

AGNIESZKA SOMPOLSKA-RZECHUŁA
MICHAŁ ŚWITŁYK
Zachodniopomorski Uniwersytet Technologiczny
Szczecin

10.5604/00441600.1225666
Tłumaczenie

CZYNNIKI WPLYWAJĄCE NA PRAWDOPODOBIENSTWO POPRAWY PRZYCHODÓW GOSPODARSTW ROLNYCH SPECJALIZUJĄCYCH SIĘ W PRODUKCJI MLEKA

Abstrakt

W pracy wykorzystano model regresji logistycznej do określenia czynników wpływających na prawdopodobieństwo poprawy przychodów gospodarstw rolnych specjalizujących się w produkcji mleka. Badanie miało charakter dynamiczny i obejmowało lata 2009-2011. Wśród zmiennych niezależnych uwzględniono zmienne ilościowe i jakościowe związane z działalnością gospodarstw rolnych. Najbardziej istotne w modelu logitowym, w każdym roku, okazały się zmienne dotyczące: powierzchni użytków rolnych i liczby krów mlecznych, mające dodatni wpływ na prawdopodobieństwo poprawy przychodu. Wszystkie oszacowane modele odznaczają się bardzo wysoką jakością, a tym samym zdolnością do poprawnej klasyfikacji gospodarstw.

Słowa kluczowe: przychody, gospodarstwa rolne, produkcja mleka, model logitowy.

O znaczeniu produkcji mleka w Polsce świadczą dane powszechnego spisu rolnego – chów bydła prowadzony w 454 tys. gospodarstw rolnych dawał utrzymanie około 2 mln mieszkańców wsi (Ziętara, 2013).

Udział produkcji zwierzęcej w strukturze produkcji globalnej rolnictwa w latach 2009-2011 zmalał z 47,7% (2009) do 43,7% (2011). Produkcja mleka, żywności wołowego i cieląt w badanych latach stanowiła około 20% produkcji globalnej. Udział produkcji zwierzęcej w strukturze produkcji towarowej rolnictwa charakteryzował się podobnymi tendencjami: z 56,5% w 2009 r. zmniejszył się do 53,4% w 2011 r. Stabilny był natomiast udział w produkcji towarowej bydła, cieląt i mleka w analizowanym okresie (23-24%, w tym mleka 17-18%).

Niniejsze opracowanie ma na celu identyfikację czynników poprawy przychodu gospodarstw rolnych specjalizujących się w produkcji mleka. Badanie przeprowadzono na podstawie informacji pochodzących z Systemu Zbierania i Wykorzystywania Danych Rachunkowych z Gospodarstw Rolnych (Polski FADN). Badana zbiorowość liczyła 670 gospodarstw rolnych, a okres badawczy obejmował lata 2009-2011. Dane dotyczyły gospodarstw typu 45, czyli gospodarstw specjalizujących się w chowie bydła mlecznego. W związku z faktem, iż stopień specjalizacji w badanych gospodarstwach był różny, obliczono dla nich procentowy udział wartości sprzedaży mleka i przetworów w wartości produkcji ogółem. Obliczonym współczynnikiem skorygowano koszty, zakładając ich proporcjonalny rozkład na poszczególne działy i gałęzie.

Wśród zmiennych dotyczących działalności gospodarstw rolnych uwzględniono zarówno zmienne ilościowe, jak i jakościowe, a do realizacji celu wykorzystano model logitowy, za pomocą którego oszacowano prawdopodobieństwo poprawy przychodu w gospodarstwach rolnych. Regresja logistyczna cieszy się coraz większym uznaniem w wielu dziedzinach nauki, takich jak: medycyna, psychologia, nauki techniczne, bankowość, ubezpieczenia, demografia oraz ekonomia. Przykładami prac z zastosowaniem modelu logitowego są: (Jackowska i Wycinka, 2011; Kmiec, 2015; Kowerski, Bielak i Długosz, 2006; Kasprzyk i Fura, 2011).

Metoda badawcza

Model regresji logistycznej (podobnie jak liniowej regresji wielorakiej) pozwala na zbadanie wpływu wielu zmiennych niezależnych X_1, \dots, X_k na zmienną zależną Y . Zmienna zależna przyjmuje jedynie dwie wartości i ma charakter dychotomiczny. Te dwie wartości kodowane są jako 1 i 0, gdzie: wartość 1 oznacza posiadanie danej cechy, a 0 – brak danej cechy (Hosmer i Lemenshow, 2000). Do opisanie takiej zależności w regresji logitowej wykorzystuje się funkcję logistyczną (Stanisz 2007):

$$f(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}, \quad z \in R. \quad (1)$$

Funkcja ta przyjmuje wartości z przedziału (0;1), a jej wykres przypomina rozciągniętą literę S. Można wyróżnić trzy etapy zmiany wartości funkcji: początkowo do pewnej wartości progowej praktycznie nie zmieniają prawdopodobieństwa, po osiągnięciu wartości progowej gwałtownie wzrasta prawdopodobieństwo do jeden i utrzymuje się na tym poziomie. Taka funkcja znajduje liczne zastosowania do opisu zjawisk w medycynie, epidemiologii, psychologii czy ekonomii, np. ryzyko zachorowania na pewną chorobę, szansa na wyzdrowienie, zdolność do znalezienia pracy itp.

