Bronisław K. Bednarek

ZASTOSOWANIE ANALIZY PROBITOWEJ I LOGITOWEJ DO BADANIA ZJAWISK SPOŁECZNO-EKONOMICZNYCH

1. Wprowadzenie

Celem każdego badania naukowego jest opis i wyjaśnienie jakiegoś wycinka rzeczywistości. Cel ten jest realizowany poprzez dążenie do udzielenia odpowiedzi na dwa rodzaje pytań[1]: 1) jaki jest ten wycinek? 2) dlaczego taki właśnie jest? Udzieleniu poprawnych odpowiedzi służy odpowiedni dobór narzędzi badawczych, czyli metod i technik właściwych danej dyscyplinie naukowej. Dążenie do możliwie pełnego opisu i objaśnienia wybranej części rzeczywistości przekracza często możliwości jednej dziedziny wiedzy i wymusza podjęcie badań w zakresie wielu dyscyplin, a nawet interdyscyplinarnych[2]. Takie postępowanie sprzyja wzajemnemu przenikaniu i adaptacji metod wypracowanych w różnych dyscyplinach naukowych, przyczyniając się zarówno do ich rozwoju, jak i do pogłębienia analizy. Dotyczy to również metod omawianych w tej pracy.

Opisu przedmiotu badania dokonujemy poprzez przypisanie mu określonych właściwości natury jakościowej bądź ilościowej, a najczęściej jednych i drugich. Każdą właściwość charakteryzuje zespół cech wymiernych lub opisowych, czyli takich, których istotę można wyrazić tylko słownie. Dysponując wybranym zespołem cech, staramy się następnie wykryć ewentualne związki i zależności między nimi. Jeżeli przedmiotem badania są zjawiska lub procesy o charakterze masowym, wtedy do ich wyjaśnienia można zastosować metody matematyczno-

---

[1] Por. K. Ajdukiewicz, *Logika pragmatyczna*, Warszawa 1965, s. 395.

[2] Na temat badań wielodyscyplinowych wypowiadał się m.in. S. Wierzchosławski w pracy: *Węzłowe problemy metodologii badań statystycznych zjawisk społeczno-ekonomicznych w Polsce*, w: *Rozwój metodologii badań statystycznych w Polsce*, Biblioteka Wiadomości Statystycznych, Warszawa 1995, t. 44, s. 32–37.

-statystyczne, a w szczególności metody analizy wielowymiarowej. Jest jednak jeden warunek: cechy opisujące właściwości obiektów w ich zbiorowościach muszą być wyrażone liczbowo.

Zbudowanie modelu matematycznego opisującego rzeczywiste kształtowanie się danego zjawiska i wyjaśniającego przyczyny takiego kształtowania się zależy głównie od właściwego doboru zmiennych i zastosowania odpowiednich metod weryfikacji. Zmienne w modelu powinny być ekwiwalentne[3], to znaczy powinny możliwie najlepiej odzwierciedlać treść badanego zjawiska (zmienne objaśniane) bądź najlepiej reprezentować czynniki oddziałujące na to zjawisko (zmienne objaśniające). Metody natomiast powinny być dostosowane do charakteru liczb służących pomiarowi tych zmiennych. Charakter i treść informacyjna liczb zależy od zastosowanych skal pomiarowych[4], ponieważ na ogół wyniki pomiaru w skali mocniejszej przenoszą więcej informacji niż w skali słabszej. Również zakres metod poszerza się wraz ze wzmocnieniem skali. Nie zawsze jednak większa ilość informacji przynosi pożądany efekt w procesie badawczym. Badania empiryczne wykazują, że w niektórych przypadkach zrezygnowanie z części informacji, poprzez przejście do pomiaru w skali słabszej, przynosi wymierne korzyści w postaci lepszego modelu i wykrycia związków nie ujawnionych wcześniej[5]. Należy przy tym zauważyć, że przekształcenie liczb wyrażonych w skali mocniejszej na liczby w skali słabszej jest zawsze możliwe, natomiast nie jest możliwe

[3] Por. J. W. Wiśniewski, *Ekonometryczne badanie zjawisk jakościowych. Studium metodologiczne*, UMK, Toruń 1986, s. 31 – 34.

[4] Wyróżnia się 4 skale pomiarowe: nominalną, porządkową, przedziałową i ilorazową. W skali nominalnej liczby służą do identyfikacji lub wyróżniania rozłącznych kategorii (typów) elementów, a w skali porządkowej liczby dodatkowo wyznaczają kolejności występowania (rangi) elementów w zbiorze. Zob. J. Steczkowski, A. Zeliaś, *Statystyczne metody analizy cech jakościowych*, PWE, Warszawa 1981, s. 15–18. Liczby w skali przedziałowej pozwalają dodatkowo na określenie odległości między elementami zbioru lub właściwościami zjawisk i mogą być przekształcane liniowo. Liczby w skali ilorazowej mogą być traktowane jako odległości od zera, a skala posiada właściwości skal poprzednich oraz naturalny punkt zerowy. Na liczbach można wykonywać działania: w skali nominalnej – liczenie, w skali porządkowej – prócz liczenia przekształcenia nie zmieniające porządku elementów, w skali przedziałowej – dodatkowo dodawanie arytmetyczne, w skali ilorazowej – także mnożenie i dzielenie. Na liczbach z poszczególnych skal dopuszczalne jest stosowanie następujących technik i procedur statystycznych: w skali nominalnej – opartych na liczeniu; w skali porządkowej – także opartych na rangowaniu (kwantyle, korelacja rang Spearmana), testy nieparametryczne; w skali przedziałowej – wyznaczanie średniej arytmetycznej i wariancji, rachunek korelacji i regresji, niektóre testy parametryczne; w skali ilorazowej – wszystkich. Por. m.in. *Problemy psychologii matematycznej*, praca zbiorowa, red. J. Kozielecki, Warszawa 1971, s. 26–28.

[5] Przykład taki omawiają J. Steczkowski i A. Zeliaś w cytowanej pracy, s. 17. Konieczność przejścia od skali mocniejszej do słabszej pokazana jest też w artykule: D. J. Aigner, A. S. Goldberger, A. J. Kalton, *On the Exploatory Power of Dummy Variable Regressions*, „International Economics" Review, t. 16, 1975, nr 2.

przekształcenie odwrotne (przez przekształcenie można stracić część informacji, lecz nie można zyskać). Zatem z jednej strony charakter danych wpływa limitująco na wybór metod statystycznych, z drugiej strony wybrana metoda może wymagać nadania liczbom odpowiedniego charakteru.

W różnych dziedzinach ludzkiej działalności dość powszechnie występują zjawiska i procesy noszące miano jakościowych. Każde zjawisko jakościowe może być w pewien sposób wyrażone za pomocą liczb. Jest to możliwe poprzez opisujące właściwości tych zjawisk cechy jakościowe, których pomiar dopuszczalny jest tylko w dwóch najsłabszych skalach: nominalnej i porządkowej[6]. Cechy te mogą przyjmować co najwyżej przeliczalną, a praktycznie skończoną liczbę wartości, stanowiąc wyodrębnioną grupę zmiennych dyskretnych (skokowych). Wśród nich na szczególną uwagę zasługuje zmienna zero-jedynkowa reprezentująca zjawisko o charakterze dychotomicznym, w którym zaobserwować można tylko dwa wykluczające się zdarzenia. Zdarzeniom tym mogą być przyporządkowane arbitralnie liczby 0 lub 1. Ta szczególna uwaga skierowana na zmienną zero-jedynkową wynika z kilku powodów:

1. Zarówno w otaczającej nas rzeczywistości, jak i w dziedzinie badań naukowych, szczególnie w naukach społecznych, zjawiska o charakterze dualnym występują bardzo często (śmierć lub przeżycie, urodzenie kolejnego dziecka lub nie, rozpad małżeństwa lub jego przetrwanie, skuteczność zastosowanej terapii lub jej nieskuteczność, upadek bądź dalsze funkcjonowanie firmy, zakup samochodu lub nie, posiadanie lub nieposiadanie własnego domu, recydywa lub nie).
2. Każdą cechę wyrażoną w skali mocniejszej (ciągłą, wielopunktową) można poddać dychotomizacji.
3. Egzogeniczna zmienna zero-jedynkowa, jako zmienna sztuczna, wprowadzona do modelu ekonometrycznego zawierającego zmienne wyrażone w skalach mocniejszych, przyczynia się często do poprawy jego jakości.
4. W ujęciu masowym zmienna zero-jedynkowa wykazuje bezpośredni związek z prawdopodobieństwem. Jeżeli pojawienie się jedynki w ciągu obserwacji jest zawsze jednakowo możliwe, to liczba zaobserwowanych jedynek ma rozkład dwumianowy, którego rozkładem granicznym jest najczęściej rozkład normalny[7].
5. Jakkolwiek wartości informacyjne pojedynczych obserwacji wyrażonych w skali dychotomicznej[8] są mniejsze niż w pozostałych skalach, to rozpatrywane w kategoriach procesów masowych nabierają podobnego znaczenia jak

---

[6] W ogólniejszym ujęciu jako cechę jakościową rozumie się cechę złożoną będącą agregatem prostych cech kwantytatywnych (mierzalnych) i kwalitatywnych (opisowych). Por.: J. W. Wiśniewski, op. cit., s. 29; T. Borys, *Jakość – pojęcie wieloznaczne*, Prace Naukowe WSE we Wrocławiu, 1972, nr 33.

[7] Wiadomości o tych rozkładach można znaleźć w każdym podręczniku do statystyki matematycznej.

[8] Skalę tę można traktować jako szczególny przypadek skali nominalnej.

w skali ilorazowej, umożliwiając stosowanie analizy korelacji i regresji. Największą wartość informacyjną zmienna zero-jedynkowa osiąga wtedy, gdy liczba obu wykluczających się możliwości jest jednakowa, w tym bowiem wypadku ma największą dyspersję. W tym sensie zmienna ta wykazuje wyższość nad zmienną rangową (porządkową)[9].

Zmienną zero-jedynkową postrzegać można szerzej niż tylko w kontekście alternatywy zdarzeń i nadać jej walor ogólności[10]:

$$X = \begin{cases} 1, & \text{gdy spełniony jest zbiór warunków } W, \\ 0, & \text{w przeciwnym wypadku.} \end{cases}$$

Na warunki te mogą składać się wartości cech kwantytatywnych lub opisowych badanego zjawiska zaobserwowanych w próbie.

Metody analizy cech jakościowych, w tym również zmiennych zero-jedynkowych, zaczęły powstawać stosunkowo wcześnie i obecnie ich zbiór jest dość zróżnicowany[11]. Metody te w głównej mierze koncentrują się na wykrywaniu związków (najlepiej przyczynowo-skutkowych) między cechami, na poszukiwaniu reguł funkcjonowania tych związków, także w wypadku, gdy zmienna endogeniczna ma charakter dychotomiczny. Miarami związku między cechami jakościowymi są różne współczynniki asocjacji (skojarzenia), które określają jego natężenie, a w wypadku współczynnika korelacji także kierunek[12]. Jeżeli jednak interesuje nas sposób wzajemnego oddziaływania cech i możliwość przewidywania ich zmian, będących efektem tego oddziaływania, powinniśmy posłużyć się modelem regresji.

W klasycznych modelach korelacji i regresji na zmienną endogeniczną nie nakłada się żadnych ograniczeń co do zakresu zmienności w jej rozkładzie warunkowym. W przypadku zmiennej zero-jedynkowej zakres ten jest ograniczony do przedziału $\langle 0, 1 \rangle$. Odpowiednim rozwiązaniem wydaje się zastosowanie liniowych modeli prawdopodobieństwa lub zlinearyzowanych modeli nieliniowych należących do grupy modeli z ograniczonymi zmiennymi zależnymi[13]. Modele te różnią się między sobą właściwościami ich parametrów strukturalnych, lecz są zbliżone pod względem możliwości stosowania technik obliczeniowych. Podstawowym modelem w tej

[9] Por. J. W. Wiśniewski, op. cit. s. 17–22, 37–43.

[10] Por. Z. Pawłowski, *Elementy ekonometrii*, Warszawa 1981, s. 170.

[11] Podstawy teoretyczne badania zjawisk jakościowych znaleźć można w książce: G. U. Yule, M. G. Kendall, *Wstęp do teorii statystyki*, PWN, Warszawa 1966. Z polskich opracowań całkowicie poświęconych tym zagadnieniom na szczególną uwagę zasługują monografie: J. W. Wiśniewski, op. cit. oraz J. Steczkowski, A. Zeliaś, op. cit.

[12] Miary te szczegółowo omawia J. W. Wiśniewski, op. cit., s. 49–67.

[13] Modele te gruntownej analizie poddaje J. W. Wiśniewski w cytowanej pracy.

grupie jest model A. S. Goldbergera, zwany też liniową funkcją prawdopodobieństwa, postaci[14]:

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots n, \tag{1}$$

gdzie y_i jako realizacje zmiennej zero-jedynkowej przyjmują wartość 1 tylko w przypadku zajścia danego zdarzenia lub ogólniej, gdy spełniony zostanie zespół warunków W_i, co wiąże się z odpowiednią konfiguracją wartości x_{ij} jako realizacji zmiennych objaśniających $X_j (j = 1, 2, ..., k)$. Parametry strukturalne $\alpha_j (j = 0, 1, ..., k)$ informują o wpływie zmiennych X_j na prawdopodobieństwo wystąpienia warunków W_i, natomiast ε_i jest składnikiem losowym wyrażającym wpływ tych czynników, które nie są reprezentowane przez te zmienne, przy czym $E(\varepsilon_i) = 0$. Zmienna zależna w modelu (1) jest traktowana tak samo jak w zwykłym modelu regresji liniowej. Jej wartość oczekiwana

$$E(y_i) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_k x_{ik}, \quad i = 1, 2, \ldots n, \tag{2}$$

jako liniowa funkcja zmiennych objaśniających, jest interpretowana jako warunkowe prawdopodobieństwo realizacji warunków W_i przy ustalonych wartościach tych zmiennych. Teoretyczna wartość y wyznaczona z (2) może być uważana za oszacowanie tego prawdopodobieństwa.

To proste rozwiązanie problemu budzi jednak wątpliwości i zastrzeżenia. Mianowicie nie jest możliwe przyjęcie założenia o homoscedastyczności składnika losowego, co jest wymagane w klasycznym modelu liniowym. Jeżeli bowiem wartość oczekiwana składnika losowego $E(\varepsilon_i) = 0$, to jego wariancja jest równa

$$D^2(\varepsilon_i) = E(y_i(1 - E(y_i))),$$

czyli zależy od wartości $E(y_i)$, a w konsekwencji zmienia się wraz ze zmianami zmiennych X_j i świadczy o jego heteroscedastyczności.

