

Użyteczna sztuka rankingów ekonomiczno-społecznych

Nadesłany: 27 stycznia 2010 r.

Zaakceptowany: 7 marca 2010 r.

Streszczenie

Artykuł identyfikuje zalety i wady różnych metod sporządzania rankingów. Temat jest ważny i aktualny, choćby wobec dyskusji o metodach oceny poziomu rozwoju i dobrobytu w poszczególnych krajach. W artykule, na przykładzie krajów Unii Europejskiej, rozpatruje się metody rangowania obiektów pochodzące z wielowymiarowej analizy statystycznej oraz metody wywodzące się z optymalizacji wielokryterialnej. Porównania metod odbywają się w dwóch płaszczyznach: wpływu na porządek obiektów zbioru zmiennych diagnostycznych przy ustalonej metodzie wykonania rankingu albo wpływu wybranej metody przy ustalonym zbiorze zmiennych diagnostycznych. Nasze spostrzeżenia i obliczenia można traktować również jako jeden z głosów w wymianie poglądów toczącej się wokół raportu Komisji ds. Mierzenia Wydajności Ekonomicznej i Postępu Społecznego kierowanej przez J.E. Stiglitz i A. Sena.

Wprowadzenie

Celem artykułu jest identyfikacja zalet i wad różnych metod sporządzania rankingów. Nasze spostrzeżenia i obliczenia można traktować również jako jeden z głosów w dyskusji toczącej się wokół raportu komisji kierowanej przez J.E. Stiglitz i A. Sena. Empiryczną ilustracją rozważań jest ranking krajów Unii Europejskiej ze względu na stopień rozwoju ekonomiczno-społecznego.

* Dr hab., prof. nadzw., Wyższa Szkoła Finansów i Zarządzania w Warszawie oraz Szkoła Główna Handlowa w Warszawie.

** Mgr, Wyższa Szkoła Finansów i Zarządzania w Warszawie, doktorantka w Szkole Głównej Handlowej w Warszawie.

Wszystkiemu jest winien Simon Kuznets¹. Jego wieloletnia praca doprowadziła do oszacowania wielkości produktu narodowego brutto (PNB) Stanów Zjednoczonych na początku lat 40. ubiegłego wieku. Z czasem pomiar PNB, a potem produktu krajowego brutto (PKB), stał się jednym ze standardowych efektów pracy instytucji statystycznych w większości krajów świata. PKB po koniecznych przeliczeniach² zaczęto używać do porównywania stopnia rozwoju gospodarek krajowych. W dalszej kolejności takie porównania zaczęto nazywać porównaniami stopnia rozwoju czy poziomu dobrobytu. O tym, że wzrostu gospodarczego nie wolno utożsamiać z rozwojem społeczeństwa i gospodarki, czy ze wzrostem dobrobytu społecznego, wiadomo od dawna³. Znane są również argumenty przeciwników używaniu wielkości PKB *per capita* do takich celów. Jedną z najważniejszych wad wymienianego miernika tkwi w jego definicji. Sposób liczenia PKB nie uwzględnia np. wartości dóbr i usług wytwarzanych oraz przekazywanych w ramach obrotu niepieniężnego w gospodarstwach domowych i innych małych wspólnotach. Te gospodarki, w których ilość wytwarzanych oraz przekazywanych w ten sposób dóbr i usług jest znaczna, mają niedoszacowaną wartość produktu krajowego. Inna wada PKB jako uniwersalnego wskaźnika polega na tym, że wartość usług świadczonych przez państwo na rzecz swoich obywateli jest liczona według nakładów, a nie efektów tych nakładów. Efekt jest często rozłożony w czasie oraz trudno mierzalny, ponieważ ma wymiar nie tylko fizyczny, ale zależny od układu preferencji społecznych, a te mogą być różne w różnych krajach.

W ostatnich miesiącach jesteśmy świadkami dyskusji w mediach o potrzebie ujednolicenia pomiaru wyników ekonomicznych i rozwoju społecznego krajów. Bezpośrednią przyczyną wymiany poglądów jest opublikowany we wrześniu 2009 r. raport z prac Komisji ds. Mierzenia Wydajności Ekonomicznej i Postępu Społecznego, pracującej pod przewodnictwem J.E. Stiglitz i A. Sena, powołanej z inicjatywy Prezydenta Francji N. Sarkozy'ego na początku 2008 r., poświęcony możliwości opracowania powszechnie akceptowanego syntetycznego miernika rozwoju ekonomiczno-społecznego kraju. Potrzeba istnienia takiego miernika i niedoskonałość powszechnie stosowanego do porównań

¹ S. Kuznets otrzymał w 1971r. nagrodę Nobla w dziedzinie ekonomii za „uzasadnioną empirycznie interpretację wzrostu gospodarczego, która umożliwiła w nowy i głęboki sposób spojrzeć na sprawy struktury społeczno-ekonomicznej i proces jej rozwoju”. W latach 30. XX w. podjął badania nad produktem narodowym Stanów Zjednoczonych. Stworzył podstawy statystyki dochodu narodowego. Badanie dochodu narodowego było już wtedy rozwijane w takich krajach, jak: Wielka Brytania, Niemcy czy Szwecja, ale dopiero działania Kuzneta doprowadziły do wdrożenia prac nad rachunkiem dochodu narodowego do regularnej działalności organów statystyki państwowej.

² Czyli po obliczeniu PKB przypadającego na 1 osobę w danym kraju (*per capita*) i sprowadzeniu tej wielkości do wartości wyrażonej według parytetu siły nabywczej. Parytet siły nabywczej jest współczynnikiem odzwierciedlającym siłę nabywczą waluty krajowej na rynku tego kraju. Parytety siły nabywczej są stosowane dla przeliczenia waluty krajowej na umowną walutę, której jednostką jest np. USD, EUR albo tzw. standard siły nabywczej.

³ Porównania międzynarodowe winny być przemyślane. Przykładowo, wiadomość podana przez „Dziennik Gazeta Prawna” w numerze 251 (2627) z 24-27 grudnia 2009 r., zatytułowana „Polska gospodarka pokonała holenderską” raczej wprowadza w błąd niż dostarcza rzetelnej informacji o miejscu polskiej gospodarki wśród krajów Unii. Po nieco dokładniejszej lekturze artykułu dowiedzieć się można, że nasz PKB liczony według parytetu siły nabywczej wyniesie w 2009 r. ok. 535,7 mld EUR, a holenderski – ok. 515,0 mld EUR. Szkopuł w tym, że ludność Holandii liczy ok. 17 mln, a Polski – ok. 38 mln.

produktu krajowego brutto *per capita*, liczonego według parytetu siły nabywczej, w opinii większości ekonomistów nie podlega dyskusji. Uwzględnianie w porównaniach międzynarodowych wielkości PKB oznacza po prostu utożsamianie rozwoju ze wzrostem gospodarczym. Słusznie zauważył francuski ekonomista F. Perroux, że „rozwój to połączenie zmian w mentalności i zmian społecznych pewnej populacji, która zmierza do skumulowania i trwałego powiększenia swojego realnego produktu globalnego” (Bremont, Salort, 1997: 142).

N. Sarkozy nie jest pierwszym, który próbuje uwolnić porównania międzynarodowe od dyktatu PKB. W 1990 r. pakistański ekonomista M. Haq zaproponował miernik, na którego wartość składały się, odpowiednio przetworzone, oczekiwana długość życia, wskaźnik skolaryzacji, wskaźnik umiejętności czytania i pisania oraz wielkość PKB *per capita* liczona według parytetu siły nabywczej. Od 1993r. program ONZ ds. rozwoju (UNDP) publikuje Human Development Index (HDI, Wskaźnik Rozwoju Społecznego albo Wskaźnik Rozwoju Ludzkiego). Wspomniana agenda ONZ publikuje również wskaźniki ubóstwa społecznego (Human Poverty Index) HPI-1 – dla krajów rozwijających się, HPI-2 – dla najbardziej rozwiniętych państw. W 2001 r. w publikacjach UNDP pojawiła się kolejna miara nazwana syntetycznym wskaźnikiem osiągnięć technologicznych (TAI, Technology Achievement Index). Każda z tych miar jest skonstruowana zgodnie z podobną formułą arytmetyczną (Ochocki, 2009) pozwalającą wyznaczyć wartość syntetyczną miernika na podstawie wartości zmiennych diagnostycznych.

