

# Rozdział 1. Modele regresji przestrzennej zmiennych ukrytych i ograniczonych

## 1.1. Definicje, klasyfikacja i budowa modeli zmiennych ukrytych i ograniczonych

Prezentując w pierwszym tomie podstawowe metody i modele analizy danych przestrzennych, zakładaliśmy, iż w badaniach empirycznych każda analiza statystyczna, w tym modelowanie i prognozowanie ekonometryczne, opiera się na kwantyfikowanych wartościach zbioru  $K$  zmiennych:  $X_1, X_2, \dots, X_K$  (lub  $Y_1, Y_2, \dots, Y_K$ ) na próbie  $N$  obserwacji ( $i = 1, \dots, N$ ).

Wartość  $i$ -tej zmiennej dla  $j$ -ej obserwacji oznaczana jest jako  $x_{ij}$  ( $y_{ij}$ ), czyli obserwacje dotyczą poszczególnych jednostek populacji odpowiadających w przypadku danych przestrzennych poszczególnym lokalizacjom lub obszarom geograficznym rozlokowanym w przestrzeni. Ponadto zakładaliśmy dotąd, iż zmienne objaśniane są mierzalne i mogą przyjmować wartości ciągłe.

Wiadomo jednak, iż już w pierwszych publikacjach dotyczących badania autokorelacji przestrzennej [Cliff, Ord, 1981] i modelowania ekonometrycznego z uwzględnieniem interakcji przestrzennych [Anselin, 1988] wskazywano na możliwości rozszerzenia zbioru zmiennych objaśnianych (i odpowiednich modeli regresji przestrzennej) na zmienne typu jakościowego.

W literaturze polskiej [Wiśniewski, 1986; Gruszczyński, 2002, s. 12; Gruszczyński (red.), 2010, s. 46] najczęściej przez nazwę ekonometryczne modele dla zmiennych jakościowych rozumie się takie modele, w których zmienne objaśniane mogą przyjmować różnego typu wartości dyskretne. Formalnie wyróżnia się trzy rodzaje zmiennych jakościowych:

- 1) zmienne dwumianowe (binarne, dychotomiczne),
- 2) zmienne wielomianowe nieuporządkowane (politochomiczne),
- 3) zmienne wielomianowe uporządkowane.

Oprócz wymienionych zmiennych jakościowych przedmiotem analiz i modelowania ekonometrycznego są również dwie dodatkowe grupy zmiennych, mianowicie:

- 1) zmienne ograniczone, które przyjmują w pewnym przedziale wartości ciągłe, a w innym stałą wartość dyskretną (najczęściej zero), oraz
- 2) zmienne licznikowe (liczby całkowite, dodatnie).

Z tego względu obecnie przyjmuje się szerszą definicję zmiennych typu jakościowego, według koncepcji tzw. zmiennych ukrytych (*latent variables*), do opisu których konstruowane są odpowiednie modele ekonometryczne.

Zmienna ukryta jest nieobserwowalna, natomiast obserwowane są efekty lub decyzje związane z jej wartościami. Przyjmuje się tutaj interpretację skłonnościową<sup>1</sup>, ponieważ reprezentuje ona skłonność (inklinację, ciężenie) danego osobnika lub danej jednostki obserwacji do podejmowania decyzji lub przyjmowania stanu odpowiadającego różnym wartościom obserwowanej jakościowej zmiennej  $Y$  (np.  $y_i = 1$ ).

Interpretacja skłonnościowa dotyczy zarówno jednostek, które świadomie dokonują wyboru jednej z dwóch kategorii, jak i jednostek, które trafiają do danej kategorii, nie decydując o tym.

Dla nieobserwowalnej zmiennej ukrytej  $Y^*$  należy zdefiniować odpowiednią transformację definiującą obserwowane wartości efektów lub decyzje wynikające z jej działania. Przyjęta formuła transformacji określa jednocześnie typ modelu dla danej zmiennej ukrytej.

Klasyfikacja modeli, w których objaśniane są zmienne ukryte i ograniczone, powinna więc być oparta na dwóch kryteriach: rodzaju zmiennej definiującej efekty lub decyzje oraz wybranej postaci rozkładu prawdopodobieństwa realizacji tych efektów.