W tej sytuacji stosowanie klasycznej metody najmniejszych kwadratów (KMNK) nie jest właściwe, bowiem oszacowany tą drogą estymator wektora parametrów równania (1), wprawdzie nieobciążony, jest mało efektywny, a estymator jego wariancji jest obciążony[15]. Może to prowadzić w analizie współzależności

[14] Por. A. S. Goldberger, *Teoria ekonometrii*, PWE, Warszawa 1972, s. 319–320; M. Rószkiewicz, *Zarys metod statystyki wielowymiarowej z wykorzystaniem programów komputerowych STATGRAPHICS w.6* oraz *SPSS w.5*, SGH, Warszawa 1998, s. 74; J. W. Wiśniewski, op. cit., s. 72.

[15] Por. m. in.: A. S. Goldberger, op. cit., s. 321; H. Theil, *Zasady ekonometrii*, Warszawa 1979, s. 635; M. A. Kooyman, *Dummy variables in Econometrics*, Tilburg University Press, The Netherlands, Rotterdam 1976, s. 66.

zjawisk do błędnych wniosków. Uwaga ta dotyczy zarówno modeli liniowych typu (1), jak i zlinearyzowanych względem parametrów strukturalnych i składnika losowego modeli nieliniowych. W takim wypadku właściwym byłby model uogólnionej regresji liniowej wymagający zastosowania uogólnionej metody najmniejszych kwadratów (UMNK). Jednak i to podejście nie rozwiązuje problemu do końca, bowiem nie zapewnia utrzymania teoretycznych wartości prawdopodobieństwa w przedziale $\langle 0, 1 \rangle$. Ponieważ funkcja liniowa jest nieograniczona, przeto mogą istnieć takie układy wartości zmiennych objaśniających, dla których $E(y_i)$ przyjmie wartość większą od 1 lub mniejszą od 0. Jest to sprzeczne z definicją zmiennej zależnej i budzi poważne wątpliwości wobec interpretowania jej wartości oczekiwanej jako prawdopodobieństwa[16].

Istnieje kilka możliwości zaradzenia tym trudnościom:

- odrzucenie obserwacji skrajnych i ponowne oszacowanie modelu[17];
- zdefiniowanie teoretycznych wartości zmiennej objaśnianej w ten sposób, aby były równe wartości oszacowanej, jeżeli należy ona do przedziału $\langle 0, 1 \rangle$, równe 1 dla wartości oszacowanych przewyższających 1 i równe 0 dla wartości oszacowanych mniejszych od 0[18];
- zastosowanie innych postaci analitycznych modelu regresji, szczególnie takich, w których konieczne jest zmniejszenie zakresu zmienności zmiennych objaśniających (np. przez logarytmowanie w modelu potęgowym[19]);
- przedefiniowanie zmiennej objaśnianej tak, aby przyjęcie przez nią wartości teoretycznych spoza przedziału $\langle 0, 1 \rangle$ było niemożliwe.

Jedną z możliwości spełnienia wymogów ostatniej propozycji stwarza oparcie się na obserwacjach zagregowanych (zamiast na obserwacjach indywidualnych i oryginalnej zmiennej zero-jedynkowej) poprzez częstości w wyróżnionych grupach, a następnie powiązanie jej, za pomocą przekształcenia częstości, z dystrybuantą odpowiedniego rozkładu prawdopodobieństwa. Spośród różnych przekształceń zmiennych zero-jedynkowych reprezentowanych przez częstości najbardziej znane i najczęściej stosowane są transformacje probitowe i logitowe.

Na świecie zainteresowanie modelami wykorzystującymi te transformacje nie osłabło w ciągu minionego półwiecza i wciąż jest żywe, o czym świadczą liczne prace teoretyczne, poświęcone ich uogólnieniu i ulepszeniu technik estymacyjnych

---

[16] Wykazuje to A. S. Goldberger w cytowanej pracy, s. 307–312.

[17] Taką sugestię znaleźć można w pracy J. Steczkowskiego: *Statystyczna procedura określania struktury zbiorowości,* Zeszyty Naukowe WSE, nr 21, Kraków 1970. Jednak odrzucenie pewnej liczby obserwacji znacznie pogarsza, szczególnie w małej próbie, dokładność oszacowań.

[18] Por. J. W. Wiśniewski, op. cit., s. 110. Jest to zabieg nieco sztuczny i mało godny polecenia przy znacznej liczbie przypadków wyjścia wartości teoretycznych poza przedział $\langle 0,1 \rangle$.

[19] Zmniejsza to wprawdzie ryzyko wyjścia wartości teoretycznych poza przedział $\langle 0,1 \rangle$, lecz go nie wyklucza.

oraz jeszcze liczniejsze prace aplikacyjne, poszerzające obszar możliwych zastosowań. W Polsce analizą probitową zaczęto interesować się na początku lat sześćdziesiątych, do czego impuls dała prawdopodobnie praca J. Tobina z roku 1955 ogłoszona drukiem w „Przeglądzie Statystycznym" roku 1960[20], jednak nie zyskała ona większej popularności. Najwięcej prac poświęconych analizie probitowej i jej zastosowaniu powstało właśnie w dekadzie lat sześćdziesiątych, później prace związane z tym zagadnieniem ukazywały się sporadycznie[21]. Analiza logitu do tej pory praktycznie nie wzbudziła w Polsce większego zainteresowania[22].

Ze względu na duże zainteresowanie tymi metodami na świecie, zarówno w sferze teorii, jak i zastosowań praktycznych, warto poświęcić im więcej uwagi. Celem pracy jest właśnie wskazanie na możliwości zastosowania analizy probitu i logitu do badania związków przyczynowo-skutkowych między cechami jakościowymi, w szczególności wyrażonymi poprzez zmienne zero-jedynkowe. Ponieważ szczegółowe opisy tych metod znaleźć można w cytowanej literaturze[23], dlatego omówione zostaną tylko skrótowo, a zasadnicza uwaga skupiona będzie na sprawach aplikacyjnych i interpretacyjnych.

2. Charakterystyka modeli

2.1. Klasyczny model regresji probitowej

Analiza probitowa służy do określania związku przyczynowego pomiędzy nasileniem bodźca (przyczyny) a reakcją na ten bodziec (skutkiem) wyrażoną w proporcjach lub częstościach przypadków. Liczby te za pomocą transformacji probitowej zamieniane są na probity (jednostki prawdopodobieństwa), które

[20] J. Tobin, *Zastosowanie analizy probitu do wyników badań statystycznych dotyczących zjawisk ekonomicznych*, „Przegląd Statystyczny", t. 7, z. 1, PWN, Warszawa 1960, s. 3–19.

[21] Bibliografię większości prac znaleźć można w monografii J. W. Wiśniewskiego, op. cit. Poza tym wymienić można prace: B. Górecki, *Analiza probitowa w badaniach zależności englowskiech*, „Przegląd Statystyczny", t. 20, z. 2, 1973, s. 153–165; także: B. K. Bednarek, *Wyposażenie w dobra trwałego użytku gospodarstw domowych w cyklu życia rodziny*, Raport merytoryczny z badania w ramach zadania badawczego 09.1.4.08.c, CPBP, ICE AE, Poznań 1986 (archiwum CBR AE, maszynopis); M. Chromińska, *Czynnik ekonomiczny a dzietność rodzin rolników indywidualnych w Polsce. Studium statystyczno-demograficzne dla lat 1926–1984*, AE, Poznań 1996, Zeszyty Naukowe, Seria II, z. 148 (zastosowanie regresji probitowej do badania motywów urodzenia kolejnego dziecka).

[22] Jej opis i przykład zastosowania znaleźć można w cytowanej monografii W. J. Wiśniewskiego.

[23] W szczególności pozycja j.w.

w przeciwieństwie do proporcji są addytywne i umożliwiają rachunek korelacyjny. Jeżeli zmienna X reprezentująca bodziec wyraża działanie dużej liczby niezależnych czynników, to zgodnie z centralnym twierdzeniem granicznym jej rozkład powinien być zbliżony do normalnego. Probity powiązane są z prawdopodobieństwem P wyrażającym proporcje poprzez dystrybuantę standaryzowanego rozkładu normalnego:

$$P = F(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{1}{2}u^2} du, \tag{3}$$

gdzie

$$Y = \frac{X - \mu}{\sigma} \tag{4}$$

jest znormalizowaną zmienną losową zwaną probitem[24], X – zmienną losową o rozkładzie normalnym, μ – wartością oczekiwaną, σ – odchyleniem standardowym. Można również powiedzieć, że probit jest funkcją odwrotną $F^{-1}(P)$ do dystrybuanty standaryzowanego rozkładu normalnego. Zatem transformacja (3) przekształca przedział prawdopodobieństwa $\langle 0, 1 \rangle$ na zbiór $(-\infty, +\infty)$, zwiększając przedział zmienności zmiennej zależnej wyrażonej przez proporcje.

Podstawiając w (4) $\alpha = -\frac{\mu}{\sigma}$ i $\beta = \frac{1}{\sigma}$, otrzymamy liniowe równanie regresji probitowej

$$Y = \alpha + \beta X. \tag{5}$$

Liniowość równania (5) wynika ze zgodności rozkładów prawdopodobieństwa wyrażonego przez probit i zmiennej X. Jeżeli rozkład zmiennej X znacznie różni się od normalnego, wtedy związek (5) nie zachowuje liniowości. Często można ją poprawić zastępując zmienną X jej logarytmem[25].

Przed przystąpieniem do oszacowania parametrów modelu (5) należy dokonać podziału zbioru obserwacji na grupy według przyjętego kryterium i obliczyć dla nich częstości wystąpienia zdarzenia, a następnie, na podstawie równości (3), zamienić je na probity. Szacowanie można przeprowadzić metodą największej wiarygodności (MNW) lub klasyczną MNK, a jego dobroć ocenić za pomocą testu χ^2.

[24] Inaczej jest to kwantyl rzędu P standaryzowanego rozkładu normalnego. Początkowo powiększano go o 5 w celu ułatwienia obliczeń. Obecnie jest to zabieg zbędny.

[25] Wiele zjawisk ma często rozkład logarytmiczno-normalny, np. taki jest najczęściej rozkład dochodów. Por. m. in. B. Górecki, *Analiza...*, op. cit., s. 154–157. Również M. Chromińska stosuje w modelu probitowym logarytm powierzchni gruntu rolnego reprezentującej dochód, op. cit., s. 182–203.

Oszacowany parametr β określa wielkość i kierunek reakcji wyrażonej w probitach na jednostkową zmianę bodźca reprezentowanego przez zmienną X.

Oszacowany model pozwala obliczyć prawdopodobieństwo realizacji zespołu warunków W przy określonej wartości zmiennej X, jak również wyznaczyć wartość X (wielkość bodźca) dla ustalanego prawdopodobieństwa realizacji warunków W. Wartości te można odczytać z tablicy dystrybuanty standaryzowanego rozkładu normalnego na podstawie równości (3) oraz zależności $F(-Y) = 1 - F(Y)$ wynikającej z symetrii rozkładu.

2.2. Model wielowymiarowej regresji probitowej

Przebieg danego zjawiska kształtowany jest przez wiele czynników, toteż model wyjaśniający z jedną tylko zmienną objaśniającą na ogół nie jest zadowalający. Zwykle bowiem chcemy poznać, jakie to są czynniki i jaki jest udział każdego z nich w kształtowaniu zjawiska. Z drugiej strony przy niemonotonicznym wpływie cech na zmienną objaśnianą prosty związek między dwiema zmiennymi może zmienić charakter po dołączeniu nowej zmiennej, która może oddziaływać zarówno na zmienną objaśnianą, jak i na zmienną objaśniającą. Odpowiedniejszy do tego celu jest wielowymiarowy model analizy probitu zaproponowany przez J. Tobina do badań ekonomicznych. Jest on na tyle ogólny, że może być też stosowany w każdym przypadku, w którym realizacja warunków W (objawiająca się przyjęciem przez zmienną objaśnianą wartości 1) zależy od osiągnięcia przez bodziec dostatecznie dużego natężenia.

Charakterystyka modelu przedstawiona zostanie na tle określonej sytuacji ekonomicznej nawiązującej do przykładu rozpatrywanego w dalszej części pracy. Niech to będzie zakup lub posiadanie wybranego dobra trwałego przez gospodarstwo domowe z pewnej populacji. W modelu zakłada się, że na posiadanie dobra trwałego wpływ wywierają czynniki reprezentowane przez zespół zmiennych $X_1, X_2, ..., X_k$[26]. Przyjmuje się, że bodźcem wywołującym efekt posiadania danego dobra jest indeks I będący liniową kombinacją tych zmiennych. Przy dostatecznie dużej wartości I^*-tego indeksu, zwanej wartością progową, gospodarstwo domowe staje się posiadaczem dobra trwałego. Badaną grupę n gospodarstw domowych dzieli się na g bardziej jednorodnych grup według wartości zmiennych X_j zgodnie z przyjętym kryterium klasyfikacyjnym. Dla każdej grupy tworzy się indeks

[26] Zmienne X_j nie muszą być bezpośrednio obserwowanymi zmiennymi, lecz mogą być także funkcjami jednej lub kilku zmiennych oryginalnych, np. logarytmami, potęgami, iloczynami. Zob. J. Tobin, op. cit., s. 6.

$$I_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}, \quad i = 1, 2, \ldots, g, \tag{6}$$

gdzie β_j są parametrami strukturalnymi modelu a x_{ij} wartościami zmiennych X_j ($j = 1, \ldots, k$) w i-tej grupie gospodarstw domowych.

Niech Y_i będzie zmienną objaśnianą opisującą posiadanie przedmiotów trwałego użytkowania przez gospodarstwa w i-tej grupie gospodarstw domowych i niech I_i^* będzie wielkością progową indeksu w tej grupie. Wtedy

$$Y_i = \begin{cases} 1 & \text{dla } I_i \geq I_i^* \\ 0 & \text{dla } I_i < I_i^* \end{cases}, \quad i = 1, 2, \ldots, g.$$

Zakłada się też, podobnie jak w modelu klasycznym, że w zbiorowości gospodarstw domowych podzielonych na grupy rozkład wartości progowych I_i^* jest standaryzowanym rozkładem normalnym. Wartości te spełniają rolę składników losowych[27] wyrażających indywidualne preferencje gospodarstw domowych, które nie znalazły odbicia w zmiennych modelu (6). Zatem prawdopodobieństwo zdarzenia, że gospodarstwo domowe w i-tej grupie posiada dane dobro jest równe

$$P\{Y_i = 1 | I_i\} = P\{I_i \geq I_i^*\} = P(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{1}{2}u^2} du, \tag{7}$$

a prawdopodobieństwo zdarzenia przeciwnego wynosi

$$Q\{Y_i = 0 | I_i\} = P\{I_i < I_i^*\} = Q(I_i) = 1 - P(I_i). \tag{8}$$

Prawdopodobieństwo $P(I_i)$ jest wartością dystrybuanty rozkładu $N(0,1)$ określoną na przedziale $(-\infty, I_i\rangle$, należy więc do przedziału $\langle 0,1\rangle$.