W 2006 r. powstała New Economics Foundation i opublikowała Happy Planet Index (Wskaźnik Szczęśliwej Planety). Miernik ten uwzględnia wartości trzech zmiennych diagnostycznych: spodziewaną długość życia, subiektywne zadowolenie z życia oraz tzw. ślad ekologiczny⁴.

W trwającej na świecie dyskusji jest również akcent polski. G.W. Kołodko (Kołodko, 2008: 270) pisze o Zintegrowanym Wskaźniku Pomyślności. Na jego wartość mają mieć wpływ zmiany zachodzące w czterech sferach życia. „Pierwsza to poziom produkcji, PKB *per capita*. Druga to sfera samopoczucia związanego ze zdrowiem własnym, stopniem zadowolenia z fazy życia za nami i realistycznych ocen co do jego długości przed nami, poziomem wykształcenia, jakością życia zbiorowego (rządzenie i demokracja). Trzecia to stan środowiska naturalnego i oszacowanie jego percepcji (wpływu na samopoczucie, zdrowie, kondycję do życia i pracy). Czwarta to ocena czasu wolnego i wartości kulturowych, którymi można go wypełniać. Kluczowy jest PKB, bo to materialna podstawa życia. Proponuję na tę sferę przeznaczyć dwie piąte całości. Na pozostałe trzy po jednej piątej.” Powyższy cytat ukazuje główną myśl autora koncepcji. Z jednej strony pozorna precyzja, ponieważ zostały podane wagi dla mierzenia wpływu poszczególnych „sfer” na ostateczną wartość miernika. Z drugiej zaś brak precyzji w definiowaniu zmiennych diagnostycznych dla sfer drugiej, trzeciej i czwartej. Analityczna postać miernika też wydaje się skomplikowana, ponieważ pomysł autora wyraźnie zmierza do konstruowania Zintegrowanego Wskaźnika Pomyślności jako miary syntetycznej z jednej zmiennej diagnostycznej (PKB *per capita*) i trzech zmiennych syntetycznych.

⁴ Ślad ekologiczny mierzy zapotrzebowanie ludności i gospodarki danego kraju na środowisko naturalne, stopień zużycia jego zasobów oraz zdolność do ich regeneracji.

Komisja kierowana przez J.E. Stiglitz i A. Sena, wskazując na niedoskonałości wykorzystywanych obecnie mierników poziomu rozwoju, przedstawia nowe propozycje podejścia do tematu. Zwraca uwagę na fakt, że jednym z powodów, dla których aktualny kryzys okazał się zaskoczeniem dla wielu rządów było skupienie uwagi na nieodpowiednich kryteriach oceny rozwoju gospodarczego. Autorzy raportu przyznają, że stworzenie idealnego wskaźnika pozwalającego przewidzieć z wyprzedzeniem podobne sytuacje może okazać się niemożliwe, uznają jednak zarazem, że załamanie gospodarcze unaczniło wady popularnie używanych statystyk. Nadal jednak pozostają one podstawą podejmowania decyzji gospodarczych i politycznych, a w odpowiedzi na wzrost zapotrzebowania, zbierane i udostępniane są coraz to nowsze informacje. Otwartą kwestią pozostaje odpowiednie ich wykorzystanie, jako że wiele obszarów życia i działalności gospodarczej pozostaje między sobą w określonej zależności, lecz mimo to nie zawsze wyniki osiągnięte w jednej dziedzinie mogą zostać w prosty sposób przełożone na wyniki osiągnięte w innej.

J.E. Stiglitz i A. Sen podnoszą również – istotny w naszym odczuciu – problem związany ze stosowaniem miar statystycznych zauważając, iż zarówno mierzone zjawisko, jak i sam sposób pomiaru nieuchronnie wpływają na podejmowane decyzje. W przypadku zdefiniowania na nowo wykorzystywanych w polityce gospodarczej statystyk i wskaźników, pożądane kierunki tej polityki i sposoby jej prowadzenia mogą okazać się odmienne od dotychczasowych. Przykładowo, wybór między promowaniem wzrostu gospodarczego (odzwierciedlonego w PKB) i ochroną środowiska może okazać się źle zdefiniowany w sytuacji, w której degradacja środowiska stanie się integralną częścią miernika rozwoju.

Kolejną kwestią, na którą warto zwrócić uwagę, jest występujący często „rozrzut” między standardowymi statystykami opisującymi ekonomię w skali makro i wynikami badań skoncentrowanych na jakości życia obywateli. Zauważmy, że sytuacja taka, chociaż może wywierać negatywny wpływ na zaufanie do oficjalnych raportów i statystyk, nie neguje mierników takich jak PKB jako miar produkcji, ale ukazuje konieczność rozważania ich wraz z innymi lub zmodyfikowania w taki sposób, by odzwierciedlały również inne aspekty rozwoju. Wydaje się więc, że w wykorzystywanych wskaźnikach statystycznych punkt ciężkości powinien zostać przesunięty z pomiaru wydajności oraz produkcji do pomiaru szczęścia i samopoczucia.

W raporcie uwzględniono słabe strony mierników powszechnie używanych w trzech głównych obszarach: PKB, jakości życia oraz zrównoważonego rozwoju i środowiska, sugerując jednocześnie nowe kierunki ich rozwoju i przebudowy.

W pierwszym obszarze komisja zwraca uwagę na kwestie związane z wykorzystaniem konsumpcji oraz dochodu jako kryterium oceny. W związku z międzynarodowym przepływem kapitału oraz różnicami cen sugeruje przeniesienie nacisku z produkcji na konsumpcję, ze szczególnym uwzględnieniem rozkładu dochodu oraz konsumpcji gospodarstw domowych. Jednym z powodów jest fakt, że dane w skali mikro umożliwiają dokładniejsze prześledzenie ewolucji standardów życia obywateli. Ponadto, jak wynika z badań przeprowadzonych w krajach OECD, realny dochód gospodarstw domowych zmienia się inaczej niż PKB *per capita* i pozostaje na niższym poziomie.

Sugerowana zmiana podejścia umożliwiłaby nie tylko odzwierciedlenie we wskaźnikach usług dostarczanych przez rząd, ale także lepszą ocenę poziomu życia obywateli. Proponuje się skorygowanie mierników dochodu ze względu na wynikające z rozwoju społeczno-gospodarczego zmiany funkcjonowania gospodarstw domowych, na skutek których część usług niegdyś zapewnianych przez członków gospodarstwa domowego obecnie nabywana jest na rynku. Konsekwencją tego faktu jest wzrost dochodów mierzonych w tradycyjny sposób, sugerujący zmianę standardu życia, podczas gdy w rzeczywistości następuje jedynie przesunięcie wykonywanych usług z sektora pozarynkowego na rynek. Podobnie, wiele prac wykonywanych w gospodarstwach domowych nie znajduje odzwierciedlenia w oficjalnych miernikach produkcji, chociaż niewątpliwie stanowią one ważny aspekt gospodarowania.

Istotną kwestią pozostaje rozróżnienie w wydatkach konsumpcyjnych środków podchodzących z bieżących dochodów i zgromadzonego majątku. Oczywiście jest, że gospodarstwo domowe przeznaczające na konsumpcję zgromadzone oszczędności powiększa wprawdzie swój dobrobyt, ale kosztem przyszłej sytuacji, a skutki takiego postępowania, jeśli jest ono masowe, znaleźć mogą odzwierciedlenie w stanie całej gospodarki. Zwraca się uwagę na pojawiający się przy tym problem odpowiedniej wyceny zasobów zarówno kapitału rzeczowego, jak i ludzkiego czy społecznego. W odniesieniu do dochodu i bogactwa gospodarstw domowych, komisja podejmuje również zagadnienie analizy czasu wolnego, sugerując konieczność nie tylko porównań międzynarodowych, ale również między regionami jednego kraju, ponieważ różne grupy obywateli mogą w odmienny sposób podejmować decyzje związane z wyborem między czasem wolnym i pracą. Odpowiednie zmierzenie czasu wolnego wymaga trudnego czasem do zdefiniowania rozróżnienia między faktycznym odpoczynkiem a pracami domowymi (kiedy określone działanie jest sposobem spędzenia czasu wolnego, a kiedy produkcją na rzecz gospodarstwa domowego). Ponadto, aktualne metody szacowania wartości czasu wolnego wymagają uwzględnienia kosztu alternatywnego (zazwyczaj stawki płac), przy czym na ogół nie uwzględnia się stosunku do pracy, mimo iż istnieją przesłanki sugerujące różnice w wycenie między poszczególnymi grupami obywateli. Otwartą kwestią pozostaje także wycena czasu wolnego w przypadku jednostek, które gdyby chciały, mogłyby podjąć dodatkową pracę, np. emerytów. Wydaje się, że w tego typu analizach konieczne jest uwzględnienie indywidualnych preferencji.

