

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

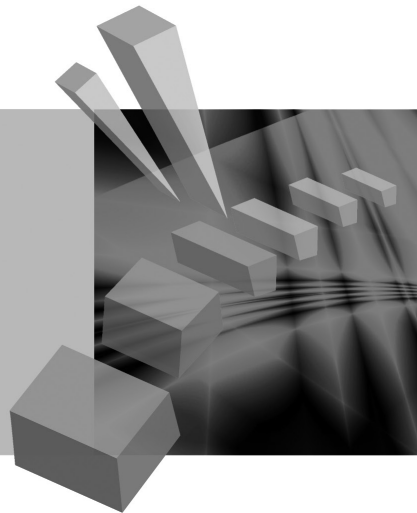
**RESEARCH PAPERS**

of Wrocław University of Economics

**279**

# Taksonomia 21

## Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

**Krzysztof Jajuga**

**Marek Walesiak**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

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)

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski:</b> Sejm VI kadencji – maszynka do głosowania .....	11
<b>Barbara Pawelek, Adam Sagan:</b> Zmienne ukryte w modelach ekonomicznych – respecyfikacja modelu Kleina I .....	19
<b>Jan Paradysz:</b> Nowe możliwości badania koniunktury na rynku pracy .....	29
<b>Krzysztof Najman:</b> Samouczące się sieci GNG w grupowaniu dynamicznym zbiorów o wysokim wymiarze .....	41
<b>Kamila Migdał-Najman:</b> Zastosowanie jednowymiarowej sieci SOM do wyboru cech zmiennych w grupowaniu dynamicznym .....	48
<b>Aleksandra Matuszewska-Janica, Dorota Witkowska:</b> Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych .....	58
<b>Iwona Foryś, Ewa Putek-Szeląg:</b> Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim .....	67
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk:</b> Klasyfikacja internetowych rachunków bankowych z uwzględnieniem zmiennych symbolicznych.....	77
<b>Marta Jaročka:</b> Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni .....	85
<b>Anna Zamojska:</b> Badanie zgodności rankingów wyznaczonych według różnych wskaźników efektywności zarządzania portfelem na przykładzie funduszy inwestycyjnych.....	95
<b>Dorota Rozmus:</b> Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i> .....	106
<b>Ewa Wędrowska:</b> Wrażliwość miar dywergencji jako mierników niepodobieństwa struktur.....	115
<b>Katarzyna Wójcik, Janusz Tuchowski:</b> Wpływ automatycznego tłumaczenia na wyniki automatycznej identyfikacji charakteru opinii konsumenckich ...	124
<b>Małgorzata Misztal:</b> Ocena wpływu wybranych metod imputacji na wyniki klasyfikacji obiektów w modelach drzew klasyfikacyjnych.....	135
<b>Anna Czapkiewicz, Beata Basiura:</b> Badanie wpływu wyboru współczynnika zależności na grupowanie szeregów czasowych .....	146
<b>Tomasz Szubert:</b> Czynniki różnicujące poziom zadowolenia z życia oraz wartości życiowe osób sprawnych i niepełnosprawnych w świetle badań „Diagnozy społecznej” .....	154

<b>Marcin Szymkowiak:</b> Konstrukcja estymatorów kalibracyjnych wartości globalnej dla różnych funkcji odległości .....	164
<b>Wojciech Roszka:</b> Szacowanie łącznych charakterystyk cech nieobserwowanych łącznie .....	174
<b>Justyna Brzezińska:</b> Metody wizualizacji danych jakościowych w programie <b>R</b> .....	182
<b>Agata Sielska:</b> Regionalne zróżnicowanie potencjału konkurencyjnego polskich gospodarstw rolnych w województwach po akcesji do Unii Europejskiej .....	191
<b>Mariusz Kubus:</b> Liniowy model prawdopodobieństwa z regularyzacją jako metoda doboru zmiennych .....	201
<b>Beata Basiura:</b> Metoda Warda w zastosowaniu klasyfikacji województw Polski z różnymi miarami odległości .....	209
<b>Katarzyna Wardzińska:</b> Wykorzystanie metody obwiedni danych w procesie klasyfikacji przedsiębiorstw .....	217
<b>Katarzyna Dębowska:</b> Modelowanie upadłości przedsiębiorstw oparte na próbach niezbilansowanych .....	226
<b>Danuta Tarka:</b> Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska ..	235
<b>Artur Czech:</b> Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim .....	246
<b>Beata Bal-Domańska:</b> Ocena relacji zachodzących między inteligentnym rozwojem a spójnością ekonomiczną w wymiarze regionalnym z wykorzystaniem modeli panelowych .....	255
<b>Mariola Chrzanowska:</b> <i>Ordinary kriging</i> i <i>inverse distance weighting</i> jako metody szacowania cen nieruchomości na przykładzie warszawskiego rynku .....	264
<b>Adam Depta:</b> Zastosowanie analizy wariancji w badaniu jakości życia na podstawie kwestionariusza SF-36v2 .....	272
<b>Maciej Beręsewicz, Tomasz Klimanek:</b> Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań .....	281
<b>Karolina Paradysz:</b> Benchmarkowa analiza estymacji dla małych obszarów na lokalnych rynkach pracy .....	291
<b>Anna Gryko-Nikitin:</b> Dobór parametrów w równoległych algorytmach genetycznych dla problemu plecakowego .....	301
<b>Tomasz Ząbkowski, Piotr Jałowiecki:</b> Zastosowanie reguł asocjacyjnych do analizy danych ankietowych w wybranych obszarach logistyki przedsiębiorstw przetwórstwa rolno-spożywczego .....	311
<b>Agnieszka Przedborska, Małgorzata Misztal:</b> Zastosowanie metod statystyki wielowymiarowej do oceny wydolności stawów kolanowych u pacjentów z chorobą zwyrodnieniową leczonych operacyjnie .....	321
<b>Dorota Perło:</b> Rozwój zrównoważony w wymiarze gospodarczym, społecznym i środowiskowym – analiza przestrzenna .....	331

<b>Ewa Putek-Szeląg, Urszula Gieraltowska, Analiza i diagnoza wielkości produkcji energii odnawialnej w Polsce na tle krajów Unii Europejskiej..</b>	342
--	-----

