

## **PORÓWNANIE MODELU REGRESJI WIELORAKIEJ ORAZ DRZEWA REGRESYJNEGO NA PRZYKŁADZIE INDEKSU KORUPCJI**

**Michał Gostkowski**

Katedra Ekonometrii i Statystyki  
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie  
e-mail: [michal\\_gostkowski@sggw.pl](mailto:michal_gostkowski@sggw.pl)

**Krzysztof Gajowniczek, Piotr Jałowiecki**

Katedra Informatyki  
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie  
e-mail: [krzysztof\\_gajowniczek@sggw.pl](mailto:krzysztof_gajowniczek@sggw.pl), [piotr\\_jalowiecki@sggw.pl](mailto:piotr_jalowiecki@sggw.pl)

**Streszczenie:** W pracy przedstawiono wyniki badań nad modelowaniem tzw. indeksu korupcji (ang. Failed States Index). Zbudowano i porównano model regresji wielorakiej z drzewem regresyjnym. Badania zostały oparte na podstawie danych publikowanych przez niezależną organizację The Fund for Peace. Jako potencjalne zmienne zostały wybrane zmienne udostępnione na stronie internetowej Banku Światowego. Wstępne wyniki jednoznacznie wskazują, że drzewo regresyjne lepiej odzwierciedla zmienność parametru niż model regresji wielorakiej.

**Słowa kluczowe:** Failed States Index, indeks korupcji, model regresji wielorakiej, drzewo regresyjne

### **WPROWADZENIE**

W ostatnich latach w większości państw świata, również w Polsce bardzo często podnoszona jest tematyka korupcji, która jako zjawisko przynosi wymierne straty we wszystkich gospodarkach narodowych. Jednocześnie jest ona zjawiskiem towarzyszącym ludzkości praktycznie „od zawsze”, przykładowo Bardhan [1997] stwierdza, że problem korupcji był już rozpowszechniony w starożytności. Dodatkowo sam termin korupcja nie jest jednoznaczny. W zależności od sytuacji, kontekstu i potrzeby jest on używany w różnych znaczeniach, a często po prostu inaczej rozumiany. To znacząco utrudnia jego precyzyjne zdefiniowanie.

Korupcję można określić jako nadużywanie sektora publicznego w celu osiągnięcia korzyści prywatnych [Lambsdorff 2007]. Termin korzyści prywatne odnosi się głównie do pieniędzy lub wartościowych aktywów, ale może także obejmować zwiększenie władzy lub statusu. Otrzymywanie obietnic przyszłych przysług lub korzyści dla członków rodziny, przyjaciół i znajomych również może być uznane za korzyści prywatne. Korupcja występuje szczególnie w sytuacjach interakcji sektorów prywatnego i publicznego. Występuje zatem przede wszystkim tam gdzie urzędnicy państwowi mają bezpośrednią odpowiedzialność za świadczone usługi publiczne [Rose-Ackerman 1997].

Tanzi [1988] stwierdza, że istnieją pośrednie i bezpośrednie czynniki stymulujące korupcję. Do czynników bezpośrednich zalicza regulacje i zezwolenia, podatki, decyzje dotyczące wydatków, dostarczenie dóbr i usług poniżej ceny rynkowej oraz finansowanie partii politycznych. Natomiast do czynników pośrednich zalicza jakość biurokracji, poziom płac w sektorze publicznym, systemy karne, kontrolę instytucjonalną, przejrzystość zasad, praw i procesów. Korupcja jest symptomem słabości instytucjonalnej i wpływa na słabe wyniki gospodarcze, społeczne i polityczne [Akçay 2006].