Aby znaleźć oceny $(b_0, b_1, \ldots, b_k)$ parametrów równania (6), należy dla każdej grupy gospodarstw obliczyć indeks I. Jeżeli w grupie liczącej n_i ($n_1 + n_2 + \ldots + n_g = n$) gospodarstw domowych r_i z nich posiada dobro trwałe, wtedy częstość $c_i = r_i/n_i$ jest oceną prawdopodobieństwa $P\{Y_i = 1 | I_i\}$. Można więc na podstawie zależności (7) obliczyć empiryczny probit jako wartość standaryzowanej zmiennej normalnej, dla której dystrybuanta jest równa c_i. Wartość obliczonego probitu jest szukaną wielkością indeksu I_i. Ponieważ w ramach każdej grupy gospodarstw domowych prawdopodobieństwo posiadania dobra trwałego (określone przez częstość) jest identyczne dla każdego gospodarstwa, więc ma rozkład dwumianowy. Różne

[27] Por. A. S. Goldberger, op. cit., s. 322.

prawdopodobieństwa występują tylko dla gospodarstw z różnych grup i z założenia są funkcją dystrybuanty indeksu (6).

Do oszacowania ocen (b_0, b_1,..., b_k) parametrów równania (6) stosuje się metodę największej wiarogodności (MNW). W tym celu poszukuje się maksimum funkcji wiarogodności $L(b_0, b_1,..., b_k)$ rozkładu dwumianowego lub częściej maksimum jej wygodniejszej obliczeniowo postaci $L^* = \ln L$. Po zróżniczkowaniu funkcji L^* względem parametrów b_j ($j = 0, 1,..., k$) otrzymuje się $k + 1$ równań największej wiarogodności $\partial L^*/\partial b_j = 0$ w postaci uwikłanej, których nie można bezpośrednio rozwiązać względem szukanych ocen b_j. Można natomiast znaleźć ich przybliżone wartości, stosując metody iteracyjne. Zaproponowana przez J. Tobina metoda pozwala obliczyć estymatory poszukiwanych parametrów z dowolną dokładnością, a także określić ich średnie błędy szacunku i ocenić dobroć modelu[28]. Oszacowane parametry określają siłę i kierunek oddziaływania poszczególnych zmiennych bezpośrednio na indeks I (probit), a pośrednio na zmienną objaśnianą. Duża bezwzględna wartość parametru świadczy o dużej sile tego oddziaływania, znak dodatni określa stymulujący charakter danej zmiennej, znak ujemny właściwość przeciwną. Parametr b_j określa zmianę indeksu spowodowaną jednostkową zmianą zmiennej X_j przy ustalonych wartościach pozostałych zmiennych modelu. Aby tę zmianę wyrazić w prawdopodobieństwie, należy, opierając się na transformacji probitowej, skorzystać z tablicy dystrybuanty rozkładu $N(0,1)$. W ten sam sposób można określić prawdopodobieństwo posiadania dobra trwałego przez gospodarstwo domowe przy określonej konfiguracji wartości zmiennych objaśniających.

Jeżeli zmienne objaśniające mają charakter opisowy, o wartościach wyrażonych w skali nominalnej, lub taki charakter został im nadany, to ich wartości określają pewne stany jakościowe. W takim przypadku bardziej adekwatne jest raczej wyspecyfikowanie związku pomiędzy różnymi stanami danej zmiennej a badanym zjawiskiem niż mierzenie jej ogólnego wpływu na przebieg tego zjawiska. Zmienna jakościowa przyjmuje pewną skończoną liczbę stanów, które mogą być zakodowane za pomocą zmiennych zero-jedynkowych. Zatem każda zmienna objaśniająca w modelu może być określona blokiem zmiennych zero-jedynkowych, których liczba jest o 1 mniejsza od liczby jej stanów[29]. Pominięte stany przenoszą swój efekt na wyraz wolny i stanowią kategorię odniesienia (bazową). Przyjęcie innego zestawu pomijanych stanów nie powoduje żadnych zmian w wartościach teoretycznych indeksu; szacowane są tylko inne parametry[30].

Jeżeli s_j ($j = 1, 2, ..., k$) oznacza liczbę stanów przyjmowanych przez zmienną X_j, to liczba zmiennych zero-jedynkowych Z jest równa

---

[28] Zob. J. Tobin, op. cit.

[29] Zabieg ten ma na celu uniknięcie dokładnej zależności liniowej zmiennych objaśniających.

[30] Por. A. Goldberger, op.cit., s. 284–286.

$$s = s_1 + s_2 + ... + s_k - k.$$

Zatem równanie indeksu stymulującego posiadanie dobra trwałego jest teraz następujące:

$$I_i = \beta_0 + \beta_1 Z_1 + ... + \beta_s Z_s, \quad i = 1, 2, ... g. \tag{9}$$

W wyrazie β_0 kumulują się efekty cząstkowe wywierane na indeks przez wszystkie pominięte stany zmiennych oryginalnych X_j. Jeżeli zmienna X_l przyjmuje $m + 1$ stanów, to oszacowane współczynniki b_1, b_2, ..., b_m przy zmiennych Z_1, ..., Z_m przedstawiają efekty cząstkowe kolejnych stanów zmiennej X_1 względem stanu bazowego (pominiętego), którego efekt ujęty jest w wyrazie wolnym. Ściślej mówiąc, współczynnik b_1 określa zmianę indeksu I spowodowaną przejściem od stanu bazowego do stanu pierwszego z nie pominiętych, przy ustalonych wartościach pozostałych zmiennych. Zmianę prawdopodobieństwa wywołaną efektem cząstkowym wyznacza się z tablicy dystrybuanty rozkładu $N(0,1)$.

2.3. Model regresji logitowej

W modelu logitowym korzysta się z transformacji logitowej, która proporcje (częstości) zamienia na pewne jednostki addytywne zwane logitami. Logity Y mogą być powiązane z prawdopodobieństwem P realizacji zespołu warunków W (zajścia zdarzenia) zależnością[31]

$$P = F(Y) = \frac{e^Y}{1 + e^Y}, \tag{10}$$

która w prosty sposób może być przekształcona do postaci

$$Y = \ln \frac{P}{1 - P}. \tag{11}$$

Logit jest zatem logarytmem naturalnym ilorazu

$$\psi = \frac{P}{1 - P} \tag{12}$$

[31] Podejście takie zapoczątkował J. Berkson. Zob. J. Berkson, *Maximum Likelihood and Minimum X^2 Estimates of the Logistic Function*, „Journal of the American Statistical Association", t. 50, 1955; S. S. Halli, K.V. Rao, *Advanced Techniques of Population Analysis*, Plenum Press, N. York – London 1992, s. 103.

prawdopodobieństw dwóch zdarzeń przeciwnych. Wyrażenie (12) można nazwać ryzykiem względnym lub szansą[32] realizacji warunków *W* (zajścia zdarzenia). Ryzyko względne (szansę) można zdefiniować również inaczej. Jeżeli w próbie *n*-elementowej jest *m* przypadków zajścia pewnego zdarzenia, to jego prawdopodobieństwo empiryczne określa stosunek *m*/*n*. Zatem

$$\psi = \frac{\frac{m}{n}}{1-\frac{m}{n}} = \frac{\frac{m}{n}}{\frac{n-m}{n}} = \frac{m}{n-m}.$$

Stąd też szansa zdarzenia może być rozumiana jako stosunek liczby przypadków zajścia danego zdarzenia do liczby przypadków, w których to zdarzenie nie zaszło.

Proste przekształcenie prawej strony równania (10) sprowadza je do postaci

$$P = F(Y) = \frac{1}{1+e^{-Y}}, \tag{13}$$

co wskazuje na powiązanie logitu z funkcją logistyczną (stąd nazwa logit). Jak wynika z formuły (11), transformacja logitowa przekształca najpierw częstości reprezentowane przez *P* na wartości z przedziału $\langle 0, +\infty)$, a następnie przez logarytmowanie zamienia wartości ψ na liczby z przedziału $(-\infty, +\infty)$. W ten sposób przedział zmienności $\langle 0, 1\rangle$ zmiennej zero-jedynkowej zostaje przekształcony na zbiór liczb rzeczywistych *R*, przy czym logit jest funkcją ciągłą w tym zbiorze. Ze względu na to, że modele logitowe, oparte na transformacji logitowej, wykorzystują funkcję logistyczną, toteż często nazywane są modelami regresji logistycznej[33]. Ponieważ równanie regresji dotyczy logitu, w pracy nazywane jest modelem logitowym.

---

[32] W języku angielskim stosunek (12) określany jest wyrazem „odds", którego sens jest powszechnie zrozumiały; w języku polskim nie ma dobrego odpowiednika. Stosunek ten nazywany jest ryzykiem względnym. Wydaje się, że jego sens dość dobrze oddaje wyraz „szansa". Jeżeli bowiem prawdopodobieństwo zajścia pewnego zdarzenia wynosi 1/2, to mówi się, że jego szansa jest jak 1 do 1. W tym sensie szansa wyciągnięcia kiera z talii kart wynosi 1/3 natomiast prawdopodobieństwo jest równe 1/4.

[33] W literaturze przedmiotu spotkać można rozróżnienie między modelem logistycznym i logitowym. W modelu logistycznym z założenia zmienne objaśniające są ciągłe. Jeżeli do modelu mają być włączone zmienne nominalne, to muszą być sztucznie zakodowane. Model logitowy nie jest funkcją zmiennych ciągłych. Także parametry modeli są nieco inaczej interpretowane. Jeśli jednak wszystkie zmienne objaśniające są nominalne, to obydwa modele dają identyczne wyniki. Por. S. S. Halli, K. V. Rao, op. cit., s. 102. Również w pakietach statystycznych dla komputerów spotyka się odrębne procedury dla tych modeli.

Jeżeli przyjmie się założenie, że probit Y jest kombinacją liniową k zmiennych X_j, to można postąpić podobnie jak przy konstrukcji modelu probitowego. W takim przypadku równanie (6) przyjmie postać

$$Y_i = \ln \frac{P_i}{1-P_i} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{ik}, \quad i = 1, 2, \ldots, g, \tag{14}$$

gdzie P_i oznacza prawdopodobieństwo posiadania określonego dobra trwałego przez gospodarstwo domowe w grupie i. W przypadku wystąpienia w modelu (14) tylko jednej zmiennej objaśniającej X stanie się on odpowiednikiem klasycznego modelu probitowego (5). W modelu tym zakłada się tylko, że dystrybuanta wartości teoretycznych jest funkcją logistyczną a wariancja jest stała.

Jeżeli zmienne objaśniające X_j nie są kategoryczne (czyli mierzone są w skali co najmniej przedziałowej), wtedy w modelu reprezentowane są przez swoje średnie $\bar{x}_{ij}$ w grupie. Również prawdopodobieństwa P_i zastępowane są przez częstości względne c_i. Zatem model przyjęty do szacowania ma teraz postać

$$Y_i = \ln \frac{c_j}{1-c_i} = \beta_0 + \beta_1 \bar{x}_{i1} + \ldots + \beta_k \bar{x}_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots g, \tag{15}$$

gdzie ε_i uwzględnia zakłócenia losowe spowodowane zastępowaniem prawdopodobieństw w modelu (14) przez częstości c_i zawierające element losowości wynikający z zastąpienia zmiennych przez ich wartości średnie będące wypadkową różnych wartości w grupie. Parametry modelu (15) szacuje się tak jak poprzednio za pomocą MNW. Miarą wiarogodności modelu może być statystyka $L^2 = -2 \ln (L_0/L_1)$[34], która posiada w przybliżeniu rozkład χ^2. Istotność parametrów modelu można określić za pomocą statystyki A.Walda.

Model (14) nie specyfikuje bezpośrednio prawdopodobieństwa P_i w zależności od określonej konfiguracji wartości zmiennych objaśniających. Prawdopodobieństwo to można wyznaczyć jako wartość funkcji F z formuł (10) lub (13). Uzyskane oceny b_j parametrów β_j ($i = 1, 2, \ldots, k$) można interpretować podobnie jak w modelu (6): określają siłę i kierunek wpływu poszczególnych zmiennych objaśniających na zmianę logitu, a pośrednio na prawdopodobieństwo zdarzenia. Współczynnik b_j określa zmianę logitu spowodowaną zmianą zmiennej X_j o jednostkę, przy ustalonych wartościach pozostałych zmiennych.

Na podstawie oszacowanego modelu można też określić ryzyko względne realizacji warunków W (np. posiadania dobra trwałego przez gospodarstwo

[34] L_0 i L_1 oznaczają funkcje wiarogodności odpowiednio dla modelu jedynie ze stałą i ze zmiennymi.

domowe). W tym celu należy przekształcić formułę (11) i przedstawić funkcję ψ w postaci

$$\psi = e^{b_0+b_1x_1+b_2x_2+\ldots+b_kx_k} = e^{b_0}e^{b_1x_1}e^{b_2x_2}\ldots e^{b_kx_k}, \qquad (16)$$

gdzie e^{b_j} wyraża zmianę ryzyka względnego spowodowaną działaniem czynnika reprezentowanego przez zmienną X_j. Wartość tę należy teraz porównywać z liczbą 1 (nie jak poprzednio z zerem), a uzyskaną różnicę można wyrazić w procentach. Jeżeli $e^{b_j} > 1$, to uznaje się, że zmienna X_j wpływa stymulująco na prawdopodobieństwo realizacji warunków W; gdy $e^{b_j} < 1$, to uważa się, że zmienna X_j działa ograniczająco na to prawdopodobieństwo; gdy $e^{b_j} = 1$, to uznaje się brak wpływu zmiennej X_j na możliwość realizacji warunków W.

Jeżeli zmienne objaśniające X_j (j = 1, 2, ..., k) uwzględnione w modelu są zmiennymi jakościowymi, mierzonymi w skali nominalnej, należy je zastąpić blokiem zmiennych zero-jedynkowych, postępując identycznie jak przy konstrukcji modelu probitowego[35]. Wtedy równanie (15) przyjmie postać

$$Y_i = \ln\frac{c_i}{1-c_i} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_s Z_s, \quad i = 1, 2, \ldots g. \qquad (17)$$

Również w tym przypadku efekty pominiętych stanów poszczególnych zmiennych kumulują się w wyrazie wolnym, a efekty cząstkowe kolejnych stanów każdej zmiennej odnoszą się do kategorii bazowej (stanu pominiętego). Zatem współczynnik b_1 wyraża przyrost logitu spowodowany przejściem od stanu bazowego do stanu pierwszego (nie pominiętego) zmiennej X_1, przy ustalonych stanach pozostałych zmiennych. Pominięcie innego stanu spowoduje szacowanie innych parametrów, lecz nie zmieni teoretycznych wartości logitu.

