

Metody ilościowe w modelowaniu zachowań konsumenckich

Małgorzata Rószkiewicz

W artykule dokonano prezentacji głównych opcji badawczych, wykorzystywanych w ilościowej analizie zachowań konsumenckich. Zwrócono uwagę na założenia o charakterze konceptualnym oraz formalnym. Prezentowany tekst koncentruje się wokół zagadnień związanych z podejściem stochastycznym, modelującym indywidualne wybory konsumenckie. Rozważono tu zarówno podejście binarne, jak i wielonominale, z uwzględnieniem oddziaływania jednowymiarowego, jak i wielowymiarowego zbioru uwarunkowań. Wskazano również na możliwość uwzględniania czasu wśród determinant wyborów konsumenckich.

1. Wprowadzenie

Analiza ilościowa zachowań konsumenckich może być prowadzona na różnych poziomach i według różnych podejść badawczych. Wybór ścieżki postępowania zależy od tego, w jaki sposób zachowania konsumenckie są opisywane, czyli w jaki sposób w budowanych modelach konceptualnych jest definiowana zmienna zależna. Można wyróżnić tu dwie charakterystyczne opcje badawcze. Jedna z nich odnosi się do wyrażanych ilościowo decyzji konsumenckich, które są opisywane takimi zmiennymi mierzalnymi jak wielkość jednorazowego zakupu lub jego wartość, co daje możliwość agregacji indywidualnych zachowań w strumień łącznych efektów dokonanych wyborów rynkowych i utożsamia efekty zachowań konsumenckich z poziomem zrealizowanej sprzedaży. Poziom sprzedaży może być rejestrowany zarówno na podstawie danych operatywnych, jak i wywiadu marketingowego lub monitoringu sprzedaży, zależnie od stopnia rozwoju systemu dystrybucji produktów i usług oferowanych przez firmę (im system ten jest bardziej rozwinięty, tym mniej użyteczne są w rejestrowaniu poziomu sprzedaży dane dostępne w przedsiębiorstwie i tym częściej sięga się do badań marketingowych w formie monitoringu sprzedaży). Wielkość lub wartość jednorazowego zakupu może być natomiast określana jedynie w badaniach marketingowych za pomocą np. rejestrów kasowych w ramach monitoringu sprzedaży lub wywiadów indywidualnych.

Drugie z wyróżnionych tu podejść odnosi się zaś do wyrażanych jakościowo indywidualnych wyborów konsumenckich. Zmienna opisująca je jest zmienną dyskretną, przyjmującą skończoną liczbę wartości, których liczba jest uzależniona

od wyróżnionych kategorii rozważanego zachowania. W zakresie zaistniałych zdarzeń jest zmienną dychotomiczną, określająca zakup lub jego brak, zaś w zakresie intencji może odnosić się do stopnia skłonności dokonania zakupu, np. w formie stwierdzeń typu: zdecydowanie tak, raczej tak, raczej nie, zdecydowanie nie, nie wiem. Warto jednak podkreślić, że z punktu widzenia metod modelowania decyzji konsumenckich rozróżnienie między zaistniałym zachowaniem a intencją zakupu nie ma znaczenia, natomiast rozróżnienie to jest istotne w przypadku interpretacji wyników oraz błędów *ex post*, czyli faktycznej dokładności symulacji dokonywanych na podstawie zbudowanych modeli. Sam pomiar zachowań indywidualnych lub intencji zakupu może dokonywać się jedynie drogą badań marketingowych, poprzez testy marketingowe lub bezpośrednie obserwacje zachowań nabywczych (jawne lub ukryte). Wówczas rola badań marketingowych w rozpoznawaniu, a następnie analizie zachowań konsumentów jest zatem podstawowa.

W przypadku modelowania metodami ilościowymi zachowań konsumenckich można wyróżnić również dwie odmienne koncepcje badawcze. Pierwsza ujmuje rozmiary fizyczne lub wartość sprzedaży jako kategorie całkowicie zdeterminowane przez określone wielkości rynkowe. Tego typu deterministycznym modelem jest, dobrze znany w literaturze, model Parfitta i Collinsa (por. Parfitt, Collins 1968: 131-145), w ramach którego udział w rynku jest dekomponowany na trzy wielkości: stopę penetracji, stopę wyłączności oraz stopę intensywności zakupów.

Innym, równie popularnym modelem deterministycznym, opisującym łączny efekt zachowań konsumenckich, jest ujęcie wartości sprzedaży jako funkcji rozmiarów rynku, pozycji firmy na rynku oraz ceny jednostkowej produktu. Modelowanie tego typu wykorzystuje się do oceny wpływu każdego z wyróżnionych w nim składników na poziom sprzedaży (Lambin 2001). Modele deterministyczne zachowań konsumenckich są wykorzystywane głównie w analizie sprzedaży (Strzyżewska, Rószkiewicz 2002).

Odmiernym względem deterministycznego jest podejście stochastyczne, które decyzje konsumenckie opisuje za pomocą zmiennych o charakterze losowym. Randomizacja wyborów konsumenckich znalazła swe podstawy w tzw. teorii użyteczności losowej.

Teoria użyteczności losowej wychodzi z założenia, że zachowanie konsumenta, mimo że jest podporządkowane regule racjonalnego wyboru, nie zawsze jest przewidywalne. Wynika to z losowych uwarunkowań samej użyteczności, która w zależności od kontekstu dla tego samego dobra może być określana na różnym poziomie. Formuła funkcji użyteczności przyjmuje zatem postać:

$$U_{ij} = \hat{U}_{ij} + \varepsilon_{ij}$$

gdzie:

U_{ij} – użyteczność i -tego obiektu dla j -tego konsumenta,

$\hat{U}_{ij} = \sum_{k=1}^p \beta_{ijk} x_{jk}$ – użyteczność i -tego obiektu, odnotowywana w przypadku

j -tego konsumenta opisanego p -wymiarowym zestawem cech niezależnych (X), zaś β_{ijk} – współczynniki kombinacji liniowej, wyrażające wpływ wyróżnionych cech niezależnych na poziom użyteczności i -tego obiektu w przypadku j -tego konsumenta.

