

# Sprawozdanie 1

Analiza szeregów czasowych zadanie spr1

Kacper Szmigielski (282255)

## Spis treści

<b>1 Zadanie 3</b>	<b>4</b>
1.1 Ilustracja działania testów białoszumowości . . . . .	4
1.1.1 Zilustrowanie działania testów dla danych ( $n = 200$ ) typu biały szum (rozkład $WN(0, \sigma^2)$ ) . . . . .	4
1.1.2 Zilustrowanie działania testów dla danych ( $n = 200$ ) typu (rozkład $Exp(2)$ ) . . . . .	5
1.1.3 Zilustrowanie działania testów dla danych ( $n = 200$ ) z autokorelacją (nie biały szum) . . . . .	6
1.2 Porównywanie testów graficznych oraz formalnych dla testowania białoszumowości . . . . .	7
1.2.1 Porównanie testów pod względem liczby powtórzeń . . . . .	7
1.2.2 Porównanie pod względem różnych długości testów . . . . .	8
1.2.3 Porównanie pod względem różnego maksymalnego opóźnienia $h$ . . . . .	9
1.2.4 Przykłady szeregów czasowych innych niż szum IID . . . . .	10
<b>2 Zadanie 4</b>	<b>14</b>
2.1 Wybór danych . . . . .	14
2.2 Zastosowanie poznanych metod graficznych, do podstawowej analizy wykresu	14
2.2.1 Wykres Liniowy . . . . .	14
2.2.2 Wykres sezonowy . . . . .	16
2.2.3 Wykres kwartalny . . . . .	17
2.2.4 Wykres ACF . . . . .	18
2.2.5 Wykres boxplot . . . . .	19
2.2.6 Wykresy rozrzutu . . . . .	20

2.2.7	Wstępne wnioski po analizie podstawowych wykresów . . . . .	20
2.3	Dekompozycje . . . . .	21
2.3.1	Dekompozycja addytywna . . . . .	21
2.3.2	Dekompozycja multiplikatywna . . . . .	24
2.3.3	Dekompozycja na podstawie modelu regresji . . . . .	27
2.3.4	Dekompozycja oparta na metodzie STL . . . . .	31
2.4	Porównanie zależności wyników od parametrów . . . . .	32
2.4.1	Porównanie modeli decompose(addytywny i multiplikatywny) . . . . .	32
2.4.2	Porównanie modeli tslm (trend jako wielomian różnych stopni) . . . . .	33
2.4.3	Porównanie modeli (stl dla różnych modeli parametrycznych trendu, parametry wygładzające) . . . . .	34
2.5	Próba rozstrzygnięcia czy zastosowanie transformacji Boxa-Coxa prowadzi do poprawy jakości dopasowania modeli dekompozycji . . . . .	35
2.5.1	Dla decompose (używamy transformacji Boxa-Coxa z lambda = 0) . . . . .	35
2.5.2	Dla tslm (używamy transformacji Boxa-Coxa z lambda = 0) . . . . .	37
2.5.3	Dla stl (używamy transformacji Boxa-Coxa z lambda = 0) . . . . .	39
2.6	Porównanie wyników eliminacji trendu i sezonowości na podstawie rozważanych metod dekompozycji z wynikami uzyskanymi poprzez odpowiednie różnicowanie danych . . . . .	40
2.6.1	Dane zróżnicowane co 4 . . . . .	40
2.6.2	Reszty z dekompozycji addytywnej z transformacją boxa-coxa . . . . .	42
2.6.3	Reszty z regresji wielomianem 5 stopnia i transformacją boxa-coxa . . . . .	43
2.6.4	Reszty z stl plus Box-Cox . . . . .	44

## Spis rysunków

1	Wykres testu graficznego dla procesu białego szumu . . . . .	4
2	Wykres testu graficznego dla $\text{Exp}(2)$ . . . . .	5
3	Wykres testu graficznego dla danych nieskorelowanych . . . . .	6
4	Wykres testu graficznego dla modelu stacjonarnego z niewielką korelacją . . . . .	11
5	Wykres testu graficznego dla modelu błędzenia losowego . . . . .	12
6	Wykres testu graficznego dla modelu szumu iid z dryfem i sezonowością . . . . .	13
7	Liniowy wykres danych . . . . .	14

8	Sezonowy wykres danych . . . . .	16
9	Kwartalny wykres danych . . . . .	17
10	ACF wykres danych . . . . .	18
11	Boxplot wykres danych . . . . .	19
12	Wykres rozrzutu danych . . . . .	20
13	Dekompozycja addytywna danych . . . . .	21
14	Acf dla dekompozycji addytywnej danych . . . . .	22
15	Boxplot dla dekompozycji addytywnej danych . . . . .	23
16	Autoplot dla dekompozycji multiplikatywnej danych . . . . .	24
17	Acf dla dekompozycji multiplikatywnej danych . . . . .	25

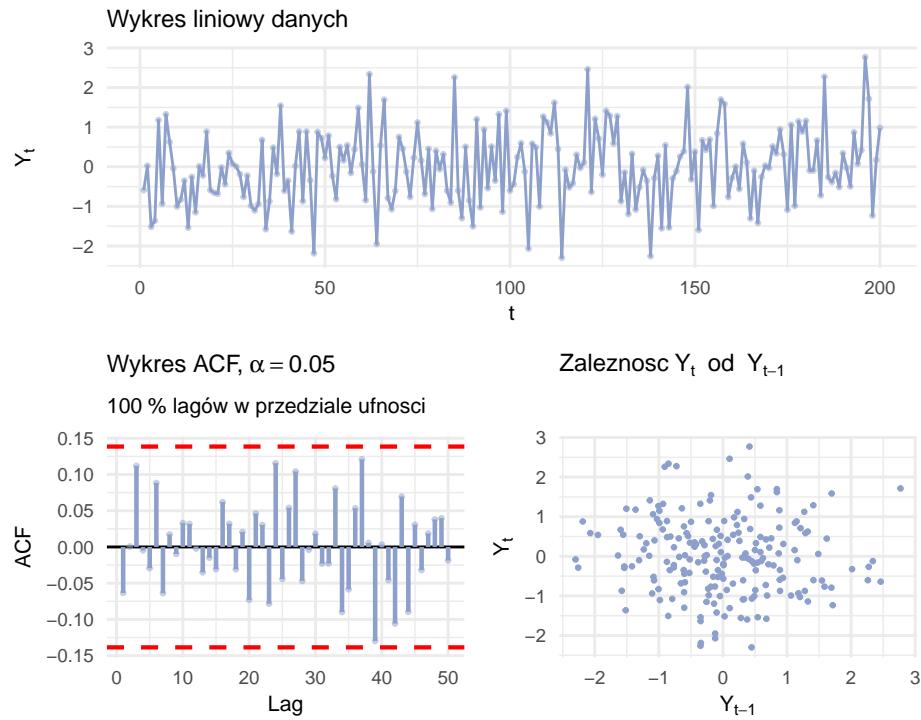
## Spis tabel

1	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych $WN(0,\sigma^2)$ . . . . .	4
2	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych $Exp(2)$ . . . . .	5
3	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych . . . . .	6
4	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	7
5	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	7
6	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	8
7	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	8
8	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	9
9	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	9
10	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	9
11	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	10
12	Wyniki procentowe kwalifikacji danych jako biały szum . . . . .	10
13	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych . . . . .	11
14	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych . . . . .	12
15	Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych . . . . .	13

# 1 Zadanie 3

## 1.1 Ilustracja działania testów białoszumowości

### 1.1.1 Zilustrowanie działania testów dla danych ( $n = 200$ ) typu biały szum (rozkład $WN(0, \sigma^2)$ )



Rysunek 1: Wykres testu graficznego dla procesu białego szumu

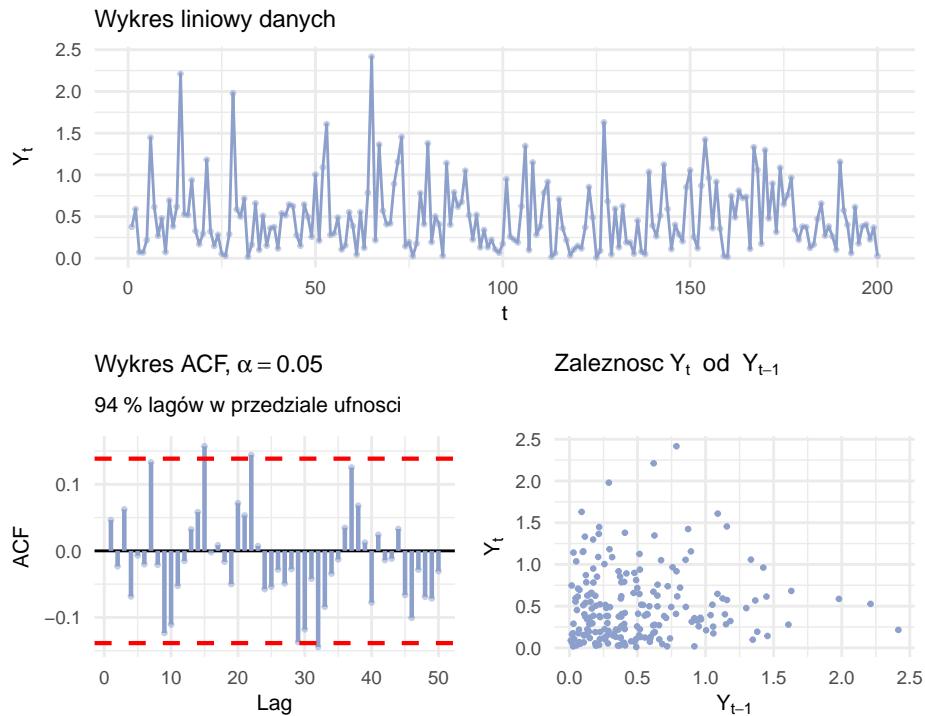
Na wykresie przedziały ufności zaznaczono czerwonymi przerywanymi liniami. Widzimy, że 100% autokorelacji próbkkowych leży w nim i nie obserwujemy punktów wyraźnie wykraczających poza granice — jest to zachowanie zgodne z założeniem białego szumu.

Tabela 1: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych  $WN(0, \sigma^2)$

Test	White_noise
Box-Pierce	TRUE
Ljung-Box	TRUE

Wnioski: Każdy test poprawnie zidentyfikował przykładowe dane jako biały szum

### 1.1.2 Zilustrowanie działania testów dla danych ( $n = 200$ ) typu (rozkład $Exp(2)$ )



Rysunek 2: Wykres testu graficznego dla  $Exp(2)$

Na rysunku przedstawiono wykres funkcji ACF dla danych z rozkładu wykładniczego  $Exp(2)$ . W tym przypadku jedynie ok. **94% autokorelacji próbkoowych** znajduje się w przedziałach ufności dla  $\alpha = 0,05$ , a ponadto widoczne są pojedyncze wartości wyraźnie wykraczające poza te granice.