Dodatkowo problem pomiaru korupcji powoduje, że jest stosunkowo niewiele dowodów, a także porozumień, jak skutecznie można ograniczyć korupcję [Olken 2005]. Niektóre rozwiązania zakładają, że właściwa kombinacja monitoringu urzędników oraz systemu kar pomaga kontrolować korupcję. W praktyce okazuje się zwykle, że te same jednostki odpowiedzialne za monitorowanie i egzekwowanie kar są jeszcze bardziej narażone na korupcję. W tym przypadku kontrola wyższych urzędników nad niższymi urzędnikami zwiększa prawdopodobieństwo przenoszenia korupcji pomiędzy urzędnikami, a nie zmniejszeniem korupcji. Alternatywne podejście, które zyskuje na znaczeniu w ostatnich latach, polega na zwiększaniu świadomości społeczeństwa na problem korupcji oraz udziału zwykłych obywateli w przeciwdziałaniu korupcji na poziomie lokalnym. W większości współczesnych społeczeństw, szczególnie tych wysoko rozwiniętych podejście takie jest uważane za klucz nie tylko do zmniejszenia korupcji, ale także do poprawy jakości świadczenia usług publicznych [Olken 2005]. Bank Światowy [2004] stwierdza wprost, że „wprowadzenie ludzi ubogich w centrum świadczonych usług, umożliwiając im monitorowanie i dyscyplinowanie usługodawców, wzmacnianiu ich głosu w kształtowaniu polityki, i wzmacnianie zachęt doprowadzi do służby usługodawców biednym”. Oczywiście takie podejście posiada również swoje wady np. może być podatne na działania lokalnych elit.

## CEL, ZAKRES I METODYKA BADAŃ

Celem badań było porównanie parametrycznych i nieparametrycznych metod przy modelowaniu zmienności indeksu korupcji FSI (ang. Failed States Index, od 2014 roku Fragile States Index). Jest on wyznaczany przez niezależną organizację o nazwie The Fund For Peace (FFP). Głównym celem organizacji jest zapobieganie

konfliktom z użyciem przemocy oraz promowanie trwałego bezpieczeństwa na świecie. W swoich badaniach FFP tworzy praktyczne narzędzia i metody, wykorzystywane w praktyce przez rządy państw, organizacje międzynarodowe i wiele innych instytucji w celu przeciwdziałania korupcji. Współpracuje ona z partnerami z ponad 50 krajów specjalizującymi się w bardzo różnych dziedzinach, np. administracji rządowej, wojska, świata nauki. Jednym z ważniejszych opracowanych przez FFP mierników jest FSI, który powstał w 2005 roku. Jest to miernik wyznaczany na skali ciągłej, a jego wyższe wartości oznaczają wyższe prawdopodobieństwo korupcji w danym kraju.

Wśród metod parametrycznych został wybrany model regresji wielorakiej, który jest jednym z najbardziej popularnych i najczęściej stosowanych narzędzi ekonometrycznych [Maddala 2006]. Model regresji wielorakiej został porównany z modelem drzewa regresyjnego, który został odkryty w latach siedemdziesiątych i był pierwszym algorytmem uczenia maszynowego. Zaliczany jest on do grupy metod określanых mianem CART (*Classification and regression tree*). Algorytmy CART cieszą się dużą popularnością i zostały uznane jako efektywne narzędzie do prognozowania i klasyfikacji [Lee i in. 2006]. Dane źródłowe, na których oparto badania zostały pobrane ze strony internetowej Banku Światowego oraz ze strony fundacji FFP. Dane dotyczące indeksu korupcji zostały pobrane dla 2011 roku i dotyczyły 87 państw z różnych części świata. Jako predyktory wybrano 31 zmiennych. Potencjalne zmienne zostały wybrane w taki sposób aby informacje, które przedstawiają były dostępne dla wszystkich wybranych państw. Pełna lista wszystkich wybranych zmiennych znajduje się w załączniku.

W modelu regresji wielorakiej zakłada się, że występuje liniowa zależność pomiędzy zmiennymi a zmienną zależną co można zapisać za pomocą następującego równania:

$$Y = \mathbf{X}\mathbf{B} + \varepsilon, \quad (1)$$

gdzie:

$\mathbf{X}$  – wektor zmiennych niezależnych,

$Y$  – zmienna zależna,

$\varepsilon$  - składnik losowy,

$\mathbf{B}$  – wektor parametrów.