3. Charakterystyka danych źródłowych i zmiennych

Omówione metody zostaną zastosowane do określenia ewentualnych związków pomiędzy stanem posiadania wyróżnionych dóbr trwałych przez gospodarstwa domowe a zespołem wybranych czynników. Jako dane źródłowe posłużyły wyniki badań budżetów gospodarstw domowych[36] prowadzonych systematycznie od roku

[35] Zastosowana do obliczeń procedura regresji logistycznej w pakiecie SPSS nie wymagała wstępnej agregacji danych, a kodowanie nastąpiło automatycznie.

[36] Dane te udostępnione przez GUS znajdują się w Banku Danych o Rodzinie i Gospodarstwie Domowym CBR AE w Poznaniu.

1957 przez Główny Urząd Statystyczny w Warszawie. Badania te, mimo swej ograniczoności i pewnych mankamentów[37], dostarczają współcześnie najbogatszych informacji o postawach konsumpcyjnych rodzin i gospodarstw domowych. W przykładzie zainteresowanie zostało ograniczone tylko do gospodarstw domowych typu pracowniczego uczestniczących w badaniu w latach 1987 i 1997. Wybór tej grupy gospodarstw domowych wynika stąd, że jest ona najliczniejsza i najbogatsza pod względem struktury w porównaniu z pozostałymi społeczno-ekonomicznymi grupami gospodarstw domowych uczestniczącymi w badaniach[38]. Dziesięcioletni dystans między obserwacjami stwarza możliwość zaobserwowania zmian w relacjach między poziomem wyposażenia gospodarstw domowych w przedmioty trwałego użytkowania a wybranymi czynnikami[39]. Do analizy danych z roku 1987 zastosowana została tylko metoda probitu, natomiast do analizy danych późniejszych – obie metody. Umożliwia to z jednej strony odniesienie się do wyników badań z różnych okresów, z drugiej strony porównanie wyników uzyskanych różnymi metodami.

Proces gromadzenia dóbr trwałych przez gospodarstwa domowe jest w swojej naturze dość złożony. Złożoność ta wynika z konkurencyjności różnorodnych potrzeb, a ta z kolei spowodowana jest ograniczonością środków będących w dyspozycji gospodarstw domowych. Z konieczności zatem niemal każde gospodarstwo domowe, postępując racjonalnie, musi ustalić hierarchię swoich potrzeb i realizować je według przyjętych priorytetów. Priorytety te są pochodną pilności odczuwanych potrzeb, które zmieniają się w czasie pod wpływem działania szerokiego spektrum czynników wewnętrznych i zewnętrznych. Ich wpływowi podlega

---

[37] Pewne uwagi krytyczne znaleźć można m.in. w pracach: S. Wierzchosławski, *Ocena źródeł statystyki ludnościowej tworzonej przez organy statystyki państwowej, z punktu widzenia demografii formalnej*, Ogólnopolskie Seminarium Demograficzne, Poznań-Puszczykowo 6–9.12.1989 r., „Studia Demograficzne" 1990, nr 3/301, s. 89–94; I. Roeske-Słomka, *Ocena budżetów gospodarstw domowych z punktu widzenia przydatności dla analizy przeobrażeń modelu gospodarstwa domowego*, Ogólnopolskie Seminarium Demograficzne, Poznań-Puszczykowo, 6–9.12.1989; B. K. Bednarek, *Zróżnicowanie wyposażenia gospodarstw domowych w dobra trwałego użytku w okresie transformacji społeczno-ekonomicznej w Polsce*, Centrum Badań nad Rodziną, Badania nad Rodziną, Materiały i Przyczynki nr 2, AE, Poznań 2000, s. 5–8.

[38] Do roku 1992 uczestniczyły w badaniach cztery społeczno-ekonomiczne grupy gospodarstw domowych, obecnie uczestniczy sześć grup. Por. *Budżety gospodarstw domowych w 1993 r.*, GUS, Warszawa 1994. Grupa pracownicza nie jest jednorodna, gdyż widoczne są wyraźne różnice zachowań konsumpcyjnych pomiędzy gospodarstwami kierowanymi przez osoby na stanowiskach robotniczych i nierobotniczych. Tu traktowana jest całościowo. Grupa ta w 1987 roku liczyła 10 658 gospodarstw domowych a w 1997–13188. Ostatecznie do obliczeń zostało zakwalifikowanych odpowiednio 9856 i 10 162 gospodarstw.

[39] Można tego oczekiwać ze względu na zmiany w poziomie i strukturze wyposażenia gospodarstw domowych w dobra trwałego użytku jakie zaszły szczególnie w pierwszej połowie lat dziewięćdziesiątych. Por. B. K. Bednarek, *Zróżnicowanie wyposażenia...*, op. cit.

również proces tworzenia infrastruktury technicznej gospodarstw domowych w postaci zasobów dóbr trwałego użytku.

Jednym z takich czynników wewnętrznych jest niewątpliwie proces rozwojowy, jakiemu w mniejszym lub większym stopniu podlega każde gospodarstwo domowe. Może on być przedstawiony w postaci swoistego cyklu, który od momentu początkowego przechodzi przez fazę progresji, stabilizacji i degresji do momentu końcowego. Proces ten jest mało dynamiczny w przypadku gospodarstwa jednoosobowego w całym cyklu. Największą dynamiką odznacza się cykl rozwojowy rodzinnego gospodarstwa domowego. Ponieważ w Polsce wciąż blisko 80% gospodarstw domowych stanowią gospodarstwa rodzinne[40], więc z dużym przybliżeniem proces rozwojowy gospodarstwa domowego może być określony przez cykl życia rodziny.

Pojęcie to obejmuje pewną wyidealizowaną sztuczną, konstrukcję, która ma uwzględnić najważniejsze wydarzenia zachodzące w życiu przeciętnej rodziny. Każda rodzina w okresie swego istnienia podlega przemianom, które, stwarzając nowe jakościowo sytuacje, przesuwają ją z jednego stanu do innego. Stany te, zwane etapami lub fazami, składają się na pewien proces określany mianem cyklu życia rodziny[41]. Nie każda rodzina przechodzi przez wszystkie fazy cyklu, jednak dla zdecydowanej większości rodzin jego przebieg jest zbliżony.

W literaturze przedmiotu istnieje wiele różnych koncepcji cyklu życia rodziny implikowanych celami, którym mają służyć. Można wszakże zaobserwować dążenie do zbudowania w miarę uniwersalnej konstrukcji, opartej na koncepcji propagowanej przez Światową Organizację Zdrowia (WHO)[42]. Chcąc dokonać empirycznej weryfikacji przyjętej koncepcji, należy mieć na uwadze nie tylko cele, lecz także możliwości, które są często ograniczone przez zbiór dostępnych informacji źródłowych. Kierując się tymi przesłankami, jak również doświadczeniem z badań własnych i innych[43], za podstawę ustalenia poszczególnych etapów cyklu, przy wykorzystaniu danych z roku 1987, przyjęto trzy cechy[44]:

---

[40] Gospodarstw rodzinnych było w Polsce w roku 1988 – 80,12%, a w roku 1995 – 78,68%. Obliczenia własne na podstawie danych z „Rocznika Demograficznego" 2000, GUS, s. 91.

[41] Szerzej na ten temat w: B. K. Bednarek, *Fazy rozwoju rodziny a wyposażenie w dobra trwałego użytku gospodarstw domowych.* Raport merytoryczny z badań w ramach IV Grupy Tematycznej CPBP 09.1k, „Uwarunkowania demograficznego rozwoju społeczno-gospodarczego Polski", Zad. bad. 09.1.4.08c, Poznań 1990 (maszynopis).

[42] Zob. H. Hansluwka, *Umieralność a cykl życia rodziny,* „Studia Demograficzne" 44, Organ Komitetu Nauk Demograficznych PAN, Warszawa 1976, s. 5.

[43] Por. B. K. Bednarek, *Wyposażenie w dobra trwałego użytku gospodarstw domowych w cyklu życia rodziny.* Raport merytoryczny badania w ramach zadania badawczego 091.408c, CPBP, Poznań 1986 (maszynopis); także M. Latuch, *Demografia społeczno-ekonomiczna*, PWN, Warszawa 1985, s. 174–177.

[44] Omawiany tu przykład pochodzi z wcześniejszej pracy autora: *Fazy...*, op. cit.

– wiek jednostki (dla osób samotnych),
– okres trwania związku małżeńskiego (dla małżeństw bezdzietnych),
– wiek najstarszego dziecka (dla małżeństw dzietnych).

Przy konstrukcji cyklu został uwzględniony tylko wiek najstarszego dziecka, bowiem, zdaniem autora, o przejściu rodziny (lub gospodarstwa domowego) w jakościowo nowy stan decydują głównie te zdarzenia, które zachodzą po raz pierwszy (narodziny pierwszego dziecka, pójście do szkoły, odejście z domu). Zdarzenia powtarzające się, szczególnie te związane z młodszymi dziećmi, już tylko powielają sytuację zaistniałą wcześniej i nie stanowią tak silnego bodźca do zmiany zachowań jak zdarzenia całkiem nowe. Przedstawiona konstrukcja cyklu[45] zawiera aspekt ekonomiczny, bowiem dzietność rodzin odnosi się tylko do dzieci na utrzymaniu[46].

Oparcie się na ważnych wydarzeniach w rodzinie pozwoliło na wyodrębnienie dziewięciu etapów cyklu rozwojowego gospodarstwa domowego:

1. stan przedmałżeński (osoby samotne do 39 lat),
2. młode małżeństwa bez dzieci (okres trwania poniżej 5 lat),
3. małżeństwa z dziećmi w wieku przedszkolnym (najstarsze dziecko poniżej 7 lat),
4. małżeństwa z dziećmi w wieku szkolnym (najstarsze dziecko poniżej 16 lat),
5. małżeństwa z dziećmi dorastającymi (najstarsze dziecko poniżej 25 lat),
6. rodziny niepełne (matka lub ojciec z dziećmi),
7. średnie małżeństwa bez dzieci (okres trwania 5–15 lat),
8. starsze małżeństwa bez dzieci (okres trwania ponad 15 lat),
9. starsze osoby samotne (od 40 lat).

Zmiany, jakie nastąpiły w metodologii badań budżetów domowych w roku 1993, nie pozwoliły na utrzymanie zasad powyższej konstrukcji cyklu rozwojowego i wymusiły zmianę kryterium wyodrębnienia niektórych jego etapów. Ponieważ nie odnotowuje się już lat trwania związku małżeńskiego kryterium stażu małżeńskiego zastąpione zostało z konieczności jego namiastką w postaci wieku żony[47]. Zmiana zasady związana jest z małżeństwami bezdzietnymi i dotyczy tylko trzech faz wyszczególnionych wyżej w punktach:

2. młode małżeństwa bez dzieci (wiek żony poniżej 25 lat),

[45] Konstrukcja ta jest zbliżona do wprowadzonych wcześniej m.in. przez A. Tymowskiego, D. Kalinina i B. Góreckiego. Więcej na ten temat w: B. K. Bednarek, *Fazy...*, op. cit., s. 21 i dalsze.

[46] Jest to tzw. dzietność ekonomiczna. W badaniu brane są pod uwagę tylko te rodziny dzietne, które mają wyłącznie dzieci na utrzymaniu (żadne nie pracuje).

[47] O przyjęciu tego kryterium zadecydował fakt, że większość kobiet wstępuje w związek małżeński w wieku 20–24 lat. Np. kobiety w tym wieku stanowiły 53,9% wszystkich kobiet, które w 1999 r. zawarły związek małżeński. Obliczenia własne na podstawie: „Rocznik Demograficzny" 2000, GUS, Warszawa , s. 159.

7. średnie małżeństwa bez dzieci (wiek żony 25–39 lat),
8. starsze małżeństwa bez dzieci (wiek żony od 40 lat).

Taka konstrukcja poszczególnych etapów miała w swym zamierzeniu możliwie pełne odtworzenie sytuacji, w jakiej może się znaleźć większość gospodarstw domowych. Stąd też prócz rodzin pełnych uwzględnione zostały rodziny niepełne i osoby samotne. Należy przy tym mieć na uwadze, że przedstawiona koncepcja cyklu rozwojowego gospodarstwa domowego jest wynikiem kompromisu między ideą przyświecającą konstrukcji rzeczywistego cyklu życia rodziny lub jednostki a możliwościami stwarzanymi przez dostępne dane źródłowe. Niestety, oficjalne badania budżetów gospodarstw domowych nie zajmują się historią rodzin. Dlatego konstruowana na ich podstawie zmienna odwzorowująca przebieg rozwojowy gospodarstwa domowego, uwzględniająca cykl życia rodziny, nie jest zmienną w pełni ekwiwalentną. Niemniej, jak się wydaje, może być uważana za dobrego reprezentanta czynnika wywołującego zmianę postaw gospodarstw domowych wynikających z zachodzących w nich przeobrażeń.

Zmienną cyklu rozwojowego (X_1) uzupełniają jeszcze dwie zmienne, które, zdaniem nie tylko autora, silnie kształtują postawy i zachowania konsumpcyjne ludności. Są to: klasa miejscowości reprezentująca czynnik środowiskowy związany ze stopniem urbanizacji (X_2) i wykształcenie głowy gospodarstwa domowego[48], a więc zmienna wyrażająca czynnik kulturowy (X_3). Miejscowości zostały podzielone na cztery klasy opisane symbolami cyfrowymi:

1. miasta o liczbie mieszkańców ponad 100 tysięcy,
2. miasta o liczbie mieszkańców 20–100 tysięcy,
3. miasta o liczbie mieszkańców poniżej 20 tysięcy,
4. wieś.

Wykształcenie reprezentowane jest przez pięć poziomów wyrażonych kodami:

1. wyższe,
2. średnie (pełne każdego typu, pomaturalne, niepełne wyższe),
3. zasadnicze (zasadnicze zawodowe, niepełne średnie),
4. podstawowe,
5. inne (podstawowe nie ukończone, samouctwo)[49].

Zmienna endogeniczna (zależna) ma charakter dychotomiczny i wyraża posiadanie co najmniej jednego dobra danego rodzaju przez gospodarstwo domowe bądź jego brak.

Wszystkie omówione zmienne mają charakter jakościowy. Jakkolwiek zmienna cyklu rozwojowego ma w swej istocie charakter dynamiczny, to jednak

[48] W badaniach budżetów gospodarstw domowych jest to osoba, która dostarcza najwięcej środków na utrzymanie albo nie mniej od innych osób i tymi środkami dysponuje.

[49] W badaniu dla roku 1997 poziom 5 nie był brany pod uwagę ze względu na znikomą liczbę przypadków.

ze względu na konstrukcję nie powinna być traktowana nawet jako zmienna porządkowa. Ma ona wyrażać przede wszystkim odmienność zachowań wynikających z faktu znajdowania się gospodarstw domowych w różnych fazach. W ten sam sposób należy odnieść się do dwóch pozostałych zmiennych.

