

PRACE NAUKOWE  
Uniwersytetu Ekonomicznego we Wrocławiu nr 309  
RESEARCH PAPERS  
of Wrocław University of Economics No. 309

# Spółeczno-gospodarcze aspekty statystyki

Redaktorzy naukowi

**Zofia Rusnak**  
**Edyta Mazurek**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2013

Redaktor Wydawnictwa: Joanna Szynal

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

[www.ibuk.pl](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com),

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

[http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się na stronie internetowej Wydawnictwa

[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Kopiowanie i powielanie w jakiegokolwiek formie wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2013

**ISSN 1899-3192**

**ISBN 978-83-7695-398-4**

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Tadeusz Bednarski:</b> Rola Jerzego Sławy-Neymana w kształtowaniu metod statystycznej analizy przyczynowości .....	11
<b>Filip Borowicz:</b> Ocena możliwości uzupełnienia danych BAEL informacjami ze źródeł administracyjnych w celu dokładniejszej analizy danych o bezrobociu .....	19
<b>Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna:</b> Przydatność testów nieparametrycznych Kruskala-Wallisa i mediany w długoterminowej ocenie parametrów kruszyw melafirowych .....	27
<b>Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna:</b> Karty kontrolne w ocenie jakości kruszyw dla budownictwa drogowego.....	42
<b>Czesław Domański:</b> Uwagi o procedurach weryfikacji hipotez z brakującą informacją.....	54
<b>Stanisław Heilpern:</b> Zależne procesy ryzyka.....	62
<b>Artur Lipieta, Barbara Pawelek, Jadwiga Kostrzewska:</b> Badanie struktury wydatków w ramach wspólnej polityki UE z wykorzystaniem analizy korespondencji.....	78
<b>Agnieszka Marciniuk:</b> Dwa sposoby modelowania stopy procentowej w ubezpieczeniach życiowych .....	90
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> Model nieproporcjonalnej intensywności Coxa w analizie bezrobocia .....	114
<b>Edyta Mazurek:</b> Statystyczna analiza podatku dochodowego od osób fizycznych.....	127
<b>Katarzyna Ostasiewicz:</b> Awersja do nierówności w modelowaniu użytkowania dóbr wspólnych.....	159
<b>Piotr Peternek:</b> Porównanie kart kontrolnych indywidualnych pomiarów uzyskanych z wykorzystaniem uogólnionego rozkładu lambda oraz krzywych Johnsona .....	179
<b>Małgorzata Podogrodzka:</b> Starzenie się ludności a płodność w Polsce w latach 1991-2010 – ujęcie regionalne .....	192
<b>Renata Rasińska, Iwona Nowakowska:</b> Jakość życia studentów w aspekcie znajomości wskaźników zrównoważonego rozwoju .....	203

<b>Maria Rosienkiewicz, Jerzy Detyna:</b> Analiza efektywności metod wyboru zmiennych objaśniających do budowy modelu regresyjnego .....	214
<b>Jerzy Śleszyński:</b> National Welfare Index – ocena nowego miernika rozwoju trwałego i zrównoważonego .....	236
<b>Maria Szmuksta-Zawadzka, Jan Zawadzki:</b> Wykorzystanie oszczędnych modeli harmonicznych w prognozowaniu na podstawie szeregów czasowych o wysokiej częstotliwości w warunkach braku pełnej informacji.....	261
<b>Anna Zięba:</b> O możliwościach wykorzystania metod statystycznych w badaniach nad stresem .....	278