W Klasycznej Metodzie Najmniejszych Kwadratów (KMNK), wektor  $\mathbf{B}$  uzyskuje się na podstawie następującego wzoru:

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2)$$

W KMNK dodatkowo zakłada się, że macierz zmiennych niezależnych jest algebraicznie niezależna, natomiast składnik losowy ma rozkład  $N(0, \sigma^2)$  [Faraway 2004].

Drzewo regresyjne jest to skierowany graf acykliczny i spójny, mający jeden dający się wyróżnić wierzchołek, nazywany korzeniem [Koronacki i Ćwik 2008]. Graf skierowany składa się z wierzchołków i łączących je krawędzi skierowanych, czyli krawędzi prowadzących od jednego wierzchołka do drugiego przy czym może

to się odbywać tylko w jednym kierunku. W przypadku drzew regresyjnych wierzchołki określa się najczęściej węzłami, natomiast krawędzie gałęziami. Jeżeli z węzła wychodzą krawędzie, wówczas dany węzeł określa się jako rodzic. Węzeł, z którego nie wychodzą krawędzie, określa się jako liść.

Model drzewa regresyjnego zawiera zbiór reguł, które mają na celu podział niejednorodnej grupy obiektów (znajdującej się w korzeniu) na mniejsze, bardziej jednorodne grupy (liście) w stosunku do pewnej zmiennej zależnej [Berry i Linoff 2004]. W tym celu należy wyznaczyć warunki postaci  $\{X_i \leq c\}$ . Jeśli warunek jest spełniony dla konkretnej obserwacji jest ona przekierowana do lewego węzła. W przeciwnym przypadku do prawego węzła.

Dla zbudowanego drzewa estymator jest wyznaczany w następujący sposób:

1. dla każdej obserwacji znajdź końcowy liść,
2. wartością teoretyczną jest empiryczna średnia ze wszystkich  $y$  należących do danego liścia.

Podział węzła  $m$  na węzły  $m_L$  i  $m_R$  jest dokonywany w sposób zapewniający najlepsze dopasowanie funkcji regresyjnej. Tym samym predyktor  $X_i$  i punkt  $c$  jest wybierany tak aby wartość  $SSE(m_L) + SSE(m_R)$  była minimalna, gdzie  $SSE$  jest sumą kwadratów reszt [Farawy 2006]. Należy tutaj zaznaczyć, że wartość  $SSE$  może obniżyć się tylko w przypadku dodania kolejnego węzła. Kierując się tylko i wyłącznie tą zasadą optymalnym rozwiązaniem jest sytuacja kiedy każdy liść zawiera tylko jeden element wówczas  $SSE=0$ . Powyższa sytuacja opisana jest w literaturze jako *overtraining* co można tłumaczyć jako przeuczenie. Oznacza sytuację kiedy estymator jest idealnie dopasowany do danych ale jednocześnie jest bardzo skomplikowany i niestabilny. W celu uniknięcia powyższego zjawiska w drzewach regresyjnych stosuje się przycinanie. Wówczas minimalizuje się następującą funkcję:

$$R(T) + a/|T| \quad (3)$$

gdzie:  $R(T) = \sum_{i=1}^T SSE_T$ ,  $|T|$  - liczba liści, natomiast  $a > 0$  jest parametrem kontrolującym przycinanie. Zdecydowaną zaletą drzew regresyjnych jest łatwość interpretacji oraz mała liczba założeń, co w przypadku regresji wielorakiej niekoniecznie jest prawdą.

