

D E B I U T Y S T U D E N C K I E

2 0 2 3

ZASTOSOWANIE METOD ILOŚCIOWYCH W EKONOMII I FINANSACH

pod redakcją
Alicji Grześkowiak
i Piotra Peterneka



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2023

Recenzja

Katarzyna Ostasiewicz

Redakcja wydawnicza

Elżbieta Żurawska-Łuczyńska

Korekta

Katarzyna Gwizda

Skład i łamanie

Beata Mazur

Projekt okładki

Beata Dębska

Na okładce wykorzystano zdjęcia z zasobów 123 Royalty Free

Praca opublikowana na licencji Creative Commons Uznanie autorstwa

Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0).

Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>



ISBN 978-83-67899-08-6 (wersja papierowa)

ISBN 978-83-67899-09-3 (wersja elektroniczna)

DOI: 10.15611/2023.09.3

Druk i oprawa: TOTEM

Dominik Głowacki

e-mail: 177530@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

Badanie zróżnicowania przeciętnego miesięcznego dochodu rozporządzalnego na osobę w województwach w Polsce

DOI: 10.15611/2023.09.3.02

JEL Classification: C12, C51, C52

Streszczenie: W artykule podjęto tematykę związaną z dochodem do dyspozycji, który jest bardzo ważną częścią codziennego życia. Kategoria ta jest znacznie bardziej reprezentatywna niż często stosowane średnie wynagrodzenie brutto. Głównym celem analizy jest identyfikacja zjawisk ekonomicznych, które mają największy wpływ na wysokość dochodu do dyspozycji. Aby osiągnąć ten cel, należy stworzyć i zweryfikować model ekonometryczny. W punkcie drugim przedstawiono dochód do dyspozycji jako propozycję miernika dobrobytu społecznego. Ponadto zaprezentowano wskaźniki inne niż tradycyjny produkt krajowy brutto na mieszkańca. Kolejny punkt poświęcono teoretycznemu aspektowi modelowania ekonometrycznego. Do analizy wybrano model liniowy z wyrazem wolnym. Zmienne zostały dobrane metodą grafową. Po sprawdzeniu istotności parametrów strukturalnych modelu przeprowadzono testy: losowości, symetrii, normalności i homoskedastyczności. Analizę danych przedstawiono w ostatnim punkcie. Zmienną zależną była wysokość dochodu do dyspozycji. Spośród jedenastu potencjalnych zmiennych niezależnych wybrano dwie i wprowadzono do modelu ekonometrycznego. Tymi zmiennymi były saldo miejsc pracy oraz przeciętna liczba osób w gospodarstwie domowym. Okazało się, że mają one istotny wpływ na zmienną objaśnianą. Udowodniono, że model wyjaśnia 70% zmienności zmiennej zależnej. Analiza wykazała zatem zjawiska wpływające na wysokość dochodu do dyspozycji, co było głównym celem badania.

Słowa kluczowe: ekonometria, modelowanie ekonometryczne, jakość życia

1. Wstęp

We współczesnym świecie pieniądź jest czynnikiem determinującym bardzo wiele kwestii związanych z życiem codziennym – od powszedniego problemu, który można wyrazić pytaniem: *stać mnie czy nie stać?*, po sferę dotyczącą prestiżu i przynależności do klasy społecznej. Wobec tak ważnej roli pieniądza w życiu człowieka można zastanowić się, jaka kategoria ekonomiczna najlepiej prezentuje zasobność portfeli. Często stosowana wielkość, czyli średnie wynagrodzenie brutto ma kilka istotnych wad. Po pierwsze, nie każda osoba wydająca pieniądze pracuje, a co za tym idzie pojęcie wynagrodzenia jej nie dotyczy. Do tego grona zaliczają się zarówno osoby nieletnie, jak i seniorzy, którzy pobierają rentę lub emeryturę. Należy mieć również na uwadze podatki oraz składki na ubezpieczenia. Ich odprowadzenie powoduje bowiem, że faktyczna kwota, która zasili budżet domowy będzie mniejsza niż średnie wynagrodzenie brutto. Sytuację materialną zdecydowanie lepiej opisuje

dochód do dyspozycji. Jego wysokość informuje o tym, ile jednostka (lub gospodarstwo domowe) może wydać na konsumpcję lub zaoszczędzić. Wysokość dochodu rozporządzalnego jest na tyle istotną kwestią w codziennym funkcjonowaniu, że można wysnuć tezę, że jest to jeden ze wskaźników poziomu życia. Intuicyjnie wydaje się, że przeciętny dochód rozporządzalny zależy w głównej mierze od zarobków. Celem analizy jest potwierdzenie bądź zaprzeczenie tej tezy oraz wytypowanie kategorii ekonomicznych mających najsilniejszy wpływ na wysokość dochodu do dyspozycji. Możliwe będzie sprawdzenie, czy wskazywany w mediach podział na Polskę A i Polskę B jest wciąż aktualny, czy jest jedynie przestarzałym stereotypem.

2. Metodyka badań

Zgodnie z definicją ekonometria to „nauka o mierzeniu i modelowaniu zjawisk ekonomicznych, w tym ich zależności od innych zjawisk ekonomicznych, demograficznych i socjologicznych oraz od zjawisk przyrodniczych i technicznych w celach poznawczych, symulacyjnych i predykcyjnych” (Dziechciarz, 2003, s. 10). Nakreślanie przyczyn oraz poszukiwanie możliwych skutków określonych zjawisk jest kluczowym aspektem i najczęstszym zastosowaniem ekonometrii w praktyce. W niniejszym rozdziale zostanie zaprezentowany teoretyczny zarys dotyczący miar, metod oraz testów statystycznych, które są niezbędne w modelowaniu ekonometrycznym.

