

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 384

**Taksonomia 24**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**

Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego  
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania  
znajdują się na stronie internetowej Wydawnictwa  
[www.pracnaukowe.ue.wroc.pl](http://www.pracnaukowe.ue.wroc.pl)  
[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Publikacja udostępniona na licencji Creative Commons  
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska  
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2015

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)  
**e-ISSN 2392-0041** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)  
**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:  
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
tel./fax 71 36 80 602; e-mail:[econbook@ue.wroc.pl](mailto:econbook@ue.wroc.pl)  
[www.ksiegarnia.ue.wroc.pl](http://www.ksiegarnia.ue.wroc.pl)

Druk i oprawa: TOTEM

## Spis treści

Wstęp.....	9
<b>Krzysztof Jajuga, Józef Pociecha, Marek Walesiak:</b> 25 lat SKAD.....	15
<b>Beata Basiura, Anna Czapkiewicz:</b> Symulacyjne badanie wykorzystania entropii do badania jakości klasyfikacji.....	25
<b>Andrzej Bąk:</b> Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie <code>pllord</code> .....	33
<b>Justyna Brzezińska:</b> Analiza klas ukrytych w badaniach sondażowych.....	42
<b>Grażyna Dehnel:</b> Rejestr podatkowy oraz rejestr ZUS jako źródło informacji dodatkowej dla statystyki gospodarczej – możliwości i ograniczenia ..	51
<b>Sabina Denkowska:</b> Wybrane metody oceny jakości dopasowania w <i>Propensity Score Matching</i> .....	60
<b>Marta Dziechciarz-Duda, Klaudia Przybysz:</b> Zastosowanie teorii zbiorów rozmytych do identyfikacji pozafiskalnych czynników ubóstwa.....	75
<b>Iwona Foryś:</b> Potencjał rynku mieszkaniowego w Polsce w latach dekonjunktury gospodarczej.....	84
<b>Eugeniusz Gatnar:</b> Statystyczna analiza konwergencji krajów Europy Środkowej i Wschodniej po 10 latach członkostwa w Unii Europejskiej.....	93
<b>Ewa Genge:</b> Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa.....	100
<b>Alicja Grześkowiak:</b> Wielowymiarowa analiza uwarunkowań zaangażowania Polaków w kształcenie ustawiczne o charakterze pozaformalnym.....	108
<b>Monika Hamerska:</b> Wykorzystanie metod porządkowania liniowego do tworzenia rankingu jednostek naukowych.....	117
<b>Bartłomiej Jefmański:</b> Zastosowanie modeli IRT w konstrukcji rozmytego systemu wag dla zmiennych w zagadnieniu porządkowania liniowego – na przykładzie metody TOPSIS.....	126
<b>Tomasz Józefowski, Marcin Szymkowiak:</b> Wykorzystanie uogólnionej miary odległości do porządkowania liniowego powiatów województwa podkarpackiego w świetle funkcjonowania specjalnej strefy ekonomicznej Euro-Park Mielec.....	135
<b>Krzysztof Kompa:</b> Zastosowanie testów parametrycznych i nieparametrycznych do oceny sytuacji na światowym rynku kapitałowym przed kryzysem i po jego wystąpieniu.....	144
<b>Mariusz Kubus:</b> Rekurencyjna eliminacja cech w metodach dyskryminacji....	154