W obszarze związanym z PKB komisja porusza również kwestię sposobu pomiaru wartości usług, w tym tych dostarczanych przez państwo. W tym przypadku oprócz obrotu sugerowane jest uwzględnienie także ich jakości (np. dostępu do sklepu, jakości obsługi, wybór produktów).

W obszarze związanym z jakością życia obywateli, komisja koncentruje się na potrzebie uwzględnienia obok obiektywnych faktów również subiektywnych odczuć badanych. Podczas gdy panuje zgodność co do tego, iż jakość życia zależy od takich cech jak zdrowie czy wykształcenie, istotna jest także rola jednostki w polityce i społeczeństwie, środowisko oraz czynniki wpływające na poczucie osobistego i ekonomicznego bezpieczeństwa. W tym obszarze konieczne jest również uwzględnienie nierówności wynikających z przynależności do określonych grup społecznych, a także identyfikacja związków pomiędzy jakością różnych wymiarów życia, a następnie agregacja ocen w poszczególnych wymiarach.

Kolejną ważną kwestią, sygnalizowaną już wyżej, pozostaje uwzględnienie w ocenie preferencji obywateli. Podczas gdy zmiany w realnych dochodach i bogactwie mogą skutkować zmianami satysfakcji zachodzącymi w tym samym kierunku, nieznane pozostają zarówno siła tej zmiany, jak i jej zróżnicowanie między konsumentami o odmiennych preferencjach. Tymczasem nie można ignorować faktu, że preferencje wywierają wpływ na postępowanie obywateli i znajdują odzwierciedlenie w dysponowaniu posiadanymi środkami finansowymi. Pozostają one również podstawą subiektywnego wnioskowania o dobrobycie w odpowiedziach na badania ankietowe. Uwzględnianie preferencji w ocenie rozwoju społeczno-ekonomicznego może pozwolić na uniknięcie nadmiernych uproszczeń wynikających z uśredniania, nadal jednak nierozwiązany pozostaje problem agregacji indywidualnych ocen i preferencji w celu oceny stanu całego społeczeństwa.

W przypadku zrównoważonego rozwoju ocena pojedynczego kraju jest szczególnie utrudniona ze względu na zależność od polityki prowadzonej przez wiele państw wartości kryteriów z tej dziedziny (np. jakości środowiska naturalnego). Komisja uznaje potrzebę dążenia do stworzenia jasnych i zrozumiałych mierników odzwierciedlających zagrożenia mogące wystąpić w tym obszarze. Za przykłady takich statystyk można uznać wskaźniki wyczerpywania zasobów i złóż lub obrazujących natężenie zachodzących zmian klimatycznych.

Większość ze zgłaszanych propozycji zastąpienia PKB *per capita* innym miernikiem rozwoju społeczno-gospodarczego sprowadza się do konstruowania miar syntetycznych, których wartości są wyznaczane na podstawie wartości zdefiniowanych zmiennych diagnostycznych. Za każdym razem wyznaczenie wartości zmiennej syntetycznej prowadzi do sporządzenia rankingu krajów. Warto więc przyrzeć się bliżej możliwościom tworzenia rankingów na podstawie wielokryterialnych ocen obiektów.

Ocena wielu obiektów według jednolitego i jednego kryterium stwarza możliwość podania dla zbioru obiektów – w naszym przypadku krajów – ich liniowego uporządkowania, zwanego popularnie rankingiem. Ranking sporządzony na podstawie jednej mierzalnej zmiennej jest łatwy do wykonania (patrz *casus* PKB *per capita*). Z przytoczonej dyskusji wynika, że ocena gospodarki i stanu społeczeństwa za pomocą jednej, chociażby najlepszej miary, nie może być w pełni zadowalająca i obejmować wszystkich dziedzin rozwoju gospodarczego i społecznego. Istnieje zatem potrzeba uwzględnienia wielu kategorii opisujących stan i perspektywy gospodarki.

Wzięcie pod uwagę kilku powszechnie uzgodnionych i zaakceptowanych kryteriów rozwoju ekonomiczno-społecznego, rodzi problemy związane ze sposobem sporządzenia ostatecznego rankingu. Ogólnie są możliwe dwie drogi postępowania. Jedną jest sporządzenie na podstawie wartości wielu zmiennych diagnostycznych jednego miernika syntetycznego i uszeregowanie krajów według wartości tegoż miernika. O takim postępowaniu była mowa w przypadku wszystkich opisywanych do tej pory mierników. Postępowanie można przeprowadzić jednak na dwa sposoby. Pierwszy z nich nazwiemy „naiwnym” i jako przykład podamy⁵ rankingi krajów Unii Europejskiej budowane przez

⁵ Sposób przygotowania wartości wspomnianych wcześniej wskaźników HDI, HPI-1, HPI-2 oraz TAI nie może być uznany za „naiwny” ze względu na metodykę przygotowania przetworzonych wartości diagnostycznych i sposób wyznaczania wartości miary syntetycznej (Ochocki, 2009).

dziennikarzy tygodnika „Polityka”. Drugim sposobem wyznaczenia wartości miernika syntetycznego może być zastosowanie metod wielowymiarowej analizy porównawczej (Panek, 2009).

Druga droga postępowania w przypadku oceny obiektu ze względu na wiele cech jest odmienna od pierwszej i polega na jednoczesnym uwzględnianiu przy ocenie wartości wszystkich branych pod uwagę kryteriów. Metody ilościowe wspomagające decyden-ta w tak wykonywanych rankingach są zaliczane do optymalizacji wielokryterialnej, jed-nej z dziedzin badań operacyjnych (Trzaskalik, 2006).

W trakcie referowania wyników rankingów otrzymanych różnymi metodami wska-zane jest ocenianie zgodności uporządkowania obiektów. W opracowaniu zastosowano dwie miary: najpopularniejszą jest współczynnik korelacji rang Spearmana oraz, dla po-równania, współczynnik tau-Kendalla.

1. Zbiór zmiennych diagnostycznych

W krótkim odstępie czasu tygodnik „Polityka” zaprezentował dwa rankingi krajów Unii Europejskiej. Dziennikarze tygodnika (Kowanda, kwiecień 2009; Smoczyński, paź-dziernik 2009) podjęli trud wykonania rankingów z okazji piątej rocznicy wstąpienia Pol-ski do Unii Europejskiej. Pierwszy z nich wyróżnił 20, a drugi – 25 mierzalnych kryte-riów oceny.

Sposób uzyskania rankingów „Polityki” nazywamy „naiwnym”, dlatego, że nie od-wołuje się on do żadnej znanej ze statystyki, ekonometrii, czy badań operacyjnych meto-dy ilościowej. Ostatnie stwierdzenie w niczym jeszcze nie deprecjonuje wartości poznaw-czej rezultatów obliczeń. Krytyce poddamy jedynie dobór zmiennych diagnostycznych.

Zmienne diagnostyczne charakteryzujące rozwój ekonomiczny i społeczny (Smo-czyński, październik 2009) podzielono na trzy podzbiory opisujące stan: państwa, społeczeństwa i gospodarki.

Łatwo zauważyć, iż większość zmiennych diagnostycznych stanowią wielkości notowane tradycyjnie w statystykach krajowych. Co prawda na liście kryteriów znajdu-ją się także takie kategorie, jak „zadowolenie z życia”, „indeks percepcji korupcji” czy „indeks wolności gospodarczej”, ale ich twórcy nadali im postać liczbową.

W początkowym etapie prac poddaliśmy weryfikacji zbiór zmiennych diagnostycz-nych. Po pierwsze, wartości destymulant⁶ (P2, P6, S3, S7, S8, S9, G2, G3) poddano trans-formacji w celu zamiany ich na stymulanty (Kolenda, 2006: 22). Po drugie, ze zbioru zmiennych diagnostycznych usunięto te zmienne, których wartości cechowały się małą zmiennością mierzoną klasycznym współczynnikiem zmienności⁷. Dotyczyło to zmien-nych P7, S2, S4, S11, G4, G5 oraz G7. Po trzecie, usunięto ze zbioru zmiennych diagno-stycznych zmienne w dużym stopniu skorelowane z innymi zmiennymi. Z tego powodu wyeliminowano zmienne S1, S6, S9 i S10. Badano również, czy istnieją w zbiorze zmien-

⁶ Im niższe wartości destymulanty, tym lepsza pozycja obiektu w rankingu. Im wyższe wartości sty-mulanty, tym lepsza pozycja obiektu w rankingu.