Na wstępnym etapie budowy modelu należy wytypować zbiór potencjalnych zmiennych objaśniających (niezależnych), spośród których zostaną wybrane te, które w największym stopniu wpływają na kształtowanie się zmiennej objaśnianej (zależnej). Należy przy tym pamiętać, aby powiązanie między zmiennymi niezależnymi i zależnymi było merytorycznie uzasadnione. Do wskazania zmiennych o niewystarczającym poziomie zmienności posłuży współczynnik zmienności opisany wzorem:

$$V_j = \frac{s_j}{\bar{x}_j} \quad (1)$$

gdzie: s_j – odchylenie standardowe zmiennej X_j , \bar{x}_j – średnia arytmetyczna zmiennej X_j .

Zwykle jako wartość krytyczną podaje się $V^* = 0,1$, a wartości od niej mniejsze eliminuje się z dalszych rozważań (Dziechciarz, 2003). Można również zastosować współczynnik zmienności oparty na rozstępie, opisany wzorem (Dziechciarz, 2003):

$$W_j = 2 \cdot \frac{\max_i x_{ij} - \min_i x_{ij}}{\max_i x_{ij} + \min_i x_{ij}} \quad (2)$$

gdzie: $\max_i x_{ij}$ – wartość maksymalna danej zmiennej, $\min_i x_{ij}$ – wartość minimalna danej zmiennej.

Należy jednak mieć na uwadze również korelację, która zachodzi między zmienną objaśnianą a poszczególnymi zmiennymi objaśniającymi. Wspomniana korelacja,

a zatem istotne w budowaniu modelu ekonometrycznego wektor i macierz korelacji, opiera się na współczynniku korelacji liniowej Pearsona o wzorze (Ostasiewicz, Rusnak i Siedlecka, 2006):

$$r_{xy} = \frac{C(X, Y)}{s_x s_y} \quad (3)$$

gdzie: s_x – odchylenie standardowe zmiennej X , s_y – odchylenie standardowe zmiennej Y , $C(X, Y)$ – kowariancja między cechami X i Y .

Oprócz interpretacji przedstawionych dla poszczególnych wartości współczynnika korelacji warto również mieć na uwadze test istotności tego współczynnika. Hipoteza zerowa to $H_0: r_{xy} = 0$, natomiast alternatywna $H_1: r_{xy} \neq 0$ (Dziechciarz, 2003). Statystyka testowa obliczana jest za pomocą wzoru:

$$t = \frac{|r_{xy}|}{\sqrt{1 - r_{xy}^2}} \sqrt{n - 2} \quad (4)$$

gdzie: $r_{x,y}$ – współczynnik korelacji liniowej Pearsona pomiędzy cechą X i Y , n – liczba obserwacji.

Dopiero po przeanalizowaniu istotności współczynnika korelacji, odpowiednim zinterpretowaniu jego wartości oraz po obliczeniu współczynnika zmienności można przystąpić do wytypowania zbioru potencjalnych zmiennych objaśniających, wśród których znajdują się zmienne, które ostatecznie trafią do modelu ekonometrycznego. Niezależnie od tego, czy wybrana zostanie do tego metoda: wskaźników pojemności informacji, grafowa, czy też analizy współczynników korelacji, niezbędne jest stworzenie macierzy oraz wektora korelacji. Spośród trzech wymienionych wcześniej metod w dalszej perspektywie będzie zastosowana opisana w niniejszej pracy metoda grafowa. Kluczowe z punktu widzenia dalszego postępowania jest wyznaczenie krytycznej wartości współczynnika korelacji r^* . Posłuży do tego wzór oparty na wartości z rozkładu t -Studenta:

$$r^* = \sqrt{\frac{t^2}{t^2 + n - 2}} \quad (5)$$

gdzie: t – wartość statystyki z tablic t -Studenta dla danego poziomu istotności α oraz $n - 2$ stopni swobody (Dziechciarz, 2003).

Następnie w macierzy korelacji wszystkie wartości mniejsze niż obliczony w poprzednim kroku współczynnik korelacji r^* należy zastąpić zerami. Na podstawie nieco zmienionej macierzy korelacji można przystąpić do budowy grafu powiązań, w którym wierzchołkami są kolejne zmienne objaśniające. Jeżeli między parą zmiennych występuje korelacja znacząca, czyli że współczynnik korelacji liniowej Pearsona nie został uprzednio zastąpiony zerem, między wierzchołkami odpowiadającymi danym zmiennym należy poprowadzić odcinek łączący. Należy tak postępować wobec

wszystkich zmiennych objaśniających. Po zbudowaniu grafu (lub grafów) można przejść do wytypowania zmiennych do modelu. Automatycznie do dalszej analizy przechodzą wierzchołki bez żadnych wiązań (grafy jednoelementowe). W przypadku wierzchołków z niezerową liczbą odcinków wychodzących do modelu ekonometrycznego należy dołączyć zmienną, która odpowiada wierzchołkowi z największą liczbą odcinków. Jeżeli istnieją dwa lub więcej takie wierzchołki grafu, należy wybrać tę zmienną, która osiąga wyższą wartość bezwzględną w wektorze korelacji.

Estymacja parametrów, jak i cała procedura tworzenia modelu ekonometrycznego zostanie przeprowadzona za pomocą klasycznej metody najmniejszych kwadratów. Ogólny opis metody, wzory oraz poszczególne przekształcenia opisano w wielu pozycjach naukowych (Dziechciarz, 2003; Gładysz i Mercik, 2007). Po zbudowaniu modelu koniecznym krokiem jest weryfikacja jego poprawności.

