Anna Barańska*

ESTYMACJA PARAMETRÓW NIELINIOWYCH MODELI FUNKCYJNYCH NA POTRZEBY PREDYKCJI RYNKOWEJ WARTOŚCI NIERUCHOMOŚCI**

1. Wprowadzenie

Podstawą modelowania jednostkowych wartości nieruchomości są informacje o cenach i cechach nieruchomości, będących przedmiotem obrotu rynkowego lub o rynkowych wartościach nieruchomości reprezentatywnych i ich atrybutach. Zbiór takich informacji stanowi podstawę konstrukcji reprezentatywnej bazy nieruchomości dla analizowanego rynku.

Baza taka powinna spełniać następujące warunki:

— liczba nieruchomości powinna być większa od 30;
— dopuszczalna różnica przeciętnej ceny z wszystkich zebranych nieruchomości analizowanego rynku i przeciętnej ceny nieruchomości reprezentatywnej bazy nie powinna przekraczać 30% jej wartości;
— dopuszczalna różnica odchylenia standardowego cen z wszystkich zebranych nieruchomości analizowanego rynku i odchylenia standardowego cen nieruchomości reprezentatywnej bazy nie powinna przekraczać 50% jej wartości;
— liczba nieruchomości z całego zebranego zbioru informacji, pochodzących z analizowanego rynku, których ceny mieszczą się w przedziałach: $\hat{c} \pm \sigma$ i $\hat{c} \mp 2\sigma$ (gdzie $\hat{c}$ oznacza wartość przeciętną, a σ oznacza odchylenie standardowe), powinny być proporcjonalne do liczby nieruchomości reprezentatywnej bazy w opowiadających przedziałach.

Reprezentatywna baza nieruchomości dla analizowanego rynku spełnia wszystkie kryteria zmiennej losowej wielowymiarowej. Do modelowania zmiennej losowej wielowymiarowej reprezentowanej przez cenę lub wartość nieruchomości i jednocześnie przez wszystkie atrybuty nieruchomości, mogą być stosowane modele liniowe oraz różnej postaci modele nieliniowe.

W dotychczasowych metodach wyceny nieruchomości *a priori* zakładano, że wartość rynkowa powstaje przez zsumowanie udziałów poszczególnych atrybutów, które to udzia-

* Akademia Górniczo-Hutnicza, Wydział Geodezji Górniczej i Inżynierii Środowiska, Katedra Informacji o Terenie
** Temat realizowany w ramach badań własnych w Katedrze Informacji o Terenie AGH w Krakowie

ły mogły być prezentowane za pomocą różnych funkcji elementarnych, również nieliniowych. Założenie to jest spełnione w szczególności przez modele liniowej regresji wielorakiej. Nasuwają się pytania, czy addytywna postać modelu zmienności wartości rynkowej, dla wszystkich rynków nieruchomości, jest optymalna, tzn. np. czy modele multiplikatywne mogą dawać wyższe dokładności wyników predykcji. Aby odpowiedzieć na to pytanie, należy dokonać analizy porównawczej różnych modeli, dla różnych rynków nieruchomości. Niniejsza praca stanowi jedynie teoretyczny wstęp, omawiający badanie postawionego problemu dla konkretnej postaci funkcji modelowej.

Spośród różnych rodzin nieliniowych funkcji wielu zmiennych, do modelowania jednostkowej ceny lub wartości nieruchomości, rozpatrzono multiplikatywną kombinację wykładniczych zmienności poszczególnych atrybutów. Wykładnicza postać funkcji zapewnia zawsze dodatnią wartość rynkową nieruchomości dla poszczególnych atrybutów oraz pozwala opisywać ich zmienność w formie funkcji monotonicznej (stale rosnącej lub stale malejącej). Dodatkową zaletą takiego modelu jest możliwość prezentacji procentowej zmienności rynkowej lub katastralnej wartości nieruchomości w odniesieniu do jej wartości przy określonych atrybutach (np. zerowych).

W dalszych badaniach będzie rozpatrywany model w formie następującej multiplikatywnej funkcji wykładniczej

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdots B_m^{x_m} \tag{1}$$

gdzie:

c – jednostkowa cena lub wartość nieruchomości z bazy,
$x_1, x_2, \ldots, x_m$ – atrybuty nieruchomości z bazy, w tym atrybut „czas transakcji",
B_j – szacowane współczynniki modelu,
B_0 – jednostkowa wartość nieruchomości, dla zerowych wartości wszystkich atrybutów.