W przypadku pierwszego kryterium Green [2003] proponuje podział modeli dla zmiennych jakościowych i ograniczonych na dwie klasy:

- I) **modele wyboru dyskretnego** (*models for discrete choice*, rozdział 21, s. 663–755), a w tym:
  - modele dwumianowe (*binary choice models*),
  - modele wielomianowe kategorii uporządkowanych (*multinomial ordered models*),
  - modele wielomianowe kategorii nieuporządkowanych (*multinomial unordered models*),
  - modele danych licznikowych (*models for count data*);
- II) **modele zmiennych ograniczonych i czasu trwania** (*limited dependent variable and duration models*, rozdział 22, s. 756–802). W ramach drugiej grupy wyróżnia się:
  - modele dla zmiennych uciętych (*truncated*),
  - modele dla zmiennych cenzurowanych (*censored*),
  - modele selekcji próby (*sample selection models*) oraz
  - modele czasu trwania (*duration models*).

Modele wyboru dyskretnego opisują wynik faktycznie dokonywanych wyborów spośród różnych możliwości. Wartości liczbowe zmiennych objaśnianych

---

<sup>1</sup> Oprócz interpretacji skłonnościowej modeli dla zmiennych ukrytych w literaturze przedmiotu stosowana jest również interpretacja oparta na koncepcji użyteczności stochastycznych [McFadden, 1974; Gruszczyński, 2002, s. 16–17]. Będzie ona wykorzystana m.in. w przypadku modeli dla danych wielomianowych nieuporządkowanych.

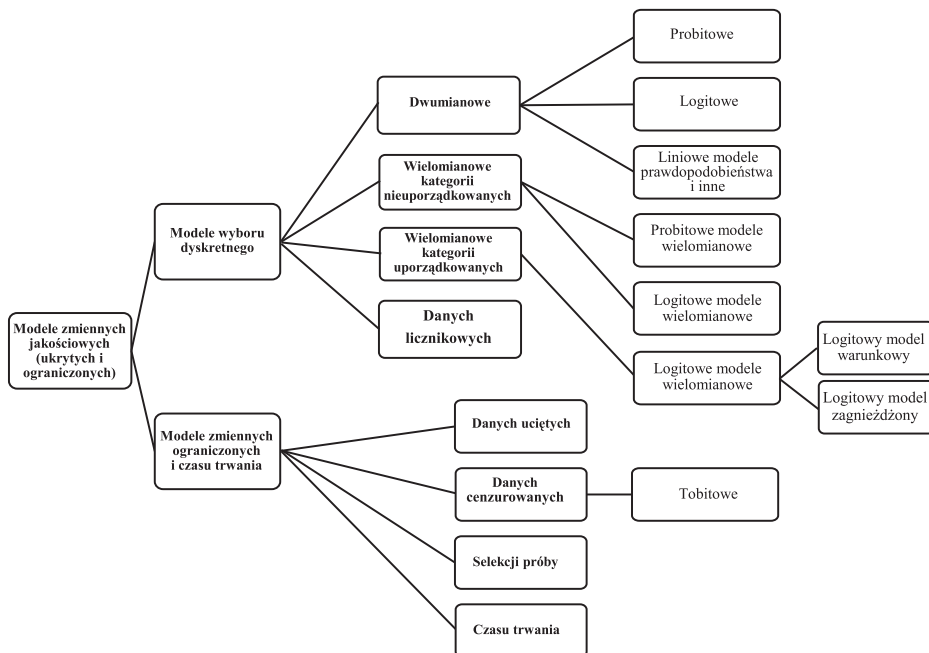
przybierają tutaj postać kodów nadawanych różnym kategoriom danej zmiennej ukrytej. Mogą to być także rangi kolejnych lub różnych wariantów tej zmiennej.

Modele zmiennych ograniczonych konstruowane są jako narzędzia opisu takich zmiennych objaśnianych, które mają albo warunkowe rozkłady dyskretno-ciągłe (np. rozkład jednopunktowy plus rozkład normalny pewnej zmiennej losowej), albo kiedy próba statystyczna jest złożoną próbą losową<sup>2</sup>.

W przypadku drugiego kryterium, jak już wspomniano, klasyfikacja modeli wynika z wyboru postaci funkcyjnej dystrybuanty rozkładu prawdopodobieństwa efektów zdarzeń lub decyzji. W praktyce najczęściej stosowane są dwa rozkłady prawdopodobieństw: normalny oraz logistyczny. Oprócz tego wymienia się rozkłady: liniowy, logarytmiczno-liniowy, Burra, krzywej Urbana i in. [Gruszczyński, 2002, s. 17–22].

