METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XIII/1, 2012, str. 82 – 95

ZAGADNIENIE REGRESJI W NAUKACH EKONOMICZNYCH

Beata Fałda, Józef Zając
Instytut Matematyki i Informatyki
Państwowa Wyższa Szkoła Zawodowa w Chełmie
Katedra Zastosowań Matematyki, Katolicki Uniwersytet Lubelski Jana Pawła II
e-mail: bfalda@kul.lublin.pl, jzajac@kul.lublin.pl

Streszczenie: Teoria regresji obejmuje zespół metod i narzędzi ścisłego opisu zależności występujących między różnego rodzaju zjawiskami. Od wielu lat jest wykorzystywana do formułowania modeli ekonomicznych i ekonometrycznych, jakkolwiek regresja jest tu rozumiana w różnoraki sposób. Przedmiotem niniejszej pracy są rozważania na temat zastosowania metod ilościowych w modelowaniu procesów ekonomicznych, ze szczególnym uwzględnieniem zależności zachodzących między nimi, wyrażonymi związkami regresyjnymi.

Słowa kluczowe: teoria regresji, regresja uogólniona

WSTĘP

Jedna z wielu definicji ekonomii mówi, że jest ona nauką obejmującą odkrywanie i formułowanie społecznych praw dotyczących działalności gospodarczej i stosunków ekonomicznych[1]. Oznacza to, że przedmiotem zainteresowania ekonomii są procesy, m.in.: produkcji, wymiany, podziału i spożycia. Aby móc je naukowo opisywać, badać i analizować potrzebne są definicje, narzędzia i odpowiednie metody.[2] Podstawowymi narzędziami

[1] Grudzewski W. M., Rosłanowska-Plichcińska K. (1984) Mierzenie wielkości i wymiarowe modelowanie zjawisk oraz procesów ekonomicznych, Zakład Narodowy im. Ossolińskich, Wydawnictwo Polskiej Akademii Nauk, Wrocław - Warszawa - ... - Łódź, str. 9.
[2] Fałda B. (2010) Modelowanie dynamiczne procesów ekonomicznych, Wydawnictwo KUL, Lublin, str. 15-16.

wykorzystywanymi w tym celu są różnego typu modele. Wśród nich zasadniczą grupę stanowią modele matematyczne i statystyczne, w tym modele regresji.

Termin „regresja" pojawił się po raz pierwszy w XIX wieku. Został on użyty przez F. Galtona (1822 - 1911) do opisu związków zachodzących w procesie dziedziczenia.[3] Z czasem metody i techniki regresji znalazły zastosowanie w różnych dziedzinach nauki, w tym w ekonomii, do opisu kształtowania się poziomu pewnego zjawiska w czasie lub do opisu zależności zachodzących między badanymi wielkościami. Analiza regresji skupiła się na estymacji parametrów równania teoretycznego regresji, które w sposób dokładny potrafi odwzorować istniejący związek między rozważanymi procesami lub zjawiskami. Na gruncie ekonomii rozwój metod regresyjnych dokonał się w obrębie ekonometrii, gdzie zasadniczy kierunek badań dotyczy stochastycznego charakteru modelowanych procesów ekonomicznych.

Przedmiotem niniejszej pracy są rozważania na temat zastosowania metod ilościowych w modelowaniu zjawisk i procesów ekonomicznych, ze szczególnym uwzględnieniem zależności zachodzących między nimi, wyrażonymi związkami regresyjnymi. Oprócz charakterystyki klasycznego podejścia do problemu regresji, zaprezentowano możliwość wykorzystania na gruncie ekonomicznym modelu regresji uogólnionej, wprowadzonego w pracy [Zając 2010].

METODY ILOŚCIOWE W NAUKACH EKONOMICZNYCH

Za prekursora stosowania metod ilościowych w ekonomii, a przede wszystkim statystyki, uznaje się W. Petty'ego (1623 - 1687). Według niego statystyka była metodą rozumowania na podstawie liczb, umożliwiającą wykrycie określonych prawidłowości wśród chaotycznych zjawisk masowych. Na uwagę zasługują również badania prowadzone przez G. Kinga (1648 - 1712), który jako pierwszy podjął próbę ilościowego opisu zależności pomiędzy zmianą cen kukurydzy a wielkością jej zbiorów przy pomocy funkcji liniowej. Z kolei G. U. Yule (1871 - 1951), w swoich pracach pochodzących z 1895 i 1896 roku, proponował zastosowanie analizy korelacji w ekonomii do badania związku pomiędzy ubóstwem a przeciwdziałaniem temu zjawisku. R. H. Hooker (1867 - 1944) w 1901 roku wykorzystał tę samą technikę badania związku między zmiennymi, analizując korelację pomiędzy współczynnikiem małżeństw i stanem koniunktury w Wielkiej Brytanii. Prowadząc swoje badania był jednak świadomy ograniczeń jakie niesie za sobą wykorzystanie analizy korelacji, szczególnie w przypadku jej stosowania do badania danych w postaci szeregów czasowych. W 1907 roku R. Benini (1862 - 1956) jako pierwszy wykorzystał w ekonomii metodę regresji wielorakiej, zaś w 1914 roku H. L. Moore (1869 - 1958) zasłynął wprowadzeniem na grunt ekonomii statystycznej estymacji parametrów