Do oceny poprawności dopasowania służy współczynnik determinacji, który oblicza się za pomocą wzoru:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (6)$$

Współczynnik ten przyjmuje wartości z przedziału od zera do jedności. Gdy R^2 osiąga wartość jeden, oznacza to idealne wyjaśnienie przez model zmienności zmiennej. Silnie związany ze współczynnikiem determinacji jest współczynnik zbieżności φ^2 . Suma tych miar daje zawsze wartość równą jeden. Jako że współczynnik determinacji jest ilorazem wyjaśnionej i całkowitej sumy kwadratów, w przypadku dobrego modelu ważne jest, aby R^2 osiągał wartość jak najbliższą jedynce (w sytuacji idealnej, aby tę jedynkę osiągnął). Gdy dopasowanie modelu ocenione współczynnikiem determinacji jest zadowalające, można rozpocząć weryfikację modelu testami statystycznymi. Pierwszym do wykonania jest test istotności statystycznej wektora parametrów strukturalnych (tzw. test Fishera). Hipoteza zerowa w tym teście mówi o tym, że żaden z parametrów nie jest istotny statystycznie, natomiast hipoteza alternatywna jest jej zaprzeczeniem, tzn. przynajmniej jeden z parametrów strukturalnych jest istotnie różny od zera. Statystyka testowa:

$$F = \frac{n-m-1}{m} \cdot \frac{R^2}{1-R^2} \quad (7)$$

ma rozkład F-Snedecora z $n_1 = m$ oraz $n_2 = n-m-1$ stopniami swobody, gdzie n to liczba obserwacji, a m to liczba zmiennych objaśniających (Dziechciarz, 2003). W celu zweryfikowania istotności statystycznej każdego kolejnego parametru należy zastosować test t-Studenta. Dla poszczególnych parametrów, hipoteza zerowa wygląda następująco:

$$t_i = \frac{|a_i|}{S(a_i)} \quad (8)$$

i ma rozkład t-Studenta z $n-m-1$ stopniami swobody, gdzie a_i to ocena parametru, natomiast $S(a_i)$ to standardowy błąd szacunku parametrów. W hipotezie zerowej

zakłada się, że badany parametr jest nieistotnie różny od zera, natomiast hipoteza alternatywna mówi, że jest on istotnie różny od zera.

Zarówno przy badaniu losowości, jak i symetrii składnika resztowego nie jest wymagana znajomość zaawansowanych operacji matematycznych, dlatego przedstawione zostaną obok siebie. Jako pierwszy zostanie przeprowadzony test losowości. Hipotezą zerową jest stwierdzenie, że ciąg reszt ma charakter losowy, natomiast hipotezą alternatywną jest brak losowości składnika resztowego. Na początku należy uporządkować reszty zgodnie z wartościami jednej ze zmiennych objaśniających. Resztom większym od zera można przypisać literę a, natomiast resztom ujemnym literę b. Statystyką testową jest liczba serii, czyli ciągów tych samych znaków. Aby móc uznać reszty za losowe wartość statystyki testowej, musi znaleźć się ona między dwoma wartościami krytycznymi, odczytanymi z tablicy testu liczby serii (Dziechciarz, 2003).

Na podobnym przypisaniu wartości reszt do liter można oprzeć test symetrii. Statystyka testowa ma postać:

$$t_{emp} = \frac{\left| \frac{m-1}{n-2} \right|}{\sqrt{\frac{\frac{m}{n} \left(1 - \frac{m}{n} \right)}{n-1}}} \quad (9)$$

gdzie: m to liczba reszt oznaczonych literą a (reszty dodatnie), natomiast n to liczba obserwacji.

Dla zadanego poziomu istotności oraz $n-1$ stopni swobody należy odczytać z rozkładu t-Studenta wartość krytyczną. Gdy statystyka testowa przyjmuje wartość mniejszą niż wartość krytyczna, należy przyjąć hipotezę zerową o symetrii składnika resztowego (Gładysz i Mercik, 2007).

Testem dobrze sprawdzającym się w przypadku badania normalności w małych próbach jest test Hellwiga. Hipoteza zerowa zakłada normalność rozkładu reszt, natomiast hipoteza alternatywna neguje to stwierdzenie. Aby przeprowadzić test, należy dokonać standaryzacji, dzieląc reszty przez odchylenie standardowe. Dla reszt uporządkowanych niemalejąco trzeba obliczyć ich wartości dystrybuanty. Tak obliczone wartości należy umieścić w celach, czyli w częściach odcinka od zera do jedynki podzielonych na tyle fragmentów, ile jest obserwacji. Kluczowe jest wyznaczenie liczby cel pustych. Z tablicy testu Hellwiga należy odczytać wartości krytyczne. W przypadku gdy liczba cel pustych jest mniejsza niż dolna wartość krytyczna lub większa niż górna wartość krytyczna, hipotezę zerową należy odrzucić (Dziechciarz, 2003).

Ostatnim, niezbędnym do weryfikacji modelu, testem jest sprawdzenie homoskedastyczności składnika resztowego. W tym celu zostanie zastosowany test Goldfelda-Quandt. Do przeprowadzenia tego testu niezbędne jest podzielenie obserwacji na dwie grupy. Hipoteza zerowa mówi o tym, że wariancje w obu podpopulacjach nie różnią się od siebie, natomiast hipoteza alternatywna jest negacją tego stwier-

dzenia. Statystyką testową jest iloraz obu wariancji liczony przez podzielenie wartości większej przez mniejszą. Wartość krytyczną testu należy wyznaczyć z rozkładu F-Snedecora z parą stopni swobody: $n_1 - m - 1$ oraz $n_2 - m - 1$. Jeżeli wartość statystyki testowej jest większa niż wartość krytyczna, hipotezę o równości wariancji w wyznaczonych podpopulacjach należy odrzucić, co oznacza występowanie zjawiska heteroskedastyczności (Dziechciarz, 2003).

Wszystkie przedstawione powyżej testy pełnią ważną rolę. Mają one za zadanie utwierdzenie badacza w przekonaniu, że stworzony za pomocą klasycznej metody najmniejszych kwadratów model jest adekwatny, a co za tym idzie analizowane zjawisko można z dużą dozą prawdopodobieństwa zrozumieć i przewidzieć.

