, ale dla różnych populacji. W przypadku BAEL są to osoby w wieku 15 lat i więcej, członkowie gospodarstw domowych, mieszkające pod wylosowanym adresem (GUS 2011a). W przypadku BBGD są to wszyscy członkowie gospodarstw domowych pod wylosowanym adresem (GUS 2011b). W BSW dane dotyczą osób pełnozatrudnionych i niepełnozatrudnionych w podmiotach gospodarki narodowej zatrudniających 10 osób i więcej, które przepracowały cały październik badanego roku (GUS 2012). Pod względem badanej populacji BAEL i BBGD są zatem bardzo podobne, a BSW różni się od pozostałych badań.

Drugą różnicą metodologiczną jest perspektywa, z której obserwowane są jednostki. W przypadku BAEL i BBGD jest to gospodarstwo domowe (GUS 2011a; 2011b), a w BSW są to podmioty gospodarki narodowej zatrudniające 10 osób i więcej (GUS 2012).

Trzecią różnicę stanowi sposób doboru obserwacji do badania. W BAEL i BBGD operatem losowania jest baza meldunkowa, a w BSW – krajowy rejestr urzędowy podmiotów gospodarki narodowej (REGON). Oba rejestry zawierają niedokładne, często nieaktualne informacje. Po zniesieniu obowiązku meldunkowego niektóre osoby nie są zameldowane w miejscu, w którym mieszkają. W przypadku rejestru REGON niepewną informacją jest liczba zatrudnionych<sup>1</sup>. Często jest to liczba deklarowana przez przedsiębiorstwo w momencie rejestracji, później nieaktualizowana. Ewentualne rozbieżności między rejestrami a stanem faktycznym są dużym problemem. Dane pochodzące z oficjalnych rejestrów stanowią podstawę obliczenia wag uogólniających wyniki badania na populację generalną (GUS 2011a; 2011b; 2012). Jest to jedną z przyczyn zróżnicowania struktury społeczno-demograficznej badanych prób wykraczającego poza błąd losowy.

Kolejną różnicą metodologiczną jest sposób przeprowadzenia badania. W BAEL informacje zbierane są w cyklu kwartalnym w sposób ciągły i dotyczą przede wszystkim aktywności ekonomicznej ludności. BBGD odbywa się w cyklu rocznym w sposób ciągły. W każdym miesiącu badana jest inna grupa gospodarstw domowych (rotacja całkowita z miesięcznym okresem wymiany). Sprawozdawane są dochody i przychody w ujęciu indywidualnym oraz rozchody w ujęciu zagregowanym na poziomie gospodarstwa domowego. Z kolei BSW ma cykl dwuletni i jest przeprowadzane w latach parzystych. Gromadzi się w nim szczegółowe informacje o wynagrodzeniach i ich składnikach.

## 2.1. Dostępne informacje

W każdym z rozpatrywanych zbiorów danych jednostkowych inny jest zakres dostępnej informacji o płacach i jej składnikach. W danych pochodzących z BSW dostępne są informacje o wynagrodzeniach brutto oraz ich składnikach w firmach zatrudniających 10 pracowników i więcej (GUS 2012). Dane są ograniczone do pracowników etatowych. Nie jest dostępna informacja o wynagrodzeniach pracowników mniejszych firm oraz osób samozatrudnionych. W BAEL respondenci deklarują wysokość płacy netto w pierwszym miejscu zatrudnienia, o ile wykonują pracę jako pracownicy najemni (GUS 2011a). Wysokość płacy podają zatem tylko osoby zatrudnione. W badaniu nie jest dostępna informacja o wysokości wynagrodzeń uzyskiwanych przez osoby samozatrudnione. BBGD dostarcza informacji na temat dochodów netto, czyli przychodów po odjęciu zaliczki na podatek dochodowy i składek na obowiązkowe ubezpieczenia społeczne oraz ubezpieczenie zdrowotne (GUS 2011b). Obejmuje między innymi do-

<sup>1</sup> W danych pojawiają się np. firmy zatrudniające jednego pracownika, chociaż w rejestrze jest ich co najmniej 10.

chody z pracy najemnej, w tym dochody ze stosunku pracy, umów cywilnoprawnych (umowa zlecenia, umowa o dzieło), pracy nakładczej, praktyk zawodowych czy dochody otrzymywane w ramach przygotowania zawodowego. W odróżnieniu od pozostałych badań dane BBGD uwzględniają wszystkie miejsca zatrudnienia respondenta.

Kolejną różnicą między badaniami jest jednostka losowana do badania. W przypadku BAEL i BBGD jest to gospodarstwo domowe, w BSW – firma zatrudniająca 10 pracowników i więcej. Uczestnikiem badania w BAEL i BBGD jest respondent, który wypełnia ankietę. W razie jego nieobecności może ją wypełnić inny członek wylosowanego gospodarstwa domowego. W badaniu BSW ankietę wypełnia dział kadr firmy. Wskazane różnice powodują, że populacja członków gospodarstw domowych może się różnić od populacji osób pracujących w firmach zatrudniających 10 pracowników i więcej. Ponadto informacje dotyczące wysokości wynagrodzeń pochodzące z działów kadr mogą się różnić od informacji przekazanych przez respondentów.

Najważniejszą konsekwencją różnic metodologicznych pomiędzy badaniami jest to, że na podstawie informacji zawartej w bazach danych nie można bezpośrednio wyznaczyć części wspólnej wszystkich trzech badań. W BBGD nie ma bowiem informacji o wielkości firmy, w której pracuje respondent. W konsekwencji bezpośrednie wyznaczenie części wspólnej BBGD i BSW jest niemożliwe. W opracowaniu zostaną więc porównane osobno próby z badania BAEL i BSW oraz BAEL i BBGD. W celu ukazania różnic między rozkładami cech mierzonych za pomocą skal nominalnych skale o większej liczbie wartości zostaną zagregowane w sposób odpowiadający skalom z mniejszą liczbą wartości. Analizy zostaną przeprowadzone dla danych z 2010 r. Dla pozostałych lat z okresu 2006–2012 rozkłady cech są bardzo podobne.

## 2.2. Część wspólna BAEL i BSW

Wspólną częścią obu badań są informacje o cechach demograficznych i wynagrodzeniach osób pracujących w firmach zatrudniających 10 pracowników i więcej. Bogatszy zestaw informacji społeczno-demograficznych gromadzony jest w BAEL. Stopień szczegółowości pytań w BSW wyznacza zatem część wspólną. Obejmuje ona takie cechy społeczno-demograficzne, jak: płeć, wiek, poziom wykształcenia, wykonywany zawód oraz doświadczenie zawodowe.

Omawiając wspólną część badań, należy uwzględnić różnice metodologiczne między nimi. BAEL jest badaniem ciągłym. Gromadzone informacje dotyczą tygodnia poprzedzającego tydzień badania. W niniejszym opracowaniu wykorzystane są dane BAEL dotyczące IV kwartału. Przyczyną takiego ograniczenia jest chęć zachowania jak największej porównywalności z BSW. Informacje demograficzne w BSW obrazują stan na 31 października roku, w którym przeprowadzono badanie. Informacje o wynagrodzeniach i czasie pracy dotyczą października lub całego roku.

Pomimo różnego sposobu zbierania danych rozkłady podstawowych cech społeczno-demograficznych w obu badaniach powinny być zbliżone. Nie muszą być identyczne przede wszystkim ze względu na błędy losowe towarzyszące badaniom.

## Zatrudnienie

W BSW liczba osób zatrudnionych jest otrzymywana bezpośrednio w badaniu. Wszystkie objęte nim osoby są zatrudnione. W przypadku BAEL określenie liczby osób zatrudnionych w dużych firmach nie jest celem badania i wartość tę należy ustalić za pośrednictwem odpowiedzi udzielonych przez respondentów. Identyfikowano ją na podstawie odpowiedzi na pytanie zadawane pracownikom najemnym: „Ile osób pracuje w Pana(i) instytucji (firmie)”.

Z porównania definicji osoby zatrudnionej przyjętych w obu badaniach wynika, że BSW nie obejmuje wielu kategorii osób zatrudnionych uwzględnionych w BAEL. W szczególności dotyczy to:

- osób zatrudnionych na podstawie umowy o pracę w celu przygotowania zawodowego,
- uczniów szkół dla niepracujących oraz słuchaczy szkół wyższych odbywających praktyki wakacyjne lub dyplomowe,
- osób wykonujących pracę na podstawie umowy zlecenia, umowy o dzieło, umowy cywilnoprawnej, umowy o pracę nakładczą, kontraktu menedżerskiego czy pracę tymczasową,
- przebywających na urlopach wychowawczych w celu sprawowania opieki nad dzieckiem,
- osób, które przebywały łącznie powyżej 10 dni roboczych na zwolnieniach lekarskich (opłaconych przez zakład pracy), urlopach wypoczynkowych, urlopach bezpłatnych,
- osób, które przebywały 1 dzień roboczy lub więcej na zwolnieniach lekarskich opłaconych ze środków ZUS,
- osób, które były na emeryturze i jednocześnie pracowały w jednostce sprawozdawczej.

W BAEL możliwa jest identyfikacja osób wykonujących pracę na podstawie innych umów niż umowa o pracę, przebywających na urlopach wychowawczych, oraz osób, które były na emeryturze i jednocześnie pracowały. Można także zidentyfikować osoby pracujące poza granicami kraju, zatrudnione przez agencje oraz skierowane na szkolenia. Nie jest natomiast możliwa identyfikacja osób, które przebywały łącznie powyżej 10 dni roboczych na zwolnieniach lekarskich (opłaconych przez zakład pracy), urlopach wypoczynkowych, urlopach bezpłatnych oraz osób, które przebywały 1 dzień roboczy lub dłużej na zwolnieniach lekarskich opłaconych ze środków ZUS. Nie są również gromadzone informacje o tym, jaka instytucja opłaca pracownika przebywającego na zwolnieniu lekarskim, ani o liczbie dni urlopu bądź zwolnienia lekarskiego. Informacje w BAEL dotyczą jednego wybranego tygodnia przypadającego przed tygodniem, w którym odbywa się wywiad. Nie zawsze jest możliwa identyfikacja uczniów lub studentów odbywających praktyki zawodowe.

Oszacowana liczba osób pracujących w firmach zatrudniających co najmniej 10 osób wynikająca z BAEL to 8 686 030 z 95-procentowym przedziałem ufności (8 590 058; 8 781 999), podczas gdy według danych BSW to 7 986 102 z 95-procentowym przedziałem ufności (7 966 851; 8 005 353). Czy różnice między tymi oszacowaniami należy uznać za niewielkie czy duże? O tym, że są one znaczne, świadczy fakt, iż przedziały ufności dla średnich nie mają części wspólnej, zatem różnice są statystycznie istotne.

Oceniając różnice, należy wziąć pod uwagę, że badania różnią się operatem losowania. Drugim elementem zniekształcającym wyniki jest zróżnicowana, niehomogeniczna stopa odpowiedzi w obu badaniach. W BAEL stopa odpowiedzi jest zdecydowanie wyższa na obszarach wiejskich niż w dużych miastach. W BSW wyższe stopy odpowiedzi są uzyskiwane w firmach dużych pod względem zatrudnienia.