[3] Sen A. K., Srivastava M. S. (1990) Regression analysis: theory, methods and applications, Springer-Verlag, New York, str. 1.

ekonomicznych. W 1900 roku L. Bachelier (1870 - 1946), wykorzystując szeregi czasowe cen akcji na paryskiej giełdzie, zauważył losowy charakter tych cen, co stało się w późniejszym okresie podstawą do rozważań na temat efektywności rynku.[4]

Wykorzystanie narzędzi i metod matematycznych oraz statystycznych w ekonomii zaowocowało rozkwitem dwóch, wzajemnie uzupełniających się nurtów zastosowań metod ilościowych: ekonomii matematycznej i ekonometrii. O ile ekonomia matematyczna została nakierowana na formułowanie analitycznych i jakościowych modeli teorii ekonomii, o tyle ekonometria, bazująca na osiągnięciach rachunku prawdopodobieństwa i statystyki, ma charakter empiryczny i ilościowy.[5]

Obserwując proces matematyzacji ekonomii trudno jednoznacznie stwierdzić, który z nurtów był pierwszy. Chociaż ekonometria ze swoim wybitnie praktycznym zabarwieniem interesowała ludzi od dawna, jednak jej faktyczny rozwój jest datowany na lata trzydzieste XX wieku.[6]

Analiza prowadzonych badań ekonometrycznych pokazuje jednak, że pomimo licznych sporów na tle zastosowania różnych metodologii, podstawą ekonometrii jest teoria regresji, modyfikowana i uzupełniana przez wybitnych ekonometryków, m.in T. Haavelmo (1911 - 1999). Jego prace badawcze, dotyczące probabilistycznego podejścia w ekonometrii, zostały docenione Nagrodą Nobla w 1989 roku. Zwracając uwagę, iż modele ekonometryczne są konstruowane w oparciu o statystykę i dane statystyczne, które bazują na teorii prawdopodobieństwa, uważał, iż należy pogodzić się z faktem, że otrzymane wyniki będą miały charakter probabilistyczny, a nie deterministyczny. Tym samym przeniósł ciężar analiz ekonomiczno - matematycznych z szacowania parametrów oraz problemu jakości danych statystycznych na testowanie teorii.[7]

ANALIZA REGRESJI

Klasyczny model matematyczno - ekonomiczny rozpatrywany jest na gruncie ekonomii matematycznej, gdzie analiza zjawisk ekonomicznych nie opiera się na badaniu wyników obserwacji empirycznych za pomocą statystycznych metod estymacji i testowania hipotez, lecz odnosi się do teoretycznych rozważań ekonomicznych. Większość wspomnianych modeli, po odpowiednim zmodyfikowaniu, może być podstawą analiz ekonometrycznych, w których

[4] Geweke J. F., Horowitz J. L., Pesaran M. H. (2006) Econometrics: A Bird's Eye View, IZA, Discussion Paper Series No. 2458, str. 3-4.

[5] Fałda B. (2010), Modelowanie dynamiczne procesów ekonomicznych, Wydawnictwo KUL, Lublin, str. 126.

[6] Stankiewicz W. (2000) Historia myśli ekonomicznej, Państwowe Wydawnictwo Ekonomiczne, Warszawa, str. 401.

[7] Gruszecki T. (2001) Nagrody Nobla w ekonomii, Verba, Lublin, str. 48-49.

punktem wyjścia jest teoria regresji. W literaturze ekonometrycznej można spotkać dwa, wzajemnie uzupełniające się, podejścia charakteryzujące model ekonometryczny: deterministyczne i stochastyczne.[8]

Pomimo, wydawałoby się, klarownego rozróżnienia koncepcji regresji występujących na gruncie ekonomii analiza literatury przedmiotu, zaprezentowana w pracy [Czerwiński 2002], wskazuje na niejednoznaczność tego pojęcia oraz błędy interpretacyjne. Poniżej przedstawiono uwagi zawarte w cytowanej pracy w odniesieniu do wspomnianego problemu wraz z prezentacją różnych podejść do regresji.[9]

Regresja jako dopasowanie funkcji określonej klasy do wyników obserwacji

Załóżmy, że poszukujemy funkcji regresji w postaci $y = f_0(x)$, gdzie f_0 jest pewną funkcją zmiennej rzeczywistej x, należącą do ustalonej klasy funkcji Φ i taką, że dla każdej funkcji $f \in \Phi$ wartość pewnego funkcjonału $H(f;\hat{y},\hat{x})$ spełnia nierówność:

$$H(f_0;\hat{y},\hat{x}) \le H(f;\hat{y},\hat{x}) \tag{1}$$

Funkcjonał H jest miarą określoną na przestrzeni Φ. Przy jego pomocy dokonujemy pomiaru odchylenia ciągu wartości mierzonych od wartości przyjmowanych przez funkcję f w punktach pomiarowych $\hat{x} = \{x_i\}$.

