

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

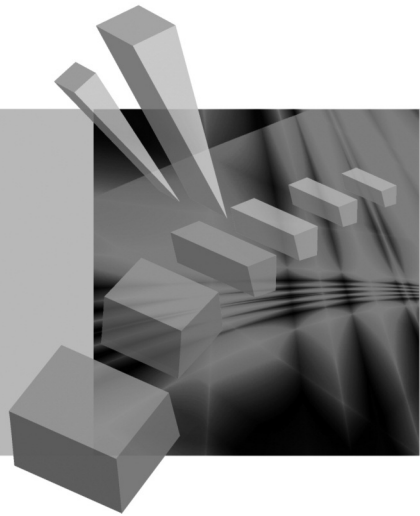
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Miroslaw Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library 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

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka, Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska, Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński, Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz, Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel, Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień, Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska, Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan, Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska, Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka, Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański, Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk, Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk, Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura, Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk, Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski, Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka, Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk, Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawelczyk, Jerzy Kołodziej, Jerzy Błaszczak , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarc , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Eugeniusz Gatnar

Uniwersytet Ekonomiczny w Katowicach
Narodowy Bank Polski

JAKOŚĆ DANYCH W SYSTEMACH STATYSTYCZNYCH BANKÓW CENTRALNYCH (NA PRZYKŁADZIE NBP)

Streszczenie: W artykule przedstawiono analizę jakości danych w systemach statystycznych banków centralnych, biorąc jako przykład Narodowy Bank Polski. Dobrym wskaźnikiem jakości tych danych jest pozycja bilansu płatniczego Polski, nazywana „saldo błędów i opuszczeń”. Dane sprawozdawane przez instytucje finansowe oraz Ministerstwo Finansów i GUS nie były dokładne i wskazywały na występowanie wysokiego deficytu na rachunku bieżącym w stosunku do PKB. Prace podjęte przez Departament Statystyki NBP spowodowały rewizję importu w bilansie płatniczym w roku 2011 w wysokości ok. 10 mln zł.

Słowa kluczowe: statystyka, finanse, jakość danych, kryzys finansowy.

1. Wstęp

Trwający kryzys finansowy ujawnił występowanie luk informacyjnych w systemach sprawozdawczych banków centralnych w Europie i na świecie. Systemy te nie rejestrowały wielu ważnych, jak się teraz okazało, charakterystyk transakcji realizowanych w sferze finansów, zwłaszcza jeśli chodzi o operacje transgraniczne.

Celem artykułu jest próba oceny jakości danych w systemach sprawozdawczych banków centralnych na przykładzie Narodowego Banku Polskiego. Jako syntetyczne odzwierciedlenie tej jakości przyjęto wielkość pozycji bilansu płatniczego Polski, która nosi nazwę „błędy i opuszczenia” (*errors and omissions*). Bilans płatniczy jest podstawowym dokumentem publikowanym miesięcznie i kwartalnie przez Departament Statystyki Narodowego Banku Polskiego.

2. Luki informacyjne

Kryzys finansowy, który przeszedł do sfery realnej, ujawnił występowanie w systemach statystyki finansowej i bankowej tzw. luk w danych (*data gaps*) lub luk informacyjnych (*information gaps*). Luki te są najczęściej wynikiem szybkiego rozwoju rynków finansowych oraz innowacyjności w zakresie nowych produktów finansowych.

Występują one wtedy, gdy pojawia się trudność w uzyskaniu odpowiednio dokładnych danych, o odpowiedniej częstotliwości w wymaganym czasie.

Zbierane do tej pory przez systemy sprawozdawcze banków centralnych w Europie dane nie obejmowały dostatecznie szczegółowo takich zjawisk, jak np. szerokie wykorzystywanie instrumentów o wysokim poziomie ryzyka. Przy czym najczęściej ryzyko to było transferowane z banków poza sektor, gdyż np. sprzedaż CDO dawała duże zyski bankom.

Nie uwzględniano także wysokiego stopnia złożoności powiązań i zależności pomiędzy firmami pośrednictwa finansowego, np. przez podmioty specjalnego przeznaczenia (*Special Purpose Investment Vehicles* – SIV) oraz firmy zarejestrowane w rajach podatkowych (*offshore companies*).

Nie zwracano dotąd należytej uwagi na działalność instytucji pozabankowych, funduszy emerytalnych, inwestycyjnych czy firm ubezpieczeniowych.

