

Dominik ŚLIWICKI, Marek RĘKLEWSKI

Wykorzystanie modeli logitowych w analizie czynników aktywności zawodowej ludności

Celem badania jest próba identyfikacji istotnych czynników wpływających na aktywność zawodową ludności w wieku 15 lat i więcej oraz osób powyżej 50. roku życia. W Badaniu Aktywności Ekonomicznej Ludności (BAEL) ludność aktywna zawodowo określana jest jako zasoby pracy, podaż pracy lub siła robocza obejmująca osoby w wieku 15 lat i więcej uznane za pracujące lub bez-

robotne. Do pracujących zalicza się osoby w wieku 15 lat i więcej, które w okresie badanego tygodnia: wykonywały pracę przez co najmniej 1 godzinę przynoszącą zarobek lub dochód; miały pracę, ale jej nie wykonywały z powodu choroby, urlopu macierzyńskiego lub wypoczynkowego czy z innych powodów (przy czym długość przerwy w pracy wynosiła do 3 miesięcy albo powyżej 3 miesięcy), ale osoby te były pracownikami najemnymi i w tym czasie otrzymywały co najmniej 50% dotychczasowego wynagrodzenia. Do pracujących zaliczani byli również uczniowie, z którymi zakłady pracy lub osoby fizyczne zawarły umowę o naukę zawodu lub przyuczenie do określonej pracy, jeżeli otrzymywali wynagrodzenie.

Do bezrobotnych zaliczono osoby w wieku 15—74 lata, które w okresie badanego tygodnia nie były osobami pracującymi, aktywnie poszukiwały pracy, były gotowe podjąć pracę w ciągu dwóch tygodni następujących po tygodniu badanym. Do tej grupy zostały zaliczone także osoby, które nie poszukiwały pracy, ponieważ miały pracę załatwioną i oczekiwały na jej rozpoczęcie przez okres nie dłuższy niż 3 miesiące oraz były gotowe tę pracę podjąć (*Aktywność...*, 2010).

Prognozowany znaczny spadek zasobów pracy i nasilające się starzenie społeczeństwa stawiają przed polityką społeczno-gospodarczą państwa zadanie podtrzymywania aktywności zawodowej ludności. W analizach rynku pracy bardzo często wykorzystuje się modele zmiennych jakościowych, do których należą m.in. modele logitowe (Wiśniewski, 2009).

TEORETYCZNE PODSTAWY MODELI LOGITOWYCH

Modele logitowe należą do grupy modeli jakościowych, tzn. takich, w których zmienna objaśniana Y jest zmienną jakościową. Najczęściej zmienna ta przyjmuje dwie wartości: 0 lub 1. Zmienna objaśniana Y w modelach logitowych przyjmuje wartości:

$Y = 1$ — osoba aktywna zawodowo,

$Y = 0$ — osoba bierna zawodowo.

Model logitowy przyjmuje postać (Gruszczyński, 2010):

$$y_i^* = \ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i$$

gdzie:

β_j — parametr strukturalny modelu,

u_i — składnik losowy,

$\ln \frac{p_i}{1 - p_i}$ — logit,

- y_i^* — zmienna nieobserwowalna,
 x_{ji} — wartości zmiennych objaśniających modelu,
 p_i — prawdopodobieństwo przyjęcia wartości 1 przez zmienną zależną y_i wyznaczone na podstawie funkcji gęstości rozkładu logistycznego:

$$p_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} = \frac{1}{1 + \exp(-x_i' \beta)} = \frac{1}{1 + e^{-y_i}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})}}$$

Zmienną nieobserwowalną y_i^* nazywa się zmienną ukrytą. Tym, co obserwujemy jest zmienna zero-jedynkowa y_i mająca postać:

$$y_i = \begin{cases} 1; & y_i^* > 0 \\ 0; & y_i^* \leq 0 \end{cases}$$

Logit to logarytm ilorazu szans przyjęcia i nieprzyjęcia wartości 1 przez zmienną y_i . Jeśli szanse są jednakowe ($p_i = 0,5$), to logit równa się zeru. Dla $p_i < 0,5$ logit jest ujemny, a dla $p_i > 0,5$ jest dodatni. Logitowa transformacja prawdopodobieństwa pozwala zastąpić wartość p_i przez liczbę z przedziału $(-\infty, +\infty)$. Jeśli oznaczmy:

$$\frac{p_i}{1 - p_i} = \exp(x_i' \beta) = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})$$

Ze wzoru wynika, że przyrost wartości X_{ji} o jednostkę wiąże się, *ceteris paribus*, z $\exp(\beta_j)$ -krotną zmianą ilorazu szans. W przypadku $\exp(\beta_j) > 1$ mamy wzrost, a w przypadku $\exp(\beta_j) < 1$ obserwujemy spadek ilorazu szans $\frac{p_i}{1 - p_i}$.