Stosując więc drugie kryterium klasyfikacyjne w badaniach empirycznych, stosowane są modele: probitowe, logitowe, liniowe, logarytmiczno-liniowe i inne.

Bardziej szczegółowa klasyfikacja modeli dla zmiennych jakościowych (ukrytych i ograniczonych) prezentowana jest na załączonym schemacie (rys. 1.1).



**Rysunek 1.1.** Klasyfikacja modeli zmiennych jakościowych

Źródło: opracowanie własne.

<sup>2</sup> Pojawienie się danej obserwacji w próbie zależy od odpowiednich wartości zmiennej objaśniającej.

Model dwumianowy (wyboru binarnego – *binary choice model*) scharakteryzujemy bardziej szczegółowo jako punkt wyjścia do opisu modeli bardziej złożonych. Stosowany jest w przypadku następującej transformacji zmiennej ukrytej:

$$y_i = \begin{cases} 1, & \text{gdy } y_i^* > c, \\ 0, & \text{gdy } y_i^* \leq c. \end{cases} \quad (1.1)$$

Zmienna binarna  $y_i$  ma oczywiście rozkład dwumianowy, a model wyboru pozwala tutaj na analizę efektów wpływu wybranych zmiennych objaśniających na decyzje podejmowane przez badane podmioty, gdy zmienna objaśniana przyjmuje tylko dwie wartości.

Jest to model prawdopodobieństwa przyjmowania przez zmienną endogeniczną wartości jeden (jeśli zostanie przekroczona wartość progowa  $c$ ) lub zero (w przeciwnym razie). Prawdopodobieństwo  $P_i$  jest funkcją wektora wartości zmiennych objaśniających (egzogenicznych) oraz wartości wektora parametrów:

$$P_i = P(y_i = 1) = F(\mathbf{x}_i\boldsymbol{\beta}), \quad (1.2)$$

gdzie:  $[\mathbf{x}_i\boldsymbol{\beta}]$  – wskaźnik określający  $i$ -tą jednostkę obserwacji (kombinacja liniowa wartości zmiennych i parametrów);  $F$  jest funkcją rosnącą tego wskaźnika; typ funkcji  $F$  określa rodzaj modelu.

Przyjmujemy więc założenie, że istnieje pewna wartość progowa tego wskaźnika, np.  $c$ , od której udział wartości  $\{y_i = 1\}$  przewyższa udział wartości  $\{y_i = 0\}$ . Zakładamy ponadto, iż reguła decyzji i ich zależność od wskaźnika  $[\mathbf{x}_i\boldsymbol{\beta}]$  nie jest deterministyczna. Oznacza to, iż przyjmujemy zależność stochastyczną wyrażoną statystycznie przez funkcję regresji.

Przyjmując liniową postać równania regresji (w zapisie macierzowym):

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.3)$$

gdzie:

$\mathbf{y}^*$  – wektor hipotetycznych obserwacji na zmiennej ukrytej,

$\mathbf{X}$  – macierz obserwacji na zmiennych objaśniających,

$\boldsymbol{\varepsilon}$  – wektor składnika losowego o rozkładzie normalnym:  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ ,

dla poszczególnych obserwacji  $i = 1, \dots, N$  mamy:

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i; \quad (1.4)$$

$\mathbf{x}_i$  –  $i$ -ty wiersz macierzy obserwacji na zmiennych objaśniających  $\mathbf{X}$ .

Kombinacje wartości zmiennych i parametrów  $[\mathbf{x}_i\boldsymbol{\beta}]$  obliczone dla poszczególnych obserwacji mogą być interpretowane jako wartości tzw. indeksu skłonności, wskazujące na pozycje (kategorie), jakie przybiera zmienna ukryta  $Y^*$  dla poszczególnych obserwacji.