W stosunku do miary H zakładamy, że:

$$H(f_0;\hat{y},\hat{x}) \le h_0, \tag{2}$$

co oznacza kryterium stosowania funkcji H.

Jeżeli istnieje funkcja spełniająca warunki (1)-(2) to możemy powiedzieć, że problem znalezienia funkcji regresji $f_0 \in \Phi$ jest poprawnie sformułowany. Procedura wyznaczenia takiej funkcji, przy założeniu liniowości i zupełności przestrzeni funkcyjnej Φ, wymaga jednak umiejętności zastosowania teorii punktów stałych i umiejętności konstrukcji ciągu przybliżającego.

Pomijając założenie zupełności przestrzeni Φ należy przyjąć dodatkowy warunek, pozwalający na to, aby - nie szukając funkcji f_0 spełniającej (1) - zadowolić się funkcją f_0^* „zbliżoną" do f_0.

Regresja - wariant 1

Przyjmujemy, iż wyniki obserwacji $(\hat{y}_i, \hat{x}_i)$ powstały w następujący sposób:

---

[8] Czerwiński Z. (1982) Matematyczne modelowanie procesów ekonomicznych, Państwowe Wydawnictwo Naukowe PWN, Warszawa, str. 81

[9] Czerwiński Z. (2002) Moje zmagania z ekonomią, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, str. 420-428.

1. wartości $\hat{x}_i$, $i = 1,2,...,n,$ zostały z góry ustalone,
2. wartości $\hat{y}_i$, $i = 1,2,...,n,$ są realizacjami zmiennej losowej η_x, która przy danym x jest pewną funkcją g zmiennej losowej ε o skończonej, choć nieznanej wariancji i znanej wartości oczekiwanej, co zapisujemy w postaci

$$\eta_x = g(x,\varepsilon). \tag{3}$$

Szczególnym i najczęściej występującym przypadkiem warunku 2, jest założenie, że istnieją stałe rzeczywiste a_0, a_1 takie, iż dla dowolnego x

$$\eta_x = a_0 + a_1 x + \varepsilon, \tag{4}$$

przy czym $E(\varepsilon) = 0.$

Ze wzoru (3) wynika, że dystrybuanta zmiennej losowej η_x jest przy danym x jednoznacznie wyznaczona przez dystrybuantę zmiennej losowej $\varepsilon.$ W przypadku warunku (4) dystrybuanta zmiennej losowej η_x jest związana, przy dowolnie ustalonym x, z dystrybuantą zmiennej losowej $\varepsilon,$ wzorem

$$F_\eta(y) = F_\varepsilon(y - a_0 - a_1 x). \tag{5}$$

Ze wzoru (3) wynika, że zaobserwowane w punktach $\hat{x}_i$ realizacje $\hat{y}_i$ są określone równaniem:

$$\hat{y}_i = g(\hat{x}_i, e_i), \tag{6}$$

gdzie e_i, $i = 1,2,...,n,$ jest realizacją zmiennej losowej $\varepsilon.$ W przypadku liniowym mamy zatem

$$\hat{y}_i = a_0 + a_1\hat{x}_i + e_i. \tag{7}$$

Równanie (6) wnosi jednak mniej informacji niż równanie (3), zaś równanie (7) - mniej informacji niż równanie (4), gdyż warunki (6) i (7) mówią tylko o określonych realizacjach.

Równaniem regresji w przedstawionej konstrukcji nazywamy równanie (3), które w szczególnym przypadku występuje w postaci (4).

Regresja - wariant 2

Przyjmujemy, iż wyniki obserwacji $(\hat{y}_i, \hat{x}_i)$ powstały w następujący sposób:

1. wartości $\hat{x}_i$, $i = 1,2,...,n,$ zostały z góry ustalone,
2. ciąg wartości $\hat{y}_i$, $i = 1,2,...,n,$ jest realizacją zmiennej losowej n-wymiarowej $(\eta_1,...,\eta_n),$ która przy danym wektorze $x = (x_1,...,x_n)$ jest pewną funkcją g tego wektora oraz n-wymiarowej zmiennej losowej $\varepsilon = (\varepsilon_1,...,\varepsilon_n)$. Zmienne losowe ε_i, $i = 1,2,...,n,$ mają rozkłady brzegowe o skończonej wariancji i zerowej wartości oczekiwanej.

Szczególnym przypadkiem założenia 2 jest istnienie liczb rzeczywistych a_0 i a_1 takich, że dla wektora x zachodzi zależność

$$\eta_i = a_0 + a_1 d_i x + \varepsilon_i, \tag{8}$$

przy czym $E(\varepsilon) = 0$.