Na podstawie powyższych trzech cech badaną zbiorowość pracowniczych gospodarstw domowych można podzielić maksymalnie na $9 \cdot 4 \cdot 5 = 180$ w miarę jednorodnych grup (kategorii). Ponieważ dla uniknięcia przypadkowości odrzucane były kategorie liczące mniej niż 4 elementy, ostatecznie w badaniu uczestniczyło 137 grup z roku 1987 i 131 z roku 1997. Dla każdej z nich wyznaczana jest frakcja gospodarstw posiadających dane dobro trwałe, która jest oceną empirycznego prawdopodobieństwa posiadania tego dobra i stanowi punkt wyjścia do dalszych obliczeń. Aby ocena ta była zbliżona do wartości rzeczywistej, w grupie nie może być zbyt mało obserwacji.

Wymagania programu obliczeniowego wielowymiarowej regresji probitowej[50] narzucały stosowanie modelu (9), z czym związane było przekształcenie oryginalnych zmiennych na zmienne zero-jedynkowe. Zatem zmienna pierwotna w modelu reprezentowana jest przez blok zmiennych zero-jedynkowych, przy czym tu pomijany jest pierwszy stan każdej zmiennej jako kategoria odniesienia[51]. Sposób przekształcania tych zmiennych na blok zmiennych zero-jedynkowych Z_j pokazany jest na przykładzie zmiennej (X_2) opisującej klasę miejscowości:

$$Z_9 = \begin{cases} 1 & \text{dla miast } 20-100 \text{ tys,} \\ 0 & \text{w przeciwnym przypadku;} \end{cases}$$

$$Z_{10} = \begin{cases} 1 & \text{dla miast poniżej 20 tys,} \\ 0 & \text{w przeciwnym przypadku;} \end{cases}$$

$$Z_{11} = \begin{cases} 1 & \text{dla wsi,} \\ 0 & \text{w przeciwnym przypadku;} \end{cases}$$

$Z_9 = Z_{10} = Z_{11} = 0$ dla miast o liczbie ludności powyżej 100 tysięcy (stan pomijany). Zatem zmienne Z_1 do Z_8 określają stany od 2 do 9 cyklu rozwojowego gospodarstwa domowego, a zmienne Z_{12} do Z_{15} – poziomy wykształcenia od 2 do 5 (dla danych z 1997 r. poziom 5 został pominięty). Oddziaływanie pominiętych stanów na zmienną zależną, którą w modelu jest indeks stymulujący posiadanie dobra

[50] Do obliczeń został wykorzystany program „Analiza probitowa" opracowany w Centrum Informatycznym UW, zmodyfikowany w CBR AE w Poznaniu.

[51] Jak już wspomniano wcześniej, wybór pomijanych stanów jest dowolny. Por. p. 2.

trwałego, kumuluje się w wyrazie wolnym. Dla każdego dobra równanie indeksu będzie teraz miało postać:

$$I = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_{15} Z_{15}. \tag{18}$$

W przypadku analizy logitowej, ze względu na charakter nadany zmiennym objaśniającym, wykorzystany został model (17). W tym przypadku dla każdego dobra równanie określające logit Y ma postać

$$Y = \ln \frac{c}{1-c} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_{14} Z_{14}, \tag{19}$$

przy czym zmienne Z_j określają te same stany co w modelu (18).

4. Prezentacja i interpretacja wyników

Wyniki estymacji dla danych z roku 1987 przedstawia tabela 1. Zawiera ona oszacowane współczynniki równań (18) dla 14 wybranych dóbr trwałego użytku. W główce tabeli liczby oznaczają kolejne stany każdej zmiennej (bez stanu pierwszego), a litery b_i (i = 1,..., 15) wyrażają oceny parametrów β_i przy kolejnych zmiennych zero-jedynkowych w równaniu (18). Gwiazdki przy współczynnikach oznaczają ich istotność przy poziomie istotności $\alpha = 0{,}05$ na podstawie testu t-studenta. Ostatnia kolumna zawiera obliczone statystyki χ^2 umożliwiające ocenę zgodności rozkładu empirycznego zmiennej objaśnianej z rozkładem $N(0,1)$. Obliczona wartość tej statystyki mniejsza od podanej wartości krytycznej dla 95% przedziału ufności nie przeczy hipotezie o zgodności rozkładów.

Na podstawie ostatniej kolumny tabeli 1 łatwo można stwierdzić dobrą zgodność rozkładów dla każdego modelu. Również większość oszacowanych parametrów jest statystycznie istotna. Daje to podstawę do wnioskowania, że wszystkie bloki zmiennych wywierają wpływ na proces gromadzenia dóbr trwałych przez gospodarstwa domowe; tylko w nielicznych przypadkach wpływ ten jest mało istotny. Stosunkowo słaby związek z poziomem zasobów pralek i wirówek elektrycznych oraz gramofonów elektrycznych wykazuje zmienna cyklu rozwojowego. Miejsce zamieszkania nie wpływa na posiadanie samochodu osobowego i tylko słabo oddziałuje na posiadanie odkurzacza i zamrażarki. Natomiast wykształcenie głowy gospodarstwa domowego, zgodnie z oczekiwaniem, zdaje się być czynnikiem wpływającym na gromadzenie wszystkich uwzględnionych w badaniu dóbr trwałych.

Tabela 1. Oceny parametrów modelu probitowego dla danych z roku 1987

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie				Chi-kw
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	5	Wart. kryt. 162
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	
Telewizor czarno-biały	0,563*	–0,456*	–0,358*	–0,407*	–0,384*	–0,233	–0,362*	–0,424*	–0,249	0,238*	0,436*	0,572*	0,241*	0,538*	0,844*	0,870*	28,44
Telewizor kolorowy	–0,379*	0,597*	0,628*	0,920*	0,883*	0,407*	0,901*	0,813*	0,178	–0,148*	–0,264*	–0,634*	–0,347*	–0,738*	–1,227*	–2,017*	118,51
Gramofon elektryczny	–0,626*	0,294	0,325*	0,363*	0,535*	0,227	0,212	0,167	–0,046	–0,122*	0,036	–0,343*	–0,399*	–0,652*	–0,821*	–1,066*	94,95
Magnetofon mono	–0,197*	–0,103	0,219*	0,455*	0,456*	0,388*	0,060	–0,202	–0,419*	0,093*	0,097*	0,021	–0,146*	–0,242*	–0,386*	–0,515*	114,11
Pralka i wirówka	–0,369*	–0,298*	–0,061	–0,010	0,030	0,174	–0,096	0,157	0,180	0,440*	0,495*	0,977*	0,338*	0,667*	1,006*	0,750*	47,06
Automat pralniczy	0,212*	0,736*	1,025*	1,133*	1,035*	0,578*	0,957*	0,855*	0,034	–0,219*	–0,248*	–0,672*	–0,570*	–1,022*	–1,491*	–1,758*	84,12
Odkurzacz elektryczny	1,345*	0,333	0,717*	0,995*	0,959*	0,590*	0,959*	1,057*	0,294*	0,035	0,027	–0,381*	–0,197*	–0,610*	–0,978*	–1,426*	16,82
Chłodziarka	1,490*	0,354	0,577*	0,897*	0,791*	0,615*	0,643*	0,745*	0,396*	0,160*	0,268*	–0,094	–0,305*	–0,414*	–0,464*	–0,809*	7,75
Zamrażarka	–1,504*	0,292	0,560*	0,804*	0,835*	0,278	0,689*	0,773*	0,015	–0,072	–0,008	0,181*	–0,295*	–0,415*	–0,653*	–1,325*	136,26
Maszyna do szycia el.	–0,536*	0,555*	0,582*	0,819*	0,853*	0,524*	0,649*	0,662*	0,189	–0,144*	–0,149*	–0,271*	–0,182*	–0,323*	–0,615*	–0,791*	86,96

cd. tab. 1

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie				Chi-kw
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	5	Wart. kryt. 162
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	
Maszyna do szycia zw.	–1,775*	–0,697*	0,072	0,347*	0,422*	0,386*	0,056	0,725*	0,503*	0,166*	0,141*	0,255*	0,118	0,256*	0,437*	0,504*	83,86
Rower	–0,984*	0,213	0,858*	1,538*	1,317*	1,098*	0,332*	0,732*	0,017	0,340*	0,514*	0,927*	–0,153*	–0,249*	–0,184*	–0,370	137,51
Motocykl, skuter, motorower	–2,804*	0,657*	0,576*	0,651*	0,761*	0,183	0,622*	0,608*	–0,250	0,439*	0,776*	0,989*	0,178*	0,390*	0,437*	0,295	131,91
Samochód osobowy	–0,935*	0,991*	1,121*	1,380*	1,363*	0,332*	1,290*	1,356*	0,269	–0,059	–0,036	0,042	–0,542*	–1,030*	–1,314*	–2,049*	140,98

*Istotność na poziomie $\alpha = 0,05$.
Źródło: opracowanie własne.

Zasadniczym celem pracy nie jest jednak szczegółowa analiza związków pomiędzy poziomem wyposażenia a wybranymi cechami gospodarstw domowych, lecz wykazanie użyteczności zastosowanych metod do wykrycia i oceny tych związków oraz interpretacja otrzymanych wyników. Dlatego szczegółowo omówiony zostanie jeden z oszacowanych modeli, co stanowić będzie podstawę do interpretacji pozostałych. Wybrany model, który dotyczy indeksu stymulującego wyposażenie gospodarstw domowych w automaty pralnicze, jak wynika z tablicy 1, ma równanie:

$$\begin{aligned}\hat{I} = {} & 0{,}212 + 0{,}736Z_1 + 1{,}025Z_2 + 1{,}133Z_3 + 1{,}035Z_4 + 0{,}578Z_5 + \\ & + 0{,}957Z_6 + 0{,}855Z_7 + 0{,}034Z_8 - 0{,}219Z_9 - 0{,}248Z_{10} - 0{,}672Z_{11} - \\ & - 0{,}570Z_{12} - 1{,}022Z_{13} - 1{,}491Z_{14} - 1{,}758Z_{15}.\end{aligned} \tag{20}$$

W równaniu tym tylko współczynnik przy zmiennej Z_8 określającej ostatni etap cyklu rozwojowego jest statystycznie nieistotny. Wyraz wolny ujmuje efekty oddziaływania na indeks kategorii bazowych poszczególnych cech, czyli gospodarstwa domowego osoby młodej samotnej, z dużego miasta, z wykształceniem wyższym. Dla takiego gospodarstwa domowego, czyli dla konfiguracji wartości (1,1,1) trzech zmiennych objaśniających, wartość indeksu (probitu) $\hat{I} = 0{,}212$, dla której z tablicy dystrybuanty rozkładu $N(0,1)$ odczytać można odpowiadające jej prawdopodobieństwo. Zatem prawdopodobieństwo posiadania automatu pralniczego przez to gospodarstwo domowe wynosi $P(\hat{I}) = F(\hat{I}) = \Phi(0{,}212) \approx 0{,}584$. Wartości oszacowanych współczynników w ramach każdego bloku należy odnosić do właściwej temu blokowi kategorii bazowej. Zatem współczynniki $b_1 - b_8$ wyrażają efekty cząstkowe poszczególnych etapów cyklu rozwojowego gospodarstwa względem młodej osoby samotnej, współczynniki $b_9 - b_{11}$ – efekty cząstkowe klas miejscowości względem miast największych, współczynniki $b_{12} - b_{15}$ – efekty cząstkowe poziomów wykształcenia względem wykształcenia wyższego, przy założeniu, że wartości zmiennych w pozostałych blokach pozostaną bez zmian. Każdy efekt cząstkowy określa zmianę wartości indeksu stymulującego I, a w konsekwencji zmianę prawdopodobieństwa posiadania dobra trwałego spowodowaną zmianą (jednostkową) danej cechy, czyli przejściem od kategorii bazowej do kategorii wyróżnionej (od stanu pominiętego do wyróżnionego). Na przykład $b_1 = 0{,}736$ określa wzrost indeksu spowodowany zmianą gospodarstwa młodej osoby samotnej na gospodarstwo młodego małżeństwa bez dzieci na utrzymaniu. Wynika to z różnicy wartości indeksu obliczonych dla konfiguracji wartości cech (2,1,1) i (1,1,1), czyli z różnicy $\hat{I}_1(b_0 + b_1 Z_1) - \hat{I}_0(b_0) = (b_0 + b_1 \cdot 1) - b_0 = b_1$ (dla pierwszej konfiguracji $Z_1 = 1$ a pozostałe zmienne są zerami, dla drugiej konfiguracji wszystkie Z_j są zerami)[52]. Zmiana prawdopodobieństwa wywołana tym efektem

[52] Zob. sposób definiowania bloków zmiennych zero-jedynkowych, s. 26.

cząstkowym jest różnicą prawdopodobieństw $P(\hat{I}_1) - P(\hat{I}_0) = \Phi(0{,}948) - \Phi(0{,}212) \approx 0{,}828 - 0{,}584 = 0{,}244$. Jeżeli do obliczeń wzięte będą inne wartości klasy miejscowości i wykształcenia, to wprawdzie cząstkowy efekt zmiany indeksu pozostanie ten sam (0,736), lecz odpowiadająca mu zmiana prawdopodobieństwa będzie inna.

Równanie (20) pozwala obliczyć prawdopodobieństwo posiadania automatu pralniczego przez dowolne gospodarstwo domowe. Niech to będzie gospodarstwo starszego małżeństwa bez dzieci (stan 8), mieszkającego w małym mieście (stan 3), którego głową jest osoba z podstawowym wykształceniem (stan 4). Ta rodzina jest określona przez konfigurację wartości cech (8,3,4), dla której zmienne zero-jedynkowe przyjmują wartości $Z_7 = Z_{10} = Z_{14} = 1$, a pozostałe są zerami. Na podstawie (20) wartość indeksu wynosi $\hat{I} = 0{,}212 + 0{,}855 - 0{,}248 - 1{,}491 = -0{,}672$, dla której prawdopodobieństwo $P(\hat{I}) = \Phi(-0{,}672) + 1 - \Phi(0{,}672) \approx 1 - 0{,}749 = 0{,}328$[53].