Model logistyczny jest bardzo dobrym narzędziem w przypadku rozpatrywania prawdopodobieństwa zaistnienia pewnego zjawiska. Pozwala przedstawić, w jaki sposób prawdopodobieństwo wystąpienia badanego zdarzenia zależy od pewnych zmiennych, które mogą być zarówno ilościowe, jak i jakościowe. Zaletą modelowania logistycznego jest otrzymanie wzoru matematycznego, za pomocą którego można określić, z jaką siłą i kierunkiem poszczególne zmienne wpływają na modelowane zjawisko. Ponadto model regresji logistycznej nie wymaga niektórych założeń koniecznych dla regresji liniowej. Wektor zmiennych objaśniających i reszty nie muszą mieć rozkładu normalnego. Dodatkową zaletą regresji logistycznej jest to, że analiza i interpretacja wyników są podobne jak w metodach klasycznej regresji.

Logistyczny model regresji dla zmiennej dychotomicznej Y określony jest następującą zależnością (Maddala, 2008; Stanisiz, 2007):

$$P(Y = 1 / X_1, \dots, X_k) = \frac{e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}}{1 + e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}} \quad (2)$$

gdzie:

$\alpha_0, \alpha_1, \dots, \alpha_k$, są współczynnikami regresji logistycznej, X_1, \dots, X_k – zmiennymi niezależnymi, które mogą mieć charakter zarówno jakościowy, jak i ilościowy.

Dokonując transformacji modelu logistycznego (2) za pomocą logarytmowania, otrzymuje się iloraz szans, przez który należy rozumieć stosunek prawdopodobieństwa wystąpienia określonego zdarzenia do prawdopodobieństwa, że taki przypadek się nie pojawi. Logarytm naturalny szansy zwany jest logitem (Stanisiz, 2007; Cramer, 2003; Kleinbaum i Klein, 2002):

$$\text{logit}P = \ln \frac{P(Y = 1 / X_1, \dots, X_k)}{1 - P(Y = 1 / X_1, \dots, X_k)} = \alpha_0 + \sum_{i=1}^k \alpha_i X_i \quad (3)$$

Równość ta jest logitową postacią modelu logistycznego. W modelu logitowym logarytm szansy wystąpienia zdarzenia jest liniową funkcją zmiennych objaśniających.

Po oszacowaniu parametrów modelu regresji logistycznej można wyznaczyć wartości teoretyczne zmiennej Y według standardowej zasady prognozy:

$$\hat{y}_i = \begin{cases} 1, & \text{gdy } 0,5 < \hat{p}_i \leq 1 \\ 0, & \text{gdy } 0 < \hat{p}_i \leq 0,5 \end{cases} \quad (4)$$

gdzie:

\hat{p}_i – prawdopodobieństwa teoretyczne uzyskane z oszacowanego na podstawie próby losowej modelu regresji logistycznej.

W sytuacji, gdy próba jest niezbilansowana, czyli liczba jedynek znacznie różni się od liczby zer, do prognozowania wartości teoretycznych można zastosować modyfikację standardowej zasady i liczyć prognozy według zasady optymalnej wartości granicznej α :

$$\hat{y}_i = \begin{cases} 1, & \text{gdy } \alpha < \hat{p}_i \leq 1 \\ 0, & \text{gdy } 0 < \hat{p}_i \leq \alpha \end{cases} \quad (5)$$

Wartość graniczną α ustala się jako udział jedynek w próbie. Można wówczas dokonać oceny poprawności oszacowanego modelu, zliczając trafnie oraz błędnie zaklasyfikowane przypadki (tab. 1).

Tabela 1

Macierz klasyfikacji przypadków

Zakwalifikowanie obiektów na podstawie modelu logitowego	Rzeczywista przynależność obiektów		Suma
	$\hat{y}_i = 1$	$\hat{y}_i = 0$	
$\hat{y}_i = 1$	n_{11}	n_{12}	$n_{\cdot 1}$
$\hat{y}_i = 0$	n_{21}	n_{22}	$n_{\cdot 2}$
Suma	$n_{\cdot 1}$	$n_{\cdot 2}$	n

Źródło: Opracowanie własne na podstawie (Dobosz, 2004).