## Summaries

<b>Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine .....</b>	18
<b>Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model .....</b>	28
<b>Jan Paradysz: New possibilities for studying the situation on the labour market .....</b>	40
<b>Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....</b>	47
<b>Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering .....</b>	57
<b>Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....</b>	66
<b>Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...</b>	76
<b>Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables .....</b>	84
<b>Marta Jarocka: The impact of the method of the selection of diagnostic variables on the result of linear ordering on the example of ranking of universities in Poland.....</b>	94
<b>Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....</b>	105
<b>Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....</b>	114
<b>Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements .....</b>	123
<b>Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis .....</b>	134
<b>Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...</b>	145
<b>Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....</b>	153
<b>Tomasz Szubert: Factors differentiating the level of satisfaction with life and the life's values of people with and without disabilities in the light of the "Social Diagnosis" survey .....</b>	162
<b>Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures .....</b>	173

<b>Wojciech Roszka:</b> Joint characteristics' estimation of variables not jointly observed.....	181
<b>Justyna Brzezińska:</b> Visualizing categorical data in $\mathbf{R}$ .....	190
<b>Agata Sielska:</b> Regional diversity of competitiveness potential of Polish farms after the accession to the European Union .....	200
<b>Mariusz Kubus:</b> Regularized linear probability model as a filter .....	208
<b>Beata Basiura:</b> The Ward method in the application for classification of Polish voivodeships with different distances.....	216
<b>Katarzyna Wardzińska:</b> Application of Data Envelopment Analysis in company classification process.....	225
<b>Katarzyna Dębowska:</b> Modeling corporate bankruptcy based on unbalanced samples .....	234
<b>Danuta Tarka:</b> Influence of the features selection method on the results of objects classification using environmental data.....	245
<b>Artur Czech:</b> Application of chosen methods for the selection of diagnostic variables in indirect consumption research.....	254
<b>Beata Bal-Domańska:</b> Assessment of relations occurring between smart growth and economic cohesion in regional dimension using panel models	263
<b>Mariola Chrzanowska:</b> Ordinary kriging and inverse distance weighting as methods of estimating prices based on Warsaw real estate market .....	271
<b>Adam Depta:</b> Application of analysis of variance in the study of the quality of life based on questionnaire SF-36v2 .....	280
<b>Maciej Beręsewicz, Tomasz Klimanek:</b> Using indirect estimation with spatial autocorrelation in dwelling price surveys.....	290
<b>Karolina Paradysz:</b> Benchmark analysis of small area estimation on local labor markets .....	300
<b>Anna Gryko-Nikitin:</b> Selection of various parameters of parallel evolutionary algorithm for knapsack problems .....	310
<b>Tomasz Ząbkowski, Piotr Jałowiecki:</b> Application of association rules for the survey of data analysis in the selected areas of logistics in food processing companies .....	320
<b>Agnieszka Przedborska, Małgorzata Misztal:</b> Using multivariate statistical methods to assess the capacity of the knee joint among the patients treated surgically for osteoarthritis .....	330
<b>Dorota Perło:</b> Sustainable development in the economic, social and environmental dimensions – spatial analysis.....	341
<b>Ewa Putek-Szeląg, Urszula Gieraltowska:</b> Analysis and diagnosis of the volume of renewable energy production in Poland compared to EU countries .....	352

**Jan Paradysz**

Uniwersytet Ekonomiczny w Poznaniu

---

## NOWE MOŻLIWOŚCI BADANIA KONIUNKTURY NA RYNKU PRACY

---

**Streszczenie:** Ogólne liczby i wskaźniki bezrobocia, aktywnych i biernych zawodowo nie wystarczają do opisu lokalnego rynku pracy ze względu na zróżnicowanie struktur ludności według cech demograficznych. Informacje z rejestrów administracyjnych i bieżącego badania siły roboczej (w Polsce BAEL) pozwalają rozwinąć metodologię z wykorzystaniem analizy demograficznej. Ten kierunek analizy jest godny uwagi także z tego powodu, że w przypadku kobiet jednolita metodologia umożliwia śledzenie trzech karier: edukacyjnej, zawodowej i rodzinnej. W artykule przedstawiono problem translacji demograficznej, czyli wzajemne relacje między analizą kohortową i transwersalną na przykładzie wybranych krajów Ameryki i Europy oraz wskazano na nowe możliwości interpretacji wskaźników reprodukcji ludności i możliwości ich aplikacji w analizie koniunktury ekonomicznej.

**Słowa kluczowe:** analiza demograficzna, analiza kohortowa, teoria translacji, rynek pracy, aktywność ekonomiczna ludności.

### 1. Wstęp

Wśród determinant rynku pracy do ważniejszych zaliczymy osobowe cechy pracowników (płeć, wiek, przygotowanie zawodowe, staż) i szeroko rozumianą koniunkturę. Postaramy się tutaj pokazać, że wszystkie te czynniki jednocześnie uwzględnia analiza demograficzna. Wykorzystanie tej metodologii przy badaniu rynku pracy powinno przyczynić się do polepszenia oceny bieżącej koniunktury gospodarczej, jak również zaowocować precyzyjniejszymi prognozami. To nowe podejście do analizy rynku pracy prezentujemy na poziomie ogólnokrajowym tylko ze względu na dostępność długiej serii danych w ujęciu kohortowym. Jednakże w przekrojach przestrzennych teoria translacji nabiera szczególnego znaczenia, gdzie międzynarodowe i międzyregionalne przepływy siły roboczej mają dodatkowy wpływ na zachowania generacyjne (ujęcie kohortowe) oraz koniunkturę gospodarczą krajów i regionów w ujęciu transwersalnym. Rozpoznanie wzajemnych relacji między ujęciem transwersalnym i kohortowym wymaga długich i szczegółowych szeregów czasowych według wieku, płci i statusu na rynku pracy. Niewiele krajów dysponuje odpowiednią bazą informacyjną. Niestety, Polska do nich nie należy, chociaż potencjalnie mo-

głaby ją mieć, gdyż mamy rejestr bezrobotnych oraz Badanie Aktywności Ekonomicznej Ludności. Brak nam odpowiedniej tabulacji danych o przepływach na rynku pracy. Z tego właśnie względu sięgnęliśmy po przykłady ze Stanów Zjednoczonych, gdzie taki kierunek badań podejmuje Nowa Ekonomia. Naszym celem jest wskazanie na niedoskonałość używanych dotąd metod i potrzeby tworzenia nowych źródeł statystycznych.

