

1 Metoda Najmniejszych Kwadratów (*MNK*)

1. Co to jest zmienna endogeniczna, a co to zmienne egzogeniczna?
2. Podaj postać macierzy obserwacji dla modelu $y_t = a + bt + \varepsilon_t$
3. Co to jest wartość dopasowana, podaj wzór wartości dopasowanej?
4. Co to jest reszta? Podaj relację między zmienną zależną, resztą i wartością dopasowaną zmiennej zależnej. Wyjaśnij różnicę między resztą a błędem losowym.
5. Dlaczego zmienne w modelu nie mogą być liniowo zależne?
6. Dlaczego liczba obserwacji musi być w *MNK* większa od ilości zmiennych w modelu?
7. Wyjaśnij dlaczego R^2 nie może służyć do porównywania modeli i podaj alternatywną miarę dopasowania, która można do tego celu użyć.

2 Interpretacja parametrów modelu

1. Kiedy mówimy, że model jest liniowy dla przekształconych zmiennych?
2. Jaka jest interpretacja parametrów w modelu odpowiadających zmiennym ciągłym w modelu liniowym a jaka w modelu logarytmiczno-liniowym?
3. Dlaczego zmienną jakościową powinniśmy przed wstawieniem do modelu rozkodować na zmienne zerojedynkowe?
4. Dlaczego wstawiając do modelu ze stałą zmienne zerojedynkowe, związane z jakościową zmienną objaśniającą, powinniśmy pominąć jeden z poziomów zmiennej jakościowej?

3 Klasyczny Model Regresji Liniowej (*KMRL*)

1. Wymień założenia *KMRL*.
2. Co to znaczy, że estymator jest nieobciążony i dlaczego jest to ważne?
3. Udowodnij, że w *KMRL* estymator b jest nieobciążony.
4. Co to jest prognoza?
5. Podaj dwa źródła błędu prognozy.
6. Podaj (słowami) treść twierdzenia Gaussa-Markowa i wyjaśnij dlaczego jest ono ważne

4 Testowanie hipotez statystycznych

1. Dla pewnej hipotezy statystycznej uzyskaliśmy p -value na poziomie α^* . Przy poziomie istotności α , kiedy hipotezę tę odrzucimy, a kiedy nie będziemy mieli do tego podstaw?
2. Jak tworzymy statystykę, która służy do testowania hipotezę o tym, że $\beta_k = 0$?
3. Jak przetestować hipotezę łączną o tym, że $H\beta = h$ na podstawie sum kwadratów reszt w modelu z ograniczeniami i bez ograniczeń?

5 Podstawowe problemy estymacji za pomocą MNK

1. Jaki skutek może mieć pominięcie istotnej zmiennej w modelu?
2. W jakim szczególnym przypadku możemy uzyskać prawidłowe oszacowania parametrów mimo, że w modelu pominięto istotne zmienne?
3. Dlaczego z modelu powinno się usuwać zmienne nieistotne?
4. Co to jest obserwacja nietypowa i czym różni się od outlier'a?
5. Kiedy można z bazy danych usuwać obserwacje nie pasujące do krzywej regresji, a kiedy nie powinno się tego robić?
6. Kiedy mówimy, że zmienne w modelu są ściśle współliniowe?
7. Dlaczego silna współliniowość zmiennych może być problemem w badaniu statystycznym?
8. Za pomocą jakiej statystyki można wykryć współliniowość w modelu?

6 Testy diagnostyczne

1. Do czego służą testy diagnostyczne?
2. Za pomocą jakiego testu i dlaczego testujemy prawidłowość formy funkcyjnej?
3. Za pomocą jakiego testu i dlaczego testujemy normalność składnika losowego w modelu?
4. Za pomocą jakiego testu i dlaczego testujemy stabilność parametrów i jak go przeprowadzamy?
5. Za pomocą jakiego testów testujemy brak heteroskedastyczności składnika losowego?
6. Za pomocą jakiego testów testujemy brak autokorelacji składnika losowego?.

7 Heteroskedastyczność i autokorelacja

1. Kiedy mówimy o występowaniu heteroskedastyczności w modelu - jakie założenie $KMRL$ nie jest spełnione w przypadku występowania heteroskedastyczności?
2. Kiedy mówimy o występowaniu autokorelacji w modelu - jakie założenie $KMRL$ nie jest spełnione w przypadku występowania autokorelacji?
3. Jaki wpływ ma heteroskedastyczność lub autokorelacja na wnioskowanie statystycznego w MNK ?

8 Porównywanie konkurencyjnych modeli

1. Czym różnią się hipotezy zagnieżdżone od niezagnieżdżonych?
2. Dlaczego wielokrotne testowanie hipotez prostych może doprowadzić nas do wyboru błędnego modelu statystycznego?
3. W jaki sposób wykorzystujemy kryteria informacyjne do porównywania modeli?