Symbol d_i oznacza tutaj i-ty n-wymiarowy wektor jednostkowy, zaś $d_i x$ jest iloczynem skalarnym wektorów d_i oraz x. Założmy teraz dodatkowo, że η_i zależy tylko od x_i. Wynika stąd, że realizacje poszczególnych składowych tych zmiennych spełniają równanie

$$\hat{y}_i = a_0 + a_1 \hat{x}_i + e_i. \tag{9}$$

gdzie e_i, $i = 1,2,...,n$, jest nieznaną realizacją i-tej składowej n-wymiarowej zmiennej losowej $\varepsilon = (\varepsilon_1,...,\varepsilon_n)$. Równaniem regresji w tej konstrukcji jest równanie (8), które wymaga umiejętności wyznaczania parametrów a_0 i a_1.

Regresja I rodzaju

Przyjmijmy, że wyniki obserwacji $(\hat{y}_i, \hat{x}_i)$ są realizacjami dwuwymiarowej zmiennej losowej (η, ς) o rozkładzie normalnym. Funkcja regresji I rodzaju jest określona równaniem $y = E(\eta \mid \varsigma = x)$.

W przypadku dwuwymiarowego rozkładu normalnego przyjmuje ono postać $y = a_0 + a_1 x$, gdzie parametry a_0 i a_1 są wyznaczone jednoznacznie.

Rozważana koncepcja regresji jest rzadko stosowana w naukach ekonomicznych, ponieważ analizowane zjawiska lub procesy nie podlegają najczęściej opisowi, w którym rozważane zmienne mają łączny rozkład normalny.

Regresja II rodzaju

Przyjmujemy teraz, że wyniki obserwacji $(\hat{y}_i, \hat{x}_i)$ są realizacjami dwuwymiarowej zmiennej losowej (η, ς) o rozkładzie normalnym. Pod pojęciem regresji rozumiemy tutaj funkcję liniową postaci $y = a_0 + a_1 x$, której parametry a_0 i a_1 spełniają nierówność

$$E(\eta - q - p\varsigma)^2 \geq E(\eta - a_0 - a_1\varsigma) \tag{10}$$

dla dowolnych rzeczywistych p, q.

Pomimo powszechności sądów, ta koncepcja regresji rzadko znajduje uzasadnienie do stosowania w ekonomii.

Ponieważ w modelach ekonometrycznych składnikowi losowemu przypisuje się dość istotną rolę rozważmy go jako zmienną losową postaci

$\varepsilon(p,q) = \eta - p - q\varsigma$ i postawmy problem znalezienia takiej pary liczb p, q, przy której zmienna ta osiąga minimalną wariancję. Jeżeli liczbami tymi będą a_0 i a_1, zaś $\{\hat{y}_i, \hat{x}_i\}$ są realizacjami dwuwymiarowej zmiennej losowej (η, ς) o rozkładzie normalnym, to zadanie oszacowania parametrów, przy których zmienna losowa $\varepsilon(p,q)$ ma minimalną wariancję, jest dobrze postawione i metoda najmniejszych kwadratów daje oceny parametrów a_0 i a_1 o pożądanych własnościach.

Niestety tak zdefiniowana zmienna losowa $\varepsilon(a_0, a_1)$ nie może być uważana za składnik losowy, wywierający wpływ na kształtowanie się zmiennej objaśnianej.

Jeżeli $\varepsilon = \varepsilon(a_0, a_1)$, zaś e_i jest nieznaną realizacją zmiennej losowej ε, to równanie

$$\hat{y}_i = a_0 + a_1\hat{x}_i + e_i \tag{11}$$

jest tożsamością. Mimo formalnego podobieństwa równań (7) oraz (9), nie oznacza to, aby wartości $\hat{y}_i$ powstawały jako liniowa funkcja wartości $\hat{x}_i$ powiększonych o nieznane wartości e_i. Jeżeli wyniki obserwacji obydwu zmiennych są realizacjami dwuwymiarowej zmiennej losowej o rozkładzie normalnym, to często spotykany komentarz, że „zmienna objaśniana kształtuje się pod wpływem zmiennej objaśniającej oraz składnika losowego" traci sens, gdyż składnik losowy jest tu wartością resztową.

Regresja z losowymi zmiennymi objaśniającymi

Przyjmujemy, że wyniki obserwacji $(\hat{y}_i, \hat{x}_i)$ spełniają następujące warunki:

1. wartości $\hat{x}_i$, $i = 1,2,\ldots,n,$ są realizacjami pewnej zmiennej losowej ς,
2. wartości $\hat{y}_i$, $i = 1,2,\ldots,n,$ spełniają warunek $\hat{y}_i = a_0 + a_1\hat{x}_i + e_i$, gdzie e_i jest realizacją pewnej zmiennej losowej ε, niezależnej od ς, a ponadto $E(\varepsilon) = 0$.

