

Małgorzata Renigier

Budowa modelu geostatystycznego z wykorzystaniem reszt

Acta Scientiarum Polonorum. Administratio Locorum 4/1/2, 83-96

2005

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

BUDOWA MODELU GEOSTATYSTYCZNEGO Z WYKORZYSTANIEM RESZT

Małgorzata Renigier

Katedra Gospodarki Nieruchomościami i Rozwoju Regionalnego
Uniwersytet Warmińsko-Mazurski w Olsztynie

Streszczenie. Diagnozowanie wartości przestrzeni na podstawie modelowania statystycznego, zawierającego reguły przyczynowo-skutkowe jest często niepełne i niewystarczające do osiągnięcia prawidłowych wniosków oraz podejmowanych na ich podstawie decyzji. Włączenie składnika niewyjaśnionego przez model (reszt) do budowy modelu, stanowi istotne dopełnienie formuł statystycznych (deterministycznych) pozbawionych elementu niepewności i losowości.

Artykuł obejmuje zagadnienia związane z budową modelu geostatystycznego z wykorzystaniem składnika resztowego.

Słowa kluczowe: reszty z modelu, model geostatystyczny, etapy budowy modelu, nieliniowość i niepewność w analizach przestrzennych.

WSTĘP

Człowiek, ilekroć się zastanawiał, jak jest zbudowany świat, zawsze starał się powiązać wiele obserwacji dotyczących różnych dziedzin, doszukiwał się ich wzajemnych związków, próbował ustalić kolejność następstw poszczególnych zjawisk. W tym momencie pojawiła się myśl budowy modeli imitujących zaobserwowane zachowania danego zjawiska [Ostasiewicz 1999].

Zgodnie z teorią Bertalanffy'ego [1984] jedynym sensownym sposobem badania organizacji jest jej badanie jako systemu, natomiast opisem danego systemu jest właśnie model. Przyjmując założenie, że system jest to byt

przejawiający egzystencję przez synergiczną interakcję swych elementów działających w określonym czasie i w przestrzeni, można stwierdzić, że model jest uproszczoną reprezentacją systemu w czasie i w przestrzeni, stworzoną z zamiarem zrozumienia zachowania systemu rzeczywistego [Cempel 2003] i przekazującego raczej uniwersalne cechy obserwowanych bytów.

Podstawą modelowania jest postulowanie istnienia praw ekonomicznych i przyrodniczych oraz możliwość ich odtworzenia na podstawie informacji historycznych. Jest to postępowanie twórcze polegające na idealizacji i syntezie zjawisk świata rzeczywistego. Istnieją dwie główne funkcje modelowania rzeczywistości. Pierwsza jest traktowana jako swoiste narzędzie, rodzaj rzemiosła do podejmowania decyzji i działań, druga zaś jest postrzegana jako fundamentalny nośnik teorii. Wobec tego bardzo ważną staje się odpowiedź na pytanie: czy model w sposób adekwatny prezentuje aspekty rzeczywistości i czy w danej sytuacji jest odpowiednim narzędziem analitycznym?

Nie zawsze dobrze skonstruowany model może być użyty w każdej sferze problemowej. Z uwagi na pracochłonność i kosztochłonność procesów analiz statystycznych starano się tworzyć modele w miarę uniwersalne i możliwe do wykorzystania przy różnych okazjach. Wówczas jednak można było napotkać na problem związany z procesem dynamicznej zmiany zjawisk, dla których wcześniejsze założenia nie są adekwatne i ze względu na ograniczenia czasowe i przestrzenne niemożliwe było przekonstruowanie formalnego obrazu (modelu). Model bowiem jest często tworzony na podstawie konkretnego zjawiska interpretowanego przecież indywidualnie. Ważną kwestią wobec tego jest umiejętność i intuicja analityka, od którego głównie zależy powodzenie operacji.

Tradycyjnie mówi się o adekwatności modelu do opisanej przez niego rzeczywistości. Może jednak zbudowany model bardziej odzwierciedla problem, sposób jego widzenia i związane z tym sposobem wyobrażenia o rzeczywistości (a więc ukształtowany wcześniej w heterogenicznej sferze rozumowania jakościowego pewien twór natury epistemologicznej) niż rzeczywistość jako taką. Problem zgodności z obiektywną rzeczywistością pojawia się zanim powstaną równania i nie może być w pełni rozstrzygany w języku matematyki [Barteczko, Bocian 1996]. Są oczywiście pewne punkty stałe odzwierciedleń rzeczywistości, których weryfikacja powinna być przeprowadzona za pomocą narzędzi statystyczno-matematycznych. Pełna i obiektywna weryfikacja problemu powinna być dokonana z uwzględnieniem natury heterogenicznej rzeczywistości i będzie tym prostsza i wiarygodniejsza, im model będzie bardziej zbliżony do natury zjawiska.