Kierując się racjonalnością, konsument dokona wyboru tej oferty, której użyteczność jest największa. Ponieważ wybór ten ma charakter stochastyczny, zatem można wyrazić go nie za pomocą maksymalizacji użyteczności oferty, lecz za pomocą maksymalizacji prawdopodobieństwa jej wyboru, gdyż:

$$P(\text{wybór } k\text{-tego obiektu}) = P(U_k > U_i \text{ dla każdego } i \neq k),$$

co pozwala wiązać prawdopodobieństwo wyboru określonego wariantu z działaniem zespołu p zmiennych niezależnych, różnicujących użyteczność poszczególnych wariantów według formuły:

$$P(k) = f(X_1, X_2, \dots, X_p) 1$$

W podejściu tym przyjmuje się, że decyzja zakupu wykazuje zależność statystyczną z zespołem czynników, których oddziaływanie kształtuje tę decyzję z dokładnością do składnika losowego. Czynniki te mogą leżeć zarówno po stronie konsumentów i dotyczyć ich cech deskryptywnych oraz behawioralnych, jak i po stronie przedsiębiorstwa i obejmować zestaw narzędzi marketingu mix. W konstruowanych modelach zmienną zależną może być prawdopodobieństwo zakupu, estymowane częstością zakupów z badania empirycznego lub też z modeli liczby przypadków dokonywania zakupu (por. Jajuga 1990). Zagadnieniem związanym z podejściem stochastycznym w modelowaniu wyborów konsumenckich poświęcono uwagę w dalszej części niniejszego artykułu¹.

W metodach ilościowych do modelowania zmiennych dyskretnych służy specjalna grupa modeli. Można je traktować jako modele prawdopodobieństwa zajścia określonego zdarzenia (w tym wypadku dokonania zakupu bądź rezygnacji z zakupu) w zależności od różnorodnych czynników warunkujących to zajście. Modele te pozwalają zilustrować wpływ wybranych czynników na możliwość wystąpienia badanych zdarzeń lub jedynie wyrażanych intencji.

Badanie zależności między możliwością wystąpienia zdarzenia i ewentualnymi czynnikami limitującymi lub stymulującymi tę możliwość może polegać na:

- porównawczym badaniu przypadków, czyli badaniu o charakterze przekrojowym,
- badaniu panelowym (longitudinalnym), czyli wzdłużnym (por. Rószkiewicz 2002a).

Badanie porównawcze przypadków polega na porównaniu dwóch zbiorowości: pierwszej, w której obserwowane zachowanie konsumenckie wystąpiło, zwanej „grupą badaną” i drugiej, zwanej „grupą kontrolną” lub „grupą odniesienia”, w której zachowanie to nie miało miejsca. W obu grupach odnotowuje się występowanie czynnika, którego wpływ na badane zachowanie podlega ocenie. Z kolei w badaniach panelowych dobiera się obie grupy według określonych kryteriów. Są one jednorodne (inaczej: ekwiwalentne) pod względem określonych cech, np. demograficznych, ekonomicznych, społecznych lub innych, które wyznacza cel badania. Różnią się zaś występowaniem czynnika, którego wpływ na zachowania jednostek jest oceniany. Może to być np. czynnik marketingowy typu trwająca kampania reklamowa lub modyfikacja jednej z cech produktu bądź organizacji sprzedaży. Populację tę klasyfikuje się według działania kontrolowanych czynników (lub tylko jednego) i dokonuje się obserwacji w czasie. Są to badania prospektywne. Obserwacji podlega zaś występowanie badanych zachowań w populacji po upływie ustalonego czasu. Krótko mówiąc, w badaniach przekrojowych zbioro-

wości podlegające obserwacji konstruuje się według wariantów zmiennej zależnej, czyli interesujących badacza zachowań konsumenckich, zaś w badaniach panelowych według oddziaływania czynników, których wpływ na zachowania konsumenckie jest obserwowany w czasie. Rozważać można również wpływ upływu czasu obok czynników testowanych, które są pod kontrolą badacza.

Techniki analityczne, które mogą być stosowane w przypadku oceny wpływu różnych czynników na możliwość wystąpienia określonych zachowań konsumenckich, zależą od liczby kategorii wyróżnianych do opisu analizowanego zachowania oraz od liczby rozważanych czynników, które mogłyby analizowane zachowanie determinować, a ponadto od sposobu ich skalowania. Można tu rozważać wariant najprostszy, gdy zachowanie konsumenckie opisywane jest przez zmienną dychotomiczną (np. wystąpił zakup produktu lub nie dokonano zakupu) i również dwuwariantowy jest czynnik, którego wpływ na to zachowanie jest testowany. Wariantami bardziej rozbudowanymi są przypadki, gdy liczba czynników poddanych ocenie zwiększa się, jak również zachowanie konsumenckie jest opisywane wielonominalnie. Przypadki te zostaną poniżej szczegółowo scharakteryzowane.

2. Dychotomiczny opis zachowań konsumenckich oraz binarny charakter analizowanego czynnika

Jeśli przedmiotem analizy jest zarówno dychotomiczny czynnik, którego wpływ jest oceniany, jak i dychotomicznie opisywane jest zachowanie konsumenckie, to wyniki pomiaru są klasyfikowane w tablicy dwudzielczej. Model takiej sytuacji badawczej opisuje łączny rozkład prawdopodobieństwa zmiennej losowej, dwuwymiarowej, w którym każda zmienna przyjmuje jedynie dwie wartości. Rozkład łączny prezentuje następująca tablica:

Zmienna X	Zmienna Y		Rozkład brzegowy X
	y1	y2	
x1	P11	P12	P1.
x2	P21	P22	P2.
Rozkład brzegowy Y	P.1	P.2	1

Tab. 1. Rozkład łączny dychotomicznej zmiennej losowej dwuwymiarowej (XY) Źródło: opracowanie własne

Z szansę wystąpienia wartości x_1 względem wartości x_2 przy zadanym Y na poziomie y_j opisuje relacja prawdopodobieństw warunkowych (Arminger, Clogg, Sobel 1995: 251-303):

$$\omega_{(1:2|j)} = \frac{P(x_1 | y_j)}{P(x_2 | y_j)} = \frac{P_{1j} / P_{.j}}{P_{2j} / P_{.j}} = \frac{P_{1j}}{P_{2j}}$$

Relacja ta daje podstawę wyznaczenia wskaźnika ryzyka względnego o postaci:

$$\Psi = \frac{\omega_{(1:2|1)}}{\omega_{(1:2|2)}} = \frac{P_{11} / P_{21}}{P_{12} / P_{22}} = \frac{P_{11} P_{22}}{P_{21} P_{12}}$$

Otrzymany wynik jest porównywany z wartością 1 i na ogół jest wyrażany w procentach. Jeśli wartość ryzyka względnego równa się 1, oznacza to, że rozkłady warunkowe zmiennej X są identyczne i tym samym zmienna Y nie ma wpływu na występowanie bądź niewystępowanie zmiennej X na poziomie x_1 . Jeśli wartość współczynnika jest mniejsza od 1, oznacza to, że zmienna Y działa ograniczająco na możliwość wystąpienia zmiennej X na poziomie x_1 . Przy czym zakres rozbieżności z wartością 1 określa siłę tego oddziaływania. Jeśli wartość współczynnika przekracza wartość 1, to uznaje się, że zmienna Y stymuluje wystąpienie zmiennej X na poziomie x_1 , przy czym siła oddziaływania jest tym większa, im wartość współczynnika bardziej odbiega od wartości 1.