Do oceny stopnia dopasowania modelu regresji logistycznej do danych empirycznych można wykorzystać miarę *zliczeniowy* R^2 , która przyjmuje wartości z przedziału $\langle 0, 1 \rangle$ zdefiniowaną następująco (Maddala, 2008):

$$\text{zliczeniowy}R^2 = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (6)$$

Im bliższa jeden wartość tej miary, tym lepsze dopasowanie modelu logistycznego do danych empirycznych badanego zjawiska, *zliczeniowy* R^2 oznacza procent trafnie zaklasyfikowanych przypadków. Model dobrze sprawdza się w prognozowaniu badanego zjawiska, kiedy *zliczeniowy* $R^2 > 50\%$. Oznacza to, że klasyfikacja na podstawie modelu jest lepsza od przypadkowej. Inne miary dopasowania można znaleźć w pracy (Sompolska-Rzechuła, Machowska-Szewczyk i in., 2014).

Innym sposobem oceny jakości modelu regresji logistycznej jest test Hosmera–Lemeshowa (Hosmer i in., 1989; Homer i in., 2008), który dla różnych podgrup danych porównuje obserwowane licznosci występowania w danej podgrupie obiektów posiadających wyróżnioną cechę O_g i oczekiwane licznosci E_g występowania wartości wyróżnionej. Jeśli O_g i E_g są wystarczająco bliskie, wówczas można założyć, że zbudowano dobrze dopasowany model. Zwykle do ob-

liczeń dzieli się obserwacje na G podgrup, wykorzystując np. decyle. Hipotezy w teście przyjmują następującą postać:

$$H_0 : O_g = E_g \text{ dla wszystkich kategorii,}$$

$$H_1 : O_g \neq E_g \text{ przynajmniej dla jednej kategorii.}$$

Wartość statystyki testowej jest wyznaczana następująco:

$$H = \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g \left(1 - \frac{E_g}{N_g}\right)}$$

gdzie:

N_g – liczba obserwacji w grupie g , $g \in \{1, \dots, G\}$.

Statystyka ta ma asymptotycznie rozkład χ^2 z $G-2$ stopniami swobody.

Do oceny jakości modelu logitowego, obok różnych miar, wykorzystuje się także krzywą ROC (*Receiver Operating Characteristic Curves*), która budowana jest na podstawie wartości zmiennej zależnej oraz przewidywanego prawdopodobieństwa zmiennej zależnej, pozwalającego na ocenę zdolności zbudowanego modelu regresji logistycznej do klasyfikacji przypadków w dwóch grupach: posiadającej wyróżnioną cechę i nieposiadającej wyróżnionej cechy. Krzywa ROC powstaje przez połączenie punktów o współrzędnych (1-swoistość, czułość). Przez czułość rozumie się zdolność wykrywania jednostek posiadających wyróżnioną cechę:

$$\text{czułość} = \frac{n_{11}}{n_{\bullet 1}}. \quad (7)$$

Natomiast swoistość opisuje zdolność wykrywania jednostek nieposiadających wyróżnionej cechy:

$$\text{swoistość} = \frac{n_{22}}{n_{\bullet 2}}. \quad (8)$$

Powstała w ten sposób krzywa, a w szczególności pole pod nią, obrazuje jakość klasyfikacyjną modelu. Gdy krzywa ROC pokrywa się z przekątną $y = x$, to decyzja o przyporządkowaniu przypadku do wybranej klasy (1) lub (0), podejmowana na podstawie modelu, jest tak samo dobra jak losowy podział badanych przypadków do tych grup. Jakość klasyfikacyjna modelu jest dobra, gdy krzywa znajduje się znacznie powyżej przekątnej $y = x$, czyli gdy pole pod krzywą ROC jest znacznie większe niż 0,5.

Charakterystyka gospodarstw rolnych

Do oceny przychodów osiągniętych przez gospodarstwa rolne przyjęto następujące zmienne:

- Y – przychody ze sprzedaży: mleka, żywca wołowego i dopłaty (tys. zł),
- X_1 – powierzchnia użytków rolnych (ha),
- X_2 – liczba krów mlecznych (LU)¹,
- X_3 – nakłady plonotwórcze, na które składają się: nasiona, nawozy i środki ochrony roślin (zł),
- X_4 – pasze z zakupu (zł),
- X_5 – pozostałe koszty bezpośrednie produkcji zwierzęcej (zł),
- X_6 – koszty utrzymania maszyn i budynków (zł),
- X_7 – koszty energii (zł),
- X_8 – pozostałe koszty ogólnogospodarcze (zł),
- X_9 – koszt amortyzacji (zł),
- X_{10} – posiadanie następcy (-1 – nie ma następcy, 0 – nie dotyczy, 1 – jest następcą),
- X_{11} – wiek (lata),
- X_{12} – płeć (1 – kobieta, 2 – mężczyzna),
- X_{13} – wykształcenie (1 – podstawowe, 2 – zawodowe, 3 – zawodowe rolnicze, 4 – średnie, 5 – średnie rolnicze, 6 – wyższe, 7 – wyższe rolnicze).