⁷ Zmienne o małej zmienności wartości nie różnicują dobrze zbioru obiektów.

Tabela 1. Lista mierników rankingu „Polityki”

Mierniki opisujące stan państwa	
P1	indeks percepcji korupcji (Transparency International)
P2	współczynnik Gini'ego koncentracji dochodów (Human Development Report)
P3	publiczne wydatki na ochronę zdrowia, % PKB (Human Development Report)
P4	publiczne wydatki na edukację, % PKB (Human Development Report)
P5	gęstość sieci autostrad, km/1000km ² (Eurostat)
P6	więźniowie na 100 tys. osób (Human Development Report)
P7	indeks demokracji („The Economist”)
Mierniki opisujące stan społeczeństwa	
S1	udział w dochodach 20% najbiedniejszych mieszkańców (Human Development Report)
S2	średnia długość życia (Human Development Report)
S3	śmiertelność niemowląt, na 1000 urodzeń (Human Development Report)
S4	indeks edukacji (Human Development Report)
S5	użytkownicy Internetu, na 1000 osób (Human Development Report)
S6	zadowolenie z życia (Happy Planet Index)
S7	rozwoły, na 1000 osób (Human Development Report)
S8	samobójstwa, na 100 tys. osób (Human Development Report)
S9	morderstwa, na 100 tys. osób (Human Development Report)
S10	nakład prasy codziennej, na 1000 osób (UNESCO)
S11	jakość życia w stolicy (Mercer)
Mierniki opisujące stan gospodarki	
G1	PKB, USD per capita, według parytetu siły nabywczej (CIA World Factbook)
G2	inflacja rok do roku (Eurostat)
G3	bezrobocie (Eurostat)
G4	zatrudnienie (Eurostat)
G5	udział sektora usług w PKB (CIA World Factbook)
G6	wydatki na badania i rozwój, % PKB (Human Development Report)
G7	indeks wolności gospodarczej (Heritage Foundation)

Źródło: Smoczyński, 2009, <http://www.polityka.pl/rynek/gospodarka/>

nych diagnostycznych pary zmiennych ujemnie ze sobą skorelowanych. Nie stwierdzono istnienia takich zmiennych. W rezultacie w zbiorze zmiennych diagnostycznych pozostało 14 zmiennych: P1-P6, S3, S5, S7, S8, G1-G3 oraz G6. Ich nazwy w tabeli 1 są zapisane kursywą.

W dalszej części pracy porównamy rankingi wykonywane na pierwotnym zbiorze zmiennych diagnostycznych. Oznaczać je będziemy „1_nazwa”. Rankingi wykonane na zmodyfikowanym zbiorze zmiennych diagnostycznych otrzymały oznaczenie „2_nazwa”.

2. Metody wykonywania rankingów

2.1. Metoda „naiwna”

Metodyka postępowania podczas sporządzania rankingu publikowanego w tygodniku „Polityka” jest nieskomplikowana. Dla każdego z uwzględnionych kryteriów jest wyłaniany kraj – zwycięzca, który uzyskuje 100 punktów. Następnie, dla wszystkich pozostałych krajów obliczono proporcjonalny dystans do zwycięzcy, stosownie do wartości danego miernika dla danego kraju. Punkty uzyskane przez kraj dla wszystkich kryteriów zsumowano i uszeregowano kraje Unii od największej do najmniejszej liczby uzyskanych łącznie punktów. Otrzymujemy w ten sposób porządkowanie liniowe zbioru obiektów. Gdyby istniał kraj idealnie spełniający wszystkie kryteria, to w pierwszym z rankingów uzyskałby 2000 punktów, a w drugim – 2500 punktów.

Dla ilustracji dalej omawianych metod sporządzania rankingów wykorzystamy dane przygotowane przez Smoczyńskiego, a proponowane metody rangowania i otrzymane za ich pomocą rankingi będą porównywane z rankingiem tygodnika „Polityka”.

2.2. Rankingi wielokryterialne w optymalizacji

Problem uszeregowania elementów z n -elementowego zbioru ze względu na wiele kryteriów jest jednym z trzech podstawowych zagadnień wyróżnianych w obszarze badań operacyjnych. Bezpośrednie porównanie ocenianych obiektów jest zadaniem skomplikowanym lub wręcz niemożliwym ze względu na dużą liczbę kryteriów albo konfliktujący charakter niektórych z nich. Rankingi wielokryterialne pozwalają uszeregować rozpatrywane obiekty przy uwzględnieniu każdego z przyjętych kryteriów również w sytuacjach, w których niektóre z kryteriów nie mają charakteru ilościowego lub też różnorodność jednostek, w których są wyrażane jest tak duża, że stworzenie zagregowanego miernika jest rozwiązaniem sztucznym i trudnym w zastosowaniu. Wykorzystane przez nas metody umożliwiają budowę rankingów także wówczas, gdy nie jest jednoznacznie określone, w jaki sposób strata poniesiona na jednym kryterium kompensowana jest zyskiem na innych (Gal, 1999). W przypadku oceny rozwoju społeczno-gospodarczego zaleta ta wydaje się szczególnie istotna. Z powyższych powodów uznajemy za zasadne zastosowanie rankingów wielokryterialnych do uszeregowania krajów.

W pracy wykorzystano trzy metody konstrukcji rankingów. Ze względu na ograniczone ramy niniejszego opracowania nie przedstawiamy pełnego omówienia algorytmów, a jedynie zwięzły opis mający na celu prezentację idei.

Każda z rozpatrywanych tu metod wymaga określenia wag poszczególnych kryteriów, odzwierciedlających ich relatywną istotność. Na potrzeby niniejszego opracowania przyjęliśmy, że wszystkie kryteria są tak samo ważne. Drugą cechą wspólną jest fakt,

że ranking nie powstaje poprzez proste zagregowanie ocen, lecz opiera się na porównywaniu rozpatrywanych alternatyw⁸ parami. Postępowanie takie pozwala na uniknięcie paradoksu Korhonen, polegającego na preferencji rozwiązań niezbilansowanych⁹.

W każdym zagadnieniu wielokryterialnym można napotkać problem niejednoznaczności. W przypadku metod wspomagających podejmowanie decyzji dotyczy ona najczęściej preferencji decydentów, które mogą być jedynie przybliżane poprzez wagi i parametry, natomiast drugim jej źródłem mogą być same dane. Wynika to z faktu, iż pomiary statystyczne mogą cechować się niepewnością i niedokładnością, czasem również nieoznacznością (Trzaskalik, 2006). Metoda rozwiązywania problemów w oparciu o takie dane z konieczności musi więc uwzględniać ich niedoskonałości. W rozważanym przez nas przypadku należy zwrócić uwagę na fakt, że dane pochodzą z różnych krajów i, w związku z tym, mogą nie oddawać w takim samym stopniu rzeczywistego obrazu sytuacji. Jest to drugi z powodów, dla których metody konstrukcji rankingów, których użyjemy opierają się na porównaniu alternatyw parami. Wprowadzone zostają w nich także relacje preferencyjne obejmujące sytuacje zarówno silnej i słabej preferencji, jak i nieporównywalności czy równoważności. Zgrupowanie tych relacji umożliwia wyznaczenie na zbiorze badanych obiektów relacji rozmytych.

Idea metody ELECTRE (Roy, 1978) uwzględnia porównywanie alternatyw parami dla każdego kryterium. Dla każdej pary podmiotów w pierwszym kroku określana jest jedna z czterech podstawowych sytuacji preferencji, a następnie jedna z pięciu sytuacji zgrupowanych, będących sumą podstawowych, jednakże bez możliwości ich rozróżnienia. Szczególną uwagę warto zwrócić przy tym na rozmytą relację nazywaną relacją przewyższenia, odpowiadającą sytuacji, w której istnieją ważne przesłanki, uzasadniające preferencję lub przypuszczenie preferencji jednego z pary porównywanych obiektów. Formalnie, relacja przewyższenia definiowana jest jako suma silnej i słabej preferencji oraz równoważności, bez możliwości ich odróżnienia. Na tej podstawie wyznaczane są dwie relacje całkowitego porządku. Przy ich uwzględnieniu wyznaczony zostaje ranking końcowy, przy czym:

- uznaje się, że wariant a jest lepszy od wariantu b , jeżeli przynajmniej w jednym z porządków a umieszczony jest wyżej niż b , a w drugim na tym samym poziomie,
- uznaje się, że warianty a i b są równoważne, jeżeli w obu porządkach są umieszczone na tym samym poziomie,
- uznaje się, że warianty a i b są nieporównywalne, jeżeli w jednym porządku a umieszczone jest wyżej, natomiast w drugim niżej niż b .