9 Modele dynamiczne

1. Podaj ogólną postać modelu $ARIMA(p, d, q)$
2. W jaki sposób generuje się prognozy za pomocą modelu $ARIMA$?
3. Poda ogólną postać modelu ADL .
4. Podaj wzory na mnożniki bezpośredni i długookresowy w modelu ADL i podaj ich interpretację
5. Wyjaśnij w jaki sposób można znaleźć położenie równowagi długookresowej w modelu ADL .

10 Niestacjonarność, kointegracja i przyczynowość w sensie Grangera

1. Kiedy mówimy, że zmienna jest $I(0)$ a kiedy, że jest $I(1)$? Podaj przykład procesu $I(0)$ i $I(1)$.
2. W jakiej sytuacji mówimy o występowaniu kointegracji między zmiennymi?
3. Wyjaśnij kiedy może wystąpić i na czym polega zjawisko regresji pozornej?
4. Opisz procedurę testowania istnienia pierwiastka jednostkowego za pomocą testu ADF
5. Zapisz model dynamiczny w formie mechanizmu korekty błędem (ECM) i wyjaśnij związek między ECM i kointegracją.
6. Podaj interpretację poszczególnych współczynników w mechanizmie korekty błędem (ECM).
7. Opisz procedurę testowania kointegracji za pomocą dwustopniowej procedury Engla-Grangera.
8. Wyjaśnij jak co należy rozumieć przez przyczynowość w sensie Grangera i jak ją testujemy.

11 Zadania

ZADANIE 11.1 Mając obserwacje $\mathbf{x}'_1 = [1, 2, -1, 4, -1]$ i $\mathbf{y}' = [7, 2, 1, 10, -3]$, policzyć estymator MNK parametru \mathbf{b} w modelu $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$. Znaleźć: $\hat{\mathbf{y}}$ oraz \mathbf{e} .

ZADANIE 11.2 Z pewnej regresji otrzymano następujący wynik:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 0 \\ 3 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$$

Policzyć estymator \mathbf{b} bezpośrednio z układu równań normalnych.

ZADANIE 11.3 Rozwiązać poniższe problemy:

1. Poniższa tabela przedstawia część wyników oszacowania KMRL (obliczenia w programie STATA). Ile wynosi współczynnik determinacji liniowej i statystyka testująca istotność całej regresji?

Source	SS	df	MS
Model	61394.8437	2	30697.4219
Residual	8055.3563	17	473.844488
Total	69450.20	19	3655.27368

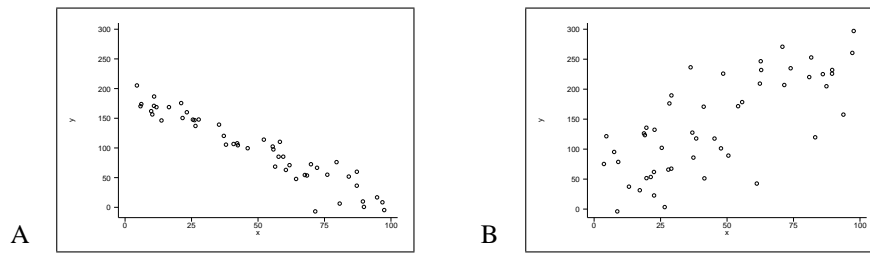
2. Oszacowano następujący KMRL: $y = Xb + e$. Następnie na poziomie istotności 5% weryfikowano hipotezę zerową mówiącą o tym, że $\beta_3 = \beta_4$ i otrzymano następujące wartości (z programu STATA):

$$F(1, 15) = 1.26$$

$$\text{Prob} > F = 0.2789$$

Jaki jest wynik weryfikacji tej hipotezy?

3. Poniższe rysunki przedstawiają wykresy rozrzutu dla zmiennych x i y . Co można na ich podstawie powiedzieć o współczynnikach determinacji liniowej i o znakach oszacowanych współczynników modeli oszacowanych w oparciu o zilustrowane dane?



4. Poniższa tabela przedstawia część wyników oszacowania KMRL (obliczenia w programie STATA). Ile współczynników regresji jest istotnie różny od zera (na poziomie 5%)?

y	Coef.	Std. Err.	t	P> t
x2	-10.97064	5.274215	-2.08	0.064
x3	1.46263	.1264963	11.56	0.000
x4	15.23092	.6048963	25.18	0.000
x5	2.605007	3.280705	0.79	0.446
_cons	20991.19	9000.29	2.33	0.042

5. Ile wynosi współczynnik determinacji liniowej w KMRL z jedną zmienną objaśniającą i stałą, jeżeli nieobciążony estymator kowariancji zmiennych objaśniającej i objaśnianej w próbie wynosi -6, a nieobciążone estymatory odchyłeń standardowych zmiennych objaśniającej i objaśnianej wynoszą odpowiednio 3 i 4?