W badanym okresie, obejmującym lata 2009-2011, zauważono zwiększenie wartości zmiennych: Y , X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 , X_8 , X_9 . Przychody ze sprzedaży wzrastały średnio z roku na rok o 16,9%. Zwiększyła się także powierzchnia użytków rolnych średnio o 1,69% oraz liczba krów – 1,61%. Wzrosły również nakłady plonotwórcze, średnio o 3,75%. Koszt związany z zakupem pasz co roku zwiększał się średnio o 9,36%. Wśród pozostałych kosztów największy wzrost odnotowano w przypadku kosztów energii (średnio o 16,62%) oraz kosztów utrzymania maszyn i budynków (o 6,37%). W najmniejszym stopniu wzrosły koszty amortyzacji (średnio o 3,54%). W analizowanych latach zmieniła się także średni wiek osoby zarządzającej gospodarstwem – z 44 lat w roku 2009 do 45, 6 lat w roku 2011.

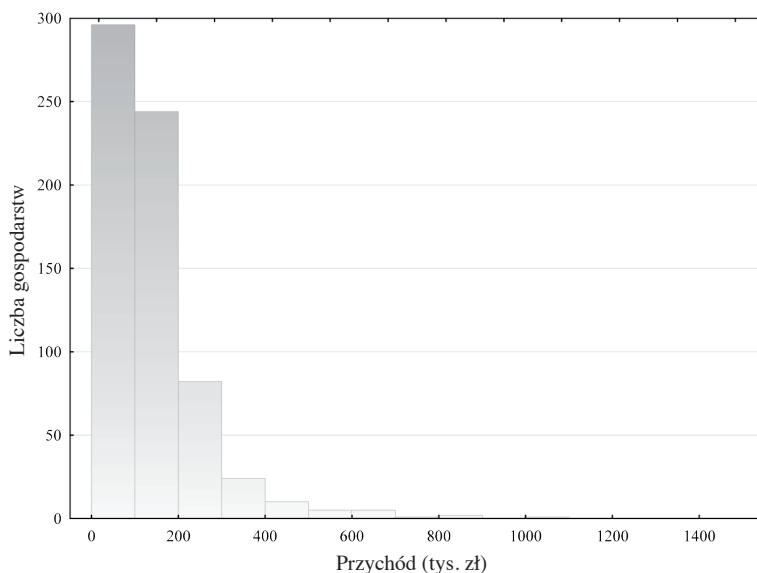
W 2009 roku 18% badanych gospodarstw nie posiadało następcy, w roku 2011 już 21%. Następcę miało 17% gospodarstw w roku 2009, w 2011 roku 20%. O jeden pkt. proc. (z 13% do 14%) zwiększył się odsetek gospodarstw zarządzanych przez kobiety. Nie uległa natomiast zmianie w badanym okresie struktura gospodarstw rolnych pod względem wykształcenia osoby kierującej

¹ LU – jednostka przeliczeniowa zwierząt (ang. *Livestock Unit*).

gospodarstwem. Najwięcej (29%) gospodarstw kierowanych było przez osoby z wykształceniem zawodowym rolniczym, 25% przez osoby o wykształceniu średnim rolniczym, kierownikami 10% gospodarstw były osoby z wykształceniem podstawowym. Tylko w 5% gospodarstw osoba kierująca posiadała wykształcenie wyższe lub wyższe rolnicze.

Zmienna przyjęta jako zależna, określająca wielkość przychodów gospodarstwa w każdym analizowanym roku, charakteryzuje się bardzo silną asymetrią prawostronną (rys. 1-3). Z tego powodu, do określenia gospodarstw o wyższym poziomie przychodów wykorzystano pozycyjną miarę określającą przeciętny poziom badanego zjawiska – medianę, która w przeciwieństwie do średniej jest bardziej odporna na występowanie wartości odstających².