## Charakterystyki demograficzne

Analizy w tym podrozdziale i w kolejnych prowadzone są na podpróbach oryginalnych danych, ograniczonych do osób pracujących w firmach zatrudniających 10 pracowników i więcej. Wykorzystano definicję osoby zatrudnionej zgodną z BSW.

Pierwszą analizowaną cechą pracowników jest ich płeć. Różnica w udziale kobiet w próbie jest znaczna i wynosi około 6 pkt proc. W BSW udział kobiet wśród pracowników wynosi 50,1%, a w BAEL 43,7%.

Wiek respondenta w BAEL jest ustalany na podstawie odpowiedzi na pytanie o rok urodzenia oraz na pytanie, czy w danym roku miał już urodziny. W danych BSW znajduje się informacja o roku i miesiącu urodzenia pracownika. Ze względu na te różnice informacja o wieku tej samej osoby w obu badaniach może być odmienna. W celu porównania struktury wiekowej oryginalny wiek respondenta zastąpiono wiekiem szacowanym na podstawie jego roku urodzenia. Dane na wykresie 1 zaprezentowano w przedziałach wieku dostosowanych do metody wyznaczania wag w BAEL.

Pod względem struktury wieku różnice między próbami badawczymi są nieznaczne, ale istotne statystycznie (wartość  $p$  statystyki  $\chi^2 = 0,000$ ). Warto zwrócić uwagę na skrajne kategorie. W BAEL liczba osób młodszych, do 34 lat, jest większa o około 1,5 pkt proc. Analogiczną różnicę obserwuje się w starszych grupach, szczególnie w przedziale 45–54 lata. Odmienna liczebność grupy 35–44 lata jest nieistotna w sensie statystycznym (wartość  $p$  statystyki  $\chi^2 = 0,752$ ). Mniejsza liczba młodych pracowników w BSW może być spowodowana wykluczeniem z badania osób przyuczających się do zawodu. Trudno jednak znaleźć uzasadnienie dla różnicy w starszej grupie wiekowej.

Kolejną analizowaną cechą pracowników jest ich poziom wykształcenia. Pytanie o poziom wykształcenia występuje w obu badaniach, ale różne są możliwości odpowiedzi. W BAEL lista odpowiedzi na pytanie o wykształcenie zawiera 12 pozycji, a w BSW 9 pozycji (brak kategorii: „dyplom ukończenia kolegium”, „niepełne podstawowe”, „bez wykształcenia szkolnego”).

W celu porównania rozkładów dostosowano kategorie wykształcenia w danych BSW do kategorii BAEL. Przy agregowaniu kategorii wykształcenia w BSW kierowano się liczbą lat, która jest niezbędna do osiągnięcia danego poziomu wykształcenia. Jest to zgodne z wytycznymi klasyfikacji ISCED<sup>2</sup>. Kategorie wykształcenia: „niepełne podstawowe” oraz „bez wykształcenia szkolnego”, dołączono do kategorii „podstawowe”. Z kolei kategorię „dyplom ukończenia kolegium” dołączono do kategorii „licencjat”.

Na wykresie 2 widzimy, że struktury prób pod względem poziomu wykształcenia są niemal identyczne. W sensie statystycznym rozkłady jednak się różnią (wartość  $p$  statystyki  $\chi^2 = 0,000$ ).

## Charakterystyki zatrudnienia

Drugą grupą charakterystyk, które zostały porównane, są cechy miejsca wykonywania pracy. W przypadku BSW dotyczy to podmiotu gospodarki narodowej, który został wylosowany do badania. W BAEL dotyczy to miejsca pracy, które respondent wskazał jako główne. Zgodnie z metodologią badania za główne miejsce pracy uważa się to, które przynosi wyższy dochód.

Pytanie dotyczące wielkości firmy w obu badaniach ma odmienną konstrukcję. W BAEL informacji tej dostarcza respondent. Lista pytań obejmuje 7 kategorii:

- respondent deklaruje liczbę pracowników, jeśli jest mniejsza niż 10,
- 11–19 pracowników,

<sup>2</sup> UNESCO, <http://www.unesco.org/education/information/nfsunesco/doc/isced\1997.htm>.

- 20–49 pracowników,
- 50–100 pracowników,
- 101–250 pracowników,
- 251 i więcej,
- nie wiem, ale mniej niż 11 osób,
- nie wiem, ale więcej niż 10 osób.

Uwzględnienie w badaniach informacji o dwóch ostatnich kategoriach nie jest możliwe. Odpowiedź „nie wiem, ale mniej niż 11 osób” wskazało 1,7% pracujących. Odpowiedzi „nie wiem, ale więcej niż 10 osób” udzieliło 6,3% pracujących. Stanowią oni 9,3% osób pracujących w dużych firmach.

W BSW znajduje się bezpośrednie pytanie o liczbę pracujących. Na tej podstawie wprowadzono podział według liczby pracowników zgodny z listą pytań w BAEL.

Pod względem liczby pracowników w firmie struktura próby BAEL, zaprezentowana na wykresie 3, znacznie się różni od struktury próby BSW. Obserwowane różnice są istotne w sensie statystycznym (wartość  $p$  statystyki  $\chi^2 = 0,000$ ). W szczególności sposób uwidacznia się to dla kategorii „251 i więcej pracowników”, która stanowi 26,5% próby BAEL oraz 36% próby BSW. Tej rozbieżności nie tłumaczy liczba odpowiedzi „nie wiem” w BAEL.

Doświadczenie zawodowe pracowników można mierzyć na podstawie ogólnego stażu pracy oraz stażu pracy w aktualnym miejscu zatrudnienia. W BAEL respondent jest pytany o to, kiedy zaczął pracować w obecnym miejscu pracy oraz jaki jest jego ogólny staż pracy. Informacja o czasie zatrudnienia jest wyrażona pełną liczbą lat. W BSW informacji o stażu pracy oraz stażu pracy w obecnym miejscu pracy dostarcza dział kadr. Staż wyrażony jest w latach i miesiącach. Aby porównać informacje z obu badań, uwzględniono jedynie lata pracy. Dla zwiększenia czytelności wykresów 4 i 5 obserwacje pogrupowano w 10-letnie przedziały.

W BAEL nieznacznie mniejsza jest liczba osób o ogólnym stażu pracy do 10 lat i jednocześnie nieznacznie więcej jest osób o najdłuższym stażu: 30 lat i więcej. Mimo że informacja w BAEL pochodzi bezpośrednio od respondentów, a w BSW wynika z oficjalnych dokumentów, rozkłady stażu pracy w obu badaniach są bardzo podobne. Różnice między rozkładami są jednak statystycznie istotne (wartość  $p$  statystyki  $\chi^2 = 0,000$ ). Jedynie w przypadku stażu pracy w miejscu zatrudnienia wynoszącego 10–19 lat różnica nie jest istotna statystycznie (wartość  $p$  statystyki  $\chi^2 = 0,990$ ).

## Wykonywany zawód

Kolejną analizowaną charakterystyką związaną z miejscem pracy jest zawód wykonywany przez respondenta. W BAEL wykorzystano 4-cyfrowy kod klasyfikacji zawodów i specjalności Ministerstwa Pracy i Polityki Społecznej, a w BSW kod 6-cyfrowy. Aby zachować przejrzystość prezentacji, na wykresie 6 porównano wielkie grupy zawodów (na poziomie 1-cyfrowym).

Struktury prób według zawodów w obu badaniach są podobne. Warto zwrócić uwagę na dużo wyższy udział specjalistów w BSW niż w BAEL. Może się to wiązać z faktem, że BSW bada próbę przedsiębiorstw, a nie populację. Z kolei na stanowiskach technicznych i robotniczych wyższe udziały odnotowano w BAEL. Występujące znaczne różnice we wskazanych grupach zawodowych powodują, że wynik formalnego testu na równość rozkładów jest negatywny (wartość  $p$  statystyki  $\chi^2 = 0,000$ ). Jedynie w przypadku czwartej wielkiej grupy zawodowej (pracownicy biurowi) obserwowana różnica nie jest statystycznie istotna (wartość  $p$  statystyki  $\chi^2 = 0,684$ ).

## Wynagrodzenia

Najważniejsze porównanie, które jest celem przeprowadzonej analizy, dotyczy rozkładu wynagrodzeń w obu badaniach. Czynnikiem utrudniającym bezpośrednie porównanie rozkładów jest odmienny sposób gromadzenia informacji statystycznej.

W BAEL respondenci są proszeni o podanie wysokości zarobków netto w poprzednim miesiącu w głównym miejscu pracy. Warto zauważyć, że jedynie około 1/3 respondentów udziela odpowiedzi na to pytanie. Dodatkowo, około 46% respondentów badania wskazuje przedział zarobków. Jednak z powodu niedostosowania przedziałów w ankiecie do rzeczywistego rozkładu wynagrodzeń ponad 20% obserwacji trafia do najwyższej kategorii (w 2010 r. 2500 zł i więcej). W rezultacie nie można wykorzystać tej informacji do tworzenia rozkładu wynagrodzeń. Pozostałe ponad 20% zatrudnionych odmawia odpowiedzi na pytanie.

Dane BSW pochodzą z działów księgowości badanych jednostek. Dostarczają one informacji o wynagrodzeniach ogółem brutto i ich składnikach. W konsekwencji BAEL informuje o płacy netto, a BSW o płacy brutto. Konstrukcja polskiego systemu podatkowego sprawia, że nie istnieje funkcja, która jednoznacznie przeliczałaby wynagrodzenie netto na wynagrodzenie brutto. Taka funkcja nie jest różnowartościowa. W ogólnym przypadku funkcja do niej odwrotna nie istnieje. Przy pewnych założeniach na podstawie wynagrodzenia brutto można wyznaczyć wynagrodzenie netto, ale będzie ono obciążone błędem. Mamy świadomość, że jest to bardzo uproszczona metoda. BSW nie zawiera jednak informacji o sytuacji rodzinnej osoby, której wynagrodzenie jest raportowane. Z tego powodu zdecydowano się przyjąć pewne upraszczające założenia.

W celu porównania informacji statystycznej z obu badań przeliczono wynagrodzenia brutto z BSW na wynagrodzenia netto. Założono, że pracownik:

- ma jedno źródło dochodu,
- samodzielnie rozlicza podatki,
- nie korzysta z odliczeń od podatku.

Pierwsze założenie jest uzasadnione, gdyż jedynie niewielka liczba osób ma drugie miejsce zatrudnienia. Według danych BAEL z IV kwartału 2010 r. jest to około 6,5% zatrudnionych (GUS 2011a), z czego 2/3 pracuje na własny rachunek, a 1/3 wykonuje pracę najemną. Słuszność drugiego i trzeciego założenia jest wątpliwa, gdyż wielu podatników korzysta z możliwości wspólnego rozliczania się lub odliczeń od podatku. W rezultacie rzeczywiście płacone podatki są niższe niż wynikające z założeń. Należy się zatem spodziewać, że rzeczywiste wynagrodzenia netto osób zatrudnionych są w rzeczywistości wyższe niż wyliczone.