Zgodnie z przyjętym kryterium oznacza to, że **test graficzny nie zaklasyfikowałby tego szeregu jako biały szum**, mimo że z konstrukcji wiemy, iż jest to ciąg niezależnych zmiennych losowych o średniej  $\frac{1}{2}$  i wariancji  $\frac{1}{4}$ , czyli teoretycznie spełniających definicję białego szumu.

Tabela 2: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych  $Exp(2)$

Test	White_noise
Box-Pierce	TRUE
Ljung-Box	TRUE

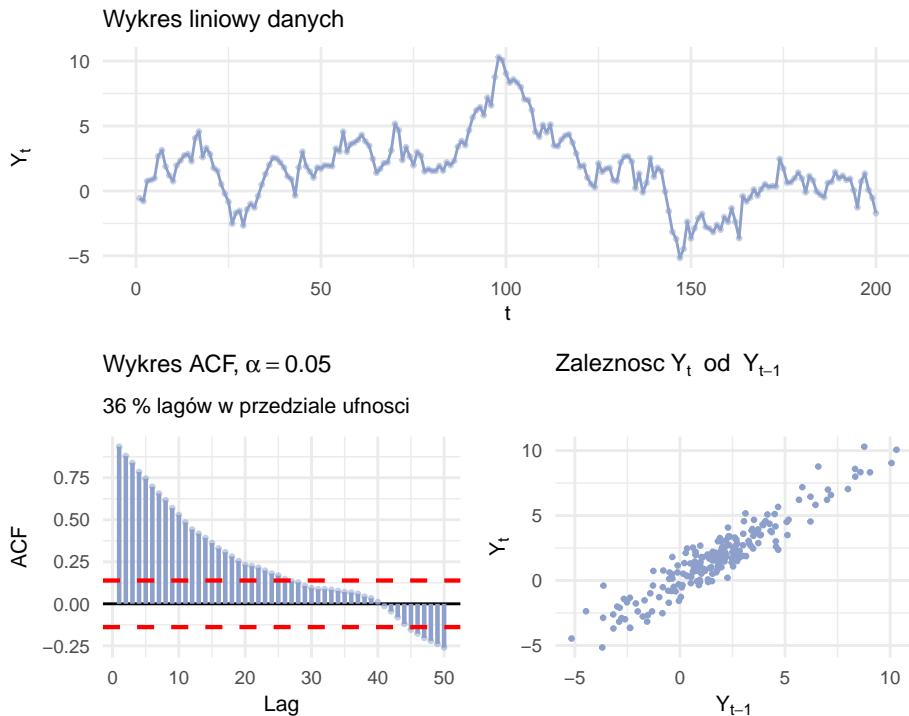
Wnioski: Dla danych testowych jedynie testy formalne poprawnie zidentyfikowały je jako biały szum

### 1.1.3 Zilustrowanie działania testów dla danych ( $n = 200$ ) z autokorelacją (nie biały szum)

Dane zostały wygenerowane jako realizacja jednowymiarowego procesu autoregresyjnego rzędu 1 (AR(1))  $\{X_t\}_{t=1}^n$ , zdefiniowanego rekurencyjnie przez

$$X_1 = \varepsilon_1, \quad X_t = \varphi X_{t-1} + \varepsilon_t, \quad t = 2, \dots, n,$$

gdzie  $\varphi = 1$  oraz  $\{\varepsilon_t\}$  jest ciągiem niezależnych zmiennych losowych o rozkładzie  $\varepsilon_t \sim \mathcal{N}(0, 1)$ .



Rysunek 3: Wykres testu graficznego dla danych nieskorelowanych

36% autokorelacji leży w przedziale ufności, więc według testu graficznego nie to jest biały szum

Tabela 3: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych

Test	White_noise
Box-Pierce	FALSE
Ljung-Box	FALSE

Wnioski: Wszystkie testy poprawnie zidentyfikowały dane jako **NIE** biały szum

## 1.2 Porównywanie testów graficznych oraz formalnych dla testowania białoszumowości

W testach przyjmiemy  $\alpha = 0.05$

### 1.2.1 Porównanie testów pod względem liczby powtórzeń

#### 1.2.1.1 Dla $WN(0, \sigma^2)$

Tabela 4: Wyniki procentowe kwalifikacji danych jako biały szum

	500	200	50	25	5
graficzny_test	0.71	0.70	0.64	0.56	0.8
test_pearsona	0.97	1.00	1.00	0.92	1.0
test_ljunga	0.91	0.93	0.96	0.88	1.0

Na podstawie wyników przedstawionych w tabeli można zauważyć, że liczba powtórzeń symulacji ma istotny wpływ na szacowany odsetek prawidłowych klasyfikacji białego szumu. Dla większej liczby replik (500 lub 200) wyniki są znacznie stabilniejsze, natomiast przy małej liczbie powtórzeń (50, 25, 5) pojawia się duża zmienność, co utrudnia rzetelną ocenę poziomu testu.

Największą zmiennością i najniższą zgodnością z teoretycznym poziomem istotności charakteryzuje się **test graficzny**, który często daje wyniki znacznie poniżej oczekiwanych 95%. Wynika to z faktu, że test graficzny jest naturalnie zależny od oceny eksperta – jego największą zaletą jest właśnie **czynnik ludzki**, czyli możliwość wizualnej interpretacji wykresu ACF, identyfikacji nieoczywistych wzorców, sezonowości czy struktury zależności, których nie da się łatwo ująć poprzez automatyczne reguły decyzyjne.

Testy formalne — Boxa–Pierce'a oraz Ljunga–Boxa — zachowują się znacznie stabilniej. Wyniki testu Boxa–Pierce'a są zwykle nieco zawyżone (częściej akceptuje  $H_0$ ), co oznacza, że test ten jest *zachowawczy*.

#### 1.2.1.2 Dla $Exp(4)$

Tabela 5: Wyniki procentowe kwalifikacji danych jako biały szum

	1000	500	200	50	25	5
graficzny_test	0.75	0.74	0.72	0.68	0.56	1
test_pearsona	0.98	0.98	0.95	1.00	0.96	1
test_ljunga	0.93	0.92	0.90	0.94	0.88	1

Dla danych wygenerowanych z rozkładu wykładniczego  $\text{Exp}(2)$  otrzymujemy analogiczne wnioski jak w przypadku szumu normalnego.

#### 1.2.1.2.1 Dla danych skorelowanych

Tabela 6: Wyniki procentowe kwalifikacji danych jako biały szum

	1000	500	200	50	25	5
graficzny\_test	0	0	0	0	0	0
test_pearsona	0	0	0	0	0	0
test_ljunga	0	0	0	0	0	0

W przypadku danych skorelowanych, generowanych zgodnie z podanym wcześniej modelem rekurencyjnym, wszystkie zastosowane testy — zarówno graficzny, jak i testy Boxa–Pierce'a oraz Ljunga–Boxa — poprawnie identyfikują je jako **niebędące białym szumem**, niezależnie od liczby przeprowadzonych powtórzeń symulacji.

#### 1.2.2 Porównanie pod względem różnych długości testów

**1.2.2.1 Dla  $WN(0, \sigma^2)$**  Tutaj dla każdej długości zostało wykonanych 100 testów i policzona z nich średnia

Tabela 7: Wyniki procentowe kwalifikacji danych jako biały szum

	400	200	50	25	10	5
graficzny\_test	0.68	0.74	0.63	0.86	0.92	1.00
test_pearsona	1.00	0.99	0.98	0.99	0.94	1.00
test_ljunga	0.85	0.95	0.95	0.93	0.92	0.92

Analiza danych zestawionych w tabeli prowadzi do wniosków zbieżnych z tymi uzyskanymi wcześniej dla różnych liczb powtórzeń symulacji.

### 1.2.2.2 Dla $\text{Exp}(4)$

Tabela 8: Wyniki procentowe kwalifikacji danych jako biały szum

	400	200	50	25	10	5
graficzny_test	0.73	0.72	0.72	0.94	0.99	1.00
test_pearsona	0.98	0.99	1.00	0.98	1.00	1.00
test_ljunga	0.91	0.94	0.95	0.95	0.98	0.99

Wnioski analogiczne, jak w poprzednim przykładzie

### 1.2.2.3 Dla danych skorelowanych

Tabela 9: Wyniki procentowe kwalifikacji danych jako biały szum

	400	200	50	25	10	5
graficzny_test	0	0	0	0.00	0.55	1
test_pearsona	0	0	0	0.02	0.76	1
test_ljunga	0	0	0	0.02	0.54	1

W tym przypadku wyraźnie widać, że mała liczba powtórzeń silnie zaburza jakość identyfikacji. Przy bardzo niskich długościach danych (5 lub 10) wszystkie testy mają tendencję do błędного klasyfikowania danych skorelowanych jako biały szum. Zjawisko to jest zgodne z klasycznymi zaleceniami Boxa i Jenkinsa, według których wiarygodna ocena autokorelacji wymaga spełnienia warunku  $h \geq 50$  oraz ograniczenia maksymalnego opóźnienia do  $h \leq \frac{n}{4}$

## 1.2.3 Porównanie pod względem różnego maksymalnego opóźnienia $h$

### 1.2.3.1 Dla $WN(0, \sigma^2)$

Tabela 10: Wyniki procentowe kwalifikacji danych jako biały szum

	199	100	50	10	5
graficzny_test	0.91	0.88	0.72	0.64	0.76
test_pearsona	1.00	1.00	0.98	0.96	0.93
test_ljunga	0.84	0.92	0.94	0.95	0.93

Przeprowadzone analizy pokazują, że wyniki testu graficznego silnie zależą od wyboru maksymalnego opóźnienia  $h_{\max}$ . Odsetek szeregów kwalifikowanych ajko biały szum rośnie dla

niego wraz ze wzrostem  $h_{\max}$ . Test Box–Pierce'a daje wyniki mieszczące się w przedziale od 95% do 100% akceptacji  $H_0$ , co oznacza, że jest nieco zachowawczy, a w miarę zmniejszania  $h_{\max}$  uzyskiwane wartości zbliżają się do teoretycznego poziomu 95%, przy czym test pozostaje stosunkowo stabilny względem wyboru liczby lagów. Z kolei test Ljunga–Boxa dla bardzo dużych opóźnień (np.  $h_{\max} = 199$ ) staje się zbyt „ostry” i zbyt często odrzuca  $H_0$ .