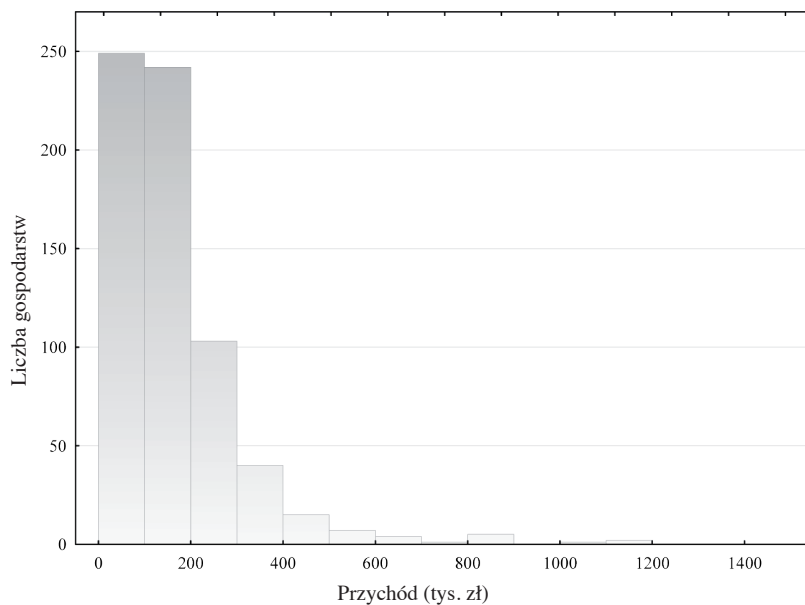
Analizując rozkłady zmiennej zależnej zauważa się bardzo silną asymetrię prawostronną (najsilniejszą w roku 2011), oznaczającą występowanie większej liczby gospodarstw o niższym przychodzie w porównaniu ze średnim. Gospodarstwa są bardzo silnie zróżnicowane pod względem wielkości przychodów, od 82% w roku 2009 do ponad 86% w roku 2011. Mediana przychodów zwiększa się z roku na rok i wynosi: 113,3 tys. zł w roku 2009, 129,2 tys. zł dla roku 2010 i 147,3 tys. zł w roku 2011. Zwiększył się także maksymalny przychód gospodarstwa z 1038 tys. zł w roku 2009 do 1413 tys. zł w roku 2011 (o 36%).



Rys. 1. Rozkład przychodu gospodarstw w roku 2009.

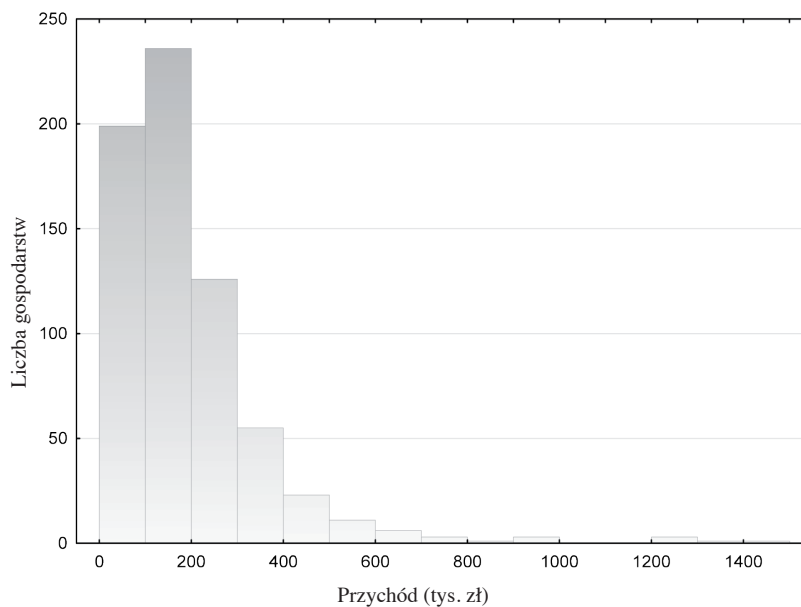
Źródło: Opracowanie własne.

² W badaniach społeczno-ekonomicznych, ze względu na asymetryczny rozkład cech, często wykorzystuje się medianę (Młodak, 2006).



Rys. 2. Rozkład przychodu gospodarstw w roku 2010.

Źródło: Opracowanie własne.



Rys. 3. Rozkład przychodu gospodarstw w roku 2011.

Źródło: Opracowanie własne.

Wyniki modelowania z wykorzystaniem regresji logitowej

Ponieważ wśród zmiennych niezależnych są zmienne zarówno ilościowe, jak i jakościowe, do realizacji celu wykorzystano model regresji logistycznej. Zmienna zależna została określona w sposób binarny jako poziom przychodów równający się co najmniej wartości mediany (wartość 1) oraz przyjmujący wartość poniżej mediany (wartość 0).

W celu znalezienia najlepszej kombinacji zmiennych istotnie wpływających na prawdopodobieństwo zwiększenia przychodu dokonano formalnej selekcji cech za pomocą regresji krokowej i otrzymano następujące zestawy zmiennych:

- dla lat 2009 i 2010: $X_1, X_2, X_3, X_4, X_5, X_9$,
- dla roku 2011: X_1, X_2, X_3, X_4, X_5 .

Otrzymane zbiory tworzą nowe listy zmiennych, które są słabo skorelowane między sobą i równocześnie silnie skorelowane z pozostałymi zmiennymi.

Oceny parametrów modelu logitowego przedstawiono w tabeli 2.