6. Oszacowano macierz wariancji-kowariancji dla KMRL:

$$\text{Var}[\mathbf{b}] = \begin{bmatrix} 0.04 & & & & \\ -0.72 & 1.21 & & & \\ -0.45 & 0.59 & 6.25 & & \\ 2.55 & -7.68 & -6.87 & 0.09 & \end{bmatrix}.$$

Który ze współczynników jest statystycznie nieistotny, jeżeli wiadomo, że jest tylko jeden taki współczynnik, a $\mathbf{b}' = [2, -1.1, 5, -1.5]$?

ZADANIE 11.4 Dany są następujące wyniki regresji:

Source	SS	df	MS	Number of obs =	94
Model	2277280.75	2	1138640.37	F(2, 91) =	5.31
Residual	19501750.4	91	214304.949	Prob > F =	0.0066
Total	21779031.1	93	234183.13	R-squared =	0.1046
				Adj R-squared =	0.0849
				Root MSE =	462.93

netpay	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

wiek		11.91639	4.126379	2.89	0.005	3.719845	20.11294
sex		-164.1205	97.10046	-1.69	0.094	-356.9986	28.75764
_cons		841.5447	210.304	4.00	0.000	423.8017	1259.288

w których *netpay* to płaca netto, *wiek* jest wiekiem respondenta, a *sex* przyjmuje wartość 0 dla mężczyzn i 1 dla kobiet.

1. Jaka jest interpretacja współczynników w tym modelu?
2. Które zmienne w modelu są istotne?
3. Czy łącznie wszystkie zmienne w modelu są istotne?
4. Jakie jest odchylenie standardowe błędu losowego w tym modelu?
5. Jaka część zmienności płac została wyjaśniona przez model?

Hipotezy testować na poziomie istotności $\alpha = 0.05$. Wskazać statystyki użyte do testowania.

ZADANIE 11.5 Na podstawie danych GUS zbudowano KMRL wyjaśniający poziom wydatków na żywność za pomocą dochodu na głowę wyrażonego w złotych przeliczonego według skali oksfordzkiej, liczby osób dorosłych i dzieci w rodzinie oraz zmiennej zero-jedynkowej przyjmującej wartość 1 dla gospodarstwa domowego wiejskiego. Szacowano następujący model:

$$\text{wydatki}_i = \text{stala} + \beta_1 \text{dochod}_i + \beta_2 \text{dorosle}_i + \beta_3 \text{dzieci}_i + \beta_4 \text{wies}_i + \varepsilon_i$$

i otrzymano wyniki:

Source	SS	df	MS	Number of obs =	29771
Model	828538978	4	207134745	F(4, 29766) =	.
Residual	956257715	29766	32125.8387	Prob > F =	.
Total	1.7848e+09	29770	59952.8617	R-squared =	0.4642
				Adj R-squared =	0.4641
				Root MSE =	179.24

wydatki	Coef.	Std. Err.	t
dochod	.1289331	.0021889	.
dorosle	135.2956	1.100277	.
dzieci	104.175	.6937367	.
wies	28.26728	2.287173	.
_cons	47.61749	3.397637	.

1. Uzupelnic brakujace wielkości w tabeli.
2. Zinterpretowac otrzymane wyniki.
3. Przeprowadzic testy istotności i łącznej istotności ($F_{0.95}(4, \infty) = 2, 37$).
4. Ocenic dopasowanie modelu do danych empirycznych.

5. Chcemy przetestować hipotezę, czy tylko liczba osób w gospodarstwie domowym ma wpływ na wydatki żywnościowe a nie to, czy są to dorośli, czy dzieci. Zapisać tę hipotezę zerową oraz hipotezę alternatywną, a także wyjaśnić w jaki sposób można ją zweryfikować.

Przyjąć poziom istotności $\alpha = 0.05$.

