

Joanna Kisielińska

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

WYKORZYSTANIE MODELI REGRESYJNYCH W PROBLEMACH Z ZAKRESU EKONOMIKI ROLNICTWA

REGRESSION MODELS IN AGRICULTURAL ECONOMICS

Słowa kluczowe: modele regresyjne, sieci neuronowe

Key words: regression model, neural networks

Synopsis. Dokonano przeglądu modeli regresyjnych opracowanych dla problemów z zakresu ekonomiki rolnictwa. Okazało się, że autorzy stosują jedynie modele liniowe i linearyzowalne. Zwrócono uwagę na możliwość zastosowania nowoczesnych narzędzi jakimi są sieci neuronowe.

Wstęp

Prawidłowo opracowane modele przyczynowo-skutkowe pozwalają opisać złożone zjawiska ekonomiczne. Daje to możliwość nie tylko lepszego zrozumienia badanego problemu, ale również przewidywania jego dalszego rozwoju. Uzyskane informacje mogą stanowić wskazówki do ewentualnego ingerowania w proces ekonomiczny w celu uzyskania pożądanych rezultatów.

Modele przyczynowo-skutkowe tworzone są głównie przy założonej z góry funkcyjnej postaci związku (najczęściej liniowej albo linearyzowalnej). Jest to ograniczenie istotne, ponieważ zwykle nie wiadomo jak kształtują się związki między zmiennymi.

Problem poszukiwania modelu w ogólnym przypadku polega na znalezieniu funkcji najlepiej¹ odwzorowującej zależności zachodzące między zmiennymi. Założenie określonej postaci funkcji sprowadza zadanie do wyznaczenia jej współczynników. Jeżeli dodatkowo przyjęta zostanie postać liniowa (lub nieliniowa, ale linearyzowalna) można w tym celu wykorzystać znane wzory macierzowe². Należy jednak pamiętać, że jest to istotne uproszczenie sformułowanego zadania. Założenie niewłaściwej postaci analitycznej może prowadzić do istotnych błędów. Z drugiej strony, brak jest uniwersalnej procedury pozwalającej na wyznaczenie funkcji najlepszej z punktu widzenia przyjętego kryterium. Problem można obejść stosując sieci neuronowe, które umożliwiają modelowanie dowolnych zależności.

Ponieważ w większości książek poświęconych ekonometrii problem poszukiwania zależności między zmiennymi sprowadzany jest do wyznaczania współczynników z góry założonej funkcji, autorka chciała zwrócić uwagę na ograniczenia wprowadzane przez taką procedurę. Przedstawiony w artykule przegląd literatury potwierdza potrzebę precyzyjnego sformułowania problemu poszukiwania modeli związków między zmiennymi i popularyzacji najbardziej uniwersalnego narzędzia do jego realizacji jakim jest sieć neuronowa.

Podstawy teoretyczne analizy regresji

Analiza regresji wielorakiej należy do metod wielowymiarowej analizy danych, która zajmuje się badaniem obiektów opisanych wieloma cechami. Informacje o obiektach umieszczane są w macierzy zwanej macierzą obserwacji, o liczbie wierszy równej liczbie obiektów i liczbie kolumn równej liczbie cech. W analizie regresji przyjmuje się, że zmienność jednej z cech (zwanej zmienną zależną) może być wyjaśniona przez cechy inne (zwane zmiennymi niezależnymi).

Najbardziej ogólnym modelem regresji jest związek funkcyjny między zmienną zależną y i wektorem zmiennych niezależnych x . Ponieważ przewidywanie wartości zmiennej y na podstawie zna-

¹ Jeżeli stosowana jest metoda najmniejszych kwadratów chcemy zminimalizować sumę kwadratów odchyłań.

² Wzór (4) w dalszej części artykułu.

jomości x obarczone jest zwykle pewnym błędem, do równania wprowadza się składnik losowy ε . Model ma więc postać:

$$y = g(x) + \varepsilon \tag{1}$$

gdzie: $g(x)$ jest funkcją o nieznannej postaci.