### 1.2.3.2 Dla $\text{Exp}(4)$

Tabela 11: Wyniki procentowe kwalifikacji danych jako biały szum

	199	100	50	10	5
graficzny_test	0.91	0.84	0.72	0.65	0.80
test_pearsona	1.00	1.00	0.99	0.98	0.96
test_ljunga	0.82	0.92	0.94	0.98	0.94

### 1.2.3.3 Dane z dryfem

Tutaj analogiczne wnioski jak w **poprzednim** przykładzie

Tabela 12: Wyniki procentowe kwalifikacji danych jako biały szum

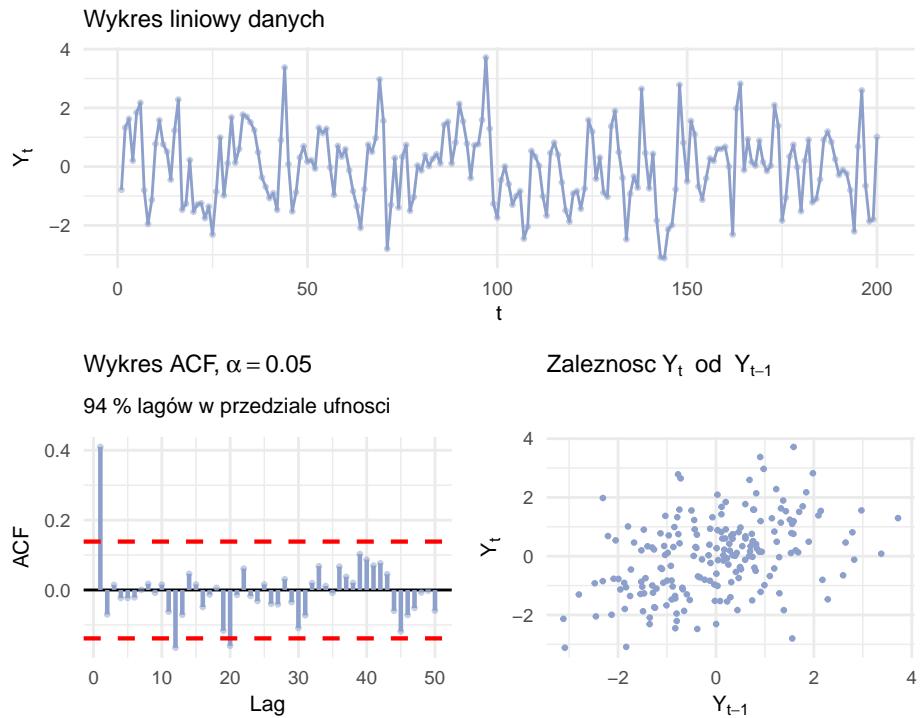
	199	100	50	10	5
graficzny_test	0	0	0	0	0
test_pearsona	0	0	0	0	0
test_ljunga	0	0	0	0	0

Dla danych z dryfem, wszystkie testy działają poprawnie, niezależnie od maksymalnego lagu.

## 1.2.4 Przykłady szeregów czasowych innych niż szum IID

### 1.2.4.1 Model stacjonarny o niewielkiej korelacji (MA(1))

Rozważamy szereg czasowy wygenerowany z modelu MA(1)  $X_t = Z_t + 0,3Z_{t-1}$ , gdzie  $Z_t \sim WN(0, 1)$ .



Rysunek 4: Wykres testu graficznego dla modelu stacjonarnego z niewielką korelacją

Około **94% wartości ACF** mieści się w przedziale ufności, natomiast **jeden lag wyraźniego przekracza**. Może to sugerować niewielkie odstępstwo od białoszumowego charakteru danych, jednak **na podstawie samego testu graficznego jest to obserwacja na granicy istotności**.

Tabela 13: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych

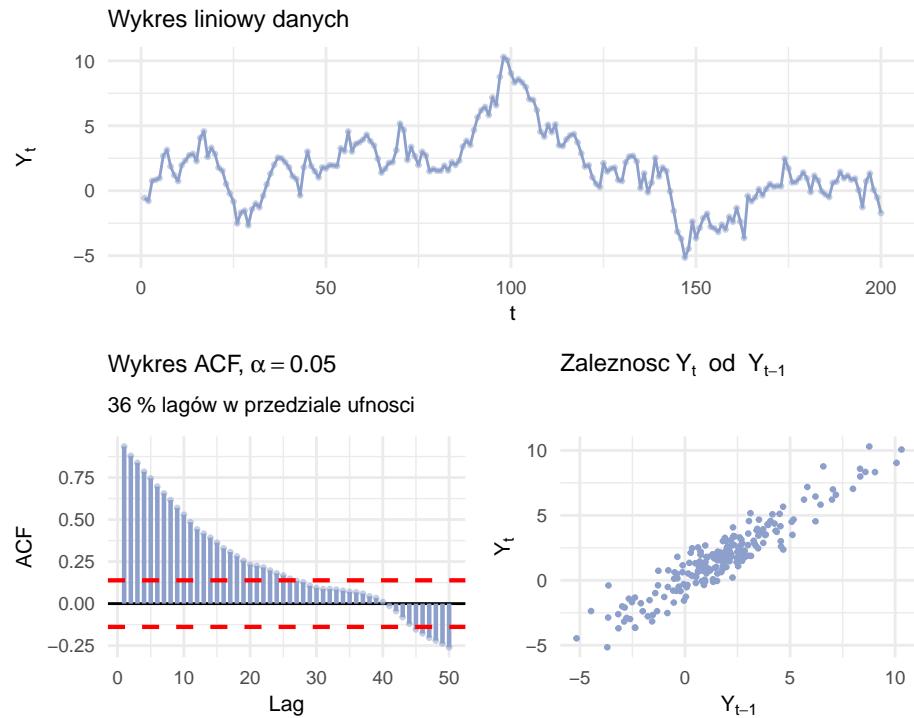
Test	White_noise
Box-Pierce	FALSE
Ljung-Box	FALSE

Co jednoznacznie wskazuje na niebiałoszumowy charakter szeregu czasowego

**1.2.4.2 Błędzenie losowe** Rozważamy proces **błędzenia losowego** (random walk) zdefiniowany jako

$$Y_0 = 0, \quad Y_t = Y_{t-1} + e_t, \quad t = 1, \dots, n,$$

gdzie  $\{e_t\}$  jest białym szumem  $e_t \sim \mathcal{N}(0, 1)$ .



Rysunek 5: Wykres testu graficznego dla modelu błędzenia losowego

Test graficzny jednoznacznie wskazuje na brak białoszumowego charakteru danych

Tabela 14: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych

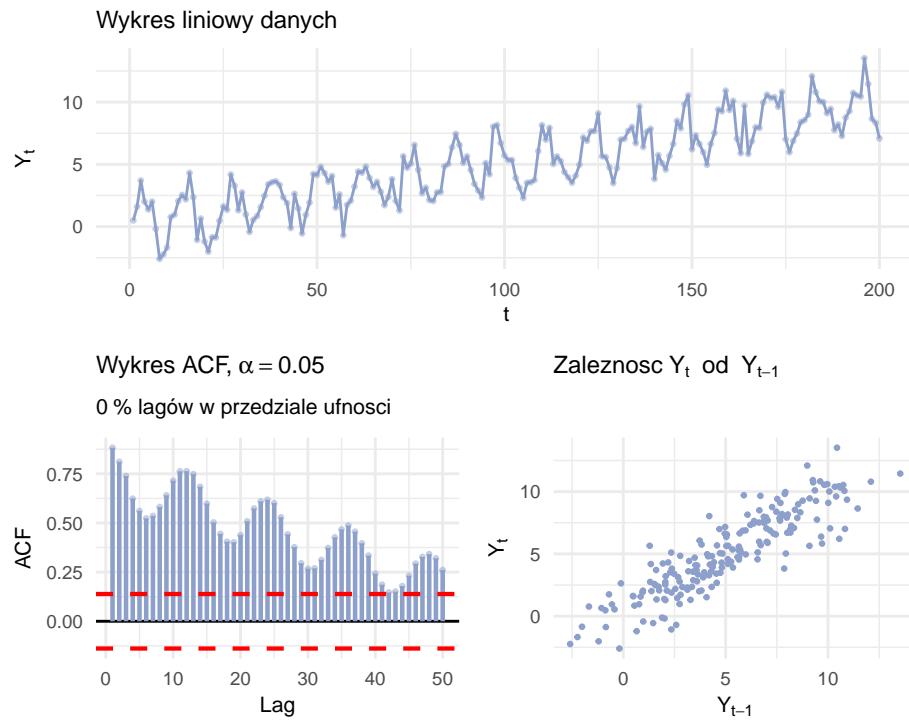
Test	White_noise
Box-Pierce	FALSE
Ljung-Box	FALSE

W tym przypadku oba testy jednoznacznie wskazują na brak charakteru białoszumowego danych

**1.2.4.3 szumm IID + trend deterministyczny + sezonowość** Rozważamy szereg czasowy złożony z i.i.d. szumu, liniowego trendu rosnącego oraz składowej sezonowej o okresie 12:

$$X_t = \varepsilon_t + 0,05 t + 2 \sin\left(\frac{2\pi t}{12}\right), \quad t = 1, \dots, 200,$$

gdzie  $\varepsilon_t \sim \mathcal{N}(0, 1)$  są niezależnymi zmiennymi losowymi o rozkładzie normalnym (szum i.i.d. o średniej 0 i odchyleniu standardowym 1).



Rysunek 6: Wykres testu graficznego dla modelu szumu iid z dryfem i sezonowością

Test graficzny jednoznacznie wskazuje na brak białoszumowego charakteru danych

Tabela 15: Porównanie wyników testów Box-Pierce'a i Ljunga-Boxa przy danych danych skorelowanych

Test	White_noise
Box-Pierce	FALSE
Ljung-Box	FALSE

Co potwierdzają również test Boxa-Persea oraz Ljunga-Boxa zwracając FALSE.

## 2 Zadanie 4

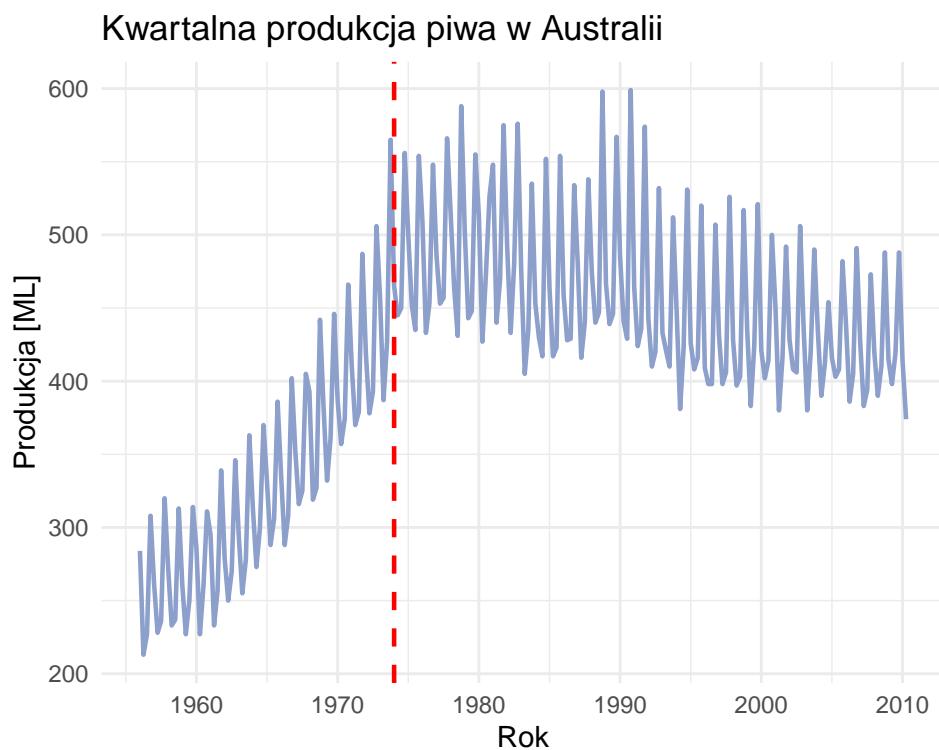
### 2.1 Wybór danych

Do analizy wykorzystano dane z zestawu **ausbeer**, przedstawiające kwartalną produkcję piwa w Australii, wyrażoną w megalitrach (milionach litrów). Zbiór obejmuje kolejne kwartały od lat 50. do lat 90., dzięki czemu pozwala na obserwację zarówno sezonowości, jak i długookresowych trendów w produkcji.

