

Anna Rosa

Politechnika Koszalińska

CZYNNIKI WPŁYWAJĄCE NA KORZYSTANIE PRZEZ ROLNIKÓW ZE ŚRODKÓW UNIJNYCH – UJĘCIE MODELOWE

*FACTORS INFLUENCING THE USE BY FARMERS OF EU FUNDS –
A MODEL APPROACH*

Słowa kluczowe: rolnicy, środki pomocowe, czynnik społeczny, czynnik ekonomiczny, powierzchnia gospodarstwa rolnego

Key words: farmers, aid measures, social factor, economic factor, status and size of a farm

Abstrakt. Celem artykułu było określenie czynników, które wpływają na prawdopodobieństwo korzystania ze środków pomocowych przez rolników. Źródłem danych do estymacji parametrów modelu logistycznego były badania terenowe zrealizowane z zastosowaniem metody sondażu diagnostycznego, a narzędziem badawczym był kwestionariusz ankiety. Zestaw zmiennych objaśniających stanowiły: status oraz powierzchnia gospodarstwa rolnego, zatrudnienie, wartość produkcji, płeć, wiek oraz wykształcenie kierującego. Analiza wykazała, że statystycznie istotny wpływ na fakt korzystania ze środków unijnych przez rolników miały powierzchnia gospodarstwa rolnego oraz wiek kierującego gospodarstwem

Wstęp

Wstąpienie Polski do Unii Europejskiej (UE) stworzyło nowe możliwości pozyskania środków finansowych z programów i działań pomocowych UE dla rolnictwa. Wdrożeniem instrumentów współfinansowanych z budżetu UE zajęła się Agencja Restrukturyzacji i Modernizacji Rolnictwa (ARiMR), która jest instytucją rządową, a jej działalność wzoruje się na praktykach stosowanych przez podobne instytucje wspierające rozwój rolnictwa w krajach UE. W latach 2002-2010 gospodarstwa rolne w Polsce w wyniku realizacji krajowej i unijnej polityki wsparcia rolnictwa zasilone zostały kwotą około 93 mld zł [Rudnicki 2013]. Przy czym najważniejszy instrument wsparcia udzielanego rolnictwu i obszarom wiejskim w Polsce od roku 2004 stanowią dopłaty bezpośrednie [Rosa 2012]. Decyzje rolników związane z zewnętrznymi funduszami, w tym środkami pomocowymi, są determinowane zarówno przez czynniki ekonomiczne, behawioralne, jak i społeczno-demograficzne [Kata 2013].

Celem artykułu było określenie czynników, które wpływają na prawdopodobieństwo skorzystania przez rolników ze środków pomocowych. Dla potrzeb badania czynniki charakteryzujące gospodarstwa rolne pogrupowano według cech: społecznych (płeć, wiek, wykształcenie) i ekonomicznych (status oraz powierzchnia gospodarstwa rolnego, zatrudnienie, wartość produkcji).

Materiał źródłowy i metodyka badań

Do wyodrębnienia czynników, które wpływają na prawdopodobieństwo korzystania ze środków unijnych przez rolników wykorzystano model regresji logistycznej. Źródłem danych do estymacji modelu prawdopodobieństwa były badania terenowe zrealizowane w 2012 roku wśród rolników z regionu Pomorza Środkowego z zastosowaniem metody sondażu diagnostycznego, a narzędziem badawczym był kwestionariusz ankiety. Wyboru respondentów dokonano wykorzystując nielosową (nieprobabilistyczną) metodę kuli śnieżnej (ang. *snowball sampling*). W badaniu podjęto próbę przeprowadzenia 300 wywiadów z rolnikami, jednak do analizy zakwalifikowano 249 kwestionariuszy.

