

Łukasz Mach

Politechnika Opolska

MODELOWANIE LOGITOWE NARZĘDZIEM WSPOMAGAJĄCYM PODEJMOWANIE DECYZJI NA RYNKU NIERUCHOMOŚCI MIESZKANIOWYCH

Streszczenie: W artykule przedstawiono proces budowy modelu logitowego, którego celem jest wspomaganie procesu podejmowania decyzji na rynku nieruchomości. Opracowany model regresji logistycznej określa prawdopodobieństwo dokonania transakcji na rynku mieszkaniowym i wskazuje statystycznie istotnie zmienne wpływające na kształtowanie się popytu. Merytorycznie przygotowany i poprawnie przeprowadzony proces podejmowania decyzji (na podstawie modelowania logitowego) jest kluczowym determinanem wpływającym na poprawę konkurencyjności przedsiębiorstw, zwłaszcza w czasie trwania globalnego kryzysu gospodarczego.

Słowa kluczowe: podejmowanie decyzji, regresja logistyczna, prawdopodobieństwo sprzedaży.

1. Wstęp

Merytorycznie przygotowany i poprawnie przeprowadzony proces podejmowania decyzji jest najważniejszym czynnikiem oddziałującym na poprawę konkurencyjności przedsiębiorstw. Dobrze przygotowany proces decyzyjny pozwala zminimalizować ryzyko źle podjętych decyzji, pozwala także wzmocnić pozycję konkurencyjną przedsiębiorstw. W dobie globalnego kryzysu gospodarczego przedsiębiorstwa powinny przywiązywać szczególną uwagę do trafnych decyzji, szczególnie w branżach wyjątkowo narażonych na oddziaływanie kryzysu. Uwzględniając teorię ekonomii, należy stwierdzić, że głównym czynnikiem wpływającym na konkurencyjność gospodarek są inwestycje, które w znacznym stopniu są implikowane przez branżę budowlaną (m.in. łatwość sprzedaży zasobu mieszkaniowego).

Celem niniejszej publikacji jest zbudowanie jakościowego modelu decyzyjnego, którego głównym zadaniem będzie określenie prawdopodobieństwa dokonania transakcji na rynku nieruchomości oraz wskazanie zmiennych istotnie wpływających na dokonanie transakcji. Zdefiniowany model bazuje na regresji logistycznej z powodzeniem wykorzystywanej w procesach decyzyjnych, w których zmienna zależna ma charakter dychotomiczny. Wybór modelu bazującego na funkcji logistycznej ma

również swoje uzasadnienie, gdyż większości zjawisk społeczno-ekonomicznych ma parametry zbliżone do przebiegu funkcji logistycznej.

2. Regresja logistyczna metodą analizy danych jakościowych

W wielu przypadkach procesy decyzyjne bazują na modelach regresji wielorakiej, tzn. takiej, w której analizujemy wpływ kilku zmiennych niezależnych na jedną zmienną zależną typu mierzalnego [Dittmann 2004; Maddala 2008]. Autor niniejszej publikacji zajmował się tym zagadnieniem we wcześniejszych pracach badawczych, w których za pomocą modelowania liniowego (regresji wielorakiej) wskazywał na istotne determinanty określające cenę nieruchomości na rynku mieszkaniowym [Mach 2008]. Gdy zaś zmienna zależna jest typu dychotomicznego, wówczas powinniśmy zastosować regresję logistyczną. W badaniach ekonomicznych bardzo popularnym przykładem wykorzystania regresji logistycznej jest analiza zdolności spłaty zaciągniętych kredytów bankowych, natomiast w badaniach społecznych – możliwość wskazania prawdopodobieństwa np. zakupu nowego produktu przez klienta z punktu widzenia określonych (statystycznie istotnych) cech dotyczących produktu i specyfiki procesu decyzyjnego nabywcy.

Autor, kontynuując wcześniejsze badania rynku mieszkaniowego (dotyczące zastosowania regresji wielorakiej), w niniejszej publikacji zastosował regresję logistyczną do określenia prawdopodobieństwa sprzedaży nieruchomości mieszkaniowej z punktu widzenia czasu oczekiwania nieruchomości na sprzedaż (prawdopodobieństwo sprzedaży nieruchomości w czasie krótszym od średniego czasu oczekiwania).