⁸ Termin „alternatywa” jest użyty stosownie do słownictwa występującego w optymalizacji wielokryterialnej. Jego synonimami są w naszym przypadku terminy „wariant”, „podmiot”, „obiekt”.

⁹ Paradoks może występować w przypadku zastosowania do budowy rankingu średniej ważonej. Można go zilustrować za pomocą następującego przykładu: założmy, że budujemy ranking trzech krajów, oceniając je na podstawie dwóch kryteriów: wysokości PKB oraz średniej przewidywanej długości życia. Przypuśćmy, że w skali 0-100 pierwsze państwo zostało ocenione na 0 w kategorii PKB i 100 w kategorii średniej przewidywanej długości życia. Drugie, odpowiednio, 100 i 0, natomiast trzecie – 45 i 45. Jeżeli do agregacji zostałaby zastosowana zwykła średnia ważona, trzecie państwo, charakteryzujące się umiarkowanymi wartościami obu kryteriów, nie mogłoby znaleźć się na pierwszej pozycji w ranking.

Najwyższe pozycje rankingu przypisywane są obiektom najlepszym, czyli takim, dla których nie można znaleźć wariantów lepszych w świetle powyższych zasad. Następnie, kolejno umieszczamy na niższych miejscach alternatywy, które są gorsze jedynie od obiektów sklasyfikowanych wyżej (Trzaskalik, 2006).

Metoda MAPPAC (Matarazzo, 1990) bazuje na porównaniu alternatyw parami dla każdej pary kryteriów, co prowadzi do rozstrzygnięcia, który z dwóch rozważanych obiektów jest lepszy lub też czy są one nierozróżnialne z punktu widzenia danych kryteriów. Na podstawie otrzymanych wyników wprowadza się na zbiorze alternatyw dwie relacje: preferencji (alternatywa *a* została oceniona lepiej niż *b*) oraz obojętności (alternatywy *a* oraz *b* zostały ocenione tak samo). W metodzie tej również dopuszcza się występowanie relacji rozmytych, pozwalających uwzględnić podczas oceny niepewność związaną z pomiarem lub wynikającą z odmiennego charakteru kryteriów. Sposób budowy tych relacji nie jest jednoznacznie określony, może zostać dopasowany do konkretnego problemu decyzyjnego, co świadczy o elastyczności metody. W ostatnim kroku preferencje są agregowane, co prowadzi do stworzenia wynikowego rankingu.

W metodzie AGREPREF (Hradílek, Juřica, Gurekćy, Krejći, 2006) dla każdej pary alternatyw sprawdza się, ze względu na które kryteria są one oceniane jako lepsze. W niniejszej pracy obliczono trzy oddzielne wskaźniki: stopień preferencji dla każdego podmiotu z rozważanej pary i stopień obojętności jako odsetki kryteriów, dla których, odpowiednio, jedna z alternatyw była oceniana lepiej i obie okazywały się równoważne. Na tej podstawie, po uszeregowaniu wariantów według liczby obiektów przez nie zdominowanych, wyznaczany jest ostateczny ranking. W przypadku, gdy oceny obiektów różnią się nieznacznie, można założyć, że różnice te są zaniedbywalne. W takich sytuacjach metoda umożliwia wprowadzenie dodatkowych parametrów, tzw. progów preferencji oraz obojętności. Wskazują one, jakie wartości stopni preferencji i obojętności muszą wystąpić dla wybranych alternatyw, by – odpowiednio – jedna z nich uznana została za lepszą lub obie za tak samo dobre. Aby uwzględnić w naszych rozważaniach wspomniany powyżej problem niejednoznaczności, przyjęliśmy, że oba parametry wynoszą 0,5.

2.3. Rankingi wielokryterialne w wielowymiarowej analizie porównawczej

Inne postępowanie od metod rangowania obiektów scharakteryzowanych w poprzednim punkcie stosuje się w wielowymiarowej analizie porównawczej (Kolenda, 2006: 133–156). W przeprowadzanej analizie wykorzystamy dwie z nich. Przypominamy, że wszystkie zmienne diagnostyczne są stymulantami. Ich wartości poddano standaryzacji tak, aby średnia arytmetyczna standaryzowanych wartości każdej zmiennej diagnostycznej była równa 0, a odchylenie standardowe równe 1.

W metodzie wzorca rozwoju Hellwiga wyznacza się abstrakcyjny obiekt, zwany wzorcem rozwoju. Przyporządkowuje mu się jako wartości zmiennych diagnostycznych maksymalne wartości każdej z cech. W kolejnym etapie oblicza się odległość każdego z rozpatrywanych obiektów (krajów) od wzorca. Otrzymane odległości od wzorca wyznaczają miejsce w rankingu. Im odległość obiektu od wzorca jest mniejsza, tym pozycja tego obiektu w rankingu jest wyższa.

Drugą z zastosowanych metod jest metoda sum standaryzowanych wartości zmiennych diagnostycznych¹⁰. Dla każdego obiektu wyznacza się średnią arytmetyczną z standaryzowanych wartości zmiennych diagnostycznych. Im większa jest wartość średniej, tym pozycja danego obiektu w rankingu jest wyższa. W metodzie tej występuje paradoks Korhonen. Jak widać, stopień skomplikowania wyznaczania rankingu w opisanych metodach wielowymiarowej analizy porównawczej nie różni się zbytnio od stopnia skomplikowania rachunków wykonywanych podczas sporządzania rankingu tygodnika „Polityka”. Zasadnicza merytoryczna różnica tkwi w sposobie przygotowania danych.

Przykładem zastosowania wielowymiarowej analizy porównawczej do otrzymania rankingu poziomu rozwoju krajów Unii Europejskiej jest opracowanie M. Stec z 2008 r. Autorka wyróżnia 23 zmienne charakteryzujące poszczególne kraje. Wszystkie zmienne diagnostyczne są mierzalne, a ich wartości publikowane w oficjalnych statystykach krajowych i Eurostatu. W większości przypadków dane pochodzą z 2005 i 2006 r., ale są też z 2003 r.

Autorka proponuje wykonanie rankingów trzema metodami: wzorca rozwoju Hellwiga, sum standaryzowanych wartości zmiennych oraz rang. Zmienne diagnostyczne spełniają kryteria formalne. Wartości ich klasycznych współczynników zmienności są większe od 0,1, cechy są parami jak najsłabiej skorelowane ze sobą. Dla porównania wyniki rankingów M. Stec zostaną zestawione z wynikami rankingu tygodnika „Polityka” – wykonanymi na potrzeby niniejszej pracy.

3. Wyniki i porównania rankingów

Dysponujemy łącznie 15 rankingami krajów Unii Europejskiej ze względu na stopień ich rozwoju. Wyniki wszystkich rankingów umieszczono w tabeli 2. Przyjmujemy symboliczne oznaczenia rankingów, ułatwiające dokonywanie porównań. Oznaczenia rankingów są następujące:

- PKBpc – ranking według wielkości PKB *per capita*, liczonego w USD, według parytetu siły nabywczej,
- Stec_wzorzec – ranking Hellwiga metodą wzorca rozwoju z pracy (Stec, 2008),
- Stec_suma – ranking metodą sum standaryzowanych wartości z pracy (Stec, 2008),
- 1_polityka – ranking tygodnika „Polityka” (Smoczyński, 2009),
- 1_electre – ranking metodą ELECTRE III przed modyfikacją zbioru zmiennych diagnostycznych,
- 1_mappac – ranking metodą MAPPAC przed modyfikacją zbioru zmiennych diagnostycznych,
- 1_agrepref – ranking metodą AGREPREF przed modyfikacją zbioru zmiennych diagnostycznych,

¹⁰ Dla zapewnienia porównywalności z rankingiem tygodnika „Polityka” konsekwentnie zakładamy, że wszystkie zmienne diagnostyczne są jednakowo ważne. Wobec tego w metodzie sum standaryzowanych wartości przyjmujemy dla każdej zmiennej wagę równą 1. Nic nie stoi na przeszkodzie, aby wartości wag różnicować stosownie do wiedzy o ważności poszczególnych zmiennych diagnostycznych.