## Summaries

<b>Tadeusz Bednarski:</b> Role of Jerzy Sława-Neyman in statistical inference for causality .....	18
<b>Filip Borowicz:</b> Assessing the possibility of supplementing the Polish LFS data with register records for more detailed unemployment data analysis.	26
<b>Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna:</b> Usefulness of nonparametric Kruskal-Wallis and median tests in long-term parameters assessment of melaphyre crushed rocks .....	41
<b>Mariusz Donocik, Bogdan Kisiała, Mirosław Mróz, Beata Detyna, Jerzy Detyna:</b> Control charts in the assessment of aggregates quality for road construction.....	53
<b>Czesław Domański:</b> Some remarks on the procedures of the verification of hypotheses under incomplete information.....	61
<b>Stanisław Heilpern:</b> Dependent risk processes .....	77
<b>Artur Lipieta, Barbara Pawelek, Jadwiga Kostrzewska:</b> Study of the structure of expenditure under the EU's common policy using correspondence analysis .....	89
<b>Agnieszka Marciniuk:</b> Two ways of stochastic modelling of interest rate in life insurances .....	113
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> The Cox non-proportional hazards model in the analysis of unemployment.....	126
<b>Edyta Mazurek:</b> Statistical assessment of Personal Income Tax .....	158
<b>Katarzyna Ostasiewicz:</b> Inequality aversion in modeling the use of common pool resources .....	178
<b>Piotr Peternek:</b> Comparison of control charts of individual measurements based on general Lambda distribution and Johnson curves.....	191
<b>Małgorzata Podogrodzka:</b> The ageing of the population and fertility in Poland in the years 1991-2010 by voivodeships.....	202
<b>Renata Rasińska, Iwona Nowakowska:</b> Students' life quality in terms of knowledge of sustainable development indicators .....	213

---

<b>Maria Rosienkiewicz, Jerzy Detyna:</b> Efficiency analysis of chosen methods of explanatory variables selection within the scope of regression model construction.....	235
<b>Jerzy Śleszyński:</b> <i>National Welfare Index</i> – assessment of a new measure of sustainable development.....	260
<b>Maria Szmuksta-Zawadzka, Jan Zawadzki:</b> The application of harmonic models in forecasting based on high frequency time series in condition of lack of full information.....	277
<b>Anna Zięba:</b> About statistical methods in the study on stress .....	284

**Filip Borowicz**

Uniwersytet Wrocławski

---

## **OCENA MOŻLIWOŚCI UZUPEŁNIENIA DANYCH BAEL INFORMACJAMI ZE ŹRÓDEŁ ADMINISTRACYJNYCH W CELU DOKŁADNIEJSZEJ ANALIZY DANYCH O BEZROBOCIU**

---

**Streszczenie:** Skuteczność przeciwdziałania bezrobociu jest uwarunkowana między innymi wiarygodną analizą czynników wpływających na czas trwania bezrobocia. W sondażowych badaniach rynku pracy poziom braków odpowiedzi jest niejednokrotnie wysoki. Ponadto badania takie są narażone na „wyczerpywanie się” danych, ponieważ część jednostek rezygnuje z udziału w trakcie cyklu badawczego. Niekompletność próby nie zawsze skutkuje obciążonością wnioskowania statystycznego. Zależy to w szczególności od wpływu posiadania zatrudnienia na prawdopodobieństwo odmowy i jest to tzw. efekt przyczynowy [van den Berg i in. 2006]. W pracy omawia się wybrane badania obciążoności spowodowanej absencją respondentów, wykorzystujące dane ankietowe uzupełnione informacjami administracyjnymi, oraz rozważa się możliwość jednoczesnego wykorzystania danych zbieranych przez GUS i urzędy pracy.

**Słowa kluczowe:** analiza danych bezrobocia, analiza braków odpowiedzi, bezrobocie rejestrowane, BAEL.

### **1. Wstęp**

Dane o rynku pracy należą do najbardziej skrupulatnie gromadzonych informacji gospodarczych. Obecnie podstawowym źródłem danych o aktywności ekonomicznej ludności są wszelkiego rodzaju reprezentatywne badania gospodarstw domowych. Inne źródła danych, takie jak rejestry administracyjne czy dane spisowe, nie dają wystarczająco kompletnego obrazu rynku pracy. W przypadku danych pochodzących ze spisów zakres zebranych danych wydaje się w pełni wystarczający, problemem jest tutaj częstotliwość odbywania się tego typu badań. Zmiany na rynku pracy mogą występować w sposób gwałtowny i nieregularny, więc są trudne do wychwycenia w badaniach spisowych odbywających się z reguły raz na kilka lub kilkanaście lat. Dane gromadzone przez urzędy pracy mimo stałej aktualizacji obejmują swoim zakresem tylko część populacji i nie są wystarczającym źródłem informacji do wielopłaszczyznowej analizy rynku pracy. Spowodowane to jest między