MODELOWANIE ZJAWISK PRZESTRZENNYCH W UJĘCIU ANALIZY RESZT

Procesy społeczno-gospodarcze często są ujmowane w postaci zaawansowanych modeli matematycznych. Posługiwanie się modelami statystycznymi ma służyć poprawie, ujednoliceniu i usprawnieniu procesów rozwoju gospodarki rynkowej.

Trudno jest zbudować model, który idealnie odzwierciedlałby realne relacje i jednoznacznie je przewidywał. Zaletą modeli jest to, iż nie odpowiadając „kategorycznie” na zadane pytanie, pozwalają na syntezę rozproszonej i niejednorodnej informacji i w konsekwencji identyfikację nowej informacji a priori, będącej swoistą kompilacją wcześniej zgromadzonych danych.

W pracy zaproponowano nazwę model geostatystyczny w celu podkreślenia źródła analizowanego zjawiska i wyodrębnienia przedmiotu badań. To pojęcie i wyodrębnienie z pozostałych modeli nie sprowadzałoby się jedynie do nazwy, ale wyodrębniłoby i usystematyzowało stosowane metody i założenia właściwe dla tej dziedziny badań, w której podstawową rolę odgrywa czynnik przestrzeni.

Stworzony w ramach geostatystycznej analizy struktur przestrzennych model należy traktować nie jako rozstrzygające kryterium, ale jako narzędzie tworzące dodatkowe informacje na kanwie wyliczeń matematycznych „nieskalanych” cechą arbitralności, zgodnie ze stwierdzeniem Johna Tukeya [1998], iż: „lepiej dać przybliżoną odpowiedź na interesujące, czasami niedokładnie zadane pytanie, niż dać poprawną odpowiedź na pytanie nieciekawe i precyzyjnie zadane”.

Jednym z lepszych narzędzi pozwalających uwzględnić wymienione aspekty związane z odzwierciedlaniem rzeczywistości w postaci modeli są reszty z modelu, reprezentujące jakość modelu oraz zawierające informacje o zjawiskach niewyjaśnionych przez model. Reszty można traktować, jako wyraz niepewności w procesie modelowania. Stanowią one, bowiem formalną miarę różnic między zjawiskami specyficznymi a uogólnieniami, jakie dają tworzone modele, zgodnie z ogólną formułą:

$$e_i = y_i - \hat{y}_i \quad (1)$$

gdzie:

e_i – reszta z modelu,

y_i – wartość empiryczna (obserwowana),

$\hat{y}_i$ – wartość estymowana.

Wartość estymowana może przyjmować różne postacie. W pracy przedstawiony zostanie proces określania wartości $\hat{y}_i$.

Reszty są obszarem zjawisk niewyjaśnionych przez model z uwagi na brak odpowiednich danych statystycznych, brak wiedzy merytorycznej i praktycznej na temat modelowanego obszaru badań lub/i z powodu dynamicznego, często nieprzewidywalnego tempa zmian badanego zjawiska. Wykluczenie zatem lub wyeliminowanie dwóch pierwszych wspomnianych wyżej możliwości interpretacji reszt pozwoli na wykorzystanie ich do ujawnienia i zgłębienia komplementarnych informacji na temat analizowanych zjawisk [Renigier i Żróbek 2005].

W pracy przyjęto założenie, że model geostatystyczny reprezentuje ogólną zmienność wynikającą z przesłanek natury deterministycznej. Włączenie reszt do procesu budowy modelu pozwala na minimalizację nieprawidłowości związanych z formalną strukturą modelu. Celem pracy było dopełnienie formuł statystycznych (deterministycznych) pozbawionych elementu niepewności i losowości przez włączenie składnika niewyjaśnionego przez model (reszt) do budowy modelu.