<b>Marta Kuc:</b> Wpływ sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego państw Unii Europejskiej pod względem poziomu życia ludności .....	163
<b>Paweł Lula:</b> Kontekstowy pomiar podobieństwa semantycznego .....	171
<b>Iwona Markowicz:</b> Model regresji Feldsteina-Horioki – wyniki badań dla Polski .....	182
<b>Kamila Migdał-Najman:</b> Ocena wpływu wartości stałej Minkowskiego na możliwość identyfikacji struktury grupowej danych o wysokim wymiarze .....	191
<b>Małgorzata Misztal:</b> O zastosowaniu kanonicznej analizy korespondencji w badaniach ekonomicznych.....	200
<b>Krzysztof Najman:</b> Zastosowanie przetwarzania równoległego w analizie skupień .....	209
<b>Edward Nowak:</b> Klasyfikacja danych a rachunkowość. Rozważania o relacjach .....	218
<b>Marcin Pelka:</b> Adaptacja metody <i>bagging</i> z zastosowaniem klasyfikacji pojęciowej danych symbolicznych.....	227
<b>Józef Pocięcha, Mateusz Baryła, Barbara Pawelek:</b> Porównanie skuteczności klasyfikacyjnej wybranych metod prognozowania bankructwa przedsiębiorstw przy losowym i nielosowym doborze prób .....	236
<b>Agnieszka Przedborska, Małgorzata Misztal:</b> Wybrane metody statystyki wielowymiarowej w ocenie jakości życia słuchaczy uniwersytetu trzeciego wieku .....	246
<b>Wojciech Roszka:</b> Konstrukcja syntetycznych zbiorów danych na potrzeby estymacji dla małych domen .....	254
<b>Aneta Rybicka:</b> Połączenie danych o preferencjach ujawnionych i wyrażonych .....	262
<b>Elżbieta Sobczak:</b> Poziom specjalizacji w sektorach intensywności technologicznej a efekty zmian liczby pracujących w województwach Polski ....	271
<b>Andrzej Sokołowski, Grzegorz Harańczyk:</b> Modyfikacja wykresu radarowego .....	280
<b>Marcin Szymkowiak, Marek Witkowski:</b> Wykorzystanie mediany do klasyfikacji banków spółdzielczych według stanu ich kondycji finansowej ..	287
<b>Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski:</b> Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych – zastosowanie testu <i>join-count</i> .....	296
<b>Dorota Witkowska:</b> Wykorzystanie drzew klasyfikacyjnych do analizy zróżnicowania płac w Niemczech .....	305
<b>Artur Zaborski:</b> Analiza niesymetrycznych danych preferencji z wykorzystaniem modelu punktu dominującego i modelu grawitacji.....	315

## Summaries

<b>Krzysztof Jajuga, Józef Pocięcha, Marek Walesiak:</b> XXV years of SKAD	24
<b>Beata Basiura, Anna Czapkiewicz:</b> Simulation study of the use of entropy to validation of clustering.....	32
<b>Andrzej Bąk:</b> Problem of choosing the optimal linear ordering procedure in the p_llord package.....	41
<b>Justyna Brzezińska-Grabowska:</b> Latent class analysis in survey research...	50
<b>Grażyna Dehnel:</b> Tax register and social security register as a source of additional information for business statistics – possibilities and limitations.....	59
<b>Sabina Denkowska:</b> Selected methods of assessing the quality of matching in Propensity Score Matching .....	74
<b>Marta Dziechciarz-Duda, Klaudia Przybysz:</b> Applying the fuzzy set theory to identify the non-monetary factors of poverty.....	83
<b>Iwona Foryś:</b> The potential of the housing market in Poland in the years of economic recessions.....	92
<b>Eugeniusz Gatnar:</b> Statistical analysis of the convergence of CEE countries after 10 years of their membership in the European Union.....	99
<b>Ewa Genge:</b> Trust to the public and financial institutions in the Polish society – an application of latent Markov models.....	107
<b>Alicja Grześkowiak:</b> Multivariate analysis of the determinants of Poles' involvement in non-formal lifelong learning .....	116
<b>Monika Hamerska:</b> The use of the methods of linear ordering for the creating of scientific units ranking.....	125
<b>Bartłomiej Jefmański:</b> The application of IRT models in the construction of a fuzzy system of weights for variables in the issue of linear ordering – on the basis of TOPSIS method .....	134
<b>Tomasz Józefowski, Marcin Szymkowiak:</b> GDM as a method of finding a linear ordering of districts of Podkarpackie Voivodeship in the light of the operation of the Euro-Park Mielec special economic zone .....	143
<b>Krzysztof Kompa:</b> Application of parametric and nonparametric tests to the evaluation of the situation on the world financial market in the pre- and post-crisis period.....	153
<b>Mariusz Kubus:</b> Recursive feature elimination in discrimination methods ...	162
<b>Marta Kuc:</b> The impact of the spatial weights matrix on the final shape of the European Union countries ranking due to the standard of living.....	170
<b>Paweł Lula:</b> The impact of context on semantic similarity.....	181
<b>Iwona Markowicz:</b> Feldstein-Horioka regression model – the results for Poland.....	190