Efekty krańcowe w modelu logitowym nie są stałe i zależą od zmiennych objaśniających:

$$\frac{\partial p_i}{\partial X_{ji}} = \beta_j \frac{\exp(x_i' \beta)}{[1 + \exp(x_i' \beta)]^2} = \beta_j p_i (1 - p_i)$$

Interpretacja otrzymanych parametrów strukturalnych modelu jest podobna do interpretacji parametrów modelu liniowego. Odczytywane są one jako przyrost prawdopodobieństwa zdarzenia $Y=1$, związany z jednostkowym przyrostem X_j . Dla dodatniego β_j wzrost X_j wiąże się ze wzrostem szans na to, że $Y=1$, natomiast spadkowi X_j towarzyszy spadek szans na to, że $Y=1$. Dla ujemnego β_j wzrost X_j wiąże się ze spadkiem szans na to, że $Y=1$, natomiast spadkowi X_j towarzyszy wzrost szans na to, że $Y=1$. Efekty krańcowe podaje się w przypadku wartości średnich wszystkich zmiennych objaśniających.

Istotność całego modelu weryfikuje się za pomocą testu ilorazu wiarygodności, którego hipotezy mają postać:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \exists_{1 \leq j \leq k} \beta_j \neq 0$$

Hipoteza zerowa mówi, że wszystkie parametry przy zmiennych objaśniających są zerami, to znaczy prawdziwy jest jedynie model z wyrazem wolnym. Statystyka testu przyjmuje postać:

$$LR = 2(\ln L_p - \ln L_{ww})$$

gdzie:

L_p — wartość funkcji wiarygodności dla pełnego modelu,

L_{ww} — wartość funkcji wiarygodności dla modelu zawierającego tylko wyraz wolny.

Jakość dopasowania modelu zmiennej zero-jedynkowej można ocenić na podstawie specjalnej wersji wskaźnika determinacji R^2 , tzw. pseudo- R^2 . Wartości tego współczynnika mieszczą się w przedziale $[0,1]$, a jego wyższe wartości mają świadczyć o lepszym dopasowaniu modelu. W badaniu podano wskaźnik pseudo- R^2 McFaddena. Opiera się on na porównaniu modelu pełnego z modelem zredukowanym tylko dla wyrazu wolnego. Oblicza się go według wzoru:

$$McFaddenR^2 = 1 - \frac{\ln L_p}{\ln L_{ww}}$$

gdzie $\ln L_p$ jest logarytmem funkcji wiarygodności modelu pełnego, a $\ln L_{ww}$ oznacza logarytm funkcji wiarygodności modelu, w którym występuje tylko wyraz wolny. W praktyce wartości R^2 McFaddena są niewielkie, bliższe 0 niż 1.

R^2 można wyznaczyć na podstawie współczynnika korelacji między y oraz $\hat{y}$. Miara ta oparta jest na resztowej sumie kwadratów (Maddala, 2008):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

W przypadku zależnej zmiennej binarnej mamy:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = n_1 - n\left(\frac{n_1}{n}\right)^2 = \frac{n_1 n_0}{n}$$

Z tego:

$$R^2 = 1 - \frac{n}{n_1 n_0} \sum_{i=1}^n (y_i - \hat{y})^2$$

gdzie:

n_1 — liczba wartości 1 w n -elementowej próbie,

n_0 — liczba wartości 0.

Dodatkowym sposobem na zbadanie jakości dopasowania modelu jest przedstawienie wyników prognozy na podstawie modelu. Prognoza opiera się na oszacowanym prawdopodobieństwie $\hat{p}_i$, które jest funkcją $F(x'_i \beta)$. Zazwyczaj przyjmuje się, że jeśli $F(x'_i \beta) \geq 0,5$, to prognoza jest równa $\hat{y}_i = 1$. Jeśli $F(x'_i \beta) < 0,5$, to prognoza z modelu równa się $\hat{y}_i = 0$. Na tej podstawie generowana jest tablica trafności:

Faktyczne	Przewidywane		Razem
	$\hat{Y} = 0$	$\hat{Y} = 1$	
R a z e m	N_{f0}	N_{f1}	N
$Y=0$	n_{00}	n_{01}	N_{p0}
$Y=1$	n_{10}	n_{11}	N_{p1}

gdzie:

n_{00} — liczba przypadków, dla których wartość rzeczywista i przewidywana są równe 0,

n_{01} — liczba przypadków, dla których wartość rzeczywista wynosi 0, a przewidywana 1,

n_{10} — liczba przypadków, dla których wartość rzeczywista wynosi 1, a przewidywana 0,

n_{11} — liczba przypadków, dla których wartość rzeczywista i przewidywana są równe 1.