Możliwość obliczenia tych prawdopodobieństw oraz ich zmian sugeruje także możliwość wykorzystania metody do celów prognostycznych. Istnieje wiele różnych sytuacji, dla których możliwe jest obliczenie prawdopodobieństwa pewnego zdarzenia przy określonej konfiguracji czynników bądź obliczenie zmiany prawdopodobieństwa spowodowanej zmianą tej konfiguracji. Zdarzeniem takim może być np. zakup jakiegoś dobra trwałego albo pakietu ubezpieczeniowego przez gospodarstwo domowe, urodzenie kolejnego dziecka w rodzinie, rozpad małżeństwa, skuteczność zastosowanego leczenia, skuteczność danego rodzaju reklamy, niespłacenie kredytu, bankructwo firmy i wiele innych. Jeżeli oszacowane na podstawie modelu prawdopodobieństwo zdarzenia jest mniejsze od 1/2, to uważa się zazwyczaj, że przy danej konfiguracji czynników zdarzenie to w przyszłości nie wystąpi, jeżeli jest większe niż 1/2, to przyjmuje się, iż zdarzenie to w przyszłości wystąpi.

Do oceny związków między badanym zjawiskiem a zespołem czynników obliczanie prawdopodobieństw nie jest konieczne. Wartości współczynników równania (20) określają względną siłę oddziaływania poszczególnych stanów zmiennych objaśniających na posiadanie pralek automatycznych, natomiast znaki wskazują kierunek tego oddziaływania względem stanów odniesienia. Pośrednio na podstawie tych wartości można wnioskować także o wpływie zmiennych pierwotnych. Wystarczy zatem ograniczyć się do analizy oszacowanego równania probitu (indeksu).

Łatwo jest zauważyć, że gospodarstwa domowe młodych i starszych osób samotnych są zdecydowanie gorzej zaopatrzone w automaty pralnicze niż inne gospodarstwa, szczególnie małżeństw mających dzieci na utrzymaniu (co jest dość oczywiste). Związek między miejscem zamieszkania a poziomem zasobów

[53] Wskaźnik wyposażenia na 100 gospodarstw domowych tej grupy wynosił 33,3 czyli na 1 gospodarstwo 0,333. Por. B. K. Bednarek, *Fazy...*, op. cit., s.104.

automatów pralniczych jest wyraźny, lecz słabszy niż w przypadku zmiennej cyklu rozwojowego (mniejsze wskaźniki), przy czym wykazuje prawidłowość: im mniejsza miejscowość, tym niższy jest ten poziom. Najsilniejszy wpływ na stan posiadania automatów pralniczych wywiera poziom wykształcenia głowy gospodarstwa domowego (największe współczynniki) i również jest ukierunkowany: im wyższy poziom wykształcenia, tym wyższy stan zasobów. Można zatem stwierdzić, że wszystkie trzy cechy wpływają znacząco na kształtowanie się zasobów pralek automatycznych w gospodarstwach domowych, przy czym wpływ zmiennej cyklu rozwojowego jest niemonotoniczny, natomiast wpływ pozostałych dwóch zmiennych jest monotoniczny.

Analiza wszystkich oszacowanych modeli pozwala potwierdzić powyższe spostrzeżenia. Jakkolwiek wszystkie trzy uwzględnione cechy wpływają na stan posiadania dóbr trwałych przez gospodarstwa domowe, to jednak wpływ ten jest zróżnicowany zarówno co do siły, jak i kierunku. Zdaje się nie podlegać dyskusji, że największą siłę sprawczą wykazuje wykształcenie głowy gospodarstwa domowego[54] (najwyższe współczynniki w większości modeli). Tylko na zasoby pojazdów jednośladowych zmienna ta wywiera nieco słabszy wpływ niż pozostałe zmienne. Natomiast klasa miejscowości zdecydowanie najsłabiej oddziałuje na poziom wyposażenia gospodarstw domowych, gdyż tylko w przypadku motocykli, skuterów i motorowerów jej wpływ jest największy. Obie zmienne mają wspólną własność: ich oddziaływanie jest monotoniczne, tzn. wzrostowi ich wartości towarzyszy (prawie zawsze) wzrost lub spadek zasobów przedmiotów trwałego użytkowania. Dla większości dóbr trwałych ze wzrostem poziomu wykształcenia lub klasy miejscowości związany jest wzrost zasobów, odwrotnie niż w przypadku dóbr starszej generacji, mających nowocześniejsze substytuty. Ta prawidłowość nie dotyczy tylko rowerów, bowiem ich zasoby rosną wraz ze wzrostem wykształcenia, lecz przy spadku klasy miejscowości.

W dziesięcioleciu 1987–1997 zaszły ważne wydarzenia w życiu społeczno-gospodarczym kraju, które wywarły silny wpływ na zachowania konsumpcyjne gospodarstw domowych: 1) duży i gwałtowny na początku lat dziewięćdziesiątych spadek dochodów, będący skutkiem kryzysu gospodarczego i restrykcyjnego programu naprawczego[55], 2) powstanie i rozwój rynku konsumenta z jego wszystkimi atrybutami, 3) korzystne dla dóbr trwałego użytku zmiany relacji cen[56]. Trudno ocenić wpływ pierwszego z tych zdarzeń na zakupy dóbr trwałych, natomiast

[54] Jest to w dużym stopniu uzasadnione silnym związkiem między poziomem wykształcenia i dochodami, a ostatecznie o realizacji zamiaru zakupu decyduje dochód.

[55] W stosunku do roku 1989 dochody realne gospodarstw domowych spadły w następnym roku o 15–20% i mniej więcej na tym poziomie utrzymywały się do roku 1993. Zob. B. K. Bednarek, *Zróżnicowanie...*, op.cit., s. 10–12.

[56] Ibidem, s. 15–17.

oddziaływanie dwóch pozostałych było bardzo silne. Jak pokazują badania[57], w pierwszej połowie lat dziewięćdziesiątych nastąpiła gruntowna modernizacja zasobów dóbr trwałego użytku w gospodarstwach domowych: przedmioty przestarzałe, zużyte fizycznie lub moralnie, w dużym stopniu zostały wycofane i zastąpione przedmiotami nowej generacji. Poza tym gospodarstwa domowe wzbogaciły się o wiele dóbr zupełnie nowych, dotąd nie spotykanych. Z tego powodu lista dóbr trwałych objętych badaniami budżetów gospodarstw domowych została w latach dziewięćdziesiątych zmodernizowana i poszerzona. Dlatego porównań wyposażenia gospodarstw domowych w trwałe dobra konsumpcyjne w latach 1987 i 1997 można dokonywać tylko w ograniczonym zakresie. Niemniej zasadnym wydaje się pytanie: czy w zmienionych warunkach rynkowych zmieniło się też oddziaływanie uwzględnionych czynników na poziom wyposażenia.

Wyniki oszacowania probitowych modeli wyposażenia gospodarstw domowych w dobra trwałego użytku w roku 1997 zawiera tabela 2. Tak jak poprzednio nie można odrzucić hipotezy o zgodności rozkładu empirycznego zmiennej endogenicznej ze standaryzowanym rozkładem normalnym dla wszystkich modeli. Również zdecydowana większość oszacowanych parametrów wykazuje statystyczną istotność, co świadczy o istniejących związkach pomiędzy wszystkimi blokami zmiennych i zasobami dóbr trwałych. Charakter tych związków w ciągu dziesięcioleci w zasadzie nie zmienił się. Tak jak poprzednio najsilniej na wielkość zasobów dóbr trwałego użytku oddziałuje wykształcenie głowy gospodarstwa domowego, następnie cykl rozwojowy, a najsłabiej klasa miejscowości. Także kierunki tego oddziaływania pozostały bez zmian. Wyraźnie widoczny jest stymulujący wpływ poziomu wykształcenia na wielkość zasobów dóbr trwałych (wyjątek stanowią motocykle, skutery i motorowery oraz czarno-białe telewizory). Oddziaływanie klasy miejscowości jest dwukierunkowe: im większe miejscowości, tym większe zasoby nowoczesnego sprzętu elektronicznego, lecz mniejsze zasoby indywidualnych środków komunikacji oraz zamrażarek.

Wpływ zmiennej cyklu rozwojowego gospodarstwa domowego zachował swój niemonotoniczny charakter, można jednak dostrzec wyraźne zmiany w relacjach między oddziaływaniem faz niektórych jego stanów. W roku 1987 najgorzej wyposażone w dobra trwałego użytku były gospodarstwa młodszych i starszych osób samotnych, a w dziesięć lat później za takie należy uznać gospodarstwa domowe rodzin niepełnych i starszych osób samotnych. Wynika to prawdopodobnie z ich stale pogarszającej się sytuacji finansowej. Natomiast wyposażenie gospodarstw domowych młodych osób samotnych relatywnie wyraźnie się poprawiło i jest nawet lepsze od wyposażenia gospodarstw młodych małżeństw bez dzieci. Zmienna cyklu rozwojowego nie wpływa na kształtowanie zasobów

[57] Ibidem, s. 22–25.

Tabela 2. Oceny parametrów modelu probitowego dla danych z roku 1997

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie			Chi-kw
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	Wart. kryt. 155
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	
Telewizor czarno- biały	–1,151*	–0,545*	–0,427*	–0,231*	–0,184*	–0,435*	–0,323*	–0,291*	–0,159	–0,049	–0,035	0,042	–0,003	–0,045	0,179*	118,21
Telewizor kolorowy	1,979*	–0,158	0,286*	0,483*	0,584*	0,079	–0,061	0,458*	–0,379*	0,063	–0,065	–0,253*	–0,069	–0,320*	–0,630*	4,45
Telewizja satelitarna	0,095	–0,084	0,191*	0,402*	0,531*	0,192*	0,186	0,487*	–0,005	0,001	–0,317*	–0,859*	–0,090*	–0,391*	–0,621*	74,57
Magnetofon	–0,844*	–0,366*	–0,103	–0,030	–0,053	–0,246*	–0,287*	–0,227*	–0,446*	–0,093*	–0,039	0,002	–0,101*	–0,187*	–0,280*	106,56
Odtwarzacz CD	–0,762*	0,082	0,202*	0,125	0,322*	–0,052	0,351*	–0,210	–0,576*	–0,157*	–0,323*	–0,442*	–0,285*	–0,611*	–0,928*	108,19
Wieża	0,010	0,452*	0,250*	0,161*	0,318*	–0,166	0,559*	–0,331*	–0,875*	–0,129*	–0,292*	–0,475*	–0,311*	–0,533*	–0,774*	99,54
Magnetowid	0,430*	0,122	0,324*	0,614*	0,657*	–0,044	0,361*	0,149	–0,470*	–0,053	–0,128*	–0,349*	–0,038	–0,219*	–0,642*	54,69
Kamera video	–1,393*	–0,298	0,213	0,206	0,139	–0,388	0,203	0,076	–0,342	–0,114	–0,286*	–0,404*	–0,371*	–0,747*	–1,197*	97,45
Komputer PC	–0,662*	0,101	0,035	0,545*	0,773*	0,278*	0,036	–0,177	–0,576*	–0,189*	–0,350*	–0,530*	–0,629*	–1,057*	–1,200*	128,64
Pralka i wirówka	–0,708*	–0,673*	–0,494*	–0,352*	–0,357*	–0,256*	–0,608*	–0,437*	–0,220*	0,454*	0,555*	0,880*	0,342*	0,681*	0,929*	81,58
Automat pralniczy	1,160*	0,116	0,416*	0,716*	0,799*	0,195*	0,399*	0,705*	–0,115	–0,166*	–0,324*	–0,712*	–0,310*	–0,859*	–1,308*	46,30

cd. tab. 2

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie			Chi-kw
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	Wart. kryt. 155
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	
Odkurzacz elektryczny	1,885*	–0,393*	0,149	0,347*	0,516*	0,115	0,050	0,521*	–0,169	–0,025	–0,013	–0,391*	–0,039	–0,372*	–0,670*	9,09
Chłodziarka	2,057*	–0,309	0,012	0,127	0,178	–0,144	0,074	–0,003	–0,169	0,061	0,188	–0,059	–0,008	–0,087	–0,247*	2,22
Zamrażarka	–0,439*	–0,085	–0,083	0,280*	0,457*	–0,184*	–0,367*	0,323*	–0,540*	–0,072*	0,170*	0,419*	–0,216*	–0,377*	–0,611*	101,68
Robot kuchenny	0,215*	0,436*	0,487*	0,522*	0,496*	0,121	0,261*	0,368*	–0,135	0,169*	0,209*	0,026	–0,202*	–0,515*	–0,731*	54,08
Maszyna do szycia	–0,016	–0,787*	–0,494*	–0,022	0,148*	–0,248*	–0,648*	0,063	–0,364*	0,001	–0,035	–0,089*	0,029	–0,080*	–0,246*	84,12
Rower	–0,159*	–0,630*	–0,427*	0,301*	0,394*	–0,109	–0,406*	–0,337*	–0,804*	0,379*	0,643*	0,822*	–0,074	–0,112*	–0,207*	81,34
Motocykl, skuter, motorower	–2,509*	–0,109	–0,068	0,089	0,122	–0,929*	–0,044	–0,090	–0,179	0,234*	0,589*	0,739*	0,240*	0,333*	0,338*	115,62
Samochód osobowy	–0,044	0,550*	0,649*	0,789*	0,838*	–0,459*	0,633*	0,833*	–0,430*	0,077*	0,097*	0,230*	–0,517*	–0,936*	–1,408*	102,12

*Istotność na poziomie $\alpha = 0,05$.
Źródło: opracowanie własne.

kamer wideo, chłodziarek oraz motocykli, skuterów i motorowerów. Chłodziarka jest dobrem najpowszechniejszym i właściwie nieodzownym w gospodarstwie domowym, toteż brak jest także wyraźnego związku między zasobami tego dobra a dwiema pozostałymi zmiennymi.

Przy konstrukcji zmiennej cyklu rozwojowego istniała obawa, czy nowe kryterium wyodrębniania faz dla gospodarstw domowych małżeństw bez dzieci na utrzymaniu będzie odpowiednie. Po oszacowanych współczynnikach równań można sądzić, że gospodarstwa domowe znajdujące się w tych fazach dość wyraźnie różnią się między sobą pod względem wyposażenia w dobra trwałego użytku. Wydaje się więc, że zastosowane kryterium jest do zaakceptowania.

Omówione wyniki znajdują potwierdzenie w oszacowanych modelach logitowych, których parametry[58] zawiera tabela 3. Ostatnia kolumna tej tabeli zawiera wartości statystyki $L^2 = -2 \ln (L_0/L_1) = (-2\ln L_0) - (-2\ln L_1)$, która jest miarą wiarogodności oszacowanych modeli. Dwukrotny logarytm wiarogodności $-2\ln L$ (lub $-2LL$) jest tym mniejszy, im większa jest wiarogodność wyników i dla modelu idealnego (gdy wiarogodność $L = 1$) osiąga wartość 0. Statystyka L^2 jest różnicą wartości statystyki dla modelu zwierającego jedynie stałą i statystyki dla modelu ze zmiennymi. Duża wartość tej statystyki, związana z małym poziomem istotności, pozwala odrzucić hipotezę o poprawności modelu zawierającego wyłącznie stałą. Ponieważ statystyka L^2 ma asymptotyczny rozkład chi-kwadrat[59] przy r stopniach swobody (r – różnica między liczbą szacowanych parametrów w obu modelach), zatem wartością krytyczną jest wartość statystyki $\chi^2 (a, r)$ dla ustalonego poziomu istotności (tu $\alpha = 0{,}05$). Obliczona wartość L^2 większa od krytycznej oznacza, że przynajmniej jeden parametr modelu (poza wyrazem wolnym) jest różny od zera, czyli jest statystycznie istotny. Istotność poszczególnych parametrów oceniana jest za pomocą statystyki Walda

$$\left[\frac{b_j}{S(b_j)}\right]^2, \quad j = 1, 2, \ldots, k, \tag{21}$$

gdzie $S(b_j)$ jest standardowym błędem szacunku współczynnika b_j, która również ma rozkład chi-kwadrat (w tablicy współczynniki istotne oznaczone są gwiazdką).