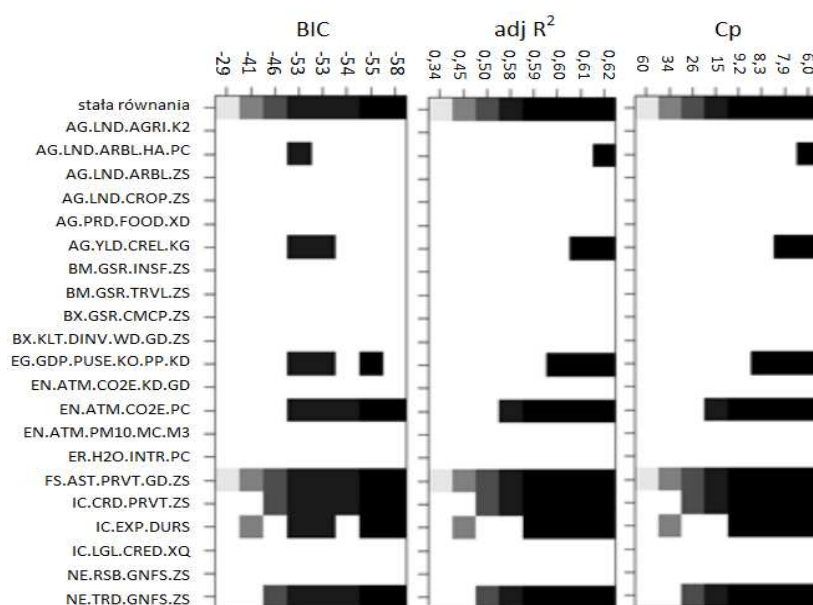
## WYNIKI BADAŃ

W pierwszym etapie badań wyznaczono macierz współczynników korelacji rang Spearmana dla wszystkich par zmiennych objaśniających w celu wyznaczenia zmiennych wysoko skorelowanych. Współczynnik korelacji rang Spearmana jest dużo bardziej odporny na obserwacje odstające niż współczynnik korelacji Pearsona [Croux i Dehon 2010]. Na podstawie macierzy współczynników korelacji z analizy usunięto 10 zmiennych.

W kolejnym etapie badań wykorzystano kryterium informacyjne Bayesa-Schwartz (BIC), kryterium informacyjne Mallowa (MC) oraz wartości

skorygowanego współczynnika  $R^2$  w celu wyboru najlepszego zestawu zmiennych w modelu regresji wielorakiej (Rysunek 2). Skorygowany współczynnik  $R^2$  jest podstawową miarą dopasowania modelu [Borkowski i in. 2003]. Jednak Greene [2003] zaznacza, że skorygowany współczynnik  $R^2$  może dostarczać błędnych wyników w sytuacji gdy liczebność próby wzrasta, dlatego do oceny modeli i wyboru najlepszego proponuje zastosowanie kryterium informacyjnego Akaike (AIC) lub Bayesa-Schwartza (BIC) jednocześnie zaznaczając, że kryterium informacyjne BIC „preferuje” modele z mniejszą liczbą zmiennych. Dodatkowo uwzględniono kryterium informacyjne Mallowa (MC) w celu porównania wyników. W wyniku tych działań zbiór zmiennych objaśniających zredukowano do 8.

Rysunek 2. Graficzna prezentacja wybranych zmiennych do modelu regresji wielorakiej według poszczególnych kryteriów



Źródło: opracowanie własne

Wartości liczbowe oznaczają poziom danego kryterium informacyjnego lub współczynnika  $R^2$  gdy w modelu uwzględnione są zaznaczone zmienne. BIC – kryterium informacyjne Bayesa-Schwartza, adj R<sup>2</sup> – skorygowany współczynnik R<sup>2</sup>, MC – kryterium informacyjne Mallowa.

W kolejnym kroku zbudowano model regresji wielorakiej i usunięto zmienne nieistotne statystycznie. Ostatecznie otrzymano model zawierający 6 zmiennych (Tabela 1). Wszystkie zmienne uwzględnione w modelu są statystycznie istotne. Przeprowadzenie odpowiednich testów potwierdziło wymagane własności reszt modelu w tym: losowość (test serii, p-value=0,1514), normalność (test Shapiro-Wilka, p-value=0,1462), homoskedastyczność (test Breuscha-Pagana, p-value

=0,8839). Przeprowadzono także odpowiednie testy w celu sprawdzenia czy pomiędzy zmiennymi nie występuje współliniowość. W tym celu wykorzystano czynnik inflacji wariancji VIF (wartości czynnika VIF dla wszystkich zmiennych były mniejsze od 2).