Zatem procentową trafność prognoz oblicza się następująco (Kufel, 2011):

$$\text{łączna (ogółem):} \quad Traf\ Pr\ og = \frac{n_{00} + n_{11}}{N} \cdot 100$$

$$\text{dla } Y=1: \quad Traf\ Pr\ og_1 = \frac{n_{11}}{N_{p1}} \cdot 100$$

$$\text{dla } Y=0: \quad Traf\ Pr\ og_0 = \frac{n_{00}}{N_{p0}} \cdot 100$$

Ponadto trafność modeli zmiennych jakościowych można przedstawić za pomocą ilorazu szans według wzoru:

$$IRS = \frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{10}}$$

DANE WYKORZYSTANE DO ESTYMACJI MODELI LOGITOWYCH

Źródłem danych do estymacji modeli prawdopodobieństwa aktywności zawodowej ludności w wieku 15 lat i więcej oraz osób powyżej 50. roku życia były dane jednostkowe z BAEL w IV kwartale 2009 r. Zestaw zmiennych objaśniających (zestawienie), opisujących społeczno-ekonomiczną sytuację ankietowanych na rynku pracy, to: poziom wykształcenia (*WYKSZ*), płeć (*PLEC*), miejsce zamieszkania z podziałem na województwa (*WOJ*), a także miasto/wieś (*MIASTO*), stopień pokrewieństwa z głową rodziny (*SP*), wcześniejszą sytuację na rynku pracy (*ROK_WCZ*) i źródło utrzymania (*ZRODLO*) w badanym miesiącu, stan cywilny (*STCYW*), fakt wystąpienia wypadku przy pracy (*WYP*), wiek (*WIEK*).

ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH

Województwo zamieszkania			
<i>WOJ_02</i>	dolnośląskie	<i>WOJ_18</i>	podkarpackie
<i>WOJ_04</i>	kujawsko-pomorskie	<i>WOJ_20</i>	podlaskie
<i>WOJ_06</i>	lubelskie	<i>WOJ_22</i>	pomorskie
<i>WOJ_08</i>	lubuskie	<i>WOJ_24</i>	śląskie
<i>WOJ_10</i>	łódzkie	<i>WOJ_26</i>	świętokrzyskie
<i>WOJ_12</i>	małopolskie	<i>WOJ_28</i>	warmińsko-mazurskie
<i>WOJ_14</i>	mazowieckie	<i>WOJ_30</i>	wielkopolskie
<i>WOJ_16</i>	opolskie	<i>WOJ_32</i>	zachodniopomorskie
Stopień pokrewieństwa z głową rodziny			
<i>SP_01</i>	głowa gospodarstwa domowego	<i>SP_06</i>	ojciec/matka/teść/teściowa
<i>SP_02</i>	mąż/zona	<i>SP_07</i>	dziadek/babcia/wnuk/wnuczka/prawnuk/prawnuczka
<i>SP_03</i>	partner/partnerka	<i>SP_08</i>	brat/siostra
<i>SP_04</i>	syn/córka	<i>SP_09</i>	wujek/ciotka/dalszy krewny
<i>SP_05</i>	zięć/synowa	<i>SP_10</i>	niespokrewniony członek gospodarstwa domowego

ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH (dok.)

Wykształcenie

WYKSZ_0	wyższe ze stopniem naukowym (co najmniej doktora)	WYKSZ_5	zasadnicze zawodowe
WYKSZ_1	wyższe (licencjat, inżynier, magister)	WYKSZ_6	gimnazjalne
WYKSZ_2	policealne	WYKSZ_7	podstawowe
WYKSZ_3	średnie zawodowe	WYKSZ_8	niepełne podstawowe
WYKSZ_4	średnie ogólnokształcące	WYKSZ_9	bez wykształcenia szkolnego