Wynika stąd, że $\hat{y}_i$ są realizacjami pewnej zmiennej losowej η postaci $\eta = a_0 + a_1\varsigma + \varepsilon,$ która jest funkcją liniową dwóch zmiennych losowych ς i ε. Stąd

$$\hat{y}_i = a_0 + a_1\hat{x}_i + e_i. \tag{12}$$

Równanie (12) różni się od równania (7) tym, że założenia dotyczące $\hat{x}_i$ mówią, iż wartości $\hat{y}_i$ zostały wylosowane a nie z góry ustalone. Natomiast równanie (12) różni się od (11) tym, że brak jest tutaj założenia o jednoczesnym losowaniu liczb $\hat{y}_i$ oraz $\hat{x}_i$. Liczby $\hat{y}_i$ są realizacjami zmiennej losowej, które na mocy założenia powstają jako wartości liniowego przekształcenia realizacji dwóch innych zmiennych losowych.

KONCEPCJA REGRESJI UOGÓLNIONEJ I JEJ ZASTOSOWANIE W EKONOMII

Model regresji uogólnionej należy do modeli regresji o charakterze deterministycznym. W modelu tym funkcje regresji powstają jako rozwiązanie pewnego problemu ekstremalnego, określonego w środowisku skończenie lub nieskończenie wymiarowej przestrzeni Hilberta. Konstrukcja tego modelu pozwala na wyznaczenie ciągu regresji, „aproksymującego" dane doświadczalne lub notowania ciągłe, w ściśle określonym sensie.

Rozważmy następującą strukturę $\Re := (A, B, \delta; x, y)$, gdzie:

- A, B są danymi niepustymi zbiorami,
- dane otrzymane w drodze eksperymentu $x : \Omega_1 \to A$ oraz $y : \Omega_2 \to B$ zachodzą dla pewnych niepustych zbiorów Ω_1 i Ω_2,
- $\delta : (\Omega_1 \to B) \times (\Omega_2 \to B) \to \overline{\mathbb{R}}$ jest kryterium odchylenia funkcji teoretycznej od funkcji empirycznej, którą nazywać będziemy strukturą regresji.[10]

Celem teorii regresji uogólnionej jest wyznaczenie takich funkcji $f_0 \in \boldsymbol{F}$, które spełniają warunek najlepszego dopasowania do danych empirycznych. Funkcje te minimalizują funkcjonał dopasowania $F(f) := \delta(f \circ x, y)$ czyli spełniają nierówność $F(f) \geq F(f_0)$ dla każdego $f \in \boldsymbol{F}$. Zbiór tych funkcji oznaczamy symbolem $M(\boldsymbol{F}, \Re)$, zaś każdą funkcję ze zbioru rozwiązań nazywamy funkcją regresji rodziny $\boldsymbol{F}$ z uwagi na strukturę $\Re$.

Uogólnienie klasycznego odchylenia kwadratowego, liczonego względem dowolnej miary μ, jest określone formułą $\delta(u, v) = \int\limits_{\Omega_1 \times \Omega_2} |u(t_1) - v(t_2)|^2 d\mu(t_1, t_2)$.

Przechodząc do pomiarów dyskretnych i przyjmując $t_1 \to k$ oraz $t_2 \to l$ widzimy, że miara $\mu : \boldsymbol{B} \to [0; +\infty]$ redukuje się do wyrażenia $\mu(\{(k, l)\}) = \rho_{k,l}$, gdzie $\rho_{k,l}$ jest dowolną nieujemną funkcją rzeczywistą. Wtedy $\delta(u, v) = \sum_{k=0}^{n} \sum_{l=0}^{m} \rho_{k,l} |u(k) - v(k)|^2$.

Przyjmując dalej, że $m = n$, zaś

[10] Zając J. (2010) Regresja uogólniona, „Miscellanea mikroekonometrii", Uniwersytet Szczeciński, Szczecin, str. 170-171.

$$\rho_{k,l} := \begin{cases} 1 & dla \quad k = l, \\ 0 & dla \quad k \neq l \end{cases} \tag{13}$$

otrzymujemy klasyczna miarę Gaussa odchylenia kwadratowego $\delta(u,v) = \sum_{k=0}^{n} |u(k) - v(k)|^2$.