W celu obliczenia wysokości wynagrodzeń netto wykorzystano skalę podatkową obowiązującą w 2010 r.:

- przychód do 85 528 zł = 18% przychodu minus 556,02 zł kwoty wolnej,
- przychód powyżej 85 528 zł = 14 839,02 zł plus 32% nadwyżki ponad 85 528 zł.

Ostatnim krokiem analizy jest porównanie rozkładu wynagrodzeń w próbie BAEL i BSW. W tym celu wyznaczono wartości wynagrodzeń dla kolejnych decyli rozkładu oraz dodatkowo dla 1. i 99. percentyla rozkładu. Graficznie zaprezentowano rozkłady na wykresie 7.

Do 20. percentyla rozkładu różnicowanie wysokości wynagrodzenia pomiędzy próbami jest nieznaczne – nie przekracza 100 zł. Powyżej tego punktu wraz z każdym decylem różnica wzrasta przeciętnie o 100 zł. Dzieje się tak aż do 8. decyla. W 9. decyli i 99. percentylu różnica jest znacznie większa.



Wskazuje to, że dla niskich zarobków, poniżej 3000 zł, z obu zbiorów płynie podobna informacja. Wadą BAEL jest nieujawnianie wysokości płac przez osoby o zarobkach przekraczających średnią dla sektora przedsiębiorstw. Uwzględnienie przy tworzeniu rozkładu dla BAEL informacji z pytania dodatkowego – o wskazanie przedziału płacy netto – nie poprawia sytuacji, gdyż w 2010 r. dolną granicą ostatniego przedziału było 2500 zł.

Problem z prawym ogonem rozkładu płac w danych BAEL został wcześniej zauważony w pracy Mycka, Morawskiego i Mycielskiego (2007). Mimo jego występowania można uznać, że rozkłady płac w obu badaniach są zbliżone. Będzie to szczególnie uzasadnione, jeśli założymy, że 9. decyl rozkładu wynagrodzeń z BAEL odpowiada 8. decylowi rozkładu wynagrodzeń z BSW.

### 2.3. Część wspólna BAEL i BBGD

Podobnie jak w przypadku BAEL i BSW wspólna część obu badań obejmuje informacje o cechach demograficznych, cechach miejsca pracy i wynagrodzeniach. Informacje z BAEL dotyczą IV kwartału 2010 r., z BBGD całego 2010 r. Wykorzystano wszystkie obserwacje, by zapewnić odpowiednio dużą liczebność próby BBGD.

#### Zatrudnienie

Liczba pracowników najemnych oszacowana na podstawie BAEL wynosi 11 882 283 z 95-procentowym przedziałem ufności (11 829 413; 11 935 152). Analogiczne szacunki dla danych BBGD wynoszą 11 940 136 z 95-procentowym przedziałem ufności (11 885 408; 11 994 864). Liczba pracowników najemnych oszacowana na podstawie obu zbiorów danych jest zatem bardzo podobna. Różnica wynosi niespełna 60 tys. osób i nie jest statystycznie istotna (wartość p statystyki  $\chi^2 = 0,15$ ); stanowi to mniej niż 0,5% ogólnej liczby. Zbliżone szacunki są wynikiem zbieżności definicyjnej BBGD oraz BAEL. W obu badaniach do dochodów z pracy najemnej zalicza się dochody zarówno ze stosunku pracy, jak i umów cywilnoprawnych, umów o pracę nakładczą, kontraktów menadżerskich itp. Pod uwagę brane są okresy tymczasowej absencji, w tym urlopy płatne, bezpłatne, macierzyńskie, wychowawcze, okresy przebywania na zwolnieniu lekarskim itp. Różnica w oszacowaniu przypuszczalnie wynika ze sposobu identyfikacji osób pracujących. W BBGD są one identyfikowane na podstawie otrzymywanych dochodów, a nie tylko deklaracji o aktywności zawodowej jak w przypadku BAEL.

#### Charakterystyki demograficzne

Procentowy udział kobiet oszacowany na podstawie BBGD i BAEL wynosi, odpowiednio, 46,2% oraz 47,5%. Różnica około 1,3 pkt proc. jest statystycznie istotna (wartość p statystyki  $\chi^2 = 0,000$ ). Może ona wynikać z nieco częstszego uwzględniania w próbie BBGD osób mniej aktywnych zawodowo, do których ankietom łatwiej dotrzeć.

Struktury wiekowe pracowników najemnych w obu badaniach są również bardzo podobne (zob. wykres 8). Obserwowane różnice między oszacowaną liczebnością każdej grupy są jednak istotne statystycznie (wartość p statystyki  $\chi^2 = 0,000$ ).

Badania BBGD oraz BAEL nieznacznie różnią się pod względem listy odpowiedzi na pytanie o wykształcenie. Oprócz poziomów wykształcenia uwzględnionych w BAEL w BBGD występują dodatkowo kategorie: „bez wykształcenia” oraz „kolegium nauczycielskie, języków obcych lub dla pracowników służb społecznych”. Aby porównać rozkłady wykształcenia, dodatkowe kategorie w BBGD przyporządkowano kategoriom występującym w obu badaniach, odpowiednio: „wykształceniu podstawowemu” oraz „licencjatowi”.

Skonstruowane w ten sposób rozkłady cechy wykształcenie zaprezentowano na wykresie 9. Są one zbliżone, jednak w bazie BBGD wyższy jest odsetek osób słabiej wykształconych. W przypadku kategorii „wykształcenie podstawowe i zasadnicze” występuje nadreprezentacja w stosunku do BAEL, a w kategorii „wykształcenie magisterskie” – niedostateczna reprezentacja. Obserwowane różnice są statystycznie istotne we wszystkich grupach wykształcenia z wyjątkiem licencjatu.

### Charakterystyki zatrudnienia

Jedyną charakterystyką zatrudnienia bezpośrednio dostępną w BBGD jest wykonywany zawód. Stopień dezagregacji zawodów z klasyfikacji zawodów i specjalności jest w przypadku BBGD niższy niż w BAEL. Dostępna jest jedynie informacja o 2-cyfrowym kodzie. Porównano rozkłady na najwyższym poziomie agregacji, czyli dla wielkich grup zawodowych (wykres 10). W obu badaniach wydają się zbliżone, z wyjątkiem udziału grupy 9, czyli pracowników zatrudnionych przy pracach prostych, których w BBGD jest o jedną trzecią więcej niż w BAEL. Różnice w oszacowaniach udziału poszczególnych grup w populacji zatrudnionych nie przekraczają 2 pkt proc., jednak okazują się statystycznie istotne (wartość  $p$  statystyki  $\chi^2 = 0,000$ ).

### Wynagrodzenia

Ostatni etap analiz stanowi porównanie rozkładów wynagrodzeń. W tym celu, podobnie jak w przypadku porównania rozkładu z BAEL i BSW, wyznaczono wartości dla kolejnych decyli oraz dodatkowo dla 1. i 99. percentyla rozkładu wynagrodzeń. Graficznie zaprezentowano rozkłady na wykresie 11.

Przed porównaniem rozkładów warto zwrócić uwagę, że w BAEL i w BBGD stosuje się inne definicje wynagrodzeń. W pierwszym badaniu respondenci podają jedynie dochody z głównego miejsca pracy, natomiast w BBGD wymagane są bardzo dokładne informacje dotyczące wszystkich źródeł (ponad 80 typów) dochodów lub przychodów. Z tego względu dochody z pracy najemnej są w BBGD systematycznie wyższe niż w BAEL. Rozkład w BBGD rozpoczyna się jednak od niższej wartości. Dla 1. i 2. decyla różnica pomiędzy badaniami nie przekracza 10 zł, a dla 3. i 4. decyla wynosi 100 zł. W 8. decyly rozkładu wynagrodzeń wynosi 300 zł, w 9. – 500 zł. Jest to tendencja bardzo podobna jak w przypadku różnic między wynagrodzeniami w BAEL i w BSW.

## 3. Konsekwencje różnic dla ocen parametrów

W tej części opracowania zostanie omówione, w jaki sposób różnice metodologiczne pomiędzy polskimi badaniami przekrojowymi wpływają na różnice między uzyskiwanymi wartościami ocen para-

metrów dla podstawowego modelu szeroko wykorzystywanego w naukach społecznych, to jest modelu kapitału ludzkiego Mincera. Pokażemy, jakie różnice między ocenami parametrów mogą być powodowane przez wykorzystanie różnych zbiorów danych.

W tym celu wykorzystamy model premii za wykształcenie oparty na teorii kapitału ludzkiego. Według tego modelu stopa zwrotu z wykształcenia to liczbowa wartość parametru przy zmiennej określającej wykształcenie w regresji logarytmu płacy względem liczby lat edukacji albo poziomu wykształcenia oraz zmiennych kontrolnych. Jako zmienne kontrolne posłużą charakterystyki, które są wspólne dla wszystkich badań. Ze względu na niemożność bezpośredniego porównywania wszystkich trzech badań zostaną oszacowane oddzielne modele dla porównania BAEL z BSW oraz BAEL z BBGD. Zgodnie z tendencjami w literaturze modele zostaną oszacowane dla całych dostępnych prób oraz dla prób ograniczonych do osób w wieku 25–45 lat.

W celu porównania ocen parametrów uzyskanych na podstawie tego samego modelu, którego parametry są uzyskiwane z różnych zbiorów danych, w tej części opracowania posłużono się charakterystykami społeczno-demograficznymi respondentów, ujednoczonymi we wszystkich trzech zbiorach. Wspólne informacje oprócz wysokości płac respondentów dotyczą: płci, wieku, poziomu wykształcenia oraz wykonywanego zawodu. Dodatkowo, zbiór zmiennych objaśniających uzupełniono o potencjalne doświadczenie zawodowe respondenta (w latach) i jego kwadrat podzielony przez 100. Potencjalne doświadczenie zawodowe obliczono, odejmując od wieku respondenta liczbę lat, jaką powinien poświęcić na uzyskanie swojego wykształcenia. Jest to postępowanie zgodne z regułami klasyfikacji ISCED. Zmienną zależną w modelu stanowi wysokość wynagrodzenia netto w pierwszym miejscu pracy. Informację tę zawierają dane BAEL oraz BBGD. Procedura jej uzyskania dla zbioru BSW została opisana w drugiej części opracowania. Jako zmienne objaśniające wykorzystano wszystkie charakterystyki wspólne dla zbiorów.

W główce każdej tabeli z wynikami umieszczono nazwę zbioru danych, z którego korzystano, by uzyskać oszacowania. Brak litery obok nazwy zbioru oznacza, że wykorzystano wszystkie dane dostępne za 2010 r. (w przypadku BAEL za IV kwartał 2010 r.). Litera P znaczy, że próbę ograniczono do osób w wieku najwyższej aktywności zawodowej, czyli 25–44 lata. Z kolei literą D oznaczono próbę osób pracujących w firmach zatrudniających 10 pracowników i więcej.