Źródło utrzymania

ZRODLO_01	praca najemna	ZRODLO_06	zasilek dla bezrobotnych
ZRODLO_02	praca we własnym gospodarstwie rolnym	ZRODLO_07	świadczenie przedemerytalne
ZRODLO_03	praca na własny rachunek poza rolnictwem indywidualnym	ZRODLO_08	inne świadczenia społeczne
ZRODLO_04	emerytura	ZRODLO_09	inne niezarobkowe źródło
ZRODLO_05	renta inwalidzka	ZRODLO_10	pozostawanie na utrzymaniu

Sytuacja na rynku pracy rok wcześniej

ROK_WCZ_1	praca	ROK_WCZ_5	niepełnosprawność
ROK_WCZ_2	bezrobocie	ROK_WCZ_6	zasadnicza służba wojskowa
ROK_WCZ_3	nauka/szkolenie	ROK_WCZ_7	obowiązki rodzinne
ROK_WCZ_4	emerytura, wcześniejsza emerytura	ROK_WCZ_8	inna forma bierności zawodowej

Stan cywilny

STCYW_01	kawaler/panna	STCYW_03	wdowiec/wdowa
STCYW_02	żonaty/zamężna	STCYW_04	rozwódziona(a)/w separacji

Wystąpienie wypadku przy pracy w 2009 r.

WYP_1	tak, bez absencji lub ze zwolnieniem lekarskim do 3 dni	WYP_3	nie był(a) poszkodowany(a) w wypadku przy pracy
WYP_2	tak, ze zwolnieniem lekarskim do 4 dni i więcej	WYP_4	nie dotyczy

Wiek w latach

WIEK

Miejsce zamieszkania (miasto/wieś)

MIASTO (miasto = 1)

Płeć (mężczyzna/kobieta)

PLEC (mężczyzna = 1)

Źródło: opracowanie własne na podstawie formularzy statystycznych o symbolu ZD, ZG.

WYNIKI ESTYMACJI MODELI

W celach porównawczych oszacowano modele logitowe aktywności zawodowej dotyczące dwóch grup wiekowych: ludności w wieku 15 lat i więcej oraz osób powyżej 50. roku życia. Wyniki oszacowań modeli znajdują się w tabl. 1—5. Zmienne nieistotne statystycznie eliminowano z modelu metodą *a posteriori*. Procedura ta polega na porównaniu wartości *p* z przyjętym poziomem istot-

ności. Jeżeli p było większe od α , zmienną eliminowano z modelu i dokonywano reestymacji. Działanie to powtarzano aż do uzyskania modelu z wartościami p mniejszymi lub równymi α . Wartość α przyjęto na poziomie 10%.

TABL. 1. WYNIKI OSZACOWANIA MODELU LOGITOWEGO AKTYWNOŚCI ZAWODOWEJ OSÓB W WIEKU 15 LAT I WIĘCEJ W POLSCE W IV KWARTALE 2009 R.