## 2. Translacja demograficzna

Translacja demograficzna analizuje wpływ zmian rozkładu zdarzeń ludnościowych w ujęciu wzdłużnym na wskaźniki w ujęciu transwersalnym<sup>1</sup>. Wbrew temu, co wcześniej przyjmowano, między innymi także w modelach Lotki, założenie o mocnej ergodyczności procesów ludnościowych jest trudne do zaakceptowania nie tylko w prognozowaniu, ale i w bieżącej ocenie koniunktury demograficznej. Okazuje się, że przeszłość w generacjach rzeczywistych ma znaczenie, a założenie o stałości cząstkowych współczynników płodności może dawać fałszywy obraz rzeczywistości.

Analiza demograficzna rozwijała się już od przeszło 200 lat, gdy w latach czterdziestych XX wieku dostrzeżono w niej pewne anomalie, które zmusiły amerykańskich demografów do refleksji nad wynikami w tradycyjnej transwersalnej analizie płodności kobiet. Efektem tej refleksji było odkrycie analizy kohortowej. Wspomnianą anomalię ilustruje rys. 1, gdzie na skali z prawej strony mamy współczynniki dzietności teoretycznej (*Total Fertility Rate* – TFR), jeden z trzech najważniejszych wskaźników rozwoju demograficznego danego kraju<sup>2</sup>. TFR jest interpretowany jako przeciętna liczba dzieci przypadających na statystyczną Amerykankę, gdyby rodziła z częstością obserwowaną w danym roku kalendarzowym.

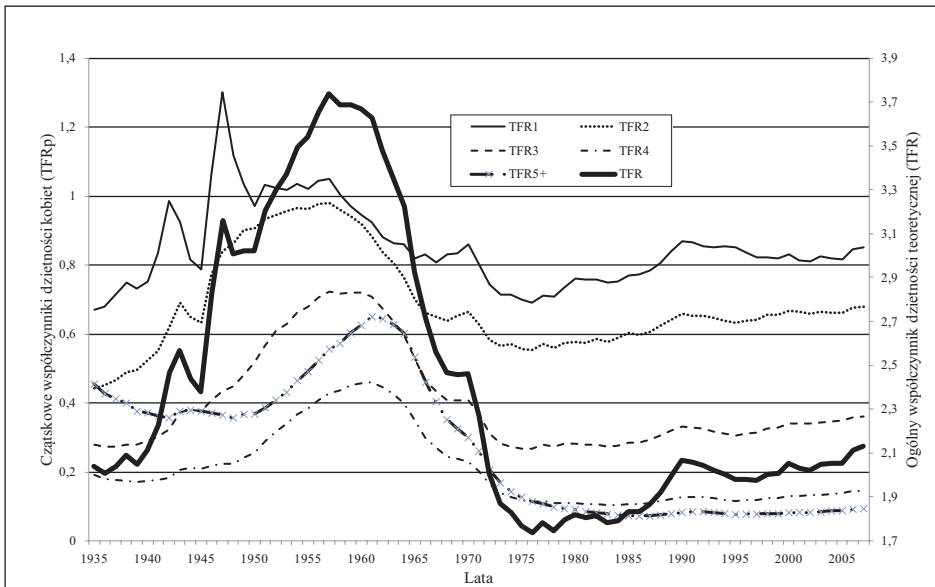
Jak widać na rys. 1, w latach czterdziestych TFR zaczyna gwałtownie rosnąć, osiągając najwyższy poziom w 1957 r., kiedy przeciętna Amerykanka mogła oczekiwać 3,74 dziecka. Potem następuje równie silny spadek dzietności do poziomu 1,75 dziecka w 1978 r. Bardzo wysoka dzietność w latach pięćdziesiątych zapewne długo nie wzbudziłyby zastrzeżeń demografów, gdyby nie zwrócono uwagi na współczynniki dzietności cząstkowej, wyrażające teoretyczne liczby dzieci poszczególnych kolejności porodu na 1 kobietę, zob. skala po lewej stronie na rys. 1.

---

<sup>1</sup> Matematyczne podstawy teorii translacji demograficznej dał N.B. Ryder [1964] i można je znaleźć w niektórych podręcznikach z zakresu demografii matematycznej [Paradysz 1985, s. 182-184]. O poprawności i żywotności teorii translacji świadczy wiele opracowań metodologicznych, najczęściej w odniesieniu do dzietności kobiet, w których proponuje się różne warianty korekt bieżących współczynników dzietności teoretycznej [Bongaarts 2006; Bongaarts, Feeney 1998; Bongaarts, Sobotka 2012].

<sup>2</sup> Dwa pozostałe to przeciętne dalsze trwanie życia noworodka oraz wskaźnik umieralności niemowląt. Wszystkie te trzy miary odnoszą się do kohort hipotetycznych w analizie transwersalnej.





**Rys. 1.** Współczynniki dzietności reoretycznej – ogólny (TFR) i częściowe (TFRp) w USA w latach 1935-2007

Źródło: <http://www.humanfertility.org/>.

Wartości współczynnika dzietności pierwszej kolejności (TFR1) wyższe od jedności świadczą, że nie może on być interpretowany jako prawdopodobieństwo urodzenia pierwszego dziecka, jak to wcześniej czyniono. Naniesienie częściowych współczynników płodności według wieku i kolejności porodu na diagram Lexisa<sup>3</sup> pozwoliło stwierdzić, że – po pierwsze – analiza transwersalna nie jest jedynym możliwym ujęciem badania losów człowieka. Po drugie, w analizie kohortowej, która ujmuje rozwój populacji według generacji rzeczywistych, nie jest możliwe uzyskiwanie wartości absurdalnych, takich jak  $TFR1 > 1$ , lub niespełnienie relacji  $TFR1 > TFR2 > TFR3 > TFR4 > \dots$ . Po trzecie, żadna z tych dwóch analiz nie jest alternatywnym modelem rozwoju demograficznego drugiego ujęcia i nie mogą się wzajemnie zastępować. Po czwarte, obydwa rodzaje analiz są niezbędne dla właściwej oceny rozwoju i polityki ludnościowej oraz prognozowania demograficznego. Obserwacja ewolucji rozkładu zdarzeń według wieku<sup>4</sup> w generacjach rzeczywistych

<sup>3</sup> Siatka demograficzna zwana jest także diagramem Lexisa, chociaż znacznie się różni od tego, co pierwotnie Lexis skonstruował<sup>101</sup>. Współcześnie posługujemy dwoma nowymi wariantami diagramu Lexisa. Pierwszy z nich, wykorzystujący pierwszą ćwiartkę kartezjańskiego układu współrzędnych, nazwałem kiedyś francuskim, bo zawdzięczamy go R. Pressatowi. Drugi, amerykański, wariant siatki nawiązuje do czwartej ćwiartki układu kartezjańskiego i jest wygodnym narzędziem obserwacji badanych procesów w arkuszu kalkulacyjnym [Paradysz 1999].