W pierwszej kolejności zostaną przeanalizowane oszacowania modeli regresji logarytmu płacy względem liczby lat edukacji i zmiennych kontrolnych. Ich wyniki przedstawiono w tabelach 1 i 2. W tabeli 1 zaprezentowano oszacowania dla danych BAEL i BSW. Ze względu na to, że BSW dotyczy dużych firm, porównywalne wyniki umieszczono w czterech ostatnich kolumnach. W tabeli 2 zrezygnowano z prezentowania oszacowań ograniczonych do dużych firm z powodu niemożności ich uzyskania dla danych BBGD.

Na wstępnym etapie analizy warto zwrócić uwagę, że wartość współczynnika determinacji liniowej jest podobna we wszystkich modelach<sup>3</sup>. Nieznacznie niższe wartości uzyskano dla danych BSW, ale jest to zrozumiałe, zważywszy na około stukrotnie większą liczebność próby.

Oszacowana wartość stopy zwrotu z dodatkowego roku edukacji jest podobna we wszystkich zbiorach danych. Najwyższe punktowe wartości oszacowania uzyskano dla danych z BBGD (0,053–0,054), przeciętne dla zbioru BSW (0,042–0,051), a najniższe na podstawie danych BAEL (0,040–0,050). Ograniczenie próby do pracowników w wieku największej aktywności zawodowej prowadzi do nieznacznego spadku wartości oszacowanej stopy zwrotu z edukacji w przypadku danych BBGD oraz do wzrostu jej wartości w przypadku danych BAEL i BSW.

<sup>3</sup> W tabeli 7 zaprezentowano przedziały ufności dla ocen parametrów.

W podobny sposób kształtują się oszacowania premii wynikających z dłuższego doświadczenia zawodowego. Najniższymi wartościami oszacowania charakteryzują się modele oparte na danych BAEL, a najwyższymi – na danych BSW. Po ograniczeniu prób do osób w wieku największej aktywności zawodowej wartości oszacowania stają się niemal identyczne we wszystkich modelach.

Współczynnik przy zmiennej zero-jedynkowej kobieta, mierzący lukę płacową między kobietami a mężczyznami, we wszystkich specyfikacjach modelu z wyjątkiem jednej mieści się między -0,2 a -0,3. Dla pełnej próby BSW jest nieznacznie mniejszy niż 0,2.

Na koniec oszacowano parametry określające wpływ zmiennych wskazujących na przynależność pracownika do grupy zawodowej. Kategorią odniesienia jest grupa 1., czyli przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy. Porównanie wyników z tabeli 2 wskazuje, że oszacowania na podstawie danych BAEL i BBGD mają zbliżone wartości. Z kolei wyniki z tabeli 1 pokazują, że istnieją znaczne różnice między punktowymi wartościami oszacowań dla grup zawodowych na podstawie danych BAEL a BSW. Wskazuje to, że na uzyskane punktowe wartości oszacowań przy zmiennych zero-jedynkowych utworzonych przez rozkodowanie zmiennej o charakterze jakościowym może wpływać rozkład tej zmiennej w próbie. Jest on podobny w badaniu BAEL i BBGD, ale występują znaczne różnice między danymi BAEL i BSW.

Na drugim etapie analizie zostaną poddane oszacowania modeli regresji logarytmu płacy względem zbioru zmiennych wskazujących poziom wykształcenia i zmiennych kontrolnych. Rezultaty oszacowań zaprezentowano w tabeli 3 i 4. Tabela 3 zawiera oszacowania dla danych BAEL i BSW. Podobnie jak w przypadku modeli z liczbą lat edukacji porównywalne wyniki znajdują się w czterech ostatnich kolumnach<sup>4</sup>. W tabeli 4 zrezygnowano z prezentowania oszacowań ograniczonych do dużych firm ze względu na niemożność ich uzyskania dla danych BBGD.

Najwyższe dopasowanie danych do przyjętej formy funkcyjnej modelu uzyskano w przypadku BAEL. Wartość współczynnika dopasowania w tych modelach wynosi od 0,366 do 0,388. W przypadku danych BBGD dopasowanie było nieco słabsze (0,309–0,357). Najślabiej dopasowany okazał się model dla danych BSW, co wiąże się z dużo większą liczebnością tego zbioru (0,249–0,285).

Analizując wyniki oszacowania premii wynikającej z dłuższego doświadczenia zawodowego, należy zauważyć, że we wszystkich modelach kształtuje się ona na zbliżonym poziomie. W przypadku danych BAEL wynosi od 0,024 do 0,037, dla BSW od 0,031 do 0,038, a w przypadku BBGD od 0,036 do 0,039. Dużo większy rozrzut oszacowań dla danych BAEL wynika z faktu, że jest to najmniejszy zbiór spośród trzech analizowanych. Ponadto na podstawie tego zbioru oszacowano cztery modele, podczas gdy na podstawie pozostałych zbiorów po dwa. Należy zaznaczyć, że wyniki uzyskane dla premii za doświadczenie zawodowe są bardzo bliskie uzyskanym za pomocą modelu, w którym miarą wykształcenia była implikowana liczba lat edukacji.

Współczynnik przy zmiennej zero-jedynkowej kobieta obrazuje lukę płacową między kobietami a mężczyznami. W przypadku wszystkich specyfikacji modelu (z jednym wyjątkiem) wynosi ona od -0,2 do -0,3. Najniższe wartości współczynnika uzyskano dla danych BSW. Wartości parametrów dla danych BBGD i BAEL są bardzo podobne. Wyniki są analogiczne do uzyskanych w modelu z implikowaną liczbą lat edukacji.

Punktowe oceny parametrów przy zmiennych wskazujących na przynależność pracownika do grupy zawodowej, tak jak w przypadku wykształcenia mierzonego latami edukacji, są podobne dla danych BAEL i BBGD, ale różne dla danych z BSW. Potwierdza to, że na uzyskane punktowe wartości

<sup>4</sup> W tabeli 8 zaprezentowano przedziały ufności dla ocen parametrów.

oszacowań przy zmiennych zero-jedynkowych utworzonych przez rozkodowanie zmiennej o charakterze jakościowym może wpływać rozkład tej zmiennej w próbie.

Na koniec przeanalizowano parametry zmiennych opisujących poziom wykształcenia. Poziomym odniesienia jest dla nich posiadanie doktoratu. W przypadku współczynników dla poszczególnych poziomów wykształcenia łatwo można dostrzec, że ich wartości są istotnie różne w każdym modelu. Dotyczy to nie tylko modeli oszacowanych na podstawie innych zbiorów danych, ale także modeli szacowanych na podstawie różnych podzbiorów tego samego zbioru danych.

Ze względu na różnice między punktowymi ocenami parametrów dla zmiennych obrazujących wykonywany zawód i brak podobieństwa ocen parametrów dla zmiennych obrazujących poziom wykształcenia wykonano dodatkowe obliczenia. Ich celem było stwierdzenie, czy wybranie grupy odniesienia w przypadku zmiennej o charakterze jakościowym wpływa na uzyskanie podobnych ocen parametrów. W modelach, których wyniki zaprezentowano w tabelach 3 i 4, poziomym odniesienia dla wykształcenia był doktorat. Taki poziom wykształcenia rzadko jednak występuje w populacji; legitymuje się nim niecały 1% osób. W celu weryfikacji hipotezy oszacowano dodatkowe modele dla danych BAEL i BSW, w których poziomym referencyjnym dla wykształcenia jest wykształcenie zawodowe, najczęściej występujące w populacji. Wyniki dodatkowych oszacowań zaprezentowano w tabeli 5. Wskazują one, że zmiana kategorii odniesienia istotnie wpływa na oszacowania. Po zmianie poziomu referencyjnego uzyskane oszacowania są bardzo podobne we wszystkich modelach. Dotyczy to zwłaszcza wykształcenia, którego oszacowania są zbliżone do poziomu odniesienia. Różnice między ocenami parametrów w modelach szacowanych na podstawie danych BAEL i danych BBGD są nieznaczne, a w przypadku modeli szacowanych na podstawie danych BSW – większe.

Wszystkie uzyskane wyniki wskazują, że jeśli analizowana jest zależność między zmiennymi o ciągłych rozkładach, to wybór zbioru danych, na podstawie którego szacowana jest zależność, ma niewielki wpływ na uzyskane punktowe wartości estymatora. Świadczą o tym zbliżone wartości oszacowań premii za doświadczenie zawodowe we wszystkich specyfikacjach modelu czy stopy zwrotu z wykształcenia w przypadku, gdy jest ono mierzone implikowaną liczbą lat edukacji. W przypadku analizowania siły związku między zmienną o rozkładzie ciągłym a zmienną o rozkładzie dyskretnym uzyskanie zbliżonych oszacowań jest możliwe tylko wtedy, gdy zmienna dyskretna ma podobny rozkład w różnych zbiorach. Na punktowe oceny parametrów wpływa również sposób wyboru kategorii odniesienia i kodowania zmiennej.

Konsekwencje różnic między zbiorami danych zależą od typu zmiennych. Dla zmiennych o rozkładach ciągłych różnice między punktowymi ocenami parametrów modeli są niewielkie. Gdy analizuje się zróżnicowanie ocen parametrów obrazujących związek między zmiennymi ciągłymi a dyskretnymi, wartości liczbowe parametrów mogą się znacznie różnić, jeśli zmienna dyskretna ma istotnie różny rozkład w zależności od zbioru danych. Dzieje się tak, ponieważ na wartości współczynników przy zmiennych o rozkładzie dyskretnym większy wpływ ma występowanie obserwacji nietypowych.

#### 4. Wnioski

W artykule przeprowadzono analizę porównawczą informacji o charakterystykach osób zatrudnionych i wysokości ich wynagrodzeń, zawartych w trzech polskich badaniach przekrojowych: badaniu aktywności ekonomicznej ludności, badaniu struktury wynagrodzeń według zawodów oraz badaniu budżetów gospodarstw domowych.

Badania te mają charakter próby losowej z populacji generalnej. Biorąc pod uwagę losowy charakter prób i związane z tym błędy losowe, a także zróżnicowanie populacji objętych badaniem, należy uznać, że charakterystyki respondentów są podobne we wszystkich trzech badaniach. Warto zwrócić uwagę, że w próbie BSW jest więcej kobiet, osób starszych, o dłuższym doświadczeniu zawodowym, pracujących w dużych przedsiębiorstwach zatrudniających powyżej 250 pracowników. W przypadku pozostałych charakterystyk nie odnotowano znacznych różnic. Dzięki dużej zbieżności definicji rozkłady cech w próbie BBGD oraz BAEL są bardzo podobne, z wyjątkiem wynagrodzeń.

Największe różnice w rozkładzie wynagrodzeń między zbiorami BAEL, BBGD i BSW są obserwowane w prawym ogonie. Prawdopodobnie jest to skutkiem odmowy odpowiedzi na pytanie o wysokość uzyskiwanego wynagrodzenia w BAEL przez osoby o potencjalnie wyższych zarobkach oraz nadreprezentacji w BBGD osób starszych i mniej aktywnych na rynku pracy. W BAEL 20% respondentów udzielających odpowiedzi na pytanie o dochód wskazuje najwyższy możliwy przedział. W BSW respondenci to osoby aktywne zawodowo i informacje o ich wynagrodzeniu są uzyskiwane bez ich wiedzy. Można zatem stwierdzić, że przyczyną odmiennego rozkładu płac są różnice metodologiczne pomiędzy badaniami.