ZADANIE 11.6 Na podstawie danych pochodzących ze strony www.economagic.com zbudowano KMRL wyjaśniający kurs USD w stosunku do Euro za pomocą podaży pieniądza w USA i indeksu cen dóbr konsumpcyjnych w USA (inflacji): $kurs = \text{stala} + \beta_1 \text{podaz} + \beta_2 \text{cpi} + \varepsilon$. Model bazuje na 108 obserwacjach od stycznia 1990 do stycznia 1999. Otrzymano następujące wyniki:

Source	SS	df	MS	Number of obs = 108		
Model	.136566483	2	.068283242	F(2, 105) =	13.06	Prob > F = 0.0000
Residual	.548871572	105	.005227348	R-squared =	0.1992	Adj R-squared = 0.1840
Total	.685438055	107	.006405963	Root MSE =	.0723	

kurs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
podaz	.0000863	.0000958	0.90	0.370	-.0001037	.0002764
cpi	-.0040808	.0010651	-3.83	0.000	-.0061927	-.001969
_cons	1.73358	.1012719	17.12	0.000	1.532776	1.934383

Durbin-Watson d-statistic(3, 108) = .1814083

Breusch-Pagan chi2(1) = 1.61 Prob > chi2 = 0.2049

Ramsey RESET F(3, 102) = 15.09 Prob > F = 0.0000

Jarque-Bera chi2(2) = 6.91 Prob > chi2 = 0.0315

oraz macierz wariancji-kowariancji estymatorów:

	podaz	cpi	_cons
podaz	9.2e-09		
cpi	-8.0e-08	1.1e-06	
_cons	2.4e-06	-.000086	.010256

Na podstawie powyższych wyników odpowiedzieć na następujące pytania. Każdą odpowiedź uzasadnić wielkościami odpowiednich statystyk testowych. Testy przeprowadzić na poziomie istotności $\alpha = 0.05$.

1. Zinterpretować parametry modelu.
2. Sprawdzić istotność oraz łączną istotność uzyskanych wyników.
3. Sprawdzić, czy w modelu występuje autokorelacja?
4. Sprawdzić, czy w modelu występuje heteroscedastyczność?
5. Sprawdzić, czy składnik losowy ma rozkład normalny?
6. Sprawdzić poprawność funkcyjną modelu.

7. Zapisać regresję pomocniczą, którą należy przeprowadzić by zbadać występowanie procesu autoregresyjnego rzędu 3. Jak nazywa się ten test?
8. Korzystając z wyników poprzednich podpunktów napisać jakiej procedury estymacyjnej należy użyć by otrzymać zgodne estymatory wariancji?

ZADANIE 11.7 Oszacowano regresję logarytmu zagregowanej konsumpcji w wyrażeniu *nominalnym* (zmienna y) na logarytmie PKB w wyrażeniu *nominalnym* (zmienna x_1) i logarytmie deflatora (zmienna x_2). Regresję przeprowadzono dla danych kwartalnych dla Polski z okresu 1995-2002. Wprowadzono także do modelu zmienne sezonowe przyjmujące wartość 1 dla odpowiedniego kwartału a 0 dla pozostałych. Wyniki regresji znajdują się w tabeli poniżej.

y	Coef.	Std.Err.	t	P> t
x1	.80	.081	9.72	0.000
x2	-.59	.117	-5.00	0.000
seasonal2	-.05	.011	-4.38	0.000
seasonal3	-.07	.011	-6.25	0.000
seasonal4	-.17	.012	-14.36	0.000
cons	1.87	.926	2.02	0.054

Number of obs = 32, F(5, 26) = 229.88 [0.0000]

R-squared = 0.9779, s = .02027

Durbin-Watson test statistic: d(6, 32) =2.261303
 Breusch-Godfrey LM statistic: Chi-sq(2) =9.894179 [.0071]
 White's general test statistic: Chi-sq(14) =25.15268 [.0331]
 Jarque-Berra test statistic: Chi-sq(2) =3.34 [.1886]
 Ramsey RESET test statistic: F(3, 23) =1.66 [.2030]
 Chow test statistic (t=1999.1): F(3, 23) =0.28 [.8359]

Przy założonym poziomie istotności $\alpha = 0.01$ przeprowadzić analizę wyników. Każdą z odpowiedzi należy uzasadnić za pomocą odpowiedniego testu.

1. Określić, czy model jest dobrze dopasowany, czy zbiór zmiennych niezależnych istotnie objaśnia zmienną zależną?
2. Podać, które zmienne w modelu są istotne?
3. Zbadać, czy w modelu występuje autokorelacja?
4. Zbadać, czy w modelu występuje heteroskedastyczność?
5. Sprawdzić, czy forma funkcyjna modelu jest prawidłowa?
6. Przetestować, czy błąd losowy w modelu ma rozkład normalny?
7. Sprawdzić, czy parametry modelu są stabilne?
8. Zinterpretować współczynnik przy zmiennej x_1 .
9. W jaki sposób należałoby weryfikować hipotezę o neutralności pieniądza w gospodarce (tzn., że ważne są jedynie wartości realne a nie nominalne)?

10. W jaki sposób należałoby weryfikować hipotezę mówiącą o tym, że zmienne sezonowe są nieistotne w modelu?

11. Jeśli model nie spełnia założeń KMRL określić:

- które założenie nie jest spełnione?
- jakie ma to konsekwencje dla wnioskowania statystycznego?
- jakie są metody radzenia sobie z tym problemem?