<sup>4</sup> Za demografią francuską nazywamy je kalendarzem danego rodzaju zdarzeń (płodności, małżeństwa, migracji, a dalej także zatrudnienia, bezrobocia itd.). Miarą kalendarza najczęściej bywa

pozwała na ustalenie ich wpływu na intensywność w ujęciu transwersalnym oraz zastosowania odpowiednich korekt.

Wracając do naszego przypadku USA, na rys. 1 należy zauważyć, że ten kraj nie był wyjątkiem w rozwoju populacyjnym krajów rozwiniętych, które w XIX bądź XX wieku przeszły przez rewolucję demograficzną. W tych krajach pod wpływem naukowo-technicznego rozwoju najpierw zmniejszyła się umieralność, a następnie po upływie 1-3 pokoleń zaczęła się obniżać rozrodczość<sup>5</sup>. Rodzice zaczynają w coraz większym stopniu decydować nie tylko o liczbie dzieci (intensywność procesu), ale o tym, kiedy mogą się one narodzić (kalendarz urodzeń). Wśród kilku wybranych krajów Ameryki i Europy widzimy interesującą zbieżność przebiegu rozrodczości, co początkowo narzucało mylną interpretację tego zjawiska. Zasadniczo mylnym tropem było interpretowanie przyczyn „baby boomu”, czyli silnego wzrostu dzietności w latach pięćdziesiątych, rekompensatą powojenną. W większości krajów europejskiego kręgu kulturowego, ale i wysoko rozwiniętych krajów azjatyckich, w młodszych generacjach kobiet dzietność była niższa. Przyczyną wysokich współczynników dzietności teoretycznej w ujęciu transwersalnym w latach pięćdziesiątych była przede wszystkim zmiana kalendarza płodności w generacjach rzeczywistych, zob. rys. 2. O ile w starszych generacjach kobiet, urodzonych przed drugą wojną światową, przeciętny wiek rodzenia pierwszego dziecka był coraz niższy, o tyle urodzone po 1940 r. rodziły swe pierwsze dziecko coraz później. Szczególnie to widać w USA i w Kanadzie, dla których mamy najdłuższe szeregi czasowe danych. Jest to efekt rozpoczynających się olbrzymich przemian w zachowaniach seksualnych młodzieży i ich wpływie na formowanie się rodzin<sup>6</sup>. Przeciętny wiek w chwili pierwszego porodu we wcześniejszych generacjach w Polsce<sup>7</sup> był podobny do tego, co widzimy w Kanadzie, później jednak w generacjach kobiet urodzonych w latach 1945-1965 ustabilizował na poziomie 23,5 roku. Dynamika tego wieku w Czechach była podobna jak w Polsce, chociaż „czeskie” średnie pozostawały stale nieco niższe.

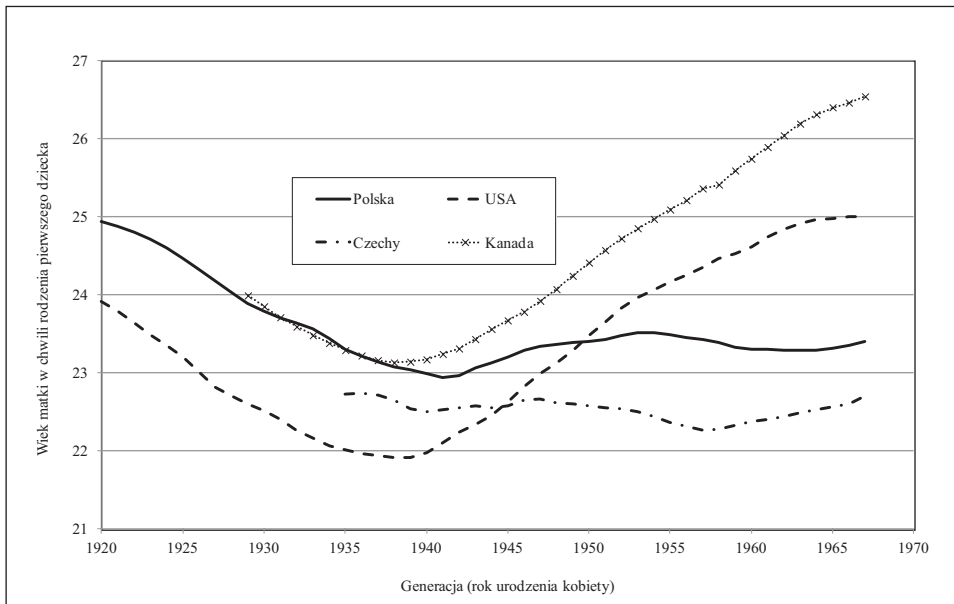
---

przeciętny wiek w momencie doznania danego zdarzenia. W szczególności jest to przeciętny wiek macierzyństwa, w chwili pierwszego małżeństwa, emigracji, pierwszej pracy, bezrobocia itd.

<sup>5</sup> Do potencjalnych rodziców docierała informacja, że nie trzeba wydawać na świat 6 dzieci, żeby połowa z nich przeżyła. Z drugiej strony coraz większej liczbie przeżywających dzieci trzeba było zapewnić „życiowy start” i trzeba było sprostać rosnącym potrzebom potomstwa.

<sup>6</sup> Przy okazji warto zwrócić uwagę na siłę zmian kalendarza płodności w Ameryce Północnej. Zapewne tutaj właśnie należy szukać początków drugiego przejścia demograficznego, a nie w krajach skandynawskich czy Europy Zachodniej. Przeciętny wiek urodzenia pierwszego dziecka w Kanadzie i USA zachowuje się podobnie z tym jednak, że stale jest on niższy o około 1 rok. Ma to zapewne związek ze strukturą etniczną obu krajów ze względu na większy odsetek ludności kolorowej i hiszpańskojęzycznej w USA.

<sup>7</sup> Bieżąca statystyka urodzeń pozwala nam na odtworzenie płodności kobiet w generacjach rzeczywistych dopiero od generacji 1935, która weszła w okres rozrodczy dopiero w 1950 r. Dla wcześniejszych generacji oszacowaliśmy przeciętny wiek rodzenia pierwszego dziecka częściowo za pomocą statystyki bieżącej, a częściowo na podstawie danych z reprezentacyjnego badania dzietności kobiet w NSP 1988, por. [Paradysz 1992].