innymi przez zależność bezrobocia rejestrowanego od zmieniających się w czasie kryteriów i przepisów. W szczególności z tych względów współczesne badania rynku pracy są realizowane jako odrębne, specjalne programy, służące tylko do analizy aktywności ekonomicznej ludności. W badaniach tych na ogół obserwuje się wybraną drogą losowania próbę osób za pomocą metody kilkukrotnego ankietowania każdej jednostki w ustalonych interwałach czasowych. Osoby włączone do takiego badania podczas cyklicznie odbywających się tzw. wywiadów określają między innymi swój aktualny status na rynku pracy. Zgodnie z rekomendacjami Międzynarodowej Organizacji Pracy [Husmanns i in. 1990; Eurostat 2004, 2012] rozważa się tutaj, czy respondent w chwili badania posiada pracę, czy jest osobą bezrobotną aktywnie poszukującą zatrudnienia bądź czy jest bierny zawodowo, tzn. nie pracuje i nie podejmuje żadnych działań w celu znalezienia pracy. W założeniach metoda kilkukrotnego ankietowania ma w szczególności umożliwić poznanie rzeczywistego czasu, w jaki badany znalazł pracę. Określenie rzeczywistego czasu poszukiwania pracy jest możliwe tylko w przypadku, gdy podczas trwania badania osoba bezrobotna znalazła zatrudnienie i fakt ten został zarejestrowany w trakcie ankietowania. Możliwych jest jednak wiele sytuacji, w których dysponujemy tylko niepełnymi informacjami o czasie poszukiwania pracy. Monitorowana osoba może w trakcie trwania badania nie znaleźć zatrudnienia, może zrezygnować z udziału w badaniu bądź z innych powodów kontakt z ankietowanym staje się niemożliwy.

Problem braku udziału respondentów w ankietowanych badaniach rynku pracy jest od wielu lat przedmiotem licznych opracowań, analiz i badań naukowych [Romeo 1997; O'Muirheartaigh, Campanelli 1999; Little, Rubin 2002; van den Berg i in. 1994, 2006; Groves 2006, Pyy-Martikainen, Rendtel 2008]. Niekompletność próby spowodowana brakiem odpowiedzi może mieć wpływ na obciążoność wnioskowania statystycznego w sytuacji, gdy fakt znalezienia zatrudnienia ma wpływ na szansę odmowy udziału w badaniach [van den Berg i in. 2006; Bednarski, Borowicz 2010, Bednarski 2013]. Związek taki w pracy van den Berga [2006] został nazwany efektem przyczynowym. Natomiast gdy niemożność kontaktu z ankietowanym nie jest bezpośrednio wynikiem znalezienia przez niego zatrudnienia, to wtedy mówimy o tzw. efekcie selektywnym [van den Berg i in. 2006]. W sytuacji gdy wywiad ankietowy nie został przeprowadzony, po pierwsze, rzadko znamy rzeczywistą przyczynę absencji, po drugie, nie dysponujemy wiedzą na temat dalszych losów badanego na rynku pracy. W ograniczonym stopniu podczas ankietowych badań rynku pracy są gromadzone dane na temat przyczyn braku udziału. Jednak zwykle są to tylko ogólne informacje, np. zmiana miejsca zamieszkania, brak osiągalności lub po prostu odmowa. Szczegółowa analiza takich informacji w małym stopniu może być wykorzystana do oceny występowania efektu przyczynowego i określenia skali obciążoności wnioskowania statystycznego [van den Berg i in. 1994, 2006]. Z tego też względu w analizie problemu absencji stosuje się między innymi dane zebrane podczas reprezentatywnych badań ankietowych, uzupełnione informacjami z innych źródeł, np. z rejestrów administracyjnych. Zestawienie danych z kilku źródeł dla tych samych jednostek umożliwia określenie losów monitorowanych osób na rynku