Wyniki prowadzonych obserwacji empirycznych klasyfikowane są w tablicy dwudzielnej w każdym z dwóch wyróżnionych typów analizy inaczej, gdyż inaczej są definiowane zarówno grupa badana, jak i kontrolna. W badaniach panelowych grupy rozróżnia oddziaływanie lub brak oddziaływania testowanego czynnika, zaś w porównawczych badaniach przypadków grupy rozróżnia występowanie lub brak występowania obserwowanego zachowania konsumenckiego. Zasady te przedstawiają kolejno tablice 2 i 3.

Czynnik	Zachowanie konsumenckie		Rozkład brzegowy czynnika
	wystąpiło	nie wystąpiło	
Wystąpił (grupa badana)	liczba przypadków a	liczba przypadków c	a+c
Nie wystąpił (grupa kontrolna)	liczba przypadków b	liczba przypadków d	b+d
Rozkład brzegowy zachowania	a+b	c+d	N

Tab. 2. Schemat klasyfikacji wyników w badaniu panelowym w tablicy dwudzielnej. Źródło: opracowanie własne

Zachowanie konsumenckie	Czynnik		Rozkład brzegowy zachowania
	wystąpił	nie wystąpił	
Wystąpiło (grupa badana)	liczba przypadków a	liczba przypadków c	a+c
Nie wystąpiło (grupa kontrolna)	liczba przypadków b	liczba przypadków d	b+d
Rozkład brzegowy czynnika	a+b	c+d	n

Tab. 3. Schemat klasyfikacji wyników w badaniu porównawczym przypadków w tablicy dwudzielnej. Źródło: opracowanie własne

Na podstawie tak pogrupowanych wyników obserwacji przedmiotem estymacji jest szansa, czyli ryzyko wystąpienia określonego zachowania konsumenckiego

w wyniku oddziaływania testowanego czynnika. Na podstawie wyników badań panelowych wskaźnik ryzyka szacuje się według wzoru (Armitage¹⁹⁷⁸: 355):

$$\hat{\psi} = \frac{\frac{a}{(a+c)}}{\frac{b}{(b+d)}} = \frac{a \cdot (b+d)}{b \cdot (a+c)}$$

Na podstawie wyników porównawczego badania przypadków szacuje się zaś iloraz szans według formuły:

$$\hat{\psi} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \cdot d}{b \cdot c}$$

Ryzyko względne można również oszacować na podstawie próby losowej, warstwowej, w której grupa badana oraz kontrolna pobierane są z odrębnych warstw. Wówczas konieczna jest ocena istotności otrzymanych wyników. W ocenie tej korzysta się z zasad estymacji przedziałowej. Dla otrzymanego wyniku z próby wyznacza się przedział wartości zgodnie z ogólnymi zasadami przedziałowej estymacji parametrycznej, na podstawie prób losowych. Ponieważ przedziały ufności są konstruowane według reguł addytywnych, korzysta się tu z rachunku logarytmicznego, który transformuje wielkości ilorazowe do addytywnych. Wariancję logarytmu wskaźnika ryzyka względnego, wyznaczonego w badaniu longitudinalnym definiuje wzór:

$$\sigma^2(\ln \hat{\psi}) = \frac{c}{a(a+c)} + \frac{d}{d(b+d)},$$

w porównawczym badaniu przypadków zaś, dla ilorazu szans, wzór:

$$\sigma^2(\ln \hat{\psi}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}.$$

Maksymalny błąd dla logarytmu wskaźnika wynosi: $1,96 \sqrt{\sigma^2(\ln \hat{\psi})}$.

Określa on połowę rozpiętości przedziału dopuszczalnych wartości logarytmu wskaźnika ryzyka względnego, czyli przedziału:

$$\left\langle \ln \hat{\psi} - 1,96 \sqrt{\sigma^2(\ln \hat{\psi})}, \ln \hat{\psi} + 1,96 \sqrt{\sigma^2(\ln \hat{\psi})} \right\rangle.$$

Krańce tego przedziału należy następnie antylogarytmować.

Jeśli tak wyznaczony przedział nie zawiera wartości 1, to wynik uznaje się za istotny statystycznie. Oszacowany przedział ma bowiem wówczas jednoznacznie interpretację, wskazując na jednoznacznie stymulujące (powyżej 1 oba krańce przedziału) lub jednoznacznie limitujące (poniżej 1 oba krańce przedziału) działanie testowanego czynnika na możliwość zajścia badanego zachowania konsumpcyjnego. Jeśli zaś zawiera wartość 1, prowadzi to do niejednoznaczności wyniku estymacji i wówczas uznaje się, że otrzymany wynik z próby jest nieistotny.

Przedstawione zasady mogą być rozbudowane do wariantu, w którym rozważa się więcej niż jeden czynnik determinujący możliwość wystąpienia rozważanego zachowania konsumenckiego. Wówczas rozważa się wielokrotne tablice kontyngencji. Wskaźnik ryzyka względnego dla rozkładu trójwymiarowej zmiennej losowej (X, Y, Z) , który definiuje formuła:

$$\Psi_k = \frac{P_{(11k)}P_{(22k)}}{P_{(21k)}P_{(12k)}} = \frac{P_{(11k)}P_{(22k)}}{P_{(21k)}P_{(12k)}}$$

jest wyznaczany dla wszystkich k kategorii trzeciej zmiennej w zdefiniowanej zmiennej trójwymiarowej. W ten sposób można w zasadzie rozważać zmienne o dowolnym wymiarze, ale w takich wielokrotnych tablicach kontyngencji rachunki stają się dość kłopotliwe. Wygodniejsze jest wówczas modelowanie tych zależności, w ramach którego możliwe jest zarówno budowanie modeli prawdopodobieństw występowania określonych wyborów konsumenckich, jak i modeli odnoszących się do liczby przypadków dokonywanych wyborów. W pierwszym wariantcie są to modele regresji statystycznej, opisujące prawdopodobieństwo wyboru konsumenckiego z dokładnością do składnika losowego, w drugim zaś są to modele opisujące liczbę zdarzeń.