Tabela 2

Oceny parametrów modelu logitowego

	Ocena parametru			p-value			Iloraz szans		
	2009	2010	2011	2009	2010	2011	2009	2010	2011
Stała	-12,3651	-12,4245	-10,3504	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
X_1	0,0872	0,0980	0,0683	0,0000	0,0000	0,0005	1,0911	1,1029	1,0707
X_2	0,2533	0,3322	0,2327	0,0000	0,0000	0,0000	1,2883	1,3940	1,2620
X_3	0,0001	0,0001	0,0001	0,0006	0,0241	0,0003	1,0001	1,0001	1,0001
X_4	0,0001	0,0001	0,0001	0,0000	0,0000	0,0156	1,0001	1,0001	1,0000
X_5	0,0002	0,0002	0,0003	0,0155	0,0142	0,0004	1,0002	1,0002	1,0003
X_9	0,0001		0,0001	0,0009		0,0211	1,0001		1,0001

Źródło: Obliczenia własne.

W modelu dodatni, statystycznie istotny wpływ na zmienną zależną mają wszystkie zmienne niezależne, we wszystkich latach. Najistotniejszy wpływ na wielkość przychodu w gospodarstwach mają dwie zmienne: X_1 – powierzchnia użytków rolnych (ha) oraz X_2 – liczba krów mlecznych.

Interpretując ilorazy szans przy tych zmiennych (zakładając, że pozostałe zmienne uwzględnione w modelu są stałe), uzyskuje się następujące informacje:

- jeżeli powierzchnia użytków rolnych ulegnie zwiększeniu o jeden hektar, to szansa na zwiększenie przychodu gospodarstwa wzrośnie w 2009, 2010 i 2011 roku odpowiednio o: 9,1, 10,3 i 7,1%,
- jeżeli zwiększy się liczba krów mlecznych o jedną, to szansa zwiększenia przychodu wzrasta w 2009, 2010 i 2011 roku, odpowiednio o: 28,8, 39,4 i 26,22%.

Ocenę poprawności oszacowanego modelu, zliczając trafność klasyfikacji gospodarstw, zawierają dane tabeli 3.

Tabela 3

Zakwalifikowanie gospodarstw na podstawie modelu logitowego	Trafność klasyfikacji modelu logitowego				Ogólna trafność klasyfikacji R^2_{zlicz}	
	Rzeczywista przynależność gospodarstw					
	$y_i = 1$		$y_i = 0$			
$\hat{y}_i = 1$	2009	309	2009	22	2009	92,97%
	2010	308	2010	23		
	2011	301	2011	22		
$\hat{y}_i = 0$	2009	25	2009	313	2010	92,68%
	2010	26	2010	311		
	2011	32	2011	312		
Czułość, swoistość	2009	92,51%	2009	93,43%	2011	91,63%
	2010	92,22%	2010	93,11%		
	2011	90,39%	2011	93,41%		

Źródło: Obliczenia własne.

Na podstawie wyników przedstawionych w tabeli 3 można stwierdzić, że oszacowane modele logitowe charakteryzują się bardzo wysoką czułością i swoistością, czyli posiadają wysoką zdolność do typowania gospodarstw o rzeczywiście wyższych bądź niższych przychodach. Wartości zliczeniowego współczynnika R^2_{zlicz} są dużo wyższe od 50%, co oznacza, że klasyfikacja na podstawie modelu jest lepsza od przypadkowej.

Wyniki testu Hosmera–Lemeshowa wskazują na brak istotnych różnic między liczebnościami empirycznymi a teoretycznymi, wynikającymi z oszacowanych modeli regresji logistycznej (tab. 4).

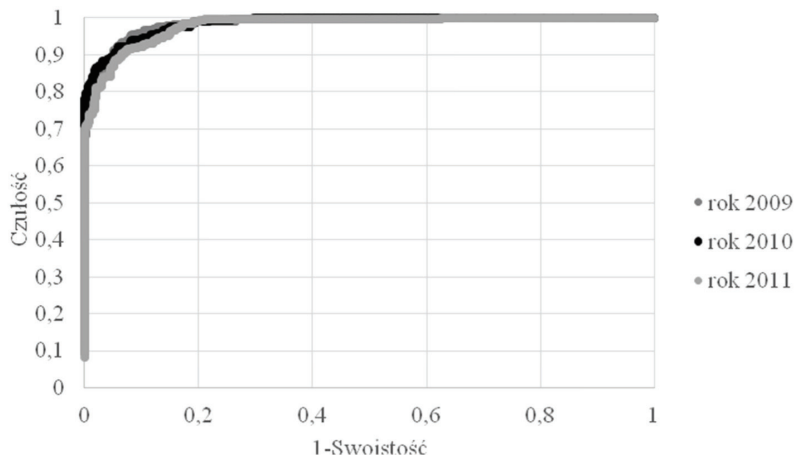
Tabela 4

Rok	Stopień dopasowania modeli logistycznych do danych empirycznych		Pole pod krzywą ROC
	Test Hosmera–Lemeshowa		
	χ^2	p	
2009	7,00	0,537	98,46%
2010	8,64	0,374	98,53%
2011	12,57	0,127	97,96%

Źródło: Obliczenia własne.