PROCEDURA TWORZENIA MODELU GEOSTATYSTYCZNEGO Z WYKORZYSTANIEM ANALIZY RESZT

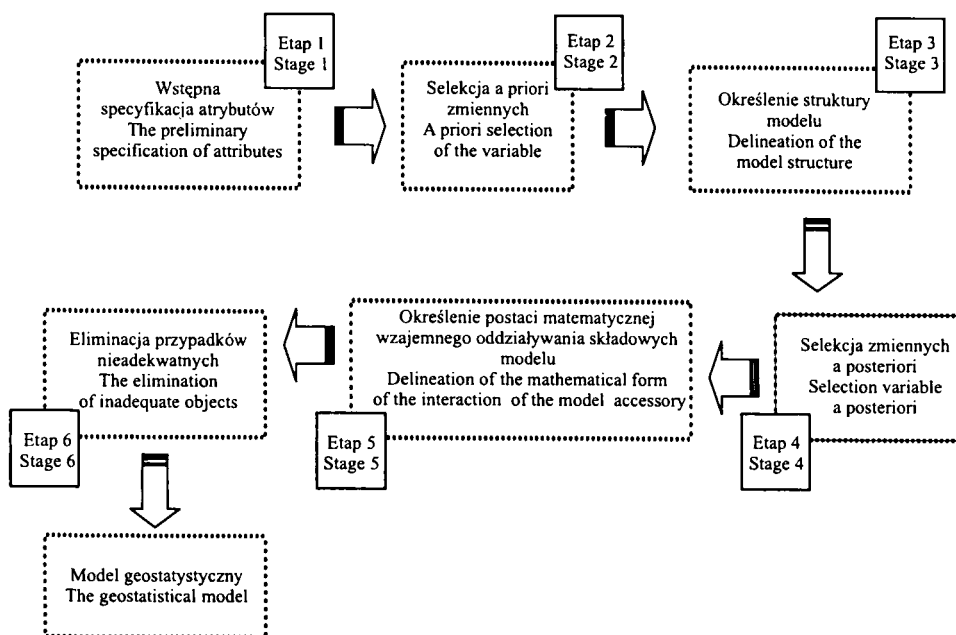
Głównym założeniem przy tworzeniu modelu geostatystycznego było najwierniejsze odzwierciedlenie czynników determinujących obraz analizowanych procesów przestrzennych. Zastosowana metodologia badawcza i narzędzia statystyczne do jego realizacji mają służyć minimalizacji możliwości popełnienia błędów formalnych związanych z budową struktury modelu. Zaproponowana metodologia jest na tyle plastyczna, iż może być stosowana do diagnozowania skutków różnych procesów przestrzennych. W artykule przyjęto założenie, że wyrazem zachodzących procesów przestrzennych jest kształtowanie się wartości przestrzeni, która jest odzwierciedlana później w osiąganej na rynku cenie za nieruchomości gruntowe.

Systemy przestrzenne są zazwyczaj niezwykle złożone i heterogeniczne w swej strukturze i działaniu. Przedstawiony niżej schemat tworzenia modelu uwzględnia różnego rodzaju powiązania zjawisk diagnozowanej struktury przestrzennej, przy dopuszczeniu zależności nieliniowych, jako że są one częstym czynnikiem występującym w relacjach złożonych zjawisk. Rysunek 1 przedstawia schemat realizacji kolejnych etapów budowy modelu geostatystycznego.

Poszczególne etapy budowy modelu zrealizowano na przykładzie olsztyńskiego rynku nieruchomości. Charakterystykę obiektu badań przedstawiono w tabeli 1.

Etap 1 – Wstępna specyfikacja atrybutów

Przed przystąpieniem do realizacji kolejnych etapów budowy modelu należy dokładnie przeanalizować badane zjawisko. W praktyce przejawia się to doбором takiego zbioru cech opisujących zjawisko, aby zmaksymalizować pojemność informacyjną wytypowanych zmiennych związanych z przedmiotem badań. Taki zbiór zmiennych powinien być odpowiednio skwantyfikowany i mieć określony poziom wiarygodności oraz zachowywać właściwą precyzję. W artykule przyjęto 25 zmiennych niezależnych (objaśniających) opisujących kształtowanie się wartości na olsztyńskim rynku nieruchomości (zmienna objaśniana – cena nieruchomości). Do dalszych badań przyjęto zmienne, których nazewnictwo przedstawiono w tabeli 2.



Rys. 1. Schemat realizacyjny etapów budowy modelu geostatystycznego z wykorzystaniem reszt

Fig. 1. Realization scheme of the stages of the construction of a geostatistical model using residuals

Źródło: opracowanie własne
Source: own elaboration

Tabela 1. Charakterystyka obiektu badań

Table 1. The characterization of the object of research

Obiekt badań Object of research	Zmienna objaśniana (zależna) Explanatory variable (dependent)	Okres analiz Term of analyses	Liczba obserwacji Number of observations	Liczba atrybutów* Number of attributes*
Olsztyn - nieruchomości gruntowe niezabudowane Olsztyn – vacant, land real estates	cena nieruchomości value of real estates	1997-2003	821	25

* szczegółowa charakterystyka przyjętych atrybutów w pracy Renigier 2004c. przy założeniu, że: $i = 1, 2 \dots k$ (poziomy czynnik klasyfikacyjny) oraz $j = 1, 2, \dots, n_j$ (numery jednostek obserwacji w i -tym poziomie czynnika klasyfikacyjnego).