Ze względu na obszerność materiału statystycznego w artykule zaprezentowano tylko wyniki dla danych z 2010 r. Analizy powtórzone dla danych z 2006, 2008 oraz 2012 r. dały rezultaty bardzo podobne do zaprezentowanych. Świadczy to, że uwidocznione podobieństwa i różnice nie mają charakteru losowego, lecz systematyczny. Ten systematyczny charakter może być powiązany z różnicami metodologicznymi między badaniami.

Różnice między zbiorami danych BSW a BAEL i BBGD są uzasadnione na gruncie nauk ekonomicznych. Mężczyźni częściej niż kobiety wybierają samozatrudnienie. Wobec tego w BAEL i BBGD mogą zostać objęci definicją osoby pracującej, ale nie są objęci badaniem BSW. Z kolei kobiety, w szczególności starsze, dużo chętniej podejmują pracę w sektorze państwowym albo dużych zakładach pracy gwarantujących stabilność zatrudnienia.

Uzyskane rezultaty w części empirycznej są zgodne z fundamentalnymi podstawami statystycznymi modelu regresji liniowej. Współczynnik korelacji Pearsona jest dobrą miarą zależności między zmiennymi, co do których można założyć, że mają rozkład ciągły. Jeśli jedna ze zmiennych jest dyskretna lub jakościowa, to wykorzystanie współczynnika korelacji Pearsona może spowodować nieprawidłową ocenę siły związku. Prowadzi to do następującej konstatacji. Uprawnione jest porównywanie rezultatów modeli oszacowanych na podstawie różnych zbiorów danych, jeżeli dotyczą one zależności między zmiennymi o charakterze ciągłym. Porównywanie zależności między zmiennymi o rozkładzie ciągłym i zmiennymi o rozkładach dyskretnych może prowadzić do błędów. Ocena siły związku zależy bowiem od arbitralnie przyjętej skali dla zmiennej dyskretniej lub sposobu jej rozkodowania na zmienne zero-jedynkowe. Dzieje się tak, gdyż na uzyskiwane punktowe oceny parametrów wpływa zarówno sposób rozkodowania zmiennej, jak i jej rozkład w danej próbie. Większość autorów, porównując wartości parametrów przy poziomach wykształcenia, z reguły nie sprawdza, czy sposób konstrukcji modelu w różnych badaniach był jednakowy.

Dalsze prace będą zmierzać w dwóch kierunkach. Dzięki wykorzystaniu statystycznych metod łączenia postaramy się uwzględnić w danych BBGD fakt pracy w firmie zatrudniającej 10 osób i więcej. Uzyskane wstępne wyniki potwierdzają, że badania zawierają różne informacje o wysokości wynagrodzeń i należy dokonać ich unifikacji. Stworzenie danych zharmonizowanych, zawierających informacje z trzech podstawowych źródeł, wzbogaci naszą wiedzę o rozkładzie wynagrodzeń w polskiej gospodarce. Jest to możliwe ze względu na podobne rozkłady zmiennych społeczno-demograficznych osób zatrudnionych, objętych tymi badaniami.

## Bibliografia

- Gajderowicz T., Grotkowska G., Wincenciak L. (2012), Premia płacowa z wykształcenia wyższego według grup zawodów, *Ekonomista*, 5, 577–603.
- GUS (2011a), *Badanie Aktywności Ekonomicznej Ludności IV kwartał 2010 r.*, Główny Urząd Statystyczny, Warszawa.
- GUS (2011b), *Budżety Gospodarstw Domowych w 2010 r.*, Główny Urząd Statystyczny, Warszawa.
- GUS (2012), *Struktura wynagrodzeń według zawodów w październiku 2010 r.*, Główny Urząd Statystyczny, Warszawa.
- Majchrowska A., Roszkowska S. (2013), Czy wykształcenie i doświadczenie zawodowe ma znaczenie? Wyniki równania Mincera dla Polski, *Roczniki Kolegium Analiz Ekonomicznych*, 30, 235–251.
- Myck M., Morawski L., Mysiński J. (2007), Employment fluctuations and dynamics of the aggregate average wage in Poland, 1996–2003, *The Economics of Transition*, 15, 759–779.
- Myck M., Nicińska A., Morawski L. (2009), *Count your hours: returns to education in Poland*, IZA Discussion Papers, 4332.
- Newell A., Socha M. (2005), *The distribution of wages in Poland, 1992–2002*, IZA Discussion Paper, 1485.
- Puhani P. (2000), *On the identification of relative wage rigidity dynamics. A proposal for methodology on cross-section data and empirical evidence for Poland in transition*, IRISS Working Paper, 2000-11.
- Rogut A., Roszkowska S. (2007), Rozkład płac i kapitału ludzkiego w Polsce, *Gospodarka Narodowa*, 11–12, 55–84.
- Strawiński P. (2006), Zwrot z inwestycji w wyższe wykształcenie, *Ekonomista*, 6, s. 805–821.
- Strawiński P. (2008), Oplacalność studiów w Polsce. Model i jego wyniki, *Ekonomista*, 4, 535–553.

## Podziękowania

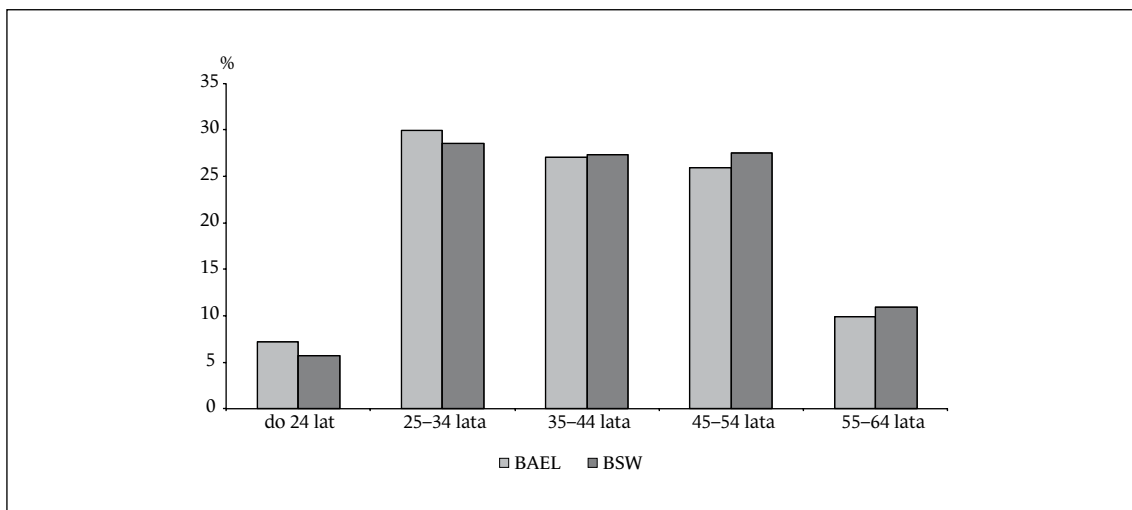
Badanie sfinansowano ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji nr DEC-2013/09/B/HS4/01304. Autor dziękuje uczestnikom Warsztatów Doktorskich z zakresu Statystyki i Ekonometrii 2014, członkom Narodowej Rady Statystycznej przy Prezesie Głównego Urzędu Statystycznego oraz anonimowym recenzentom za uwagi i wskazówki, które wzbogaciły tekst. Dodatkowe podziękowania dla Karoliny Konopczak, która wyjaśniła niejasności metodologiczne dotyczące danych BBGD.

Wyrażone w artykule opinie są poglądami autora i nie powinny być łączone z instytucjami.

## Aneks

Wykres 1

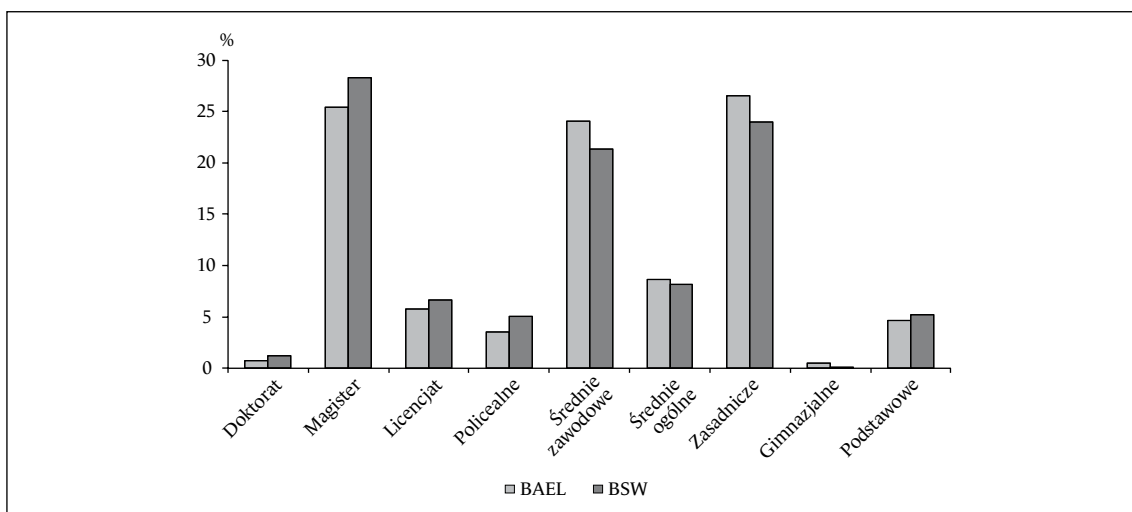
Struktura według wieku respondentów: BAEL a BSW



Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 2

Struktura według wykształcenia respondentów: BAEL a BSW

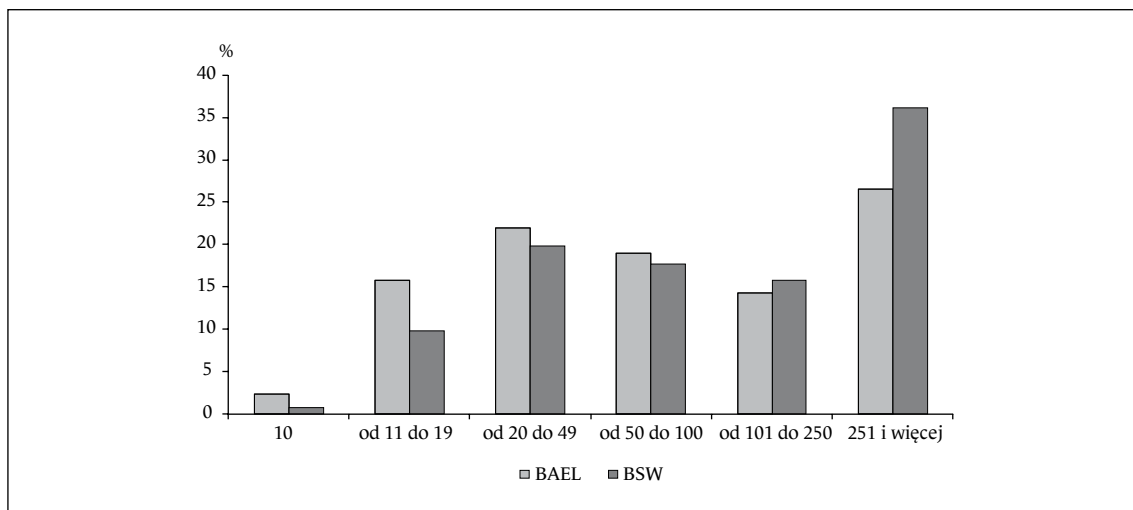


Źródło: opracowanie na podstawie BAEL i BSW.