pracy pomimo braku ich udziału w badaniu ankietowym. W szczególności możemy stwierdzić, czy osoba, która we wcześniejszych wywiadach była zarejestrowana jako aktywnie poszukująca pracę, w momencie nieudanej próby kontaktu posiadała zatrudnienie. Uzupełnienie danych pochodzących z reprezentatywnych badań rynku pracy innymi źródłami danych wymaga połączenia dwóch lub większej liczby baz danych na poziomie jednostkowym. W praktyce może być to trudne, gdyż wymusza użycie spójnego i jednoznacznego sposobu identyfikacji pojedynczej jednostki we wszystkich bazach. Proces gromadzenia i przetwarzania danych przez różne instytucje odbywa się zwykle niezależnie, każda z instytucji może stosować własne rozwiązania, identyfikatory czy definicje, co może powodować niespójność, np. w kontekście określenia statusu danej osoby na rynku pracy.

Problem brakujących danych jest również powszechny w statystyce publicznej i w szczególności dotyczy zagadnień ze statystyki małych obszarów, tam także pojawia się możliwość estymacji w oparciu o dane pochodzące z kilku źródeł. Przykładem są tutaj badania prowadzone przez konsorcjum EURAREA [Project 2004] oraz praca *Estymacja pośrednia na lokalnym rynku pracy* [Gołata 2004], gdzie używa się informacji zagregowanych pochodzących z kilku źródeł do estymacji bezrobocia na lokalnym rynku pracy. Prezentowane w niniejszej pracy zagadnienia dotyczą jednak danych jednostkowych, których używa się do analizy mechanizmów wpływających na stopę bezrobocia.

## **2. Przykłady zastosowania danych z kilku źródeł w badaniach rynku pracy**

Dane pochodzące z kilku źródeł i obejmujące te same jednostki zostały wykorzystane m.in. przez van den Berga, Lindebooma i Doltona [2006] w szczegółowej analizie problemu absencji w ankietowych badaniach rynku. W opracowaniu tym zostały użyte dane dotyczące brytyjskiego rynku pracy z przełomu lat osiemdziesiątych i dziewięćdziesiątych XX wieku. Pierwszym rozpatrywanym źródłem informacji były dane zebrane przez Policy Studies Institut w trakcie oceny skuteczności wprowadzonego w Wielkiej Brytanii w 1987 roku programu pomocy dla bezrobotnych „Restart”. Jednym z założeń udziału w tym programie było obowiązkowe spotkanie każdego bezrobotnego z konsultantem z urzędu pracy po sześciu miesiącach od utraty zatrudnienia. Ocena skuteczności programu „Restart” została przeprowadzona w marcu i kwietniu 1989 roku na podstawie próby losowej około 9 tys. bezrobotnych. Podstawowe cechy o charakterze demograficzno-społecznym opisujące każdą osobę włączoną do badania zostały zaczerpnięte z danych rejestrowanych przez urzędy pracy. Ponadto wykorzystano dwa dodatkowe źródła: Joint Unemployment and Vacancies Operating System (JOVOS) i National Online Manpower System (NOMIS), które umożliwiły oszacowanie między innymi rzeczywistego czasu poszukiwania pracy dla wszystkich osób włączonych do badania skuteczności programu „Restart”. W przeprowadzonej analizie problemu absencji respondentów wykorzystano rów-