<b>Kamila Migdal-Najman:</b> The assessment of impact value of Minkowski's constant for the possibility of group structure identification in high dimensional data.....	199
<b>Małgorzata Misztal:</b> On the use of canonical correspondence analysis in economic research.....	208
<b>Krzysztof Najman:</b> The application of the parallel computing in cluster analysis.....	217
<b>Edward Nowak:</b> Data classification and accounting. A study of correlations	226
<b>Marcin Pelka:</b> The adaptation of bagging with the application of conceptual clustering of symbolic data.....	235
<b>Józef Pocięcha, Mateusz Baryła, Barbara Pawelek:</b> Comparison of classification accuracy of selected bankruptcy prediction methods in the case of random and non-random sampling technique.....	244
<b>Agnieszka Przedborska, Małgorzata Misztal:</b> Selected multivariate statistical analysis methods in the evaluation of the quality of life of the members of the University of the Third Age.....	253
<b>Wojciech Roszka:</b> Construction of synthetic data sets for small area estimation.....	261
<b>Aneta Rybicka:</b> Combining revealed and stated preference data.....	270
<b>Elżbieta Sobczak:</b> Specialization in sectors of technical advancement vs. effects of workforce number changes in Poland's voivodships.....	279
<b>Andrzej Sokółowski, Grzegorz Harańczyk:</b> Modification of radar plot.....	286
<b>Marcin Szymkowiak, Marek Witkowski:</b> Classification of cooperative banks according to their financial situation using the median.....	295
<b>Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski:</b> The influence of classification method selection on the identification of spatial dependence – an application of join-count test.....	304
<b>Dorota Witkowska:</b> Application of classification trees to analyze wages disparities in Germany.....	314
<b>Artur Zaborski:</b> Asymmetric preference data analysis by using the dominance point model and the gravity model.....	323

**Aneta Rybicka**

Uniwersytet Ekonomiczny we Wrocławiu

e-mail: aneta.rybicka@ue.wroc.pl

---

## POŁĄCZENIE DANYCH O PREFERENCJACH UJAWNIONYCH I WYRAŻONYCH

---

**Streszczenie:** W pomiarze preferencji wykorzystuje się obserwacje historyczne oraz dane o charakterze antycypacyjnym. Rozróżniamy zatem metody analizy preferencji ujawnionych i preferencji wyrażonych. **Preferencje ujawnione** stanowią odbicie rzeczywistych decyzji rynkowych konsumentów. Podstawą analizy jest tu materiał statystyczny zgromadzony w wyniku rejestracji danych o rzeczywistych wyborach rynkowych konsumentów. **Preferencje wyrażone** dotyczą natomiast hipotetycznych zachowań rynkowych konsumentów. Jednakże w badaniach preferencji konsumentów można połączyć źródła danych o preferencjach ujawnionych i wyrażonych. W artykule przedstawiono charakterystykę preferencji ujawnionych i wyrażonych oraz połączenie ich mocnych stron. Zaprezentowano również typy połączenia danych oraz modele ekonometryczne wykorzystywane w estymacji łącznej.

**Słowa kluczowe:** preferencje ujawnione, preferencje wyrażone, połączenie źródeł danych o preferencjach.

DOI: 10.15611/pn.2015.384.28

### 1. Wstęp

W pomiarze preferencji wykorzystuje się obserwacje historyczne oraz dane o charakterze antycypacyjnym, opisujące intencje konsumentów. W związku z tym różniamy metody analizy preferencji ujawnionych i metody analizy preferencji wyrażonych.

**Preferencje ujawnione** (*Revealed Preferences* – RP) stanowią odbicie rzeczywistych decyzji rynkowych konsumentów. Najważniejszą zaletą danych o preferencjach ujawnionych jest to, że bazują na aktualnych, rzeczywistych wyborach. Natomiast główną wadą tego typu podejścia jest to, że nowe produkty są poza zasięgiem danych historycznych.