Regresja logistyczna jest popularną techniką wielowymiarowej analizy danych wykorzystywaną w wielu dziedzinach nauki, np. w medycynie, naukach przyrodniczych i ekonomicznych¹. Jest ona przydatna w sytuacji, gdy wartości zmiennej dwustanowej typu: posiadanie brak oszczędności, niewypłacalny lub wypłacalny będą objaśniane za pomocą innych zmiennych oraz gdy dokonuje się oceny siły i kierunku wpływu tych zmiennych na analizowane zjawisko. Budowany model regresji logistycznej pozwala na zbadanie wpływu wielu zmiennych X_1, \dots, X_k na dychotomiczną zmienną zależną Y . Zmienna zależna przyjmuje jedynie dwie wartości, które są oznaczane jako jeden i zero, gdzie: 1 – wartość wyróżniona – posiadanie danej cechy, 0 – brak danej cechy. Funkcja, na której oparty jest model regresji logistycznej uwzględnia prawdopodobieństwo przyjęcia przez zmienną dychotomiczną wartości równej 1, określonej następująco:

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

gdzie $\alpha_0, \alpha_1, \dots, \alpha_k$ są parametrami modelu, X_1, \dots, X_k są zmiennymi niezależnymi przyczynowymi, które mogą mieć charakter zarówno jakościowy (skategoryzowany), jak i ilościowy.

Współczynniki modelu regresji logistycznej poszukuje się metodą największej wiarygodności [Dobosz 2004] oraz uogólnioną metodą najmniejszych kwadratów [Gruszczyński, Podgórska 1996]. Jednak dla niektórych wartości zmiennych prawdopodobieństwo p może być poza przedziałem $[0, 1]$, co jest sprzeczne z podstawową własnością prawdopodobieństwa. Aby nie było takich sprzeczności, wartości prawdopodobieństwa poddaje się transformacji [Rusnak 2011]. Najczęściej spotykaną transformacją jest funkcja logitowa [Stanisz 2007, Cramer 2003, Kleinbaum, Klein 2002]:

$$\text{logit } p = \ln \left(\frac{p}{1-p} \right)$$

gdzie: p – prawdopodobieństwo występowania określonego zdarzenia.

W wyniku transformacji otrzymuje się model logitowy:

$$\text{logit } p = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k$$

Do interpretacji wykorzystuje się iloraz szans rozumiany jako stosunek prawdopodobieństwo wystąpienia wyróżnionego zdarzenia do prawdopodobieństwa, że ten przypadek się nie pojawi. Określa się zatem o ile większa bądź mniejsza jest szansa wystąpienia zdarzenia, tj. $\frac{p}{1-p}$. Po oszacowaniu parametrów modelu logitowego dokonuje się oceny jego poprawności, zliczając trafnie oraz błędnie zaklasyfikowane przypadki (tab. 1).

Do oceny stopnia dopasowania modelu regresji logistycznej do danych empirycznych wykorzystuje się miarę zliczeniową R^2 , która przyjmuje wartości z przedziału $\langle 0, 1 \rangle$ [Maddala 2008]:

$$R^2_{zlicz} = \frac{n_1 + n_2}{n}$$

Im wartość tej miary bliższa jedności, tym lepsze dopasowanie modelu logistycznego do danych empirycznych badanego zjawiska, R^2_{zlicz} oznacza procent trafnie zaklasyfikowanych przypadków. Model dobrze sprawdza się w wyjaśnianiu badanego zjawiska, gdy $R^2_{zlicz} > 50\%$.

Tabela 1. Macierz klasyfikacji przypadków

Table 1. Classification Matrix cases

Wartości oczekiwane/ The values expected values	Wartości obserwowane/ The values observed		Suma/ Sum
	$y_i = 1$	$y_i = 0$	
$\hat{y}_i = 1$	n_{11}	n_{12}	$n_{1\bullet}$
$\hat{y}_i = 0$	n_{21}	n_{22}	$n_{2\bullet}$
Suma/Sum	$n_{\bullet 1}$	$n_{\bullet 2}$	n

Źródło: opracowanie własne na podstawie [Dobosz 2004]

Source: own study based on [Dobosz 2004]

¹ Regresja logistyczna jest wykorzystywana w badaniach przez m.in. [Zawadzka, Ardan 2010, s. 430-441, Franc-Dąbrowska 2010, Śliwicki, Ręklewski 2012, s. 23-35].