Stosując podstawienia $u := f \circ x$ oraz $v := y$ widzimy, że

$$F(f) := \delta(f \circ x, y) = \sum_{k=0}^{n} |f \circ x(k) - y(k)|^2 = \sum_{k=0}^{n} |f(x_k) - y_k|^2, \tag{14}$$

gdzie x_k oraz y_k są danymi empirycznymi. Prezentowane uogólnienie klasycznie rozumianej regresji polega na tym, że zamiast funkcji specjalnych *f*, występujących w różnych znanych typach regresji rozważamy dowolną rodzinę $L_1(\Re)$ funkcji $f : A \to B$, takich że $f \circ x(t_1)$ jest funkcją $\boldsymbol{B}$ - mierzalną i taką że $\int\limits_{\Omega_1 \times \Omega_2} |f \circ x(t_1)|^2 d\mu(t_1, t_2) < +\infty$, jak również zbiór $L_2(\Re)$, składający się ze wszystkich funkcji $g : \Omega_2 \to B$, takich że $g(t_2)$ jest funkcją $\boldsymbol{B}$ - mierzalną, a ponadto $\int\limits_{\Omega_1 \times \Omega_2} |g(t_2)|^2 d\mu(t_1, t_2) < +\infty$.

Zatem funkcjonał

$$L_1(\Re) \ni u \to g^*(u) := \int\limits_{\Omega_1 \times \Omega_2} u \circ x(t_1) \overline{g \circ y}(t_2) d\mu(t_1, t_2) \tag{15}$$

jest liniowy i ograniczony w dowolnej przestrzeni Hilberta, zaś jego norma supremum spełnia nierówność

$$\sup\{|g^*(f)| : f \in L_1(\Re) \wedge \|f\| \leq 1\} \leq \left\{ \int\limits_{\Omega_1 \times \Omega_2} |g \circ y(t_2)|^2 d\mu(t_1, t_2) \right\}^{1/2}. \tag{16}$$

Stosując nierówność Schwarza widzimy, że

$$|g^*(f)| \leq \left(\int\limits_{\Omega_1 \times \Omega_2} |g \circ y(t_2)|^2 d\mu(t_1, t_2) \right)^{1/2} \|f\|, \tag{17}$$

gdzie $\|f\|$ jest normą f w przestrzeni $L_1(\Re)$.

Zatem problem regresji uogólnionej, rozumiany jako określony powyżej problem ekstremalny, polega na znalezieniu wszystkich funkcji $f_0 \in \boldsymbol{F}$, które minimalizują funkcjonał

$$F_g(f)=\delta(f\circ x, g\circ y)=\int\limits_{\Omega_1\times\Omega_2}\left|f\circ x(t_1)-g\circ y(t_2)\right|^2 d\mu(t_1,t_2), \tag{18}$$

gdzie $f\in \boldsymbol{F}$, zaś $g\in L_2(\Re)$.

Podstawowym wynikiem dotyczącym funkcji regresji jest informacja mówiąca, że jest to zbiór liniowy, który składa się z funkcji będących kombinacją liniową dowolnie wybranych, unormowanych funkcji bazowych.

Jeżeli $p\in \mathbb{N}\cup\{\infty\}$ jest wymiarem poszukiwanego ciągu regresji, niech $h_k\in L_1(\Re)$, gdzie $k=1,2,\dots,p$, będzie ciągiem funkcji liniowo niezależnych, zaś $\boldsymbol{F}$ przestrzenią liniową rozpiętą na wektorach h_k. Wykorzystując metodę ortogonalizacji Gramma-Schmidta z ciągiem tym łączymy ciąg $\{h_k^{'}\}$, którego elementy określone są wzorem

$$h_1^{'}:=h_1 \text{ oraz } h_n^{'}:=h_n-\sum_{k=1}^{n-1}\frac{\langle h_n\mid h_k^{'}\rangle}{\left\|h_k^{'}\right\|^2}h_k^{'}, \text{ dla } n=2,3,\dots,p. \tag{19}$$

Wtedy

$$Reg(\boldsymbol{F},\Re_g)=\sum_{k=1}^{p}\frac{\overline{g^*(h_k^{'})}}{\left\|h_k^{'}\right\|^2}h_k^{'}, \tag{20}$$

gdy p jest skończoną liczbą naturalną, oraz $Reg(\boldsymbol{F},\Re_g)=\sum_{k=1}^{\infty}\frac{\overline{g^*(h_k^{'})}}{\left\|h_k^{'}\right\|^2}h_k^{'}$, gdy $p=\infty$.[11]

Dwuwymiarowy przykład funkcji regresji otrzymujemy w następujący sposób. Załóżmy, że dana jest dowolna funkcja $g\in L_2(\Re)$, zaś $\boldsymbol{F}=lin(\{h_1,h_2\})$ jest przestrzenią liniową rozpiętą na liniowo niezależnych funkcjach $h_1,h_2\in L_1(\Re)$. Wtedy $Reg(\boldsymbol{F},\Re_g)=\sum_{k=1}^{2}\frac{\overline{g^*(h_k^{'})}}{\left\|h_k^{'}\right\|^2}h_k^{'}$, gdzie $h_1^{'}:=h_1$ oraz

$$h_2^{'}:=h_2-\frac{\langle h_2\mid h_1^{'}\rangle}{\left\|h_1^{'}\right\|^2}h_1^{'}=h_2-\frac{\langle h_2\mid h_1\rangle}{\left\|h_1\right\|^2}h_1.$$

[11] Partyka D., Zając J. (2010) Generalized problem of regression, Bulletin De La Societe Des Sciences Et Des Lettres De Łódź, vol. LX, no. 1, str. 86-87.