W ocenie jakości otrzymanych modeli logitowych wykorzystano także krzywą ROC oraz pole pod krzywą (tab. 4 i rys. 4). Krzywe ROC dla każdego badanego roku są zbliżone kształtem i wielkością pola pod krzywą. Wartości pola

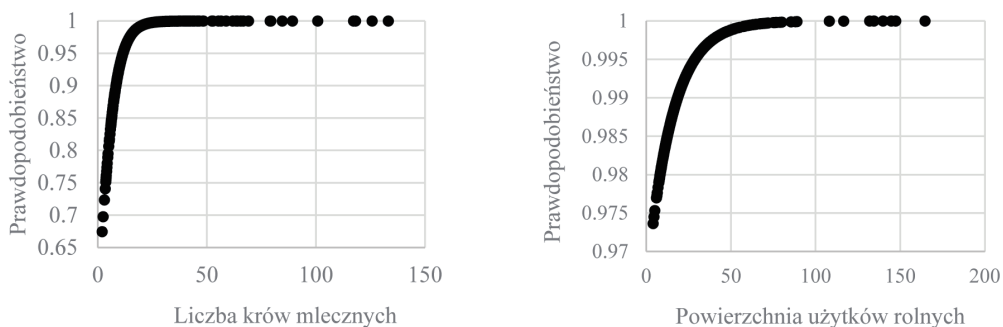
pod krzywą ROC, zbliżające się do 100%, świadczą o bardzo wysokiej jakości oszacowanych modeli w każdym roku. Pole jest istotnie większe od 0,5 (na poziomach istotności większych od 0,000001 w każdym modelu), można zatem na podstawie zbudowanych modeli klasyfikować gospodarstwa.



Rys. 4. Krzywe ROC dla modeli logitowych.

Źródło: Opracowanie własne.

Na podstawie oszacowanych modeli logitowych można określić, jak zmienna niezależna wpływa na prawdopodobieństwo osiągnięcia przez gospodarstwo wyższego przychodu. Wszystkie zmienne mają na to istotny dodatni wpływ. Oznacza to, że im wyższe wartości zmiennych niezależnych, tym wyższe prawdopodobieństwo uznania gospodarstwa za osiągające wyższy poziom przychodów. Na przykładzie liczby krów mlecznych i powierzchni użytków rolnych ilustruje to rys. 5.



Rys. 5. Prawdopodobieństwo uznania gospodarstwa za gospodarstwo o wyższych przychodach w zależności od liczby krów mlecznych i powierzchni użytków rolnych.

Źródło: Opracowanie własne.

Na podstawie rys. 5 można stwierdzić, że prawdopodobieństwo uznania gospodarstwa za zdolne do osiągania wyższych przychodów w dużym stopniu zależy zarówno od liczby krów mlecznych, jak i powierzchni użytków rolnych. Wraz ze zwiększeniem liczby krów mlecznych wzrasta prawdopodobieństwo uznania gospodarstwa za posiadające wyższe przychody. Na przykład, dla gospodarstwa o czterech krowach mlecznych prawdopodobieństwo wynosi 0,7, a przy 20 krowach zwiększa się do 0,99. Podobnie jest w przypadku powierzchni użytków rolnych, lecz wzrost wartości prawdopodobieństwa jest łagodniejszy. Dla gospodarstwa posiadającego najmniejszą powierzchnię użytków rolnych, wynoszącą około 4 ha, prawdopodobieństwo uznania go za obiekt posiadający wyższy przychód jest wysokie, w granicach 0,97, a przy zwiększeniu powierzchni użytków rolnych np. do 20 ha, badane prawdopodobieństwo wzrasta do 0,99.

Podsumowanie

Zastosowanie modelowania logitowego pozwoliło na określenie czynników poprawiających przychody gospodarstw rolnych specjalizujących się w produkcji mleka oraz wyznaczenie prawdopodobieństwa zakwalifikowania go do grupy o wyższych przychodach.

Przeprowadzone badanie daje podstawy do sformułowania następujących wniosków:

1. W całym badanym okresie wartości cech przyjętych jako niezależne w modelu logitowym wykazywały tendencje wzrostowe – prawidłowość ta dotyczy zmiennych o charakterze ilościowym.
2. Przychód określony jako zmienna zależna także zwiększał średnie wartości z roku na rok, a jego rozkład charakteryzował się bardzo silną, prawostronną asymetrią.
3. Największa zmiana wiązała się z posiadaniem następcy. W 2011 roku 20% gospodarstw zadeklarowało posiadanie następcy i był to wzrost o 3 pkt. proc. w porównaniu do roku 2009.
4. Struktura badanych gospodarstw nie zmieniała się znacząco pod względem płci osoby zarządzającej gospodarstwem. 87% gospodarstw było kierowanych przez mężczyzn.
5. Stabilna była także struktura gospodarstw pod względem wykształcenia osoby kierującej gospodarstwem. Największy odsetek dotyczył osób z wykształceniem średnim rolniczym, a najmniejszy – z wykształceniem wyższym.
6. Spośród zmiennych niezależnych przyjętych do modelu regresji logitowej istotny wpływ na prawdopodobieństwo zwiększenia dochodu miały w latach 2009-2011: powierzchnia użytków rolnych, liczba krów mlecznych, nakłady plonotwórcze, pasze z zakupu, pozostałe koszty bezpośrednie produkcji zwierzęcej oraz koszt amortyzacji.
7. Najbardziej istotne w modelach, w każdym badanym roku, okazały się zmienne niezależne: powierzchnia użytków rolnych i liczba krów mlecznych, które miały dodatni wpływ na wielkość przychodów.