W zapisie probabilistycznym mamy więc:

$$\begin{aligned} P(y_i = 1) &= P[(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})_i > c] = P(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i > c) = \\ &= 1 - P(\varepsilon_i < c - \mathbf{x}_i\boldsymbol{\beta}), \\ P(y_i = 0) &= P[(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})_i \leq c] = P(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \leq c) = \\ &= P(\varepsilon_i \leq c - \mathbf{x}_i\boldsymbol{\beta}). \end{aligned} \tag{1.5}$$

Należy zauważyć, iż wartość progowa  $c$  jest identyczna dla wszystkich obserwacji. W przeciwnym razie należałoby przyjąć, iż różne wartości progowe związane są z odpowiednimi obserwacjami, a więc zależą od wartości charakterystyk tych obserwacji. Zakładając stałość wartości progowej i przyjmując postać modelu regresji (1.3) z wyrazem wolnym, nie mamy możliwości ich rozróżnienia i identyfikacji. Można więc przyjąć arbitralnie  $c = 0$ , tak jak dla wartości oczekiwanej składnika losowego  $E(\varepsilon_i) = 0$ .

Należy więc podkreślić, iż zapis probabilistyczny reguł decyzyjnych zależy wyłącznie od rozkładu składnika losowego równania regresji (1.3).

Wymienia się różne cele modelowania zmiennych jakościowych [Gruszczyński, 2002, s. 14], a w szczególności:

- weryfikacja hipotezy (hipotez) w kwestii mechanizmu generującego zmienną  $Y$ , czyli ustalenie zbioru zmiennych  $\mathbf{X}$ , które są istotne dla określenia wartości prawdopodobieństwa  $P$  w danej zbiorowości,
- prognoza  $P(y_i = 1)$  prawdopodobieństwa zdarzenia bądź stanu, że zmienna  $Y$  przyjmuje wartość  $y_i = 1$ , lub prognoza zmiany prawdopodobieństwa  $P$  wywołanej zmianą wartości jednej ze zmiennych egzogenicznych  $\mathbf{X}$ .

Do osiągnięcia tych celów konieczne jest oszacowanie parametrów przyjętego równania regresji z uwzględnieniem odpowiedniej postaci funkcyjnej rozkładu składnika losowego. W praktyce najczęściej stosowane są dwa rozkłady prawdopodobieństwa: normalny oraz logistyczny. W przypadku badań z zastosowaniem modelu dwumianowego mamy więc najczęściej model probitowy lub logitowy.

Oznaczając przez  $F(\cdot)$  funkcję dystrybuanty oraz przez  $f(\cdot)$  funkcję gęstości wybranego rozkładu dla składnika losowego modelu oraz ponieważ wartość progowa powinna być normalizowana do zera ( $c = 0$ ), prawdopodobieństwa zdarzeń w modelu zmiennej binarnej można zapisać następująco:

$$\begin{aligned} P(y_i = 1) &= P(\varepsilon_i > \mathbf{x}_i\boldsymbol{\beta}) = 1 - F(-\mathbf{x}_i\boldsymbol{\beta}), \\ P(y_i = 0) &= P(\varepsilon_i \leq -\mathbf{x}_i\boldsymbol{\beta}) = F(-\mathbf{x}_i\boldsymbol{\beta}). \end{aligned} \tag{1.6}$$

W przypadku rozkładów symetrycznych<sup>3</sup> mamy zawsze:

$$f(\mathbf{x}_i\boldsymbol{\beta}) = f(-\mathbf{x}_i\boldsymbol{\beta}) \quad \text{oraz} \quad F(\mathbf{x}_i\boldsymbol{\beta}) = 1 - F(-\mathbf{x}_i\boldsymbol{\beta}). \tag{1.7}$$

<sup>3</sup> W literaturze przedmiotu można znaleźć także przykłady zastosowań rozkładu niesymetrycznego, np. model gompit, w którym przyjmuje się rozkład Gomperta [Thomas, 2000, s. 55; Gruszczyński, 2002, s. 19].

W przypadku różnych modeli przyjmujemy założenie dotyczące homoskedastyczności składnika losowego z dowolną wartością poziomą wariancji. W celu ujednoczenia zapisu stosowane jest przekształcenie rozkładu normalnego o wartości oczekiwanej zero i wariancji równej jeden:  $\varepsilon \sim N(0, 1)$ . Jest to również konieczne do zredukowania argumentu funkcji prawdopodobieństwa dla  $y_i = 1$ . Dlatego też punktem wyjścia postaci estymacyjnej modelu probitowego jest następująca formuła prawdopodobieństwa:

$$P_i \left( \frac{\varepsilon_i}{\sigma} < \frac{\mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) = \Phi \left( \frac{\mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right). \quad (1.8)$$

Uogólnieniem modelu dwumianowego jest model wielomianowy zmiennej uporządkowanej. Przyjmuje się tutaj następującą transformację zmiennej ukrytej:

$$y_i = \begin{cases} 1, & \text{gdy } c_0 < y_i^* \leq c_1, \\ 2, & \text{gdy } c_1 \leq y_i^* < c_2, \\ \dots & \dots \dots \\ J, & \text{gdy } c_{J-1} \leq y_i^* < c_J, \end{cases} \quad (1.9)$$

gdzie:  $c_0 = -\infty$ ,  $c_J = \infty$ ,  $J > 2$ .