okres: 1956 Q1 – 2008 Q2

### 2.2 Zastosowanie poznanych metod graficznych, do podstawowej analizy wykresu

#### 2.2.1 Wykres Liniowy



Rysunek 7: Liniowy wykres danych

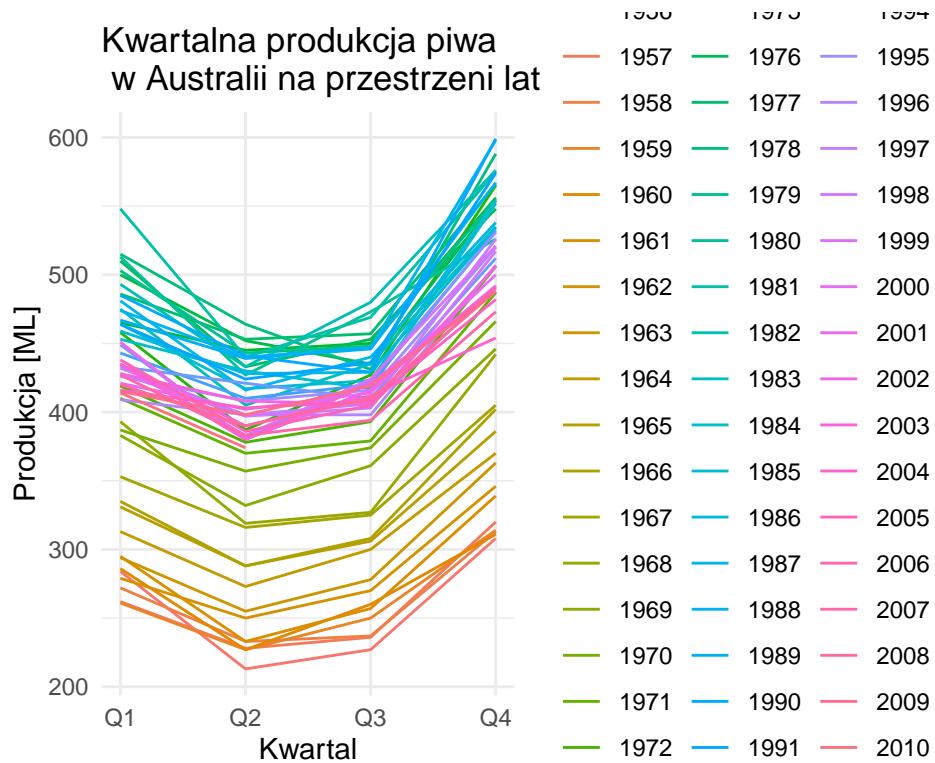
Z wykresu można zauważyć wyraźną sezonowość w krótszych przedziałach czasowych. Dane wykazują również dwa odmienne trendy:

- **trend rosący** od około 1950 roku do około 1974,

- **trend malejący** od około 1974 do 2010 roku.

Ponadto wykres sugeruje **niejednorodność wariancji** – wartości w pierwszej części szeregu (lata 1950–1974) charakteryzują się wyraźnie mniejszą zmiennością niż obserwacje po 1974 roku.

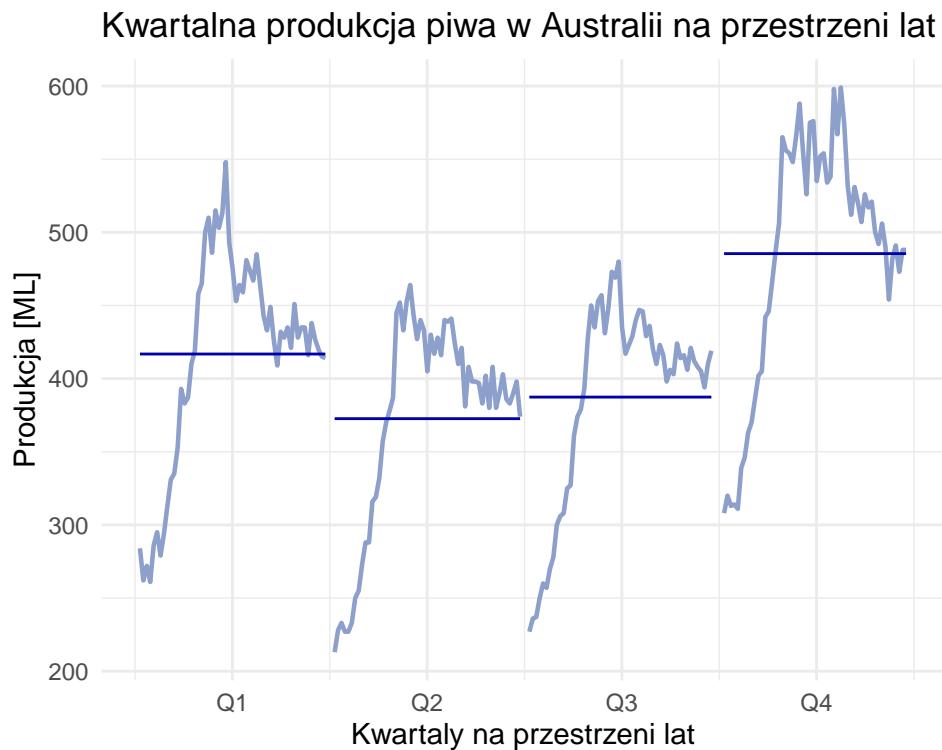
## 2.2.2 Wykres sezonowy



Rysunek 8: Sezonowy wykres danych

Na podstawie wykresu seasonplot widać, że na przestrzeni lat produkcja rosła, aż do ustabilizowania się na poziomie pomiędzy 400 a 500 na przestrzeni lat 1974 i 2010

### 2.2.3 Wykres kwartalny

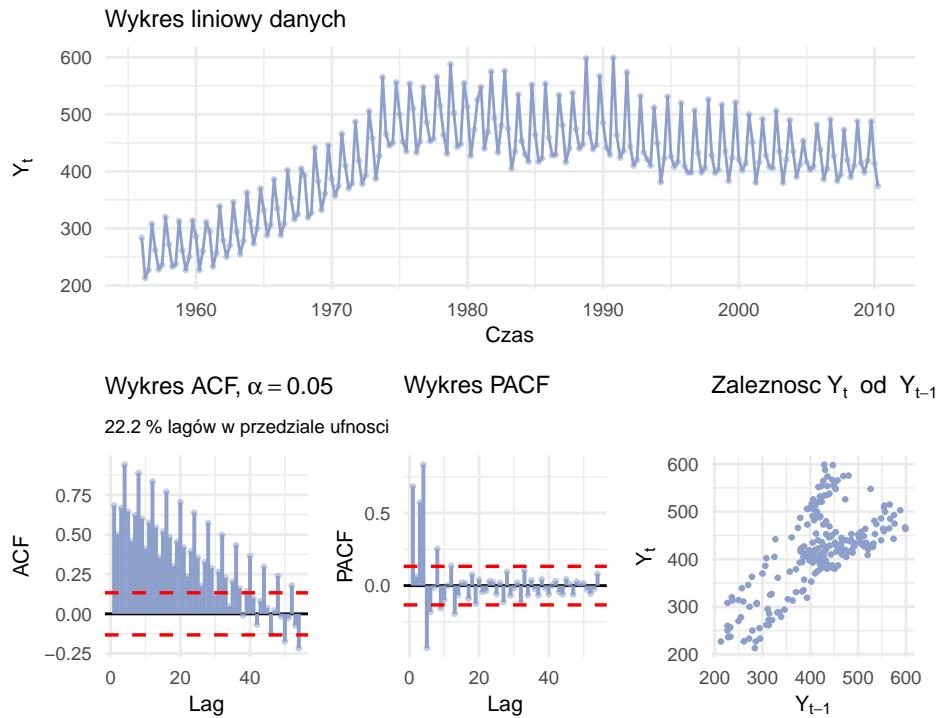


Rysunek 9: Kwartalny wykres danych

Na podstawie wykresu można zauważyć:

- **Trend długookresowy** w produkcji piwa na przestrzeni lat: początkowo rosnący, a następnie przechodzący w łagodny spadek (zgodnie z obserwacjami z poprzednich wykresów).
- **Sezonowość** – najwyższa średnia produkcja przypada na **pierwszy** oraz **czwarty kwartał** roku.

#### 2.2.4 Wykres ACF

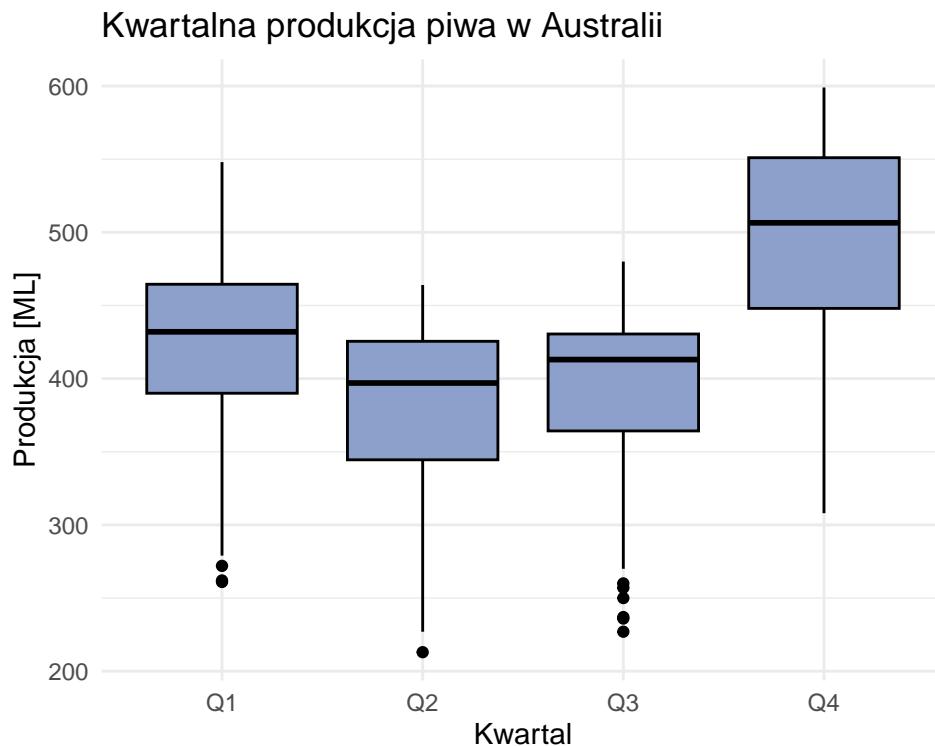


Rysunek 10: ACF wykres danych

Na podstawie wykresu ACF można zauważać:

- silną sezonowość kwartalną, co widać jako duże piki na lagach 4, 8, 12, 16 itd.;
- wolny zanik autokorelacji, wskazujący na obecność trendu i niestacjonarność sze-regu;
- wiele istotnych wartości ACF, co potwierdza zarówno trend, jak i stabilną sezono-wość.