ZADANIE 11.8 Dla modelu *ADL* postaci dla zależności zmian bezrobocia w/g bael od inflacji dla Polski w latach 1994.1 – 2003.4 otrzymaliśmy następujące oszacowanie parametrów (przy liczeniu zaokrąglaj do setnych)

Source	SS	df	MS	Number of obs = 38			
Model	399.156095	3	133.052032	F(3, 34)	=	142.94	
Residual	31.6488362	34	.930848124	Prob > F	=	0.0000	
				R-squared	=	0.9265	
				Adj R-squared	=	0.9201	
Total	430.804931	37	11.6433765	Root MSE	=	.9648	

bael	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
bael						
L1	.8823577	.1672545	5.28	0.000	.5424556	1.22226
L2	.0364489	.1618563	0.23	0.823	-.2924825	.3653804
inf	-.0395884	.0202393	-1.96	0.059	-.0807196	.0015429
_cons	1.83276	1.012992	1.81	0.079	-.2258878	3.891407

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
4	19.272	4	0.0007

- policz i zinterpretuj mnożnik bezpośredni
- policz i zinterpretuj mnożnik długookresowy
- Jaka będzie wartość oczekiwana bezrobocia jeśli:
 - inflacja będzie stale równa 0%
 - inflacja będzie stale równa 40%

Co implikują te wartości w odniesieniu do polityki gospodarczej?

- co implikuje wynik testu Breuscha-Godfrey'a jeśli chodzi własności estymatorów w tej regresji?

ZADANIE 11.9 Estymacja modelu $AR(2)$ na pierwszych różnicach dla próby 100 obserwacji dała następujący wynik (w nawiasach błędy standardowe):

$$\Delta y_t = -0,18 - 0,14 y_{t-1} + 0,24 \Delta y_{t-1} + \epsilon_t$$

(0.12) (0.05) (0.10)

Przetestować na poziomie istotności $\alpha = 0,05$ hipotezę o pierwiastku jednostkowym. Podać postaci hipotezy zerowej i alternatywnej.

ZADANIE 11.10 Dany jest model:

$$\Delta y_t = \alpha(y_{t-1} - \beta x_{t-1}) + \delta \Delta y_{t-1} + \epsilon_t$$

i wyniki następujących regresji dla 100 obserwacji (poziom istotności $\alpha = 0,05$):

$$\Delta y_t = 0,4 y_{t-1} + 0,2 \Delta y_{t-1},$$

(0.2) (0.02)

$$\Delta x_t = -0,8 x_{t-1} + 0,2 \Delta x_{t-1},$$

(0.5) (0.1)

$$y_t = 0,8 + 3,2 x_t.$$

(0.1) (0.1)

1. Czy w tym przypadku ma sens (i dlaczego) testowanie kointegracji pomiędzy x_t i y_t ?
2. Reszty MNK \hat{u}_t z regresji y_t na x_t zostały użyte w regresji $\Delta \hat{u}_t$ na \hat{u}_{t-1} , która dała następujący wynik (błędy standardowe w nawiasach):

$$\Delta \hat{u}_t = -0,8 \hat{u}_{t-1}.$$

(0.2)

Jaki jest wynik testu na kointegrację? Jaką postać ma ewentualny wektor kointegrujący?

3. Przeprowadzono regresję Δy_t na \hat{u}_{t-1} i Δy_{t-1} i otrzymano

$$\Delta y_t = -0,4 \hat{u}_{t-1} + 0,3 \Delta y_{t-1}$$

(0.1) (0.1)

Jaki model został ten sposób oszacowany? Zinterpretuj parametr stojący przed \hat{u}_{t-1} .

Rozwiązania zadań

Zadanie 11.1 W modelu występuje stała, więc macierz: X

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \\ 1 & 4 \\ 1 & -1 \end{bmatrix}.$$

Podstawiając do wzorów $b = (X'X)^{-1}X'y$, $\hat{y} = Xb$, $e = y - \hat{y}$ otrzymuje się:

$$b = \begin{bmatrix} 1.4 \\ 2 \end{bmatrix}, \hat{y} = \begin{bmatrix} 3.4 \\ 5.4 \\ -0.6 \\ 9.4 \\ -0.6 \end{bmatrix}, e = \begin{bmatrix} 3.6 \\ -3.4 \\ 1.6 \\ 0.6 \\ -2.4 \end{bmatrix}$$

Zadanie 11.2 Układ równań normalnych jest następującej postaci:

$$X'Xb = X'y,$$

co można zapisać jako:

$$\begin{aligned} 6b_1 + 3b_2 + 0b_3 &= 2 \\ 3b_1 + 4b_2 + 2b_3 &= 4 \\ 0b_1 + 2b_2 + 1b_3 &= 2 \end{aligned}.$$

Rozwiązaniem układu równań jest wektor $[0, \frac{2}{3}, \frac{2}{3}]$.