Porównanie tablic 2 i 3 pozwala stwierdzić wysoką zgodność oszacowanych modeli probitowych i logitowych pod względem istotności współczynników oraz ich znaków i wzajemnych relacji, różnią się tylko bezwzględnymi wartościami.

[58] Obliczenia wykonane zostały przy użyciu procedury pakietu statystycznego SPSS dla mikrokomputerów.

[59] Por. J. Tobin, op. cit., s. 10–13; S. S. Halli, K. V. Rao, op. cit., s. 104.

Tabela 3. Oceny parametrów modelu logitowego dla danych z roku 1997

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie			L^2
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	Wart. kryt. 23,09
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	
Telewizor czarno-biały	–1,949*	–1,180*	–0,853*	–0,443*	–0,351*	–0,870*	–0,599*	–0,567*	–0,298	–0,099	–0,064	0,086	–0,007	–0,096	0,336*	51,46
Telewizor kolorowy	3,753*	–0,386	0,661*	1,096*	1,366*	0,153	–0,104	1,031*	–0,825*	0,174	–0,142	–0,530*	–0,158	–0,784*	–1,456*	164,60
Telewizja satelitarna	0,153	–0,192	0,302*	0,648*	0,860*	0,301*	0,307	0,787*	–0,008	–0,001	–0,515*	–1,411*	–0,141*	–0,632*	–1,005*	1094,41
Magnetofon	–1,397*	–0,750*	–0,179	–0,046	–0,087	–0,453*	–0,460	–0,406*	–0,841*	0,169*	–0,065	0,005	–0,181*	–0,339*	–0,521*	75,96
Odtwarzacz CD	–1,290*	0,102	0,405*	0,256	0,620*	–0,055	0,703*	–0,367	–1,113*	–0,289*	–0,586*	–0,843*	–0,501*	–1,117*	–1,778*	509,57
Wieża	0,017	0,768*	0,414*	0,266*	0,525*	–0,272	0,935*	–0,554*	–1,515*	–0,211*	–0,479*	–0,786*	–0,508*	–0,869*	–1,282*	873,88
Magnetowid	0,727*	0,123	0,516*	1,002*	1,077*	–0,096	0,611*	0,225	–0,779*	–0,091	–0,215*	–0,583*	–0,069	–0,374*	–1,062*	785,84
Kamera video	–2,448*	–0,642	0,459	0,567	0,323	–0,903	0,418	0,177	–0,764	–0,251	–0,633*	–0,917*	–0,772*	–1,636*	–2,870*	232,69
Komputer PC	–1,174*	0,227	0,085	1,031*	1,440*	0,530*	0,219	–0,279	–1,095*	–0,333*	–0,640*	–0,999*	–1,085*	–1,888*	–2,209*	1164,82
Pralka i wirówka	–1,184*	–1,065*	–0,800*	–0,570*	–0,578*	–0,408*	–0,989*	–0,706*	–0,347*	0,752*	0,914*	1,440*	0,579*	1,130*	1,529*	1334,45
Automat pralniczy	2,084*	0,171	0,701*	1,225*	1,374*	0,294	0,760*	1,218*	–0,233	–0,320*	–0,599*	–1,250*	–0,599*	–1,582*	–2,312*	1379,77

cd. tab. 3

Nazwa dobra trwałego	Wyraz wolny	Stany cyklu rozwojowego								Klasa miejscowości			Wykształcenie			L^2
		2	3	4	5	6	7	8	9	2	3	4	2	3	4	Wart. kryt. 23,09
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	
Odkurzacz elektryczny	3,544*	–0,764*	0,309	0,750*	1,110*	0,212	0,113	1,123*	–0,384	–0,074	–0,039	–0,860*	–0,084	–0,851*	–1,419*	221,61
Chłodziarka	3,902*	–0,849	0,028	0,315	0,440	–0,351	0,222	–0,006	–0,400	0,162	0,470	–0,138	–0,010	–0,214	–0,615*	30,08
Zamrażarka	–0,706*	–0,133	–0,151	0,456*	0,742*	–0,325*	–0,593*	0,525*	–0,959*	–0,112*	0,286*	0,688*	–0,358*	–0,626*	–0,996*	633,96
Robot kuchenny	0,349*	0,674*	0,794*	0,854*	0,813*	0,202	0,437*	0,601*	–0,209	0,278*	0,349*	0,047	–0,338*	–0,850*	–1,203*	495,44
Maszyna do szycia	–0,024	–1,212*	–0,805*	–0,038	0,232*	–0,405*	–1,045*	0,096	–0,589*	–0,001	–0,060	–0,147*	0,050	–0,127*	–0,373*	395,91
Rower	–0,268*	–1,057*	–0,691*	0,496*	0,650*	–0,178	–0,644*	–0,535*	–1,314*	0,615*	1,050*	1,366*	–0,118	–0,182*	–0,339*	1489,06
Motocykl, skuter, motorower	–4,954*	–0,404	–0,211	0,166	0,227	–2,631*	–0,162	–0,222	–0,489	0,575*	1,398*	1,697*	0,576*	0,784*	0,776*	215,20
Samochód osobowy	–0,078	0,882*	1,070*	1,300*	1,379*	–0,767*	1,098*	1,371*	–0,746*	0,122*	0,157*	0,374*	–0,853*	–1,532*	–2,291*	1481,53

*Istotność na poziomie $\alpha = 0,05$.
Źródło: opracowanie własne.

Znajduje potwierdzenie głoszony w literaturze pogląd, że w bardzo wielu przypadkach obie metody dają zbieżne wyniki, a parametry równań różnią się tylko co do czynnika skalarnego[60]. Istotnie pomnożenie, np. równania modelu probitowego dla automatu pralniczego, przez 1,8 daje równanie, którego współczynniki niewiele się różnią od współczynników odpowiedniego równania modelu logitowego:

$$1{,}8\,\hat{I} = 2{,}088 + 0{,}209Z_1 + 0{,}749Z_2 + 1{,}289Z_3 + 1{,}438Z_4 + 0{,}351Z_5 + \\ + 0{,}718Z_6 + 1{,}269Z_7 - 0{,}207Z_8 - 0{,}299Z_9 - 0{,}583Z_{10} - 1{,}282Z_{11} - \\ - 0{,}558Z_{12} - 1{,}546Z_{13} - 2{,}354Z_{14}, \tag{22a}$$

$$Y = 2{,}084 + 0{,}171Z_1 + 0{,}701Z_2 + 1{,}225Z_3 + 1{,}374Z_4 + 0{,}294Z_5 + \\ + 0{,}760Z_6 + 1{,}218Z_7 - 0{,}233Z_8 - 0{,}320Z_9 - 0{,}599Z_{10} - 1{,}250Z_{11} - \\ - 0{,}599Z_{12} - 1{,}582Z_{13} - 2{,}312Z_{14}. \tag{22b}$$

Interpretacja parametrów modelu logitowego jest podobna do interpretacji w modelu probitowym. Każdy oszacowany współczynnik określa zmianę wartości logitu, a w konsekwencji zmianę prawdopodobieństwa posiadania dobra trwałego spowodowaną jednostkowym przyrostem danej cechy (od 0 do 1), czyli przejściem od kategorii bazowej (0) do innej (1), przy założeniu, że wartości pozostałych cech nie ulegną zmianie. Znaki przy współczynnikach pozwalają wnioskować, czy prawdopodobieństwo posiadania dobra trwałego przez gospodarstwo domowe z wyróżnioną kategorią jest większe (znak +), czy mniejsze (znak –) od takiego prawdopodobieństwa dla gospodarstwa domowego z kategoria bazową. Z równania (22b) wynika, że dla gospodarstwa młodej osoby samotnej z wyższym wykształceniem, mieszkającej w dużym mieście, logit $Y = 2{,}084$, a prawdopodobieństwo posiadania automatu pralniczego $P = 1/(1 + e^{-2{,}084}) = 0{,}889$. Przejście do kategorii trzeciej, czyli do gospodarstwa domowego młodego małżeństwa z dziećmi przedszkolnymi, spowoduje przyrost logitu o współczynnik $b_2 = 0{,}701$ do wartości $Y = 2{,}084 + 0{,}701 = 2{,}785$, której odpowiada prawdopodobieństwo $P = 1/(1 + e^{-2{,}785}) = 0{,}942$ i w efekcie przyrost prawdopodobieństwa o 0,053. Wartości bezwzględne i znaki parametrów modelu pozwalają wnioskować o sile i kierunku oddziaływania poszczególnych kategorii (stanów) cech na kształtowanie zasobów poszczególnych dóbr trwałego użytku w gospodarstwach domowych.

W analizach modeli logitowych łatwiej i prościej jest rozpatrywać szansę (ryzyko względne) zdarzenia niż logarytm szansy (logit) ze względu na konkretną

[60] Możnik jest liczbą zbliżoną do 2. D. J. Finney w pracy *Probit Analysis*, University Press, Cambridge 1971, s. 47 proponuje możenie przez 2, natomiast Halli i Rao w cytowanej pracy (s. 112) sugerują mnożenie przez 1,8. Porównując wartości logitów i probitów z prawdopodobieństwami można ocenić, że prawdopodobieństwu około 0,5 odpowiada mnożnik 1,6 a prawdopodobieństwu bliskiemu 0 lub 1 mnożnik 2.

interpretację. Wymaga to przekształcenia oszacowanego równania do postaci (16), która dla automatu pralniczego jest następująca:

$$\psi = e^{2,084} e^{0,171Z_1} e^{0,701Z_2} e^{1,225Z_3} e^{1,374Z_4} e^{0,294Z_5} e^{0,706Z_6} e^{1,218Z_7} \cdot e^{-0,233Z_8} e^{-0,320Z_9} e^{-0,599Z_{10}} e^{-1,250Z_{11}} e^{-0,599Z_{12}} e^{-1,582Z_{13}} e^{-2,312Z_{14}}. \quad (23)$$

Każda potęga e^{b_j} wyraża mnożnikowy (multiplikatywny) efekt zmiany ryzyka względnego spowodowany przyrostem zmiennej Z_j o jednostkę przy nie zmienionych wartościach pozostałych zmiennych, czyli zmiany stanu bazowego cechy na stan wyróżniony. Zatem $e^{0,701} = 2,016$ oznacza, że przejście od gospodarstwa domowego młodej osoby samotnej do gospodarstwa małżeństwa z dziećmi przedszkolnymi na utrzymaniu zwiększa ponad dwukrotnie szansę posiadania automatu pralniczego.

Powyższa interpretacja bierze się stąd, że e^{b_j} jest stosunkiem dwóch rodzajów ryzyka względnego obliczonych dla $Z_j = 1$ i $Z_j = 0$ przy dowolnych, lecz ustalonych wartościach pozostałych zmiennych. Dla rozpatrywanych poprzednio dwóch typów gospodarstw domowych, czyli dla konfiguracji zmiennych (3,1,1) i (1,1,1), stosunek ten jest następujący:

$$\frac{\psi(3,1,1)}{\psi(1,1,1)} = \frac{e^{2,084} e^{0,701 \cdot 1}}{e^{2,084} e^{0,701 \cdot 0}} = e^{0,701} = 2,016.$$

Z drugiej strony, korzystając z obliczonych wcześniej prawdopodobieństw posiadania pralki automatycznej przez te gospodarstwa, otrzymuje się ten sam wynik. Istotnie, $P(3,1,1) = 0,942$ i $P(1,1,1) = 0,889$, zatem $\Psi(3,1,1) = 0,942/(1 - 0,942) = 16,241$ i $\Psi(1,1,1) = 0,889/(1 - 0,889) = 8,009$, stąd $\Psi(3,1,1)/\Psi(1,1,1) = 16,241/8,009 = 2,028$ (nadmiar jest wynikiem zaokrągleń). Potraktowanie e^{b_j} jako stosunku ryzyk względnych pozwala stwierdzić, że jeżeli:

- $e^{b_j} > 1$, to kategoria zmiennej reprezentowana przez Z_j zwiększa szansę (a także prawdopodobieństwo) wystąpienia badanego zjawiska (np. posiadania dobra trwałego) w porównaniu z kategorią bazową;
- $e^{b_j} < 1$, to kategoria zmiennej reprezentowana przez Z_j zmniejsza szansę wystąpienia badanego zjawiska w porównaniu z kategorią bazową;
- $e^{b_j} = 1$, to obie te kategorie dają jednakowe szanse wystąpienia badanego zjawiska.

Mnożąc różnicę $e^{b_j} - 1$ przez 100, przyrost lub spadek szansy wystąpienia danego zdarzenia można wyrazić w procentach. Nawiązując do poprzednich rozważań, można powiedzieć, że szansa posiadania automatu pralniczego przez gospodarstwa domowe małżeństw z dziećmi w wieku przedszkolnym jest o $(e^{0,7801} - 1) \cdot 100\% = 101,6\%$ większa niż przez gospodarstwa młodych osób samotnych.