* detailed characterization of accepted attributes in dissertation Renigier 2004c

Źródło: badania własne

Source: own studies

Etap 2 – Selekcja wytypowanych zmiennych a priori

Przystępując do etapu selekcji zmiennych a priori należy przeprowadzić wstępną analizę wytypowanych zmiennych zgodnie z przyjętymi założeniami konstrukcji modelu. Nie negując, ale podważając fakt katerycznego założenia powiązań liniowych między zmiennymi w modelu, optymalny zbiór zmiennych można wytypować, stosując analizę **jednoczynnikowej wariancji**, wykorzystując zależność:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (2)$$

gdzie:

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ – to całkowita suma kwadratów odchyłeń od średniej ogólnej (*SST*),

$\sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i$ – to suma kwadratów odchyłeń średnich grupowych od średniej ogólnej (*SSB*),

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ – to suma kwadratów odchyłeń wartości cechy od średniej grupowej (*SSE*).

Analizę jednoczynnikowej wariancji wykorzystano w celu określenia występowania istotnego wpływu analizowanej zmiennej niezależnej na zmienną objaśnianą. Wskaźnik korelacji Pearsona, który jest odporny na związki krzywoliniowe zastosowano w celu określenia siły wzajemnych powiązań między zmiennymi. Wskaźnik ten liczony jest według formuły:

$$e_{yx} = \sqrt{\frac{SSB}{SST}} \quad (3)$$

W pracy, po zastosowaniu wyżej wymienionych miar, do dalszych badań wyselekcjonowano 18 zmiennych (szczegółowo temat omówiono w pracy Renigier 2004b i 2004c). Wyniki przedstawiono w tabeli 2.

Etap 3 – Określenie struktury modelu

Kolejnym etapem budowy modelu, po wyłonieniu wstępnego (*a priori*) zbioru zmiennych, jest określenie zależności między cechami i określenie postaci funkcyjnej modelu. W pracy zaproponowano dobór postaci funkcyjnej poprzez analizę indywidualnych relacji między parą zmiennych (czyli zmien-

Tabela 2. Wynik wstępnej selekcji zmiennych
Table 2. Result of the preliminary selection of variables

MS – metody selekcji, AW – analiza wariancji, AK – analiza korelacyjna, W – wynik selekcji
MS – methods of selection, AW – analysis variation, AK – correlational analysis, W – result of selection

MS	Zmienne – Variables																								
	Pow	Data	Liczba dz	F wład	Front	Głęb	Kształt	Usyt	E	W	G	T	K	UT	DD	Ud	Uk	Ui	PZP	F trans	Atr-Woda	Atr-Las	Odl H	Odl C	Odl K
AW	*	*			*	*			*	*	*	*	*	*	*	*			*		*	*	*	*	*
AK	*	*			*	*			*	*	*	*	*	*	*	*			*		*	*	*	*	*
W	*	*			*	*			*	*	*	*	*	*	*	*			*		*	*	*	*	*

Źródło: badania własne
Source: own studies

symbole zmiennych niezależnych zgodnie z oznaczeniami:
independent variables symbols according to designation

pow – powierzchnia;
pow – area;
data – data transakcji;
data – date of transaction;
liczba dz – liczba działek będąca przedmiotem jednej transakcji;
liczba dz – number of parcels;
F wład – forma władania nieruchomością;
F wład – form of possession the real estate;
front – front działki;
front – front of parcel;
głęb – głębokość działki;
głęb – depth of parcel;
kształt – kształt działki;
kształt – shape of parcel;
usyt – usytuowanie działki;
usyt – location of parcel;
E – wyposażenie w sieć elektroenergetyczną;
E – equipment into electrical energy;
W – wyposażenie w sieć wodociagową;
W – equipment into water – supply system;
G – wyposażenie w sieć gazową;
G – equipment into gas system;
T – wyposażenie w sieć telekomunikacyjną;
T – equipment into phone;

K – wyposażenie w sieć kanalizacyjną;
K – equipment into sewage system;
UT – ukształtowanie terenu;
UT – form of area;
DD – dojazd i dostęp do działki;
DD – approach and access to parcel;
Ud – uciążliwość sąsiedztwa związana z ruchem kołowym ;
Ud – difficulty of neighborhood with traffic related;
Uk – uciążliwość sąsiedztwa związane z liniami kolejowymi;
Uk – difficulty of neighborhood with rail lines related;
Ui – inne uciążliwości sąsiedztwa związane np. z zakładem przemysłowym;
Ui – other difficulty of neighborhood;
PZP – przeznaczenie w planie;
PZP – allocation in local plan;
F trans – forma transakcji;
F trans – form of transaction;
Atr woda – atrakcyjność położenia w stosunku do obszarów leśnych i parków;
Atr woda – the attractiveness of position in relation to forest areas and parks;
Atr las – atrakcyjność położenia w stosunku do obszarów zbiorników wodnych;
Atr las – the attractiveness of position in relation to areas of water reservoirs;
Odl H – lokalizacja związana z odległością od najważniejszego ośrodka handlowo – usługowego;
Odl H – localization connected with the distance from the main commerce;
Odl C – lokalizacja związana z odległością od centrum miasta;
Odl C – localization connected with the distance from the city centre;
Odl K – lokalizacja związana z odległością od komunikacji ogólnej;
Odl K – localization connected with the distance from the main means of transport.

ną zależną i zmienną niezależną), tworząc model wielofunkcyjny [Renigier 2004a].