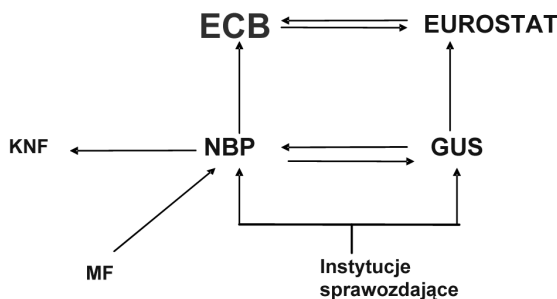
Systemy sprawozdawcze banków centralnych i instytucji nadzorczych nie zawierały danych:

- o odpowiedniej częstotliwości, co najmniej kwartalnej,
- o odpowiednim stopniu szczegółowości lub odpowiednim stopniu agregacji, np. skonsolidowanego bilansu całego przedsiębiorstwa, a nie tylko jego oddziałów w poszczególnych krajach,
- o walutach, w jakich wycenione są poszczególne pozycje aktywów,
- o geograficznym pochodzeniu lub położeniu aktywów,
- o ratingu posiadanych aktywów itd.

3. System statystyki finansowej w Polsce

W systemie statystyki finansowej Narodowy Bank Polski odgrywa najważniejszą rolę. Realizuje on zadania statystyczne, publikując:

- Bilans płatniczy Polski: miesięczny, kwartalny i roczny.
- Międzynarodową pozycję inwestycyjną Polski.
- Zadłużenie zagraniczne Polski.



Rys. 1. System statystyki finansowej w Polsce

Źródło: opracowanie własne.

NBP prowadzi także rachunki finansowe w ramach rachunków narodowych (*National Accounts*), a następnie przekazuje te dane do GUS. Bank centralny także współpracuje z Ministerstwem Finansów (MF) i Komisją Nadzoru Finansowego (KNF).

System powiązań pomiędzy instytucjami w ramach systemu statystyki finansowej w Polsce przedstawiono na rys. 1.

4. Gromadzenie danych statystycznych w NBP

Departament Statystyki NBP gromadzi dane statystyczne dla potrzeb statystyki finansowej przekazywane przez podmioty gospodarcze za pomocą systemu PeGaZ.

Miesięcznie sprawozdaje się 640 przedsiębiorstw (w tym 73 banki), kwartalnie zaś sprawozdaje się 10 017 przedsiębiorstw (w tym 2506 kantorów). W sumie w roku 2011 sprawozdawało się do Departamentu Statystyki NBP 10 657 podmiotów.

System sprawozdawczy PeGaZ ma progi sprawozdawcze dla podmiotów sprawozdających się kwartalnie w wysokości:

- Przedsiębiorstwa – od 10 mln do 300 mln zł.
- Osoby fizyczne – od 7 mln zł.
- Kredyty handlowe – od 3 mln zł.
- Kantory – bez progu.

Z kolei dla podmiotów sprawozdających się miesięcznie znajdują się one na poziomie:

- Przedsiębiorstwa – powyżej 300 mln.
- Banki, fundusze, firmy inwestycyjne – bez progu.

5. Saldo błędów i opuszczeń

Najważniejszym dokumentem produkowanym miesięcznie i kwartalnie przez Departament Statystyki NBP jest bilans płatniczy Polski [*Bilans płatniczy...* 2011]. Stanowi on zestawienie wartości wszystkich transakcji gospodarczych, które w ciągu miesiąca lub kwartału zostały zrealizowane między rezydentami kraju a resztą świata (nierezydentami). Jest to więc zapis stosunków gospodarczych kraju z zagranicą: po jednej stronie zapisuje się w nim wszelkie transakcje powodujące dochody dewizowe, po drugiej – wszystkie wydatki dewizowe.

Bilans płatniczy składa się z czterech części. Pierwszą jest bilans bieżący, który obejmuje transakcje eksportu i importu towarów (bilansu handlowego), wymianę usług, transfery i inne bieżące transakcje. Wymiana aktywów finansowych, które powodują zobowiązania w przyszłości, jest natomiast księgowana w drugiej części bilansu płatniczego, czyli na rachunku finansowym. Trzecią część stanowią oficjalne transakcje rezerwowe. Dotyczą one zmian w rezerwach walutowych państwa. Błędy w sprawozdawczości sprawiają, że części transakcji nie udaje się rozpoznać,

Tabela 1. Bilans płatniczy Polski w lipcu 2011 r.