Ocena przy zmiennej Produkt Krajowy Brutto w przeliczeniu na jednostkę zużytej energii jest ujemna co jest całkowicie uzasadnione. Produkt krajowy brutto określa wartość wszystkich dóbr i usług wytworzonych w gospodarce w danym okresie [Begg i in. 2003] i może być utożsamiony jako miara bogactwa danego państwa. Niższe wartości Produktu Krajowego Brutto dotyczą państw słabo rozwiniętych, gdzie występuje zjawisko korupcji, co zostało odzwierciedlone w modelu poprzez ujemną ocenę parametru. Zależność pomiędzy korupcją a produktem krajowym brutto została także potwierdzona w innych pracach [Mauro 1995], [Abed i Davoodi 2000]. Gospodarki słabo rozwinięte charakteryzują się oprócz występowania zjawiska korupcji dużym udziałem sektora pierwszego, który obejmuje rolnictwo, leśnictwo i rybołówstwo w strukturze gospodarki narodowej. Powyższa struktura gospodarki charakteryzuje się niskim poziomem emisji CO<sup>2</sup>, co także zostało uwzględnione w modelu poprzez ujemną ocenę parametru przy zmiennej EN.ATM.CO2E.PC. Dodatnią ocenę parametru posiada natomiast zmienna IC.CRD.PRVT.ZS i dotyczy czasu realizacji eksportu liczonego w dniach. Długi czas realizacji eksportu może oznaczać niejasność procedur związanych z eksportem i skłania firmę do wykorzystywania korupcji w celu przyspieszenia procesu. Ujemna ocena parametru przy zmiennej dotyczącej wymiany handlowej (NE.TRD.GNFS.ZS) została również potwierdzona w innych badaniach [Ades i Tella 1999]. Ostatnie dwie zmienne związane są z funkcjonowaniem systemu bankowego: kredyty sektora prywatnego (FS.AST.PRVT.GD.ZS) oraz procent ludności w biurach informacji kredytowej (IC.CRD.PRVT.ZS). Jednak w literaturze brak jest potwierdzenia powyższej zależności, co zmusza do przeprowadzenia dalszej analizy w celu potwierdzenia.

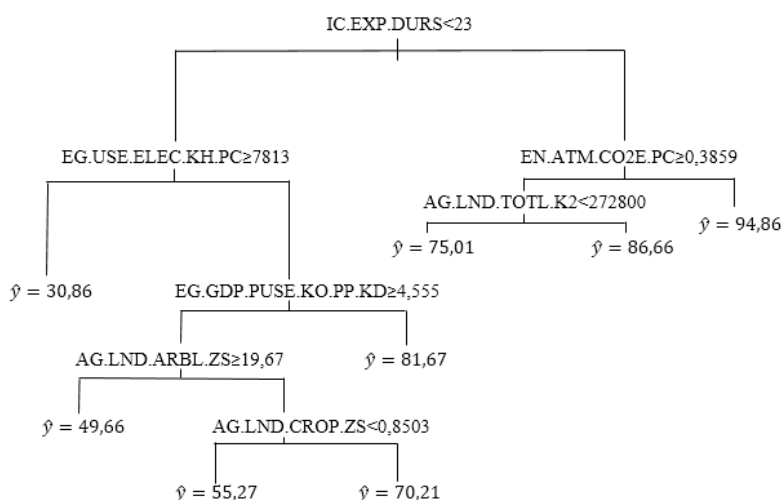
Tabela 1. Oceny parametrów wraz z wyliczonymi odpowiednimi statystykami w modelu regresji wielorakiej dla Failed States Index