Przyjmując

$$a_2 := \frac{\overline{g^*\left(h_2^{'}\right)}}{\left\|h_2^{'}\right\|^2} \text{ oraz } a_1 := \frac{\overline{g^*\left(h_1^{'}\right)}}{\left\|h_1^{'}\right\|^2} - \frac{\langle h_2 \mid h_1 \rangle}{\left\|h_1^{'}\right\|^2} a_2 \tag{21}$$

widzimy, że

$$Reg(\boldsymbol{F}, \Re) = a_2 h_2 + a_1 h_1, \tag{22}$$

gdzie

$$a_2 = \frac{\overline{g^*(h_2)}\|h_1\|^2 - \overline{g^*(h_1)\langle h_2 \mid h_1 \rangle}}{\|h_2\|^2\|h_1\|^2 - \left|\langle h_2 \mid h_1 \rangle\right|^2} \text{ oraz } a_1 = \frac{\overline{g^*(h_1)} - \langle h_2 \mid h_1 \rangle a_2}{\|h_1\|^2}. \tag{23}$$

Jeżeli dodatkowo przyjmiemy, że $h_2(t) = t$, zaś $h_1(t) = 1$, a ponadto $g(t) = t$ dla $t \in$ ✡ to wyrażenia określające a_1 i a_2 przyjmują znaną postać

$$a_2 = \frac{(n+1)\sum_{k=0}^{n} x_k y_k - \sum_{k=0}^{n} x_k \sum_{k=0}^{n} y_k}{(n+1)\sum_{k=0}^{n} x_k^2 - \left(\sum_{k=0}^{n} x_k\right)^2} \text{ oraz } a_1 = \frac{\sum_{k=0}^{n} y_k - a_2 \sum_{k=0}^{n} x_k}{n+1} \tag{24}$$

współczynników regresji liniowej drugiego rodzaju postaci $f_0(t) = a_2 t + a_1$.

Poniżej przedstawiona zostanie ilustracja funkcjonowania teorii regresji uogólnionej w opisie przebiegu zmian przeciętnego miesięcznego wynagrodzenia nominalnego w Polsce w sektorze przedsiębiorstw.

W analizie wykorzystano dane za okres 01.1998-12.2011 dostępne na www.money.pl. Ciąg funkcji regresji $\{r_k\}$ dla $k = 0,1,2,3,4,5,6,7$ wyznaczony zostanie w ośmiowymiarowej przestrzeni liniowej, z wektorami bazowymi postaci:

$$h_0(x) = 1,\ h_1(x) = x,\ h_2(x) = \cos\frac{x\pi}{6},\ h_3(x) = \sin\frac{x\pi}{6},$$

$$h_4(x) = x^2,\ h_5(x) = x^3,\ h_6(x) = \frac{1}{x+1},\ h_7(x) = x^4.$$

Wybór wektorów bazowych został podyktowany niestabilnym charakterem rozważanych danych.

Stosownie do przyjętej bazy otrzymujemy:

$r_0(x) = 2574{,}60,$

$r_1(x) = 41496{,}61 + 12{,}75x,$

$r_2(x) = 1497{,}37 + 12{,}74x + 42{,}48\cos\frac{x\pi}{6},$

$r_3(x) = 1498{,}11 + 12{,}73x + 42{,}49\cos\frac{x\pi}{6} - 10{,}93\sin\frac{x\pi}{6},$

$$r_4(x) = 1519{,}9 + 11{,}97x + 0{,}0045x^2 + 42{,}43\cos\frac{x\pi}{6} - 10{,}95\sin\frac{x\pi}{6},$$

$$r_5(x) = 1314{,}97 + 26{,}27x - 0{,}20x^2 + 0{,}00083x^3 + 40{,}09\cos\frac{x\pi}{6} - 2{,}23\sin\frac{x\pi}{6},$$

$$r_6(x) = 1536{,}72 + 18{,}57x - 0{,}12x^2 + 0{,}00055x^3 - \frac{1137{,}95}{1+x} + 44{,}42\cos\frac{x\pi}{6} +$$

$$+ 3{,}09\sin\frac{x\pi}{6},$$

$$r_7(x) = 1962{,}63 - 0{,}007x + 0{,}1x^2 - 0{,}00022x^3 - 9{,}35\cdot 10^{-9}x^4 - \frac{2505{,}56}{1+x} +$$