Współczynniki B_j przyjmują wartości bliskie jedności i mają charakter czynników, które pomnażają bazową wartość nieruchomości o atrybutach zerowych. Jeżeli od wartości współczynnika B_j odejmiemy 1, to różnica ta określa współczynnik zmiany ceny nieruchomości na jednostkę j-tego atrybutu.

Estymację parametrów modelu (1) można wykonać różnymi metodami, czyli według różnych algorytmów, innymi słowy – stosując różne funkcje celu.

Można korzystać z następujących metod estymacji:
- metody najmniejszych kwadratów,
- metody Markowa (w postaci ważonej metody najmniejszych kwadratów),
- metody największej wiarygodności,
- metody momentów.

W niniejszym opracowaniu przedstawiony zostanie schemat postępowania dla ważonej metody najmniejszych kwadratów, zastosowanej do estymacji parametrów modelu (1).

2. Estymacja parametrów modelu ważoną metodą najmniejszych kwadratów

Do estymacji wartości współczynników B_j funkcję (1) trzeba doprowadzić do postaci liniowej. W tym celu logarytmujemy stronami równanie (1), przy wykorzystaniu logarytmu naturalnego, otrzymując

$$\ln c = \ln B_0 + x_1 \cdot \ln B_1 + x_2 \cdot B_2 + \ldots + x_m \cdot \ln B_m \tag{2}$$

Układ równań postaci (2) ma cechy modelu probabilistycznego, który w zapisie macierzowym przyjmuje następującą postać

$$\{\ln \delta\} + \{A\} \cdot \{\ln B\} = \{\ln c\} \tag{3}$$

gdzie:

$\{A\}$ – macierz prostokątna pionowa, zawierająca jedynki i wartości atrybutów nieruchomości z bazy,

$\{\ln c\}$ – macierz jednokolumnowa, zawierająca logarytmy naturalne cen nieruchomości z bazy,

$\{\ln B\}$ – macierz jednokolumnowa, zawierająca estymowane wartości logarytmów naturalnych współczynników B_j modelu (1),

$\{\ln \delta\}$ – macierz jednokolumnowa, zawierająca odchyłki losowe do logarytmów naturalnych cen nieruchomości z bazy,

przy czym n oznacza liczbę rozważanych nieruchomości w bazie, zaś m oznacza liczbę rozważanych atrybutów.

Układ równań (3), zapisany w pełnej formie macierzowej przybiera następująca postać

$$\begin{Bmatrix} \ln \delta_1 \\ \ln \delta_2 \\ \ln \delta_3 \\ \ldots \\ \ln \delta_n \end{Bmatrix} + \begin{Bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ 1 & x_{21} & x_{22} & \ldots & x_{2m} \\ 1 & x_{31} & x_{32} & \ldots & x_{3m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{Bmatrix} \times \begin{Bmatrix} \ln B_0 \\ \ln B_1 \\ \ln B_2 \\ \ldots \\ \ln B_m \end{Bmatrix} = \begin{Bmatrix} \ln c_1 \\ \ln c_2 \\ \ln c_3 \\ \ldots \\ \ln c_n \end{Bmatrix} \tag{4}$$

Wagę ufności dla każdej nieruchomości z bazy można ustalić np. na podstawie podobieństwa atrybutów cenotwórczych do ich średnich wartości w bazie danych (np. za pomocą odchyłek $x_i - \hat{x}_i$) lub na podstawie wartości atrybutu „źródło informacji o nieruchomościach rynkowych":

p = 0,1 – informacja własna,

p = 0,7 – przetarg, licytacja,

p = 0,6 – akt notarialny, urząd skarbowy,

p = 0,5 – oferty.

Po przyporządkowaniu wszystkim nieruchomościom występującym w bazie odpowiednich wag ufności, można zestawić macierz wag {P} w formie następującej macierzy diagonalnej

$$\{P\} = \begin{Bmatrix} p_1 & 0 & \dots & 0 & 0 \\ 0 & p_2 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & p_{n-1} & 0 \\ 0 & 0 & \dots & 0 & p_n \end{Bmatrix} \tag{5}$$

Po zastosowaniu ważonej metody najmniejszych kwadratów, przy uwzględnieniu zasady Gaussa–Markowa, otrzymuje się następujące wzory na estymowane parametry modelu nieliniowego

$$\{\ln \hat{B}\} = (\{A^T\}\{P\}\{A\})^{-1} \times \{A\}^T \{P\}\{\ln c\} \tag{6}$$