BILANS PŁATNICZY

w mln EUR

	2010	2011		Dane skumulowane **	
	VII	VI*	VII*	VIII 2009-VII 2010	VIII 2010-VII 2011*
A. Rachunek bieżący	-1 710	-1 596	-1 611	-12 930	-16 476
Saldo obrotów towarowych	-808	-1 065	-1 200	-6 425	-10 843
Eksport f.o.b.	10 162	11 888	10 721	113 934	132 470
Import f.o.b.	10 970	12 953	11 921	120 359	143 313
Saldo usług	213	452	349	3 102	3 232
Przychody	2 150	2 348	2 377	22 554	25 942
Rozchody	1 937	1 896	2 028	19 452	22 710
Saldo dochodów	-1 333	-1 127	-1 122	-12 411	-12 392
Przychody	484	590	528	4 924	6 283
Rozchody	1 817	1 717	1 650	17 335	18 675
Saldo transferów bieżących	218	144	362	2 804	3 527
Przychody	495	424	725	7 749	8 538
Rozchody	277	280	363	4 945	5 011
B. Rachunek kapitałowy	911	511	832	5 131	7 187
Przychody	932	526	850	5 376	7 677
Rozchody	21	15	18	245	490
C. Rachunek finansowy	5 804	2 561	-3 257	34 161	18 873
Polskie inwestycje bezpośrednie za granicą	22	-930	-3	-2 030	-5 980
Zagraniczne inwestycje bezpośrednie w Polsce	149	-405	-359	9 240	5 483
Inwestycje portfelowe - aktywa	-92	-92	120	-753	-800
Udziałowe papiery wartościowe	-111	7	131	-723	-639
Dłużne papiery wartościowe	19	-99	-11	-30	-161
Inwestycje portfelowe - pasywa	2 721	2 891	630	18 075	14 424
Udziałowe papiery wartościowe	415	968	-167	4 094	4 304
Dłużne papiery wartościowe	2 306	1 923	797	13 981	10 120
Pozostałe inwestycje - aktywa	391	8	-2 046	-1 339	-4 189
NBP	0	0	0	-313	0
Sektor rządowy	-2	15	-3 006	-206	-3 270
Monetarne Instytucje Finansowe (z wyłączeniem NBP)	398	271	-221	422	-132
Pozostałe sektory	-5	-278	1 181	-1 242	-787
Pozostałe inwestycje - pasywa	2 526	990	-1 529	11 150	10 294
NBP	1 151	496	-604	3 190	-404
Sektor rządowy	1 021	-97	744	2 545	2 576
Monetarne Instytucje Finansowe (z wyłączeniem NBP)	992	227	-1 443	3 449	3 729
Pozostałe sektory	-638	364	-226	1 966	4 393
Pochodne instrumenty finansowe	87	99	-70	-182	-359
D. Saldo błędów i opuszczeń	-1 210	-85	1 849	-9 111	-4 916
Razem A - D	3 795	1 391	-2 187	17 251	4 668
E. Oficjalne aktywa rezerwowe	-3 795	-1 391	2 187	-17 251	-4 668

* Dane wstępne

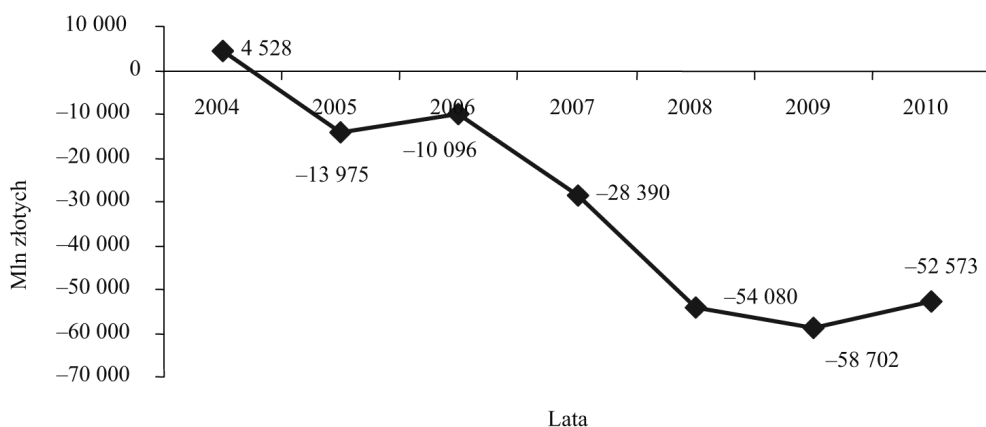
** Dane skumulowane są prezentowane krocząco w ujęciu 12 miesięcy.

Źródło: NBP.

dlatego w bilansie pojawia się też czwarta pozycja: saldo błędów i opuszczeń, która jest składnikiem bilansującym (*balance item*) [Fausten, Brooks 1996].

Bardzo ważne dla oceny kondycji gospodarki jest saldo obrotów bieżących, wyliczane na podstawie transakcji bilansu bieżącego. W ciągu ostatnich kilkunastu lat było ono zazwyczaj w Polsce ujemne. Przykładowy bilans płatniczy za lipiec 2011 r. znajduje się w tab. 1.