## 2.2.5 Wykres boxplot

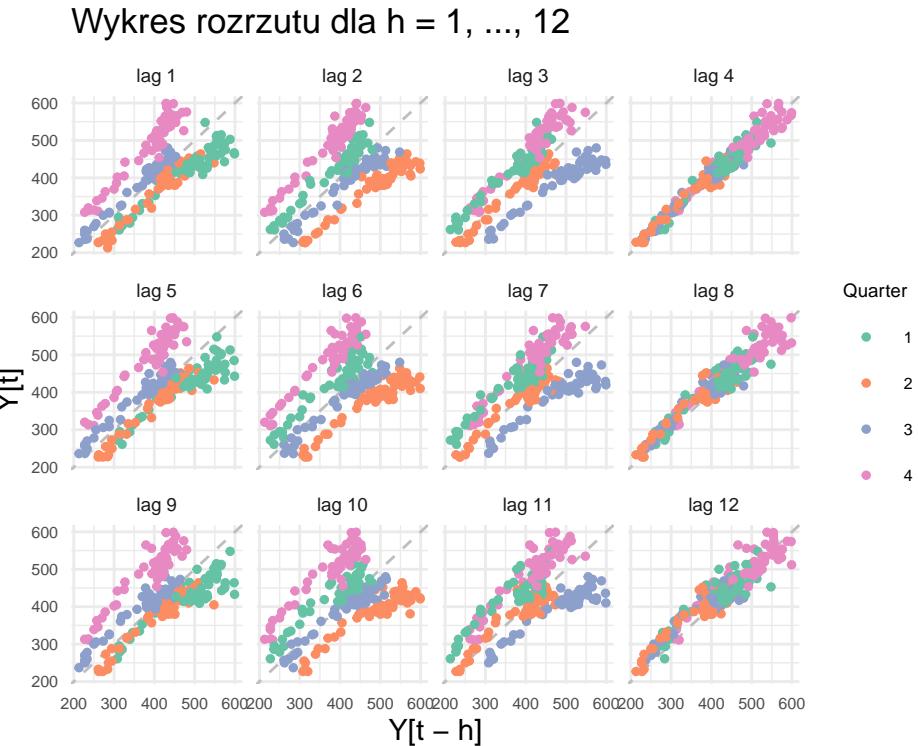


Rysunek 11: Boxplot wykres danych

Na podstawie wykresu typu *boxplot* można zauważyc:

- w kwartale **Q1**, **Q2** oraz **Q3** pojawiają się **wartości odstające**, co wskazuje na epizody produkcji znaczaco odbiegajace od typowego poziomu w tych okresach
- kwartalny rozkład produkcji różni się między okresami — kwartalny **Q4** charakteryzuje się najwyższą medianą oraz największym rozrzutem wartości
- rozkład w Q1–Q3 jest bardziej skupiony, natomiast Q4 wskazuje na większą zmienność sezonową

## 2.2.6 Wykresy rozrzutu



Rysunek 12: Wykres rozrzutu danych

Wykres autokorelacji pokazuje, że:

- **lag 4** wykazuje wyraźnie istotną korelację liniową, co jest zgodne z charakterem danych kwartalnych — obserwacje oddalone o cztery okresy reprezentują ten sam kwartał
- dla **lag 8** oraz **lag 12** również widoczne są istotne zależności, co wynika z faktu, że są to wielokrotności 4, a zatem odzwierciedlają strukturę kwartalną

## 2.2.7 Wstępne wnioski po analizie podstawowych wykresów

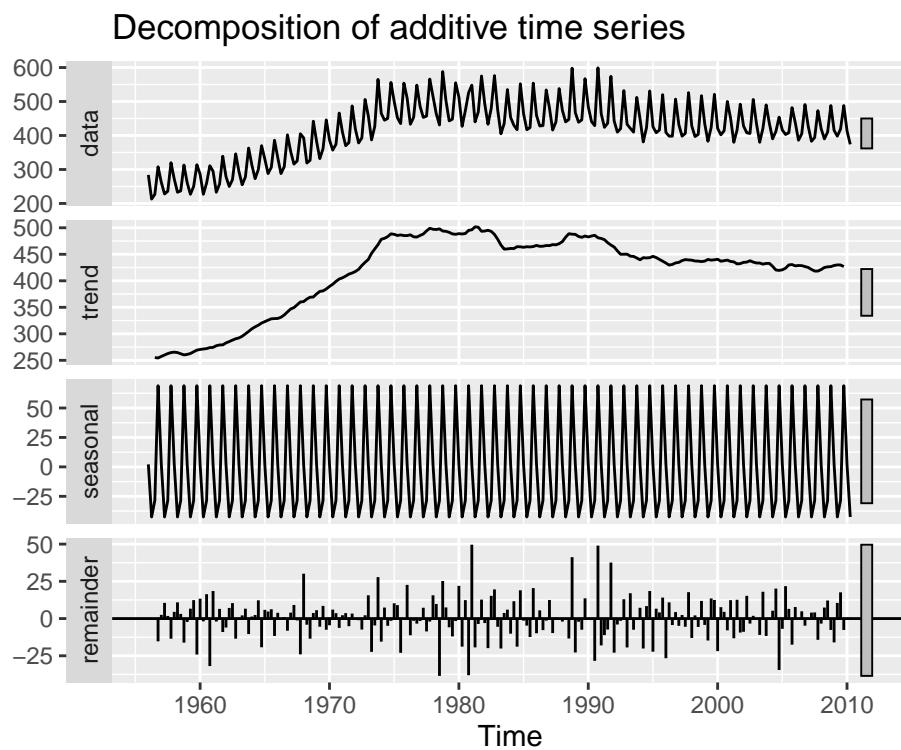
Na podstawie wstępnej analizy szeregu czasowego można sformułować następujące obserwacje:

- do około **1974 roku** widoczny jest **trend rosący**, po czym pojawia się **łagodny trend malejący**
- dane wykazują **wyraźną sezonowość** o okresie \*\*4 miesięcy
- funkcja autokorelacji (ACF) **zanika powoli oraz wykazuje strukturę okresową**, co potwierdza obecność trendu i sezonowości

- ACF nie odpowiada charakterystycze białego szumu, ponieważ wiele wartości przekracza granice przedziałów ufności
- w danych występują obserwacje odstające, szczególnie w pierwszych trzech kwartałach
- największą korelację obserwuje się dla opóźnienia 4, co jest zgodne z kwartalnym charakterem danych (na podstawie wykresu rozrzutu).

## 2.3 Dekompozycje

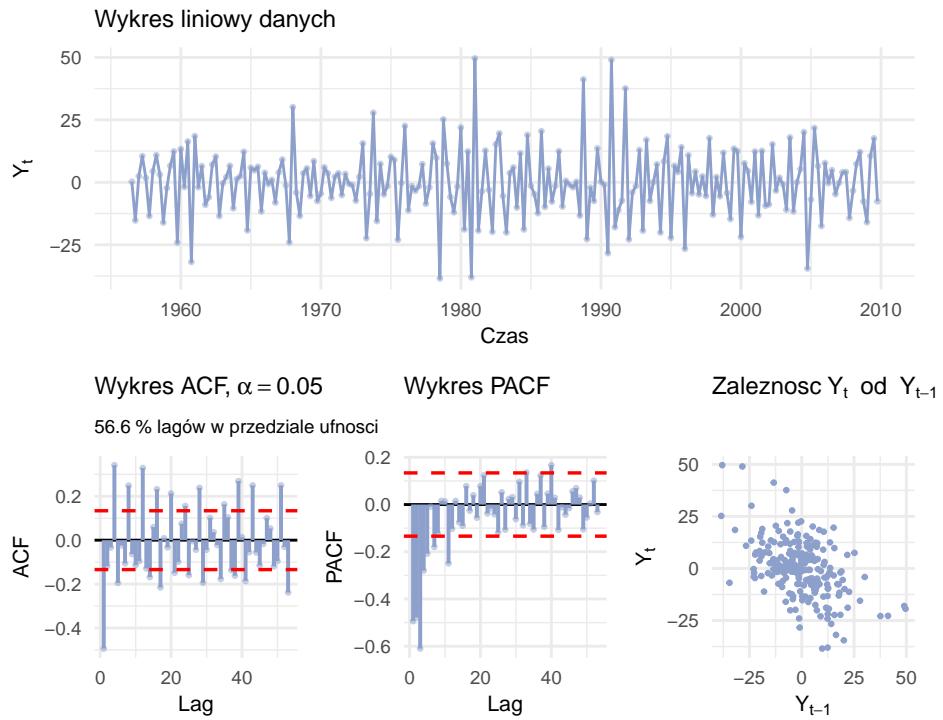
### 2.3.1 Dekompozycja addytywna



Rysunek 13: Dekompozycja addytywna danych

**2.3.1.1 Wnioski z dekompozycji addytywnej** Dekompozycja addytywna potwierdza wcześniejsze obserwacje dotyczące badanego szeregu czasowego:

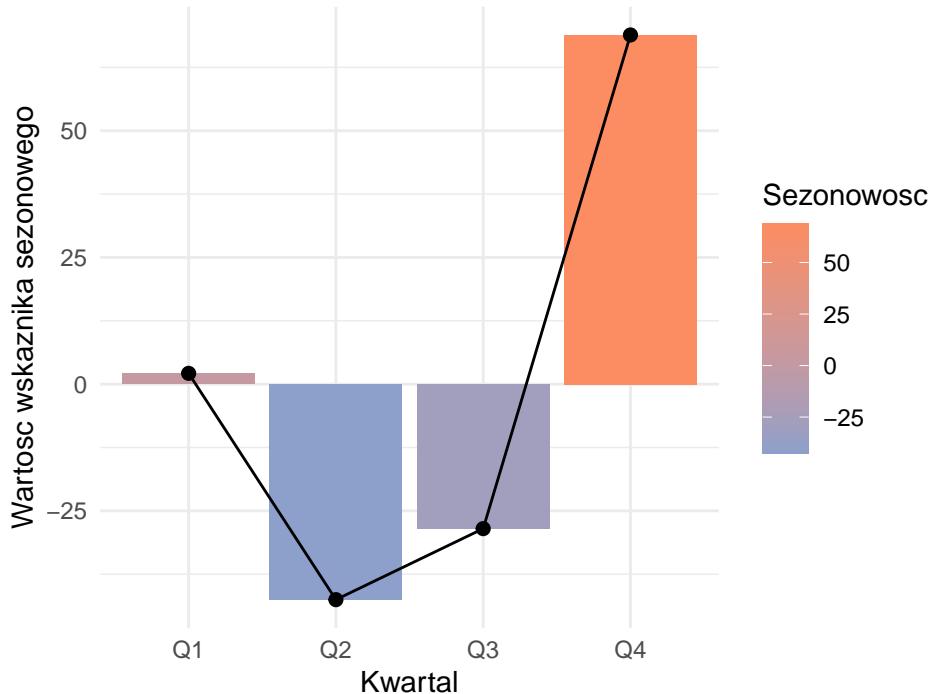
- występuje silna sezonowość danych, widoczna jako regularnie powtarzający się komponent sezonowy;
- komponentu trendu wykazuje wyraźny podział na dwa segmenty:
  - trend rosnący w pierwszym okresie (do około 1974 roku),
  - trend malejący po 1974 roku.