Macierz kowariancji dla estymowanych współczynników regresji wyznacza się według wzoru

$$\mathrm{Cov}\{\ln \hat{B}\} = \hat{\sigma}_0^2(\{A\}^T \{P\}\{A\})^{-1} \tag{7}$$

przy czym $\hat{\sigma}_0^2$ oznacza wariancję estymacji nieliniowego modelu multiplikatywnego, której estymator określa się według wzoru

$$\hat{\sigma}_0^2 = \frac{\{\ln c\}^T \{P\}\{\ln c\} - \{\ln \hat{B}\}^T \{A\}^T \{P\}\{\ln c\}}{n - m - 1} \tag{8}$$

Elementy na głównej przekątnej macierzy (7) stanowią kwadraty odchyleń standardowych logarytmów naturalnych poszczególnych współczynników nieliniowego modelu (1), czyli $\sigma^2(\ln B_0)$, $\sigma^2(\ln B_1)$, $\sigma^2(\ln B_2)$, ..., $\sigma^2(\ln B_m)$.

3. Procedura weryfikacji wyestymowanego modelu

W celu sprawdzenia przydatności wyestymowanego modelu należy dokonać jego weryfikacji za pomocą stosownych metod statystycznych.

Badanie dopuszczalności modelu ze względu na wartości współczynników zmienności oraz zbieżności

— Współczynnik zmienności V

$$V = \frac{\sigma_\delta}{\hat{c}} \tag{9}$$

gdzie:

σ_δ – błąd standardowy reszt,

$\hat{c}$ – wartość przeciętna cen nieruchomości z bazy,

— Współczynnik zbieżności ϕ^2

$$\phi^2 = \frac{\sum_{i=1}^{n} \delta_i^2}{\sum_{i=1}^{n} (c_i - \hat{c})^2} \tag{10}$$

gdzie:

$\delta_i = c_i - c_{i(p)}$ – reszta dla *i*-tej nieruchomości (różnica ceny obserwowanej i prognozowanej na podstawie modelu),

$\hat{c}$ – wartość przeciętna cen nieruchomości z bazy.

Wartości graniczne dla V i ϕ^2 ustala się arbitralnie. Najczęściej są to: $V_0 = 0{,}10$ oraz $\phi_0^2 = 0{,}40$.

Badanie istotności współczynników modelu

— Weryfikacja istotności układu współczynników w oparciu o statystykę *Fishera–Snedecora,* za pomocą następującej hipotezy H_0: $\sum_{j=0}^{m} B_j^2 = 0$ przeciwko hipotezie alternatywnej H_1: $\sum_{j=0}^{m} B_j^2 \neq 0$; postać statystyki

$$F = \frac{1-\phi^2}{\phi^2} - \frac{n-m-1}{m} \tag{11}$$

która przy prawdziwości hipotezy zerowej ma rozkład *F*-Snedecora o (m, $n - m - 1$) stopniach swobody.

— Weryfikacja istotności poszczególnych współczynników regresji w oparciu o statystykę *T*-Studenta, przy hipotezie zerowej H_0: $B_j = 0$ przeciwko hipotezie alternatywnej H_1: $B_j \neq 0$; postać statystyki

$$T = \frac{\ln \hat{B}_j}{\sigma(\ln \hat{B}_j)} \tag{12}$$

która przy prawdziwości hipotezy zerowej ma rozkład *T*-Studenta o ($n - m - 1$) stopniach swobody.

Jeżeli dla którejś ze zmiennych objaśniających test statystyczny nie wykaże podstaw do odrzucenia hipotezy zerowej – usuwamy tę zmienną z modelu i wykonujemy ponowną estymację parametrów.

Badanie symetrii składnika losowego

Niech k oznacza liczbę reszt dodatnich (lub ujemnych) δ_i. Weryfikujemy hipotezę dotyczącą wskaźnika struktury reszt o ustalonym znaku (tzw. frakcji p) następującej postaci: H_0: $p = 0{,}5$, przeciw hipotezie alternatywnej H_1: $p \neq 0{,}5$.

Statystyka testowa ma postać

$$T = \frac{\left|\frac{k}{n} - 0{,}5\right|}{\sqrt{\dfrac{\frac{k}{n}\left(1 - \frac{k}{n}\right)}{n-1}}} \tag{13}$$

która dla prób mało licznych (n 60 30) ma rozkład T-Studenta, a dla prób licznych – rozkład normalny.

Jeżeli test wskaże na konieczność odrzucenia hipotezy zerowej – należy zmodyfikować postać analityczną modelu.