	Ocena parametru	Błąd standardowy	Wartość statystyki T-Studenta	p-value
Stała równania	95,57818	5,9436	16,081	<0,00001
EG.GDP.PUSE.KO.PP.KD	-0,82128	0,32884	-2,498	0,01467
EN.ATM.CO2E.PC	-1,34555	0,30478	-4,415	3,30E-05
FS.AST.PRVT.GD.ZS	-0,08211	0,0278	-2,953	0,00418
IC.CRD.PRVT.ZS	-0,23641	0,0451	-5,242	1,38E-06
IC.EXP.DURS	0,17638	0,08483	2,079	0,04096
NE.TRD.GNFS.ZS	-0,0623	0,02763	-2,255	0,02701

Źródło: opracowanie własne

Na podstawie tego samego zestawu danych zbudowano drzewo regresyjne składające się z 8 końcowych węzłów (liści) (Rysunek 3). Jako kluczowe zmienne zostały wybrane: czas realizacji eksportu (IC.EXP.DURS), poziom emisji CO2 (EN.ATM.CO2E.PC), zużycie energii elektrycznej (EG.USE.ELEC.KH.PC), Produkt Krajowy Brutto w przeliczeniu na jednostkę zużytej energii (EG.GDP.PUSE.KO.PP.KD), powierzchnia danego kraju (AG.LND.TOTL.K2), udział trwałych użytków rolnych (AG.LND.CROP.ZS) oraz udział gruntów ornych (AG.LND.ARBL.ZS). W pierwszym liściu (dla którego  $\hat{y}=30,86$ ) znalazła się Islandia, Katar, USA, Kanada, Austria, Norwegia, Korea Południowa, Japonia, Finlandia, Luksemburg, Belgia. W drugim liściu (dla którego  $\hat{y}=49,66$ ) znalazła się

Rysunek 3. Drzewo regresyjne dla zmiennej Failed States Index



Źródło: opracowanie własne

Słowacja, Polska, Włochy, Grecja, Turcja, Węgry, Malta, Barbados. W trzecim liściu (dla którego  $\hat{y}=55,27$ ) znalazła się Argentyna, Szwajcaria, Chile, Peru, Estonia, Bahamy, Senegal, Gabon. W czwartym liściu (dla którego  $\hat{y}=70,21$ ) znalazły się następujące państwa: Izrael, Meksyk, Brazylia, Dominikana, Panama, Ekwador, Cypr, Malediwy, Granada, Antigua i Barbuda, Tunezja, Belize, Seszele, Jamajka. W piątym liściu (dla którego  $\hat{y}=81,67$ ) znalazła się Kuba, Arabia Saudyjska, Korea Północna, Chiny, Białoruś, Liberia, Gujana. W szóstym liściu (dla którego  $\hat{y}=75,01$ ) znalazła się Nikaragua, Fidzi, Armenia, Ghana, Jordan, Macedonia, Brunei, Sao Tome, Albania, Dżibuti, Surinam, Benin, Syria, Azerbejdżan, Samoa. W siódmym liściu (dla którego  $\hat{y}=86,66$ ) znalazła się Kolumbia, Boliwia, Rosja, Indie, Egipt, Ukraina, Iran, Uzbekistan, Pakistan, Nowa Gwinea, Wenezuela, Wybrzeże Kości Słoniowej, Irak, Turkmenistan. W ostatnim liściu (dla którego  $\hat{y}=94,86$ ) znalazła się Uganda, Eritrea, Laos, Komory, Rwanda, Gwinea Bissau, Kenia, Somalia, Mali, Afganistan.