W praktyce należy przeprowadzić analizę wrażliwości wzajemnych relacji pary zmiennych. W pracy wykorzystano 7 najczęściej wykorzystywanych funkcji w tego typu analizach: liniową, hiperboliczną, logarytmiczną, wykładniczą, kwadratową oraz wielomian trzeciego i czwartego stopnia. Za kryterium jakości wzajemnych relacji uznano wyniki mierników opierających się w swej strukturze na resztach z modelu, tj.:

a) sumie kwadratów reszt (SSe):

$$SSe = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

b) adekwatność modelu (A):

$$A = \frac{\sum_{i=1}^n \sqrt{\left(\frac{y_i - \hat{y}_i}{y_i} \right)^2}}{n} \times 100\% \quad (5)$$

c) błędzie średniokwadratowym (SSq):

$$SSq = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (6)$$

gdzie:

y_i – zmienna obserwowana,

$\hat{y}_i$ – zmienna obliczona,

n – liczba przypadków.

Do estymacji modelu wykorzystano algorytm Levenberga-Marguardta, czyli gradientową, nieliniową metodę najmniejszych kwadratów, która jest adekwatna do warunków przeprowadzenia analiz w ramach prac wstępnych.

Postać modelu wielofunkcyjnego przedstawia się następująco:

$$\begin{aligned} \hat{y} = & a + \sum_{i=pow, gleb, front} a_i \log x_i + \sum_{j=UT} a_j x_j + \sum_{k=data, G, T, DD, Ud, PZP} (a_k x_k + b_k x_k^2) \\ & + \sum_{l=K, W, E} (a_l x_l + b_l x_l^2 + c_l x_l^3) + \sum_{m=atrW, atrL, odlH, odlC, odlK} (a_m x_m + b_m x_m^2 + c_m x_m^3 + d_m x_m^4) \end{aligned} \quad (7)$$

Dopasowanie i efektywność modelu (7) porównano z innymi modelami regresji. Wyniki analiz przedstawia tabela 3.

Tabela 3. Analiza efektywności i dopasowania modelu wielofunkcyjnego
 Table 3. The analysis of the effectiveness and adjustment of a multifunctional model

Rodzaje modeli Sort of models	Mierniki dopasowania modelu Measures of the model adjustment			
	<i>Sse</i>	R^2 *	<i>A</i>	<i>SSq</i>
Model wielofunkcyjny regresji nieliniowej The multifunctional model of the non-linear regression	922.588	0,57	57	33
Model regresji liniowej The model of the linear regression	1.369.591	0,38	87	42
Model regresji wykładniczej, nieliniowej The model of the exponential regression, non-linear	2.157.337	0,30	91	51
Model regresji kwadratowej, nieliniowej The model of the square regression, non-linear	1.006.813	0,52	74	35

* R^2 (współczynnik determinacji) zastosowano jako dodatkowy miernik dopasowania

* R^2 (coefficient of determination) applied as additional measure of the adjustment

Źródło: badania własne

Source: own studies

Etap 4 – Selekcja zmiennych a posteriori

Wybór optymalnego zbioru cech jest jednym z najistotniejszych etapów przy budowie modelu. Nie zawsze wytypowany optymalny zbiór zmiennych przed budowaniem modelu jest optymalnym zbiorem po określeniu jego struktury. W związku z tym zaproponowano dodatkową procedurę selekcji zmiennych a posteriori, z wykorzystaniem reszt z modelu (jako mierników dopasowania), zgodnie z zasadą maksymalizacji pojemności informacyjnej i minimalizacji ilościowej optymalnego zbioru.

W praktyce polega to na analizie wrażliwości mierników określających dopasowanie modelu, w przypadku eliminacji poszczególnych zmiennych (pojedynczo). Z modelu wyeliminowano te zmienne, które po symulacyjnym usunięciu nie zmieniały lub poprawiały mierniki dopasowania modelu. Proces ten miał na celu wyeliminowanie zjawiska nadmiarowości zmiennych. Wyniki analiz przedstawiono w tabeli 4 (zmienne oznaczone pogrubioną czcionką odrzucono z dalszych analiz).

Etap 5 – Określenie postaci funkcyjnej wzajemnego oddziaływania składowych modelu

Inspiracją do stworzenia tego etapu była analiza tendencji rozwojowych w teorii prognozowania. Można założyć, iż wzajemne relacje występujące między zmiennymi objaśniającymi nie są tylko addytywne. Występować mogą także zależności multiplikatywne (wskaznikowe). Wobec tego selekcję wyboru optymalnej postaci składników do analizowanego modelu statystycznego można wyznaczyć po przeprowadzeniu analizy wrażliwości wyników dopasowania, w reakcji na multiplikatywną formę występowania zmiennych w modelu.