3. Wyniki badań

W poprzednich punktach przedstawiono teoretyczne aspekty problematyki dochodu rozporządzalnego oraz wprowadzenie do konstruowania i testowania modelu ekonometrycznego. Głównym celem prezentowanego badania jest określenie zróżnicowania wysokości dochodu rozporządzalnego w województwach w Polsce. Posłuży do tego modelowanie ekonometryczne, na podstawie którego wytypowane zostaną zjawiska mające największy wpływ na kształtowanie się poziomu dochodu do dyspozycji. Dane, które będą podstawą analizy, pochodzą z Banku Danych Lokalnych, prowadzony przez Główny Urząd Statystyczny [GUS].

Tabela 1. Dochód do dyspozycji i potencjalne zmienne niezależne w 2020 roku

Zmienna Województwo	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Dolnośląskie	1974,72	2,50	1,33	5693,69	5,60	3,10	23,20	2805,19	2484,28	2123,02	219,30	382,22
Kujawsko-Pomorskie	1803,20	2,61	1,31	4831,73	9,00	3,60	9,10	2100,75	2327,44	1779,98	413,10	512,43
Lubelskie	1629,68	2,71	1,27	4914,95	8,20	2,80	2,20	2107,05	2250,32	1893,18	289,50	431,22
Lubuskie	1931,93	2,57	1,31	4832,07	6,30	3,00	-0,10	1009,52	2297,82	1976,14	357,60	616,44
Łódzkie	1818,11	2,59	1,34	5148,38	6,20	3,60	8,60	2431,48	2314,46	1716,82	280,30	422,44
Małopolskie	1888,14	2,83	1,45	5536,07	5,30	4,30	17,10	3795,38	2425,04	2103,64	269,00	589,10
Mazowieckie	2190,56	2,49	1,53	6581,81	5,20	3,30	31,50	6264,11	2590,44	1872,98	251,10	611,32
Opolskie	1653,76	2,65	1,23	5078,51	6,90	3,10	-1,00	863,92	2426,30	1969,30	247,20	557,67
Podkarpackie	1556,77	3,03	1,30	4707,81	9,10	3,20	1,40	2194,59	2217,10	1720,97	338,50	566,02
Podlaskie	1917,16	2,63	1,31	4929,64	7,80	3,00	2,40	1167,34	2244,27	1851,78	376,20	731,82
Pomorskie	1741,89	2,56	1,53	5484,46	5,90	3,40	7,90	2637,72	2447,75	1976,61	285,70	452,85
Śląskie	1995,67	2,48	1,32	5450,86	4,90	3,60	15,90	4432,47	2948,28	2506,22	200,40	428,81
Świętokrzyskie	1696,50	2,73	1,18	4800,21	8,50	3,50	1,20	1149,10	2232,37	1797,12	374,20	482,87
Warmińsko-Mazurskie	1849,02	2,56	1,27	4709,12	10,20	3,60	2,70	1434,95	2275,67	1814,76	490,50	591,58
Wielkopolskie	1753,29	2,69	1,51	4985,75	3,70	3,50	15,50	3976,36	2428,30	1940,28	272,30	468,72
Zachodniopomorskie	1813,33	2,40	1,25	5099,49	8,40	3,60	3,40	1605,56	2370,50	1981,31	289,60	569,37

Źródło: opracowanie własne na podstawie danych z GUS.

W tabeli 1 zaprezentowano dane, na podstawie których zostanie przeprowadzona analiza. Zmienną objaśnianą (Y) jest wysokość przeciętnego, miesięcznego dochodu do dyspozycji przypadającego na jedną osobę w gospodarstwie domowym. Zmienne X_1 – X_{11} są potencjalnymi zmiennymi objaśniającymi:

X_1 – średnia liczba osób w gospodarstwie domowym,

X_2 – współczynnik dietności,

X_3 – przeciętne, miesięczne wynagrodzenie brutto (w zł),

X_4 – stopa bezrobocia (w %),

X_5 – wskaźnik cen towarów i usług (w %),

X_6 – saldo miejsc pracy (w tys.),

X_7 – wydatki województw na program „Rodzina 500+” (w mln zł),

X_8 – średnia miesięczna emerytura brutto (w zł),

X_9 – przeciętna miesięczna renta brutto (w zł),

X_{10} – liczba osób, którym przyznano świadczenia pomocy społecznej w przeliczeniu na 10 tys. mieszkańców,

X_{11} – wydatki województw w przeliczeniu na jednego mieszkańca (w zł).

Aby wyselekcjonować zmienne objaśniające, które wejdą do modelu ekonometrycznego, należy zwrócić uwagę nie tylko na ich merytoryczne powiązanie ze zmienną objaśnianą, ale również na korelację z tą zmienną, korelacje między poszczególnymi zmiennymi niezależnymi oraz współczynniki zmienności tych zmiennych. Dobór zmiennych do modelu zostanie przeprowadzony za pomocą metody grafowej.

Tabela 2. Statystyki opisowe

Zmienna	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Miara											
Śr. arytmetyczna	2,63	1,34	5174,03	6,95	3,39	8,81	2498,47	2392,52	1939,01	309,66	525,93
Odchylenie standardowe	0,15	0,11	485,86	1,84	0,36	9,40	1476,89	181,67	193,22	76,53	93,48
Współczynnik zmienności (%)	5,79	8,04	9,39	26,44	10,61	106,71	59,11	7,59	9,97	24,71	17,78
Max	3,03	1,53	6581,81	10,20	4,30	31,50	6264,11	2948,28	2506,22	490,50	731,82
Min	2,40	1,18	4707,81	3,70	2,80	-1,00	863,92	2217,10	1716,82	200,40	382,22
Rozstęp	0,63	0,35	1874,00	6,50	1,50	32,50	5400,19	731,18	789,40	290,10	349,60
Współczynnik zmienności (rozstęp) (%)	23,20	26,14	33,20	93,53	42,25	213,11	151,52	28,31	37,39	83,98	62,76

Źródło: opracowanie własne.