3. Dychotomiczny opis zachowań konsumenckich oraz wielowymiarowy charakter rozważanych czynników

W przypadkach gdy zachowanie konsumenckie opisywane jest przez zmienną o charakterze binarnym, natomiast wieloelementowy jest zbiór determinant (czynników) tego zachowania, wygodnym narzędziem analizy są modele regresji. Formalny opis rozważanego tu zagadnienia sprowadza się do analizowania rozkładu warunkowego zmiennej binarnej o rozkładzie Bernoulliego. Zakładając zatem, że Y jest zmienną o rozkładzie Bernoulliego, przyjmującą dwie wartości – odpowiednio 0 i 1 – z wartością oczekiwaną P_p , czyli $E(Y) = P_p$, analizie poddaje się zespół czynników opisanych wektorem zmiennych $[X_1, X_2, \dots, X_p] = X$, które kształtują możliwość zajścia zdarzenia $Y=1$. Zapis formalny tych założeń przyjmuje postać:

$$P(Y = 1 | X) = E(Y | X) = F(X^T \beta) = F\left(\sum_{i=1}^p \beta_i x_i\right)$$

gdzie:

$\beta = [\beta_1, \beta_2, \dots, \beta_p]$ – wektor współczynników funkcji regresji.

Tak przedstawiona formuła modelu prawdopodobieństwa dokonania zakupu odpowiada koncepcji maksymalizacji użyteczności przy dokonywaniu wyboru jednej z dwóch alternatywnych propozycji. F jest dystrybuantą zmiennej losowej, definiowanej przez różnicę składników losowych: $(\xi_0 - \xi_1)$, z których każdy jest składnikiem losowym liniowego równania regresji wielorakiej użyteczności określonych wariantów decyzyjnych (odpowiednio: $u_0 = f(X) + \xi_0$ oraz $u_1 = f(X) + \xi_1$) i w zależności od postaci dystrybuanty F można definiować różne modele regresji dla prawdopodobieństwa zdarzenia (Jajuga 1990). Najczęściej rozważa się:

1. Model liniowy prawdopodobieństwa:

$$p_i = \alpha_{i0} + \alpha_{i1}X_{i1} + \alpha_{i2}X_{i2} + \dots + \alpha_{ik}X_{ik}$$

2. Model probitowy:

$$F(\mathbf{X}^T\beta) = \int_{-\infty}^{\mathbf{X}^T\beta} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt$$

3. Model logitowy:

$$F(\mathbf{X}^T\beta) = \frac{\exp(\mathbf{X}^T\beta)}{1 + \exp(\mathbf{X}^T\beta)}$$

Trudności estymacyjne liniowych modeli prawdopodobieństwa, a szczególnie heteroscedastyczność składnika losowego, spowodowały, iż najczęściej przyjmuje się, że funkcja opisująca wartości prawdopodobieństwa jest formułą zgodną z krzywą Gaussa lub funkcją logistyczną. Modele te można stosować do zdarzeń, które charakteryzuje zgodność kierunku zależności prawdopodobieństwa zdarzenia względem wartości zmiennej niezależnej oraz dystrybuanty dla tego rozkładu. A dokładniej, kiedy wraz ze wzrostem wartości zmiennej niezależnej następuje wzrost wartości prawdopodobieństwa wystąpienia zdarzenia. Taka sytuacja ma miejsce np. w przypadku możliwości zakupu produktu wraz ze wzrostem dochodów gospodarstw domowych, wzrostu ryzyka zachorowania wraz z wiekiem lub wystąpienia awarii wraz z okresem eksploatacji urządzenia itp. Poniżej zostanie omówiony model regresji logistycznej oraz zasady jego estymacji. W przypadku modeli probitowych zasady postępowania i interpretacja wyników jest analogiczna, należy posłużyć się jedynie innym wzorem matematycznym funkcji rozkładu.

Niestety w modelach probitowych i logitowych, wartości oszacowanych współczynników regresji β , nie mają żadnej interpretacji. Możliwości interpretacyjne stwarza natomiast przekształcenie oszacowanego równania w iloraz:

$$\psi = \frac{P_i}{1 - P_i}$$

Iloraz ten, zwany ilorazem szans, określa relatywną możliwość wystąpienia zdarzenia. W przypadku formuły regresji logistycznej, za pomocą której przykładowo dokonuje się modelowania wartości prawdopodobieństwa, można oszacować poziom tego ryzyka jako funkcji zmiennych niezależnych. Iloraz ten upraszcza się bowiem do postaci:

$$\psi = e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}$$

Wyrażenie e^{α_j} wyraża relatywną zmianę możliwości wystąpienia zdarzenia w wyniku działania czynnika opisanego przez zmienną X_j , przy założeniu kontrolowania, czyli stabilności, pozostałych zmiennych uwzględnionych w równaniu. Wartość tę interpretuje się, porównując ją z wartością 1 i wyrażając uzyskaną różnicę w procentach. Jeżeli: $e^{\alpha_j} > 1$, to uznaje się, że czynnik opisywany przez zmienną niezależną X_j działa stymulująco na prawdopodobieństwo (możliwość) wystąpienia badanego zjawiska, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu; $e^{\alpha_j} < 1$, to uznaje się, że czynnik opisywany przez zmienną niezależną X_j działa limitująco na prawdopodobieństwo (możliwość) wy-

stąpienia badanego zjawiska, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu; $e^{\alpha_j} = 1$, to uznaje się, że czynnik opisywany przez zmienną niezależną X_j nie ma wpływu na prawdopodobieństwo (możliwość) wystąpienia badanego zjawiska, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu.