Wszystkie transakcje, których nie można zaliczyć do żadnej pozycji bilansu płatniczego, tworzą tzw. saldo błędów i opuszczeń (*errors and omissions*). Problem, jaki się pojawił od roku 2004, to wysokie, ujemne i silnie rosnące saldo błędów i opuszczeń, co wywołało niepokój MFW oraz analityków rynkowych (rys. 2).



Rys. 2. Saldo błędów i opuszczeń w bilansie płatniczym Polski w latach 2004-2011

Źródło: NBP.

W raportach banków inwestycyjnych w marcu 2011 r. pojawiły się obawy, że w Polsce występuje zjawisko tzw. podwójnego deficytu (*twin deficit*), co oznacza sytuację, gdy deficyt budżetowy oraz deficyt na rachunku bieżącym w stosunku do PKB są na zbliżonym poziomie (ok. 7-8% PKB) [Flash Comment... 2011; Emerging Markets... 2011].

Banki inwestycyjne, np. UBS, sugerowały od razu, że tak duża wielkość tej pozycji bilansu płatniczego wynika z niedoszacowania importu w wysokości ok. 10-20 mld dol. [Tip Notes 2011]. Wskazywały one również na niezgodności w niemieckich statystykach eksportu i polskich statystykach importu [CEEMEA Monitor 2011].

Pozycja bilansu płatniczego „saldo błędów i opuszczeń” jest tak uważnie obserwowana, ponieważ wpływa na kształtowanie się wielkości PKB w ujęciu wydatkowym, zgodnie z równaniem:

$$\text{PKB} = I + C + G + \text{Eksport}_{\text{netto}},$$

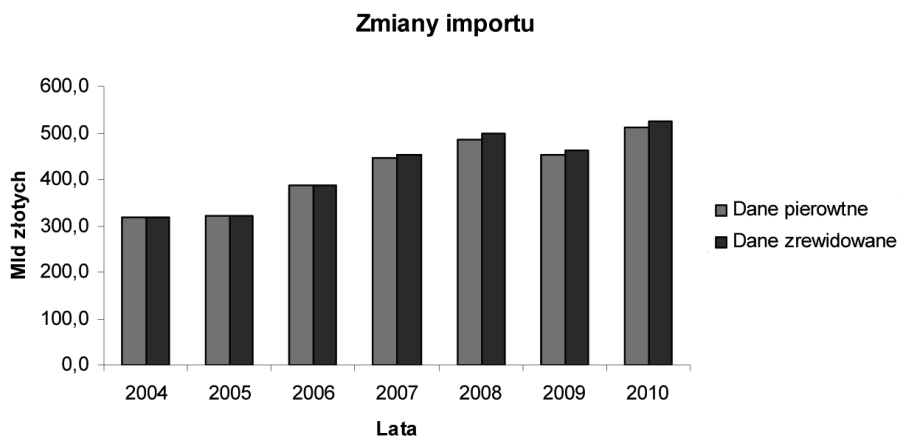
gdzie: I – inwestycje, C – konsumpcja, G – wydatki rządowe, $\text{Eksport}_{\text{netto}} = \text{Eksport} - \text{Import}$.

Większy import oznacza niższą wielkość pozycji Eksport_netto i niższy PKB, a to oznacza większy deficyt i większy dług publiczny w relacji do PKB.

Wysokie saldo błędów i opuszczeń (ok. 3% PKB w 2010 r.) wskazywało na to, że został nieoszacowany import. W efekcie głębokich studiów prowadzonych przez Departament Statystyki NBP i weryfikacji danych zidentyfikowano luki w danych obejmujące [*Handel zagraniczny...* 2011]:

- Import samochodów używanych z obszaru UE.
- Pozostałe transfery (transfery przekazywane do Polski z tytułów innych niż dochody z pracy czy też środki z Unii Europejskiej).
- Branding (import usług z zagranicy, np. w postaci opłaty za korzystanie ze znaku towarowego).
- Transakcje z przyrzeczeniem odkupu (dane o transakcjach warunkowych: repo, sell/buy back, buy/sell back, nie powinny być pokazywane jako napływ lub odpływ inwestycji portfelowych).
- Aktywa zagraniczne podmiotów sektora niefinansowego (SPV – podmiotów specjalnego przeznaczenia, np. na Cyprze).
- Operacje finansowe najbogatszych Polaków.