**Rys. 2.** Przeciętny wiek matki w chwili rodzenia pierwszego dziecka w generacjach rzeczywistych 1920-1967 w Polsce, Kanadzie, USA i Czechach

Źródło: opracowanie własne na podstawie danych z portalu <http://www.humanfertility.org> oraz dla Polski bieżącej statystyki urodzeń i [Paradysz 1992, s. 252].

Wyniki powyższych przemian w generacjach rzeczywistych występują także w ujęciu transwersalnym, dla poszczególnych lat kalendarzowych. Zatem i w Czechach, i w Polsce należy oczekiwać dalszego wzrostu przeciętnego wieku macierzyństwa w momencie rodzenia pierwszego dziecka w generacjach urodzonych po 1970 r.

### 3. Translacja procesów ekonomicznych na przykładzie rynku pracy

Jak zauważono powyżej, kariera zawodowa i rodzinna są ze sobą ściśle powiązane. Zatem ich śledzenie powinno się odbywać równoległe i za pomocą tej samej metodologii. Dodatkowym argumentem dla tej tezy jest aktualność mechanizmu translacji na rynku pracy. Albowiem, podobnie jak niegdyś w badaniach demograficznych, w analizie rynku pracy dominuje ujęcie transwersalne i założenie o mocnej ergodyczności badanego procesu. Zapewne decydującym motywem jest praktyczne znaczenie wyników analizy w modelowaniu i prognozowaniu sytuacji bieżącej, rozumianej jako coś, co dzieje się w danym roku kalendarzowym lub okresach w niezbyt odległej przeszłości.

W ostatnich latach ukazało się wiele opracowań dotyczących rynku pracy w ujęciu kohortowym zarówno w Europie<sup>8</sup>, jak i w Stanach Zjednoczonych<sup>9</sup>, w których wskazuje się na przenikanie się obu analiz. W Europie badania rynku pracy w coraz większym stopniu uwzględniają efekty generacyjne zatrudnionych oraz czynniki koniunkturalne. Dla lepszej ilustracji problemu warto tutaj przedstawić kohortowy model cząstkowych współczynników zatrudnienia (*Labour Force Participation Rate* – LFPR) [Fallick, Pingle 2007, s. 8]:

$$LFRP_{a,t} = \alpha_a^* X_t^{\lambda_a} \beta_{b=t-a} \exp(\varepsilon_{a,t}^*) \quad (1)$$

lub, po zlogarytmowaniu,

$$\log LFRP_{a,t} = \log \alpha_a^* + \log \beta_{b=t-a} + \lambda_a^* \log X_t + \varepsilon_{a,t}^*, \quad (2)$$

gdzie:  $t$  – rok kalendarzowy  
 $a$  – wiek w przedziałach pięcioletnich,  
 $b$  – rok urodzenia,  
 $\alpha_a^*$  – ustalony efekt wieku  $a$ ,  
 $\beta_{b=t-a}$  – ustalony efekt generacji (kohorty urodzeniowej),  
 $\lambda_a^*$  – cząstkowy (zależny od wieku) współczynnik wrażliwości na cykl koniunkturalny,  
 $X_t$  – macierz zmiennych objaśniających, które w podstawowym modelu uwzględniają tylko fazę cyklu koniunkturalnego,  
 $\varepsilon_{a,t}^*$  – składnik losowy.

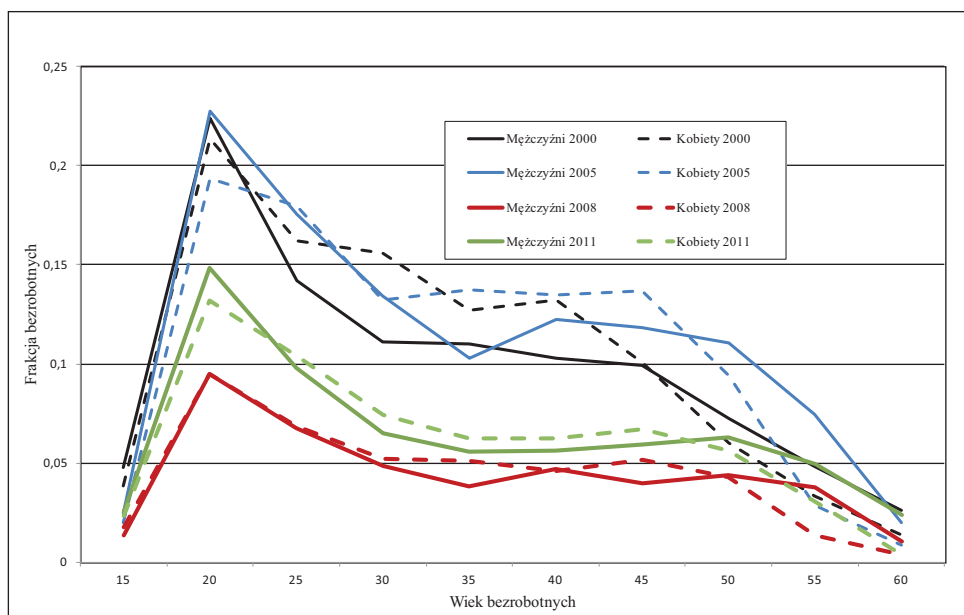
We wzorach (1) i (2) zostały pominięte oznaczenia na płeć zatrudnionych, chociaż w konkretnych weryfikacjach modelu, między innymi [Balleer, Gómez-Salvador, Turunen 2009], jest ona uwzględniana. Jak piszą sami autorzy modelu (1), jest on trudno agregowany i poszczególne cząstkowe współczynniki zatrudnienia szacuje się oddzielnie dla każdej grupy wieku. Czy można jednak mówić o pełnej zgodności aplikacyjnej między modelami demograficznymi, w tym typu APC (*age, period, cohort*), a modelem 1? Wydaje się, że jednak nie ma takiej zgodności, chociaż ich praktyczne znaczenie wydaje się potwierdzać zastosowanie w prognozowaniu siły roboczej. Tradycyjne modele demograficzne małżeńskości, płodności, umieralności i migracji uwzględniają przepływy między poszczególnymi stanami, a modele zatrudnienia, i *par excellence* aktywności zawodowej, bezrobocia, dotyczą tylko

<sup>8</sup> A. Balleer, R. Gómez-Salvador, J. Turunen [2009] na podstawie danych badania siły roboczej (LFS) z 5 największych krajów strefy euro (Niemcy, Francja, Włochy, Hiszpania i Holandia) wykorzystali kohortowy model zatrudnienia autorstwa B. Fallicka i J. Pingle'a [2007] dla ostatnich kilkunastu lat.