Tabela 2. Rankingi krajów Unii Europejskiej ze względu na poziom rozwoju

Kraj	Stec_wzorzec	Stec_suma	1_polityka	1_electre	1_mappac	1_agrepref	1_wzorzec	1_suma	2_polityka	2_electre	2_mappac	2_agrepref	2_wzorzec	2_suma	PKBpc
Austria	7	9	6	4	4	4	4	6	7	3	4	4	5	6	4
Belgia	11	17	8	7	9	12	8	8	8	11	10	13	9	9	6
Bułgaria	26	26	26	25	26	26	26	25	25	24	21	22	24	24	26
Cypr	20	21	15	17	15	14	14	15	15	16	12	14	15	15	16
Czechy	10	5	19	14	18	17	16	17	19	18	15	18	18	18	17
Dania	3	3	1	1	1	1	1	2	1	2	1	1	1	2	3
Estonia	16	16	22	22	22	21	22	22	22	25	19	23	23	23	19
Finlandia	2	2	5	6	5	5	5	4	4	9	3	6	8	5	5
Francja	5	10	12	15	11	10	11	11	11	10	7	11	10	10	11
Grecja	23	25	16	21	19	19	20	19	17	21	17	19	21	19	14
Hiszpania	18	19	17	13	13	13	13	13	18	14	14	17	14	16	13
Holandia	4	6	2	2	3	3	3	3	3	1	2	3	2	3	8
Irlandia	12	11	9	10	10	9	12	10	13	8	9	10	13	12	2
Litwa	19	14	24	23	24	23	24	24	23	23	22	24	25	25	24
Luksemburg	13	4	4	12	7	7	6	5	5	6	5	12	4	4	1
Łotwa	25	24	25	26	25	24	25	26	26	26	23	26	26	27	23
Malta	9	8	14	20	16	16	17	16	14	19	13	16	16	14	18
Niemcy	6	7	7	8	6	6	7	7	9	13	6	7	7	7	9
Polska	24	23	23	24	23	25	23	23	24	22	20	27	22	22	25
Portugalia	21	20	18	18	17	18	18	18	16	15	14	15	17	17	20
Rumunia	27	27	27	27	27	27	27	27	27	27	23	25	27	26	27
Słowacja	22	22	20	16	21	22	21	21	20	17	16	20	20	20	21
Słowenia	14	13	10	3	12	11	10	12	6	4	6	5	6	8	15
Szwecja	1	1	3	5	2	2	2	1	2	5	1	2	3	1	7
W.Brytania	8	12	11	9	8	8	9	9	12	12	8	8	11	11	10
Węgry	15	15	21	19	20	20	19	20	21	20	18	21	19	21	22
Włochy	17	18	13	11	14	15	15	14	10	7	11	9	12	13	12

Źródło: opracowanie własne.

- 1_wzorzec – ranking Hellwiga metodą wzorca rozwoju przed modyfikacją zbioru zmiennych diagnostycznych,
- 1_suma – ranking metodą sum standaryzowanych wartości przed modyfikacją zbioru zmiennych diagnostycznych,
- 2_polityka – ranking metodą tygodnika „Polityka” po modyfikacji zbioru zmiennych diagnostycznych, itd.

Możliwe są następujące strategie porównań rankingów:

- ze względu na miejsce danego kraju w rankingu; tym porównaniom, ze względu na ograniczoną objętość tekstu, poświęcimy najmniej miejsca,
- ze względu na zbiór zmiennych diagnostycznych przy ustalonej metodzie,
- ze względu na metodę wykonania rankingu przy ustalonym zbiorze zmiennych diagnostycznych.

Porównaniom może podlegać pozycja poszczególnych krajów w różnych rankingach, jak i zgodność ogólna rankingów zmierzona odpowiednim współczynnikiem korelacji. Możemy również sprawdzić, który kraj ma pozycję rankingową najbardziej wrażliwą na zmiany metody rangowania, czy zmiany zbioru zmiennych diagnostycznych.

Ze statystycznego punktu widzenia wartości w wybranym rankingu są wartościami cechy mierzalnej na skali porządkowej. Miernikami statystycznymi służącymi do badania współzależności takich zmiennych są m.in. współczynnik Spearmana korelacji rang lub współczynnik tau-Kendalla. Współczynnik Spearmana jest miarą pochodną od współczynnika Pearsona korelacji liniowej. Ta ostatnia miara jest przeznaczona do badania siły i zwrotu zależności cech pozostających w związkach liniowych, bądź zbliżonych do liniowych. Jest wrażliwa na obserwacje odstające i brak normalności rozkładu wartości cechy. Współczynnik Spearmana dziedziczy wady miary Pearsona. Niektórzy statystycy stwierdzają, że nie może być on stosowany do oceny podobieństw uporządkowań (Stanisz, 2006: 337), w zamian proponują stosowanie współczynnika tau-Kendalla. Ponieważ w literaturze do podobnych porównań najczęściej jest stosowany współczynnik Spearmana, a tau-Kendalla rzadko, zamieszczamy w tabelach 3 i 4 zestaw wartości obu miar dla porównywanych rankingów. Wartości obu miar pochodzą z przedziału $\langle -1, 1 \rangle$. Wartość 1 oznacza pełną zgodność dwóch rankingów.

Przytoczone porównania rankingów wskażą Czytelnikowi tylko możliwości w tym zakresie, ponieważ „nie rościmy pretensji” do kompletności spostrzeżeń wynikających z takich porównań. Według nas, najistotniejsze są następujące wnioski:

1. Na podstawie różnych rankingów dotyczących tego samego ogólnego kryterium, ale wykonywanych różnymi metodami lub na podstawie różnych zestawów zmiennych diagnostycznych, otrzymuje się różne sposoby uporządkowania krajów.
2. Poprzez odpowiedni dobór metody rankingu i zbioru zmiennych diagnostycznych istnieje możliwość manipulowania wynikiem rankingu. Warto jednocześnie zauważyć, że zmienność pozycji rankingowej nazywana rozstępem, mierzona różnicą między najgorszym a najlepszym miejscem, dla poszczególnych krajów jest inna i zależy również od zestawu zmiennych diagnostycznych. Dla początkowego zestawu zmiennych (por. tabela 1) dla Rumunii rozstęp wynosi 0, zaś dla Bułgarii, Danii, Estonii, Holandii, Litwy i Portugalii rozstęp wynosi 1. Dla zmo-

dyfikowanego – zgodnie z regułami sztuki – zestawu zmiennych diagnostycznych rozstęp równy 1 odnotowano tylko dla Danii. Łatwo zauważyć, że mały rozstęp dotyczy przeważnie krajów zajmujących najwyższe i najniższe pozycje w dokonanych rankingach.

3. Warto odnotować, bazując na powyższym spostrzeżeniu, że kraje zajmujące pierwsze trzy pozycje powtarzają się w większości rankingów. Dania najniżej została sklasyfikowana na trzeciej pozycji, a więc można uznać, że niezależnie od wykorzystanej metody i zbioru kryteriów, kraj ten oceniany jest jako najlepiej rozwinięty. Może to wskazywać na wysokie oceny w każdej z rozpatrywanych dziedzin. Pozycje pozostałych krajów z czołówki ulegają większym zmianom. Z państw uznawanych za dobrze rozwinięte pewną niespodzianką stanowić może siódma pozycja Szwecji w rankingu zbudowanym na podstawie PKB, co wskazuje na fakt, że wysokie oceny kraj ten zawdzięcza w większym stopniu czynnikom społecznym, środowiskowym i opisującym stan państwa niż produktywności gospodarki. Zauważmy również, że w obu rankingach zbudowanych metodą ELECTRE Szwecja znajduje się poza pierwszą trójką. Zbliżona sytuacja dotyczy Holandii, która oceniona jedynie pod kątem PKB, również nie osiąga miejsca w ścisłej czołówce.
4. Największe zróżnicowanie pozycji rankingowej – tylko dla pierwotnego zbioru danych – osiągnęły Luksemburg (8 pozycji) i Słowenia (9 pozycji).
5. Z kolei dla zmodyfikowanego zestawu danych najbardziej różnią się pozycje Luksemburga (8 miejsc), Niemiec i Polski (po 7 pozycji). Ogółem, dla wszystkich rankingów największe zróżnicowanie pozycji osiągnęły Czechy (14 pozycji), Luksemburg, Malta oraz Słowenia (po 12 pozycji). Oceny tych krajów są najbardziej zależne od wybranych kryteriów oraz metod rankingowych.
6. Dla rankingów wykonanych metodami wielowymiarowej analizy porównawczej największe zmiany dotyczą Czech i Litwy (odpowiednio 13 i 11 pozycji). Dla rankingów wielokryterialnych maksymalne różnice są mniejsze. W ich przypadku najbardziej, bo o 9 miejsc, zmieniła się pozycja Słowenii.
7. Zmodyfikowany zestaw zmiennych diagnostycznych generował mniej zgodne rankingi. Suma rozstępów dla 27 krajów wyniosła 120, a dla rankingów z pierwotnego zestawu zmiennych tylko 88. Dla rankingów skonstruowanych metodami wielowymiarowej analizy porównawczej odpowiednie sumy wynosiły 18 dla pierwotnego zestawu danych oraz 22 dla zmiennych zmodyfikowanych. Natomiast dla rankingów zbudowanych przy użyciu metod wielokryterialnych odpowiednio 65 dla pierwotnego i 100 dla zmodyfikowanego zestawu danych.
8. Na podstawie wartości współczynników tau-Kendalla (tabela 4) zgodność rankingów można ocenić bardziej krytycznie niż na podstawie wartości współczynników Spearmana (tabela 3). Ponadto wartości współczynników tau-Kendalla wyraźniej różnicują zgodność rankingów wykonywanych za pomocą różnych metod.
9. Można dostrzec pewne różnice pomiędzy rankingami zbudowanymi na podstawie zestawu kryteriów zaproponowanych w artykule M. Stec z 2008 r., uszeregowaniem pod względem PKB, a rankingami zbudowanymi w oparciu o dane