Zaletą regresji logistycznej jest to, że interpretacja wyników jest bardzo podobna do metod stosowanych w regresji klasycznej. Jednakże należy również zaznaczyć, że regresja logistyczna w porównaniu z regresją wieloraką jest bardziej skomplikowana obliczeniowo, a wyliczenie wartości i sporządzenie wykresów reszt często nie wnosi nic nowego do modelu [Stanisz 2007].

Logistyczny model regresji dla zmiennej dychotomicznej wyrażony jest wzorem (1) [Gruszczyński, Kuszewski, Podgórska (red.) 2009; Maddala 2008; Stanisz 2007].

$$P(Y = 1 / x_1, x_2, \dots, x_k) = \frac{e^{a_0 + \sum_{i=1}^k a_i x_i}}{1 + e^{a_0 + \sum_{i=1}^k a_i x_i}}, \quad (1)$$

gdzie: a_i , $i = 0, 1, 2, \dots, k$ – parametry strukturalne modelu logitowego; x_1, x_2, \dots, x_k – zmienne niezależne, które mogą być zarówno ilościowe, jak i jakościowe.

Do oszacowania współczynników regresji nie możemy użyć popularnej metody najmniejszych kwadratów, gdyż wymaga ona założenia o stałości wariancji. Z tego

względu do estymacji parametrów w regresji logistycznej wykorzystuje się metodę największej wiarygodności (MNV) [Gruszczyński, Kuszewski, Podgórska (red.) 2009; Maddala 2008; Stanisz 2007].

W regresji logistycznej, prócz interpretacji współczynników regresji, dochodzi jeszcze jeden parametr, tj. iloraz szans. Jest to stosunek prawdopodobieństwa, że jakieś zdarzenie wystąpi, do prawdopodobieństwa, że ten przypadek się nie pojawi. Dla określonego przykładu A możemy to zapisać w postaci wzoru (2) [Stanisz 2007]:

$$S(A) = \frac{P(A)}{1 - P(A)}. \quad (2)$$

Interpretacja szans wyrażonych wzorem (2) jest przydatna do wyjaśnienia oszacowanego modelu logitowego. Możemy udowodnić, że gdy wybrana zmienna niezależna wzrośnie o jednostkę, to iloraz szans zmieni się $\exp(a_i)$ razy. Jeśli $\exp(a_i) > 1$, to należy się spodziewać wzrostu ilorazu szans, natomiast gdy $\exp(a_i) < 1$, to należy oczekiwać jego spadku.

Gdy zmienna niezależna jest zmienną zero-jedynkową, wówczas $\exp(a_i)$ oznacza, ile razy wzrasta iloraz dla wartości zmiennej zależnej równej jedności [Gruszczyński, Kuszewski, Podgórska (red.) 2009].

W kolejnym kroku procesu budowy modelu logistycznego powinniśmy określić miary dopasowania modelu i dokonać weryfikacji poprawności modelu (testowania modelu). Miary dopasowania modelu możemy wyrazić za pomocą wartości pseudo- R^2 McFaddena, jak również za pomocą miar R -kwadrat Cragga-Uhlera, Nagelkerke czy Coxa-Snella.

Ze względu na nieco inne podejście obliczeniowe należy zachować szczególną uwagę przy porównywaniu modeli logistycznych, w których zostały określone miary dopasowania różnych autorów [Stanisz 2007].

Procedura testowania wyników estymacji w modelowaniu logistycznym jest dokonywana za pomocą wartości statystyki testu ilorazu wiarygodności. Hipoteza zerowa dla tak zdefiniowanych założeń mówi, że wszystkie parametry modelu, bez wyrazu wolnego są równe zero. Statystyka testu jest wyrażona wzorem (3) [Gruszczyński, Kuszewski, Podgórska (red.) 2009; Maddala 2008; Stanisz 2007].

$$2(\ln L_{MP} - \ln L_{MZ}), \quad (3)$$

gdzie: $\ln L_{MP}$ – logarytm funkcji wiarygodności dla modelu pełnego; $\ln L_{MZ}$ – logarytm funkcji wiarygodności dla modelu zredukowanego i z założenia ma rozkład χ^2 z liczbą stopni swobody równą liczbie zmiennych objaśniających modelu pełnego.

Należy również zaznaczyć, że dobór zmiennych do modelu jest poddany tym samym wymaganiom, co w klasycznym modelu liniowym.