<sup>9</sup> Do najciekawszych z nich należy dzieło pod red. K.G. Abraham, J.R. Spletzera i M. Harpera, *Labor in the New Economy*, jako pokłosie konferencji badania dochodów i bogactwa pt. „Praca w Nowej Ekonomii”, która się odbyła w miejscowości Bethesda, Maryland, w dniach 16-17 listopada 2007.

stanów. W dychotomicznym podziale na cząstkowe współczynniki pierwszej i drugiej kategorii nie ma miejsca ani na współczynniki zatrudnienia, ani na analogiczne współczynniki bezrobocia. Zilustrujemy to na podstawie danych OECD<sup>10</sup> w ujęciu transwersalnym.

Ujednolicone dane statystyczne dotyczące siły roboczej w OECD, które są dostępne na stronie internetowej OECD, istnieją dopiero od 2000 r. i nie są wystarczające dla odtworzenia kompletnych rzeczywistych generacji zatrudnionych, bezrobotnych i biernych zawodowo. Gdyby nawet było już to możliwe, to i tak wartość obliczonych na ich podstawie cząstkowych współczynników aktywności i bierności zawodowej oraz bezrobocia nie jest możliwe z tego powodu, że nie są one współczynnikami natężenia, ale struktury.



**Rys. 3.** Rozkłady cząstkowych współczynników bezrobotnych według wieku i płci w Polsce

Źródło: opracowanie własne na podstawie danych OECD [http://stats.oecd.org/Index.aspx?DatasetCode=LFS\\_D](http://stats.oecd.org/Index.aspx?DatasetCode=LFS_D); odczyt w dniu 7 sierpnia 2012.

Ich uzupełnieniem do jedności są w każdej grupie wieku współczynniki zatrudnienia i bierności zawodowej. Polskie rozkłady cząstkowych współczynników bezrobocia pod względem kształtu i wielkości przypominają nieco krzywe płodności i zawierania małżeństw. Jest to jednak pozorne podobieństwo i z funkcjami macie-

<sup>10</sup> Organizacja na rzecz współpracy i rozwoju ekonomicznego (OECD) posiada ujednoliconą statystykę rynku pracy w przekroju według płci i wieku dla wszystkich krajów w niej zrzeszonych.

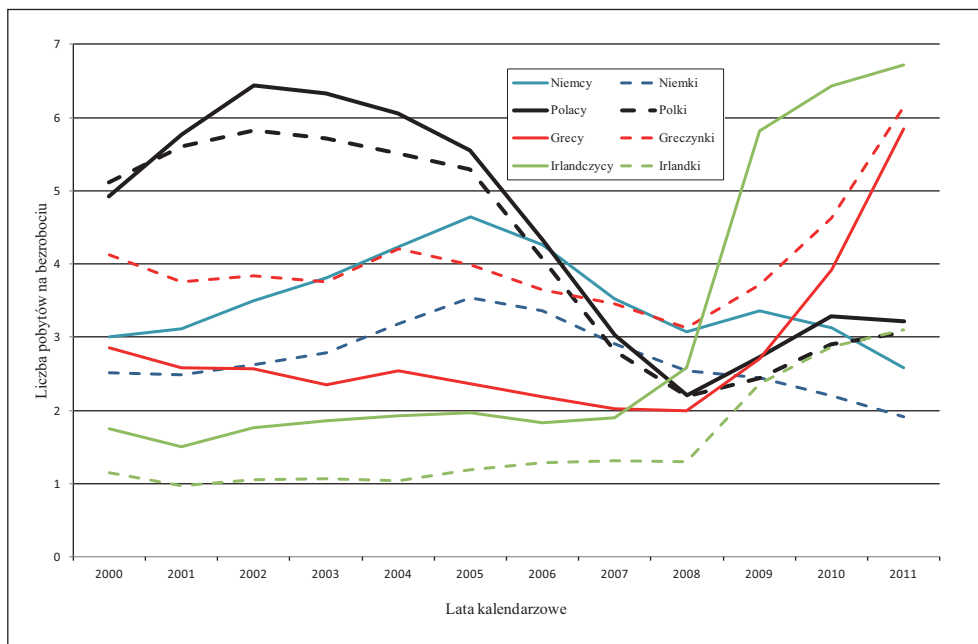
rzyństwa i małżeńskości nie mają wiele wspólnego. Najwyższa frakcja bezrobotnych jest w przedziale 20-25, po czym z wiekiem opada. Od początku 2000 r. Polska, charakteryzująca się początkowo wysokim bezrobociem, najniższy jego poziom osiągnęła w 2008 r. Nie wchodząc w problem analizy momentów tak skonstruowanej „funkcji bezrobocia”, policzyliśmy ich sumy, które mogłyby odpowiadać teoretycznej liczbie pobytów na bezrobociu, gdyby były wskaźnikami natężenia, a nie struktury. Jak widać na rys. 4, „teoretyczne stopy” bezrobocia (w skrócie TSB) dość dobrze odzwierciedlają momenty kryzysowe. Od 2008 r. TSB rosną w większości krajów europejskich z wyjątkiem Niemiec. Nie wydaje się czymś niewłaściwym odtwarzanie cząstkowych współczynników bezrobocia i ich sum w generacjach rzeczywistych i analizowanie efektów translacyjnych. Jednakże przed „do końca demograficzną” interpretacją stóp bezrobocia przestrzegają nas konsekwencje przed analogicznym potraktowaniem „teoretycznych” stóp zatrudnienia (TSZ). TSZ będące od kilku do kilkunastu razy większe od TSB mogłyby dać wynik rzędu 30-50 na jedną osobę, co trudno byłoby traktować w kategoriach mobilności pracowniczej. Wspomniane tutaj informacje o częstości przebywania na bezrobociu i zmianach pracy nie są jedynymi, którymi byłaby zainteresowana polityka zatrudnienia i zarządzanie kadrami. Na wzór tablic demograficznych opartych na cząstkowych współczynnikach pierwszej i drugiej kategorii znacznym zainteresowaniem różnych służb społecznych, gospodarczych i politycznych cieszyłyby się informacje o długotrwałości pracy, bezrobocia i innych form nieaktywności zawodowej (ojcostwo, macierzyństwo), a także o tym, ile razy zmieniamy pracę bądź przebywamy na bezrobociu, z jakiego powodu, gdzie i jak długo pracujemy. Bardzo istotnym wzbogaceniem systemu informacyjnego byłoby powiązanie tych tablic z mobilnością społeczną i przestrzenną, w tym z dojazdami do pracy, oraz ich integracja z wielostanowymi i wielowymiarowymi tablicami demograficznymi.