Rysunek 14: Acf dla dekompozycji addytywnej danych

**2.3.1.2 Analiza reszt** Na podstawie wykresu reszt oraz odpowiadających im funkcji ACF, PACF i wykresu zależności  $Y_t$  od  $Y_{t-1}$  nie można jednoznacznie stwierdzić, że reszty mają charakter białego szumu. Co prawda nie obserwuje się wyraźnego trendu ani sezonowości, jednak jedynie ok. 60% wartości ACF mieści się w przedziałach ufności, a kilka lagów wyraźnie od nich odstaje. Wskazuje to na obecność istotnej autokorelacji, a więc na dodatkową strukturę w resztach.

### Wskaźniki sezonowości



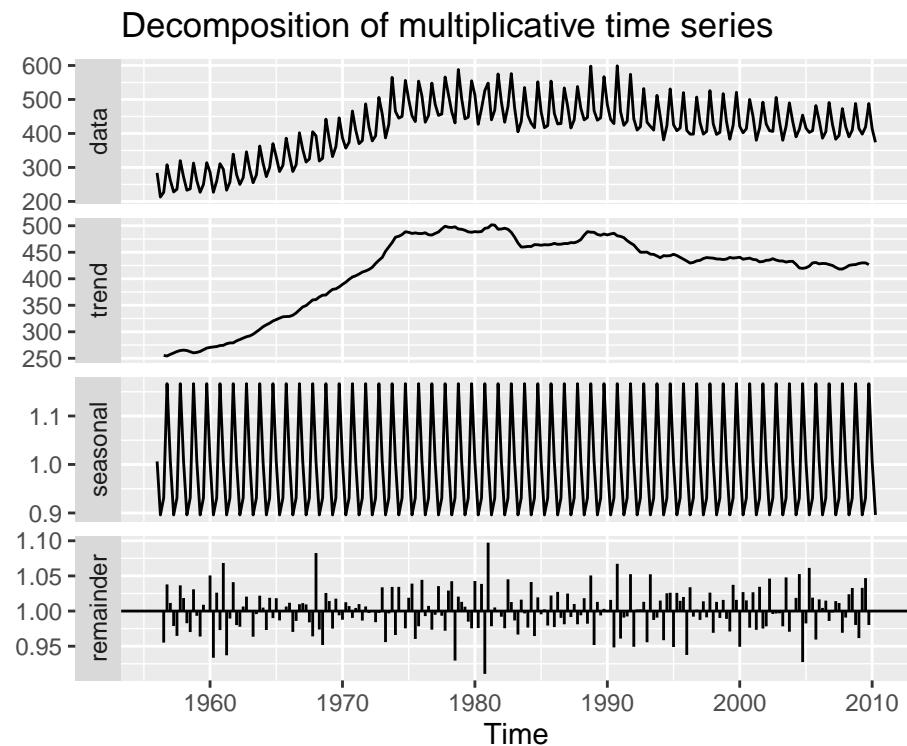
Rysunek 15: Boxplot dla dekompozycji addytywnej danych

#### 2.3.1.3 Wykres boxplot

**2.3.1.4 Wnioski z analizy wskaźników sezonowości** Na podstawie wykresu wskaźników sezonowości można stwierdzić, że:

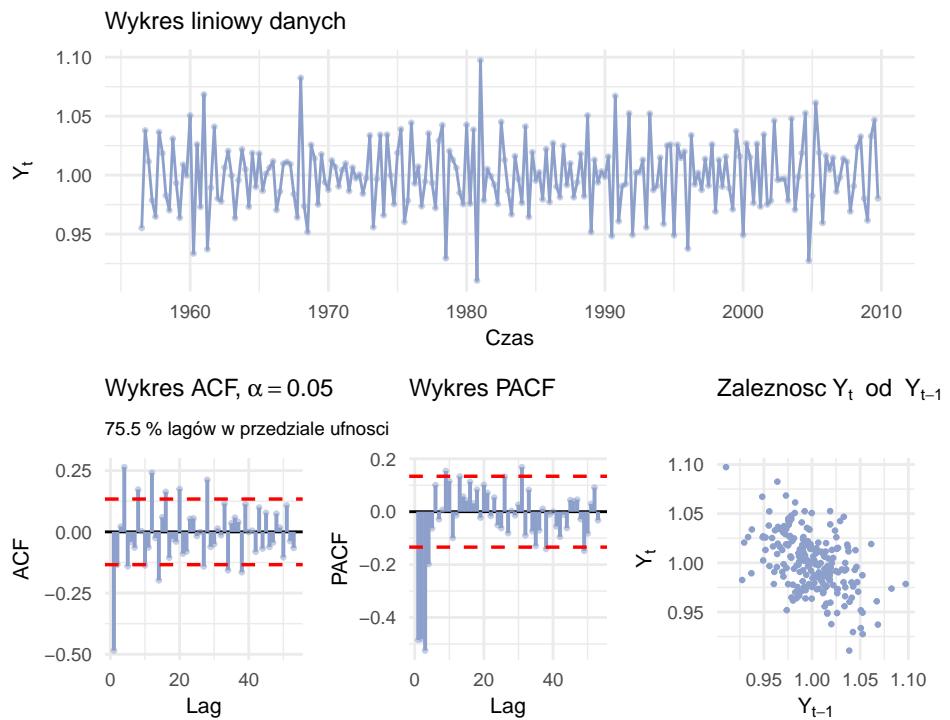
- **I kwartał (Q1)** charakteryzuje się produkcją nieznacznie wyższą od wartości przeciętnej,
- **II kwartał (Q2)** wykazuje **najniższy poziom sezonowy** – produkcja jest wyraźnie poniżej średniej,
- **III kwartał (Q3)** nadal pozostaje poniżej średniej, lecz wartości są wyższe niż w Q2,
- **IV kwartał (Q4)** osiąga **najwyższy dodatni wskaźnik sezonowy**, co oznacza największą produkcję w skali roku.

### 2.3.2 Dekompozycja multiplikatywna



Rysunek 16: Autoplot dla dekompozycji multiplikatywnej danych

**2.3.2.1 Autoplot** Wykres typu autoplot daje **społjne** wnioski z poprzednim rodzajem dekompozycji.

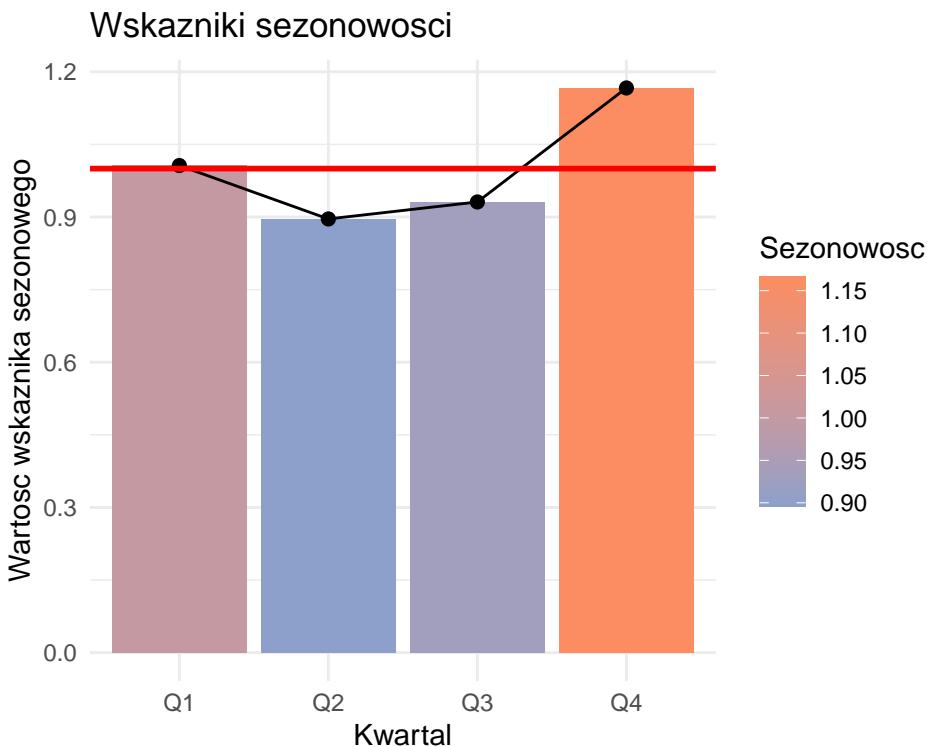


Rysunek 17: Acf dla dekompozycji multiplikatywnej danych

**2.3.2.2 Analiza reszt** Na podstawie wykresu reszt oraz odpowiadających im funkcji ACF, PACF i wykresu zależności  $Y_t$  od  $Y_{t-1}$  można stwierdzić, że dekompozycja multiplikatywna prowadzi do wyraźnie lepszych rezultatów niż dekompozycja addytywna.

- reszty **nie wykazują żadnego trendu ani sezonowości**
- około **75,5% wartości ACF** znajduje się w przedziałach ufności, pojawiają pojedyncze odstające wartości, co oznacza, że reszty **nie mają jeszcze w pełni białoszumowego charakteru**
- także wykres PACF pokazuje wyraźną poprawę: liczba odstających lagów jest mniejsza
- wykres zależności  $Y_t$  względem  $Y_{t-1}$  ma **bardziej losowy charakter** niż w przypadku dekompozycji addytywnej

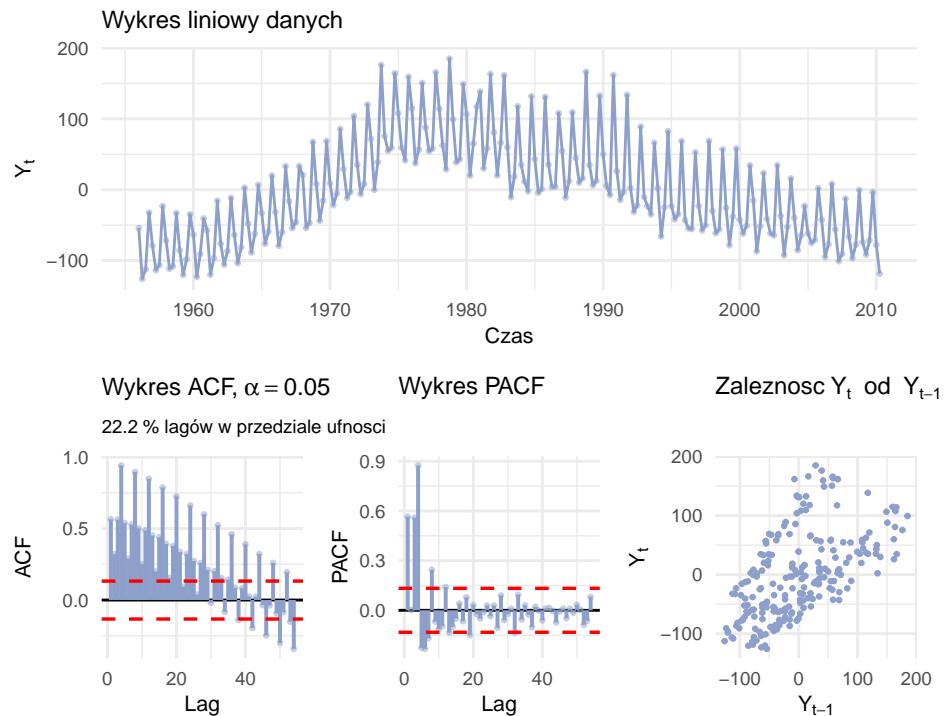
### 2.3.2.3 Barplot



Wnioski z barplotu są spójne z rezultatami dekompozycji addytywnej.