**Preferencje wyrażone** (*Stated Preferences* – SP) dotyczą natomiast hipotetycznych zachowań rynkowych konsumentów. Zaletą podejścia preferencji wyrażonych jest jego elastyczność. Metody te mogą być wykorzystywane w badaniach

dotyczących nowych (hipotetycznych) produktów lub usług (profilów wyboru). Często tylko intencje respondentów są źródłem informacji. Główną wadą danych wyrażonych jest to, że mają „hipotetyczną” naturę.

Praca ma charakter wprowadzający do tematyki połączenia danych, gdzie dokonuje się syntetycznej oceny stanu wiedzy. Artykuł nie zawiera przykładów, ma bowiem charakter przeglądowy i teoretyczny, mieści się w obszarze badań preferencji. Celem artykułu jest dokonanie przeglądu w zakresie rozwiązań zaprezentowanych w literaturze przedmiotu, które pozwalają na połączenie informacji z danych o preferencjach wyrażonych i ujawnionych. Zaprezentowano w nim również ich wady i zalety.

## **2. Połączenie mocnych stron danych o preferencjach ujawnionych i wyrażonych**

Połączenie danych oraz łączna estymacja ujawnionych oraz wyrażonych preferencji są przeprowadzane tak, by wykorzystać zalety obu grup danych przy jednoczesnej minimalizacji wad. Połączenie obu typów danych zostało określone jako paradygmat „wzbogacenia danych” [Louviere i in. 2000].

Dane o preferencjach ujawnionych mogą zostać „wzmocnione” poprzez dane o preferencjach wyrażonych. Wybory reprezentujące preferencje ujawnione ograniczone są tylko do aktualnych na rynku i dotyczą tylko obecnych konsumentów. Badania preferencji wyrażonych pozwalają na wybór między opcjami (profilami) nie tylko historycznymi. Połączenie danych pozwala zatem na rozszerzenie modelu zachowań poza ograniczony zakres danych historycznych. Wadą danych o preferencjach ujawnionych jest również mała efektywność gromadzenia danych. Często ankiety dotyczące preferencji ujawnionych gromadzą pojedyncze informacje. Zgromadzenie większej liczby informacji w badaniach przekrojowych lub panelowych znacząco zwiększa koszt badań oraz powoduje, że w badaniu możemy tracić respondentów (co prowadzi do problemów z doбором próby).

Dodatkowe wybór/wybory (wyrażone), jakich dokonują respondenci, pozwalają na zgromadzenie większej liczby informacji na temat preferencji każdego z nich, a więcej informacji od każdego z respondentów może prowadzić do poprawy jakości modelu. Zatem jeśli w badaniu, oprócz danych o preferencjach ujawnionych, wykorzystamy również dane o preferencjach wyrażonych, uzyskamy lepszą jakość modelu, a zatem przy mniejszej próbie otrzymamy te same przedziały ufności, które by otrzymano w badaniu wykorzystującym tylko dane o preferencjach ujawnionych z większą próbą.

Dane o preferencjach wyrażonych mogą zostać wzmocnione poprzez dane o preferencjach ujawnionych. W wielu przypadkach wybory hipotetyczne nie biorą pod uwagę budżetu lub innych istotnych ograniczeń. W badaniach respondenci dokonują hipotetycznych wyborów, jednakże gdy dokonywane są rzeczywiste wy-



bory na rynku, respondenci biorą pod uwagę pewne ograniczenia i w rzeczywistości dokonują wyborów różniących się od deklarowanych. Połączenie danych pozwala zatem na to, by po konfrontacji z rzeczywistością hipotetyczne wybory zostały „sprowadzone na ziemię” [Louviere i in. 2000].

Wiele przykładów badań przedstawionych jest w pracy autorów J.C. Whitehead, S.K. Pattanayak, G.L. Houtven, B.R. Gelso (2007), *Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science*, Journal of Economic Surveys, Vol. 22, Issue 5, 872-908.

W standardowym badaniu preferencji wyrażonych profile są zbudowane bez odniesienia do wyboru respondenta reprezentującego podejście ujawnione (nawet gdy dane o preferencjach ujawnionych są łączone z danymi o preferencjach wyrażonych). Jednakże naukowcy zaczęli wykorzystywać eksperymenty z wykorzystaniem danych o preferencjach wyrażonych, które są projektowane (konstruowane) na podstawie danych o preferencjach ujawnionych (na podstawie wyborów respondentów). Na przykład w badaniu dotyczącym podróży [Rose i in. 2005] zwrócono się do respondentów o opisanie ostatniej podróży. Hipotetyczne trasy (preferencje wyrażone) są opisane przez czas i koszty, które różnią się pewną wartością (poniżej lub powyżej) opisaną. Następnie respondenci proszeni są o dokonanie wyboru spośród hipotetycznych profilów dróg.