niez dane zebrane w ramach badania ankietowego przeprowadzonego na tej samej grupie bezrobotnych, która została zakwalifikowana do badania kontrolnego programu „Restart”. Badanie to zostało przeprowadzone przez instytucję badawczą Social and Community Planning około sześć miesięcy po zaplanowanych spotkaniach bezrobotnych z próby z konsultantami z urzędu pracy. Z całej próby 9 tys. bezrobotnych tylko około 58% osób wzięło udział w tym badaniu, a więc poziom odmów był bliski 40%. W pracy podkreśla się, że ankieterzy nie poprzestawali na jednej próbie kontaktu. Próby skontaktowania się z badanymi odbywały się wielokrotnie, w różnych dniach tygodnia i o różnych godzinach. W oparciu o powyższe zbiory danych autorzy mieli do dyspozycji dla wszystkich osób z próby dane na temat rzeczywistego czasu poszukiwania pracy, ponadto dysponowali wiedzą na temat udziału lub jego braku w badaniu ankietowym. Taki zasób danych umożliwił między innymi szczegółową analizę porównawczą rozkładu czasu poszukiwania pracy przez respondentów, którzy wzięli udział w badaniu, z grupą osób, które nie partycypowały w wywiadach ankietowych. Rozbieżności szczególnie były widoczne w momencie odpowiadającym dacie przeprowadzenia wywiadu ankietowego, co sugeruje, według autorów badania, że część osób, które nie wzięły udziału w badaniu ankietowym, w momencie próby kontaktu posiadało już zatrudnienie. Autorzy proponują metodę umożliwiającą identyfikację występowania efektu przyczynowego oraz wykazując, że w przypadku analizowanych przez nich danych fakt znalezienia zatrudnienia jest związany z szansą odmowy udziału w badaniu i ma to bezpośredni wpływ na obciążoność wnioskowania statystycznego.

Dane jednostkowe pochodzące z wielu źródeł zostały również wykorzystane przez Pyy-Martikainen i Rendtela [2008] w badaniach wpływu absencji respondentów w ankietowych badaniach rynku pracy na obciążoność wnioskowania statystycznego w szczególności w kontekście modelu Coxa. Przedstawione w opracowaniach rezultaty zostały oparte na analizie danych rynku pracy w Finlandii z końca lat dziewięćdziesiątych XX wieku. W badaniach wykorzystano rejestry urzędów pracy oraz informacje z badania ankietowego European Community Household Panel (ECHP) w Finlandii. Czas poszukiwania pracy i podstawowe cechy demograficzno-społeczne każdej osoby z analizowanego zbioru danych zostały zaczerpnięte z danych urzędów pracy. Dodatkowo do danych dołączono informacje o udziale bądź odmowie w badaniu ankietowym EHCP. Połączenie obydwu źródeł danych zostało dokonane za pomocą numeru ewidencyjnego stosowanego na terenie Finlandii. Badania Pyy-Martikainen i Rendtela zostały w głównej mierze oparte na analizie porównawczej wyników estymacji dla trzech następujących grup badanych: (I) z pełną informacją o czasie poszukiwania pracy, gdzie rzeczywisty czas poszukiwania pracy został wyznaczony na podstawie danych rejestrowanych; (II) z częściową informacją o czasie poszukiwania pracy – z grupy osób z pełną informacją o czasie poszukiwania pracy zostały wyłączone jednostki, z którymi nigdy nie został nawiązany kontakt w trakcie trwania badania ankietowego; (III) z zaobserwowanym czasem poszukiwania pracy – dla grupy osób z częściową informacją rzeczywisty czas po-

szukiwania pracy został zastąpiony czasem zaobserwowanym w trakcie trwania badania ankietowego (rzeczywistym bądź cenzurowanym). Na podstawie porównania wyników estymacji dla wymienionych grup autorzy wykazują, że absencja respondentów ma istotny wpływ na jakość statystycznej analizy danych rynku pracy.