### 2.3.3 Dekompozycja na podstawie modelu regresji

#### 2.3.3.1 Trend liniowy

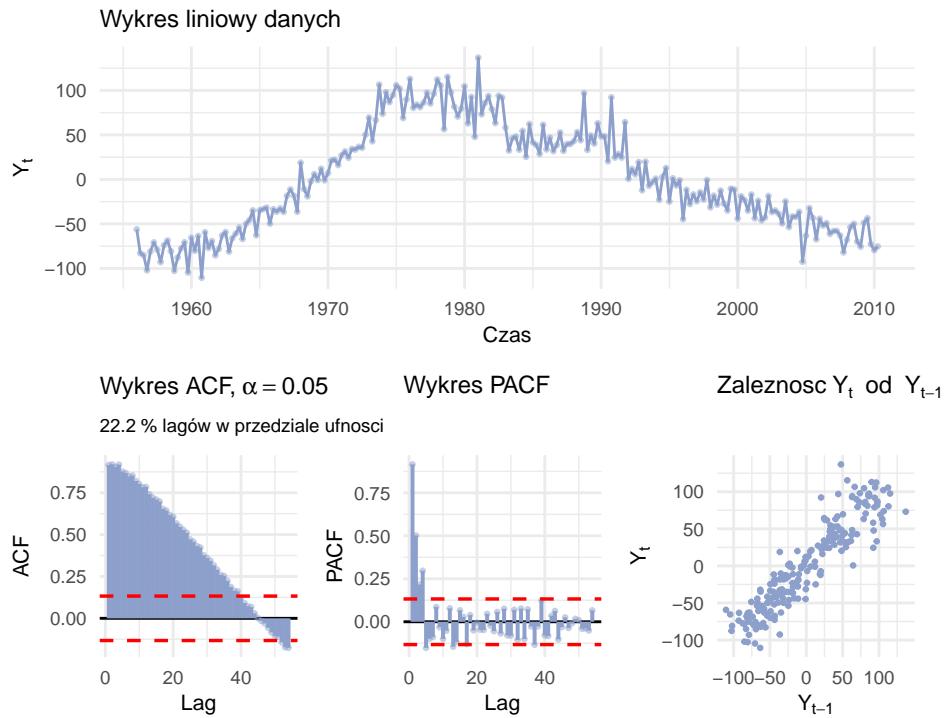


Na podstawie wyników estymacji modelu `tslm(dane ~ trend)` można stwierdzić, że:

- współczynnik determinacji  $R^2 \approx 0,27$  wskazuje, że trend liniowy wyjaśnia jedynie około **27% zmienności** danych – znaczna część wariancji pozostaje w resztach

Wnioski te wskazują, że sam trend liniowy nie wystarcza do pełnego opisu szeregu

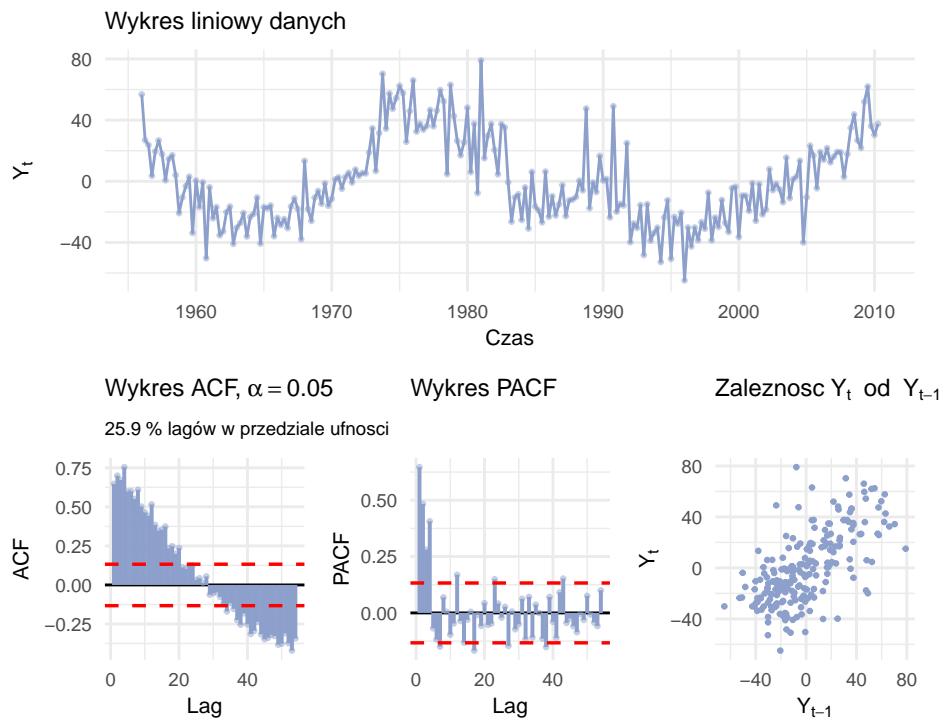
### 2.3.3.2 Trend liniowy plus sezonowość



### 2.3.3.3 Porównanie modeli: tylko trend liniowy vs trend liniowy + sezonowość

- jakość dopasowania nie poprawia się istotnie:
  - w modelu z trendem i sezonowością  $R^2$  pozostaje dalej ok 27%, nie widać, znacznej poprawy

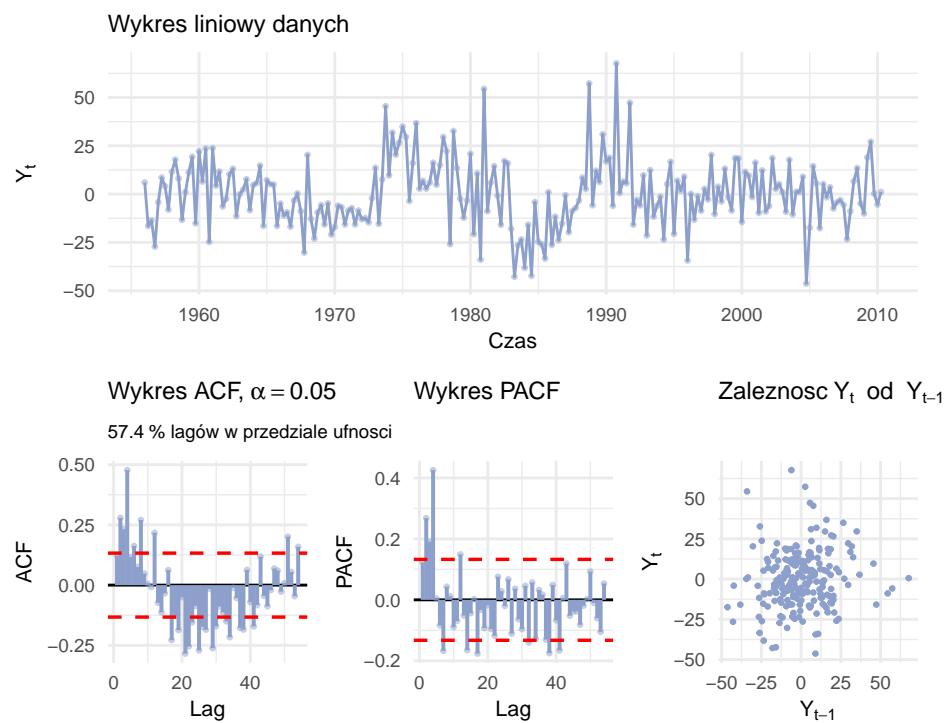
### 2.3.3.4 Trend kwadratowy plus sezonowość



Model `tslm(dane ~ season + trend + I(trend^2))` prowadzi do następujących obserwacji:

- współczynnik determinacji  $R^2 \approx 0.8909$ , a skorygowany  $R^2 \approx 0.8883$ , co oznacza, że model z trendem kwadratowym wyjaśnia **~89% zmienności** szeregu — wyraźnie więcej niż poprzedni model (27%)

### 2.3.3.5 Trend piątego stopnia sezonowość

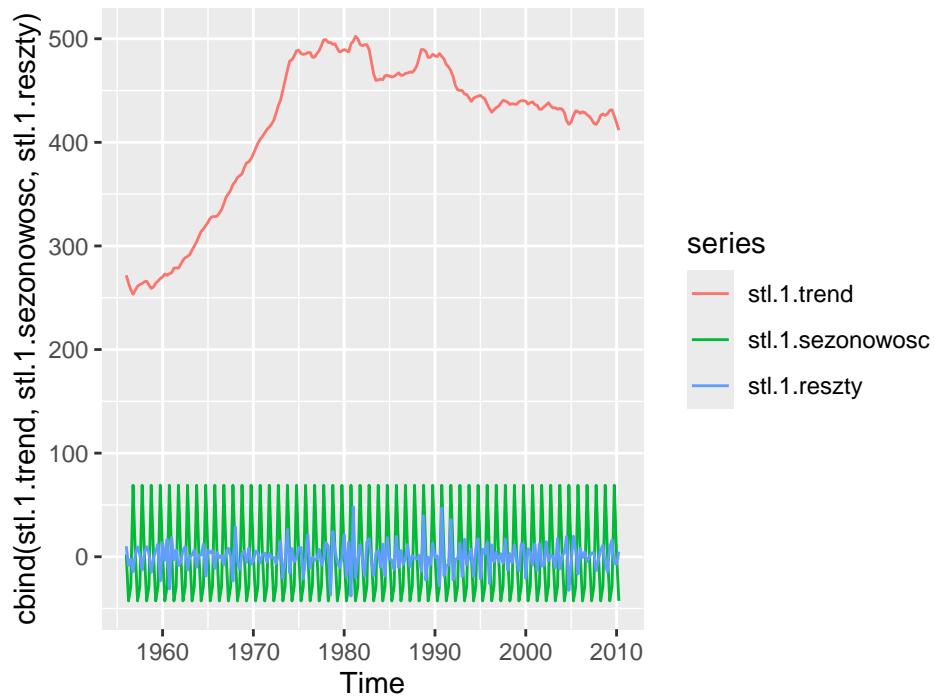


Model `tslm(dane ~ season + I(trend^5))` charakteryzuje się bardzo dobrym dopasowaniem:

- współczynnik determinacji wynosi  $R^2 \approx 0,96$ , a skorygowany  $R^2 \approx 0,956$ , co oznacza, że model wyjaśnia ponad 90% zmienności danych

W porównaniu z modelem `trend + season + trend^2` model z trendem piątego stopnia wypada o wiele lepiej. Aż 60% ACF znajduje się w przedziale ufności co jest dużym przeskokiem z 26%, gdy uwzględnialiśmy trend drugiego stopnia.