Zadanie 11.3

1.

$$\begin{aligned} R^2 &\approx 88\%, \\ F &\approx 65. \end{aligned}$$

2. Nie ma podstaw do odrzucenia hipotezy zerowej. Współczynniki regresji nie różnią się od siebie istotnie.

3.

$$\begin{aligned} R_A^2 &> R_B^2, \\ \hat{\beta}_A &< 0, \\ \hat{\beta}_B &> 0. \end{aligned}$$

4. O istotności współczynników regresji świadczą wartości p -value znajdujące się w kolumnie $P > |t|$. Przy danym poziomie istotności wartości te powinny być mniejsze od 5%. Jest tak w przypadku trzech zmiennych: stałej, x_3 , x_4 .

5.

$$R^2 = 25\%.$$

6. Na podstawie macierzy wariancji-kowariancji i wektora b' można obliczyć:

$$t = [10, -1, 2, -5].$$

Najmniejszą, co do wartości bezwzględnej, wartością statystyki t wyróżnia się zmienna x_2 .

Zadanie 11.4

1. Współczynnik przy zmiennej *wiek* wskazuje o ile przeciętnie zwiększa się płaca netto z każdym rokiem życia. Współczynnik przy zmiennej *sex* wskazuje o ile mniej przeciętnie niższa jest płaca netto kobiet.
2. Na poziomie 5% istotna w modelu jest stała ($t = 4.00$, $pvalue=0.000$) oraz zmienna *wiek* ($t = 2.89$, $pvalue=0.005$).
3. Na poziomie 5% cała regresja jest istotna ($F = 5.31$, $pvalue=0.0066$).
4. $s \approx 462.93$.
5. Niecałe 10.5% ($R^2 = 0.1046$) zmienności płac zostało wyjaśnione przez model.

Zadanie 11.5

1. Statystyki t oblicza się ze wzoru $t = \frac{\beta}{se(\beta)}$. Statystykę F oblicza się ze wzoru $F = \frac{(S_R - S)/g}{S/(N-K)}$. W modelu w którym występuje tylko stała $b = \bar{y}$ a $S_R = (\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}}) = TSS$. Zatem $F = \frac{(TSS - RSS)/(K-1)}{RSS/(N-K)} = \frac{N-K}{K-1} \frac{ESS}{RSS}$

Source	SS	df	MS	Number of obs =	29771
Model	828538978	4	207134745	F(4, 29766) =	6448.17
Residual	956257715	29766	32125.8387	Prob > F	= 0.0000
Total	1.7848e+09	29770	59952.8617	R-squared	= 0.4642
				Adj R-squared	= 0.4641
				Root MSE	= 179.24

wydatki	Coef.	Std. Err.	t
dochod	.1289331	.0021889	58.90
dorosle	135.2956	1.100277	122.97
dzieci	104.175	.6937367	150.17
wies	28.26728	2.287173	12.36
_cons	47.61749	3.397637	14.01

2. Wzrost dochodu o złotówkę powoduje przeciętny wzrost wydatków na żywność o 12 groszy. Dodatkowa osoba dorosła w gospodarstwie domowym przeciętnie zwiększa wydatki na żywność o 135.30 zł. Dodatkowe dziecko zwiększa wydatki na żywność o 104.18 zł. Gospodarstwa domowe na wsi przeciętnie wydają na żywność o 28.27 zł więcej niż pozostałe.
3. Wszystkie zmienne są indywidualnie istotne, ponieważ wszystkie statystyki t są co do wartości bezwzględnej większe od 2, zmienne są również łącznie istotne bowiem wartość statystyki testowej jest znacznie większa od wartości krytycznej równej 2.37.

4. Model jest w 46% dopasowany do danych empirycznych.
5. Liczba osób w gospodarstwie domowym to suma dorosłych oraz dzieci. Zatem hipotezy dotycząca braku wpływu liczby osób na wydatki żywnościowe wygląda następująco: $H_0 : \beta_2 = \beta_3$ przy hipotezie alternatywnej $H_0 : \beta_2 \neq \beta_3$. Szacowane są dwa modele regresji, jeden z narzuconymi ograniczeniami (w modelu tym zmienną objaśniającą jest zmienna dzieci+dorośle) i drugi bez ograniczeń i za pomocą testu F sprawdza się czy różnica sum kwadratów reszt w obu modelach jest statystycznie istotna.