Jakość zbudowanego modelu regresji logistycznej można również ocenić innymi miarami, np. za pomocą krzywej ROC (ang. *Receiver Operating Characteristic Curves*), a w szczególności pola pod krzywą ROC. Krzywa powstaje przez połączenie punktów w kartezjańskim układzie o współrzędnych (1– swoistość, czułość) [Sompolska-Rzechuła i in. 2014]. Czułość jest to zdolność wykrywania jednostek posiadających wyróżnioną cechę:

$$\text{czułość} = \frac{n_{11}}{n_{\bullet 1}}$$

Swoistość natomiast opisuje zdolność wykrywania jednostek nieposiadających wyróżnionej cechy. Wyznaczana jest jako iloraz obserwacji nieposiadających wyróżnionej cechy i sklasyfikowanych poprawnie do liczby wszystkich obserwacji nieposiadających wyróżnionej cechy:

$$\text{swoistość} = \frac{n_{22}}{n_{\bullet 2}}$$

Jakość klasyfikacyjna modelu jest dobra, gdy pole pod krzywą ROC jest większe niż pole pod prostą $x = y$, zatem większe niż 0,5.

Wyniki badań

Do oceny prawdopodobieństwa korzystania ze środków pomocowych (unijnych) przez rolników wzięto do badań następujący zestaw zmiennych:

Y – korzystanie ze środków unijnych (tak – 1, nie – 0),

X_1 – status gospodarstwa (T – towarowe, NT – nietowarowe),

X_2 – powierzchnia gospodarstwa rolnego w ha,

X_3 – zatrudnienie osób do pracy w gospodarstwie rolnym (tak – 1, nie – 0),

X_4 – wartość produkcji w zł (poniżej 29 999 – 1; od 30 000 do 49 999 – 2, od 50 000 do 99 999 – 3, od 100 000 do 500 000 – 4, powyżej 500 000 – 5),

X_5 – płeć kierującego gospodarstwem (mężczyzna – 1, kobieta – 2),

X_6 – wiek kierującego gospodarstwem w latach,

X_7 – wykształcenie (podstawowe – 1, zasadnicze zawodowe – 2, niepełne średnie – 3, średnie – 4, średnie rolnicze – 5, wyższe – 6).

Za pomocą regresji krokowej w tył określono zmienne przyczynowe wpływające istotnie na zmienną zależną Y . Otrzymano zbiór składający się z następujących zmiennych przyczynowych: powierzchnia gospodarstwa rolnego, wartość produkcji, wiek kierującego gospodarstwem. Oceny parametrów modelu logitowego z uwzględnieniem wymienionych zmiennych przedstawiono w tabeli 2.

Tabela 2. Oceny parametrów modelu logitowego z istotnymi zmiennymi przyczynowymi
Table 2. Logit model parameter estimation with significant causal variables

Zmienna/ Variable	Nazwa zmiennej/ Variable name	Ocena parametru/ Parameter estimate	<i>p</i> -value	Iloraz szans/ The odds ratio
	stała/ <i>constans</i>	-0,7916	0,2905	-
X_2	powierzchnia/area [ha]	0,0740	0,0004	1,0769
X_4	wartość produkcji [zł]/value of production [PLN]	-0,5582	0,0392	0,5722
X_6	wiek [lata]/age [years]	0,0470	0,0039	1,0482

Źródło: badania własne

Source: own study

Oszacowany model logistyczny przybiera następującą postać:

$$P(y = 1) = p = \frac{e^{-0,7916+0,0740x_2-0,5582x_4+0,0470x_6}}{1 + e^{-0,7916+0,0740x_2-0,5582x_4+0,0470x_6}}$$