4. Wielonominalny opis zachowań konsumenckich oraz wieloraki charakter analizowanych czynników

W przypadku gdy liczba wariantów zachowań konsumenckich jest większa niż 2, model przyjmuje postać modelu wielonominalnego. Model wielonominalny jest uogólnieniem modelu dla zmiennej zależnej o charakterze binarnym. Jego postać definiuje formuła:

$$P_{ij} = \frac{e^{z_{ij}}}{1 + \sum_{s=1}^k e^{z_{is}}}$$

gdzie: $z_{ij} = \alpha_{i0} + \alpha_{i1}X_{1j} + \alpha_{i2}X_{2j} + \dots + \alpha_{ik}X_{kj}$

P_{ij} – prawdopodobieństwo zajścia zachowania opisanego kategorią j -tą przy i -tej konfiguracji zmiennych niezależnych, k – liczba kategorii nominalnych, zdefiniowanych jako warianty zachowania konsumenckiego.

Wówczas:

$$P_{i0} = \frac{e^{z_{i0}}}{1 + \sum_{s=1}^k e^{z_{is}}}$$

Podobnie jak w przypadku binarnej zmiennej zależnej parametry modelu wielonominalnego szacuje się metodą największej wiarygodności i w podobny sposób interpretuje.

Modele prawdopodobieństwa pozwalają analizować jedynie relatywny wpływ wyróżnionych czynników na prawdopodobieństwo zakupu.

Jeśli wszystkie zmienne niezależne są zmiennymi dyskretnymi (jakościowymi), to model regresji logistycznej sprowadza się do modelu log-liniowego, którego podstawą jest rozkład wielowymiarowy. W modelach tych wszystkie wyróżnione zmienne, zwane tu czynnikami, są uznawane za zmienne niezależne, zmienną zależną jest zaś liczba przypadków, dla których obserwuje się określoną konfigurację czynników niezależnych. Liczebność ta może zależeć od każdego z obserwowanych czynników oddzielnie oraz od interakcji między nimi. Wówczas są to modele nasycone, w których nie występuje składnik losowy.

Formułę modelu log-liniowego wyprowadza się z warunku niezależności stochastycznej:

$$\hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

który po zlogarytmowaniu przyjmuje postać równania log-liniowego:

$$\log \hat{n}_{ij} = -\log n + \log n_i + \log n_j$$

co w postaci symbolicznej zapisuje się jako:

$$\log n_{ij} = \mu + \lambda_{x_i} + \lambda_{y_j}$$

gdzie:

n_{ij} – liczba przypadków jednoczesnego występowania i -tej kategorii czynnika X i j -tej kategorii czynnika Y , dla $i = 1, 2, \dots, k$, oraz $j = 1, 2, \dots, l$;

$$\mu - \text{średnia liczebność, przy czym } \mu = \frac{1}{k \cdot l} \sum_{i=1}^k \sum_{j=1}^l \ln n_{ij};$$

λ_{x_i} – zakres zmiany wartości średniej w wyniku oddziaływania i -tej kategorii czynnika X , przy czym $\sum_{i=1}^k \lambda_{x_i} = 0$;

λ_{y_j} – zakres zmiany wartości średniej w wyniku oddziaływania j -tej kategorii czynnika Y , przy czym $\sum_{j=1}^l \lambda_{y_j} = 0$;

Modele logarytmiczno-liniowe nasycone określa z kolei formuła (dla oceny wpływu np. dwóch czynników X i Y o ustalonej liczbie kategorii):

$$\log n_{ij} = \mu + \lambda_{x_i} + \lambda_{y_j} + \lambda_{x_i y_j}$$

gdzie:

λ_{A, B_j} – zakres zmiany wartości średniej w wyniku oddziaływania interakcji między i -tą kategorią czynnika X i j -tą kategorią czynnika Y , przy czym

$$\sum_{i=1}^k \lambda_{x_i y_j} = 0 \text{ oraz } \sum_{j=1}^l \lambda_{x_i y_j} = 0.$$

Modele logarytmiczno-liniowe nasycone nie zawierają składnika losowego, gdyż wyrażają efekty średnie oraz efekty wynikające z interakcji między czynnikami. Mogą przybierać postać formuł podstawowych, uwzględniających wszystkie możliwe oddziaływania czynników oraz wszystkich możliwych interakcji między nimi lub modeli hierarchicznych, jeśli niektóre oddziaływania zostaną pominięte. W zależności od znaku oszacowanego parametru λ_{x_i} , czynnik X uznaje się za stymulantę (znak dodatni) lub destymulantę (znak ujemny) liczby obserwowanych przypadków.

Model log-liniowy jest dobrze dopasowany do rzeczywistości, jeśli ma miejsce równość liczebności teoretycznych z obserwowanymi lub, co jest równoważne, jeśli odchylenia między tymi liczebnościami mają jedynie charakter przypadkowy (losowy). Sprawdzenie tego założenia odbywa się za pomocą jednego z dwóch równoważnych testów: testu niezależności chi-kwadrat Pearsona lub chi-kwadrat największej wiarygodności. Jeśli model jest zdefiniowany przez dwie zmienne (tak jak przedstawiono to wyżej) i w rezultacie obserwacji ustala się, że występuje zależność między nimi, to wówczas nie występuje równość liczebności teoretycznych z empirycznymi i model założony nie jest dobrze dopasowany. Wówczas taki model powinien być odrzucony na rzecz modelu opisującego zależność.

W przypadku modelu hierarchicznego ocenie podlegają różne stopnie złożoności cech, czyli różne interakcje między cechami. Dąży się przy tym do jak najprostszej formuły modelu. W tym celu przeprowadza się kolejno testy weryfikujące istotność zależności na każdym poziomie złożoności modelu. Dla każdego poziomu złożoności cech przeprowadza się ocenę statystycznej zależności za pomocą testu niezależności chi-kwadrat, zaczynając od efektów głównych.

W rozpoznawaniu relacji między badanym typem zachowań konsumpcyjnych i analizowanymi czynnikami użyteczna może być również analiza korespondencji. Analiza korespondencji rozszerza informację o stwierdzonej zależności stochastycznej między cechami. W ramach tej procedury w pierwszej kolejności należy zatem stwierdzić, np. posługując się przywoływanym już wcześniej testem niezależności chi-kwadrat, czy zależność stochastyczna między badanymi cechami występuje. Brak takiej zależności wyklucza bowiem prowadzenie dalszej eksploracji zmiennych.