Co potrzebujemy, żeby tego rodzaju tablice można było opracować? Oprócz dalszego doskonalenia struktur według płci, wieku i (ewentualnie) trwania w określonym stanie, przede wszystkim potrzebujemy odpowiednich statystyk strumieni aktywizacji i dezaktywizacji zawodowej. Dane statystyczne o podobnym charakterze zaczęto już zbierać w niektórych krajach o lepiej zorganizowanej statystyce pracy. Do takich źródeł należy amerykański Job Openings and Labor Turnover Survey (JOLTS) oraz integracja różnych administracyjnych źródeł danych<sup>11</sup>.

Dla celów analizy kohortowej muszą to być długie szeregi czasowe. Dla odtworzenia tylko jednej kohorty zatrudnieniowej obserwacja musi trwać ponad 50 lat. Teoria translacji w zakresie zatrudnienia jest jeszcze bardziej wymagająca, a okres obserwacji jest już liczony powyżej 100 lat. Oczywiście modelowanie i estymacja pośrednia może nam tutaj przyjść z pomocą, ale prace nad odpowiednim systemem informacyjnym należy rozpocząć już dzisiaj. Natomiast w bieżącej analizie rynku

---

<sup>11</sup> O integracji informacji w zakresie rynku pracy zob. [Fallick, Fleischman 2004; Fallick, Fleischman, Pingle 2010; Davis i in. 2010].



**Rys. 4.** Teoretyczna roczna stopa bezrobocia mężczyzn w wybranych krajach OECD w latach 2000-2011

Źródło: opracowanie własne na podstawie danych OECD [http://stats.oecd.org/Index.aspx?DatasetCode=LFS\\_D](http://stats.oecd.org/Index.aspx?DatasetCode=LFS_D); odczyt w dniu 7 sierpnia 2012.

pracy w ujęciu transwersalnym można wyjść od cząstkowych współczynników zatrudnienia  $[Z(x, p, t)]$  jako funkcji wieku  $x$  oraz kolejności zatrudnienia  $p$  w roku kalendarzowym  $t$ :

$$Z(x, p, t) = \frac{LPP(x, p, t)}{L(x, t)}. \quad (3)$$

W liczniku (3) występuje liczba podejmujących pracę  $p$ -ty raz w wieku  $x$ , w roku  $t$ , a w mianowniku są wszystkie osoby w tymże wieku  $x$  w połowie okresu  $t^{12}$ . Wzorem demografii [Paradysz 1985, s. 43-47] funkcje (3) można scharakteryzować za pomocą  $n + 1$  pierwszych momentów zwykłych nieunormowanych:

$$R(n, p, t) = h \cdot \sum_{x=15}^{67} Z(x, p, t) \cdot (x + 0,5 \cdot h)^n, \quad (4)$$

gdzie  $h$  jest interwałem klasowym wieku, najczęściej  $h = 1$  lub  $5$ , oraz  $n = 0, 1, \dots, 4$ .

<sup>12</sup> Zgodnie z zasadą konstrukcji cząstkowych współczynników demograficznych drugiej kategorii [Kędelski, Paradysz 2006, s. 83-91].



Moment zerowy –  $R(0, p)$  – informuje nas o intensywności danego zdarzenia, a momenty wyższego rzędu o kształcie funkcji  $p$ -tego zatrudnienia. Na przykład liczby  $R(0, 1, t) = 0,79$ ;  $R(0, 2, t) = 0,42$ ;  $R(0, 3, t) = 0,15$  oznaczają, że 79% w danej kohorcie hipotetycznej podejmuje pierwszą pracę, 42% – drugą, 15% – trzecią. Na podstawie tych danych można też ustalić strukturę zatrudnienia według liczby miejsc pracy [Kędełski, Paradysz 2006, s. 196]. Przy założeniu stacjonarności procesu zatrudnienia w naszym hipotetycznym przykładzie 21% osób nigdy nie podejmie pracy, 37% pozostanie wiernym jednemu pracodawcy, a 27% zatrudnionych ograniczy się tylko do pracy w dwóch firmach. Wykorzystując wzór (4), można obliczyć przeciętny wiek w chwili  $p$ -tego zatrudnienia oraz odpowiednie odchylenia standardowe, a także współczynniki asymetrii i koncentracji tych funkcji [Kędełski, Paradysz 2006, s. 191-192].

Rodzina funkcji  $Z(x, p, t)$ , co do kształtu i intensywności zdeterminowana przez kolejność zatrudnienia, wiek i płeć, daje możliwość bardziej pogłębionej analizy. Po uwzględnieniu zwolnień i innych odejść będzie można ustalić przerwy między pracą u kolejnych pracodawców. Pozwoli to wprowadzić także pojęcie bezrobocia oraz przyczyni się do pogłębieniem studiów nad wiekiem w chwili ostatniego zatrudnienia. Nawiązując do naszego oryginalnego opracowania [Paradysz 1995], gdzie połączyliśmy dwa systemy analizy demograficznej: klasyczny Lotki-Kuczyńskiego z tablicami wejścia-wyjścia [Paradysz 1985, s. 138-139], zdekomponujemy funkcję  $Z(x, p, t)$  – wzór (3) na dwie części. Pierwsza z nich –  $Z'(x, p, t)$  – odpowiada tym pracującym, dla których  $p$ -ta praca nie jest ostatnia. Druga –  $Z''(x, p, t)$  – dotyczy tych, którzy zakończą swoją karierę zawodową na pracy kolejności  $p$ . W tym celu wyjdziemy od pojęcia prawdopodobieństwa  $p$ -tego zatrudnienia w wieku  $x$ :