W tabeli 2 przedstawione zostały podstawowe statystyki opisowe. Analizując przedstawione w tabeli 2 wartości współczynników zmienności, można zauważyć, że w każdym z przypadków współczynnik oparty na rozstępie był większy od tego, który uwzględnia odchylenie standardowe. Zestawiając obliczone współczynniki z często stosowaną wartością krytyczną klasycznego współczynnika zmienności $V_j = 10\%$,

należałoby odrzucić zmienne X_1 , X_2 , X_3 , X_8 oraz X_9 . Biorąc pod uwagę również współczynnik oparty na rozstępie, w przypadku każdej zmiennej przekracza on graniczną wartość 10%. Należy również zapoznać się ze stopniem skorelowania potencjalnych zmiennych objaśniających ze zmienną objaśnianą. W tym celu utworzony został wektor korelacji \mathbf{R}_0 – tabela 3.

Tabela 3. Wektor korelacji \mathbf{R}_0

Zmienna	Korelacja ze zmienną Y
X_1	-0,623
X_2	0,405
X_3	0,716
X_4	-0,447
X_5	0,151
X_6	0,718
X_7	0,580
X_8	0,557
X_9	0,403
X_{10}	-0,214
X_{11}	0,206

Źródło: opracowanie własne.

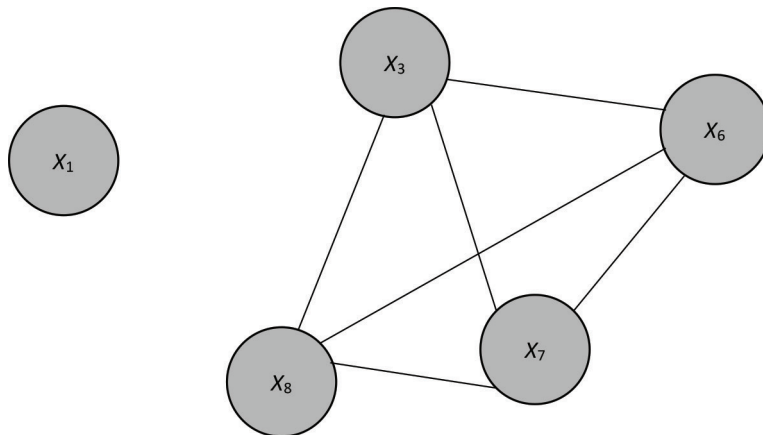
Na podstawie wartości w wektorze korelacji można zauważyć, że aż osiem spośród jedenastu potencjalnych zmiennych objaśniających jest skorelowanych dodatnio ze zmienną objaśnianą. Zależność ujemna występuje dla zmiennych X_1 , X_4 oraz X_{10} . Na podstawie wzoru (5) obliczono, że aby uznać wartość współczynnika korelacji liniowej Pearsona za istotny statystycznie na poziomie istotności $\alpha = 0,05$ oraz dla czternastu stopni swobody, musi on osiągnąć wartość 0,498. Powoduje to usunięcie sześciu zmiennych z grona potencjalnych zmiennych objaśniających. Oznacza to, że do dalszej analizy przejdą zmienne: X_1 , X_3 , X_6 , X_7 oraz X_8 . Tabela 4 przedstawia macierz korelacji \mathbf{R} , ilustrująca współczynniki korelacji liniowej Pearsona między zmiennymi, które na podstawie wektora korelacji zostały wytypowane do dalszej analizy.

W tym momencie należy wyznaczyć wartość krytyczną współczynnika korelacji. Oznacza to, że wszystkie wartości współczynników korelacji mniejsze lub równe r^* w macierzy \mathbf{R} należy zastąpić zerami. Na podstawie utworzonej w ten sposób macierzy \mathbf{R}' utworzone zostały grafy widoczne na rysunku 1.

Tabela 4. Macierz korelacji R

Zmienna \ Zmienna	X_1	X_3	X_6	X_7	X_8
X_1	1	-0,392	-0,290	-0,136	-0,462
X_3	-0,392	1	0,869	0,805	0,624
X_6	-0,290	0,869	1	0,904	0,623
X_7	-0,136	0,805	0,904	1	0,679
X_8	-0,462	0,624	0,623	0,679	1

Źródło: opracowanie własne.



Rysunek 1. Grafy powiązań między zmiennymi

Źródło: opracowanie własne.

Jak można zauważyć na powyższym rysunku, powstały dwa grafy, w tym jeden jednoelementowy. Zmienna X_1 , tworząca graf bez wiązań, wejdzie automatycznie do modelu ekonometrycznego. Spośród zmiennych, tworzących graf wieloelementowy, każda ma po trzy wiązania. W tym przypadku należy wybrać tę zmienną, która jest najsilniej skorelowana ze zmienną objaśnianą. Z grona tych trzech zmiennych największą wartość w wektorze korelacji osiąga zmienna X_6 . Oznacza to, że z grona jedenastu potencjalnych zmiennych objaśniających wytypowane zostały dwie zmienne: liczba osób w gospodarstwie domowym oraz saldo miejsc pracy.

Na początku analizy założono, że pomiędzy zmiennymi objaśniającymi a zmienną objaśnianą występuje zależność liniowa. Możliwe będzie zweryfikowanie tego założenia przez analizę współczynnika determinacji. Niskie wartości mogą bowiem wskazywać na zależność nieliniową. Na tym etapie badania można uznać, że jeśli wartość współczynnika determinacji przekroczy 0,5 – wyjaśnianie zmienności zmiennej obja-

śnianej przez model będzie w miarę satysfakcjonujące. Po wytypowaniu zmiennych, które utworzą model, należy przeprowadzić estymację parametrów, a następnie sprawdzić ich istotność. Na etapie dobrania zmiennych, model ma postać:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_6 x_6 + \varepsilon \quad (10)$$

Estymacja parametrów zostanie przeprowadzona za pomocą klasycznej metody najmniejszych kwadratów, przy zastosowaniu programu Microsoft Excel. W wyniku zastosowania modułu analizy danych uzyskane zostały wartości parametrów strukturalnych, które przedstawia poniższy wzór:

$$\hat{y} = 2980,07 - 472,62x_1 + 9,90x_6 \quad (11)$$

Należy w tym miejscu sprawdzić istotność statystyczną wektora parametrów strukturalnych oraz wartość współczynnika determinacji.