Wykres 3

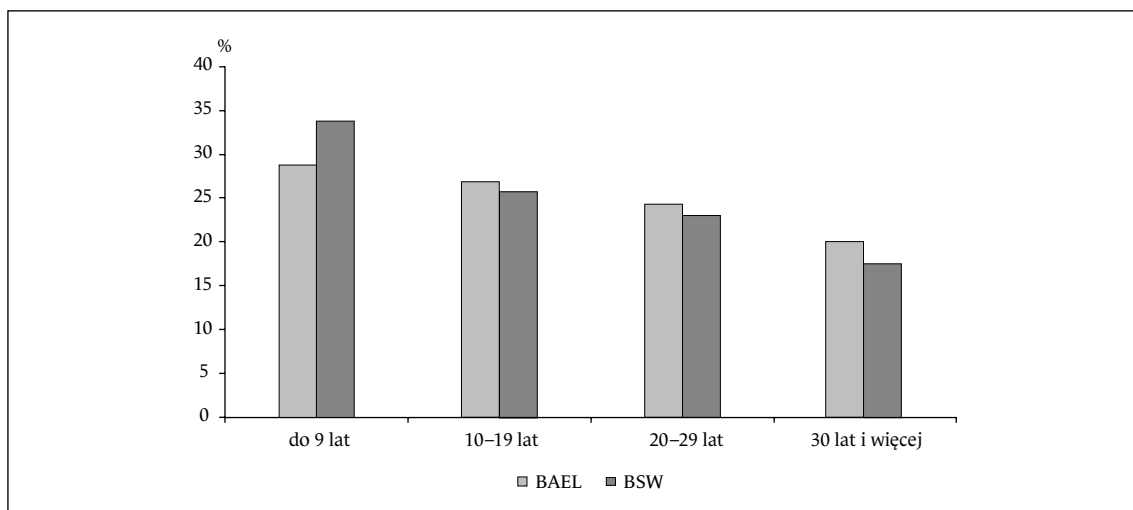
Struktura według wielkości przedsiębiorstwa: BAEL a BSW



Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 4

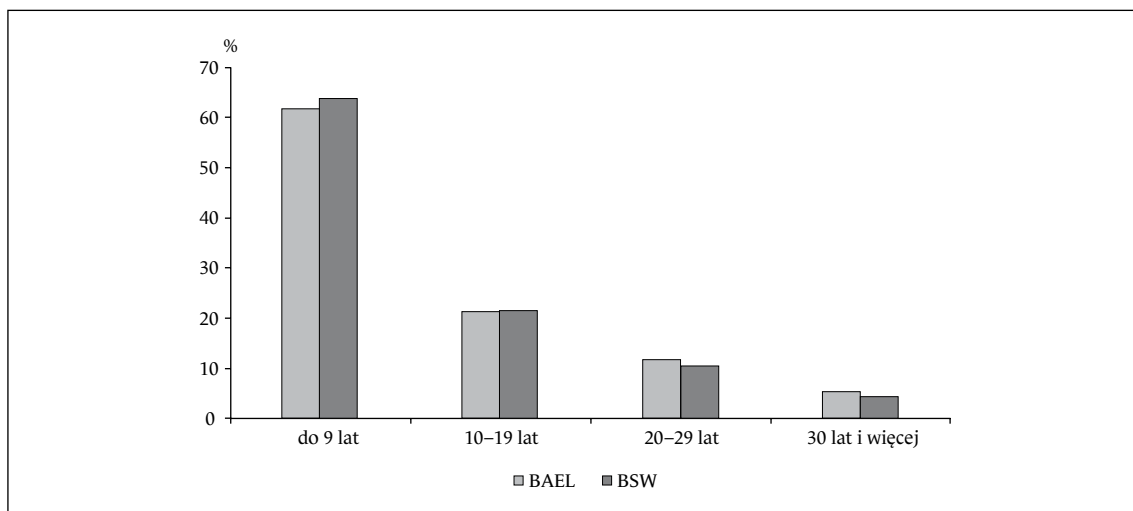
Struktura według stażu pracy ogółem: BAEL a BSW



Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 5

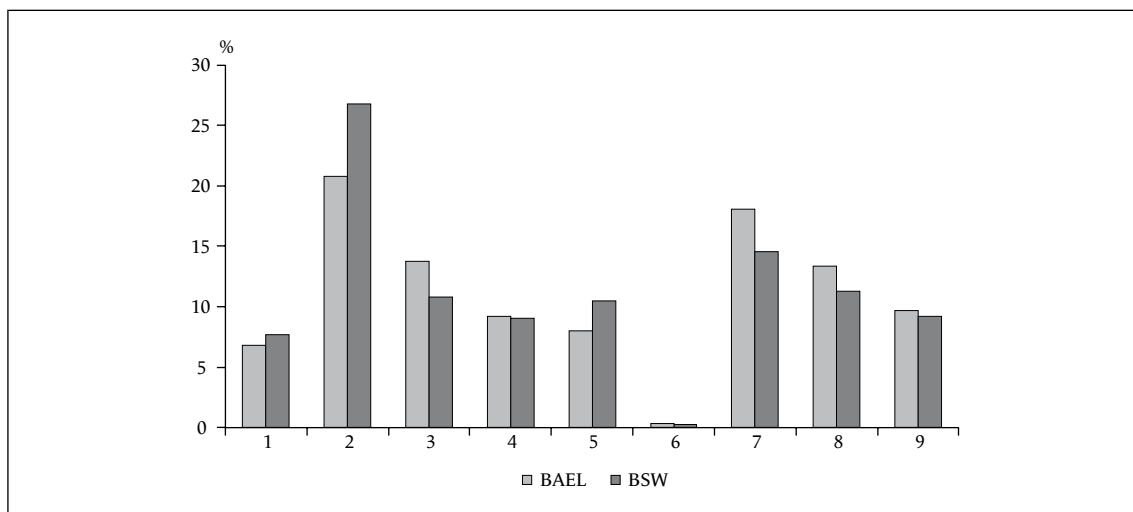
Struktura według stażu pracy w miejscu zatrudnienia: BAEL a BSW



Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 6

Struktura według wielkich grup zawodowych: BAEL a BSW

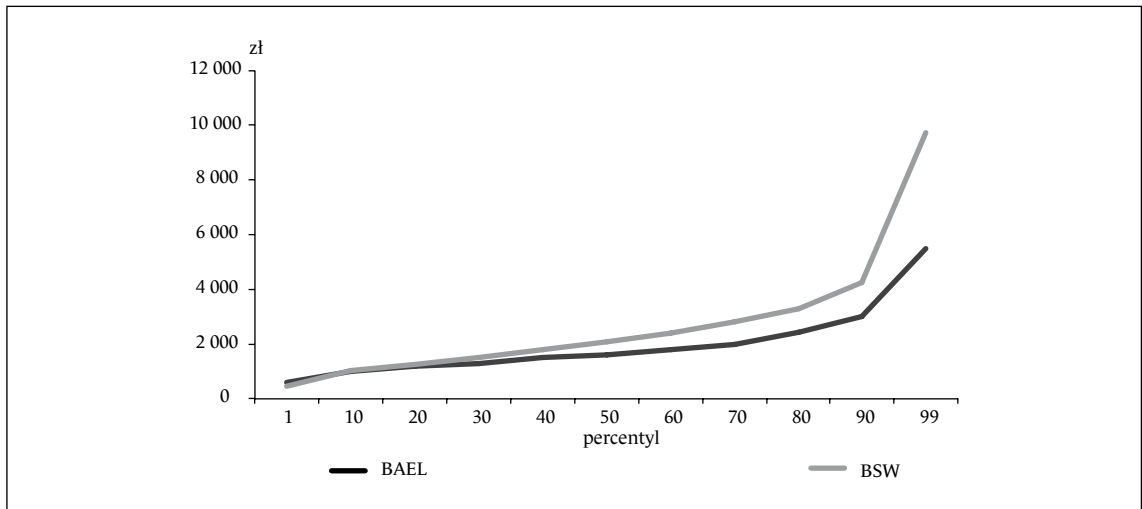


1 – przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy, 2 – specjaliści, 3 – technicy i inny średni personel, 4 – pracownicy biurowi, 5 – pracownicy usług i sprzedawcy, 6 – rolnicy, ogrodnicy, leśnicy i rybacy, 7 – robotnicy przemysłowi i rzemieślnicy, 8 – operatorzy i monterzy maszyn i urządzeń, 9 – pracownicy przy pracach prostych, 10 – siły zbrojne.

Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 7

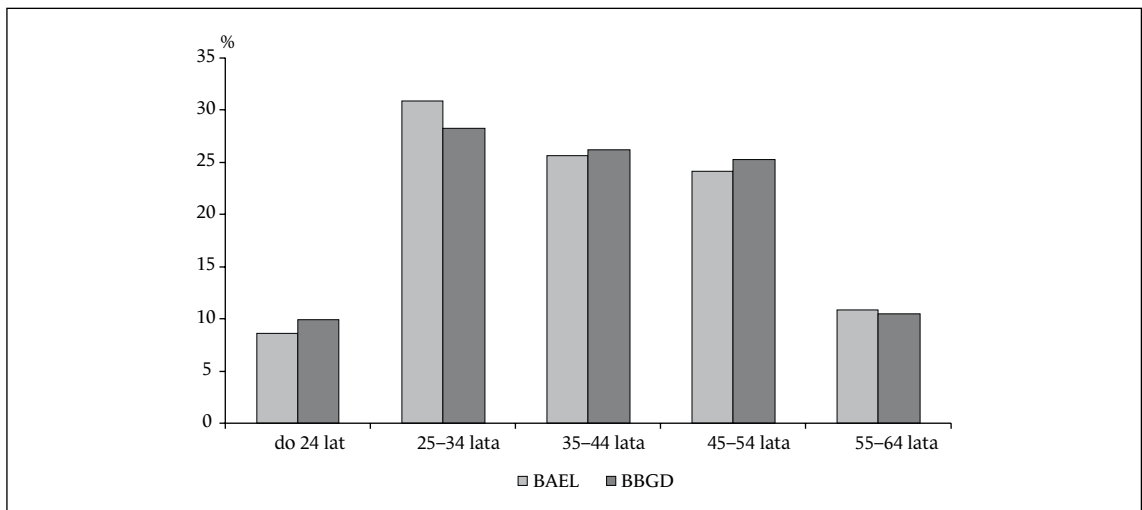
Wynagrodzenia respondentów: BAEL a BSW



Źródło: opracowanie na podstawie BAEL i BSW.