Celem analizy korespondencji jest wyznaczenie przestrzeni o możliwie małym wymiarze, w której można dokonać efektywnej prezentacji jednoczesnych wystąpień poszczególnych kategorii badanych cech. Współrzędne tej przestrzeni są wyznaczane w wyniku dekompozycji według wartości osobliwych macierzy:

$$(W - w_i w_j^T)$$

gdzie:

$W = [w_{ij}]$ – macierz częstości względnych wyznaczonych w tablicy kontyngencji rozkładu cech wielonominalnych, w_i , w_j – częstości brzegowe, odpowiednio wierszy i kolumn tej macierzy.

Dekompozycja ta wymaga wyznaczenia macierzy standaryzowanych różnic oraz jej dekompozycji. Macierz standaryzowanych różnic przyjmuje postać:

$$A = D_r^{-1/2} (W - w_i w_j^T) D_c^{-1/2}$$

gdzie:

D_r – diagonalna macierz o elementach w_i na głównej przekątnej, D_c – diagonalna macierz o elementach w_j na głównej przekątnej.

Dekompozycja macierzy A przebiega zaś według wzoru:

$$A = U \Gamma V^T$$

gdzie:

$\Gamma = [\gamma_{ii}]$ – macierz diagonalna wartości własnych, U – macierz lewych wektorów osobliwych macierzy A , V – macierz prawych wektorów osobliwych macierzy A .

Współrzędne kategorii obu cech są obliczane według wzoru:

– dla cechy zapisanej w wierszach tablicy kontyngencji: $Z = D_r^{-1/2} U \Gamma$

– dla cechy zapisanej w kolumnach tablicy kontyngencji: $Y = D_c^{-1/2} V \Gamma$

Zachodzi następująca zależność:

$$\chi^2 = n \gamma = n \sum_{k=1}^K \gamma_k^2$$

gdzie:

χ^2 – statystyka mierząca siłę zależności stochastycznej badanych cech zgodnie z testem niezależności chi-kwadrat, $\gamma = \sum_{k=1}^K \gamma_k^2$ – inercja całkowita, będąca sumą wszystkich wartości własnych macierzy A . Jest ona miarą rozproszenia punktów w rzeczywistej przestrzeni rzutowania (Greenacre 1984, 1993).

5. Identyfikacja czynników determinujących zachowania konsumentów

Odrębnym problemem jest identyfikacja tych zmiennych niezależnych, które w modelowaniu prawdopodobieństwa wystąpienia zachowania konsumentkiego (zgodnie z przedstawionymi wyżej zasadami) powinny zostać wyróżnione. W tym wypadku dogodnym narzędziem wydaje się być analiza dyskryminacji, gdyż obie grupy metod – analiza regresji logistycznej oraz analiza dyskryminacji – wymagają podobnych założeń co do zbioru zmiennych niezależnych. Zastosowanie obu metod prowadzi zatem do spójnych wyników.

Analiza dyskryminacji stwarza możliwości konstrukcji formuły matematycznej, identyfikującej przynależność obiektów (np. konsumentów czy odbiorców) do jednej z kilku wyróżnionych grup. Najprostszym przypadkiem jest podział na dwie grupy, np. nabywcy produktu oraz rezygnujący z zakupu. Konstrukcja reguły przynależności opiera się na wielowymiarowej informacji o każdej jednostce badanej (obiekcie) oraz o jej przynależności do określonej grupy. W analizie dyskryminacji definiuje się zmienną grupującą, określaną mianem zmiennej zależnej, której wartości pozwalają na zdefiniowanie poszczególnych grup. Definiuje się również zmienne niezależne, które charakteryzują poszczególne obiekty należące do każdej z wyróżnionych grup. Reguła przynależności jest formułą matematyczną, zbudowaną ze zmiennych niezależnych. Zmienne te noszą również nazwę zmiennych diagnostycznych, gdyż mając do dyspozycji skonstruowaną wcześniej regułę przynależności, można jedynie na ich podstawie rozstrzygnąć o przynależności obiektu do określonej grupy. Stwarza to dodatkowo możliwości przewidywania kształtowania się struktur rynkowych.

Na podstawie wielowymiarowej informacji o każdym z obserwowanych obiektów, oraz o jego przynależności do określonej grupy, konstruuje się funkcję, zwane funkcjami dyskryminacji, przekształcającą wielowymiarową informację o badanym obiekcie w skalar. Funkcje te wyznacza się w taki sposób, by maksymalizować zróżnicowanie ich wartości między grupami względem zróżnicowania wewnątrz grup. Poszukuje się zatem takiego rozwiązania, by wewnątrz istniejących grup jednostek uzyskać jak największą jednorodność wartości funkcji przynależności, między grupami zaś jak największą heterogeniczność.

Postać tych funkcji jest liniowa:

$$Y_i = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p$$

Dla określonej liczby funkcji dyskryminacji l (przy czym liczba funkcji jest zawsze o 1 mniejsza od liczby kategorii zmiennej grupującej) ich współczynniki to współrzędne wektorów a_j , będących rozwiązaniami równania (Garbiński 1992: 57):

$$(B - \lambda_j W) a_j = 0,$$

gdzie:

λ_j – nieujemne pierwiastki równania wyznacznikowego: $|B - \lambda W| = 0$, uporządkowane w ciąg nierosnący; B – macierz wariancji międzygrupowych zmiennych niezależnych, W – macierz wariancji wewnątrzgrupowych zmiennych niezależnych; j – numer funkcji dyskryminacji, $j = 1, 2, \dots, l$.

Suma zmienności między- i wewnątrzgrupowej wartości funkcji dyskryminacji wyznacza zmienność całkowitą wartości tej funkcji. Dla modelu z wieloma funkcjami dyskryminacyjnymi warunek ten przyjmuje postać:

$$T = B + W$$

Współczynnik λ , skonstruowany według wzoru:

$$\lambda = \frac{\det W}{\det(B + W)},$$

nosi nazwę współczynnika λ Wilksa i informuje o tym, jaka część zmienności funkcji dyskryminacji nie jest wyjaśniana różnicami między grupami. Współczynnik ten określa, w jakim stopniu wartości funkcji dyskryminacji są zakłócone innymi czynnikami niż przynależność do wyróżnionych dwóch grup. Jeśli wartość tego współczynnika jest wysoka, wskazuje to na brak podstaw do klasyfikacji obiektów według grup. Natomiast im niższa jest wartość statystyki λ , tym poprawniejsza jest formuła dyskryminacyjna.

