

STUDIA METODOLOGICZNE

Aleksandra HAŁKA

Determinanty wydatków gospodarstw domowych na usługi związane z wypoczynkiem¹

Podstawowym problemem badawczym postawionym w artykule jest określenie, jakie czynniki wpływają na wysokość wydatków gospodarstw domowych na produkty związane z wypoczynkiem. Zgodnie z teorią ekonomii na wydatki danego podmiotu wpływają przede wszystkim dochód oraz ceny². Różne wzorce zachowań konsumentów zależą m.in. od fazy cyklu życia danego gospodarstwa domowego związanego z liczbą członków, wiekiem i wykształceniem rodziny, liczbą dzieci czy ze statusem cywilnym. W badaniach empirycznych wskazuje się dodatkowo inne determinanty wydatków, np.: typ gospodarstwa domowego, miejsce zamieszkania czy inne czynniki socjoekonomiczne.

W badaniu oszacujemy wpływ wymienionych zmiennych na poziom wydatków gospodarstw domowych. Dane zaczerpnięto z badania budżetów gospodarstw domowych GUS z 2008 r.

Warto podkreślić, że znaczna część (ok. 65%) gospodarstw domowych wykazuje zerowe wydatki na wypoczynek, co stwarza klasyczny problem ocenzero-

¹ Autorka pragnie podziękować Jarosławowi Jakubikowi za wskazówki przekazane podczas pisanie artykułu oraz Danucie Kołodziejczyk za pomoc w napisaniu kodów SAS wykorzystanych do stworzenia bazy danych na potrzeby badania. Szczególne podziękowania autorka chciałaby przekazać drowi Pawłowi Baranowskiemu za cenne uwagi i wskazówki, które zostały wykorzystane w trakcie prowadzonego badania.

² Wykorzystanie tej zmiennej nie było możliwe z uwagi na charakter próby (przekrojowa). Pominiecie cen w analizie oznacza de facto, że przyjęto, iż w badanym okresie (2008 r.) terytorialne różnice cen były pomijalne.

wania zmiennej objaśnianej. Najczęściej proponowanym rozwiązaniem tego problemu jest zastosowanie modelu Tobina (tobitowego). Alternatywę stanowi tutaj zastosowanie metody najmniejszych kwadratów (MNK) na próbie, z której wyłączono obserwacje zerowe. W artykule zostaną porównane wyniki analiz otrzymane na podstawie tych metod.

WYKORZYSTANE METODY ESTYMACJI

W badaniach wydatków gospodarstw domowych, szczególnie na dobra wyższego rzędu, badacze spotykają się z problemem tzw. „wartości zerowych”, czyli braku zarejestrowanych wydatków na dane dobro lub usługę. Takie „wartości zerowe” mogą wynikać z kilku przyczyn. Po pierwsze, może to być dobro rzadko kupowane i gospodarstwo poniosło nań wydatki przed lub po badaniu. Drugim powodem może być fakt, że dany podmiot nie może sobie pozwolić na zakup danego dobra lub też nie kupuje go, ponieważ nie należy ono do jego koszyka dóbr preferowanych.

Modele, które wykorzystują takie dane i nie mają poprawki ze względu na „wartości zerowe” mogą być obciążone (Maddala, 1983). Proste usunięcie obserwacji zerowych sprawia, że próba przestaje być losowa, co zwykle pogarsza własności estymatora (Curry Raper i in., 2002). Wśród metod używanych w celu rozwiązania tego problemu badacze często wykorzystują metodę zaimplementowaną przez Tobina (Tobin, 1958)³. Model ten ma na celu wyjaśnienie relacji pomiędzy obserwowalną zmienną zależną y_i (przyjmującą wartości nieujemne) a wektorem zmiennych wyjaśniających x_i . W modelu tym zakłada się, że istnieje zmienna ukryta (*latent variable*) y_i^* , przyjmująca wartości ze zbioru liczb rzeczywistych⁴. Zmienna ta zależy liniowo od obserwowanych wartości zmiennych niezależnych x_i .

$$y_i^* = \beta x_i + u_i \quad u_i \sim IID(0, \sigma^2) \quad (1)$$

gdzie:

y_i^* — zmienna ukryta (nieobserwowalna),

x_i — k -wymiarowy wektor obserwowanych regresorów,

β — k -wymiarowy wektor parametrów,

u_i — składnik losowy (niezależna zmienna losowa o zerowej wartości oczekiwanej i stałej wariancji).

Z kolei relacja pomiędzy zmienną nieobserwowalną a obserwowalną jest zdefiniowana następująco:

$$y_i = \begin{cases} y_i^* & \text{dla } y_i^* > 0 \\ 0 & \text{dla } y_i^* \leq 0 \end{cases}$$

gdzie y_i — obserwowana zmienna zależna.

³ Cytat za Yen S. T., Jensen H. (1995).

⁴ M. Verbeek (2008, s. 231 i 232) podaje, że w przypadku analizy wydatków można interpretować tę zmienną jako pożądaną przez podmiot wielkość wydatków na dane dobro.

W celu oszacowania parametrów β stosuje się metodę największej wiarygodności (w tym celu dodatkowo zakłada się, że składnik losowy ma rozkład normalny). Metoda ta zapewnia asymptotyczną normalność oszacowań parametrów, w przypadku dużych prób do badania modelu można stosować konwencjonalne testy.

Funkcja wiarygodności w modelu Tobina ma następujący wzór (Greene, 2003; Gruszczyński, 2010):

$$L(\theta) = \prod_0 \left[1 - \Phi \left(\frac{X_i \beta}{\sigma} \right) \right] \prod_1 \sigma^{-1} \varphi \left[\frac{(Y_i - X_i \beta)}{\sigma} \right]$$

gdzie:

Π_0 — iloczyn po tych wartościach i , dla których $y_i^* \leq 0$,

Π_1 — iloczyn po tych wartościach i , dla których $y_i^* > 0$,

$\Phi(\cdot)$ oraz $\varphi(\cdot)$ są odpowiednio skumulowaną dystrybucją i funkcją gęstości prawdopodobieństwa rozkładu normalnego.