Przyjmuje się, iż uporządkowanie może mieć różną interpretację w zależności od rodzaju zjawiska, np. od najmniejszego do największego, od najmniej pewnego do najbardziej pewnego [Gruszczyński (red.), 2010, s. 49].

Natomiast opisanie modeli wielomianowych nieuporządkowanych za pomocą zmiennej ukrytej jest bardzo trudne. Dlatego też przyjmuje się tutaj koncepcję i interpretację użyteczności stochastycznych<sup>4</sup>. Zakłada się mianowicie, iż każda jednostka dokonująca wyboru (firma, konsument) czyni to w sposób racjonalny, czyli tak, aby zmaksymalizować użyteczność podjętej decyzji. Chcąc analizować te decyzje, mamy jednak tylko ograniczone informacje i niepewność odnośnie do poziomu użyteczności podejmowanych decyzji. Z tego względu istnieje konieczność przyjęcia zależności stochastycznej. Dla  $N$  jednostek podejmujących decyzje ( $i = 1, \dots, N$ ) oraz ustalonego zbioru  $C_r$  alternatywnych decyzji  $r = 1, \dots, R$  zapisujemy funkcję użyteczności jako sumę składowej deterministycznej (systematycznej)  $V_{ir}$  oraz składnika losowego określającego stopień niepewności  $\varepsilon_{ir}$ :

$$U_{ir} = V_{ir} + \varepsilon_{ir}. \quad (1.10)$$

Składowa systematyczna może zawierać różne zmienne objaśniające charakteryzujące zarówno podjęte decyzje w ocenach poszczególnych decydentów, jak i cechy osobiste tych jednostek.

<sup>4</sup> Jak podaje Gruszczyński [2002, s. 17], teorię użyteczności stochastycznych jako podstawę modeli jakościowych zmiennych wielomianowych sformułował McFadden w 1974 r. Zastosowanie tej teorii do formułowania przestrzennych modeli zmiennych jakościowych można także znaleźć w opracowaniu Mohammadiana i Kanaroglou [2003] oraz w artykule Smirnova [2010].

W przypadku zbioru  $C_r$  składającego się tylko z dwóch możliwości (wybór  $A$  lub  $B$ ) i przyjętej liniowej postaci funkcyjnej, stochastyczny model użyteczności można zapisać jako układ dwóch równań:

$$\begin{aligned} u_{iA} &= \mathbf{x}_{iA}\boldsymbol{\beta}_A + \varepsilon_{iA}, \\ u_{iB} &= \mathbf{x}_{iB}\boldsymbol{\beta}_B + \varepsilon_{iB}. \end{aligned} \quad (1.11)$$

Przyjmując zmienną zero-jedynkową  $Y$  dla zaobserwowanych decyzji wyboru ( $y_i = 1$  w sytuacji wyboru  $A$  oraz  $y_i = 0$  w sytuacji wyboru  $B$ ), otrzymujemy alternatywny do (1.5) zapis probabilistyczny modelu dwumianowego, określający prawdopodobieństwo zdarzenia, że użyteczność wyboru  $y_i = 1$  jest wyższa dla  $i$ -tej jednostki od użyteczności innego wyboru:

$$\begin{aligned} P(y_i = 1) &= P(u_{iA} > u_{iB}) = P(\mathbf{x}_{iA}\boldsymbol{\beta}_A + \varepsilon_{iA} - \mathbf{x}_{iB}\boldsymbol{\beta}_B - \varepsilon_{iB} < 0) = \\ &= P(\varepsilon_{iB} - \varepsilon_{iA} < \mathbf{x}_{iA}\boldsymbol{\beta}_A - \mathbf{x}_{iB}\boldsymbol{\beta}_B) = \\ &= F(\mathbf{x}_{iA}\boldsymbol{\beta}_A - \mathbf{x}_{iB}\boldsymbol{\beta}_B), \end{aligned} \quad (1.12)$$

gdzie:  $F$  jest dystrybuantą rozkładu składnika resztowego  $[\varepsilon_{iB} - \varepsilon_{iA}]$ .