**Tabela 1.** Porównanie charakterystyk danych o preferencjach ujawnionych i danych o preferencjach wyrażonych

Dane o preferencjach ujawnionych	Dane o preferencjach wyrażonych
Oparte są na rzeczywistych zachowaniach na rynku	Oparte są na scenariuszach hipotetycznych (deklaracjach, intencjach)
Atrybuty są skorelowane	Atrybuty, dzięki eksperymentowi, nie są skorelowane
Zakres atrybutów ograniczony jest rynkiem (atrybuty rzeczywiste)	Dowolny zakres atrybutów (rzeczywiste i hipotetyczne)
Trudno zmierzyć (z badać) wartości niematerialne i prawne	Wartości niematerialne i prawne mogą być włączone do badania
Wskaźnikiem preferencji jest wybór opcji, profilu	Wskaźnikiem preferencji może być porządkowanie, ocena lub wybór hipotetycznych opcji, profilu (intencji)

Źródło: [Combining Revealed... 2000, s. 211].

Zastosowania tego typu konstrukcji i połączenia danych przedstawiają prace:

1. D. Hensher, W. Greene (2003), *Mixed logit models: State of practice*, Transportation 30, 133-176,
2. D. Hensher (2004), *Accounting for stated choice design dimensionality in willingness to pay for travel time savings*, Transportation Research B 38, 425-446.
3. D. Hensher (2006), *How do respondents process stated choice experiments? Attribute consideration under varying information load, forthcoming*, Journal of Applied Econometrics, Vol. 21, Issue 6, 861-878.

4. S. Caussade, de D. Ort'uzar J., Rizzi L. and D. Hensher (2005), *Assessing the influence of design dimensions on stated choice experiment estimates*, Transportation Research B 39, 621-640.

Zaletą tego typu konstrukcji połączenia danych jest zaprojektowanie eksperymentu, który w sposób bardziej realistyczny przedstawia profile (ponieważ profile te opisane są poziomami atrybutów „bliskimi” poziomom przedstawiającym preferencje ujawnione). Tabela 1 przedstawia porównanie charakterystyk danych o preferencjach ujawnionych i wyrażonych.

### 3. Typy połączenia danych i modele

W literaturze zaproponowanych jest kilka sposobów połączenia danych. W celu wyodrębnienia typów połączenia danych zaproponowano dwa kryteria [Whitehead i in., s. 11]:

1. Czy dane mogą być rozcięte (*stacked data*)? Dane mogą być „rozcięte”, jeśli obserwacje o preferencjach ujawnionych i wyrażonych mają podobne zmienne zależne i niezależne. Badacz traktuje dane o preferencjach wyrażonych jako dodatkowe obserwacje i zwiększa rozmiar badania poprzez włączenie ich do danych o preferencjach ujawnionych. Zakłada się tu równość oszacowanych parametrów dla preferencji ujawnionych i wyrażonych.

2. Czy jest spełnione założenie o niezależności i identyczności błędu pomiaru? Założenie w analizie regresji o błędzie pomiaru jest takie, że jest on niezależny i identyczny. Gdy łączymy dane o preferencjach ujawnionych i wyrażonych, często założenie to jest niespełnione, ponieważ błędy pomiaru wśród respondentów są skorelowane.

Tabela 2 przedstawia 4 typy połączenia danych.