Zwykle poszukiwania właściwego modelu zależności ogranicza się do konkretnej postaci funkcji, która może być w szczególności funkcją nieliniową. Mówimy wówczas o zadaniu regresji nieliniowej o następującej postaci.

$$y = f(x, \beta) + \varepsilon \tag{2}$$

gdzie: β jest wektorem parametrów, zaś $f(x, \beta)$ jest wielowymiarową funkcją o założonej postaci.

Zadanie polega na właściwym doborze wektora parametrów β . Problem ten zasadniczo różni się od zadania określonego regułą (1), gdzie postać funkcyjna zależności nie była znana.

Jeżeli w formule (2) założymy, że funkcja f jest liniowa, problem nazywamy regresją liniową. Zmienna zależna jest wówczas określona jako kombinacja liniowa zmiennych niezależnych z dokładnością do składnika losowego ε . Model jest wówczas następujący:

$$y = \beta^T \cdot x + \varepsilon \tag{3}$$

Pod względem obliczeniowym, jak i teoretycznym (weryfikacji statystycznej) najlepiej rozpracowany jest problem liniowy. Wektor parametrów β w literaturze ekonometrycznej wyznaczany jest metodą najmniejszych kwadratów (MNK). W algebrze natomiast stosowana jest metoda pseudoinwersji umożliwiająca rozwiązanie nadokreślonych układów równań³. Obydwie metody dają tę samą formułę macierzową:

$$b = (X^T \cdot X)^{-1} \cdot X^T \cdot y \tag{4}$$

gdzie: wektor b – wartość jest oszacowań wektora poszukiwanych parametrów β , wektor y – zaobserwowane wartości zmiennej y , natomiast w macierzy X umieszczone są obserwacje zmiennych niezależnych. Jeżeli w formule (3) wprowadzamy wyraz wolny, macierz X należy uzupełnić o kolumnę zawierającą same jedynki.

Zadania regresji nieliniowej (2) dzieli się na dwie grupy. Pierwszą stanowią zadania, które można sprowadzić do problemu liniowego, przez transformacje zmiennych. Grupa druga obejmuje zadania ściśle nieliniowe, czyli niesprowadzalne do postaci liniowej. Zadanie w tym przypadku polega na oszacowaniu wektora parametrów β , aby zminimalizować sumę kwadratów błędów (MNK). Tak sformułowany problem jest typowym zadaniem optymalizacji nieliniowej bez ograniczeń, który da się rozwiązać przy pomocy odpowiednich algorytmów najszerzej opisanych w pracy Findeisen i in. [1977]. W Excelu można użyć narzędzia nazwanego przez autorów opracowania Solverem⁴. Niestety, przy problemach o dużych wymiarach algorytmy optymalizacji nieliniowej stają się często zawodne.

Zadanie regresji sformułowane w sposób najbardziej ogólny polega na dobraniu funkcji, która zminimalizuje sumę kwadratów błędów (MNK). Niestety dotychczas nie opracowano metod, które pozwoliłyby tak sformułowany problem rozwiązać. Do zadania można natomiast podejść w sposób nieco odmienny. Zamiast poszukiwać funkcji o nieznannej postaci można opracować model reprezentujący dowolną zależność. Zadanie takie pozwala zrealizować sieć neuronowa, w postaci perceptronu, czy sieci o radialnej funkcji bazowej. Udowodniono bowiem, że sieci te są aproksymatorami uniwersalnymi – umożliwiającymi odwzorowanie dowolnej funkcji⁵.

W przypadku sieci perceptronowej zależność między wektorem zmiennych niezależnych x i zmienną y jest następująca:

$$y = \sum_{j=1}^p \left[s_j \cdot \varphi \left(\sum_{i=1}^q (w_i^T \cdot x + w_{0i}) \right) \right] \tag{5}$$

gdzie: w_i^T , w_{0i} oraz s_j są wagami połączeń, zaś φ funkcją aktywacji (najczęściej tangens hiperboliczny).

³ Opis metody znaleźć można np. w pracy Fortuny i in. [1998]

⁴ Opis Solvera zawiera np. praca Kisielińska, Skórnik-Pokarowska [2005]

⁵ Opis sieci neuronowych i wykaz literatury im poświęcony zawiera praca Kisielińskiej [2008].