Miarą siły dyskryminacyjnej całego zbioru zmiennych diagnostycznych może być współczynnik korelacji kanonicznej:

$$R_c = \sqrt{1 - \lambda},$$

który dla jednej funkcji dyskryminacji, rozdzielającej badaną zbiorowość jedynie na dwie grupy (czyli gdy $k=2$), przyjmuje postać współczynnika znanego w statystyce jako współczynnik *eta* i jest wyznaczany według wzoru:

$$R_c = \eta = \sqrt{\frac{q_G}{q_G + q_R}}$$

Współczynnik ten przyjmuje wartości z przedziału $<0 ; 1>$ i jest tym większy, im większa jest siła dyskryminacyjna całego zbioru zmiennych diagnostycznych. Symptomami poprawnego rozwiązania są zatem:

- wysoka wartość współczynnika korelacji kanonicznej R_c ,
- niska wartość współczynnika λ Wilksa.

Jeśli badanie jest prowadzone na próbie losowej, należy ocenić uzyskane wartości z punktu widzenia ich istotności statystycznej. W procedurze analizy dyskryminacji weryfikuje się zgodność współczynnika λ Wilksa z wartością 1. Jeśli brak jest podstaw do odrzucenia takiej hipotezy, to klasyfikacja metodą analizy dyskryminacji nie daje istotnych statystycznie wyników, jeśli zaś można uznać, że współczynnik λ Wilksa istotnie różni się od 1, to klasyfikację według analizy dyskryminacji można uznać za istotną statystycznie.

Miarą rozbieżności wartości współczynnika λ Wilksa z wartością 1 jest statystyka χ^2 obliczana według wzoru:

$$\chi^2 = -\left(n - \frac{k+p}{2} - 1\right) \ln\left(\frac{1}{1+\lambda}\right)$$

gdzie:

n – liczba obserwacji;

k – liczba grup, w przypadku decyzji zakupu $k = 2$;

p – liczba zmiennych diagnostycznych;

λ – wartość statystyki λ Wilksa.

Współczynniki skonstruowanej funkcji przynależności informują o sile i kierunku oddziaływania poszczególnych zmiennych diagnostycznych na przynależność jednostki do jednej z wyróżnionych grup. Współczynniki te nie są porównywalne, chyba że wyznaczono ich wystandaryzowane wartości. Wygodną miarą siły dyskryminacyjnej poszczególnych zmiennych niezależnych jest współczynnik korelacji, wyznaczany dla każdej ze zmiennych mierzalnych i wartości funkcji przynależności. Kwadrat tego współczynnika, choć faktycznie oznacza stopień wyjaśniania zmienności danej zmiennej niezależnej przez zmienną dyskryminującą, to z racji swej symetryczności może służyć do określania zakresu determinacji wartości funkcji dyskryminacji przez poszczególne zmienne niezależne. Im korelacja ta jest większa, tym za mocniejszą można uznać siłę dyskryminacyjną analizowanej zmiennej niezależnej, co oznacza, że odgrywa ona znaczną rolę w decydowaniu o przynależności badanych obiektów do wyróżnionych grup.

6. Identyfikacja czynników mających wpływ na zmiany wyborów konsumenckich wraz z upływem czasu

W przypadku gdy dane dotyczące zachowań konsumenckich mają charakter panelowy lub retrospektywny, a szczególnie, gdy monitoring obejmuje dłuższy czas, możliwe jest modelowanie wyborów konsumenckich z punktu widzenia oddziaływania wyróżnionych czynników jako funkcji czasu. Wówczas dogodnym narzędziem analizy jest model regresji logistycznej Cox'a. Regresja Cox'a należy do narzędzi analizy historii zdarzeń (Frątczak 1997). W analizach tych przedmiotem oceny są prawdopodobieństwa wystąpienia zdarzenia (zmiany stanu) nie wcześniej niż po upływie określonego okresu. Prawdopodobieństwa te opisuje funkcja dożycia. Zmienność tej funkcji opisuje zaś funkcja hazardu. W analizie historii zdarzeń definiuje się m. in. modele semi-parametryczne, w których zakłada się występowanie zmiennych objaśniających intensywność przejść między stanami, czyli zakłada się podatność funkcji hazardu na działanie zmiennych niezależnych. Modele te stwarzają możliwość estymacji funkcji hazardu dla utraty klienta w zależności od upływu czasu i innych zmiennych o charakterze niezależnym. Formułę modelu określa równanie:

$$h(t) = h_0(t) \cdot e^{\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k}$$

gdzie:

$h_0(t)$ – część bazowa funkcji hazardu, będąca funkcją czasu, identyczna dla wszystkich jednostek,

α_i – współczynnik regresji opisujący rolę zmiennej X_i w kształtowaniu ryzyka utraty klienta wraz z upływem czasu.

Formuła modelu jest zbieżna z modelem regresji logistycznej, a dokładniej z możliwością oszacowania na podstawie tego modelu ryzyka względnego utraty klienta. Wyrażenie e^{α_j} określa relatywny wpływ zmiennej X_j na możliwość utrzymania klienta wraz z upływem czasu. Jeśli $e^{\alpha_j} = 1$, to zmienna X_j jest neutralna wobec możliwości utrzymania klienta wraz z upływem czasu, jeśli $e^{\alpha_j} > 1$, to zmienna X_j działa stymulująco, zwiększając możliwość utrzymania klienta, jeśli zaś $e^{\alpha_j} < 1$, to zmienna X_j działa limitująco, obniżając tę możliwość i zwiększając ryzyko utraty klienta. Dla każdego wyniku otrzymanego z próby wyznacza się średni błąd szacunku i statystykę Walda, oceniającą jego istotność.

Zagadnienia opisane w modelu Cox'a odnoszą się wprost do zagadnień związanych z przywiązaniem do marki w zachowaniach konsumenckich, czyli lojalnością nabywców. Model Cox'a wymaga badań longitudinalnych, które są jednakże dość kosztowne i trudne ze względów organizacyjnych.