Współczynnik R^2 wynosi 0,70 – co oznacza, że model w 70% wyjaśnia zmienność zmiennej objaśnianej. Jak można zauważyć, współczynnik zbieżności φ^2 ma wartość 0,30. Jego interpretacja jest dopełnieniem informacji zawartej we współczynniku determinacji, tzn. zmienność zmiennej objaśnianej nie została wyjaśniona przez model w 30%. Taki rezultat pozwala sądzić, że model dość dobrze opisuje badane zjawisko. W kolejnym kroku przeprowadzono test istotności wektora parametrów strukturalnych (test Fishera). Statystyka testowa ma wartość 15,41, co przy poziomie istotności $\alpha = 0,05$, parze stopni swobody $n_1 = 3$ i $n_2 = 12$ oraz wartości krytycznej równej 3,49 powoduje, że należy odrzucić hipotezę zerową i przyjąć alternatywną. Oznacza to, że przynajmniej jeden z parametrów jest istotnie różny od zera. Następny zastosowany test bada istotność statystyczną poszczególnych parametrów modelu – jest to test Studenta. Z rozkładu Studenta wyznaczono wartość krytyczną, która na poziomie istotności $\alpha = 0,05$ oraz czternastu stopniach swobody wynosi 2,15. Statystyki testowe obliczone dla poszczególnych parametrów strukturalnych wyniosły odpowiednio: 6,76 dla wyrazu wolnego, $-2,87$ dla zmiennej X_1 oraz 3,72 dla zmiennej X_6 . Z przeprowadzonego testu istotności statystycznej parametrów modelu wynika, że wszystkie istotnie wpływają na zmienną objaśnianą. Oznacza to, że do dalszych testów wykorzystany zostanie model podany we wzorze (11).

W testach, w których niezbędne będzie porządkowanie reszt, zostanie ono przeprowadzone względem zmiennej X_6 . Ma ona bowiem najwyższą wartość statystyki testowej w teście istotności, najmniejsze p -value, a co za tym idzie największy wpływ na zmienną objaśnianą. Pierwszym z testów jest test liczby serii, czyli test losowości. Spośród szesnastu reszt, sześć jest dodatnich, natomiast dziesięć ujemnych. Zaobserwowanych zostało osiem serii, których liczba jest statystyką testową. Z tablic testu serii, dla poziomu istotności $\alpha = 0,05$, odczytane zostały wartości krytyczne, które wyniosły odpowiednio: $K_1 = 5$ oraz $K_2 = 12$. W związku z tym, że liczba serii jest

między dolną i górną wartością krytyczną, należy przyjąć hipotezę zerową o losowości reszt. Bazując na liczbie reszt dodatnich przeprowadzono test symetrii. Na podstawie wzoru (9) wyznaczona została statystyka testowa, która wynosi 1. Wartość krytyczna wyznaczona z rozkładu Studenta na poziomie istotności $\alpha = 0,05$ oraz przy piętnastu stopniach swobody przyjmuje wartość 2,131. Biorąc pod uwagę obie obliczone wielkości, należy przyjąć hipotezę zerową, która mówi o tym, że rozkład reszt jest symetryczny. Podobnie jak losowość, tak i symetria reszt są korzystnymi zjawiskami dla uznania modelu za poprawny. Kolejnym etapem weryfikacji modelu jest sprawdzenie, czy rozkład reszt jest zgodny z rozkładem normalnym. Posłuży do tego test Hellwiga. W tabeli 5 zaprezentowano uporządkowane niemalejąco reszty, reszty standaryzowane, a także wartości dystrybuanty rozkładu normalnego dla wartości reszt standaryzowanych.

Tabela 5. Test Hellwiga – reszty

Reszty e_i	Reszty standaryzowane e'_i	Wartości dystrybuanty $F(e'_i)$
-108,98	-1,30	0,09
-106,53	-1,27	0,10
-91,39	-1,09	0,14
-66,14	-0,79	0,21
-63,97	-0,76	0,22
-53,62	-0,64	0,26
-33,48	-0,39	0,34
-23,07	-0,27	0,39
-5,21	-0,06	0,47
-5,14	-0,06	0,47
30,19	0,36	0,64
52,10	0,62	0,73
75,28	0,89	0,81
76,19	0,91	0,82
156,29	1,87	0,97
167,47	1,99	0,98

Źródło: opracowanie własne.

Obliczone wartości dystrybuanty należało umieścić w celach, czyli przedziałach utworzonych na odcinku od zera do jedynki. Przedziały mają tę samą długość, czyli $\frac{1}{n}$. W przypadku tej analizy, była to $\frac{1}{16}$, czyli 0,0625. Ponadto przyjęto, że cele są lewostronnie domknięte.

