Zadanie 1

Zakładając liniową relację między wydatkami na obuwie a dochodem oszacować MNK parametry modelu:

$$y_t = \alpha_0 + \alpha_1 X_{1t} + \epsilon_t$$

gdzie:

y_t - roczne wydatki na obuwie mierzone w zł na osobę

X_t - roczny dochód netto w tys. zł na osobę.

Tabela 1. Dane

Lp.	X_t	Y_t
1	6	700
2	8	950
3	10	1100
4	12	1250
5	14	1450
6	16	1700
7	18	1950
8	20	2100

1. Zinterpretuj uzyskane oceny parametrów oraz dokonaj oceny merytorycznej.
2. Wyznacz i zinterpretuj średni błąd równania.
3. Wyznacz i zinterpretuj błędy średnie ocen parametrów
4. Wyznacz i zinterpretuj R^2
5. Wyznacz wektor reszt
6. Zweryfikuj hipotezy dotyczące istotności parametrów;
7. Zweryfikuj hipotezy dotyczące autokorelacji składnika losowego;
8. Oceń model na podstawie uzyskanych wyników

Rozwiązanie

Zakłądamy, że założenia oraz waunki stosowalności MNK są spełnione. Aby oszacować parametry modelu, budujemy macierz obserwacji zmiennej objaśniającej (**X**) oraz zmiennej objaśnianej (**Y**):

$$\boldsymbol{X}=\begin{bmatrix}1 & 6\\ 1 & 8\\ 1 & 10\\ 1 & 12\\ 1 & 14\\ 1 & 16\\ 1 & 18\\ 1 & 20\end{bmatrix} \qquad \boldsymbol{Y}=\begin{bmatrix}700\\ 950\\ 1100\\ 1250\\ 1450\\ 1700\\ 1950\\ 2100\end{bmatrix}$$

jedynki w pierwszej kolumnie macierzy **X** oznaczają specjalną, pomocniczą zmienną różną tożsamościowo jedności, przy której oszacowany parametr będzie w istocie wyrazem wolnym modelu.

Wektor parametrów modelu obliczamy ze wzoru:

$$\boldsymbol{a}=\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\right)^{-1}\cdot\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

Obliczamy poszczególne macierze:

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}=\begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20\end{bmatrix}\cdot\begin{bmatrix}1 & 6\\ 1 & 8\\ 1 & 10\\ 1 & 12\\ 1 & 14\\ 1 & 16\\ 1 & 18\\ 1 & 20\end{bmatrix}=\begin{bmatrix}8 & 104\\ 104 & 1520\end{bmatrix}$$

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}=\begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20\end{bmatrix}\cdot\begin{bmatrix}700\\ 950\\ 1100\\ 1250\\ 1450\\ 1700\\ 1950\\ 2100\end{bmatrix}=\begin{bmatrix}11200\\ 162400\end{bmatrix}$$

Aby zastosować wzór na obliczenia parametrów strukturalnych, należy obliczyć macierz odwrotną do macierzy $\boldsymbol{X}^T\boldsymbol{X}$. W przypadku macierzy 2×2 stosujemy uproszczony schemat obliczeń.

Obliczamy wyznacznik macierzy:

$$det\, \boldsymbol{X}^T \boldsymbol{X} = \begin{vmatrix} 8 & 104 \\ 104 & 1520 \end{vmatrix} = 8 \cdot 1520 - 104^2 = 1344$$

Macierz odwrotną wyznaczamy wg schematu:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det} \cdot \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} = \frac{1}{1344} \cdot \begin{bmatrix} 1520 & -104 \\ -104 & 8 \end{bmatrix}$$

$$\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} = \begin{bmatrix} 1{,}1310 & -0{,}07738 \\ -0{,}07738 & 0{,}005952 \end{bmatrix}$$