Tabela 4. Analiza wrażliwości modelu na odrzucenie poszczególnych zmiennych
 Table 4. The sensitivity analysis of the model to the rejection of individual variables

Zmienna Variable	Wyniki mierników dopasowania po odrzuceniu zmiennej Results of adjustment measures after variable rejection		Symbol wyniku analizy wrażliwości Symbol of sensitivity analysis result
Pow	SSe= 940.377 $R^2 = 0,54$	A = 59 SSq= 34	↓
Data	SSe= 958.321 $R^2 = 0,53$	A = 64 SSq= 34	↓
Front	SSe= 922.598 $R^2 = 0,57$	A = 58 SSq= 33	=
Głęb	SSe= 923.346 $R^2 = 0,57$	A = 57 SSq= 34	=
E	SSe= 932.717 $R^2 = 0,57$	A = 61 SSq= 34	↓
W	SSe= 923.749 $R^2 = 0,57$	A = 65 SSq= 33	↓
G	SSe= 932.383 $R^2 = 0,57$	A = 54 SSq= 34	↑
T	SSe= 955.461 $R^2 = 0,55$	A = 93 SSq= 34	↓
K	SSe= 958.321 $R^2 = 0,53$	A = 64 SSq= 34	↓
UT	SSe= 922.833 $R^2 = 0,57$	A = 55 SSq= 34	↑
DD	SSe= 937.843 $R^2 = 0,56$	A = 78 SSq= 34	↓
Ud	SSe= 929.147 $R^2 = 0,57$	A = 58 SSq= 34	=
PZP	SSe= 962.248 $R^2 = 0,55$	A = 57 SSq= 34	↓
Atr woda	SSe= 930.969 $R^2 = 0,57$	A = 78 SSq= 34	↓
Atr las	SSe= 958.631 $R^2 = 0,55$	A = 80 SSq= 34	↓
Odl H	SSe= 962.036 $R^2 = 0,55$	A = 61 SSq= 34	↓
Odl C	SSe= 993.486 $R^2 = 0,53$	A = 132 SSq= 34	↓
Odl K	SSe= 928.259 $R^2 = 0,57$	A = 57 SSq= 34	=

↓ symbol spadku dopasowania modelu po odrzuceniu zmiennej

↓ symbol of the decrease of the adjustment of the model after the rejection of variable

↑ symbol wzrostu dopasowania modelu po odrzuceniu zmiennej,

↑ symbol of the increase of the adjustment of the model after the rejection of variable

= symbol braku większego wpływu na dopasowanie modelu po odrzuceniu zmiennej

= symbol of the lack of the greater influence on the adjustment of the model after the rejection of variable

Źródło: badania własne

Source: own studies

W praktyce oznacza to, iż te zmienne występują w modelu w postaci multiplikatywnej, dla których przeprowadzona symulacja wskazała poprawę wartości mierników dopasowania modelu. W efekcie przeprowadzanych analiz część zmiennych modelu zapisano w formie addytywnej. Pozostałe zaś, czyli te, które nakładają się na wymieniony „trend” wskaźnikowo, zapisano w formie multiplikatywnej – wzór 8.

$$\hat{y} = a + \sum f(x_{pow}, x_E, x_W, x_K, x_{DD}, x_{atrW}, x_{atrL}) \cdot \Pi(1 + f(x_T, x_{data}, x_{PZP}, x_{odlH}, x_{odlC})) \quad (8)$$

Wartość mierników dopasowania modelu, po zastosowaniu wyżej przedstawionej procedury, przedstawia tabela 5. Efektywność zastosowanej procedury potwierdzono także poprzez jej zastosowanie w modelu liniowym.

Tabela 5. Wartość mierników dopasowania modeli po etapie 5
Table 5. The value of adjustment measures of the models after stage 5

Rodzaj modelu ze względu na postać matematyczną jego części składowych Type of model for the mathematical form of its component parts	Typ modeli regresji Type of regression models							
	Model regresji wielorakiej liniowej The model of the lineal regression				Model wielofunkcyjny regresji nieliniowej The multifunctional model of non-linear regression			
	SSe	R2	A	SSq	SSe	R2	A	SSq
Model addytywny Additive model	1.238.974	0,41	58	39	947.941	0,56	51	34
Model z „trendem” Model with "trend"	1.118.961*	0,48	51	43	793.986	0,69	38	31

* model regresji liniowej po wprowadzeniu multiplikatywnych zmiennych (czyli po utworzeniu modelu z „trendem” można uważać za model nieliniowy)

* model of the lineal regression after the introduction multiplicative variable (that is after the creation of the model with "trend" can consider as a non-linear model)