Dodajmy, że model tobitowy jest często stosowany do badania wydatków gospodarstw domowych na dobra wyższego rzędu lub dobra trwałe. Takie zastosowanie modelu tobitowego przedstawił już Tobin w swojej pionierskiej pracy. Następnie rozwiązanie to stosowało wielu badaczy (Atkinson i in., 1990; Soberton-Ferrer, Dardis, 1991; Lee, Tan, 2006; Barquet i in., 2010).

W modelu tobitowym najpierw szacujemy prawdopodobieństwo, że zmienna zależna przyjmie wartość równą zero, a następnie determinanty jej wielkości (dla wartości niezerowych) (Gruszczyński, 2010). Standardowo przyjmuje się, że te same zmienne wyjaśniają zarówno sam fakt braku wydatków, jak i ich wielkość — w przypadku gdy są one dodatnie. W ostatnich latach model Tobina został uogólniony, tak aby ominąć to restrykcyjne założenie. Po pierwsze, ujęto możliwość, że zerowa wartość wydatków może wynikać z innych przyczyn (*double-hurdle model*). Po drugie, w modelu rzadkich zakupów (*the infrequency of purchase model*) dopuszcza się zera jako wynik sytuacji, kiedy okres badania jest krótszy niż częstotliwość zakupu danego dobra.

W klasycznym modelu regresji liniowej rozwiązaniem „wartości zerowych” jest wyeliminowanie gospodarstw domowych, które nie notowały wydatków na dane dobro lub usługę. W przypadku gdy występuje problem cenzurowania, zarówno pozostawienie „obserwacji zerowych” jak i ich eliminacja z próby prowadzi do uzyskania obciążonych oszacowań za pomocą klasycznej regresji (Greene, 2003). Wynika z tego, że w przypadku standardowych założeń w warunkach zmiennej ocenzonej model tobitowy stanowi lepsze narzędzie badawcze niż regresja klasyczna. Jednakże w szczególnych przypadkach klasyczna regresja, wykorzystująca próbę „obserwacji niezerowych”, może być lepszym rozwiązaniem niż model tobitowy.

Po pierwsze — klasyczny model tobitowy zakłada, iż składnik losowy w równaniu (1) ma rozkład normalny. W przypadku niespełnienia tego założenia estymator MNW zaproponowany przez Tobina traci zgodność (Arabmazar, Schmidt, 2003). W przypadku klasycznej regresji uchylenie założenia o normalności nie ma wpływu na oszacowanie parametrów. Po drugie — jeśli próg ocen-zurowania nie wynosi zero, wówczas standardowy estymator modelu tobitowego jest również obciążony, a obciążenie to rośnie wraz ze wzrostem wartości progu ocen-zurowania (Zuehlke, 2003).

Szczególnie w przypadku naszych analiz możemy oczekiwać, że próg ocen-zurowania może być znacząco wyższy od zera, ze względu na fakt, iż w wielu przypadkach nie istnieje możliwość wydania dowolnie niskiej kwoty na rekreację (np. minimalny koszt wycieczki zagranicznej sięga ok. 500 zł).

Ze względu na większą odporność MNK na uchylenie niektórych założeń w badaniu zdecydowano się na zastosowanie modelu Tobina i porównanie otrzymanych rezultatów z wynikami uzyskanymi z prostego modelu klasycznej regresji liniowej. Porównanie to pozwoli ocenić, czy w przypadku analiz wydatków gospodarstw domowych w Polsce zastosowanie różnych narzędzi badawczych różnicuje otrzymywane wyniki.

DANE

Analiza wydatków gospodarstw domowych, jak już wspomniano, została oparta na badaniu budżetów gospodarstw domowych GUS z 2008 r. Badanie to jest prowadzone metodą reprezentacyjną, która daje możliwość uogólnienia wyników na wszystkie gospodarstwa domowe w kraju. Od 1993 r. zastosowano miesięczną rotację gospodarstw domowych (w każdym miesiącu badane są inne podmioty⁵). Każde gospodarstwo domowe prowadzi przez miesiąc zapisy rozchodów i przychodów w specjalnych książeczkach budżetowych. Dodatkowo ankieterzy GUS zbierają, w formie wywiadu, informacje m.in. o członkach gospodarstwa domowego, zajmowanym mieszkaniu czy jego wyposażeniu.

Na potrzeby badania budżetów gospodarstw domowych zastosowano dwustopniowe losowanie warstwowe, z różnymi prawdopodobieństwami wyboru na pierwszym stopniu. Jednostkami pierwszego stopnia były terenowe punkty badań, na drugim stopniu losowane były mieszkania.

Pełna próba⁶, wykorzystana w modelu Tobina, liczyła 37155 obserwacji przekrojowych. Dla potrzeb klasycznego modelu regresji liniowej z próby wyeliminowano obserwacje o zerowej wartości zmiennej zależnej, w efekcie otrzymując próbę liczącą 13002 obserwacje.

⁵ Każde wylosowane gospodarstwo domowe badane jest w tym samym miesiącu w dwóch kolejnych latach.

⁶ W analizie pominięto obserwacje, w których wykazane dochody gospodarstwa domowego były mniejsze od zera.