Podejście statystyczne oferuje wiele wygodnych procedur do identyfikacji poziomu preferencji konsumenckich oraz czynników, które mają zasadniczy wpływ na kształtowanie się układów tych preferencji. Jednakże, obok dużych walorów użytkowych, podejście to stawia wiele formalnych wymagań, z których najtrudniejsze do spełnienia jest zachowanie statystycznej niezależności względem czynników decydujących o zakupie. Niespełnienie tych założeń lub spełnienie ich w niewystarczającym stopniu przekłada się na niski stopień dokładności modelowania oraz w konsekwencji na znaczne błędy *ex post* dokonywanych symulacji. Drugim ważnym czynnikiem, który może również ograniczać wykorzystanie przedstawionych procedur, jest często wymagana znaczna liczebność obserwacji, co przekłada się wprost na znaczne koszty pozyskania poszukiwanych danych. Warto natomiast podkreślić, że poważna bariera stosowania przedstawionych metod oceny preferencji i ich struktury, wynikająca ze skomplikowania rachunków niezbędnych do realizacji opisanych procedur, została praktycznie przełamana dzięki znacznej dostępności oprogramowania komputerowego.

Jak już wspomniano wyżej, przedstawione podejścia pomiaru i analizy udziału w rynku mogą być stosowane zarówno do obserwowanych zachowań konsumentów, jak i jedynie wyrażanych przez nich intencji zakupu, pozwalających przewidywać wielkość sprzedaży oraz udział w rynku w okresie późniejszym bądź też dla produktów, których wprowadzenie na rynek jest dopiero rozważane. W takim podejściu zakłada się, że obecna sytuacja jest jedynie punktem wyjścia do wizji przyszłości, którą kreuje samodzielnie konsument. Składana przez niego deklaracja przyszłych zachowań konsumpcyjnych jest wynikiem indywidualnej prognozy swej sytuacji „konsumpcyjnej” w przyszłości. Podejście to, oparte na deklaracjach zakupu, zakłada agregację indywidualnych prognoz konstruowanych w świadomości konsumenta na podstawie jego indywidualnych kryteriów optymalizacyjnych i indywidualnych zależności przyczynowo-skutkowych. Z pomiarem potencjalnego lub przyszłego udziału w rynku na podstawie jedynie intencji zakupu wiąże się jednak pewne trudności. Podejście to jest bowiem obciążone brakiem pewności co do prawdziwości składanych przez konsumentów deklaracji. Problem „prawdomówności” respondentów stanowi wówczas jedno z ważniejszych zagadnień w analizie uzyskanych wyników. Po to, by wyeliminować wpływ czynników skłaniających respondentów do składania deklaracji „na wyrost”, a które można określić mianem czynników zakłócających, stosuje się warunki pomiaru w największym stopniu zbliżone do

rzeczywistej sytuacji rynkowej. Sama organizacja badania nie może jednakże wyeliminować wpływu takich czynników, które stoją po stronie respondenta, a które sprawiają, że w przeszłości zachowa się on odwrotnie do złożonych deklaracji.

Praktyka wskazuje, że rozbieżności między deklaracjami składanymi przez respondentów w trakcie badania ankietowego a obserwowanymi później zachowaniami rynkowymi nie mają charakteru losowego i w efekcie rozbieżności *in plus* i *in minus* nie „znoszą się” wzajemnie. Wyniki estymacji charakteryzuje znaczny błąd *ex post*, gdyż na ogół deklaracje respondentów istotnie odbiegają od późniejszych zachowań rynkowych.

Problem tu poruszany koncentruje się wokół zagadnienia możliwości przewidywania kierunku i stopnia zniekształceń deklaracji co do intencji zakupu względem późniejszych zachowań rynkowych. Najczęściej rekomendowanym podejściem jest tu podejście bayesowskie, uwzględniające dodatkową informację na temat możliwości zachowań konsumentów. Przy czym istnieją w tym wypadku różne sposoby estymacji prawdopodobieństwa *a priori*, odnoszących się do dodatkowej informacji o prawdziwości respondentów (Rószkiewicz 2002b).

Informacje o autorce

Dr hab. Prof. SGH Małgorzata Rószkiewicz, Zakład Metod Badań Marketingowych, Instytut Statystyki i Demografii, Szkoła Główna Handlowa w Warszawie,
e-mail: MROSZKIEWICZ@qdnnet.pl

Przypisy:

- 1 W niniejszym artykule dokonano prezentacji podstawowych podejść w modelowaniu zachowań konsumenckich w ujęciu teoretycznym. Przykłady zastosowań poszczególnych modeli formalnych można znaleźć w: Rószkiewicz 2002a i 2002b.
- 2 Przy 5-procentowym ryzyku przekroczenia tej wartości wartość statystyki u o rozkładzie normalnym wynosi 1,96.

Bibliografia:

- Arminger G., Clogg C.C., Sobel M.E. 1995. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York and London: PLENUM PRESS.
- Armitage, P. 1978. *Metody statystyczne w badaniach medycznych*, Warszawa: PZWL.
- Frątczak E. 1997. *Analiza historii zdarzeń – elementy teorii, wybrane przykłady zastosowań z wykorzystaniem pakietu TDA*, Warszawa: SGH.
- Garbiński T. 1992. *Metody taksonometrii*. Kraków: Akademia Ekonomiczna w Krakowie.
- Greenacre M.J. 1984. *Theory and Applications of Correspondence Analysis*, London: Academic Press.
- Greenacre M.J. 1993. *Correspondence Analysis in Practice*, London, Academic Press.
- Jajuga K., 1990. Modele z dyskretną zmienną objaśnianą, w: *Estymacja modeli ekonometrycznych*, red. nauk. S. Bartosiewicz, Warszawa: PWE.
- Kotler Ph. 1991. *Marketing Management*, New Jersey: Prentice Hall.
- Lambin J.J. 2001. *Strategiczne zarządzanie marketingowe*, Warszawa: PWN.
- Ostaszewicz W. (red.) 1999. *Metody ilościowe w ekonomii*, Wrocław: Wyd. Akademii Ekonomicznej im. Oskara Łangego we Wrocławiu.
- Parfitt J.H., Collins B.J.K. 1968. The Use of consumer Panels for Brand Share Prediction, *Journal of Marketing Research*, nr 5, May, s.131-145.
- Rószkiewicz M. 2002 a. *Metody ilościowe w badaniach marketingowych*, Warszawa: PWN.
- Rószkiewicz M. 2002 b. *Narzędzia statystyczne w analizach marketingowych*, Warszawa: C.H. Beck.
- Rószkiewicz M. 1998. *Zarys metod statystyki wielowymiarowej z wykorzystaniem programów komputerowych*, Warszawa: SGH.
- Sagan A. 1995. Podjęcie do badań segmentacji rynku. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, nr 457. Kraków.
- Strzyżewska M., Rószkiewicz M. 2002. *Analizy marketingowe*, Warszawa: Difin.