Wektor ocen parametrów strukturalnych:

$$\boldsymbol{a} = \frac{1}{1344} \cdot \begin{bmatrix} 1520 & -104 \\ -104 & 8 \end{bmatrix} \cdot \begin{bmatrix} 11200 \\ 162400 \end{bmatrix} = \frac{1}{1344} \cdot \begin{bmatrix} 134400 \\ 134400 \end{bmatrix}$$

$$\boldsymbol{a} = \begin{bmatrix} 100 \\ 100 \end{bmatrix}$$

Zatem

$a_0 = 100$

$a_1 = 100$

Wyestymowana postać modelu jest następująca:

$$y_t = 100 + 100 \cdot X_t$$

ad. 1

Uzyskane oceny parametrów posiadają następującą interpretację:

$a_1 = 100$ - zwiększenie się rocznego dochodu netto o 1 tys. zł na osobę powoduje wzrost rocznych wydatków na obuwie (na osobę) średnio o 100 zł

$a_0 = 100$ - wartości wyrazu wolnego zazwyczaj nie interpretuje się – w tym przypadku można powiedzieć, że gdyby roczny dochód na osobe wynosił 0 zł, to roczne wydatki na obuwie (na osobę) wyniosłyby 100 zł – tylko, że jest to abstrakcyjna interpretacja, gdyż skąd przy zerowych dochodach wziętoby 100 zł na buty.

Aby wyliczyć reszty równania, oraz błąd równania, wyznaczam wartości teoretyczne zmiennej objaśniającej w oparciu o wyliczone równanie regresji. Najlepiej zrobić to korzystając z formuły macierzowej:

$$\hat{\boldsymbol{Y}}=\boldsymbol{X}\cdot\boldsymbol{a}$$

$$\hat{\boldsymbol{Y}}=\begin{bmatrix}1 & 6\\1 & 8\\1 & 10\\1 & 12\\1 & 14\\1 & 16\\1 & 18\\1 & 20\end{bmatrix}\cdot\begin{bmatrix}100\\100\end{bmatrix}=\begin{bmatrix}700\\900\\1100\\1300\\1500\\1700\\1900\\2100\end{bmatrix}$$

ad 2.

Średni błąd równania (odchylenie standardowe reszt) wyznaczamy ze wzoru:

$$S_e=\sqrt{\frac{1}{n-k}\cdot\sum_{t=1}^{n}e_t^2}$$

gdzie $e_t=y_t-\hat{y}_t$ - reszty modelu.

n – liczba obserwacji

k – liczba szacowanych parametrów

W celu wykonania dalszych obliczeń (współczynnika determinacji, statystyki testowej Jarque-Bera) obliczenia przeprowadzone zostaną w tabeli:

Tabela 2 Obliczenia pomocnicze na resztach

	y_t	$\hat{y}_t$	e_t	e_t^2	e_t^3	e_t^4	e_t-e_{t-1}	$(e_t-e_{t-1})^2$	$y_t-\bar{y}$	$(y_t-\bar{y})^2$
	700	700	0	0	0	0			-700	490000
	950	950	50	2500	125000	6250000	50	2500	-450	202500
	1100	1100	0	0	0	0	-50	2500	-300	90000
	1250	1250	-50	2500	-125000	6250000	-50	2500	-150	22500
	1450	1450	-50	2500	-125000	6250000	0	0	50	2500
	1700	1700	0	0	0	0	50	2500	300	90000
	1950	1950	50	2500	125000	6250000	50	2500	550	302500
	2100	2100	0	0	0	0	-50	2500	700	490000
Σ	**11200**		**0**	**10000**	**0**	**25000000**		**15000**		**1690000**

Suma reszt wynosi zero – oznacza to, że model spełnia założenia MNK.

Średni błąd równania (odchylenie standardowe reszt):

$$S_e=\sqrt{\frac{10000}{8-2}}=\sqrt{\frac{10000}{6}}$$

$$S_e=40{,}825$$

Obliczona wartość oznacza, że szacując roczne wydatki na obuwie na osobę na podstawie niniejszego modelu myslimy się w okresie próby średnio o 40,825 zł.

ad 3.