Zmienna objaśniana

W opracowaniu zmienną poddaną analizie jest suma wydatków gospodarstwa domowego na usługi związane z wypoczynkiem w przeliczeniu na jednego członka gospodarstwa. Do wydatków na te usługi zostały zaliczone, według terminologii GUS, wydatki na turystykę zorganizowaną, restauracje i hotele oraz wydatki w zakresie rekreacji i sportu. Znaczna część badanych gospodarstw domowych (ok. 65%) nie przeznaczala pieniędzy na takie wydatki. Może to oznaczać z jednej strony, że wydatki związane z wypoczynkiem są traktowane przez większość społeczeństwa jako dobro luksusowe, na które mogą sobie pozwolić tylko rodziny zamożne. Z drugiej strony, może istnieć grupa gospodarstw domowych, które mogą nie mieć potrzeby takich wydatków. Inną grupą mogą być takie podmioty, które dysponują odpowiednimi zasobami, mają potrzebę takich wydatków, ale nie podejmują ich ze względu na czynniki obiektywne, np. utrudniony dostęp do infrastruktury.

W dalszej części artykułu ilekroć będzie mowa o wydatkach na usługi związane z wypoczynkiem lub też w skrócie o wydatkach na usługi, będzie to oznaczało wydatki na turystykę zorganizowaną, indywidualną (zawartą w kategorii restauracje i hotele) oraz wydatki w zakresie rekreacji i sportu w przeliczeniu na jednego członka gospodarstwa domowego. W przypadku gdy wyrażenie „usługi” będzie się odnosić do szeroko rozumianego zakresu usług, będzie to wyrażenie zaznaczone.

Zmienne objaśniające

Wybór zmiennych został zdeterminowany dwoma czynnikami. Z jednej strony wzięto pod uwagę czynniki mogące wpływać na decyzje konsumentów, a wynikające z teorii ekonomii. Poziom dochodów wpływa na zmianę relacji w wydatkach i oszczędnościach gospodarstw domowych. Wraz z wyższym dochodem wzrasta ilość i różnorodność konsumowanych dóbr.

Z teorii wynika, że wydatki danego gospodarstwa domowego zależą nie tylko od dochodów, ale również od tego, w której fazie cyklu życia znajduje się dane gospodarstwo domowe. Innego rodzaju wydatki będzie preferowało gospodarstwo składające się z dwojga młodych ludzi, a inne rodzina z dziećmi na utrzymaniu. Nie jest to jedynie kwestia wysokości dochodu, lecz także kwestia preferencji, składu, potrzeb rodziny itp.

Z drugiej strony natomiast, przy wyborze opierano się na dotychczasowych badaniach wydatków gospodarstw domowych. Ze względu na niewielką liczbę dostępnych publikacji na temat wydatków konsumpcyjnych polskich gospodarstw domowych wybór zmiennych objaśniających został oparty na informacjach zawartych w dostępnych badaniach dla innych krajów. Do takich badań można zaliczyć m.in. badanie wydatków na usługi prowadzone przez H. Soberton-Ferrer i F. Dardis (1991), badania wydatków na alkohol (A. B. Atkinson, J. Gomułka, N. H. Stern, 1990; S. Yen, H. Jensen, 1995) czy też badania nad wydatkami na jedzenie poza domem prowadzone przez H. Lee i A. Tan (2006).

Jak wskazują badania, wiek głowy rodziny jest jednym z istotniejszych determinant wydatków gospodarstw domowych. Młode osoby przeważnie są bardziej

aktywne niż osoby starsze. Tak więc należałoby przypuszczać, że wyniki badania wskażą ujemną zależność wydatków i wieku głowy gospodarstwa domowego.

Formalne wykształcenie głowy rodziny zostało włączone do analizy ze względu na możliwy wpływ na styl życia i wiedzę dotyczącą przestrzegania zdrowego trybu życia. Z tego względu należy spodziewać się pozytywnego wpływu wykształcenia na wydatki na usługi związane z wypoczynkiem. Zmienne terytorialne zostały uwzględnione, aby zbadać czy istnieje zróżnicowanie preferencji i nawyków konsumentów ze względu na miejsce zamieszkania. Przykładowo, gospodarstwa domowe zamieszkujące obszary wiejskie mogą mieć utrudniony dostęp do infrastruktury związanej z uprawianiem sportu.

W efekcie zdecydowano się uwzględnić następujące cechy gospodarstwa domowego:

- 1) wiek głowy rodziny,
- 2) płeć głowy rodziny,
- 3) wykształcenie głowy rodziny,
- 4) skład gospodarstwa domowego: małżeństwo z dziećmi, małżeństwo bez dzieci, samotni rodzice, osoby bez dzieci na utrzymaniu,
- 5) dochód całkowity gospodarstwa domowego w przeliczeniu na jednego członka gospodarstwa oraz podstawowe źródło tego dochodu,
- 6) wielkość i region zamieszkania,
- 7) pora roku, w której przeprowadzono badanie.

ZMIENNE OBJAŚNIAJĄCE

Zmienna oryginalna	Nazwa zmiennej	Opis	Typ zmiennej
Miejsce zamieszkania	<i>duze_miasto</i>	1 — gospodarstwo w mieście powyżej 200 tysięcy mieszkańców, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>srednie_miasto</i>	1 — gospodarstwo w mieście poniżej 200 tysięcy mieszkańców, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>wies</i>	1 — gospodarstwo na wsi, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
Wiek	<i>wiek</i>	Wiek głowy rodziny w pełnych latach	Zmienna ciągła
Płeć	<i>plec</i>	1 — głowa rodziny jest kobietą, 0 — mężczyzną	Zmienna zero-jedynkowa
Poziom wykształcenia	<i>wyzsze^a</i>	1 — głowa rodziny posiada wykształcenie wyższe, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>srednie</i>	1 — głowa rodziny posiada średnie wykształcenie, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>podstawowe</i>	1 — głowa rodziny posiada podstawowe wykształcenie, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
Główne źródło utrzymania	<i>pracownik^a</i>	1 — podstawowym źródłem utrzymania są dochody z pracy najemnej, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>rolnik</i>	1 — podstawowym źródłem utrzymania są dochody z gospodarstwa rolnego, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa

^a Kategoria odniesienia w modelu.

ZMIENNE OBJAŚNIAJĄCE (dok.)