Działanie sieci o radialnej funkcji bazowej natomiast określa formuła:

$$y = \sum_{j=1}^p s_j \cdot \varphi(\|\mathbf{x} - \mathbf{c}_j\|) \quad (6)$$

gdzie: c_j są wektorami centrów, s_j wagami połączeń, zaś φ funkcją aktywacji (najczęściej funkcja Gaussa).

Zaletą sieci neuronowych, poza uniwersalnością, jest umiejętność doboru zmiennych niezależnych nawet gdy związki są nieliniowe, co pokazała w pracy Kisielińska [2003a]. Jeżeli potrzebna jest postać funkcyjna, zastosowanie sieci pozwala zidentyfikować nie tylko czynniki istotne, ale również przybliżyć postać zależności (co znakomicie ułatwia dobór właściwej funkcji).

Niestety, narzędzie to rzadko stosowane jest w badaniach z zakresu ekonomiki rolnictwa (co omówiono dalej) mimo, że większość pakietów statystycznych zawiera sieci neuronowe. Oprogramowanie jest przyjazne dla użytkownika i nie wymaga znajomości skomplikowanych być może zasad konstrukcji i trenowania sieci.

Wybrane przykłady zastosowań analizy regresji

Przykłady zastosowania analizy regresji w zagadnieniach z zakresu ekonomiki rolnictwa przedstawiono w tabeli 1. Przeglądu dokonano biorąc pod uwagę lata ostatnie i najbardziej popularne czasopisma i wydawnictwa związane z rozważaną tematyką. Były to pozycje następujące: *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, *Roczniki Nauk Rolniczych PAN*, *Zagadnienia Ekonomiki Rolnej*, *Więś i Rolnictwo*, *Przegląd Statystyczny* oraz *Zeszyty Naukowe SGGW*. Okazało się, że jedynie trzy pierwsze zawierały publikacje opisujące badania wykorzystujące regresję w ekonomice rolnictwa.

Opracowane modele cechowały w większości przypadków wysokie wartości współczynników determinacji, wskazujące na dobrą jakość modeli. W kilku przypadkach otrzymano wartość nawet ponad 0,90 (w jednym przypadku nawet 1), co oznacza, że model wyjaśnia zmienność cechy w ponad 90%. Jedynie w nielicznych modelach współczynnik determinacji był poniżej 0,50. Rezultat taki oznacza, że ponad 50% zmienności zjawiska kształtują czynniki nieuwzględnione. Wskazania modelu należy wówczas traktować jedynie orientacyjnie, natomiast w celu pełniejszego obrazu problemu konieczne jest podniesienie jego jakości. Jeżeli wykluczona zostanie możliwość niewłaściwego doboru postaci analitycznej, należy rozszerzyć listę zmiennych niezależnych.

W większości badań modele wykorzystywano do określania czynników kształtujących badane zjawisko. Analiza regresji, zwłaszcza w wersji krokowej, pozwala dokonać eliminacji cech słabo przyczyniających się do określania poziomu zmiennej zależnej. Warto w takich przypadkach rozszerzyć listę zmiennych o czynniki dodatkowe, przeprowadzając ewentualne redukcje cech na etapie opracowywania modelu.

W pracach przedstawionych w tabeli 1 liczba cech uwzględniona w modelach wynosiła średnio 6,5. Pięć i mniej cech wykorzystano w 10 spośród nich. Warto w tym miejscu zwrócić uwagę, że zjawiska ekonomiczne są procesami złożonymi. Kształtuje je zwykle wiele, często różnorodnych czynników. Wzięcie pod uwagę niewielkiej ich liczby jest właściwe, jeśli badacz ma pewność, że inne elementy nie mają znaczenia. Budowa modelu na podstawie określonego zestawu zmiennych niezależnych oznacza bowiem założenie, że to ich wartości determinują poziomy zmiennej zależnej. Warto wobec tego dokonać wstępnej merytorycznej analizy, uzasadniając pominięcie czynników, które przynajmniej potencjalnie mogą mieć pewne znaczenie.