Źródło: badania własne

Source: own studies

Etap 6 – Eliminacja przypadków nieadekwatnych

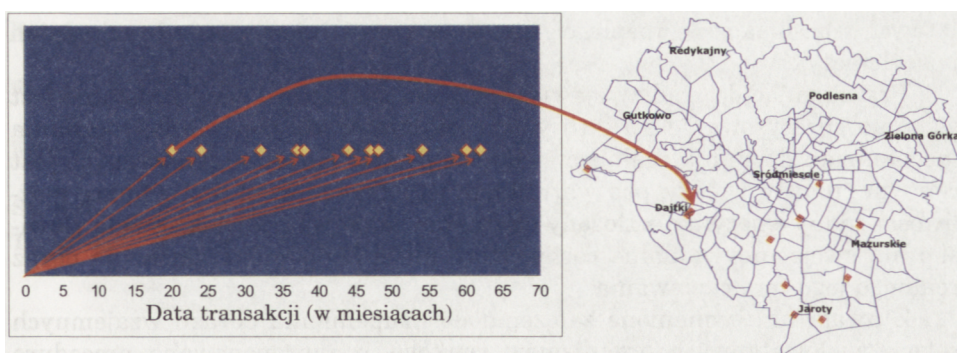
Nie zawsze przypadki odstające są obiektami „nieadekwatnymi”, często są to zjawiska związane z właściwością badanego zjawiska oraz z losowością systemu rynkowego. Przyjęte miary odstawania wykorzystujące reszty jako kryterium dopasowania, powinny rozstrzygnąć wymienione wątpliwości. Zakłada się bowiem, iż najlepszymi obiektami do ustalenia przypadków odstających są reszty [Statsoft 2003].

Do wyłonienia przypadków odstających wykorzystano analizę reszt. Posłużono się tzw. „regułą trzech sigm” (plus/minus trzy odchylenia standardowe od średniej) oraz technikami taksonomicznymi (w tym techniką aglome-

racyjną i analizą skupień), które wykorzystano do klasyfikacji odstających jednostek przestrzennych. Technikę aglomeracyjną zastosowano do określenia liczby skupień. W tej technice wykorzystano tzw. diagram drzewa zbudowany efektywną metodą Warda, która wykorzystuje do oszacowania odległości między skupieniami podejście analizy wariancji. Przy wyodrębnianiu liczby grup miano na uwadze jak najogólniejszy podział zbioru. Samą analizę skupień przeprowadzono poprzez wykorzystanie metodą grupowania k -średnich, które należą do tzw. optymalizacyjnych procesów. Procedury analizy skupień pozwalają utworzyć grupy (skupienia) „najmniej odległych od siebie” lub „najbardziej podobnych do siebie” obiektów traktowanych jako punkty wielowymiarowej przestrzeni. Podstawowym elementem analizy skupień jest pojęcie odległości taksonomicznej, przez którą rozumie się odległość między punktami wielowymiarowej przestrzeni.

Ostatecznym potwierdzeniem zakwalifikowania wyłonionych przypadków do grupy przypadków nieadekwatnych (hiperodstających) jest przeprowadzenie analizy przestrzenno-czasowego występowania przypadków. W zgromadzonym zbiorze obserwacji przypadki uznaje się ostatecznie za odstające, jeśli wyłonione wcześniej obiekty nie występują w jednym czasie ani w jednym miejscu w przestrzeni, czyli są efektem działania czynników losowych. W pracy wyłoniono 11 przypadków (rys. 2), które uznano za hiperodstające i wyeliminowano z dalszych analiz.

Wyniki wyłączenia przypadków „hiperodstających” na wartość mierników dopasowania modelu przedstawia tabela 6. Eliminacja wytypowanych 11 przypadków spośród 821, zmniejszyła odchylenie reszt (SSE) o prawie 50%, co wpłynęło jednocześnie na poprawę pozostałych mierników dopasowania.



Rys. 2. Analiza przestrzenno-czasowa występowania przypadków określonych jako odstające

Fig. 2. Spatial-temporary analysis of the occurrence of definite objects as outlying

Źródło: opracowanie własne

Source: own elaboration

Tabela 6. Wyniki wpływu odrzucenia przypadków hiperodstających na wartość mierników modelu

Table 6. The results of influence of hiperoutlying objects rejection on value of model measures

Geostatystyczny model z „trendem” Geostatistical model with „trend”				
Mierniki dopasowania modelu Measures of the model adjustment	SSe	R ²	A	SSq
Geostatystyczny model z „trendem” przed eliminacją przypadków hiperodstających Geostatistical model with "trend" before hiperoutlying objects rejection.	793.986	0,69	38	31
Geostatystyczny model z „trendem” po eliminacji przypadków hiperodstających Geostatistical model with "trend" after hiperoutlying objects rejection.	421.397	0,76	29	22

Źródło: badania własne

Source: own studies

PODSUMOWANIE

Istnieją opinie o braku miarodajności wyników modelowania i ogólne zniechęcenie do posługiwania się tym narzędziem nauki do przedstawiania „prawdy”. Wynika to być może ze złego interpretowania funkcji modelowania i przypisywaniu modelom zbyt dużej odpowiedzialności. Tymczasem model jest właśnie po to, by móc pokazać nie tyle wyniki, co warunki ich uzyskania, założenia i ich konsekwencje oraz logikę rozumowania.