Zmienna oryginalna	Nazwa zmiennej	Opis	Typ zmiennej
Główne źródło utrzymania (dok.)	<i>wlfirma</i>	1 — podstawowym źródłem utrzymania są dochody z pracy na własny rachunek, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>emeryt</i>	1 — podstawowym źródłem utrzymania są dochody z emerytury, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>in_zatr</i>	1 — podstawowym źródłem utrzymania są inne dochody niepieniężne, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
Typ biologiczny (skład osobowy) gospodarstwa domowego	<i>singiel^a</i>	1 — gospodarstwo jednoosobowe, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>malze</i>	1 — gospodarstwo dwuosobowe, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>malze_dz</i>	1 — małżeństwo z dziećmi na wychowaniu, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>sam_dz</i>	1 — samotny rodzic z dzieckiem na wychowaniu, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>in_typr</i>	1 — inny typ rodziny, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
Pora roku badania	<i>la_zim^a</i>	1 — gospodarstwo domowe było badane w miesiącach letnich lub zimowych, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>wio_je</i>	1 — gospodarstwo domowe było badane w miesiącach wiosennych lub jesiennych, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
Region Polski	<i>pol</i>	1 — gospodarstwo domowe z południowego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>centr^a</i>	1 — gospodarstwo domowe z centralnego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>wsch</i>	1 — gospodarstwo domowe ze wschodniego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>plnzach</i>	1 — gospodarstwo domowe z północno-zachodniego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>polzach</i>	1 — gospodarstwo domowe południowo-zachodniego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>pln</i>	1 — gospodarstwo domowe z północnego regionu Polski, 0 — w pozostałych przypadkach	Zmienna zero-jedynkowa
	<i>doch_cap</i>	dochody gospodarstwa domowego w przeliczeniu na członka rodziny w zł	Zmienna ciągła

^a Kategoria odniesienia w modelu.

WYNIKI BADAŃ EMPIRYCZNYCH

Model klasycznej regresji liniowej

Budowę modelu przeprowadzono dla gospodarstw domowych, w przypadku których zanotowano wydatki na usługi związane z wypoczynkiem. Wszystkie gospodarstwa, w których wydatki były równe zero zostały usunięte z analizy⁷.

⁷ Analogiczny zabieg zastosowali w swoich badaniach K. Curry Raper, M. W. Namakhoye, R. N. Nayga (2002).

W takim modelu, aby otrzymać zgodne estymatory zmienna zależna musi przyjmować wartości większe od zera.

Budowę modelu oparto na doborze zmiennych „od ogółu do szczegółu”, czyli włączeniu wszystkich zmiennych do modelu, a następnie kolejno usuwaniu zmiennych nieistotnych. W pierwotnej specyfikacji modelu przyjęto nieliniową zależność wydatków na usługi związane z wypoczynkiem od dochodu oraz zmiennej *wiek*. W przypadku tej zmiennej wynikało to z faktu, że osoby młode oraz osoby będące na emeryturze mają więcej wolnego czasu, a zatem mogą w większym stopniu (*ceteris paribus*) korzystać z usług związanych z wypoczynkiem.

Z kolei w przypadku zmiennej *doch_cap* założenie o nieliniowości zależności przyjęto na podstawie teorii Engla (Frank, 2007), zgodnie z którą do pewnego poziomu wzrost dochodu będzie przeznaczany na dobra podstawowe. Po przekroczeniu tej granicy, wraz ze wzrostem dochodu udział dóbr podstawowych w koszyku będzie malał, a rósł będzie udział dóbr luksusowych. Nieliniowość tej zależności może również wynikać z faktu, że młodzi ludzie mają mniejsze dochody, co wpływa na mniejsze wydatki na usługi, a wraz ze wzrostem dochodów będzie rosła kwota wydatków na wypoczynek (do pewnego momentu).

Pierwotna specyfikacja modelu wyglądała następująco:

$$wyd_usl_cap_i = \alpha + \sum_{m=1}^3 \beta_m \cdot doch_cap_i^m + \sum_{j=1}^3 \gamma_j \cdot wiek_i^j + \sum_{k=1}^{19} \delta_k \cdot u_{k,i} + \varepsilon_i \quad (2)$$

gdzie:

- wyd_usl_cap* — wydatki na usługi związane z wypoczynkiem w przeliczeniu na jednego członka rodziny,
- doch_cap* — wydatki gospodarstwa domowego w przeliczeniu na jednego członka rodziny,
- wiek* — wiek głowy rodziny liczony w pełnych latach,
- α — stała,
- β, γ, δ — oszacowania parametrów zmiennych,
- ε — składnik losowy,
- u* — zmienne zero-jedynkowe zgodnie z zestawieniem zmiennych objaśniających.

W celu uniknięcia problemów ze współliniowością zmiennych w modelu dla każdej zmiennej oryginalnej, dla której utworzono kilka zmiennych zero-jedynkowych, usunięto jedną zmienną jako kategorię odniesienia.

Ze względu na potencjalne problemy z heteroskedastycznością, typowe dla prób przekrojowych, zdecydowano się na oszacowanie błędów standardowych parametrów modelu według estymatora HC (*Heteroskedasticity Consistent*). W przypadku wystąpienia problemu heteroskedastyczności składnika losowego

estymator MNK pozostaje nieobciążony i zgodny, ale nie jest najefektywniejszy. Jednocześnie ocena wariancji estymatora jest obciążona. W takim przypadku stosuje się estymator HC, który koryguje macierz wariancji i kowariancji.