Jeżeli przyjrzymy się rodzajom zmiennych zależnych w modelach, stwierdzimy, że analiza regresji pozwala wyjaśniać różnorodne zjawiska z zakresu ekonomiki rolnictwa. Najwięcej modeli dotyczyło kształtowania się kosztów oraz nakładów w produkcji rolniczej (4 publikacje w tabeli 1) i dochodu rolniczego (3), a następnie wartości dodanej w rolnictwie (2), wielkości produkcji (2) i wskaźników finansowych (2). W pojedynczych pracach zmienną zależną były: wartość użytków rolnych, wydatki na spożycie żywności w gospodarstwach domowych i wartość użytków rolnych.

Biorąc pod uwagę rodzaj produkcji, najwięcej prac związanych było z produkcją rolniczą łącznie (5) oraz mleczarstwem (3). Po jednej publikacji poświęcono produkcji drobiu, pszenicy, produkcji żywa wieprzowego i uprawom szklarniowym.

Modele opracowywano przede wszystkim dla gospodarstw rolniczych (7), a następnie dla województw (2), spółdzielni (2) i przedsiębiorstw spożywczych (2). Dla gmin i gospodarstw domowych zbudowano po jednym modelu.

Tabela 1. Charakterystyka problemów rozwiązywanych przy użyciu analizy regresji

Autor	Rok publ.	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda	Współczynnik determinacji
Popielek	1999	czynniki kształtujące rozwój i efektywność produkcji drobiarskiej	43 rolnacje brojlerów w 3 farmach	nakłady energetyczne na 1 kg masy ciała	5 cech	regresja liniowa	-
Heller, Sass	1999	ocena potencjału ekonomicznego obszarów wiejskich w ujęciu przestrzennym	49 województw	wartość dodana rolnictwa na 1 ha. użytków rolnych	10 cech charakterystycznych rolnictwa w województwie	regresja liniowa i funkcja potęgowa	0,48; 0,53
Gołbiewska, Klepacki	2000	czynniki kształtujące dochód rolnicy w gospodarstwach o zróżnicowanym poziomie towarowości	1055 gospodarstw (IERiGZ)	dochód rolniczy	8, 5 i 6 cech	regresja liniowa	0,73; 0,74; 0,41
Szuk	2001	czynniki warunkujące nakłady pracy przy produkcji pszenicy ozimej	120 gospodarstw	nakłady pracy ludzkiej i mechanicznej	5 cech	regresja liniowa	0,14-0,50
Kowalczyk	2002	model wartości rynkowej gruntów rolniczych	375 gospodarstw	wartość użytków rolnych	7 cech	funkcja potęgowa	0,72
Kisielńska	2003	ocena dochodu rolniczego na podstawie analizy finansowej	1062 gospodarstw (IERiGZ)	dochód rolniczy	12 wskaźników finansowych	regresja liniowa i sieć neuronowa	0,66; 0,72
Pietrzak	2004	efektywność ekonomiczna spółdzielni mleczarskich	69 spółdzielni mleczarskich	wartość dodatkowa ze sprzedaży na litr skupu mleka	3 cechy	regresja liniowa	0,58
Plonka, Sobczyński	2004	ocena zdolności kredytowej gospodarstw rolniczych	365 gospodarstw	ryzyko utraty zdolności płatniczej	13 wskaźników finansowych	funkcja kwadratowa	0,78
Stachowiak	2004	wpływ rodzajów i technik wykonywanych zabiegów ochrony roślin na koszty całkowite ochrony roślin	dane z 59, 26 i 26 bloków szklarniowych	koszty całkowite ochrony	10 cech	regresja liniowa	1,00
Majewski	2005	regionalne zróżnicowanie skupu mleka w Polsce oraz czynniki je determinujące	16 województw	wielkość skupu mleka na 1 hektar UR	5 cech	regresja liniowa	0,92
Pawłowska, Piereligin	2005	analiza dochodów przedsiębiorstw przemysłu piekarniczego	12 przedsiębiorstw przemysłu piekarskiego w Zylomierzu	zysowność	4 cechy	regresja liniowa	0,67-0,96
Czekaj	2006	analiza dochodowości materialnych czynników produkcji w gospodarstwach rolnych	403 gospodarstw (IERiGZ)	dochód rolniczy	3 cechy-praca, ziemia, kapitał obrotowy	regresja liniowa i funkcja potęgowa	0,65; 0,68
Stępień	2006	oszacowanie kosztów jednostkowych w gospodarstwach o zróżnicowanej strukturze organizacyjnej	748 gospodarstw z Wielkopolski	koszty jednostkowe bezpośrednie, pośrednie i całkowite na 1 kg żywności wieprzowego	4 cechy	regresja liniowa	0,37; 0,52; 0,68
Wysocki, Kurzawa	2006	kształtowanie się preferencji konsumpcyjnych artykułów żywnościowych w relacji miasto-wieś	32 452 gospodarstw domowych	wydatki na spożycie	2 cechy ciągle i 4 binarne	funkcja potęgowa-wykładnicza	0,79-0,92

c.d. tabeli 1. Charakterystyka problemów rozwiązywanych przy użyciu analizy regresji