Zadanie 11.6

1. Wzrost podaży pieniądza o jednostkę powoduje wzrost kursu dolara w stosunku do euro o 0.00008. Wzrost wskaźnika cen konsumpcyjnych o jednostkę powoduje spadek kursu o 0.004.
2. Statystyka $F(3, 105) = 13.06$ oraz jej $p\text{-value}=0$ wskazuje na łączną istotność zmiennych w modelu. Pojedynczo istotna jest *stala*, ponieważ statystyka $|t_{stala}| = 17.12 > 1,96$ oraz *cpi* ponieważ $|t_{CPI}| = 3,83 > 1,96$, natomiast nieistotna jest *podaz* ponieważ $p\text{-value}$ statystyki t dla tej zmiennej jest większe od 0,05.
3. W modelu występuje autokorelacja, ponieważ statystyka Durбина-Watsona jest bliska 0. Dolna wartość krytyczna odczytana z tablic dla 2 zmiennych i 100 obserwacji wynosi 1.63. Wobec tego $DW < d_L$.
4. Na podstawie testu Breuscha-Pagana nie ma podstaw do odrzucenia hipotezy o homoscedastyczności ponieważ $p\text{-value}$ statystyki testowej wynosi 0.20 i jest większe od 0.05.
5. Na podstawie statystyki Jarque-Berra równej 6,91 odrzucamy hipotezę zerową o normalności składnika losowego (por. $p\text{-value}$).
6. Na podstawie testu RESET $F(3, 105) = 15,09$, widzimy że statystyka testowa jest duża, więc odrzucamy hipotezę zerową o poprawności formy funkcyjnej modelu.
7. Test Breuscha-Godfrey'a. $e_t = \gamma_0 + \gamma_1 e_{t-1} + \gamma_2 e_{t-2} + \gamma_3 e_{t-3} + \xi_t$.
8. Ponieważ w modelu występuje autokorelacja pierwszego rzędu i nie ma heteroscedastyczności należałoby sprawdzić występowanie autokorelacji wyższych rzędów. Jeśli jej nie ma, to można zastosować procedurę Cochran-Orcutt lub procedurę Newey'a-Westa, natomiast w przypadku występowania autokorelacji wyższego rzędu jedynie procedura Newey'a-Westa da poprawne wyniki.

Zadanie 11.7

1. Oszacowany model objaśnia prawie 98% zmienności zmiennej zależnej (współczynnik $R^2 = 0.9779$). Zbiór zmiennych objaśniających (bez stałej) jest łącznie istotny ($F_{5,26} = 229.88$) [$0.0000 < 0.01$].
2. Wszystkie zmienne objaśniające poza stałą są istotne na zadanym poziomie istotności ($p\text{-value}$ dla statystyk t są mniejsze od 0.01, nie dotyczy stałej).
3. W modelu nie występuje autokorelacja I rzędu ponieważ statystyka $DW \approx 2.26 \in (1.597, 2.403)$ przyjmuje wartość z przedziału przyjęć hipotezy zerowej. Odpowiednie wartości krytyczne dla tego testu wynoszą ($d_L = 0.917, d_U = 1.597$). W modelu może jednak występować autokorelacja wyższego rzędu, o czym świadczy $p\text{-value}$ dla testu Breuscha-Godfrey'a [$.0071 < 0.01$].
4. Nie ma podstaw do odrzucenia hipotezy o homoscedastyczności składnika losowego w związku z wartością $p\text{-value}$ dla testu White'a [$.0331 > 0.01$].

5. Forma specyfikacji modelu jest prawidłowa, o czym świadczy p -value dla testu RESET [.2030 > 0.01].
6. Błąd losowy ma rozkład normalny, o czym świadczy p -value dla testu Jarque'a-Berra [.1886 > 0.01].
7. Parametry modelu są stabilne, o czym świadczy p -value dla testu Chowa [.8359 > 0.01].
8. Współczynnik przy zmiennej x_1 jest elastycznością zagregowanej konsumpcji względem PKB, zatem wzrost PKB o 1% spowoduje wzrost zagregowanej konsumpcji o 0.8%.
9. Model można zapisać jako

$$\log(\text{cons}) = \alpha_1 + \alpha_2 \log(\text{pkb}) + \alpha_3 \log(\text{defl}) + \varepsilon$$

Konsumpcja i PKB w wyrażeniu realnym są równe odpowiednio $\text{cons}^* = \frac{\text{cons}}{\text{defl}}$, $\text{pkb}^* = \frac{\text{pkb}}{\text{defl}}$. Odejmując od obu stron $\log(\text{defl})$ uzyskujemy

$$\log(\text{cons}^*) = \alpha_1 + \alpha_2 \log(\text{pkb}^*) + (\alpha_2 + \alpha_3 - 1) \log(\text{defl}) + \varepsilon$$

a więc na konsumpcję w wyrażeniu realnym wpływa jedynie dochód w wyrażeniu realnym jeśli prawdziwe jest $H_0 : \alpha_2 + \alpha_3 = 1$.