### 3. Źródła danych o rynku pracy w Polsce

Podstawowymi źródłami informacji o polskim rynku pracy są dane gromadzone przez Główny Urząd Statystyczny w ramach Badania Aktywności Ekonomicznej Ludności (BAEL) oraz rejestry powiatowych urzędów pracy (PUP). Mimo że źródła te stanowią ogromny zasób informacji o osobach bezrobotnych, skojarzenie danych jednostkowych z obydwu jednocześnie w kontekście np. statystycznej analizy danych rynku pracy jest bardzo utrudnione. Jednym z problemów jednoczesnego wykorzystania danych jednostkowych jest odmienna definicja osoby bezrobotnej stosowana w BAEL i PUP. W przypadku badania GUS definicja ta jest zgodna z zaleceniami Międzynarodowej Organizacji Pracy oraz Eurostatu [Husmanns i in. 1990; Eurostat 2004; ONZ 2009; GUS 2012], według których przez bezrobotnego rozumie się osobę w wieku 15-74 lata, która spełnia jednocześnie trzy warunki:

- 1) w okresie badanego tygodnia nie była osobą pracującą,
- 2) aktywnie poszukiwała pracy,
- 3) była gotowa podjąć pracę w ciągu dwóch tygodni następujących po tygodniu badania.

Do bezrobotnych zalicza się również osoby, które nie poszukiwały pracy, ponieważ miały pracę załatwioną i oczekiwały na jej rozpoczęcie przez okres nie dłuższy niż 3 miesiące oraz były gotowe podjąć tę pracę.

W urzędach pracy stosuje się odmienną definicję osoby bezrobotnej, która jest określona w ustawie z dnia 20 IV 2004 roku o promocji zatrudnienia i instytucjach rynku pracy [DzU z 2004 r. nr 99, poz. 1001 z późniejszymi zmianami]. Według tej ustawy do bezrobotnych zarejestrowanych zalicza się osoby niezatrudnione i niewykonyjące innej pracy zarobkowej, zdolne i gotowe do podjęcia zatrudnienia w pełnym wymiarze czasu pracy (bądź jeśli są to osoby niepełnosprawne – zdolne i gotowe do podjęcia zatrudnienia co najmniej w połowie tego wymiaru czasu pracy), nieuczące się w szkole, z wyjątkiem szkół dla dorosłych lub szkół wyższych w systemie wieczorowym albo zaocznym, zarejestrowane we właściwym dla miejsca zameldowania (stałego lub czasowego) powiatowym urzędzie pracy oraz poszukujące zatrudnienia lub innej pracy zarobkowej, jeżeli m.in.:

- ukończyły 18 lat i jednocześnie nie ukończyły: kobiety – 60 lat, a mężczyźni – 65 lat,
- nie nabyły prawa do emerytury lub renty z tytułu niezdolności do pracy, renty szkoleniowej, renty socjalnej, nie pobierają świadczenia rehabilitacyjnego, świadczenia lub zasiłku przedemerytalnego, zasiłku chorobowego lub macierzyńskiego,

- nie są właścicielami lub posiadaczami (samoistnymi lub zależnymi) nieruchomości rolnej o powierzchni użytków rolnych powyżej 2 ha przeliczeniowych, nie podlegają ubezpieczeniu emerytalnemu i rentowemu z tytułu stałej pracy jako współmałżonek lub domownik w gospodarstwie rolnym o analogicznej powierzchni,
- nie uzyskują miesięcznego przychodu w wysokości przekraczającej połowę minimalnego wynagrodzenia za pracę, z wyłączeniem przychodów od środków pieniężnych zgromadzonych na rachunkach bankowych,
- nie pobierają, na podstawie przepisów o świadczeniach rodzinnych, świadczenia pielęgnacyjnego lub dodatku do zasiłku rodzinnego z tytułu samotnego wychowywania dziecka.

Wykorzystanie danych BAEL uzupełnionych informacjami z urzędów pracy wymaga użycia jednoznacznego identyfikatora, umożliwiającego na poziomie jednostkowym połączenie obydwu zbiorów danych. W przypadku danych administracyjnych naturalnym identyfikatorem jest numer PESEL, który podaje każdy bezrobotny podczas rejestracji w powiatowym urzędzie pracy. W badaniach BAEL nie wykorzystuje się numeru ewidencyjnego PESEL tylko identyfikatory specjalnie stworzone do celów badania. Są to wielocyfrowe kody służące do identyfikacji konkretnej osoby w gospodarstwie domowym zawierające symbole województw, powiatów, oddziałów terenowych, rejonów spisowych itp. Stworzenie odpowiedniego spójnego identyfikatora dla danych jednostkowych z BAEL i z rejestrów administracyjnych wymagałoby sięgnięcia do podstawowych danych osobowych. Informacje takie w przypadku BAEL nie są częściowo zbierane lub ich uzyskanie jest bardzo ograniczone w związku z ustawą o statystyce, w myśl której:

*Zbierane i gromadzone w badaniach statystycznych statystyki publicznej dane jednostkowe (dane osobowe dające się powiązać z konkretną osobą fizyczną i dane indywidualne dające się powiązać z podmiotem gospodarczym) są poufne i podlegają szczególnej ochronie. Mogą być wykorzystywane wyłącznie do opracowań, zestawień i analiz statystycznych (tajemnica statystyczna — art. 10 i art. 38 Ustawy z dnia 29 czerwca 1995 r. o statystyce publicznej).*

Kolejną przeszkodą w jednoczesnym wykorzystaniu danych GUS i PUP jest reprezentatywność badana BAEL tylko w odniesieniu do poszczególnych województw, z tego względu według GUS wyodrębnianie danych dla mniejszych jednostek podziału administracyjnego, np. powiatów, jest niemożliwe. Z kolei urzędy pracy zbierają dane jednostkowe w skali powiatów, które są następnie przekazywane do wojewódzkich urzędów pracy tylko w postaci zagregowanej. Przeprowadzona analiza danych BAEL dla województwa dolnośląskiego z lat 2006-2009 oraz rejestrów Powiatowego Urzędu Pracy we Wrocławiu z lat 2006-2010 wykazała, że informacje obecnie udostępniane przez Główny Urząd Statystyczny są niewystarczające do jednoznacznego połączenia obydwu zbiorów na poziomie jednostkowym. W badaniach wykorzystano takie informacje z BAEL, jak: rok urodzenia, płeć, po-

ziom wykształcenia, klasa miejscowości zamieszkania oraz wskaźnik zarejestrowania w powiatowym urzędzie pracy. Dane te zostały zestawione z odpowiednimi informacjami z Powiatowego Urzędu Pracy we Wrocławiu z uwzględnieniem daty ankietowania w badaniu BAEL. W rezultacie każdej jednostce z BAEL odpowiadało od kilku do kilkunastu osób z rejestrów urzędów pracy. Uzyskanie pełnej jednoznaczności wymagałoby użycia bardziej szczegółowych informacji, do których dostęp jest organiczny w myśl tajemnicy statystycznej. W przytoczonych przykładach badanie skojarzenie danych jednostkowych odbyło się za pomocą jednoznacznego identyfikatora (numer ewidencyjny, numer ubezpieczenia społecznego). Nie są znane powody, dla których w tych przypadkach nie natrafiono na ograniczenia wynikające z ochrony danych osobowych. Mogło to się odbyć w taki sposób, że udostępniono dane już skojarzone, bez podania szczegółowych informacji osobowych. Taka metoda wydaje się być również najskuteczniejsza w odniesieniu do danych Głównego Urzędu Statystycznego i rejestrów powiatowych urzędów pracy.

Problem wykorzystania informacji o rynku pracy pochodzących z kilku obecnie dostępnych źródeł danych w Polsce został poruszony w opracowaniach Ministerstwa Pracy i Polityki Społecznej [Liwiński, Sztanderska 2006; Góra, Sztanderska 2006], w których podkreśla się, że w Polsce gromadzi się wiele informacji, przydatnych do oceny sytuacji na rynku pracy, ale zdecydowanym problemem jest ich spójność, reprezentatywność oraz częstotliwość uzyskiwania. W wymienionych publikacjach Ministerstwa Pracy i Polityki Społecznej proponuje się stworzenie systemów uzupełniających się wzajemnie podstawowych analiz funkcjonowania rynku pracy, opartych na danych pochodzących z rejestrów administracyjnych, z badań prowadzonych przez GUS czy też instytucje edukacyjne.