Aby obliczyć srednie błędy ocen parametrów, obliczamy macierz wariancji-kowariancji MNK-estymatora:

$$\boldsymbol{D^2}(\boldsymbol{a})=S_e^2\cdot\left(\boldsymbol{X^T}\boldsymbol{X}\right)^{-1}$$

$$S_e^2=\frac{10000}{6}$$

$$\boldsymbol{D^2}(\boldsymbol{a})=\frac{10000}{6}\cdot\frac{1}{1344}\cdot\begin{bmatrix}1520 & -104\\ -104 & 8\end{bmatrix}$$

$$\boldsymbol{D^2}(\boldsymbol{a})=\begin{bmatrix}1884{,}92 & -128{,}968\\ -128{,}968 & 9{,}9206\end{bmatrix}$$

Średnie błędy ocen parametrów obliczamy jako pierwiastki z diagonalnych elementów powyższej macierzy:

$$S(a_i)=\sqrt{c_{ii}}$$

w naszym przypadku:

$$S(a_0)=\sqrt{1884{,}92}=43{,}416$$

$$S(a_1)=\sqrt{9{,}9206}=3{,}1497$$

powyższe wartości informują nas o ile rzeczywiste wartości poarametrów modelu średnio różnią się od wyestymowanych.

ad 4.

Mając wyznaczone reszty, można obliczyć wartość współczynnika determinacji. Można wyznaczyć go np. ze wzoru:

$$R^2=1-\frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2}$$

$$\bar{y}=\frac{11200}{8}=1400$$

$$R^2=1-\frac{10000}{1690000}$$

$$R^2=0{,}99408$$

wartość ta oznacza, że zmiany rocznych wydatków na obuwie (na osobę) w 99,41% zostały wyjaśnione przez powyższy model (w 99,41% zależą od rocznych dochodów).

ad 5.

Wektor reszt odczytujemy z tabeli 2:

$$\boldsymbol{e}=\begin{bmatrix} 0 \\ 50 \\ 0 \\ -50 \\ -50 \\ 0 \\ 50 \\ 0 \end{bmatrix}$$

ad 6.

Hipotezy o istotności parametrów weryfikujemy w oparciu o statystykę testową:

$$t_i=\frac{a_i}{S(a_i)}$$

Hipotezą zerową jest hipoteza zakładająca nieistotność parametru:

$$H_0\colon a_i=0$$

któa weryfikujemy przeciwko hipotezie alternatywnej:

$$H_1\colon a_i\neq 0$$

odrzucenie hipotezy zerowej na korzyść alternatywnej świadczy o istotności *i*-tego parametru.

Dla parametru a_0 :

$$t_0 = \frac{100}{43{,}416}$$

$$t_0 = 2{,}303$$

Dla parametru a_1 :

$$t_1 = \frac{100}{3{,}1497}$$

$$t_1 = 31{,}749$$

Obliczone wartości statystyki testowej porównujemy z wartością krytyczną rozkładu Studenta o $n-k$ stopniach swobody dla przyjętego poziomu istotności. Zazwyczaj przyjmujemy poziom istotności $a=0{,}05$. Wartość krytyczna rozkładu Studenta o 6 stopniach swobody dla $a=0{,}05$ wynosi:

$$t_\alpha = 2{,}4469$$

Jak widać:

$$t_0 < t_\alpha$$

na poziomie istotności 0,05 brak jest podstaw do odrzucenia hipotezy zerowej. Wyraz wolny nie jest istotny statystycznie.

$$t_1 > t_\alpha$$

na poziomieistotności 0,05 odrzucamy hipotezę zerową na korzyść alternatywnej. Parametr stojący przy zmiennej X jest statystycznie istotny.

Zamiast takiej weryfikacji jak powyżej, można za pomocą dowolnego programu statystycznego (np. MS Excell) wyznaczyć minimalny poziom istotności (wartość p), dl aktórego następuje odrzucenei hipotezy zerowej. Wartości te (wyznaczone w Excelu za pomocą funkcji ROZKŁAD.T) wynoszą:

$$p(t_0) = 0{,}0608$$

$$p(t_1) = 6{,}49 \times 10^{-8}$$

wartości p mniejsze od 0,05 zazwyczaj prtzyjmuje się jako potwierdzające istotność parametrów.

ad 7.