Zmienne	Współczynniki	Błąd standardowy	z	Wartość p	Efekt krańcowy	Iloraz szans
<i>const</i>	-3,9403	0,3110	-12,6700	0,0000 ***	x	x
<i>WOJ_06</i>	0,5525	0,0992	5,5670	0,0000 ***	0,0549	1,7376
<i>WOJ_10</i>	0,2218	0,1087	2,0400	0,0413 **	0,0245	1,2484
<i>WOJ_12</i>	0,3065	0,1081	2,8360	0,0046 ***	0,0329	1,3586
<i>WOJ_14</i>	-0,3455	0,1101	-3,1380	0,0017 ***	-0,0454	0,7078
<i>WOJ_18</i>	0,8734	0,0936	9,3360	0,0000 ***	0,0783	2,3951
<i>WOJ_22</i>	-0,2901	0,1208	-2,4020	0,0163 **	-0,0377	0,7482
<i>WOJ_26</i>	0,2799	0,1070	2,6160	0,0089 ***	0,0303	1,3229
<i>WOJ_28</i>	-0,2321	0,1162	-1,9970	0,0458 **	-0,0296	0,7929
<i>WOJ_30</i>	0,2865	0,1065	2,6910	0,0071 ***	0,0310	1,3318
<i>PLEC</i>	0,3847	0,0563	6,8340	0,0000 ***	0,0452	1,4692
<i>SP_01</i>	-0,1929	0,0981	-1,9660	0,0493 **	-0,0231	0,8246
<i>SP_02</i>	-0,4074	0,0987	-4,1270	0,0000 ***	-0,0518	0,6654
<i>SP_04</i>	0,1893	0,1058	1,7890	0,0737 *	0,0216	1,2084
<i>ZRODLO_01</i>	7,7128	0,4120	18,7200	0,0000 ***	0,6900	2236,8189
<i>ZRODLO_02</i>	5,1544	0,2574	20,0200	0,0000 ***	0,1732	173,1988
<i>ZRODLO_03</i>	5,4599	0,3926	13,9100	0,0000 ***	0,1700	235,0786
<i>ZRODLO_04</i>	0,6932	0,1228	5,6450	0,0000 ***	0,0726	2,0001
<i>ZRODLO_06</i>	0,2363	0,1323	1,7860	0,0740 *	0,0257	1,2665
<i>ZRODLO_07</i>	-0,7547	0,2423	-3,1150	0,0018 ***	-0,1152	0,4702
<i>ZRODLO_08</i>	-0,2575	0,0955	-2,6950	0,0070 ***	-0,0332	0,7730
<i>ROK_WCZ_1</i>	2,0541	0,1567	13,1100	0,0000 ***	0,2404	7,8001
<i>ROK_WCZ_2</i>	2,2367	0,1449	15,4300	0,0000 ***	0,1324	9,3626
<i>ROK_WCZ_3</i>	-0,3103	0,1683	-1,8440	0,0652 *	-0,0399	0,7332
<i>ROK_WCZ_4</i>	-1,5120	0,1778	-8,5040	0,0000 ***	-0,2349	0,2205
<i>ROK_WCZ_5</i>	-1,1470	0,1727	-6,6410	0,0000 ***	-0,1890	0,3176
<i>ROK_WCZ_6</i>	2,6998	0,5242	5,1510	0,0000 ***	0,1269	14,8762
<i>ROK_WCZ_7</i>	-0,3050	0,1698	-1,7970	0,0724 *	-0,0399	0,7371
<i>MIASTO</i>	-0,3182	0,0552	-5,7600	0,0000 ***	-0,0371	0,7275
<i>WIEK</i>	0,0846	0,0117	7,2260	0,0000 ***	0,0100	1,0883
<i>WIEK^</i>	-0,0011	0,0001	-8,7060	0,0000 ***	-0,0001	0,9989
<i>WYKSZ_0</i>	1,6915	0,6604	2,5610	0,0104 **	0,1091	5,4276
<i>WYKSZ_1</i>	1,8096	0,1403	12,9000	0,0000 ***	0,1387	6,1080
<i>WYKSZ_2</i>	1,7201	0,1895	9,0790	0,0000 ***	0,1142	5,5849
<i>WYKSZ_3</i>	1,3292	0,1311	10,1400	0,0000 ***	0,1192	3,7782
<i>WYKSZ_4</i>	0,8853	0,1301	6,8060	0,0000 ***	0,0813	2,4237
<i>WYKSZ_5</i>	1,1289	0,1304	8,6550	0,0000 ***	0,1104	3,0922
<i>WYKSZ_7</i>	0,9408	0,1286	7,3150	0,0000 ***	0,0909	2,5621
<i>WYP_4</i>	-0,7426	0,0851	-8,7300	0,0000 ***	-0,0848	0,4759

U w a g a. Poziom istotności parametrów: *** $\alpha = 0,01$, ** $\alpha = 0,05$, * $\alpha = 0,10$.

Ź r ó d ł o: obliczenia własne (wykonane z wykorzystaniem pakietu komputerowego Gretl).

**TABL. 2. TABLICA TRAFNOŚCI ZMIENNEJ
„AKTYWNOŚĆ ZAWODOWA LUDNOŚCI W WIEKU 15 LAT I WIĘCEJ”**

Faktyczne	Przewidywane		Razem
	$\hat{Y}=0$	$\hat{Y}=1$	
R a z e m	21778	22259	44036
<i>Y</i> =0	20486	957	21443
<i>Y</i> =1	1292	21301	22593

Ź r ó d ł o: obliczenia własne.

Współczynnik korelacji jako miara dopasowania modelu aktywności zawodowej ludności w wieku 15 lat i więcej był bardzo wysoki i wynosił $r=0,924$ (kwadrat współczynnika korelacji $r^2=0,854$). Wysoki współczynnik korelacji świadczy o dobrym dopasowaniu modelu opisującego zmienność wartości zmiennej „aktywność ekonomiczna”.