Tabela 3. Współczynniki Spearmana korelacji rang dla rankingów krajów Unii Europejskiej

	Stec_wzorzec	Stec_suma	1_polityka	1_electre	1_mappac	1_agrepref	1_wzorzec	1_suma	2_polityka	2_electre	2_mappac	2_agrepref	2_wzorzec	2_suma	PKBpc
Stec_wzorzec	1.000	0,930	0,846	0,810	0,878	0,890	0,879	0,879	0,834	0,744	0,867	0,827	0,832	0,856	0,769
Stec_suma	0,930	1.000	0,786	0,726	0,795	0,822	0,807	0,811	0,775	0,690	0,803	0,733	0,772	0,797	0,718
1_polityka	0,846	0,786	1.000	0,917	0,980	0,971	0,974	0,983	0,981	0,922	0,977	0,936	0,969	0,985	0,933
1_electre	0,810	0,726	0,917	1.000	0,932	0,922	0,943	0,926	0,924	0,943	0,931	0,943	0,942	0,922	0,842
1_mappac	0,878	0,795	0,980	0,932	1.000	0,993	0,993	0,995	0,957	0,918	0,978	0,946	0,971	0,977	0,924
1_agrepref	0,890	0,822	0,971	0,922	0,993	1.000	0,987	0,986	0,946	0,909	0,976	0,949	0,961	0,967	0,920
1_wzorzec	0,879	0,807	0,974	0,943	0,993	0,987	1.000	0,993	0,960	0,921	0,978	0,938	0,980	0,980	0,910
1_suma	0,879	0,811	0,983	0,926	0,995	0,986	0,993	1.000	0,963	0,915	0,978	0,933	0,972	0,984	0,930
2_polityka	0,834	0,775	0,981	0,924	0,957	0,946	0,960	0,963	1.000	0,938	0,979	0,953	0,974	0,985	0,882
2_electre	0,744	0,690	0,922	0,943	0,918	0,909	0,921	0,915	0,938	1.000	0,943	0,939	0,957	0,934	0,853
2_mappac	0,867	0,803	0,977	0,931	0,978	0,976	0,978	0,978	0,979	0,943	1.000	0,968	0,982	0,992	0,885
2_agrepref	0,827	0,733	0,936	0,943	0,946	0,949	0,938	0,933	0,953	0,939	0,968	1.000	0,947	0,949	0,838
2_wzorzec	0,832	0,772	0,969	0,942	0,971	0,961	0,980	0,972	0,974	0,957	0,982	0,947	1.000	0,988	0,876
2_suma	0,856	0,797	0,985	0,922	0,977	0,967	0,980	0,984	0,985	0,934	0,992	0,949	0,988	1.000	0,893
PKBpc	0,769	0,718	0,933	0,842	0,924	0,920	0,910	0,930	0,882	0,853	0,885	0,838	0,876	0,893	1.000

Źródło: opracowanie własne.

Tabela 4. Współczynniki tau-Kendalla korelacji rang dla rankingów krajów Unii Europejskiej

	Stec_wzorzec	Stec_suma	1_polityka	1_electre	1_mappac	1_agrepref	1_wzorzec	1_suma	2_polityka	2_electre	2_mappac	2_agrepref	2_wzorzec	2_suma	PKBpc
Stec_wzorzec	1.000	0,812	0,681	0,641	0,732	0,732	0,715	0,726	0,664	0,561	0,713	0,641	0,670	0,692	0,578
Stec_suma	0,812	1.000	0,618	0,544	0,635	0,658	0,641	0,652	0,601	0,533	0,650	0,533	0,607	0,630	0,550
1_polityka	0,681	0,618	1.000	0,778	0,903	0,858	0,886	0,909	0,903	0,766	0,891	0,801	0,875	0,920	0,783
1_electre	0,641	0,544	0,778	1.000	0,806	0,795	0,823	0,801	0,806	0,806	0,799	0,818	0,823	0,778	0,675
1_mappac	0,732	0,635	0,903	0,806	1.000	0,954	0,949	0,960	0,852	0,761	0,897	0,829	0,880	0,880	0,778
1_agrepref	0,732	0,658	0,858	0,795	0,954	1.000	0,926	0,926	0,829	0,749	0,885	0,840	0,846	0,846	0,789
1_wzorzec	0,715	0,641	0,886	0,823	0,949	0,926	1.000	0,943	0,858	0,778	0,891	0,801	0,897	0,886	0,772
1_suma	0,726	0,652	0,909	0,801	0,960	0,926	0,943	1.000	0,869	0,755	0,897	0,801	0,886	0,909	0,783
2_polityka	0,664	0,601	0,903	0,806	0,852	0,829	0,858	0,869	1.000	0,783	0,908	0,840	0,880	0,915	0,721
2_electre	0,561	0,533	0,766	0,806	0,761	0,749	0,778	0,755	0,783	1.000	0,799	0,783	0,835	0,778	0,675
2_mappac	0,713	0,650	0,891	0,799	0,897	0,885	0,891	0,897	0,908	0,799	1.000	0,880	0,908	0,948	0,719
2_agrepref	0,641	0,533	0,801	0,818	0,829	0,840	0,801	0,801	0,840	0,783	0,880	1.000	0,823	0,835	0,664
2_wzorzec	0,670	0,607	0,875	0,823	0,880	0,846	0,897	0,886	0,880	0,835	0,908	0,823	1.000	0,920	0,715
2_suma	0,692	0,630	0,920	0,778	0,880	0,846	0,886	0,909	0,915	0,778	0,948	0,835	0,920	1.000	0,726
PKBpc	0,578	0,550	0,783	0,675	0,778	0,789	0,772	0,783	0,721	0,675	0,719	0,664	0,715	0,726	1.000

Źródło: opracowanie własne.

pochodzące z tygodnika „Polityka”. Dla tych ostatnich, zarówno w przypadku pierwotnego, jak i zmodyfikowanego zestawu zmiennych, rankingi są do siebie podobne zarówno pod kątem współczynnika Spearmana, jak i współczynnika tau-Kendalla.

10. Z rankingiem, w którym jedyną zmienną diagnostyczną jest wartość PKB *per capita*, najbardziej zgodne są rankingi wykonywane metodą AGREPREF i sumy standaryzowanych wartości (dla pierwotnego zestawu zmiennych), a także metodą tygodnika „Polityka”, co dla autorów publikacji z „Polityki” nastawionych krytycznie do PKB jako powszechnie stosowanej miary rozwoju musi być dużym zaskoczeniem.
11. Metody ELECTRE, MAPPAC czy AGREPREF dają rankingi mniej ze sobą zgodne niż rankingi wykonywane metodami wielowymiarowej analizy porównawczej, co potwierdza obserwacje poczynione przy analizie rozstępów dla poszczególnych uszeregowień.
12. Miejsce Polski w przeprowadzonych rankingach znajduje się około 22–23 pozycji, oprócz rankingów 2_mappac i 2_agrepref, w których wynik dla Polski jest wyraźnie różny od pozostałych.