Po szczegółowej analizie zmiennych z modelu usunięto zmienne związane z głównym źródłem utrzymania, które okazały się być nieistotne statystycznie⁸. W efekcie uzyskano następujący model:

TABL. 1. WYNIKI OSZACOWANIA MODELU MNK

Zmienne	Ocena	Odchylenie standardowe	Statystyka t-Studenta	Wartość p	95% przedział ufności	
					minimum	maksimum
<i>duze_miasto</i>	19,05321	4,48293	4,25	0	10,26601	27,84041
<i>wies</i>	-8,27295	2,975536	-2,78	0,005	-14,10544	-2,440463
<i>wiek</i>	-0,4706842	0,132899	-3,54	0	-0,7311858	-0,2101826
<i>plec</i>	-22,74929	4,03045	-5,64	0	-30,64956	-14,84902
<i>srednie</i>	-18,77559	5,651804	-3,32	0,001	-29,85396	-7,697227
<i>podstawowe</i>	-21,88402	5,634116	-3,88	0	-32,92772	-10,84033
<i>malze</i>	-53,87594	8,38482	-6,43	0	-70,31141	-37,44046
<i>malze_dz</i>	-70,52609	7,298148	-9,66	0	-84,83153	-56,22065
<i>sam_dz</i>	-49,36822	7,890035	-6,26	0	-64,83384	-33,90259
<i>in_typr</i>	-76,52248	6,957328	-11	0	-90,15987	-62,8851
<i>wio_je</i>	-19,429	3,184425	-6,1	0	-25,67094	-13,18706
<i>pol</i>	-3,801611	4,921581	-0,77	0,44	-13,44863	5,845411
<i>wsch</i>	-14,91626	6,213276	-2,4	0,016	-27,09519	-2,737325
<i>plnzach</i>	-14,52878	4,9559	-2,93	0,003	-24,24307	-4,814488
<i>polzach</i>	-8,614296	5,767845	-1,49	0,135	-19,92012	2,691526
<i>pln</i>	-19,28528	5,43464	-3,55	0	-29,93797	-8,632584
<i>doch_cap</i>	0,0683017	0,004678	14,6	0	0,0591322	0,0774712
<i>doch_cap2</i>	-1,02E-06	7,51E-08	-13,52	0	-1,16E-06	-8,69E-07
<i>_cons</i>	128,4493	14,62544	8,78	0	99,78131	157,1173

U w a g a. Jakość dopasowania modelu: statystyka $F(18,12983) = 60,49$; $\text{prob} > F = 0,0000$ ($\text{prob} > F$ — empiryczny poziom istotności testu F); $R^2 = 0,1923$; pierwiastek błędu średniokwadratowego = 181,79.

Ź r ó d ł o: opracowanie własne.

Otrzymane oszacowania parametrów w tym modelu wydają się być zgodne z intuicją ekonomiczną. Zgodnie z oczekiwaniami model wykazał, że gospodarstwa jednoosobowe wydają więcej na usługi związane z wypoczynkiem niż pozostałe typy gospodarstw domowych. Również w tych podmiotach, w których głowa rodziny ma wyższe wykształcenie wydatki są wyższe niż w przypadku, gdy głowa rodziny ma inne wykształcenie. Odnotowano dodatnią zależność w wydatkach na usługi związane z wypoczynkiem w przypadku gospodarstw zlokalizowanych w dużych miastach. Podmioty, które są zlokalizowane w centralnym regionie Polski wydają więcej na usługi niż gospodarstwa z pozostałych

⁸ Wszystkie zmienne zero-jedynkowe związane z regionem pozostawiono w modelu ze względu na statystyczną istotność oszacowań parametrów łącznie dla wszystkich zmiennych (takie restrikcje testowano za pomocą klasycznego testu opartego na statystyce F). Usunięcie zmiennych nieistotnych *pol* oraz *polzach* nie zmieniało znacząco oszacowań pozostałych parametrów modelu.

części kraju. W przypadku badania odbywającego się w miesiącach letnich lub zimowych wydatki na usługi były większe niż w przypadku pozostałych miesięcy.

W celu oceny poprawności specyfikacji modelu przeprowadzono test LINK. Test ten pozwala zweryfikować poprawność funkcyjną modelu, czyli potwierdzić, czy zależność między zmiennymi w modelu jest liniowa oraz czy przy budowie modelu nie pominięto istotnych zmiennych objaśniających. Z kolei brak współliniowości zmiennych objaśniających zbadano wyliczając wartość VIF (*variance inflation factor* — ubytek wariancji) dla każdej ze zmiennych. Pierwiastek kwadratowy z wielkości VIF będzie określał, ile razy większy błąd standardowy oceny parametru danej zmiennej niezależnej otrzymaliśmy niż wynosiłby on, gdyby ta zmienna nie była skorelowana z inną zmienną objaśniającą.

Testowanie poprawności specyfikacji modelu oparto na założeniu, że dodanie dodatkowych zmiennych objaśniających nie zwiększy mocy wyjaśniającej modelu, a dodane zmienne będą nieistotne (Cleves i in., 2008). W przypadku testu LINK buduje się regresję wartości teoretycznych i ich kwadratu (równanie (3)). W przypadku poprawnej specyfikacji modelu parametr α_1 powinien być bliski jedności, natomiast parametr przy dodatkowej zmiennej (α_2) nie powinien być statystycznie istotny⁹. W przypadku testowanego modelu oszacowanie parametru α_1 było statystycznie istotne (statystyka t wyniosła 28,21), natomiast przy kwadracie wartości teoretycznych było nieistotne (α_2 — wartość statystyki 1,85).

$$LRH = \alpha_1(\hat{y}) + \alpha_2(\hat{y})^2 \quad (3)$$

gdzie $\hat{y}$ — wartości teoretyczne.

Jeśli chodzi o analizę wartości VIF, wykazała ona, że dla wszystkich zmiennych wielkość ta znalazła się poniżej wartości przyjmowanej przez badaczy jako wartość progowa (uznaje się, że wartości VIF większe od 10 wskazują na poważne problemy ze współliniowością zmiennych). Największą wartość współczynnika VIF przyjęła zmienna *malze_dz* i wyniosła 2,87.