Autor	Rok publ.	Problem	Obiekty	Zmienna zależna	Zmienne niezależne	Metoda	Współczynnik determinacji
Pietrzak	2006	korzyści skali w przemyśle mleczarskim	170 spółdzielni mleczarskich	wielkość produkcji	liczba pracownik ^{ów} i aktywa trwałe netto	funkcja potęgowa	0,50
Mańko, Sass, Sobczyński	2007	czynniki kształtujące aktywność inwestycyjną rolników w podregionie bydgoskim	53 gminy podregionu bydgoskiego	liczba inwestor ^{ów} w gminie	11 cech	regresja liniowa i wielomianowa	0,68-0,74
Bieniasz, Czerwińska-Kayzer, Golaś	2007	płynność finansowa przedsiębiorstw branży spożywczej	22 spółki branży spożywczej notowane na GPW	liczba inwestor ^{ów} , szybkość bieżąca, szybkość natychmiastowa	5 cech	regresja liniowa	0,66

Źródło: opracowanie własne.

Najczęściej stosowana jest regresja liniowa (13 modeli), ale znaleźć można też kilka przykładów wykorzystania regresji nieliniowej. Modele potęgowe i potęgowo wykładnicze użyto w 5 pracach, zaś wielomiany w dwóch. Sieci neuronowe stosowano bardzo rzadko – jedynie w jednym przypadku. Wprawdzie wadą sieci jest brak funkcyjnej postaci zależności, jednakże jej uniwersalność w pełni mankament ten rekompensuje. Warto tu nadmienić, że sieci neuronowe są idealnym narzędziem przynajmniej wstępnego szacowania modelu (np. w zakresie doboru cech), co pokazano w pracy Kisielińska [2003a]. Informacje tam uzyskane dostarczają cennych wskazówek co do rodzaju zależności między zmiennymi ułatwiając dobór właściwej funkcji.

Podsumowanie

Modele regresyjne stają się docenianym przez ekonomistów zajmujących się problemami rolnictwa narzędziem analizy. Przegląd publikacji dokonany w niniejszym artykule potwierdza możliwość ich stosowania do wyjaśniania różnorodnych zjawisk w tej dziedzinie. Wydaje się jednak, że zbyt małą uwagę poświęca się właściwemu doborowi cech uwzględnianych w modelach oraz ich postaci funkcyjnej. Artykuł niniejszy stanowi próbę zachęcenia badaczy do ogólniejszego spojrzenia na formułowane problemy i stosowania do ich rozwiązywania bardziej uniwersalnych narzędzi w postaci sieci neuronowych.