**TABL. 3. WYNIKI OSZACOWANIA MODELU LOGITOWEGO AKTYWNOŚCI ZAWODOWEJ
OSÓB POWYŻEJ 50. ROKU ŻYCIA W IV KWARTALE 2009 R.**

Zmienne	Współczynniki	Błąd standardowy	<i>z</i>	Wartość <i>p</i>	Efekt krańcowy	Iloraz szans
<i>const</i>	-6,3966	2,4446	-2,6170	0,0089 ***	x	x
<i>WOJ_06</i>	0,5047	0,1508	3,3470	0,0008 ***	0,1168	1,6566
<i>WOJ_12</i>	0,4005	0,1621	2,4710	0,0135 **	0,0917	1,4926
<i>WOJ_14</i>	-0,6186	0,1776	-3,4830	0,0005 ***	-0,1198	0,5387
<i>WOJ_16</i>	-0,2873	0,1730	-1,6600	0,0968 *	-0,0593	0,7503
<i>WOJ_18</i>	1,3398	0,1361	9,8430	0,0000 ***	0,3208	3,8181
<i>WOJ_22</i>	-0,3929	0,1950	-2,0150	0,0439 **	-0,0793	0,6751
<i>WOJ_24</i>	-0,3172	0,1754	-1,8080	0,0705 *	-0,0652	0,7282
<i>WOJ_28</i>	-0,4338	0,1880	-2,3070	0,0210 **	-0,0869	0,6481
<i>WOJ_30</i>	-0,3288	0,1918	-1,7140	0,0866 *	-0,0673	0,7198
<i>PLEC</i>	0,1936	0,0862	2,2450	0,0247 **	0,0422	1,2136
<i>SP_02</i>	-0,3161	0,0917	-3,4450	0,0006 ***	-0,0668	0,7290
<i>STCYW_01</i>	-0,3469	0,1786	-1,9430	0,0520 *	-0,0706	0,7068
<i>ZRODŁO_01</i>	7,8860	1,0198	7,7330	0,0000 ***	0,8911	2659,7303
<i>ZRODŁO_02</i>	5,3528	0,5230	10,2400	0,0000 ***	0,7131	211,2009
<i>ZRODŁO_03</i>	6,3390	1,0460	6,0610	0,0000 ***	0,7214	566,2468
<i>ZRODŁO_04</i>	0,4131	0,1500	2,7550	0,0059 ***	0,0888	1,5114
<i>ZRODŁO_07</i>	-0,4849	0,2701	-1,7950	0,0726 *	-0,0952	0,6158
<i>ROK_WCZ_1</i>	3,3426	0,1665	20,0700	0,0000 ***	0,6833	28,2915
<i>ROK_WCZ_2</i>	3,1281	0,1450	21,5700	0,0000 ***	0,6098	22,8296
<i>ROK_WCZ_3</i>	2,4523	1,2330	1,9890	0,0467 **	0,5263	11,6144
<i>ROK_WCZ_4</i>	-0,2986	0,1781	-1,6770	0,0935 *	-0,0648	0,7418
<i>MIASTO</i>	-0,6366	0,0922	-6,9050	0,0000 ***	-0,1399	0,5291
<i>WIEK</i>	0,1445	0,0751	1,9250	0,0543 *	0,0313	1,1555
<i>WIEK^</i>	-0,0015	0,0006	-2,6820	0,0073 ***	-0,0003	0,9985
<i>WYKSZ_0</i>	1,6362	0,9275	1,7640	0,0777 *	0,3875	5,1357
<i>WYKSZ_1</i>	1,7811	0,6156	2,8930	0,0038 ***	0,4178	5,9365
<i>WYKSZ_2</i>	1,8000	0,6443	2,7940	0,0052 ***	0,4212	6,0494
<i>WYKSZ_3</i>	1,2335	0,6093	2,0240	0,0429 **	0,2890	3,4331
<i>WYKSZ_4</i>	1,2524	0,6210	2,0170	0,0437 **	0,2997	3,4987
<i>WYKSZ_5</i>	1,0001	0,6067	1,6480	0,0993 *	0,2301	2,7184
<i>WYKSZ_6</i>	2,5583	1,2984	1,9700	0,0488 **	0,5399	12,9136
<i>WYKSZ_7</i>	1,0755	0,6014	1,7880	0,0737 *	0,2436	2,9316
<i>WYP_2</i>	-3,7619	1,5822	-2,3780	0,0174 **	-0,3082	0,0232
<i>WYP_4</i>	-1,0410	0,1288	-8,0800	0,0000 ***	-0,2399	0,3531

Ź r ó d ł o: jak przy tabl. 2.