9. Wszystkie oszacowane modele regresji logistycznej charakteryzują się bardzo dobrą jakością, o czym świadczą wartości: współczynnika R^2_{licz} wynoszące ponad 90%, test Hosmera–Lemeshowa wskazujący na brak istotnych różnic między liczebnościami empirycznymi a teoretycznymi oraz pole pod krzywymi ROC, wynoszące ponad 90%, informujące o wysokiej zdolności klasyfikacyjnej modeli.

Literatura:

- Cramer, J.S. (2003). *Logit Models from Economics and Other Fields*. Cambridge: Cambridge University Press.
- Dobosz, M. (2004). *Wspomagana komputerowo statystyczna analiza wyników badań*. Warszawa: Akademicka Oficyna Wydawnicza EXIT.
- Gołaś, Z., Kurzawa, I. (2014). Zastosowanie uporządkowanego modelu logitowego w analizie rentowności branż przemysłu spożywczego. *Zagadnienia Ekonomiki Rolnej, nr 1(338)*, s. 78-96.
- Hosmer, D.W., Lemeshow, S. (2000). *Applied logistic regression* (wyd. 2). New York: Wiley & Sons.
- Jackowska, B., Wycinka, E. (2011). Wykorzystanie regresji logistycznej w analizie czynników wpływających na aktywne poszukiwanie pracy przez osoby długotrwale bezrobotne. W: *Modelowanie i prognozowanie gospodarki narodowej*. Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, Gdańsk, s. 393-403.
- Kasprzyk, B., Fura, B. (2011). Wykorzystanie modeli logitowych do identyfikacji gospodarstw domowych zagrożonych ubóstwem. *Wiadomości Statystyczne*, nr 6, s. 1-16.
- Kleinbaum, D.G., Klein, M. (2002). *Logistic Regression*. New York: Springer.
- Kmieć, D. (2015). Zastosowanie modelu logitowego do analizy czynników wpływających na bezrobocie wśród ludności wiejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego Ekonomika i Organizacja Gospodarki Żywnościowej*, nr 110, s. 33-42.
- Kowerski, M., Bielak, J., Długosz, D. (2006). Zastosowanie logitowych modeli mikro-makro do prognozowania kondycji ekonomiczno-finansowej przedsiębiorstw. *Barometr Regionalny. Analizy i prognozy*, nr 6, s. 56-66.
- Maddala, G.S. (2001). *Introduction to Econometrics*. Third Edition. Chichester, New York: Wiley and Sons Ltd.
- Młodak, A. (2006). *Analiza taksonomiczna w statystyce regionalnej*, Warszawa: Difin.
- Sompolska-Rzechuła, A., Machowska-Szewczyk, M., Chudecka-Głaz, A., Cymbaluk-Płoska, A., Menkiszak, J. (2014). The Use of Logistic Regression in the Ovarian Cancer Diagnostics. *Ekonometria*, nr 3(45), s. 151-164.
- Stanisz, A. (2007). *Przystępny kurs z zastosowaniem Statistica PL na przykładach z medycyny*. T. 3. Kraków: StatSoft.
- Ziętara, W. (red.) (2013). *Polskie gospodarstwa z chowem bydła na tle wybranych krajów*. Program Wieloletni 2011-2014, nr 86. Warszawa: IERiGŻ-PIB.

AGNIESZKA SOMPOLSKA-RZECHUŁA
MICHAŁ ŚWITŁYK
West Pomeranian University of Technology
Szczecin

FACTORS AFFECTING PROBABILITY OF INCOME INCREASE IN AGRICULTURAL HOLDINGS SPECIALISED IN MILK PRODUCTION

Summary

The paper uses logistic regression model to determine factors affecting probability of income increase in agricultural holdings specialised in milk production. Analysis had a dynamic character and covered the period between 2009 and 2011. Independent variables included both quality and quantity features associated with farm activities. The analysis showed that, in each year, the most important variables in the logit model concerned the utilised agricultural area and the number of dairy cows, which had a positive impact on probability of income increase. All estimated models are characterised by high quality and thus can be used to correctly classify agricultural holdings.

Keywords: income, agricultural holdings, milk production, logit model.

Zaakceptowano do druku – Accepted for print: 01.12.2016.