Do weryfikacji hipotezy o autokorelacji składnika losowego wykorzystujemy zazwyczaj statystykę testową Durbina-Watsona:

$$d=\frac{\sum_{i=2}^{n}(e_t-e_{t-1})^2}{\sum_{t=1}^{n}e_t^2}$$

Wartość tę (bądź wartość 4 – *d* w przypadku, gdy *d* > 2) porównujemy z wartością krytyczną statystyki Durbina Watsona odczytaną z tablic. Test Durbina Watsona stosujemy wówczas, gdy reszty mają rozkład normalny. W pierwszej kojelności testuje się więc normalność reszt obliczając statystykę testową Jarque-Bera:

$$JB=n\cdot\left(\frac{1}{6}B_1+\frac{1}{24}(B_2-3)^2\right)$$

gdzie:

$B_1=A^2$ - kwadrat współczynnika asymetrii reszt, przy czym:

$A=\frac{M_3}{s^3}$

$M_3=\frac{1}{n}\cdot\sum_{t=1}^{n}e_t^3$ - trzeci moment centralny – wzór uproszczony z uwagi na zerowanie się sumy reszt – moment centralny tożsamy jest z momentem zwykłym.

$s=\sqrt{\frac{1}{n}\cdot\sum_{t=1}^{n}e_i^2}$ - obciążony estymator odchylenia standardowego reszt (pierwiastek z drugiego momentu centralnego reszt)

$B_2=\frac{M_4}{s^4}$ - kurtoza reszt. Przy czym:

$M_4=\frac{1}{n}\cdot\sum_{t=1}^{n}e_t^4$ - czwarty moment centralny

$$s=\sqrt{\frac{10000}{8}}=\sqrt{1250}=35{,}355$$

$$B_1=A^2=\frac{M_3^2}{s^6}$$

$M_3=0$ (bo suma sześcianów reszt jest równa zero), stąd:

$$B_1=0$$

$$B_2=\frac{M_4}{s^4}$$

$$M_4=\frac{25000000}{8}=3125000$$

$$B_2=\frac{3125000}{(1250)^2}=2$$

$$JB=8\cdot\left(0+\frac{(2-3)^2}{24}\right)$$

$$JB=\frac{1}{3}=0{,}3333$$

Hipotezą zerową jest hipoteza zakładająca normalność reszt, zas alternatywną hipoteza zakładająca, że reszty nie mają rozkładu normalnego. W tym wypadku pozytywnym wynikiem testu jest nie odrzucenie hipotezy zerowej.

Wartość krytyczną porównujemy z wartością krytyczną rozkładu χ^2 o 2 stopniach swobody, dla przyjętego poziomu istotności – najczęściej 0,05. Wartość ta wynosi w tym przypadku:

$$\chi^2_\alpha=5{,}991$$

ponieważ:

$$JB<\chi^2_\alpha$$

zatem brak jhest podstaw do odrzucenia hipoetzy zerowej. Reszty mają rozkłąd normalny.

Wyznaczam wartość statystyki testowej Durbina-Watsona (obliczenia w tabeli 2):

$$d=\frac{15000}{10000}$$

$$d=1{,}5$$

W tablicach Durbina-Watsona dla danej ilości obserwacji (n = 8) oraz ilości zmiennych objaśniających (k = 1) odczytujemy dwie wartości krytyczne:

$$d_L=0{,}763 \qquad d_U=1{,}332$$

ponieważ:

$$d>d_U$$

zatem stwierdzamy brak podstaw do odrzucenia hipotezy o braku autokorelacji reszt. Reszty nie wykazują autokorelacji.

ad. 8

Model jest dobrze dopasowany do danych empirycznych – świadczy o tym bardzo wysoka wartość współczynnika determinacji.

Wartość parametru przy zmiennej objaśniającej jest istotna. Wyraz wolny nie przeszedł wprawdzie testu istotności, lecz akurat istotność wyrazu wolnego nie jest bardzo ważnym kryterium.

Reszty nie wykazują autokorelacji, zatem uzyskane estymatory MNK są zgodne, nieobciążone i efektywne.