Wykres 8

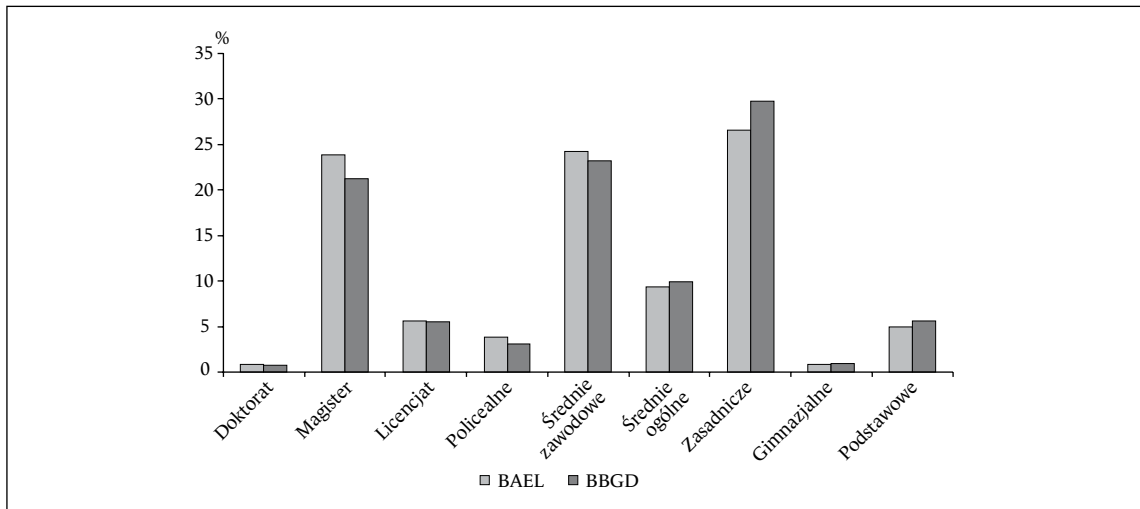
Struktura według wieku respondentów: BAEL a BBGD



Źródło: opracowanie na podstawie BAEL i BBGD.

Wykres 9

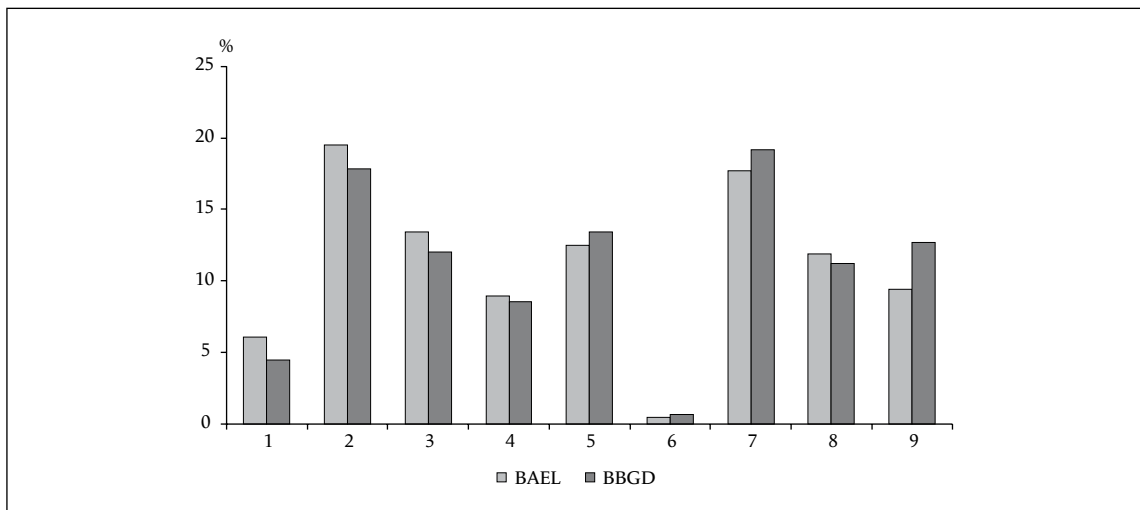
Struktura według wykształcenia respondentów: BAEL a BBGD



Źródło: opracowanie na podstawie BAEL i BBGD.

Wykres 10

Struktura według wielkich grup zawodowych: BAEL a BBGD

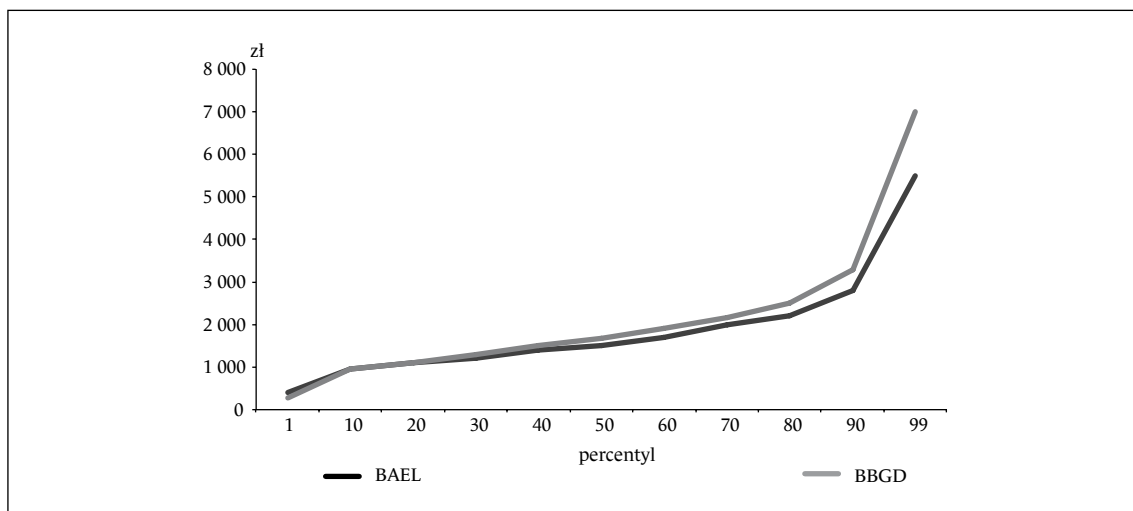


1 – przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy, 2 – specjaliści, 3 – technicy i inny średni personel, 4 – pracownicy biurowi, 5 – pracownicy usług i sprzedawcy, 6 – rolnicy, ogrodnicy, leśnicy i rybacy, 7 – robotnicy przemysłowi i rzemieślnicy, 8 – operatorzy i monterzy maszyn i urządzeń, 9 – pracownicy przy pracach prostych, 10 – siły zbrojne.

Źródło: opracowanie na podstawie BAEL i BBGD.

Wykres 11

Wynagrodzenia respondentów: BAEL a BBGD



Źródło: opracowanie na podstawie BAEL i BBGD.

Tabela 1

Modele dla liczby lat edukacji: BAEL a BSW

Zmienna	BAEL	BAEL P	BAEL D	BAEL DP	BSW D	BSW DP
Lata edukacji	0,050***	0,046***	0,040***	0,044***	0,042***	0,051***
Doświadczenie	0,029***	0,035***	0,021***	0,032***	0,027***	0,034***
Doświadczenie <sup>2</sup>	-0,059***	-0,090***	-0,041***	-0,082***	-0,052***	-0,073***
Kobieta	-0,255***	-0,289***	-0,260***	-0,290***	-0,193***	-0,221***
Grupa 2	-0,225***	-0,204***	-0,206***	-0,197***	-0,424***	-0,423***
Grupa 3	-0,303***	-0,290***	-0,303***	-0,277***	-0,462***	-0,447***
Grupa 4	-0,470***	-0,478***	-0,463***	-0,441***	-0,574***	-0,578***
Grupa 5	-0,634***	-0,612***	-0,602***	-0,570***	-0,875***	-0,824***
Grupa 6	-0,631***	-0,684***	-0,624***	-0,666***	-0,856***	-0,881***
Grupa 7	-0,502***	-0,506***	-0,519***	-0,498***	-0,691***	-0,674***
Grupa 8	-0,454***	-0,467***	-0,487***	-0,464***	-0,602***	-0,587***
Grupa 9	-0,706***	-0,683***	-0,718***	-0,660***	-0,900***	-0,855***
Stała	7,025***	7,092***	7,259***	7,156***	7,467***	7,301***
N	10 590	5 555	7 537	4 031	688 383	394 179
R <sup>2</sup>	0,349	0,360	0,376	0,367	0,278	0,242

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do osób w wieku 25–44 lata; D – próba ograniczona do dużych firm.

\* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01

Tabela 2

Modele dla liczby lat w edukacji: BAEL a BBGD

Zmienna	BAEL	BAEL P	BBGD	BBGD P
Lata edukacji	0,050***	0,046***	0,054***	0,053***
Doświadczenie	0,029***	0,035***	0,040***	0,034***
Doświadczenie <sup>2</sup>	-0,059***	-0,090***	-0,084***	-0,074***
Kobieta	-0,255***	-0,289***	-0,268***	-0,294***
Grupa 2	-0,225***	-0,204***	-0,231***	-0,208***
Grupa 3	-0,303***	-0,290***	-0,304***	-0,286***
Grupa 4	-0,470***	-0,478***	-0,433***	-0,444***
Grupa 5	-0,634***	-0,612***	-0,622***	-0,594***
Grupa 6	-0,631***	-0,684***	-0,942***	-0,963***
Grupa 7	-0,502***	-0,506***	-0,534***	-0,530***
Grupa 8	-0,454***	-0,467***	-0,450***	-0,455***
Grupa 9	-0,706***	-0,683***	-0,773***	-0,749***
Stała	7,025***	7,092***	6,873***	6,960***
N	10 590	5 555	32 197	18 003
R <sup>2</sup>	0,349	0,360	0,339	0,306

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

\* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01.