Model po przekształceniu można przedstawić następująco (w nawiasach podano standardowe błędy szacunku tych parametrów):

$$\text{logit } p = -0,7916 + 0,0740x_2 - 0,5582x_4 + 0,0470x_6$$

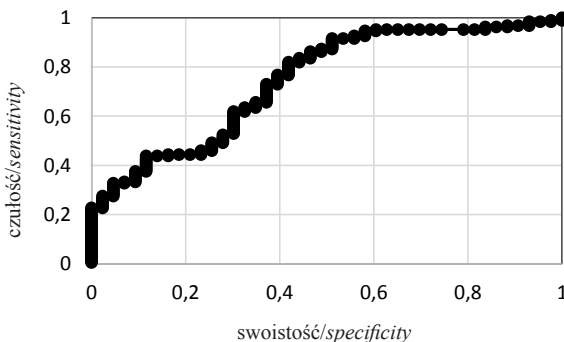
(0,7489)
(0,0208)
(0,2708)
(0,0163)

W modelu dodatni, statystycznie istotny wpływ na zmienną zależną Y mają: powierzchnia gospodarstwa rolnego i wiek kierującego gospodarstwem. Zmienna wartość produkcji ma statystycznie ujemny wpływ na korzystanie ze środków pomocowych dotowanych przez UE dla rolników.

Interpretując ilorazy szans przy i -tej zmiennej (zakładając, że pozostałe zmienne uwzględnione w modelu pozostaną bez zmian) uzyskuje się następujące informacje:

- 1) jeśli powierzchnia gospodarstwa rolnego zwiększy się o 1 ha, to szansa skorzystania ze środków unijnych zwiększy się 7,7%,
- 2) jeśli wartość produkcji będzie wyższa, to prawdopodobieństwo skorzystania ze środków unijnych zmniejszy się o 42,8%,
- 3) im starszy kierujący gospodarstwem tym szansa na korzystanie ze środków pomocowych dotowanych przez UE większa o 4,8%.

Następnie dokonano oceny poprawności modelu logitowego, zliczając trafnie oraz błędnie zaklasyfikowane przypadki (tab. 3).



Rysunek 1. Krzywa ROC

Figure 1. The ROC curve

Źródło: opracowanie własne

Source: own study

Tabela 3. Macierz klasyfikacji przypadków

Table 3. Classification Matrix cases

Wartości oczekiwane/ The expected values	Wartości obserwowane/ The values observed		Suma/ Sum	Ogólna trafność klasyfikacji/ General classification accuracy [%]
	$y_i = 1$	$y_i = 0$		
$\hat{y}_i = 1$	$n_{11} = 186$	$n_{12} = 40$	$n_{1\bullet} = 226$	81,47
$\hat{y}_i = 0$	$n_{21} = 3$	$n_{22} = 3$	$n_{2\bullet} = 6$	
Suma	$n_{\bullet 1} = 189$	$n_{\bullet 2} = 43$	$n = 232$	
Czułość, swoistość/ Sensitivity, specificity [%]	98,41	6,98		

Źródło: opracowanie własne na podstawie [Dobosz 2004]

Source: own study based on [Dobosz 2004]

Do oceny jakości oszacowanego modelu wykorzystano współczynnik R^2_{zlicz} . Ogólna trafność klasyfikacji przypadków wyrażona R^2_{zlicz} jest wysoka, zatem model ma dobrą jakość. Model odznacza się bardzo wysoką czułością, czyli zdolnością do rozpoznawania czy rolnik korzysta ze środków pomocowych. Natomiast zdolność do rozpoznawania rolników niekorzystających ze środków pomocowych dotowanych przez UE jest bardzo niska.

Do oceny jakości oszacowanego modelu wykorzystano także krzywą ROC oraz pole pod krzywą, które wynosi 74,6%, co potwierdza dobrą jakość modelu (rys. 1).

Podsumowanie i wnioski

Powierzchnia gospodarstwa rolnego, wartość produkcji oraz wiek kierującego gospodarstwem są głównymi czynnikami mającymi istotny wpływ na korzystanie przez rolnika ze środków pomocowych. Statystycznie istotny pozytywny wpływ na zmienną zależną Y miały wśród czynników ekonomicznych – powierzchnia gospodarstwa rolnego, a wśród czynników społecznych – wiek kierującego gospodarstwem. Oznacza to, że im większa powierzchnia gospodarstwa rolnego oraz starszy kierujący gospodarstwem, tym większe prawdopodobieństwo skorzystania przez gospodarstwo ze środków unijnych.