**TABL. 4. TABLICA TRAFNOŚCI ZMIENNEJ
„AKTYWNOŚĆ ZAWODOWA LUDNOŚCI POWYŻEJ 50. ROKU ŻYCIA”**

Faktyczne	Przewidywane		Razem
	$\hat{Y}=0$	$\hat{Y}=1$	
R a z e m	14159	5916	20074
$Y=0$	13595	334	13929
$Y=1$	564	5581	6145

Ź r ó d ł o: jak przy tabl. 3.

W tablicach 2 i 4 przedstawiono liczbę trafień wartości prognozowanych 0 i 1 względem wartości rzeczywistych. Zmienną prognozowaną zdefiniowano jako $\hat{Y}$, która jest zero-jedynkowa i taka, że jeśli: $\hat{y}_i > 0,5$, to przyjmuje się $\hat{Y}=1$, a kiedy $\hat{y}_i < 0,5$, to $\hat{Y}=0$.

Współczynnik korelacji między wartościami teoretycznymi wyznaczonymi na podstawie modelu do danych rzeczywistych zmiennej aktywności zawodowej powyżej 50. roku życia jest bardzo wysoki i wyniósł $r=0,918$ (kwadrat współczynnika korelacji $r^2=0,842$).

TABL. 5. STATYSTYKA DOPASOWANIA MODELI LOGITOWYCH

Wyszczególnienie	Model (1) dla osób w wieku 15 lat i więcej	Model (2) dla osób powyżej 50. roku życia
Współczynnik korelacji R między Y i $\hat{Y}$	0,924	0,918
R -kwadrat McFaddena	0,819	0,808
Skorygowany R -kwadrat	0,818	0,805
Test ilorazu wiarygodności	49976,8	19986,5
Logarytm wiarygodności	-5520	-2371,52
Trafność prognozy (ogółem)	0,949	0,955
Trafność prognozy (dla $Y=1$)	0,943	0,908
Trafność prognozy (dla $Y=0$)	0,955	0,976
Iloraz szans	352,9	402,8

Ź r ó d ł o: jak przy tabl. 2.

Wyniki testu ilorazu wiarygodności (wartość $p=0,000<0,10$) oznaczają istotność oszacowanych modeli. Ogólna trafność wartości prognozowanych przez modele była bardzo wysoka i wynosiła ponad 90%: model (1) — 94,9%, model (2) — 95,5%.

Podsumowanie

Wyniki estymacji modeli logitowych z tabl. 1 i 3 pokazują, że największą wartość poznawczą dają efekty krańcowe oraz ilorazy szans. Efekty krańcowe (obliczane dla średnich wartości zmiennych objaśniających) interpretuje się jako wpływ zmiennej objaśniającej na prawdopodobieństwo sukcesu (czyli przyjęcia wartości 1 przez zmienną objaśnianą). Ilorazy szans natomiast interpretuje się

jako procentowy wpływ jednostkowej zmiany wartości zmiennej objaśniającej na iloraz szans liczony jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki.

Z grupy zmiennych dotyczących województw w obu modelach istotne okazały się te, które odnoszą się do województw: lubelskiego, małopolskiego, mazowieckiego, podkarpackiego, pomorskiego, warmińsko-mazurskiego i wielkopolskiego (ponadto w modelu dotyczącym osób w wieku 15 lat i więcej — województwa łódzkie i świętokrzyskie, a w modelu grupy osób powyżej 50. roku życia — województwa opolskie i śląskie).

W modelu aktywności dotyczącym grupy osób powyżej 50. roku życia największy wpływ na zwiększenie prawdopodobieństwa bycia aktywnym zawodowo (o 32,1 p.proc., wobec 7,8 p.proc. w grupie osób w wieku 15 lat i więcej) wywiera zamieszkiwanie w woj. podkarpackim. Zamieszkiwanie w woj. mazowieckim zmniejsza prawdopodobieństwo bycia aktywnym zawodowo (o 12,0 p.proc., wobec 4,5 p.proc. w grupie osób w wieku 15 lat i więcej).

Analizując płeć respondenta, prawdopodobieństwo bycia aktywnym zawodowo na rynku pracy wzrasta w przypadku mężczyzn (4,2 p.proc. wśród osób powyżej 50. roku życia oraz 4,5 p.proc. wśród ludności w wieku 15 lat i więcej). Prawdopodobieństwo aktywności zawodowej maleje w stosunku do kwadratu wieku (o 0,03 p.proc. wśród osób powyżej 50. roku życia, wobec 0,01 p.proc. dla osób w wieku 15 lat i więcej).