W przypadku modeli wielomianowych nieuporządkowanych, funkcja  $F$  objaśnia nie pojedyncze prawdopodobieństwa, lecz ilorazy prawdopodobieństw, po przyjęciu jednej z kategorii jako bazowej, stanowiącej punkt odniesienia dla pozostałych.

Oprócz modeli dwumianowych oraz wielomianowych uporządkowanych koncepcja zmiennych ukrytych może być także przydatna do konstrukcji i interpretacji modeli zmiennych ograniczonych i modeli czasu trwania.

W przypadku modeli zmiennych ograniczonych – cenzurowanych, zwanych również modelami tobitowymi<sup>5</sup>, stosujemy następującą transformację zmiennej ukrytej na zmienną obserwowalną:

$$y_i = \begin{cases} y_i^*, & \text{gdy } y_i^* \geq c, \\ c, & \text{gdy } y_i^* < c. \end{cases} \quad (1.13)$$

Oznacza to, iż obserwacje na zmiennej  $Y^*$  są dostępne, gdy wartości są większe od pewnej stałej  $C$ , a w przeciwnym razie obserwujemy tylko stałą  $c$ .

Natomiast w przypadku modelu czasu trwania należy przyjąć, że:

$$y_i = \begin{cases} y_i^*, & \text{gdy } y_i^* \leq c_i, \\ c_i, & \text{gdy } y_i^* > c_i. \end{cases} \quad (1.14)$$

W modelach czasu trwania<sup>6</sup> (*duration models*) wartości zmiennej ukrytej są dostępne tylko wtedy, gdy są mniejsze lub równe wartościom zmiennej  $C$ .

<sup>5</sup> Od nazwiska J. Tobina [1958], który był autorem propozycji takich modeli.

<sup>6</sup> Modele tego typu służą do opisu sytuacji, gdy zmienna ukryta ma interpretację czasu trwania lub „życia” danego obiektu. Ich koncepcja wywodzi się od zastosowań biologicznych przy ana-

W przeciwnym razie znane są tylko wartości  $c_i$ . W porównaniu z modelem tobitowym mamy tutaj cenzurowanie prawostronne, a ponadto każda obserwacja może mieć inną wartość cenzurowania  $c_i$ .

Reasumując, w dalszych prezentacjach i rozważaniach, zgodnie z nomenklaturą Greena [2003, 2011] w niniejszej pracy będą używane nazwy wyróżniające dwie grupy zmiennych i modeli (typu „jakościowego”), mianowicie: zmienne ukryte i zmienne ograniczone oraz odpowiednio: modele dla zmiennych ukrytych i modele dla zmiennych ograniczonych.

## 1.2. Efekty przestrzenne w modelach dla zmiennych ukrytych i ograniczonych

Modele regresji przestrzennej dla zmiennych ukrytych i ograniczonych są proponowane w pracach z zakresu ekonometrii przestrzennej już od kilkunastu lat. Zainteresowanie tymi modelami wzrosło jednak wyraźnie po roku 2001 [Kelejian, Prucha, 2001; LeSage, 2002; Beron, Murdoch, Vijverberg, 2003, 2004]. Aktualne badania dotyczą w szczególności opracowania odpowiednich metod estymacji parametrów takich modeli oraz ich zastosowania w badaniach empirycznych [Holloway i in., 2002; Murdoch i in., 2003].

Uwzględnienie, a następnie praktyczne zastosowanie interakcji przestrzennych w modelach i analizach dla zmiennych jakościowych jest jednak dość skomplikowane.

Ogólnie, w badaniach przestrzennych wyróżniamy dwie grupy efektów: heterogeniczność oraz autokorelację przestrzenną. W przypadku danych jakościowych dotyczących wyborów indywidualnych podstawowe znaczenie ma heterogeniczność, czyli zróżnicowanie wyborów indywidualnych oraz zróżnicowanie możliwości wyboru. W przeciwieństwie do tego zależności w postaci autokorelacji przestrzennych pociągają za sobą współzależności pomiędzy decydentami, ich preferencjami i dokonanymi wyborami [Smirnov, 2010, s. 2].