Istnieje możliwość określenia bezpośrednich relacji między siłą oddziaływania dwóch dowolnych kategorii na zajście rozpatrywanego zdarzenia. Relację tę określa stosunek mnożników e^{b_j} dla wybranych kategorii cechy. Ponieważ e^{b_j} jest stosunkiem ryzyka względnego dla kategorii wyróżnionej i bazowej Ψ_j / Ψ_0, więc

$$\frac{e^{b_j}}{e^{b_i}} = e^{b_j - b_i} = \frac{\psi_j / \psi_0}{\psi_i / \psi_0} = \frac{\psi_j}{\psi_i}.$$

Zatem wyrażenie $e^{b_j - b_i}$ jest stosunkiem dwóch względnego ryzyka i określa, ile razy szansa zajścia badanego zdarzenia dla kategorii określanej przez Z_j jest większa (lub mniejsza) od szansy dla kategorii reprezentowanej przez Z_i. Wynik można

Tabela 4. Porównanie prawdopodobieństw oszacowanych na podstawie modelu probitowego i logitowego i częstości względnych posiadania automatu pralniczego przez gospodarstwa domowe w roku 1997

Konfiguracja cech (X_1, X_2, X_2)	Częstość względna	Prawdopodobieństwo		Odchylenie w %	
		model probitowy	model logitowy	model probitowy	model logitowy
(1,1,1)	0,929	0,877	0,889	–5,597	–4,306
(2,1,1)	0,750	0,899	0,905	19,867	20,667
(3,1,1)	0,949	0,943	0,942	–0,632	–0,738
(4,1,1)	0,986	0,971	0,965	–1,521	–2,130
(5,1,1)	0,967	0,975	0,969	0,827	0,207
(6,1,1)	0,921	0,912	0,915	–0,977	–0,651
(7,1,1)	0,837	0,942	0,945	12,545	12,903
(8,1,1)	0,939	0,969	0,964	3,195	2,662
(9,1,1)	0,797	0,852	0,864	6,901	8,406
(1,2,1)	0,857	0,840	0,854	–1,984	–0,350
(1,3,1)	0,875	0,798	0,815	–8,800	–6,857
(1,4,1)	0,778	0,673	0,697	–13,496	–10,411
(1,1,2)	0,778	0,802	0,815	3,085	4,756
(1,1,3)	0,818	0,618	0,623	–24,450	–23,839
(1,1,4)	0,700	0,441	0,443	–37,000	–36,714

Źródło: opracowanie własne.

wyrazić także w procentach. Na przykład szansa posiadania automatu pralniczego przez gospodarstwo domowe małżeństwa z dziećmi w wieku szkolnym jest $e^{1,225-0,701} = 1,69$ razy lub o 69% większa od szansy posiadania tego dobra przez gospodarstwo domowe małżeństwa z dziećmi przedszkolnymi na utrzymaniu.

Ze względu na cel pracy interesująca jest konfrontacja otrzymanych wyników z wartościami empirycznymi. W pewnym stopniu porównanie oszacowanych prawdopodobieństw posiadania automatu pralniczego przez niektóre typy gospodarstw domowych z prawdopodobieństwami empirycznymi, wyrażonymi przez częstości względne, umożliwia tabela 4. Przede wszystkim uwagę zwraca wzajemna, bardzo wysoka zgodność prawdopodobieństw obliczonych na podstawie obu modeli, ponieważ różnice między prawdopodobieństwami są niewielkie. Można również mówić o daleko idącej zgodności tych prawdopodobieństw z częstościami empirycznymi, bowiem tylko w kilku przypadkach różnice są stosunkowo duże. Nie można jednak oczekiwać od wyników estymacji, bazujących na wielkościach uśrednionych, aby idealnie pasowały do wyników obserwacji. Zatem dopasowanie obu modeli do danych empirycznych uznać można za dobre. Świadczy to o przydatności omawianych metod analizy do kwantyfikacji związków między wyposażeniem gospodarstw domowych w dobra trwałego użytku a różnymi charakterystykami badanej zbiorowości.

5. Podsumowanie

Zarówno w świecie „realnym", jak i w dziedzinie badań naukowych dość powszechnie występują zjawiska określane mianem zjawisk jakościowych. Zjawiska te są charakteryzowane za pomocą cech opisowych (kwalitatywnych) o praktycznie skończonej liczbie kategorii, których pomiar dokonywany jest w skali nominalnej lub porządkowej. Wiele z tych zjawisk ma charakter dychotomiczny opisany przez cechę mającą tylko dwie kategorie (wyrażone np. słowami TAK lub NIE), której odpowiednikiem jest zmienna zero-jedynkowa. Ze zjawiskami tego rodzaju szczególnie często mamy do czynienia w dziedzinie nauk społecznych, takich jak psychologia, socjologia i demografia, a także w dziedzinie ekonomii, medycyny, biologii, w badaniach marketingowych i wielu innych.

Do badania związków między współwystępującymi zjawiskami jakościowymi służą metody statystyczne, w wyniku których otrzymujemy różne współczynniki skojarzenia (asocjacji) cech jakościowych opisujących te zjawiska. Współczynniki skojarzenia mierzą siłę tych związków a niektóre także ich kierunek. Jeżeli związki między zjawiskami istnieją, to zazwyczaj chcemy również wiedzieć, czy dane zjawisko może być uznane za efekt działania jednego bądź wielu innych zjawisk

i poznać jego zasadę. Reguła ta powinna określać nie tylko sposób i zakres tego działania, lecz także umożliwiać przewidywanie zmian cech opisujących te zjawiska w wyniku ich wzajemnego oddziaływania. Celowi temu służą modele regresji.

W wypadku badania zjawisk o charakterze dychotomicznym reprezentująca je zmienna przyjmuje wartości 0 lub 1, zatem najprostszym podejściem wydaje się zastosowanie jako modelu regresji liniowej funkcji prawdopodobieństwa. Parametry tego modelu mają prostą interpretację w kategoriach prawdopodobieństwa, bowiem przyrost prawdopodobieństwa wystąpienia zjawiska następuje w wyniku wzrostu zmiennej niezależnej o jednostkę. Niestety, dla zmiennej dychotomicznej założenia leżące u podstaw modeli liniowych są rzadko spełnione, co znacznie ogranicza możliwości stosowania tego modelu. Pominąwszy pewne niedogodności związane z metodami szacowania parametrów, wynikające z heteroscedastyczności składnika losowego, głównym problemem jest możliwość wyjścia oszacowanych wartości prawdopodobieństw poza przedział $\langle 0, 1 \rangle$. Bezpośrednim rozwiązaniem tego problemu jest przyjęcie stosownych założeń o rozkładzie prawdopodobieństwa i takie przekształcenie oryginalnego modelu, aby otrzymanie wartości prawdopodobieństwa spoza przedziału $\langle 0, 1 \rangle$ było niemożliwe. Spośród kilku monotonicznych przekształceń najczęściej stosuje się transformację probitową i logitową. W pierwszym wypadku zakłada się, że dystrybuanta teoretycznych wartości prawdopodobieństwa jest dystrybuantą standaryzowanego rozkładu normalnego, w drugim, że jest funkcją logistyczną. Otrzymane w wyniku tych transformacji modele regresyjne noszą odpowiednio nazwę modeli probitowych lub logitowych (logistycznych) i są modelami liniowymi. Ich parametry są również interpretowane jako przyrost, lecz tym razem probitu lub logitu, a nie prawdopodobieństwa, spowodowany jednostkowym wzrostem określonej zmiennej niezależnej, przekształcany następnie na przyrost prawdopodobieństwa.

Badanie przeprowadzone na przykładzie problemu z pogranicza demografii i ekonomii w pełni wykazało przydatność obu metod do identyfikacji współzależności między zmiennymi jakościowymi w wypadku dychotomicznej zmiennej zależnej. Konkretne wnioski wypływające z badania można ująć w kilku następujących punktach:

1. Wszystkie uwzględnione zmienne, tj. cykl rozwojowy gospodarstwa domowego, klasa miejscowości i wykształcenie głowy gospodarstwa, wpływają istotnie na poziom wyposażenia gospodarstw domowych w dobra trwałego użytku, przy czym wpływ ten jest zróżnicowany. Największą moc sprawczą posiada wykształcenie głowy gospodarstwa domowego, co może być skutkiem dwojakiego charakteru tej zmiennej: z jednej strony jest ona nośnikiem czynnika kulturowego, a z drugiej jest reprezentantem czynnika dochodowego (poziom dochodu osobistego jest dość silnie skorelowany z poziomem wykształcenia). Jedynie na wyposażenie gospodarstw domowych w pojazdy jednośladowe jej wpływ jest słabszy niż pozostałych zmiennych. Natomiast

najsilniej na poziom zasobów tych pojazdów wpływa klasa miejscowości, wykazując jednak zdecydowanie najsłabsze oddziaływanie w wypadku zasobów pozostałych przedmiotów trwałego użytkowania.

2. Wpływ zmiennych objaśniających różni się także pod względem charakteru i kierunku. Oddziaływania zmiennych poziomu wykształcenia i klasy miejscowości są monotoniczne i prawie zawsze zgodne co do kierunku. W przypadku większości dóbr trwałych wzrostowi poziomu wykształcenia głowy gospodarstwa domowego lub klasy miejscowości towarzyszy również przyrost ich zasobów, natomiast w odniesieniu do dóbr starszej generacji, wychodzących z użytkowania, sytuacja jest odwrotna. Wyjątek stanowią rowery i samochody osobowe, których zasoby wprawdzie rosną wraz z poziomem wykształcenia, lecz przy spadku klasy miejscowości. Oddziaływanie zmiennej cyklu rozwojowego ma wyraźnie charakter niemonotoniczny: na ogół, wyjąwszy przedmioty starszej generacji, stan zasobów zwiększa się wraz z rozwojem rodziny, by później zacząć maleć. Prawidłowość ta w większym stopniu dotyczy wyników z roku 1987 niż wyników z roku 1997.
3. W latach 1987–1997 nastąpiła radykalna zmiana zarówno poziomu, jak i struktury zasobów dóbr trwałego użytku w gospodarstwach domowych. Nie zmienił się natomiast prawie zupełnie charakter związków między poziomem ich zasobów a zmiennymi objaśniającymi. Zmienne wykształcenia głowy gospodarstwa i klasy miejscowości utrzymały swój monotoniczny charakter oddziaływania, a zmienna cyklu rozwojowego zachowała charakter niemonotoniczny. Można jednak zauważyć dość wyraźne zmiany w relacjach między oddziaływaniem niektórych faz zmiennej cyklu. Jeżeli w roku 1987 najgorzej wyposażone były gospodarstwa domowe młodych osób samotnych, a w dalszej kolejności gospodarstwa starszych osób samotnych, to w roku 1997 zdecydowanie najgorzej wyposażone były te ostatnie, a następnie gospodarstwa domowe rodzin niepełnych z dziećmi na utrzymaniu oraz młodych małżeństw bez dzieci. Niewątpliwie jest to skutek stale pogarszającej się sytuacji materialnej tych gospodarstw. Natomiast wyposażenie gospodarstw domowych młodych osób samotnych w tym okresie relatywnie znacznie się poprawiło.
4. Zaniechanie notowania lat trwania związku małżeńskiego w badaniach budżetów gospodarstw domowych wymusiło zmianę zasad konstrukcji zmiennej cyklu rozwojowego gospodarstwa domowego. Staż małżeński na potrzeby tej pracy zastąpiony został wiekiem żony. Skonstruowana według nowej koncepcji zmienna zachowała charakter swej poprzedniczki, a określone na nowy sposób fazy pokazały wyraźnie moc różnicującą. Można zatem uznać, że nowa zmienna cyklu rozwojowego gospodarstwa domowego posiada podobny zasób informacji jak poprzednia i jej wykorzystanie w dalszych badaniach jest uzasadnione.
5. Wyniki uzyskane na podstawie modeli regresji probitowej i logitowej wykazują bardzo dużą zgodność i różnią się właściwie tylko czynnikiem skalarnym, co

wskazuje na istnienie obszaru wspólnych zastosowań. Istnieje pogląd, że dla zjawisk społecznych i ekonomicznych odpowiedni jest model probitowy, natomiast dla zjawisk spotykanych w biologii i medycynie właściwy jest model logitowy. Jakkolwiek przeprowadzone badanie nie potwierdziło tego poglądu, to jednak nie znaczy, że jest on całkowicie niesłuszny. Może natomiast oznaczać, że obszar wspólnych zastosowań omawianych modeli jest bardzo szeroki, co jednocześnie nie wyklucza sytuacji, w których jeden z nich będzie lepszy. Przy braku merytorycznych przesłanek o wyborze powinny decydować kryteria formalne, stosowane powszechnie do oceny poprawności oszacowanych modeli regresji liniowej. Pewnym mankamentem jest brak dobrej miary wyjaśniającej zmienność, będącej odpowiednikiem R^2, pomimo prób jej znalezienia[61].

Uzyskane wyniki nie są sprzeczne ze stanem wiedzy o procesie kształtowania się zasobów dóbr trwałego użytku w gospodarstwach domowych, wykazują przy tym dużą zgodność z danymi empirycznymi i nie stwarzają trudności interpretacyjnych. Pozwala to na ocenę wpływu każdego stanu zmiennych objaśniających na zmienną objaśnianą, a także na porównanie siły oddziaływania tych zmiennych. W wielu przypadkach omawiane modele mogą być wykorzystywane także do celów prognostycznych, czyli do oceny prawdopodobieństwa wystąpienia w przyszłości określonego zjawiska, a w przypadku modelu logitowego także do określenia jego ryzyka względnego.

Podstawową zaletą omawianych metod jest możliwość ich stosowania w sytuacjach, w których nie są spełnione założenia wymagane przy stosowaniu innych, alternatywnych metod, jak regresja liniowa czy analiza dyskryminacyjna. Niemniej spełnienie tych założeń poprawia jakość predykcji, natomiast współliniowość zmiennych objaśniających może prowadzić do obciążenia estymatorów i zwiększenia błędów standardowych. Należy mieć na uwadze, że możliwości stosowania modeli probitowych i logitowych podlegają również ograniczeniu do tych sytuacji, w których przewidywane prawdopodobieństwo ma, przynajmniej w przybliżeniu, rozkład normalny lub logistyczny (w obu wypadkach dystrybuanta rozkładu jest krzywą ogiwalną, czyli krzywą typu *S*). Być może faktyczny zakres ich stosowalności jest szerszy, ponieważ badania symulacyjne wykazują, że niespełnienie założeń dotyczących rozkładów nie powoduje tak groźnych skutków, jakich powszechnie oczekiwano.

Należy jeszcze dodać, że model logitowy może mieć zastosowanie również wtedy, gdy zmienna zależna ma więcej niż dwie kategorie (zastosowanie w tym wypadku modelu probitowego jest wprawdzie logicznie możliwe, lecz niepraktyczne[62]), nie będzie zatem bezpodstawnym stwierdzenie, że skala możliwych zastosowań modeli probitowych i logitowych jest bardzo szeroka.

---

[61] Por . S. S. Halli, K V. Rao, op. cit., s. 102.

[62] Ibidem, s. 115.

Bronisław K. Bednarek

APPLICATION OF THE PROBIT AND LOGIT ANALYSES TO STUDIES ON SOCIO-ECONOMIC PHENOMENA

SUMMARY

Many phenomena in the surrounding world are of qualitative character which can be expressed by zero-one variables. Only a few statistical methods can by used for the studies of this kind of phenomena. The article presents two of them, namely the probit analysis and the logit analysis. The author discusses formal conditions of their application as well as advantages and disadvantages as compared with other methods. The main trend of considerations refers to application and interpretation matters.

The possibility to apply the above-mentioned methods to study cause and effect relation between the qualitative features is shown on the example from the field of socio-economic demography. The results obtained indicate that those methods are extremely useful to study a wide spectrum of socio-economic phenomena as well as other phenomena occurring in different field of human activity.