Model Tobina

Analogicznie jak w przypadku modelu klasycznej regresji liniowej, dobór zmiennych w modelu Tobina został oparty na procedurze „od ogółu do szczegółu” (*general-to-specific*). Jednakże w przypadku modelu tobitowego oszacowań dokonano na całym dostępnym zbiorze danych (37155 obserwacji).

⁹ Idea testu jest zbliżona do testu RESET (rzędu drugiego), jednak procedura testowania przebiega nieco inaczej.

Oto wyniki regresji w modelu tobitowym po usunięciu nieistotnych zmiennych¹⁰.

TABL. 2. WYNIKI OSZACOWANIA MODELU TOBITOWEGO

Zmienne	Ocena	Odchylenie standardowe	Statystyka t-Studenta	Wartość p	95% przedział ufności	
					minimum	maksimum
<i>duze_miasto</i>	35,54	4,208151	8,45	0	27,2919	43,78809
<i>wies</i>	-29,00681	3,360279	-8,63	0	-35,59305	-22,42057
<i>wiek</i>	-4,267327	0,6003876	-7,11	0	-5,444104	-3,090551
<i>wiek2</i>	0,0156279	0,0061419	2,54	0,011	0,0035897	0,0276661
<i>plec</i>	-25,19595	3,607194	-6,98	0	-32,26615	-18,12574
<i>srednie</i>	-59,85944	4,927678	-12,15	0	-69,51782	-50,20105
<i>podstawowe</i>	-91,93531	5,569588	-16,51	0	-102,8519	-81,01876
<i>rolnik</i>	-49,47074	7,749793	-6,38	0	-64,66055	-34,28093
<i>wlfirma</i>	33,64313	6,857921	4,91	0	20,20141	47,08485
<i>emeryt</i>	-40,26001	5,225493	-7,7	0	-50,50213	-30,0179
<i>in_zatr</i>	26,81664	7,486096	3,58	0	12,14368	41,48959
<i>malze</i>	-63,35441	6,702521	-9,45	0	-76,49154	-50,21729
<i>malze_dz</i>	20,58401	6,459611	3,19	0,001	7,922995	33,24503
<i>sam_dz</i>	34,98286	7,354419	4,76	0	20,56799	49,39772
<i>in_typr</i>	-51,37154	6,340097	-8,1	0	-63,79831	-38,94477
<i>wio_je</i>	-8,719418	2,865119	-3,04	0,002	-14,33513	-3,103705
<i>pol</i>	-12,60864	4,464324	-2,82	0,005	-21,35884	-3,858442
<i>wsch</i>	-31,23229	4,915481	-6,35	0	-40,86677	-21,59781
<i>plnzach</i>	-28,81418	4,813746	-5,99	0	-38,24926	-19,37911
<i>polzach</i>	-6,289565	5,282275	-1,19	0,234	-16,64297	4,063841
<i>pln</i>	-32,70908	4,935212	-6,63	0	-42,38223	-23,03593
<i>doch_cap</i>	0,0936238	0,0056577	16,55	0	0,0825345	0,1047131
<i>doch_cap2</i>	-2,43E-06	0,000000471	-5,17	0	-0,00000336	-0,00000151
<i>doch_cap3</i>	1,28E-11	3,2E-12	3,99	0	6,5E-12	1,9E-11

U w a g a. Jakość dopasowania modelu: pseudo $R^2 = 0,0411$; logarytm funkcji wiarygodności = -98175,857.

Ź r ó d ł o: opracowanie własne.

Podobnie jak w przypadku modelu klasycznej regresji liniowej, dokonano oszacowania poprawności specyfikacji wybranego modelu. Analiza testem LINK wykazała, że przedstawiony poniżej model zawiera wszystkie istotne zmienne oraz jego postać funkcyjna jest właściwa. Statystyka testowa t dla wartości teoretycznych z modelu wniosła 86,19, natomiast dla kwadratu wartości teoretycznych -1,82, co wskazuje, że na 10% poziomie istotności nie ma podstaw do odrzucenia hipotezy o poprawności specyfikacji.

¹⁰ Podobnie jak w przypadku klasycznego modelu liniowej regresji, błędy standardowe oszacowano przy pomocy estymatora HC. Wszystkie zmienne zero-jedynkowe związane z regionem pozostawiono w modelu ze względu na statystyczną istotność oszacowań parametrów łącznie dla wszystkich zmiennych (takie restrykcje testowano za pomocą klasycznego testu opartego na statystyce F). Usunięcie nieistotnej zmiennej *polzach* nie zmieniało znacząco oszacowań parametrów modelu.

MODEL TOBITOWY I MODEL KLASYCZNEJ REGRESJI LINIOWEJ — PORÓWNANIE WYNIKÓW

Wyboru postaci funkcyjnej jednego i drugiego modelu dokonano na drodze weryfikacji statystycznej, kierując się testem RESET, jak również statystyczną istotnością parametrów. Testowano funkcje liniowe, kwadratowe oraz trzeciego stopnia dla zmiennych *wiek* oraz dochody na głowę członka gospodarstwa domowego (*doch_cap*). W porównaniu do klasycznego modelu regresji liniowej, model tobitowy wymagał dołączenia zmiennej *wiek* w drugiej potędze oraz zmiennej dochody na osobę (*doch_cap*) w drugiej i trzeciej potędze. Oznacza to, że włączenie „obserwacji zerowych” wymagało zastosowania innej postaci funkcyjnej modelu.

Wyniki otrzymane z regresji tobitowej i modelu klasycznej regresji liniowej są do siebie zbliżone. Modele wskazały, że wydatki gospodarstw domowych na usługi związane z wypoczynkiem były największe w przypadku, gdy: gospodarstwo domowe znajduje się w dużym mieście, głową gospodarstwa domowego jest mężczyzna, głowa rodziny ma wyższe wykształcenie, gospodarstwo domowe jest jednoosobowe, znajduje się w regionie centralnym Polski, a badanie zostało przeprowadzone w miesiącach letnich lub zimowych. Natomiast wraz ze wzrostem wieku głowy gospodarstwa domowego, wydatki na usługi związane z wypoczynkiem w przeliczeniu na członka gospodarstwa maleją.