Tabela 3

Modele dla poziomów wykształcenia: BAEL a BSW

Zmienna	BAEL	BAEL P	BAEL D	BAEL DP	BSW	BSW P
Magister	-0,319***	-0,554***	-0,355***	-0,584***	-0,054***	0,001
Licencjat	-0,378***	-0,613***	-0,411***	-0,634***	-0,140***	-0,083***
Policealne	-0,567***	-0,831***	-0,619***	-0,859***	-0,287***	-0,246***
Średnie zawodowe	-0,553***	-0,794***	-0,585***	-0,817***	-0,317***	-0,295***
Średnie ogólne	-0,561***	-0,782***	-0,580***	-0,805***	-0,321***	-0,298***
Zasadnicze	-0,639***	-0,870***	-0,663***	-0,886***	-0,421***	-0,397***
Gimnazjalne	-1,229***	-0,890***	-0,668***	-1,062***	-0,412***	(brak)
Podstawowe	-0,719***	-0,977***	-0,720***	-0,953***	-0,413***	-0,439***
Doświadczenie	0,029***	0,037***	0,024***	0,035***	0,031***	0,038***
Doświadczenie <sup>2</sup>	-0,060***	-0,096***	-0,046***	-0,091***	-0,058***	-0,084***
Kobieta	-0,257***	-0,288***	-0,260***	-0,288***	-0,200***	-0,226***
Grupa 2	-0,226***	-0,209***	-0,219***	-0,206***	-0,427***	-0,422***
Grupa 3	-0,298***	-0,283***	-0,292***	-0,270***	-0,434***	-0,429***
Grupa 4	-0,458***	-0,460***	-0,441***	-0,424***	-0,538***	-0,551***
Grupa 5	-0,603***	-0,581***	-0,565***	-0,541***	-0,810***	-0,769***
Grupa 6	-0,596***	-0,639***	-0,575***	-0,622***	-0,794***	-0,819***
Grupa 7	-0,452***	-0,460***	-0,465***	-0,454***	-0,609***	-0,598***
Grupa 8	-0,419***	-0,429***	-0,439***	-0,428***	-0,523***	-0,514***
Grupa 9	-0,665***	-0,639***	-0,672***	-0,622***	-0,834***	-0,789***
Stała	8,192***	8,414***	8,287***	8,464***	8,223***	8,143***
N	10 590	5 555	7 537	4 031	688 383	394 179
R <sup>2</sup>	0,366	0,369	0,386	0,380	0,285	0,249

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata; D – próba ograniczona do dużych firm. \* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01.



Tabela 4

Modele dla poziomów wykształcenia: BAEL a BBGD

Zmienna	BAEL	BAEL P	BBGD	BBGD P
Magister	-0,319***	-0,554***	-0,212***	-0,166***
Licencjat	-0,378***	-0,613***	-0,338***	-0,306***
Policealne	-0,567***	-0,831***	-0,468***	-0,419***
Średnie zawodowe	-0,553***	-0,794***	-0,475***	-0,447***
Średnie ogólne	-0,561***	-0,782***	-0,474***	-0,450***
Zasadnicze	-0,639***	-0,870***	-0,536***	-0,513***
Gimnazjalne	-1,229***	-0,890***	-1,414***	-0,859***
Podstawowe	-0,719***	-0,977***	-0,613***	-0,649***
Doświadczenie	0,029***	0,037***	0,039***	0,036***
Doświadczenie <sup>2</sup>	-0,060***	-0,096***	-0,082***	-0,079***
Kobieta	-0,257***	-0,288***	-0,271***	-0,295***
Grupa 2	-0,226***	-0,209***	-0,233***	-0,211***
Grupa 3	-0,298***	-0,283***	-0,303***	-0,281***
Grupa 4	-0,458***	-0,460***	-0,431***	-0,432***
Grupa 5	-0,603***	-0,581***	-0,606***	-0,569***
Grupa 6	-0,596***	-0,639***	-0,919***	-0,927***
Grupa 7	-0,452***	-0,460***	-0,502***	-0,491***
Grupa 8	-0,419***	-0,429***	-0,430***	-0,421***
Grupa 9	-0,665***	-0,639***	-0,745***	-0,715***
Stała	8,192***	8,414***	8,039***	8,034***
N	10 590	5 555	32 197	18 003
R <sup>2</sup>	0,366	0,369	0,357	0,309

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

\* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01.

Tabela 5

Poziom wykształcenia: BAEL a BSW

Zmienna	BAEL	BAEL P	BSW	BSW P
Doktorat	0,663***	0,886***	0,421***	0,397***
Magister	0,308***	0,301***	0,366***	0,398***
Licencjat	0,252***	0,252***	0,281***	0,314***
Policealne	0,044*	0,026	0,134***	0,151***
Średnie zawodowe	0,078***	0,068***	0,104***	0,102***
Średnie ogólne	0,082***	0,081***	0,100***	0,100***
Gimnazjalne	-0,005	-0,176	0,009	(brak)
Podstawowe	-0,057***	-0,068***	-0,008***	-0,042***
Doświadczenie	0,024***	0,035***	0,031***	0,038***
Doświadczenie <sup>2</sup>	-0,046***	-0,091***	-0,058**	-0,084***
Kobieta	-0,260***	-0,288***	-0,200***	-0,226***
Grupa 1	0,441***	0,424***	0,538***	0,551***
Grupa 2	0,222***	0,218***	0,112***	0,129***
Grupa 3	0,149***	0,153***	0,104***	0,122***
Grupa 5	-0,124***	-0,118***	-0,271***	-0,218***
Grupa 6	-0,135***	-0,199***	-0,255***	-0,268***
Grupa 7	-0,025	-0,030	-0,070***	-0,047***
Grupa 8	0,001	-0,004	0,016***	0,037***
Grupa 9	-0,321***	-0,198***	-0,295***	-0,238***
Stała	7,184***	7,155***	7,263***	7,194***
N	10 590	5 555	688 383	349 179
R <sup>2</sup>	0,366	0,369	0,285	0,249

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

\* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01.

Tabela 6

Poziom wykształcenia: BAEL a BSW

Zmienna	BAEL	BAEL P	BBGD	BBGD P
Doktorat	0,663***	0,886***	0,536***	0,513***
Magister	0,308***	0,301***	0,324***	0,347***
Licencjat	0,252***	0,252***	0,198***	0,207***
Policealne	0,044*	0,026	0,068***	0,094***
Średnie zawodowe	0,078***	0,068***	0,061***	0,066***
Średnie ogólne	0,082***	0,081***	0,062***	0,063***
Gimnazjalne	-0,005	-0,176	-0,878***	-0,346***
Podstawowe	-0,057***	-0,068***	-0,077***	-0,136***
Doświadczenie	0,024***	0,035***	0,039***	0,036***
Doświadczenie <sup>2</sup>	-0,046***	-0,091***	-0,082***	-0,079***
Kobieta	-0,260***	-0,288***	-0,271***	-0,295***
Grupa 1	0,441***	0,424***	0,431***	0,432***
Grupa 2	0,222***	0,218***	0,197***	0,221***
Grupa 3	0,149***	0,153***	0,128***	0,151***
Grupa 5	-0,124***	-0,118***	-0,175***	-0,137***
Grupa 6	-0,135***	-0,199***	-0,489***	-0,495***
Grupa 7	-0,025	-0,030	-0,071***	0,059***
Grupa 8	0,001	-0,004	0,000	0,011
Grupa 9	-0,321***	-0,198***	-0,314***	-0,283***
Stała	7,184***	7,155***	7,072***	7,088***
N	10 590	5 555	32 197	18 003
R <sup>2</sup>	0,366	0,369	0,357	0,309

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

\* p &lt; 0,1; \*\* p &lt; 0,05; \*\*\* p &lt; 0,01.

Tabela 7

Przedziały ufności dla oszacowań parametrów

Zmienna	BAEL P		BSW P		BBGD P	
Lata edukacji	0,040	0,052	0,050	0,052	0,049	0,058
Doświadczenie	0,028	0,041	0,033	0,035	0,030	0,038
Doświadczenie <sup>2</sup>	-0,110	-0,069	-0,077	-0,070	-0,088	-0,060
Kobieta	-0,311	-0,267	-0,224	-0,217	-0,309	-0,279
Grupa 2	-0,254	-0,154	-0,430	-0,416	-0,241	-0,175
Grupa 3	-0,342	-0,238	-0,455	-0,439	-0,322	-0,251
Grupa 4	-0,533	-0,422	-0,586	-0,569	-0,482	-0,406
Grupa 5	-0,665	-0,559	-0,832	-0,815	-0,631	-0,558
Grupa 6	-0,834	-0,534	-0,922	-0,840	-1,069	-0,856
Grupa 7	-0,558	-0,455	-0,682	-0,666	-0,566	-0,494
Grupa 8	-0,521	-0,413	-0,595	-0,578	-0,492	-0,417
Grupa 9	-0,741	-0,626	-0,865	-0,846	-0,789	-0,710
Stała	6,977	7,208	7,282	7,321	6,879	7,041
N	5 555		394 197		18 003	
R <sup>2</sup>	0,369		0,249		0,309	

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

Tabela 8

Przedziały ufności dla oszacowań parametrów

Zmienna	BAEL P		BSW P		BBGD P	
Magister	-0,730	-0,379	-0,015	0,018	-0,248	-0,085
Licencjat	-0,792	-0,434	-0,101	-0,066	-0,391	-0,222
Policealne	-1,014	-0,648	-0,265	-0,228	-0,510	-0,328
Średnie zawodowe	-0,972	-0,616	-0,313	-0,278	-0,531	-0,362
Średnie ogólne	-0,963	-0,602	-0,316	-0,279	-0,536	-0,364
Zasadnicze	-1,050	-0,690	-0,415	-0,380	-0,599	-0,427
Gimnazjalne	-1,186	-0,595	(brak)		-1,113	-0,605
Podstawowe	-1,161	-0,793	-0,459	-0,419	-0,740	-0,557
Doświadczenie	0,031	0,044	0,037	0,040	0,032	0,041
Doświadczenie <sup>2</sup>	-0,117	-0,075	-0,087	-0,080	-0,094	-0,065
Kobieta	-0,309	-0,266	-0,229	-0,222	-0,310	-0,281
Grupa 2	-0,259	-0,158	-0,429	-0,415	-0,244	-0,178
Grupa 3	-0,334	-0,231	-0,437	-0,421	-0,316	-0,245
Grupa 4	-0,515	-0,405	-0,559	-0,543	-0,470	-0,394
Grupa 5	-0,635	-0,528	-0,777	-0,760	-0,607	-0,532
Grupa 6	-0,788	-0,489	-0,860	-0,778	-1,034	-0,820
Grupa 7	-0,513	-0,407	-0,607	-0,590	-0,528	-0,453
Grupa 8	-0,484	-0,374	-0,523	-0,505	-0,460	-0,382
Grupa 9	-0,698	-0,580	-0,799	-0,779	-0,756	-0,674
Stała	8,232	8,597	8,124	8,161	7,945	8,123
N	5 555		394 197		18 003	
R <sup>2</sup>	0,369		0,249		0,309	

Uwagi:

Zmienna zależna: logarytm płacy miesięcznej netto. P – próba ograniczona do wieku 25–44 lata.

## Cross-comparison of wage data from Polish survey studies

---

### Abstract

The article presents a comparative analysis of data concerning employees and their wages collected in three main socio-economic cross-sectional studies conducted in Poland. The aim is to identify possible discrepancies between datasets and to provide explanation. An additional aim is to analyze the impact of the differences between representative sets of labour market data on the estimated parameter values of economic models. Parameter estimates in an exemplary model illustrating the rate of return to education show that in the case of explanatory variables with continuous distribution slight differences in the data do not result in differences in parameter estimate values. On the other hand, in the case of discrete variables the differences in estimates may be significant.

---

**Keywords:** individual data, labour force survey, household budget survey, survey of wages and salaries, rate of return to education