Zakończenie

Rankingi interesują nas dlatego, ponieważ dzięki liniowemu uporządkowaniu udzielają jednoznacznej odpowiedzi na pytanie o miejsce obiektu (u nas kraju) wśród innych obiektów. Z tego samego powodu najbardziej czytelny jest ranking wykonywany na podstawie wartości jednej zmiennej diagnostycznej. Kilkadziesiąt lat temu do porównań międzynarodowych zaczęto stosować wartość PKB *per capita*, liczoną według parytetu siły nabywczej. Miara ta dobrze odwzorowuje zróżnicowanie wzrostu gospodarek poszczególnych krajów. Nazywanie jej miarą dobrobytu, czy miarą rozwoju, jest jednak znacznym uproszczeniem. Pomiar dobrobytu zakłada pomiar konsumpcji dokonywanej przez gospodarstwa domowe. Powyższe wymaga przyjęcia co najmniej trzech założeń. Po pierwsze, poziom PKB *per capita* stanowi odpowiednie odzwierciedlenie konsumpcji przeciętnego gospodarstwa domowego, niezależnie od zróżnicowania koszyków towarów i usług, jak i wewnętrznego zróżnicowania sytuacji materialnej gospodarstw domowych. Po drugie, wielkość konsumpcji wystarcza do oceny dobrobytu z pominięciem wszystkich innych aspektów życia prywatnego (czas wolny, oszczędności) oraz życia społecznego (wydatki krajowe na edukację, bezpieczeństwo, zbrojenia). Po trzecie i chyba najważniejsze, związek między wartością PKB *per capita* a użytecznością czerpaną z uzyskanej konsumpcji jest jednakowy we wszystkich porównywanych krajach (Dievert, 1999). Ekonomisci radzą sobie z tymi problemami przysposabiając mikroekonomiczną Samuelsona teorię ujawnionych preferencji do zastosowań makroekonomicznych oraz stosując wielostronne indeksy ilościowe skonstruowane przez Afriata (Afriat, 1967).

Inna droga w porównaniach poziomu dobrobytu prowadzi do zgromadzenia informacji dotyczących wielu zmiennych opisujących różne aspekty poziomu życia. W tym kierunku idą propozycje Komisji ds. Mierzenia Wydajności Ekonomicznej i Postępu Spo-

łecznego, pracującej pod przewodnictwem J.E. Stiglitz i A. Sena. Złudzeniem jest jednak wiara w zbudowanie wielowymiarowego miernika dobrobytu, przydatnego do porównań międzynarodowych. Po pierwsze, trudno bowiem uzgodnić jednolity zbiór zmiennych diagnostycznych. Dla różnych krajów i społeczeństw ich rozwój i dobrobyt powinien być mierzony w odmienny sposób, w zależności od doświadczeń historycznych i społecznych. Wpływ na pomiar ma także aktualnie najpopularniejsza w danym kraju „szkoła ekonomii” (Stacewicz, 1991). Po drugie, gdyby takie jednolite zbiory zmiennych diagnostycznych istniały, to należałoby uzgodnić metodykę wykonywania rankingu. Pokazaliśmy, że w zależności od zbioru zmiennych i metody rangowania wyniki rankingu mogą się różnić.

Przyjęcie jako zasady wyznaczania z wielu zmiennych jednej wartości syntetycznej i porządkowania obiektów według jej wartości jest rozwiązaniem nie do przyjęcia. Każda ze zmiennych składowych konstituuje wartość zmiennej syntetycznej w stopniu zgodnym z nadaną jej wagą. Powstaje problem wyznaczania wag i odpowiedzi na pytanie, czy dla każdego kraju wagi mają mieć takie same wartości, czy wartości te w zależności od kraju mogą się różnić. Przyjęcie wszystkich wag jednakowych nie jest również wyjściem z sytuacji, ponieważ deklaruje się wtedy jednakową substytucję między wybranymi zmiennymi diagnostycznymi, czyli substytucję między odwzorowywanymi przez zmienne diagnostyczne celami gospodarowania.

Z zebranych doświadczeń wynika, że metodami przeprowadzania rankingów pozbawionymi wad posługiwania się zmienną syntetyczną są metody wywodzące się z optymalizacji wielokryterialnej. Uzasadniliśmy wcześniej ich zalety. Są to metody stosunkowo mało spopularyzowane i często pracochłonne technicznie. Bez wsparcia odpowiednich programów komputerowych metody rankingów wielokryterialnych nie są możliwe do zastosowania. Rosnąca popularność wszelkiego rodzaju porównań skłania do zastosowania takich narzędzi rangowania, które mają ugruntowane podstawy teoretyczne, dobrze odwzorowują rzeczywistość i pozwalają na uwzględnienie w procesie wykonywania rankingu subiektywizmu.

Pamiętając, że „od tego, co mierzymy, zależy to, co robimy”¹¹ należy do badań będących tematem niniejszej pracy przykładać coraz większą wagę.

Bibliografia

- Afriat S., *The Construction of a Utility Function from Expenditure Data*, “International Economic Review”, nr 1, 1967 (v.8).
- Bremond J., Salort M., *Leksykon wybitnych ekonomistów*, Wydawnictwo Naukowe PWN, Warszawa 1997.
- Diewert W.E., *Axiomatic and Economic Approaches to International Comparisons*, [w:] *International and Interarea Comparisons of Income, Output and Prices*, Heston A. (ed.), University of Chicago Press 1999.
- Gal T., Stewart T. J., Hanne T. (ed.), *Multicriteria decision making: advances in MCDM models, algorithms, theory and applications*, Kluwer Academic Publishers, Boston 1999.

¹¹ Przytoczone powiedzenie pochodzi z raportu Komisji ds. Mierzenia Wydajności Ekonomicznej i Postępu Społecznego.

- Hradílek Z., Juřica L., Gurecký J., Krejčí P., *New method Agregref for the priority location of remote controlled disconnectors in the distribution network*, Proceedings of PEA, 506(7)-506(14), 2006.
- Kolenda M., *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2006.
- Kołodko G.W., *Wędrujący świat*, Prószyński i S-ka, Warszawa 2009.
- Kowanda C., *Ranking europejski: „Polska w Unii. Pierwsze 5 lat”*, dodatek do tygodnika „Polityka”, 25.04.2009.
- Matarazzo B., *A pairwise criterion comparison approach: The MAPPAC and PRAGMA methods*, [w:] Bana e Costa C. (ed.), *Readings in Multiple Criteria Decision Aid*, Springer Verlag, Berlin 1990.
- Ochocki A., *Ludność, edukacja, poziom życia – problemy globalne*, „Wiadomości Statystyczne”, nr 8, 2009.
- Roy B., *ELECTRE III: Un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples*, Cahiers du Centre d’Etudes en Recherche Opérationnelle, 20(1):3–24, 1978
- Smoczyński W., *Od ogona do głowy*, „Polityka”, 13.10.2009.
- Stacewicz J., *Stereotypy rozwoju a ekonomia*, PWE, Warszawa 1991.
- Stanisz A., *Przystępny kurs statystyki*, t.1, StatSoft, Kraków 2006.
- Stec M., *Ranking poziomu rozwoju krajów Unii Europejskiej*, „Gospodarka Narodowa”, nr 7–8, 2008.
- Stiglitz J.E., Sen A., Fitoussi J-P., *Report by the Commission on the Measurement of Economic Performance and Social Progress*, opublikowany 14.09.2009, <http://www.stiglitz-sen-fitoussi.fr/en/documents.htm>, grudzień 2009.
- Trzaskalik T. (red.), *Metody wielokryterialne na polskim rynku finansowym*, PWE, Warszawa 2006.

The Useful Art of the Social-Economic Rankings

Summary

The purpose of the paper is to identify advantages and disadvantages of various methods of constructing rankings. The subject of our study is important due to the international debate on development and welfare measurement methods and ways of comparing results obtained for different countries. Because GDP per capita does not allow sufficient assessment, countries are compared on the basis of many criteria and results are usually presented in form of rankings.

We discuss different outranking methods originating from multidimensional statistical analysis and multicriteria optimization and compare them taking into consideration the effect of each method and each set of criteria on the final result. Our remarks are illustrated by rankings of development and economic performance built for European Union countries.

Our observations and results can be regarded also as an opinion in the discussion on the report of the International Commission on Measurement of Economic Performance and Social Progress chaired by J.E. Stiglitz and A. Sen.