10. Hipotezę o łącznej nieistotności zmiennych sezonowych należy testować za pomocą testu F , nakładając łączne ograniczenie na parametry przy zmiennych $\text{seasonal}_2, \text{seasonal}_3, \text{seasonal}_4$, - przyrównując je jednocześnie do zera.
11. (a) Model nie spełnia założenia braku autokorelacji składnika losowego.
 (b) Macierz wariancji-kowariancji jest niewłaściwa, co powoduje, że niewłaściwe są błędy standardowe parametrów i tym samym niewłaściwe są statystyki t .
 (c) Jeśli nie chcemy usunąć autokorelacji, należy posłużyć się tzw. odporną macierzą wariancji-kowariancji Newey'a-Westa. Można też podjąć próbę usunięcia autokorelacji np. za pomocą metody Prais'a-Winstena.

Zadanie 11.8

1. Mnożnik bezpośredni równy jest -0.04 (zaokrąglamy do setnych). Jednoprocentowy wzrost inflacji prowadzi do natychmiastowego spadku bezrobocia o -0.04%
2. Mnożnik długookresowy liczymy ze wzoru $\beta = \frac{\beta_0}{1 - a_1 - a_2} = \frac{-0.04}{1 - 0.88 - 0.04} = -0.5$. W długim okresie wzrost inflacji o 1% spowoduje spadek bezrobocia o 0.5%
3. Jeśli założymy, że wartość inflacji jest stała to można ją traktować w modelu tak jak stałą. W rezultacie rozwiązanie długookresowe ma postać

$$E(y_t) = \frac{\mu + \beta_0 x_t^*}{1 - a_1 - a_2}$$

dla inflacji równej zero otrzymujemy

$$E(y_t) = \frac{1.83}{1 - 0.88 - 0.04} = 22.88$$

przy inflacji równej 40%

$$E(y_t) = \frac{1.83 - 40 \times 0.04}{1 - 0.88 - 0.04} = 2.88$$

Sprowadzenie bezrobocia do niskiego poziomu (3%) wymagałoby ogromnego zwiększenia inflacji (do 40%)

4. Z testu wynika, że błąd losowy jest skorelowany. W modelu występują opóźnione zmienne zależne. W rezultacie wystąpi problem równoczesności, co implikuje, że oszacowanie parametrów nie będą zgodne. W związku z tym prezentowane wyniki regresji są prawdopodobnie nieprawidłowe.

Zadanie 11.9

Ogólną postać modelu:

$$\Delta y_t = \alpha + \beta y_{t-1} + \gamma \Delta y_{t-1} + \epsilon_t.$$

Należy zweryfikować hipotezę:

$$H_0 : \beta = 0,$$

$$H_1 : \beta < 0.$$

Wartość statystyki testowej wynosi $t = \frac{-0.14}{0.05} = -2.8$. Wartość krytyczną odczytana z tablic testu Dickey'a-Fullera wynosi $t_{DF}(100) = -2.9$. Ponieważ $t > t_{DF}$ brak jest podstaw do odrzucania hipotezy zerowej o istnieniu pierwiastka jednostkowego.

Zadanie 11.10

1. Wartość krytyczna testu Dickey'a-Fullera $t_{DF}(100) = -2,9$.

$$t_y = \frac{-0.4}{0.2} = -2.$$

Ponieważ statystyka testowa $t_y = -2 > t_{DF}$ należy uznać, że istnieje pierwiastek jednostkowy, czyli szereg y_t jest zintegrowany stopnia jeden (I(1)).

$$t_x = \frac{-0.8}{0.5} = -1.6.$$

Ponieważ statystyka testowa $t_x = -1.6 > t_{DF}$ należy uznać, że istnieje pierwiastek jednostkowy, czyli szereg y_t jest zintegrowany stopnia jeden (I(1)). Ponieważ obydwa szeregi są I(1), testowanie kointegracji ma sens, bowiem model może wskazywać na związek pomiędzy trendami zawartymi w zmiennych, a nie na relację pomiędzy zmiennymi x_t i y_t .

- 2.

$$t = \frac{-0.8}{0.2} = -4.$$

Ponieważ statystyka testowa $t_x = -4 < t_{DF}$ odrzucamy hipotezę zerową o *braku kointegracji* należy uznać, że zmienne są skointegrowane. Postać wyestymowanego wektora kointegrującego: $a[1, -\beta]' = a[1, -0.8, -3.2]$ dla dowolnego a .

3. Oszacowano model korekty błędem ECM. Wielkość -0.4 jest związana z szybkością korygowania błędów. 40% odchylenia od relacji długookresowej jest korygowana w ciągu roku.