**Tabela 2.** Typy połączenia danych o preferencjach

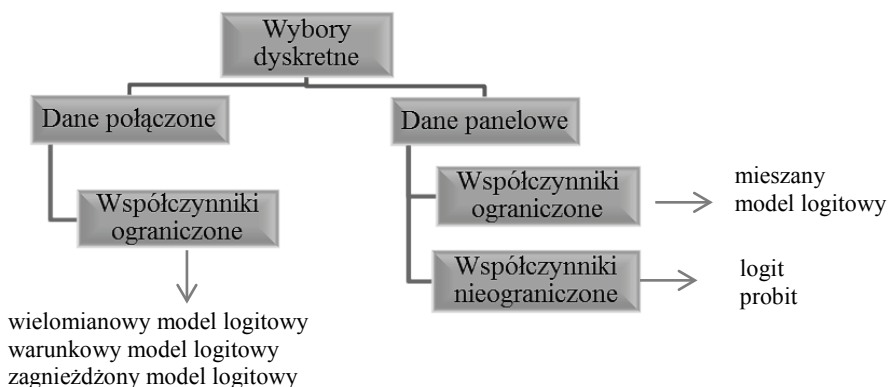
	Niezależne i identyczne rozkładu błędów	Skorelowana struktura błędów
Dane rozcięte ( <i>stacked data</i> )	studia na podstawie danych połączonych ( <i>pool data</i> )	studia na podstawie danych panelowych
Pozostałe formy	studia porównawcze	studia na podstawie danych mieszanych

Źródło: [Whitehead i in 2007, s. 54].

Trzy z czterech typów badań z zastosowaniem połączenia danych są szacowane wspólnie. Wspólna estymacja występuje, gdy relację między zmiennymi niezależnymi a zmiennymi zależnymi są szacowane w jednym, dwóch lub więcej równaniach modelu w tym samym czasie. Wyjątkiem są badania porównawcze.

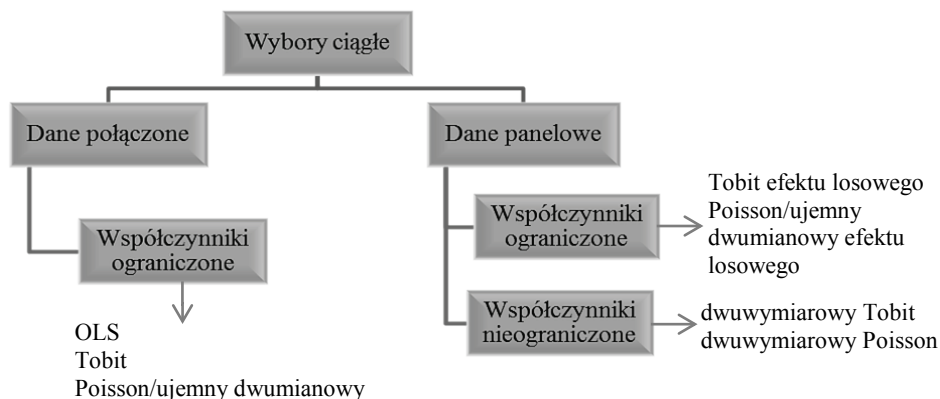
#### 4. Typy modeli ekonometrycznych wykorzystywanych w estymacji łącznej

Dane o preferencjach ujawnionych i wyrażonych mogą zostać połączone i analizowane z wykorzystaniem trzech głównych modeli ekonometrycznych: modele wyborów dyskretnych (np. logit, probit), modele wyborów ciągłych (np. Tobit, Poisson) oraz modele wyborów mieszanych [Whitehead i in. 2007, s. 13]. Rysunki 1-3 przedstawiają modele ekonometryczne wykorzystywane w analizie 3 typów połączenia danych.



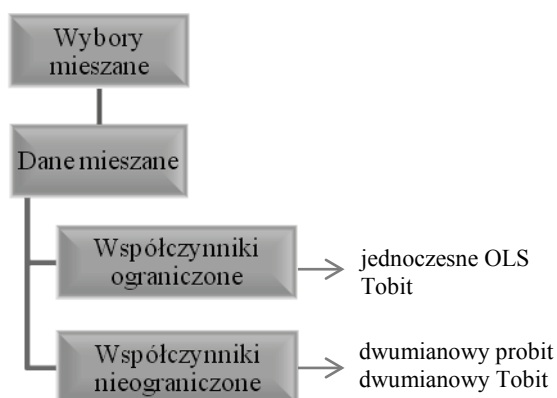
**Rys. 1.** Modele ekonometryczne wykorzystywane w łącznej estymacji danych o preferencjach ujawnionych i wyrażonych – modele wyborów dyskretnych

Źródło: [Whitehead i in. 2007, s. 55].