Czynnik ekonomiczny – zmienna wartość produkcji miał statystycznie ujemny wpływ na korzystanie ze środków UE przez rolników, czyli im większa była wartość produkcji, tym prawdopodobieństwo skorzystania przez rolnika ze środków pomocowych było niższe.

Literatura

- Cramer J.S. 2003: *Logit Models from Economics and Other Fields*, Cambridge University Press, Cambridge.
- Dobosz M. 2004: *Wspomagana komputerowo statystyczna analiza wyników badań*, Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- Franc-Dąbrowska J. 2010: *Teoretyczne i praktyczne aspekty gospodarowania zyskiem w przedsiębiorstwach rolniczych*, Wydawnictwo SGGW, Warszawa.
- Gruszczyński M., Podgórska M. (red). 1996: *Ekonometria*, Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa, 139-141.
- Kata R. 2013: *Czynniki behawioralne i demograficzne wpływające na korzystanie przez rolników z kredytów inwestycyjnych*, Zesz. Nauk. SGGW, Ekonomia i Organizacja Gospodarki Żywnościowej, nr 103, 54.
- Kleinbaum D.G., Klein M. 2002: *Logistic Regression*, Springer, New York.
- Maddala G.S. 2008: *Ekonometria*, PWN, Warszawa.
- Rosa A. 2012: *Unijne fundusze jako źródło wsparcia rolnictwa i obszarów wiejskich w Polsce*, Roczn. Nauk. SERiA, t. XIV, z. 2, 135-140.
- Rudnicki R. 2013: *Zróżnicowanie przestrzenne absorpcji funduszy Unii Europejskiej w rolnictwie polskim jako problem badawczy i aplikacyjny*, Acta Universitatis Lodzianis, Folia Geographica Socio-Oeconomica, 13, 73.
- Rusnak Z. 2011: *Regresja logistyczna a analiza oszczędności i zobowiązań finansowych gospodarstw domowych*, Zesz. Nauk. Wyższej Szkoły Bankowej we Wrocławiu, nr 25, Wrocław, 169-188.
- Śliwicki D., Ręklewski M. 2012: *Wykorzystanie modeli logitowych w analizie czynników aktywności zawodowej ludności*, Wiadomości Statystyczne, 2, Warszawa, 23-35.
- Sompolska-Rzechuła A., Machowska-Szewczyk M., Chudecka-Głaz A., Cymbaluk-Płoska A., Menkiszak J. 2014: *The Use of Logistic Regression in the Ovarian Cancer Diagnostic*, *Ekonometria Econometrics* 3(45), Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 151-164.
- Stanisz A. 2007: *Przystępny kurs z zastosowaniem Statistica PL na przykładach z medycyny, Tom 3*, Statsoft Kraków,
- Zawadzka D., Ardan R., 2010: *Modelowe ujęcie czynników pozacenowych wpływających na prawdopodobieństwo ubiegania się o kredyt bankowy przez małe przedsiębiorstwa*, [w:] A. Kopiński (red.), *Finanse przedsiębiorstw*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 98, Wrocław, 430-441.

Summary

This article aims to identify the factors that affect the likelihood benefit from the EU funds by farmer. The source of data to estimate the probability model field tests were carried out using the method of diagnostic survey and research tool was a questionnaire. Sets of explanatory variables were: status and size of a farm, employment, production value, gender, age and education farmer. The analysis showed that the larger the size of the holding and head of the holding older the more likely to benefit from EU funds.

Adres do korespondencji
dr Anna Rosa

Politechnika Koszalińska, Wydział Nauk Ekonomicznych
ul. Kwiatkowskiego 6e, 75-343 Koszalin
tel. (94) 34 39 164, email: anna.rosa@tu.koszalin.pl