Zależności autoregresyjne i autokorelacyjne są szczególnie ważne w przestrzennych modelach wyboru dyskretnego. Współzależności przestrzenne w decyzjach obserwowanych jednostek pokazują nowy aspekt ich zachowania się, który nie występuje w innych modelach. Mianowicie wpływają one na preferencje jednostek, tworząc zjawisko społecznych wpływów w podejmowaniu decyzji w taki sposób, że ani jednostki nie działają niezależnie, ani też nie podejmują decyzji łącznie (wspólnie). Badania efektów interakcji przestrzennych w analizach wyborów dyskretnych są ważne dla rozszerzenia modeli zmiennych ukrytych o społeczne aspekty podejmowania indywidualnych decyzji [Smirnov, 2010, s. 3].

lizowaniu zależności czasu życia organizmów od czynników wewnętrznych i cech środowiska. W zastosowaniach ekonomicznych znane są np. modele czasu trwania strajków, bezrobocia, umów abonenckich i in. [Gruszczyński (red.), 2010, s. 50].



Oprócz trudności i problemów estymacyjnych zasadniczą komplikacją w modelowaniu zmiennych ukrytych z uwzględnieniem interakcji przestrzennych jest fakt, że wiele współzależności nie jest bezpośrednio obserwowanych. Dodatkowo wielu decydentów działa w warunkach niepełnej informacji, skłaniając do przyjmowania założeń o stochastycznej maksymalizacji ich oczekiwań. Jeśli to dotyczy przestrzenie rozproszonych jednostek, stawiana jest hipoteza, iż przestrzenne współzależności są formą adaptacji rozmieszczenia przestrzennego badanych podmiotów w celu maksymalizacji użyteczności ich decyzji. Dokładny mechanizm, według którego przestrzenna zależność przekształca się w obserwowane wybory, zależy od specyfiki obserwacji i różni się w zależności od definicji decydentów (gospodarstwa domowe, firmy lub inne podmioty) oraz jest wyraźnie zależny od istoty poznawczych i racjonalnych aspektów indywidualnych decydentów.

Należy jednak zauważyć, iż odpowiednie uwzględnienie i pomiar przestrzennych zależności w modelu wymaga odpowiednich technik i ostrożności w interpretacji rezultatów.

Chcąc uwzględnić aspekty społeczne zachowania się indywidualnych decydentów, można przyjąć, iż interakcje przestrzenne są modelowane w celu wyjaśnienia indywidualnych preferencji w decyzjach wyboru. Przestrzenne współzależności nie są bezpośrednio obserwowane. Zamiast tego badacz obserwuje indywidualne wybory, które są kodowane dla pomiaru efektów współzależności przestrzennych.

Według O. Smirnova [2010, s. 4–8] można wyróżnić kilka grup modeli w zależności od rodzaju efektów przestrzennych i sposobu ich wprowadzenia do modelu zmiennej ukrytej, np.:

- 1) modelowanie przestrzennych zależności pomiędzy możliwościami,
- 2) przestrzenne zależności w liniowym modelu prawdopodobieństwa,
- 3) podejście mechanizmu statystycznego,
- 4) endogeniczne wagi przestrzenne,
- 5) przestrzenna heterogeniczność,

Ad 1. Modele typu pierwszego pozwalają na uwzględnianie różnego typu zależności przestrzennych pomiędzy alternatywnymi wyborami. Ma tutaj zastosowanie zazwyczaj specyfikacja modelu logitowego ze stałą elastycznością substytucji w składniku losowym. Pozwala to na ocenę efektu substytucji w indywidualnych preferencjach pomiędzy alternatywami zgodnie z ich wzajemnym rozmieszczeniem przestrzennym.

Ad 2. Zmienna binarna decyzji wyboru jest tutaj sumą zmiennych dyskretnych i ciągłych (w przeciwieństwie do reguł decyzyjnych wyrażonych przez zmienne wyłącznie dyskretne). Zastosowano jednorównaniowy liniowy model prawdopodobieństwa, który estymowany był uogólnioną metodą momentów [Pinkse, Slade, Shen, 2006]. Ten uproszczony model może jednak dawać informacje mylące pomiędzy zmiennymi wyboru dyskretnego a warunkowym prawdopodobieństwem wyboru, co ogranicza użycie tego modelu w analizach. Dodatkową wadą są asymptotyczne obciążenia estymatorów parametrów (np. UMM).