4. Prognozowanie jednostkowej wartości rynkowej nieruchomości

Prognozę jednostkowej ceny c_p (rynkowej wartości) dowolnej nieruchomości wyznacza się poprzez podstawienie do wyznaczonego modelu postaci (2) wartości atrybutów a_j. Po przekształceniu uzyskujemy wartość prognozy dla analizowanej nieruchomości

$$c_p = \exp(\ln B_0 + a_1 \cdot \ln B_1 + a_2 \cdot \ln B_2 + ... + a_m \cdot \ln B_m) \tag{14}$$

Odchylenie standardowe logarytmu naturalnego prognozowanej ceny wyznacza się na podstawie macierzy kowariancji (7) dla współczynników zlinearyzowanego modelu, według następującego wzoru

$$\sigma^2(\ln c_p) = \left(\begin{Bmatrix} 1 \\ a_1 \\ a_2 \\ \ldots \\ a_m \end{Bmatrix}^T \mathrm{Cov}\{\ln \hat{B}\} \begin{Bmatrix} 1 \\ a_1 \\ a_2 \\ \ldots \\ a_m \end{Bmatrix} \right) \tag{15}$$

Odchylenie standardowe prognozy, zgodnie z prawem przenoszenia się błędów, wyliczamy według wzoru

$$\sigma(c_p) = c_p \cdot \sigma(\ln c_p) \tag{16}$$

Na podstawie wykonanej analizy wariancji można stwierdzić, że prognozowana cena (wartość) analizowanej nieruchomości, na poziomie ufności $(1 - \alpha) = 0{,}95$, powinna zawierać się w następującym przedziale

$$p_p \pm t(0{,}975; n - m - 1) \cdot \sigma(c_p) \tag{17}$$

przy czym $t(0{,}975;\ n - m - 1)$ oznacza stosowny kwantyl rozkładu T-Studenta.

Aby prognoza była wiarygodna i możliwa do wykorzystania w wycenie nieruchomości, należy przestrzegać następujących dwóch zasad:

1) wartości atrybutów a_j wycenianej nieruchomości muszą zawierać się w przedziałach zmienności odpowiadających atrybutów nieruchomości w bazie,
2) szerokość przedziału ufności nie powinna przekraczać połowy szacowanej wartości rynkowej nieruchomości.

Wskaźnik wiarygodności wyceny, obejmujący niedokładność wszystkich parametrów występujących w modelowaniu ceny nieruchomości, jest wyrażony przez współczynnik rozproszenia w postaci:

$$\lambda = \frac{t\left(1-\frac{\alpha}{2}; n-m-1\right)\cdot \sigma(c_p)}{c_p} \tag{18}$$

Na podstawie zaproponowanej skali dla współczynnika rozproszenia, rzeczoznawca może dokonać oceny stopnia wiarygodności szacowanej wartości nieruchomości, czyli:

$\lambda \le 0{,}05$ – wiarygodność bardzo wysoka,
$0{,}05 < \lambda \le 0{,}10$ – wiarygodność wysoka,
$0{,}10 < \lambda \le 0{,}15$ – wiarygodność dość wysoka,
$0{,}15 < \lambda \le 0{,}20$ – wiarygodność dostateczna,
$0{,}20 < \lambda \le 0{,}25$ – wiarygodność dopuszczalna.

Jeżeli $\lambda > 0{,}25$, to należy stwierdzić, że wycena nieruchomości na podstawie estymowanych współczynników nieliniowej regresji jest niedopuszczalna. W takim przypadku należałoby powtórnie zweryfikować bazę i dokonać ponownej estymacji parametrów modelu.

5. Podsumowanie

Podobną analizę można wykonać dla innych metod estymacji parametrów modelu. Równość wyestymowanych różnymi metodami parametrów można zweryfikować za pomocą stosownych testów statystycznych. W ramach porównania metod estymacji należałoby również porównać macierze kowariancji dla parametrów uzyskanych z różnych metod.

Postawiony problem będzie przedmiotem dalszych badań autorki.

Literatura

[1] Barańska A.: *Kryteria stosowania modeli stochastycznych w predykcji rynkowej wartości nieruchomości*. Kraków, Akademia Górniczo-Hutnicza, Wydział Geodezji Górniczej i Inżynierii Środowiska 2003

[2] Czaja J.: *Metody szacowania wartości rynkowej i katastralnej nieruchomości*. Kraków, 2001

[3] Czaja J., Preweda E.: *Analiza statystyczna zmiennej losowej wielowymiarowej w aspekcie korelacji i predykcji*. Półrocznik AGH Geodezja, t. 6, z. 2, 2000

[4] Krysicki W. i in.: *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. Warszawa, PWN 1986