$$+ 50{,}89\cos\frac{x\pi}{6} + 4{,}32\sin\frac{x\pi}{6}.$$

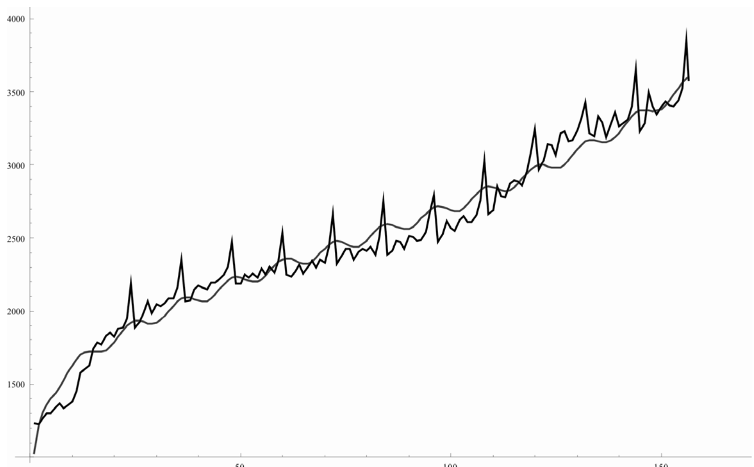
W celu sprawdzenia jakości powyższych modeli regresji zostały obliczone miary dopasowania $\delta_k(x)$:

$$\delta_0(x) = 8246{,}38,\ \delta_1(x) = 1922{,}74,\ \delta_2(x) = 1882{,}9,\ \delta_3(x) = 1880{,}24,$$

$$\delta_4(x) = 1876{,}14,\ \delta_5(x) = 1610{,}6,\ \delta_6(x) = 1538{,}04,\ \delta_7(x) = 1740{,}38.$$

Z przedstawionych danych wynika, że najlepsze dopasowanie do danych empirycznych realizuje funkcja $r_6(x)$. Wykres wskazanej funkcji przedstawiono na rysunku poniżej.

Rysunek 1. Funkcja regresji $r_6(x)$



Źródło: Opracowanie własne na podstawie danych www.money.pl, przy wykorzystaniu programu Wolfram Mathematica 6.0

PODSUMOWANIE

Teoria regresji stanowi podstawę wielu analiz prowadzonych na gruncie ekonomii. Jak zaprezentowano w niniejszej pracy pojęcie to nie ma charakteru jednoznacznego, jednak, co warto podkreślić, zawsze ma na celu przedstawienie związków między badanymi wielkościami.

Uzupełnieniem klasycznej teorii regresji jest teoria regresji uogólnionej. Już teraz wiele analiz wskazuje, iż może być ona wykorzystywana w modelowaniu procesów i zjawisk ekonomicznych. Jej niewątpliwym atutem jest możliwość elastycznego doboru bazy w rozważanej przestrzeni wektorowej pozwala na uwypuklenie tych cech przedmiotu analiz, które są według badacza najbardziej istotne.

BIBLIOGRAFIA

Czerwiński Z. (1982) Matematyczne modelowanie procesów ekonomicznych, Państwowe Wydawnictwo Naukowe PWN, Warszawa, str. 81.

Czerwiński Z. (2002) Moje zmagania z ekonomią, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań, str. 420-428.

Fałda B. (2010) Modelowanie dynamiczne procesów ekonomicznych, Wydawnictwo KUL, Lublin, str. 15-16, 126.

Geweke J. F., Horowitz J. L., Pesaran M. H. (2006) Econometrics: A Bird's Eye View, IZA, Discussion Paper Series No. 2458, str. 3-4.

Grudzewski W. M., Rosłanowska-Plichcińska K. (1984) Mierzenie wielkości i wymiarowe modelowanie zjawisk oraz procesów ekonomicznych, Zakład Narodowy im. Ossolińskich, Wydawnictwo Polskiej Akademii Nauk, Wrocław - Warszawa - ... - Łódź, str. 9.

Gruszecki T. (2001) Nagrody Nobla w ekonomii, Verba, Lublin, str. 48-49.

Partyka D., Zając J. (2010) Generalized problem of regression, Bulletin De La Societe Des Sciences Et Des Lettres De Łódź, vol. LX, no. 1, str. 78-80, 86-87.

Sen A. K., Srivastava M. S. (1990) Regression analysis: theory, methods and applications, Springer-Verlag, New York, str. 1.

Stankiewicz W. (2000) Historia myśli ekonomicznej, Państwowe Wydawnictwo Ekonomiczne, Warszawa, str. 401.

Zając J. (2010) Regresja uogólniona, „Miscellanea mikroekonometrii", Uniwersytet Szczeciński, Szczecin, str. 170-171.

www.money.pl [dostęp: luty 2012]

THE REGRESSION PROBLEM IN ECONOMIC SCIENCES

Abstract: The regression theory involves methods and tools of exact description of relations between various types of phenomena. Since many years, it is used for the economic and econometric models formulation, however, as the analyze of literature indicates, regression is understood in different ways. This article is a reflection on the application of quantitative methods in economics processes modeling, with particular emphasis on the relations between them, expressed by regression.

Key words: regression theory, generalized regression