Jako miarę do porównania dwóch zbudowanych modeli wybrano błąd średniokwadratowy (MSE) oraz średni błąd bezwzględny (MAE) (Borkowski i in. 2003):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2, \quad MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|.$$

Tabela 2. Wartości błędów średniokwadratowych oraz średnich błędów bezwzględnych dla zbudowanych modeli dla indeksu korupcji w 2011 roku

	<i>MAE</i>	<i>MSE</i>
Model regresji wielorakiej	8,88	117,34
Drzewo regresyjne	7,60	97,47

Źródło: opracowanie własne

Interpretując otrzymane wyniki należy stwierdzić, że w przypadku modelowania indeksu korupcji znacząco lepsze wyniki można uzyskać wykorzystując drzewo regresyjne (Tabela 2). Wartość średniego błędu bezwzględnego dla drzewa regresyjnego była niższa o ok. 14%, natomiast wartość błędu średniokwadratowego o ok. 16%.

## PODSUMOWANIE

Przeprowadzone badania wykazały, że w przypadku modelowania indeksu korupcji znacząco lepsze wyniki można osiągnąć stosując drzewa regresyjne niż model regresji wielorakiej. Jest to oczywiście wniosek o charakterze wstępnym, który wymaga potwierdzenia na drodze przeprowadzenia kolejnych badań obejmujących chociażby inne lata. Niemniej należy stwierdzić, że różnica jest na tyle znacząca, że stanowi poważne wskazanie na model drzewa regresyjnego. Taki rezultat porównania modeli może wynikać z faktu, że w przypadku modelu regresji wielorakiej zakłada się występowanie liniowej zależności pomiędzy zmiennymi. Natomiast w przypadku modelowania indeksu korupcji zależność może być nieliniowa, co skutkuje wyższością modelu drzewa regresyjnego nad modelem regresji wielorakiej.

## BIBLIOGRAFIA

- Abed G., Davoodi H. (2000) Corruption, Structural Reforms, and Economic Performance in the Transition Economies. IMF, Working Paper No 00/132.
- Ades A., Di Tella R. (1999) Rents, competition and corruption. *American Economic Review*, 89, 982–993.
- Akçay S. (2006) Corruption and Human Development. *Cato Journal*, Vol. 26, No. 1.
- Bank Światowy (2004) World Development Report 2004: Making Services Work For Poor People. Washington.



- Bardhan P. (1997) Corruption and Development: A Review of Issues. *Journal of Economic Literature*, Vol. XXXV, s. 1320-1346.
- Begg D., Fisher S. i in. (2003) *Ekonomia: mikroekonomia*. Warszawa, Polskie Wydawnictwo Ekonomiczne.
- Berry M., Linoff G. (2004) *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd Edition. Wiley Publishing.
- Borkowski B., Dudek H. i in.. (2003) *Ekonometria: wybrane zagadnienia*. Warszawa, Wydaw. Naukowe PWN.
- Croux C., Dehon C. (2010) Influence Functions of the Spearman and Kendall Correlation Measures. *CentER Discussion Paper Series No. 2010-40*.
- Mauro, P. (1995), Corruption and growth, *Quarterly Journal of Economics* 110, p. 681-712
- Faraway J. (2004) *Linear models with R*. Chapman & Hall/CRC.
- Faraway J. (2006) *Extending the linear models with R: generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall/CRC.
- Greene W.H. (2003) *Econometric Analysis*, Fifth Edition. Prentice Hall, New Jersey.
- Koronacki J., Ćwik J. (2008) *Statystyczne systemy uczące się*. Warszawa, Akademicka Oficyna Wydawnicza EXIT.
- Lambsdorff J. (2007) *The institutional economics of corruption and reform, Theory, Evidence, and Policy*. Cambridge.
- Lee T.-S., Chiu S.-C., Lu S.-J. (2006) Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis* 50(4)
- Olken B. A. (2005) *Monitoring corruption: evidence from a field experiment in Indonesia*. NBER Working Paper 11753.
- Rose-Ackerman, S. (1997) *The Political Economy of Corruption*. w: K. A. Elliot (ed.) *Corruption and the Global Economy*, 31–60. Washington: Institute for International Economics.
- Tanzi, V. (1998) *Corruption around the World: Causes, Consequences, Scope, and Cures*. IMF Working Paper No 63. Washington, International Monetary Fund.