W czerwcu 2011 r. dokonano rewizji Bilansu Płatniczego Polski, w której najważniejsze zmiany dotyczyły importu. Import uległ zwiększeniu, gdyż według danych Ministerstwa Finansów w 2010 r. sprowadzono ponad 740 tys. samochodów za łączną kwotę 11,3 mld zł. Zrewidowane wielkości importu pokazano na rys. 3.



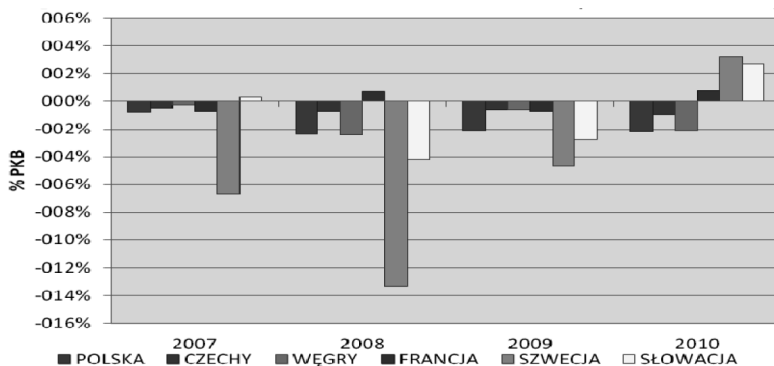
Rys. 3. Rewizje wielkości importu Polski w latach 2004-2010

Źródło: NBP.

Powodem tak dużego niedoszacowania importu były zbyt niskie prognozy sprawozdawcze systemu statystyki celnej INTRA-STAT i mało elastyczny system sprawozdawczy GUS.

Dokonana rewizja nie miała wpływu na wielkość PKB i wielkość deficytu, ponieważ jednocześnie zwiększyła się również konsumpcja.

Omawiany powyżej problem występuje też w innych krajach Unii Europejskiej. Na rysunku 4 pokazano względne wielkości salda błędów i opuszczeń w stosunku do PKB. Jak widać, Polska ma podobny poziom tego wskaźnika jak Węgry i lokuje się w środku przedstawionej grupy krajów. Warto dodać, że prawie wszystkie kraje mają ujemną wielkość tej pozycji bilansu płatniczego.



Rys. 4. Wielkość salda błędów i opuszczeń w stosunku do PKB w wybranych krajach UE

Źródło: NBP.

6. Podsumowanie

Kryzys finansowy, najsilniej oddziałujący na gospodarkę USA i strefy euro, ujawnił występowanie braków wielu istotnych informacji. Obejmowały one dane dotyczące nowych instrumentów opartych na ryzyku, aktywów w ujęciu rodzajowym i geograficznym oraz waluty, w jakiej denominowane są aktywa. To ostatnie jest ważne ze względu na znaczenie ryzyka kursowego.

Okazało się także, że banki centralne nie zwracały należytej uwagi na instytucje finansowe znajdujące się poza sektorem bankowym oraz na instytucje o charakterze globalnym.

Te ujawnione braki danych spowodowały, że banki centralne rozpoczęły rozbudowę swoich systemów sprawozdawczych, np. *Security Holdings Statistics* oraz *Insurance Statistics*.

Literatura

- Bilans płatniczy Polski w lipcu 2011*, Departament Statystyki, NBP, Warszawa 2011.
 CEEMEA Monitor, Morgan Stanley Research, 26.03.2011.
 Emerging Markets Biweekly Report, ING, 25.03.2011.

- Fausten D.K., Brooks R.D., *The balancing item in Australia's balance of payment accounts: an impressionistic view*, „Applied Economics” 1996, no 28.
- Flash Comment, Danske Markets, Danske Bank, 23.03.2011.
- GEMs Daily, Bank of America/Merrill Lynch, London Edition, 29.03.2011.
- Handel zagraniczny w statystykach lustrzanych*, Departament Statystyki, NBP, Warszawa, 30.03.2011.
- Poles apart*, „Financial Times”, 11.04.2011.
- Tip Notes, UBS Investment Research, UBS Limited, 21.02.2011.

DATA QUALITY IN CENTRAL BANKS' STATISTICAL SYSTEMS (NBP EXAMPLE)

Summary: The aim of this paper is to assess the quality of statistical data collected and processed by the statistical systems of the central banks. A good example of this quality is the position of the Poland's balance of payment named “errors and omissions”. Data sent to the National Bank of Poland by the Ministry of Finance, Statistical Office and Supervision Authority is not complete and led to large deficit on the current account to GDP. Investigations performed by the Department of Statistics in the National Bank of Poland resulted in revision of the balance of payments in year 2011 of 10 bln PLN.

Keywords: statistics, finance, data quality, financial crisis.