Znaczącą różnicą jest fakt, że w modelu tobitowym istotne okazały się zmienne zero-jedynkowe, związane z głównym źródłem utrzymania gospodarstwa domowego (w modelu klasycznej regresji liniowej zmienne te nie były statystycznie istotne). Z kolei współczynniki ocen parametrów dla zmiennej określającej małżeństwa z dziećmi na utrzymaniu oraz samotnych rodziców z dziećmi na utrzymaniu miały w nim przeciwne znaki, w porównaniu z modelem klasycznej regresji liniowej.

Podobnie jak w przypadku klasycznej regresji liniowej, zależność pomiędzy wydatkami na usługi a dochodami na głowę członka rodziny nie była liniowa. Dla modelu tobitowego zależność ta przyjmuje formę wielomianu trzeciego stopnia, a w modelu klasycznym jest to zależność kwadratowa. W przypadku modelu tobitowego wysokość wydatków na usługi związane z wypoczynkiem do pewnego momentu rośnie szybciej wraz ze wzrostem dochodów gospodarstw domowych niż w modelu klasycznej regresji liniowej. Gdy roczny dochód na członka rodziny przekroczył kwotę ok. 24 tys. zł, wielkość wydatków związanych z wypoczynkiem zaczynała się zmniejszać. Dla modelu klasycznej regresji liniowej ta granica ujawniła się przy większych dochodach (ok. 33 tys.). Uzasadnieniem takiego zachowania może być obserwowane „nasylenie” usługami związanymi z wypoczynkiem. Wzrost dochodów powyżej pewnej wielkości nie będzie skutkował wzrostem wydatków na te usługi, ze względu na ograniczone zasoby wolnego czasu.

Kolejną różnicą między modelami jest zaobserwowana w modelu tobitowym nieliniowa zależność wydatków na usługi oraz wieku głowy gospodarstwa domowego. Jak wynika z wyk. 2, wysokość wydatków na usługi związane z wypoczynkiem wraz z wiekiem maleje wolniej w modelu tobitowym niż w modelu regresji liniowej.

Porównania obu modeli można dokonać przy użyciu statystyki R^2 (standardowa miara dopasowania dla modelu klasycznej regresji liniowej). W przypadku modelu tobitowego uznaje się, że standardowo wyliczona statystyka pseudo R^2 McFaddena nie jest najlepszą miarą dopasowania¹¹. W tej sytuacji lepszym rozwiązaniem jest oszacowanie wartości R^2 pomiędzy wartościami teoretycznymi a obserwowanymi. W przypadku modelu klasycznej regresji liniowej wielkość R^2 wyniosła 0,19, natomiast dla modelu tobitowego 0,13. Jednakże należy pamiętać, że te dwie wielkości zostały oszacowane na dwóch różnych próbach i dlatego porównania powinny być dokonywane niezwykle ostrożnie.

Porównanie znaków i wartości oszacowań parametrów dla poszczególnych zmiennych, otrzymanych dla modeli regresji liniowej i tobitowego, zawiera tabl. 3.

TABL. 3. UDZIAŁ WARTOŚCI OSZACOWANEGO PARAMETRU Z MODELU REGRESJI LINIOWEJ W OSZACOWANIU PARAMETRU Z MODELU TOBITOWEGO

Zmienna	Znaki ocen parametrów	Iloraz ocen parametrów w %
<i>duze_miasto</i>	jednakowe	0,536
<i>wies</i>	jednakowe	0,285
<i>plec</i>	jednakowe	0,903
<i>srednie</i>	jednakowe	0,314
<i>podstawowe</i>	jednakowe	0,238
<i>malze</i>	jednakowe	0,850
<i>malze_dz</i>	różne	-3,426
<i>sam_dz</i>	różne	-1,411
<i>in_typr</i>	jednakowe	1,490
<i>wio_je</i>	jednakowe	2,228
<i>pol</i>	jednakowe	0,302
<i>wsch</i>	jednakowe	0,478
<i>plnzach</i>	jednakowe	0,504
<i>polzach</i>	jednakowe	1,370
<i>pln</i>	jednakowe	0,590

Źródło: obliczenia własne.

Na podstawie tabl. 3 można stwierdzić, że w przypadku większości zmiennych (11 na 13) znaki ocen parametrów były jednakowe. Znaki ocen różniły się w przypadku zmiennych: *malz_dz* i *sam_dz*. Ponadto 8 spośród 13 zmiennych wskazuje na zbliżone¹² oszacowania dla obu modeli. Na tej podstawie można ocenić, że wyniki są dość odporne ze względu na przyjętą metodę estymacji.

Jednocześnie należy podkreślić, że przyjęcie modelu tobitowego pozwoliło na włączenie zmiennych związanych z głównym źródłem utrzymania oraz wymagało określenia innej postaci funkcyjnej modeli.

¹¹ http://www.ats.ucla.edu/stat/stata/output/Stata_tobit.htm.

¹² Na potrzeby tego porównania przyjęto, że oszacowania parametrów z różnicami w granicach +/-60% są zbliżone.

Podsumowanie

W artykule podjęto próbę odpowiedzi na pytanie, jakie czynniki determinowały wydatki na usługi związane z wypoczynkiem w polskich gospodarstwach domowych. Dla celów badawczych przyjęto jednorównaniowy model ekonometryczny oparty na próbie przekrojowej pochodzącej z badań gospodarstw domowych w 2008 r., w którym parametry zostały oszacowane przy użyciu MNK i modelu Tobina.