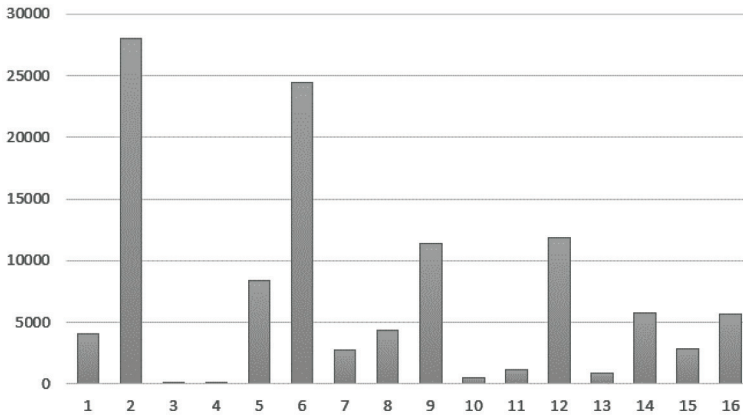
Tabela 6. Test Hellwiga – cele

Cela	Początek	Koniec	Trafienia
1	0	0,0625	0
2	0,0625	0,125	2
3	0,125	0,1875	1
4	0,1875	0,25	2
5	0,25	0,3125	1
6	0,3125	0,375	1
7	0,375	0,4375	1
8	0,4375	0,5	2
9	0,5	0,5625	0
10	0,5625	0,625	0
11	0,625	0,6875	1
12	0,6875	0,75	1
13	0,75	0,8125	0
14	0,8125	0,875	2
15	0,875	0,9375	0
16	0,9375	1	2

Źródło: opracowanie własne.

W tabeli 6 zaprezentowano wartości graniczne cel oraz liczbę trafień, czyli liczbę wartości dystrybuant reszt standaryzowanych, które mieściły się między dolną i górną wartością każdej z cel. Można zauważyć, że liczba cel pustych, która jest zarazem statystyką testową, wynosi w tym przypadku 5. Z tablic testu Hellwiga dla poziomu istotności $\alpha = 0,05$ wartości krytyczne wynoszą: $K_1^* = 3$ oraz $K_2^* = 8$. Statystyka testowa nie jest zatem mniejsza niż K_1^* ani większa niż K_2^* . Ponieważ nie ma podstaw do odrzucenia hipotezy zerowej, można uznać, że reszty mają rozkład normalny. Ostatnim z testów, jaki zostanie przeprowadzony, jest test Goldfelda–Quandt, czyli test homoskedastyczności. Na początku utworzono wykres kwadratów reszt, który może okazać się pomocny przy podzieleniu danych na dwie podpopulacje. Z wykresu na rysunku 2 wynika, że optymalne będzie podzielenie populacji na dwie, równe części.

Zgodnie z tym, co zostało napisane wcześniej, także w przypadku testu homoskedastyczności obserwacje uporządkowano niemalejąco zgodnie z wartościami zmiennej X_1 . Statystyka testowa, czyli iloraz wariancji w obu podpopulacjach, wyniosła 2,683. W przypadku tego testu należy podzielić większą z wariancji przez mniejszą. Wartość krytyczna, wyznaczona z rozkładu Snedecora na poziomie istotności $\alpha = 0,05$ oraz z parą stopni swobody: $n_1 = 5$ oraz $n_2 = 5$, przyjmuje wartość 5,05. Oznacza to, że nie ma podstaw do odrzucenia hipotezy zerowej, mówiącej o homoskedastyczności składnika resztowego. Z punktu widzenia oceny modelu jest to bardzo dobra informacja, gdyż zjawisko heteroskedastyczności jest wysoce niepożądane.



Rysunek 2. Wykres kwadratów reszt

Źródło: opracowanie własne.

Dzięki korzystnym rozstrzygnięciom wszystkich etapów weryfikacji modelu ekonometrycznego można wysnuć kilka najważniejszych wniosków. Co najistotniejsze, dwie zmienne, które utworzyły model, miały istotny wpływ na zmienną objaśnianą, czyli wielkość dochodu rozporządzalnego na osobę. Analizując parametry przy poszczególnych zmiennych, można zauważyć, że gdy liczba osób w gospodarstwie domowym zwiększa się o jedną osobę, dochód rozporządzalny na osobę maleje przeciętnie o niecałe 472 zł, przy założeniu *ceteris paribus*. Podobny wniosek można wysnuć także dla salda miejsc pracy. W przypadku powstania tysiąca nowych miejsc pracy dochód do dyspozycji wzrośnie średnio o 9,9 zł, przy założeniu *ceteris paribus*. W przypadku analizowanego modelu interpretacja wyrazu wolnego jest praktycznie niemożliwa, gdyż trudno mówić o przeciętnym dochodzie rozporządzalnym w gospodarstwie domowym, którego liczba mieszkańców wynosi zero. Kierując się wyłącznie parametrami strukturalnymi modelu, można dojść do wniosku, że najlepsza sytuacja pod względem dochodu do dyspozycji występuje w gospodarstwach domowych o małej liczbie mieszkańców, które znajdują się w rozwijających się województwach. Szybki rozwój jest tu rozumiany jako rosnąca liczba miejsc pracy.

4. Zakończenie

Głównym celem przeprowadzonej analizy było wytypowanie tych kategorii ekonomicznych, które mają największy wpływ na wysokość dochodu rozporządzalnego. Zmienne, które weszły do modelu ekonometrycznego i istotnie wpływały na zmienną objaśnianą, to: średnia liczba osób w gospodarstwie domowym oraz saldo miejsc pracy. Już na tym etapie można stwierdzić, że postawiona we wstępie teza o istotnym wpływie wysokości średniego wynagrodzenia brutto na średni dochód do dys-

pozycji, była nieprawdziwa. Pomimo że została wykazana znacząca korelacja między zmienną opisującą przeciętne wynagrodzenie a zmienną objaśnianą, którą była wysokość dochodu do dyspozycji, zmienna ta nie weszła do modelu. Warto jednak podkreślić, że zarówno średnia liczba osób w gospodarstwie domowym, jak i saldo miejsc pracy są kategoriami, które również w sposób dość oczywisty można powiązać z dochodem.