W końcowym etapie budowy modelu logistycznego można obliczyć prognozę *ex post* wartości zmiennej zależnej dla każdej obserwacji. Przyjmuje się, że trafność prognozy *ex post* wygodnie jest przedstawić za pomocą tablicy trafności. Obliczając

prognozę *ex post*, należy się zastosować do dwóch zasad, tj. zasady standardowej stosowanej przy próbie zbilansowanej oraz zasady optymalnej wartości granicznej [Gruszczyński, Kuszewski, Podgórska (red.) 2009].

Założenia poprawności analizy regresji logistycznej [Stanisz 2007] są następujące:

- wybór próby należy przeprowadzić w sposób losowy,
- w procesie przygotowania danych należy usunąć dane odstające,
- należy zastosować odpowiednie kodowanie, tzn. zmiennej zależnej wartość jeden przypisać dla przypadku nas interesującego,
- powinno się włączyć do modelu wszystkie zmienne istotne statystycznie,
- należy wyłączyć z modelu wszystkie nieistotne zmienne,
- trzeba pamiętać o liniowej zależności transformacji logitowej od zmiennych niezależnych,
- należy unikać zjawiska współliniowości,
- powinno się pamiętać o dużej liczebności próby, tj. powinno być spełnione założenie $N > 10(k + 1)$, gdzie k jest liczbą parametrów, jednakże minimalną liczebność powinno się określić na 100 elementów.

3. Proces badawczy – zastosowanie regresji logistycznej w praktyce

Budowa modelu logitowego bazowała na próbie 120 transakcji sprzedaży dokonanych w Opolu od I kw. 2008 r. do III kw. 2009 r. Przy budowie modelu logitowego w pierwszej kolejności sprawdzono, czy występują obserwacje odstające. Proces eliminacji obserwacji odstających przeprowadzono w odniesieniu do każdej zmiennej mierzalnej. W tym celu na bazie wykresu *plot-box* oraz obliczonych miar położenia (mediana, kwartył1, kwartył3) odrzucono wszystkie obserwacje podejrzane o odstające (za kryterium klasyfikacji zmiennej jako odstającej przyjęto $\pm 1,5$ IQR). Następnie zdefiniowano cel badania, którym jest zbudowanie modelu logistycznego sprawdzającego prawdopodobieństwo sprzedaży mieszkania w czasie krótszym niż średnia czasu oczekiwania na sprzedaż ze wszystkich ofert wystawionych do sprzedaży. Badania empiryczne realizowane w niniejszym artykule pozwolą m.in. odpowiedzieć na następujące hipotezy badawcze:

- Mieszkania zlokalizowane na pierwszym piętrze mają największe szanse na sprzedaż.
- W rozpatrywanym zasobie mieszkaniowym (mieszkania jedno- dwu- oraz trzy-pokojowe) każdy dodatkowy pokój zwiększa atrakcyjność mieszkania.
- Zwiększanie powierzchni mieszkania ponad powierzchnię przeciętną w próbie badawczej zmniejsza szanse mieszkania na sprzedaż.

Dla rekordów wykorzystanych w niniejszym badaniu średni czas oczekiwania nieruchomości na sprzedaż wynosi 88,64 dnia. Wejściowy model logitowy został zapisany w postaci formalnej we wzorze (4):

$$P(Y = 1 / x_1, x_2, \dots, x_{10}) = \frac{e^{a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5 + a_6 x_6 + a_7 x_7 + a_8 x_8 + a_9 x_9 + a_{10} x_{10}}}{1 + e^{a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5 + a_6 x_6 + a_7 x_7 + a_8 x_8 + a_9 x_9 + a_{10} x_{10}}}, \quad (4)$$

gdzie: x_1 – zmienna niezależna – logarytm naturalny ceny mieszkania; x_2 – zmienna niezależna – powierzchnia mieszkania; x_3 – zmienna niezależna – liczba pokoi w mieszkaniu; x_4 – zmienna niezależna – piętro, na którym znajduje się mieszkanie; x_5 – zmienna niezależna – kuchnia w mieszkaniu; x_6 – zmienna niezależna – ocena położenia mieszkania; x_7 – zmienna niezależna – standard wykończenia mieszkania; x_8 – zmienna niezależna – liczba kondygnacji w budynku; x_9 – zmienna niezależna – technologia budowy; x_{10} – zmienna niezależna – lokalizacja mieszkania w budynku; $a_0, a_1, a_2, \dots, a_{10}$ – parametry strukturalne modelu.