## ZAŁĄCZNIK

AG.LND.AGRI.K2	Grunty rolne (km <sup>2</sup> )
AG.LND.ARBL.HA.PC	Grunty orne (hektar na osobę)
AG.LND.ARBL.ZS	Grunty orne (% powierzchni gruntów)
AG.LND.CROP.ZS	Trwałe pola uprawne (% powierzchni gruntów)
AG.LND.TOTL.K2	Powierzchnia kraju (km <sup>2</sup> )
AG.PRD.CROP.XD	Indeks produkcji roślinnej (1999-2001 = 100)
AG.PRD.FOOD.XD	Indeks produkcji żywności (1999-2001 = 100)
AG.PRD.LVSK.XD	Indeks produkcji zwierzęcej (1999-2001 = 100)
AG.SRF.TOTL.K2	Całkowita powierzchnia kraju (km <sup>2</sup> )
AG.YLD.CREL.KG	Plon zbóż (kg na hektar)
BM.GSR.INSF.ZS	Ubezpieczenia i usługi finansowe (% importu usług, BOP)
BM.GSR.TRVL.ZS	Usługi turystyczne (% importu usług, BOP)
BX.GSR.CMCP.ZS	Komunikacja, komputer, itp (% eksportu usług, BOP)
BX.KLT.DINV.WD.GD.ZS	Bezpośrednie inwestycje zagraniczne, napływ netto (% PKB)

EG.GDP.PUSE.KO.PP	PKB w przeliczeniu na jednostkę zużycia energii (PPP \$ za kg ekwiwalentu ropy)
EG.GDP.PUSE.KO.PP.KD	PKB w przeliczeniu na jednostkę zużycia energii (stałe 2005 \$ PPP na kg ekwiwalentu ropy)
EG.USE.COMM.KT.OE	Zużycie energii (kt ekwiwalentu ropy)
EG.USE.COMM.GD.PP.KD	Zużycie energii (kg ekwiwalentu ropy) za każdy 1.000 USD PKB (PPP stała 2005)
EG.USE.ELEC.KH.PC	Zużycie energii elektrycznej (kWh na mieszkańca)
EN.ATM.CO2E.KD.GD	Emisja CO2 (kg na 2000 US \$ PKB)
EN.ATM.CO2E.PC	Emisje CO2 (ton na mieszkańca)
EN.ATM.PM10.MC.M3	PM10 poziomu kraju (mikrogramów na metr sześcienny)
ER.H2O.INTR.K3	Odnawialne wewnętrzne zasoby słodkiej wody ogółem (mld metrów sześciennych)
ER.H2O.INTR.PC	Odnawialne wewnętrzne zasoby słodkiej wody na mieszkańca (w metrach sześciennych)
FM.LBL.MQMY.GD.ZS	Pieniądze i quasi pieniądze (M2) jako % PKB
FS.AST.PRVT.GD.ZS	Kredyty krajowe sektora prywatnego (% PKB)
IC.CRD.PRVT.ZS	Procent ludności w biurach informacji kredytowej
IC.EXP.DURS	Czas realizacji do wywozu (dni)
IC.LGL.CRED.XQ	Siła indeksu praw (0 = słaby na 10 = silny)
NE.RSB.GNFS.ZS	Saldo zewnętrzne towarów i usług (w % PKB)
NE.TRD.GNFS.ZS	Handel (% PKB)
Failed States Index	Indeks korupcji

## CORRUPTION INDEX ANALYSIS USING MULTIPLE REGRESSION MODEL AND REGRESSION TREE

**Abstract:** This paper presents the results of research on corruption index modeling (Failed States Index). The multiple regression model was constructed and compared with the regression tree. The research was based on the data published by an independent organization The Fund for Peace. Predictors were selected from a set of variables available on the website of the World Bank. The preliminary results clearly indicate that the regression tree better reflects the variation of the parameter than the multiple regression model.

**Keywords:** Failed States Index, corruption index, multiple regression model, regression tree