Z technicznego punktu widzenia celem było stworzenie modelu, który pomyślnie przejdzie weryfikację testami statystycznymi. Tak właśnie się stało, gdyż okazało się, że zmienne objaśniające, które weszły do modelu ekonometrycznego miały istotny wpływ na zmienną objaśnianą. Ponadto testy: losowości, symetrii, normalności oraz homoskedastyczności przyniosły korzystne rezultaty. Wyznaczone współczynniki determinacji i zbieżności potwierdzają poprawność modelu. Potwierdziły się również przypuszczenia o słuszności zastosowania modelu liniowego z wyrazem wolnym. Nie zostały bowiem odnotowane żadne przesłanki, które sygnalizowałyby konieczność wykorzystania modelu nieliniowego w tym badaniu.

We wstępie została przedstawiona teza, że przeciętny miesięczny dochód do dyspozycji może być dobrym wskaźnikiem poziomu życia. Należy tu odseparować obiektywnie uznawany poziom od subiektywnego komfortu życia. W ostatnich latach porzucono bowiem koncepcję, która uznawała osoby bogatsze za szczęśliwsze. Poszukiwane są kategorie, które lepiej oddają sytuację w danym państwie niż średnie wynagrodzenie lub PKB *per capita*. Opracowania, takie jak Human Development Report czy World Happiness Report, koncentrują się na kwestiach uznawanych niegdyś za nieistotne.

Sprawą otwartą pozostają dalsze badania dotyczące dochodu rozporządzalnego. W kolejnych latach możliwe będzie sprawdzenie, czy utrzymają się zależności pomiędzy zmiennymi. Niewykluczone jest, że w następnych okresach większy wpływ na kształtowanie się dochodu do dyspozycji będą miały czynniki, które dotąd okazywały się nieistotne. Z kolei w obszarze analiz dotyczących mierników dobrobytu można się zastanowić, czy dochód do dyspozycji nie byłby lepszym reprezentantem kategorii ekonomicznych w przywoływanych wskaźnikach niż produkt krajowy brutto na osobę, który jest stosowany w obu wspomnianych raportach.

Literatura

- Bałut, M., Pelikan-Krupińska, M. i Świerczyńska, J. (2007). *Podstawy makroekonomii. Ćwiczenia*. Krakowskie Towarzystwo Edukacyjne.
- Begg, D., Fischer, S. i Dornbusch R. (2007). *Mikroekonomia* (wyd. 4). Polskie Wydawnictwo Ekonomiczne.
- Begg, D., Vernasca, G., Fisher, S. i Dornbusch, R. (2014). *Makroekonomia* (wyd. 5 popr.). Polskie Wydawnictwo Ekonomiczne.
- Brain Mastery Coaching (b.d.). *Piramida potrzeb Masłowa*. Pobrane 8.02.2022 z <https://brainmastery-coaching.com/pl/piramida-potrzeb-maslowa/>
- Dziechciarz, J. (red.). (2003). *Ekonometria. Metody, przykłady, zadania* (wyd. 2 popr.). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.

- Gładysz, B. i Mercik, J. (2007). *Modelowanie ekonometryczne. Studium przypadku* (wyd. 2). Oficyna Wydawnicza Politechniki Wrocławskiej.
- Główny Urząd Statystyczny. (2013). *Budżety gospodarstw domowych w 2012 r.* Główny Urząd Statystyczny. Pobrane 2.02.2022 z <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2012-r-,9,7.html>
- Helliwell, J., Layard, R., Sachs, J., De Neve, J-E., Aknin, L. i Wang, S. (2021). *World Happiness Report 2021*. Instytut Gallupa. Pobrane 14.02.2022 z <https://worldhappiness.report/ed/2021/>
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka: Elementy teorii i zadania* (wyd. 6). Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Sompolska-Rzechuła, A. (2013). Jakość życia jako kategoria ekonomiczna. *Oeconomica*, (71), 127–140.
- UNIC Warsaw. (2004). *Raport o rozwoju społecznym Polska 2004: W trosce o pracę*. Ośrodek Informacji ONZ w Warszawie. Pobrane 11.02.2022 z <http://unic.un.org.pl/nhdr/nhdr2004.php>
- United Nations Development Programme. (2020). *Human Development Report 2020*. Nowy Jork: ONZ – United Nations Development Programme. Pobrane 10.02.2022 z <https://hdr.undp.org/en/2020-report/download>

Study of Differentiation of the Average Monthly Disposable Income Per Capita in Voivodships in Poland

Abstract: The amount of disposable income is very important part of everyday life. This amount shows how much an average person can spend or save. This category is much more representative than often used average gross salary. Despite the importance of this issue, it is still not very popular economic category. Main purpose of the analysis is to identify economic phenomena that have the greatest impact on the amount of disposable income. To achieve this aim, an econometric model must be created and verified. In the second chapter, disposable income is presented as proposal for a measure of social well-being. Moreover, indicators other than the traditional gross domestic product per capita will be presented. The situation of Poland will be analysed compared to other European countries in terms of HDR and WHR reports. Among other things, the percentage structure of Poles' spending will also be shown. It will be presented how the amount of disposable income has changed over the years 1999-2020. The third chapter will be devoted to the theoretical aspect of econometric modelling. For the analysis was selected a linear model with an intercept. The variables were selected using the graph method. After checking the significance of the structural parameters of the model, the following tests will be carried out: randomness, symmetry, normality and homoscedasticity. The data analysis is presented in the last chapter. The dependent variable was the amount of disposable income. Out of eleven potential independent variables, two were selected and entered the econometric model. These variables were: the balance of jobs and the average number of people in household. They turned out to have a significant impact on the explained variable. The model with these two variables passed statistical tests. It has been proven that the model explains 70 percent of the variability of the dependent variable. This can be considered a good result. To sum up, it was possible to combine theoretical considerations concerning disposable income and econometric modelling. The analysis showed the phenomena affecting the amount of disposable income, which was the main purpose of the study.

Keywords: disposable income, social well-being, econometric modelling, statistical tests