Spośród zmiennych charakteryzujących stopień pokrewieństwa z głową rodziny w modelu dotyczącym grupy osób powyżej 50. roku życia jako istotna pozostała zmienna oznaczająca bycie mężem/zoną głowy gospodarstwa domowego. Oszacowanie parametru przy tej zmiennej oznacza, że powoduje ona spadek prawdopodobieństwa aktywności o 6,68 p.proc. W grupie osób w wieku 15 lat i więcej zmienna ta również powoduje spadek prawdopodobieństwa aktywności o 5,18 p.proc.

Stan cywilny kawaler/panna wpływa ujemnie na prawdopodobieństwo aktywności zawodowej tylko w grupie wiekowej 50 lat i więcej o 7,1 p.proc. W przypadku osób w wieku 15 lat i więcej żadna ze zmiennych dotyczących stanu cywilnego nie okazała się istotna.

Każdy poziom wykształcenia wpływa dodatnio na zwiększenie prawdopodobieństwa aktywności zawodowej. Zaskakujące jest to, że w grupie osób powyżej 50. roku życia największy wpływ na wzrost prawdopodobieństwa aktywności wywiera posiadanie wykształcenia gimnazjalnego. W grupie osób w wieku 15 lat i więcej największy wpływ na wzrost prawdopodobieństwa aktywności wywiera posiadanie wykształcenia wyższego, a następnie policealnego.

Analizując źródło utrzymania, prawdopodobieństwo bycia aktywnym wzrasta wśród osób pracujących najemnie w obu badanych grupach wieku (o 89,1 p.proc. w grupie osób powyżej 50. roku życia oraz o 69,0 p.proc. wśród osób w wieku 15 lat i więcej). Zmniejszenie tego prawdopodobieństwa następuje, gdy

jest pobierane świadczenie emerytalne (odpowiednio: o 9,5 p.proc. i o 11,52 p.proc.). W grupie osób w wieku 15 lat i więcej na zmniejszenie prawdopodobieństwa bycia aktywnym zawodowo (o 3,3 p.proc.) wpływa pobieranie innych świadczeń społecznych. Wcześniejszy status respondenta na rynku pracy wpływał istotnie na aktywność zawodową. Pobieranie zaś emerytury w grupie osób powyżej 50. roku życia zmniejsza prawdopodobieństwo aktywności zawodowej o 6,5 p.proc. Na zmniejszenie prawdopodobieństwa aktywności zawodowej w grupie osób powyżej 50. roku miało wpływ wystąpienie wypadku przy pracy (ze zwolnieniem lekarskim 4 dni lub więcej w 2009 r. — o 30,8 p.proc.). Zaskakującym spostrzeżeniem w przypadku obu grup wiekowych jest to, że zamieszkiwanie w miastach zmniejszało prawdopodobieństwo aktywności zawodowej (o 14,0 p.proc. wśród osób powyżej 50. roku życia, a w grupie osób w wieku 15 lat i więcej o 3,7 p.proc.).

dr Dominik Śliwicki — *Urząd Statystyczny w Bydgoszczy, Wyższa Szkoła Gospodarki w Bydgoszczy,*

mgr Marek Ręklewski — *Urząd Statystyczny w Bydgoszczy*

LITERATURA

Aktywność Ekonomiczna Ludności Polski. IV kwartał 2009 (2010), GUS

Gruszczyński M. (2010), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, red., Wolters Kluwer, Warszawa

Kufel T. (2011), *Ekonometria — Rozwiązywanie problemów z wykorzystywaniem programu Gretl*, PWN, Warszawa

Maddala G. S. (2008), *Ekonometria*, PWN, Warszawa

Wiśniewski Z. (2009), *Determinanty aktywności zawodowej ludzi starszych*, TNOiK, Dom Organizatora, Toruń

SUMMARY

The aim of the study is to identify important determinants affecting the economic activity of population aged 15 and older and those over 50 age on the labor market in Poland. The analysis provides an example of the use of logit models to identify and analyze the factors of population activity. The projected decline in labor resources and the increasing population aging confront the socio-economic policy of the State the task of sustaining activity of these populations. Analyses of labor market models often use variables, which include logit models.

РЕЗЮМЕ

Целью обследования была попытка идентификации важных факторов влияющих на профессиональную активность населения в возрасте 15 лет и больше, а также лиц старше 50 лет. Анализ является примером использования логитных моделей для идентификации и анализа факторов профессиональной активности населения. Прогнозируемое снижение трудовых ресурсов и ускоренное старение общества ставят перед социально-экономической политикой государства задачи поддержки профессиональной активности обсуждаемых групп населения. В анализе рынка труда используются модели качественных переменных, к которым принадлежат в частности логитные модели.