Literatura

- Bieniasz A., Czerwińska-Kayzer D., Golaś Z.** 2007: Czynniki kształtujące płynność finansową przedsiębiorstw branży spożywczej. *Zagadnienia Ekonomiki Rolnej*, nr 4, 62-67.
- Czekaj T.** 2006: Obserwacje odstające i wpływowe w analizie regresji – analiza dochodowości materialnych czynników produkcji w gospodarstwach rolnych. *Roczniki Naukowe SERiA*, t. VII, z. 5, 12-15.
- Findeisen W., Szymanowski J., Wierzbicki A.** 1977: Teoria i metody obliczeniowe optymalizacji. PWN, Warszawa.
- Fortuna Z., Macukow B., Wąsowski J.** 1998: Metody numeryczne. Wyd. Naukowo-Techniczne, Warszawa.
- Gołębiowska B., Klepacki B.** 2000: Czynniki kształtujące dochód rolniczy w gospodarstwach o zróżnicowanym poziomie towarowości. *Roczniki Nauk Rolniczych, seria G*, t. 88, z. 2, 59-68.
- Heller J., Sass R.** 1999: Syntetyczna metoda oceny potencjału ekonomicznego obszarów wiejskich w ujęciu przestrzennym. *Zagadnienia Ekonomiki Rolnej* 1,10-22.
- Kisielińska J.** 2003a: Sieci neuronowe i modele regresyjne w analizach zależności przyczynowo-skutkowych. *Wiadomości Statystyczne*, nr 3, 79-89.
- Kisielińska J.** 2003b: Ocena dochodu rolniczego na podstawie analizy finansowej, przy użyciu sieci neuronowej i analizy regresji. *Roczniki Nauk Rolniczych, seria G*, t. 90, z. 1,114-124.
- Kisielińska J., Skórnik-Pokarowska U.** 2005: Podstawy statystyki z przykładami w Excelu. Wyd. SGGW, Warszawa.
- Kisielińskiej J.** 2008: Modele klasyfikacyjne – prognozowanie sytuacji finansowej gospodarstw rolniczych. Wyd. SGGW, Warszawa.
- Kowalczyk S.** 2002: Model wartości rynkowej gruntów rolniczych. *Roczniki Nauk Rolniczych, seria G*, t. 89, z. 1, 43-52.
- Majewski J.** 2005: Regionalne zróżnicowanie skupu mleka w Polsce oraz czynniki je determinujące. *Roczniki Naukowe SERiA*, t. VII, z. 5, 56-60.
- Mańko S., Sass R., Sobczyński T.** 2007: Czynniki kształtujące aktywność inwestycyjną rolników w podregionie bydgoskim. *Roczniki Naukowe SERiA*, t. IX, z. 2, 233-237.
- Pawlowska L., Piereligin M.** 2005: Czynniki ekonometryczna analiza dochodów przedsiębiorstw przemysłu piekarniczego. *Roczniki Naukowe SERiA*, t. VII, z. 7B, 13-17.
- Pietrzak M.** 2004: Efektywność ekonomiczna spółdzielni mleczarskich – koncepcja pomiaru oraz czynniki wzrostu w świetle badań empirycznych. *Zagadnienia Ekonomiki Rolnej*, 3.

- Pietrzak M.** 2006: Efektywność finansowa spółdzielni mleczarskich – koncepcja oceny. Wyd. SGGW, Warszawa.
- Płonka R., Sobczyński T.** 2004: Przydatność wybranych wskaźników finansowych w ocenie zdolności kredytowej gospodarstw rolniczych. *Roczniki Naukowe SERiA*, t. VI, z. 5, 76-81.
- Popiołek W.** 1999: Czynniki kształtujące rozwój i efektywność produkcji drobiarskiej. *Roczniki Naukowe SERiA*, t. I, z. 2.
- Stachowiak M.** 2004: Wpływ rodzajów i technik wykonywanych zabiegów ochrony roślin na koszty całkowite ochrony na przykładzie modeli kosztów ochrony pomidorów szklarniowych oraz ogórków szklarniowych wczesnych i późnych. *Roczniki Naukowe SERiA*, t. VI, z. 1, 199-205.
- Stępień S.** 2006: Koszty jednostkowe w gospodarstwach o zróżnicowanej strukturze organizacyjnej. *Roczniki Naukowe SERiA*, t. VIII, z. 1, 191-193.
- Szuk T.** 2001: Czynniki warunkujące nakłady pracy przy produkcji pszenicy ozimej. *Roczniki Naukowe SERiA*, t. III, z. 5.
- Wysocki F., Kurzawa I.** 2006: Kształtowanie się preferencji konsumpcyjnych artykułów żywnościowych w relacji miasto – wieś. *Zagadnienia Ekonomiki Rolnej*, nr 2, 47-67.

Summary

This paper reviews the regression models developed for issues of the economics of agriculture. It turned out that the authors use only the linear models and linearysing. Attention was drawn to the possibility of using modern tools to which neural networks are.

Adres do korespondencji:

dr hab. Joanna Kisielińska
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Wydział Nauk Ekonomicznych
ul. Nowoursynowska 166
02-787 Warszawa
tel. 691 712 782
e-mail: joanna_kisielinska@sggw.pl