**Rys. 2.** Modele ekonometryczne wykorzystywane w łącznej estymacji danych o preferencjach ujawnionych i wyrażonych – modele wyborów ciągłych

Źródło: [Whitehead i in. 2007, s. 55].



**Rys. 3.** Modele ekonometryczne wykorzystywane w łącznej estymacji danych o preferencjach ujawnionych i wyrażonych – modele wyborów mieszanych

Źródło: [Whitehead i in. 2007, s. 55].

Dokładniej modele wykorzystywane w estymacji łącznej przedstawia praca [Whitehead i in. 2007].

## 5. Przykłady w literaturze

W literaturze przedmiotu przedstawionych jest wiele przykładów połączenia danych o preferencjach ujawnionych i wyrażonych. Do popularniejszych należą [Paccagnan 2007, s. 10-11]:

1. T.A. Cameron (1992), *Combining contingent valuation and travel cost data for the valuation of nonmarket goods*, Land Economics, 68, 302-317.

Praca ta była pierwszą, która przedstawiała połączenie wyrażonych i ujawnionych zachowań na rynku. Oszacowano parametry modelu, wykorzystując oba typy danych o preferencjach oraz funkcję największej wiarygodności. Badano rybołówstwo rekreacyjne. Poproszono respondentów o wybór między: zaprzestaniem korzystania z zasobów (a tym samym uniknięcia podatków) a kontynuowaniem połowów (uzyskując niższe dochody). Wykorzystano informacje na temat aktualnych wyborów (podjętych decyzji) oraz pytano o chęci do zapłaty.

2. W. Adamowicz i in. (1994), *Combining revealed and stated preferences methods for valuing environmental amenities*, Journal of Environmental Economics and Management, 26, 271-292.

Praca przedstawia połączenie danych o preferencjach, wykorzystując podstawę użyteczności losowej. Oszacowano dwa osobne modele dla danych o preferencjach ujawnionych i wyrażonych, a następnie wspólny model, łączący oba typy danych, oraz porównano wiarygodności estymacji łącznej oraz sumę wiarygodności oddzielnych. Badanie przeprowadzono w celu oceny wartości walorów środowi-

skowych. Respondentów poproszono o wybór jednego z trzech profili: rekreacja nad wodą „stojącą”, rekreacją nad wodą „płynącą” – cieki wodne oraz zajęcia rekreacyjne niewymagające wody (atrybuty to m.in.: odległość od miejsca, jakość wody, opłata za rybołówstwo). Również zgromadzono dane na temat aktualnych zachowań, wyborów rynkowych respondentów.

3. J. Englin, T. Cameron (1996), *Augmenting travel cost models with contingent behaviour data*, Environmental and Resource Economics.

W pracy tej przedstawiono po raz pierwszy wykorzystanie danych panelowych w połączeniu danych o preferencjach.

4. J.C. Whitehead i in. (2000), *Measuring recreation benefits of quality improvement with revealed and stated behaviour data*, Resource and Energy Economics, 22, 339-354.

Praca przedstawia połączenie danych w pomiarze korzyści rekreacyjnych wynikających z poprawy jakości wody.

5. J.C. Whitehead, S.K. Pattanayak, G.L. Houtven, B.R. Gelso (2007), *Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: an Assessment of the State of the Science*.

Praca ta przedstawia wiele przykładów badań wykorzystujących połączenie danych.

## 6. Podsumowanie

Możliwość połączenia danych o preferencjach ujawnionych i preferencjach wyrażonych pozwala np. na oszacowanie wpływu obu typów preferencji na indywidualną skłonność do zapłaty za zmiany w produktach lub usługach.

Zarówno badania preferencji ujawnionych, jak i badania preferencji wyrażonych mają swoje wady. Jednakże połączenie obu podejść pozwala, by wady jednych zostały zminimalizowane zaletami drugich. Tzn. urealniamy hipotetyczne preferencje respondentów, a preferencje ujawnione rozszerzamy na respondentów i rynek wykraczający poza rzeczywistość. Połączenie danych pozwala również na to by:

- zmniejszyć koszt pozyskania, otrzymania niezbędnych informacji (ponieważ znacznym kosztem w badaniu jest koszt gromadzenia danych, a zgromadzenie dodatkowych informacji przez pytanie/pytania wyboru odbywa się niewielkim kosztem i czasem),
- zmniejszyć szerokość przedziału ufności „chęci do zapłaty” (chęci wyboru danego profilu, zapłaty za niego),
- poprawić dopasowanie modelu.