Po zbudowaniu formalnego modelu logitowego przeprowadzono estymację jego parametrów strukturalnych. W tabeli 1 przedstawiono wyniki estymacji z zastosowaniem metody Quasi-Newtona.

Tabela 1. Wstępne wyniki estymacji modelu logitowego

Wyszczególnienie	Stała	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Ocena	4,21	-0,09	-0,09	1,65	-0,37	0,48	-1,03	-0,65	0,10	-0,08	-0,15
Błąd standardowy	14,08	1,24	0,03	0,77	0,13	0,73	0,57	0,56	0,13	0,69	0,56
$t(110)$	0,30	-0,07	-2,77	2,15	-2,78	0,65	-1,82	-1,16	0,81	-0,11	-0,27
Poziom p	0,77	0,94	0,01	0,03	0,01	0,52	0,07	0,25	0,42	0,91	0,79

Źródło: opracowanie własne.

Analizując przedstawione wyniki, zauważa się, że zmiennymi istotnie wpływającymi na zmienną Y są x_2, x_3, x_4, x_6 . Po kolejnym odrzuceniu zmiennych nieistotnych statystycznie ostateczna postać modelu logitowego przyjęła postać wyrażoną wzorem (5).

$$P(Y) = \frac{e^{3,673 - 0,104x_2 + 1,732x_3 - 0,300x_4 - 1,180x_6}}{1 + e^{3,673 - 0,104x_2 + 1,732x_3 - 0,300x_4 - 1,180x_6}}, \quad (5)$$

gdzie: x_2 – powierzchnia mieszkania; x_3 – liczba pokoi w mieszkaniu; x_4 – piętro, na którym znajduje się mieszkanie; x_6 – ocena położenia mieszkania.

Podsumowanie procesu estymacji przedstawiono w tab. 2.

Wartość statystyki p dla całego modelu przyjęła wartość 0,0000502, co świadczy o istotności modelu w porównaniu z modelem tylko z wyrazem wolnym, co potwierdza cel badania – udowodnienie, że zbudowany model wnosi coś nowego. Ponadto należy poddać interpretacji, tzw. logarytm wiarygodności, który jest miarą dopasowania całego modelu. Logarytm ten obliczany jest za pomocą statystyki

$-2 \log$ z maksimum wiarygodności zbudowanego modelu i modelu zawierającego tylko wyraz wolny.

Tabela 2. Wyniki procesu estymacji dla zmiennych istotnych

Wyszczególnienie	Stała	X_2	X_3	X_4	X_6
Ocena	3,67	-0,10	1,73	-0,30	-1,18
Błąd standardowy	0,95	0,03	0,62	0,10	0,55
$t(110)$	3,88	-3,48	2,80	-2,92	-2,16
Poziom p	0,00	0,00	0,01	0,00	0,03
-95%CL	1,79	-0,16	0,50	-0,50	-2,26
+95%CL	5,55	-0,04	2,96	-0,10	-0,10
Chi-kwadrat Walda	15,02	12,13	7,82	8,51	4,68
Poziom p	0,00	0,00	0,01	0,00	0,03
Iloraz szans z jend.	39,35	0,90	5,65	0,74	0,31
-95%CL	6,02	0,85	1,66	0,61	0,10
+95%CL	257,36	0,96	19,27	0,91	0,91
Iloraz szans zakr.		0,00	5 753,88	0,04	0,31
-95%CL		0,00	12,46	0,00	0,10
+95%CL		0,00	2 656 223,00	0,35	0,91

Źródło: opracowanie własne.

W zbudowanym modelu wartości te wynoszą odpowiednio 122,3 oraz 147,3. Duża różnica między tymi statystykami ma rozkład zbliżony do chi-kwadrat. Statystyka ta to pierwszy krok weryfikacji istotności modelu. Na bazie zaprezentowanych wartości obliczono pseudo- R^2 ; wyniósł on 0,17.

Dokonując interpretacji otrzymanych wyników, możemy wnioskować, że:

- każdy dodatkowy pokój w mieszkaniu zwiększa 5,65 razy prawdopodobieństwo sprzedaży nieruchomości;
- mamy 35% szans na sprzedaż mieszkania, jeśli zwiększymy jego powierzchnię o 10 m², natomiast 12% szans na sprzedaż, gdy powierzchnię zwiększymy o 20 m² w stosunku do powierzchni bazowej;
- z tab. 3 wynika, że mieszkanie na trzecim piętrze w stosunku do mieszkania na 1 piętrze ma o połowę mniejszą szansę na sprzedaż, natomiast jeszcze mniejsze szanse na sprzedaż (0,30) ma mieszkanie znajdujące się na piętrze piątym (w porównaniu z mieszkaniem znajdującym się na pierwszym piętrze).