## Literatura

- Aktywność ekonomiczna ludności polski – I kwartał 2012*, Główny Urząd Statystyczny, Warszawa 2012.
- Bednarski T., Borowicz F., *Analysis of non-response causality in labor market surveys*, Acta Universitatis Lodzianensis, Folia Oeconomica 253, Łódź 2010, ss. 217-224.
- Bednarski T., *On Robust Causality Nonresponse Testing in Duration Studies under the Cox Model*, Statistical Papers, DOI 10.1007/s00362-013-0523-0, 2013.
- Gołata E., *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, Prace Habilitacyjne nr 11, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2004.
- Góra M., Sztanderska U., *Wprowadzenie do analizy lokalnego rynku pracy – Przewodnik*, Ministerstwo Pracy i Polityki Społecznej, Warszawa 2006.
- Groves R., *Nonresponse Rates and Nonresponse Bias in Household*, Surveys Public Opinion Quarterly, Vol. 70, No. 5, Special Issue 2006, ss. 646-675.
- Handbook on measuring the economically active population and related characteristics in population censuses*, United Nations, New York 2009.
- Hussmanns R., Mehran F., Vijay V., *Surveys of economically active population, employment, unemployment and underemployment: An ILO manual on concepts and methods (1990 Edition)*, International Labour Office, Geneva 1990.

- Labour force survey in the acceding countries Methods and definitions – 2002*, Office for Official Publications of the European Communities, Eurostat, Luxembourg 2004.
- Labour force survey in the EU, candidate and EFTA countries Main characteristics of national surveys*, Publications Office of the European Union, Eurostat, Luxembourg 2012.
- Little R., Rubin D., *Statistical Analysis with Missing Data, 2nd Edition*, Wiley Series in Probability and Statistics, New York 2002.
- Liwiński J., Sztanderska U., *Analiza sytuacji na wybranych powiatowych rynkach pracy oraz stworzenie metodologii badania lokalnego rynku pracy w Polsce – Raport końcowy*, Ministerstwo Pracy i Polityki Społecznej, Warszawa 2006.
- O’Muircheartaigh C., Campanelli P., *A Multilevel exploration of the role of interviewers in survey non-response*, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 162 Issue 3, 1999, ss. 437-446.
- Project Reference Volume*, EURAREA Consortium, 2004.
- Pyy-Martikainen M., Rendtel U., *Assessing the impact of initial nonresponse and attrition in the analysis of unemployment duration with panel surveys*, Advances in Statistical Analysis, Vol. 92, 2008, ss. 297-318.
- Romeo C., *Measuring information loss due to inconsistencies in duration data from longitudinal surveys*, Journal of Econometrics, Vol. 78, Issue 2, 1997, ss. 159-177.
- Van den Berg G., Lindeboom M., Dolton P., *Survey nonresponse and the duration of unemployment*, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 169, Number 3, 2006, ss. 585-604.
- Van den Berg G., Lindeboom M., Ridder G., *Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour*, Journal of Applied Econometrics, Vol. 9, 1994, ss. 421-435.

## ASSESSING THE POSSIBILITY OF SUPPLEMENTING THE POLISH LFS DATA WITH REGISTER RECORDS FOR MORE DETAILED UNEMPLOYMENT DATA ANALYSIS

**Summary:** Effective unemployment prevention is conditioned by reliable recognition and analysis of factors affecting the duration of unemployment. The quality of such an analysis, in case of survey studies, depends on the unbiasedness of a random sample. In social studies, where the absence of randomly selected individuals may be very high, the condition is practically never satisfied. Large scale studies of economic activity such as Polish LFS (BAEL) are additionally affected by data attrition. Random sample incompleteness does not have to result in an biased analysis if so called causality effect [van den Berg et al. 2006] is absent – employment does not affect the chance of initial non-response and attrition. One way to assess the size of bias due to non-response and attrition is using combined survey-register data. This paper discusses the possibility of combining the Polish LFS data with register records.

**Keywords:** unemployment data analysis, non-response analysis, registered unemployment.