## Literatura

- Adamowicz W., Louviere J., Williams M., 1994, *Combining revealed and stated preference methods for valuing environmental amenities*, Journal of Environmental Economics and Management 26(3), 271-292.
- Ben-Akiva M., Morikawa T., 1990, *Estimation of switching models from revealed preferences and stated intentions*, Transportation Research A 24, 485-495.
- Bhat C., Castelar S., 2002, *A unified mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco bay area*, Transportation Research B 36, 577-669.
- Brownstone D., Bunch D., Train K., 2000, *Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles*, Transportation Research B 34, 315-338.
- Cameron T.A., 1992, *Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods*, Land Economics 68(3), 302-317.
- Combining Revealed and Stated Preference Data*, 2000, Quantitative Choice Analysis Workshop, Econometrics Laboratory, University of California at Berkeley, 22-24 May 2000 [http://elsa.berkeley.edu/eml/qca\\_reader/9.combin.pdf](http://elsa.berkeley.edu/eml/qca_reader/9.combin.pdf), (22.08.2014).
- Earnhart D., 2001, *Combining revealed and stated preference methods to value environmental amenities at residential locations*, Land Economics 77(1), 12-29.
- Earnhart D., 2002, *Combining revealed and stated data to examine housing decisions using discrete choice analysis*, Journal of Urban Economics 51(1), 143-169.
- Franses P.H., Montgomery A.L. (red.), 2002, *Econometric models in marketing*, Vol. 16, Elsevier Science.
- Hensher D., Bradley M., 1993, *Using stated response data to enrich revealed preference discrete choice models*, Marketing Letters 4, 39-152.
- Hensher D., Louviere J., Swait J., 1999, *Combining sources of preference data*, Journal of Econometrics, 89, 197-221.
- Hensher D.A., Rose J.M., Green W.H., 2005, *Applied Choice Analysis. A Primer*, Cambridge University Press.
- Louviere J.J., Hensher D.A., Swait J., 2000, *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*, Cambridge University Press.
- McConnell K.E., Quinn W., Ivar E.S., 1999, *Joint Estimation of Contingent Valuation and Truncated Recreational Demands*, [in:] *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, ed. by J.A. Herriges, C.L. Kling, Cheltenham, UK: Edward Elgar.
- Paccagnan V., 2007, *On combining stated preferences and revealed preferences approaches to evaluate environmental resources having a recreational use*, MPRA Paper no. 5867.
- Rose J., Bliemer M., Hensher D., Collins A., 2005, *Designing efficient stated choice experiments involving respondent based reference alternatives*, Working paper, Institute of Transport and Logistics Studies, University of Sydney.
- Train K., Wilson W.W., 2006, *Estimation on Stated-Preference Experiments Constructed from Revealed-Preference Choices*.
- Whitehead J.C., Pattanayak S.K., Van Houtven G.L., Gelso B.R., 2007, *Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: an Assessment of the State of the Science*, <http://econ.appstate.edu/RePEc/pdf/wp0519.pdf> (12.08.2014).

## COMBINING REVEALED AND STATED PREFERENCE DATA

**Summary:** In the preference measurement usually we use historical and hypothetical data that describe the consumers' intentions. Taking into consideration these preference types we can distinguish revealed preferences (RP) and stated preferences (SP). **Revealed preferences** reflect real market choices. We gather the data about real market choices made by consumers in time. Other data can be provided by using questionnaires where we ask about choices that have been made. Revealed preferences are based on historical data. **Stated preferences** are the hypothetical market (buying) behaviours of consumers. Methods that evaluate stated preference data use prior data gathered during interview. However, two different types of preference data (real and hypothetical) can be combined. The paper presents the characteristics of revealed and stated preferences, and what we get by combining them. Also methods and models that allow to combine and estimate such data are presented.

**Keywords:** revealed preferences, stated preferences, combining different preference data sources.