Tabela 3. Iloraz szans na sprzedaż mieszkania w zależności od piętra, na którym to mieszkanie się znajduje

Piętro	1	2	3	4	5	6	7	8	9	10
Wielkość zmniejszenia		0,74	0,55	0,41	0,30	0,22	0,16	0,12	0,09	0,07

Źródło: opracowanie własne.

Tabela 4. Tablica trafności

Wyszczególnienie	Przewidywane 0	Przewidywane 1	Procent poprawności
0,000000	14,00	25,00	35,90
1,000000	8,00	68,00	89,47

Źródło: opracowanie własne.

W tabeli 4 przedstawiono poprawnie i niepoprawnie zakwalifikowane przypadki dotyczące wyliczonego modelu. Obliczono również iloraz szans jako stosunek iloczynu poprawnie zaklasyfikowanych przypadków do iloczynu niepoprawnie zakwalifikowanych; wynosi on 4,76. Wartość większa od jedności oznacza, że ta klasyfikacja jest lepsza od tej, która zostałaby przeprowadzona przez przypadek.

4. Wnioski z badań

Przeprowadzony proces badawczy potwierdza przydatność zbudowanego modelu logitowego w procesie podejmowania decyzji na rynku mieszkaniowym. Zbudowany model może być stosowany jako narzędzie wspomagające podjęcie trafnych decyzji inwestycyjnych zarówno przez deweloperów, jak i indywidualnych sprzedających. Przyjmując za kryterium czas sprzedaży mieszkania (liczbę dni potrzebną na sprzedaż mieszkania), przy określonych parametrach wartościujących mieszkanie możemy wskazać prawdopodobieństwo powodzenia transakcji, możemy też przeprowadzić wielowymiarową interpretację parametrów.

W niniejszym artykule przedstawiono szczegółową analizę modelu logitowego, którego celem jest określenie prawdopodobieństwa sprzedaży nieruchomości w czasie krótszym od średniego czasu oczekiwania na sprzedaż.

Autor publikacji podjął również próbę budowy modeli logitowych, w których celem było określenie prawdopodobieństwa sprzedaży nieruchomości w czasie krótszym niż jeden miesiąc oraz w czasie krótszym niż jeden tydzień. Niestety na bazie posiadanych danych w obydwu modelach wszystkie zmienne niezależne okazały się nieistotne statystycznie. Niepokojącym parametrem w zdefiniowanym modelu logitowym jest niska wartość współczynnika $R^2_{McFadden}$, a także to, że reszty modelu

nie mają rozkładu normalnego (test Shapiro-Wilka). W dalszych etapach badawczych, próbując poprawić jakość zbudowanego modelu logitowego, będzie podjęta próba polepszenia jakości danych wejściowych użytych do budowy modelu (np. poprzez zwiększenie rekordów wykorzystywanych do analizy).

Literatura

- Dittmann P., *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*, Oficyna Ekonomiczna, Kraków 2004.
- Gruszczyński M., Kuszewski T., Podgórska M. (red.), *Ekonometria i badania operacyjne*, PWN, Warszawa 2009.
- Mach Ł., *Econometric model structure as a support tool in real property market parameters featuring*, The 19th International Daaam Symposium „Intelligent Manufacturing & Automation: Focus on Next Generation of Intelligent Systems and Solutions”, 22-25 October 2008, Vienna 2008.
- Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2008.
- Stanisz A., *Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny*, t. 2, *Modele liniowe i nieliniowe*, Statsoft, Kraków 2007.

LOGIT MODELLING AS A TOOL SUPPORTING DECISION MAKING IN STOCK HOUSING

Summary: The article includes the logit model construction process to support the process of decision making in the real estate market. The model of logistic regression, which has been elaborated, shall define the probability of transactions in the real estate market and it will indicate statistic variables which influence the demand significantly. The process of decision making (based on logit models) – substantially prepared and correctly executed – is a key determinant having influence on improving the competitiveness of companies, especially during the global economic crisis.