$$PZ(x, p, t) = \frac{Z(x, p, t)}{l(x, p-1, t)}, \quad (5)$$

gdzie  $l(x, p-1, t)$  oznacza dożywających wieku  $x$ , którzy byli zatrudnieni w pracy  $p-1$ . W przypadku osób, które dotychczas nie pracowały –  $l(x, 0, t)$ . Na początku trwania życia zawodowego, to znaczy dla wieku 15 lat, przyjmuje się  $l(15, p, t) = 10^c$ , gdzie najczęściej  $c$  przyjmuje wartości 4, 5 lub 6. Następnie, rekurencyjnie, rozwija się liczby przeżywających w pracy kolejności  $p$  według następującej formuły [Kędełski, Paradysz 2006, s. 216-217]:

$$l(x+1, p, t) = l(x, p, t) \cdot p(x, t) + LPP(x, p, t) + LPP(x, p+1, t), \quad (6)$$

gdzie  $p(x, t)$  jest prawdopodobieństwem przeżycia według tablic trwania życia w roku  $t$ . Funkcja (6), podobnie jak w przypadku bieżących odstępów między urodzeniami [Paradysz 1995], może posłużyć do oszacowania bieżących odstępów między pracą  $p$  oraz  $p+1$ . W tym celu wprowadzamy pojęcie prawdopodobieństwo pozostania w  $p$ -tej firmie do końca swojej aktywności zawodowej, czyli niepracowania  $p+1$  raz:



$$Q(x+1, p+1, t) = \prod_{y=x}^{67} [1 - PZ(y, p+1, t)], \quad (7)$$

gdzie  $y$  jest dodatkowym określeniem wieku w pracy  $p+1$ .

Wspomnianej wyżej dekompozycji funkcji zatrudnienia (3) na dwa składniki dokonujemy według wzorów (8) i (9):

$$Z'(x, p, t) = Z(x, p, t) \cdot [1 - Q(x+1, p+1, t)], \quad (8)$$

$$Z''(x, p, t) = Z(x, p, t) \cdot Q(x+1, p+1, t). \quad (9)$$

Na podstawie obydwu funkcji można wyznaczyć średni wiek zatrudnienia tych osób, dla których  $p$ -ta praca nie jest ostatnią w ich życiu

$$\bar{x}'(p, t) = \frac{\sum_{x=15}^{67} Z(x, p, t) \cdot (x + 0,5 \cdot h)}{\sum_{x=15}^{67} Z(x, p, t)}, \quad (10)$$

oraz tych, którzy pozostaną u  $p$ -tego dla nich przedsiębiorcy do końca swojej aktywności zawodowej

$$\bar{x}''(p, t) = \frac{\sum_{x=15}^{67} Z(x, p, t) \cdot (x + 0,5 \cdot h)}{\sum_{x=15}^{67} Z(x, p, t)}. \quad (11)$$

Analogicznie przeprowadzona analiza zwolnień i odejść z pracy (w tym z powodu śmierci i na rentę) pozwoli na ustalenie długotrwałości zatrudnienia i przerw między kolejnymi pracodawcami, w tym długotrwałości bezrobocia. Wzbogaci to zasób metod badania bieżącej koniunktury gospodarczej, a w dalszej perspektywie pozwoli na analizę rynku pracy z uwzględnieniem teorii translacji demograficznej.

## Literatura

- Balleer A., Gómez-Salvador R., Turunen J. (2009), *Labour force participation in the euro area. A cohort based analysis*. European Central Bank, Working Paper no. 1049.
- Bongaarts J., Feeney G. (1998), *On the quantum and tempo of fertility*. Population and Development Review no. 2, s. 271-291.
- Bongaarts J. (2006), *Quantum and tempo of life cycles events*. Vienna Yearbook of Population Research, s. 115-151.
- Bongaarts J., Sobotka T. (2012), *A demographic explanation for the recent rise in European fertility*. Population and Development Review no. 1, s. 83-120.

- Davis S.J., Faberman R.J., Haltiwanger J.C., Rucker I., (2010), *Adjusted Estimates of Worker Flows and Job Openings in JOLTS*, [w:] Pod red. K.G. Abraham, J.R. Spletzer, and M. Harper, *Labor in the New Economy*. University of Chicago Press, Chicago-London, s. 187-216.
- Fallick B., Fleischman C. (2004), *Employer-to-employer flows in the U.S. labor market: The complete picture of gross worker flows*. Finance and Economics Discussion Series, no. 2007-9. Federal Reserve Board, Washington, DC.
- Fallick B., Pingle J. (2007), *A cohort-based model of labor force participation*. Finance and Economics Discussion Series, no. 2007-9. Federal Reserve Board, Washington, DC.
- Fallick B., Fleischman C., Pingle J. (2010), *The effect of population aging on the aggregate labor market*. [w:] Pod red. K.G. Abraham, J.R. Spletzer, and M. Harper, *Labor in the New Economy*. University of Chicago Press, Chicago-London, s. 377-417.
- Kędeliski M., Paradysz J. (2006), *Demografia*. Wyd. Akademii Ekonomicznej w Poznaniu, Poznań.
- Paradysz J. (1985), *Wielowymiarowa analiza reprodukcji ludności*. Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, Poznań.
- Paradysz J. (1992), *Dzietność kobiet w Polsce*, Główny Urząd Statystyczny, Warszawa.
- Paradysz J. (1995), *Odstępy między urodzeniami jako miara koniunktury demograficznej*, [w:] J. Paradysz (red.) *Badanie koniunktury demograficznej*. Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 1995 s. 24-33.
- Paradysz J. (1999), *Aplikacja metod szkoły francuskiej w nauczaniu demografii w Polsce*. [w:] T. Kowaleski (red.) *Nauczanie demografii w szkołach wyższych w świetle potrzeb dnia dzisiejszego*. Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1999, s. 25-34.
- Ryder N.B. (1964), *The process of demographic translation*. *Demography*, no. 1, s. 74-82.

## NEW POSSIBILITIES FOR STUDYING THE SITUATION ON THE LABOUR MARKET

**Summary:** General numbers and unemployment rates, persons active and inactive are generally not sufficient indicators to characterize the local labour market, which is characterized by a high diversity of the population structure of basic demographic characteristics. Using the abundant information contained in the records of the administrative register and the Labour Force Survey (LFS Poland) can greatly expand the research methodology using demographic analysis. This line of analysis is also worth paying special attention because of a uniform methodology for women to keep track of three careers: educational, employment and family. This paper presents the problem of the demographic translation, the relationship between the cohort and transversal analysis based on selected countries of America and Europe and points to new possibilities for the interpretation of population reproduction rates and their potential application in the economic situation.

**Keywords:** demographic analysis, cohort analysis, translation theory, labor market, economic activity of the population.