Oszacowania parametrów w obu modelach co do kierunku oddziaływania na zmienną zależną były zgodne z wynikami innych badań:

- gospodarstwa domowe zlokalizowane w dużych miastach wydawały więcej na usługi związane z wypoczynkiem, w stosunku do podmiotów z mniejszych miast lub wsi;
- wpływ poziomu wykształcenia był dodatni, tzn. im wyższe wykształcenie głowy gospodarstwa domowego, tym większa skłonność danego podmiotu do korzystania z usług związanych z wypoczynkiem;
- mniejszą skłonność do wydatków na usługi mają gospodarstwa domowe składające się z małżeństw niż jednoosobowe podmioty. Z kolei gospodarstwa, w skład których wchodzi dzieci wskazują na większą skłonność przeznaczania dochodu na wydatki związane z wypoczynkiem.

Wpływ dochodów i wieku okazał się być nieliniowy. Wybór postaci funkcyjnej zależności pomiędzy tymi zmiennymi a wydatkami na usługi związane z wypoczynkiem był zależny od metody estymacji. W przypadku klasycznej regresji liniowej wystarczające było przyjęcie funkcji kwadratowej dla zmiennej związanej z dochodami gospodarstwa na jedną osobę oraz liniowej dla zmiennej *wiek* głowy gospodarstwa domowego. Z kolei model Tobina wymagał zastosowania wielomianu trzeciego stopnia dla zmiennej opisującej dochody oraz funkcji kwadratowej dla zmiennej *wiek*. Zatem na podstawie wyników analizy z modelu Tobina można zakładać, że wzrost wydatków na usługi związane z wypoczynkiem będzie szybciej reagował na wzrost dochodów gospodarstwa domowego niż w przypadku regresji liniowej. Jednak po przekroczeniu pewnego poziomu dochodów w przeliczeniu na członka gospodarstwa domowego wydatki te będą malały szybciej niż w modelu klasycznej regresji liniowej. W przypadku zmiennej *wiek*, model Tobina wskazuje na wolniejszy spadek wydatków na usługi związane z wypoczynkiem wraz ze wzrostem wieku głowy gospodarstwa domowego niż mogłoby to wynikać z modelu regresji liniowej.

Uzyskane wyniki okazały się w znacznym stopniu odporne ze względu na metodę estymacji. Większość oszacowań parametrów jest w obu modelach zbliżona. W przypadku dwóch zmiennych (zmienne zero-jedynkowe oznaczające małżeństwo z dziećmi na utrzymaniu oraz samotni rodzice z dzieckiem na utrzymaniu) oszacowania parametrów miały przeciwne znaki. Jednocześnie model Tobina wskazał na potrzebę dodatkowego uwzględnienia zmiennych związanych z głównym źródłem utrzymania, które w przypadku modelu klasycznej regresji liniowej były w większości statystycznie nieistotne.

Wydaje się, że rozwinięciem przedstawionego badania mogłoby być zastosowanie procedury Heckmana lub uogólnionego modelu Tobina (*double hurdle model*) czy też modelu nieczęstych zakupów (*the infrequency of purchase model*).

mgr Aleksandra Halka — *Instytut Ekonomiczny NBP*

LITERATURA

- Arabmazar A., Schmidt P. (1982), *An Investigation of the Robustness of the Tobit Estimator to Non-Normality*, *Econometrica*, vol. 50, No. 4
- Atkinson A. B., Gomułka J., Stern N. H. (1990), *Spending on Alcohol; Evidence from the Family Spending Expenditure Survey 1970—1983*, „*Economic Journal*”, vol. 100, No. 402
- Budżety Gospodarstw Domowych w 2008 r.* (2009), GUS, Warszawa
- Cleves M., Gutierrez R., Gould W., Marchenko Y. (2008), *An Introduction to Survival Analysis Using Stata*, Stata Press, Texas
- Curry Raper K., Namakhoye W. M., Nayga R. N. (2002), *Food expenditures and household demographic composition in the US: a demand systems approach*, „*Applied Economics*” vol. 34, No. 8
- Frank R. H. (2007), *Mikroekonomia jakiej jeszcze nie było*, GWP Gdańskie Wydawnictwo Psychologiczne
- Greene W. H. (2003), *Econometric Analysis*, Prentice Hall, New Jersey
- Gruszczyński M. (2010), *Mikroekonometria*, Wolters Kluwer Polska, Warszawa
- Heckman J. J. (1979), *Sample Selection Bias as a Specification Error*, *Econometrica*, vol. 47
- Lee H., Tan A. (2006), *Determinants of Malaysian Household Expenditures of Food-Away-From-Home*, Contributed paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia
- Maddala G. S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge
- Tobin J. (1958), *Estimation of Relationships for Limited Dependent Variables*, *Econometrica*, vol. 26, No. 1
- Verbeek M. (2008), *A Guide to Modern Econometrics*, Jon Wiley & Sons Ltd
- Zuehlke T. W. (2003), *Estimation of Tobit model with unknown censoring threshold*, *Applied Economics*, vol. 35, No. 10
- Yen S. T., Jensen H. (1995), *Determinants of Household Expenditures on Alcohol*, Working Paper 95-WP 144

SUMMARY

The question, which factors determine household expenditures for rest services (touristics, sports) is discussed in the article. The classical linear regression as well as the Tobin Model (the right estimation way in case of censored samples according to the econometry theory) are used in the survey. Results of both named methods are in line with economic intuition. The result comparison indicates their good resistance due to the used estimation method but the Tobin model needs more explanatory variables.

РЕЗЮМЕ

В обследовании была предпринята попытка ответа на вопрос, какие факторы детерминируют затраты домашних хозяйств на услуги связанные с отдыхом (туризм и спортивные занятия). Для этой цели использовалась классическая линейная регрессия и модель Тобина, которая согласно теории эконометрии является соответствующим способом оценивания по отношению к попыткам являющимся предметом цензуры. Результаты полученные в обоих методах соответствуют экономической интуиции. Сопоставление этих результатов указывает на их довольно хорошую устойчивость в отношении к принятому методу оценивания, хотя модель Тобина указала на необходимость учета большего числа объясняющих переменных.