Modelowanie nie jest gotową odpowiedzią, ale raczej sposobem poszukiwania nowych tez i pytań. To narzędzie badań jest nie tyle uzyskiwaniem łatwych rozwiązań, ile inspiracją do poszukiwań innowacyjnych zagadnień i problematyk.

Przystępując do analizy zjawisk gospodarczo-przestrzennych, istotne jest wstępne sprecyzowanie struktury ich działania i rozwoju. Próba poznania mechanizmu działania realnego świata w postaci modeli statystycznych jest cennym źródłem wiedzy, pod warunkiem jednak, że model ten w miarę możliwości oddaje specyfikę i złożony charakter badanych zjawisk. Każde zjawisko ma swoje indywidualne cechy, które nie uwzględnione zamazują obraz realnego jego funkcjonowania.

Z uwagi na wymienione założenia do uzupełnienia obrazu wzajemnych relacji wielowątkowych organizmów rynkowych zaproponowano procedurę tworzenia modeli geostatystycznych z wykorzystaniem reszt diagnozujących struktury przestrzenne jako twory złożone i heterogeniczne wewnętrznie. Cechą szczególną przedstawionego modelu geostatystycznego jest jego plastyczność i elastyczność. Dopuszcza on różnego rodzaju powiązania i związki

między analizowanymi zjawiskami, nie narzucając z góry sztywnej struktury konstrukcji modelu jak w przypadku np. modelu regresji liniowej. Efektem przeprowadzonych analiz, zgodnie z przyjętą wyżej procedurą budowy modelu jest model wynikający z indywidualnej struktury zjawiska. Struktura modelu, kształtuje się zatem w trakcie stopniowego odkrywania informacji zaszyfrowanej w danych, w przeciwieństwie do modelu, np. regresji liniowej, gdzie do konkretnej formy modelu są dopasowywane dane.

Istnieją jednak także pewne ograniczenia w stosowaniu wymienionej procedury modelowania struktur i systemów przestrzennych. Niezbędna jest duża liczba danych do analiz, aby uzyskać wiarygodne wnioski. Wymienione analizy są także w pewnym stopniu czasowo- i kosztochłonne, są częstym elementem towarzyszącym skomplikowanym analizom statystycznym. Pokonanie jednak wymienionych barier i dzięki temu włączenie reszt do procesu diagnozowania, a następnie planowania przestrzeni, pozwoli na wykorzystanie reguł z przeszłości, w celu minimalizacji niepewności w procesie gospodarowania nieruchomościami.

PIŚMIENNICTWO

- Barteczko K., Bocian A., 1996. Modelowanie polityki makroekonomicznej. Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Bertalanffy L., 1984. Ogólna teoria systemów. PWN, Warszawa.
- Cempel Cz., 2003. Teoria i inżynieria systemów. <http://neur.am.put.poznan.pl>
- Ostasiewicz W., 1999. Statystyczne metody analizy danych. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.
- Renigier M., 2004a. Reszty jako kryterium dopasowania modelu ekonometrycznego - referat wygłoszony na konferencji naukowej: "Analiza i modelowanie rynku nieruchomości na potrzeby wyceny" – 19-24.05.2004 r.
- Renigier M., 2004b. Selekcja zmiennych a priori przy nieklasycznych założeniach. Acta Sci. Pol., Administratio Locorum, 3 (2).
- Renigier M., 2004C. Zastosowanie metod analizy reszt z modeli statystycznych w gospodarce nieruchomościami. Rozprawa doktorska, Olsztyn.
- Renigier M., Żróbek R. 2005. Model Reszt a innowacje na rynku nieruchomości – nie publikowane.
- StatSoft. 2003 – www.statsoft.pl
- Tukey J., 1998. – www.mrs.umn.edu/~sugurea/introstat/history/w98/Tukey.htm

CONSTRUCTION OF A GEOSTATISTICAL MODEL USING RESIDUALS

Abstract. Spatial value diagnosis based on statistical modeling involving causative and effect rules is often incomplete and insufficient for correct conclusions and decision-making. The inclusion of the component unexplained by the model (residuals) to the construction of the model constitutes an essential complement of statistical formulae (deterministic) devoid of the element of uncertainty and randomization.

The present article includes the problems connected with the construction of geostatistical model using a residuals component.

Key words: residuals from model, geostatistical model, stages of model building, nonlinearity and uncertainty in spatial analyses.

Zaakceptowano do druku 2005.04.11

Accepted for print