#### 2.3.4 Dekompozycja oparta na metodzie STL

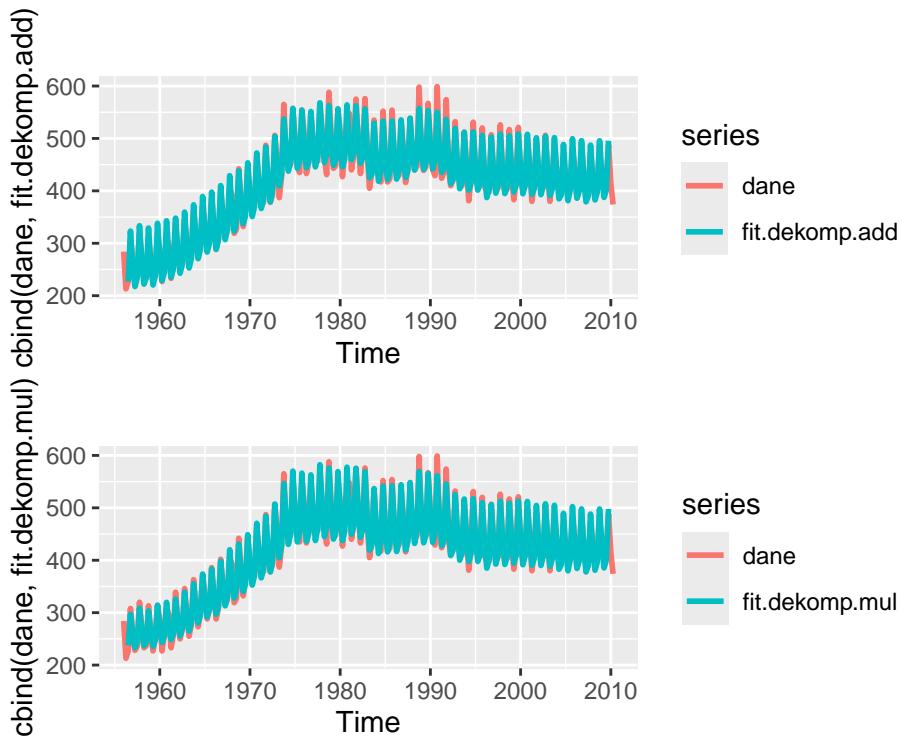


Na podstawie wykresu STL można zauważać:

- **trend wygładzony metodą LOESS** wykazuje nieliniową strukturę: wyraźny wzrost do 1974, a następnie stopniowy spadek;
- **sezonowość jest stabilna w czasie**, zgodnie z parametrem `s.window = "periodic"` — amplituda wahań sezonowych pozostaje stała, co potwierdza silny i powtarzalny charakter sezonowości kwartalnej;
- **reszty nie wykazują ani trendu, ani sezonowości**, co oznacza, że model STL skutecznie oddzielił główne komponenty szeregu.

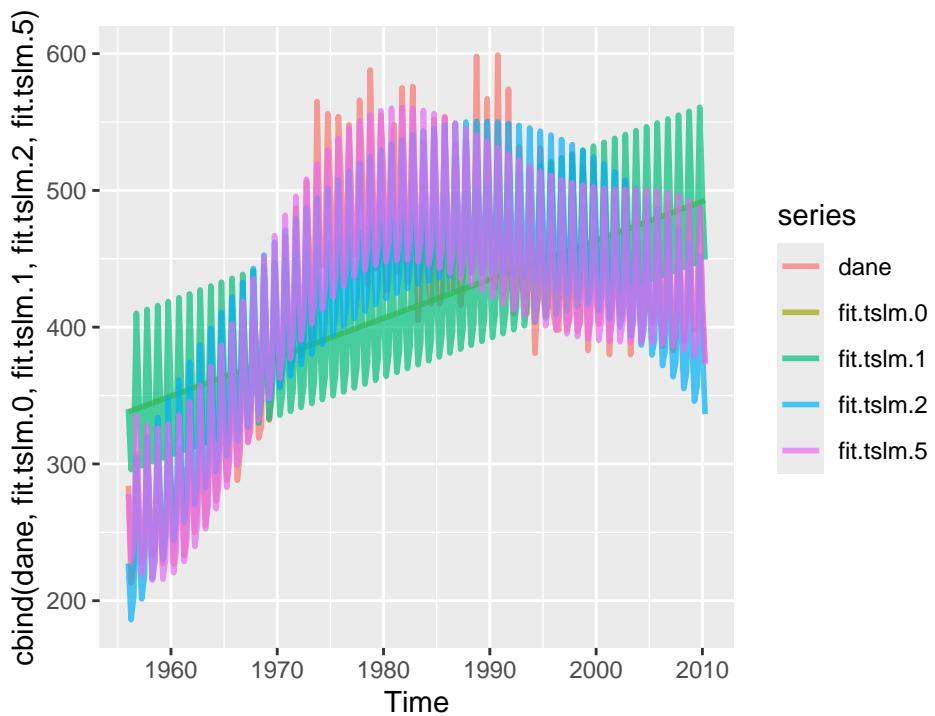
## 2.4 Porównanie zależności wyników od parametrów

### 2.4.1 Porównanie modeli decompose(addytywny i multiplikatywny)



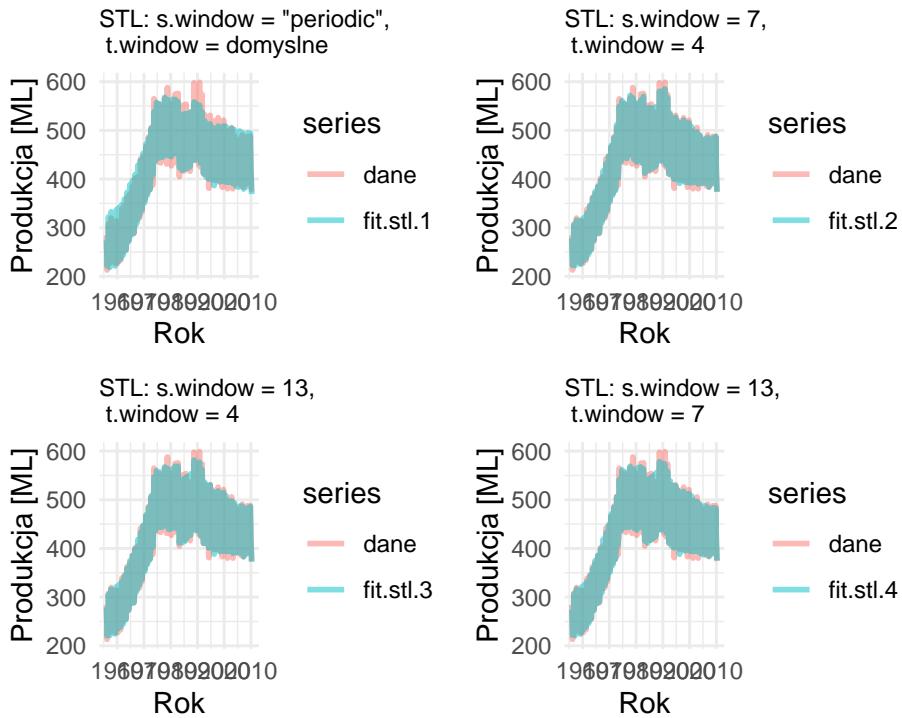
**2.4.1.1 Wniosek** Model multiplikatywny jest wyraźnie lepiej dopasowany do danych niż model addytywny. W przeciwieństwie do dekompozycji addytywnej, metoda multiplikatywna prawidłowo odwzorowuje multiplikatywny charakter sezonowości – amplituda wahań sezonowych rośnie .

## 2.4.2 Porównanie modeli tslm (trend jako wielomian różnych stopni)



Trend jest najlepiej odwzorowany, przy użyciu wielomianu stopnia **piątego**.

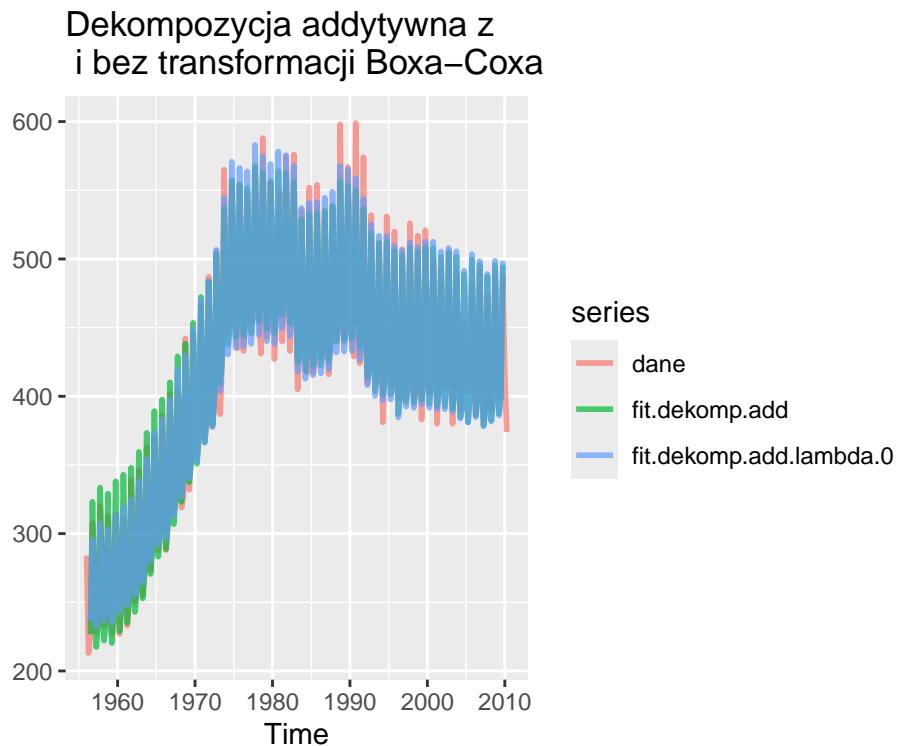
### 2.4.3 Porównanie modeli (stl dla różnych modeli parametrycznych trendu, parametry wygładzające)



Parametry wygładzające mają istotny wpływ na wyniki, z wykresów bezpośrednio wynika, że najlepsze dopasowanie uzyskujemy, przy parametrach wygładzających  $s.\text{window} = 7$  oraz  $t.\text{window} = 4$ . Takie parametry pozwalają na dobre odwzorowanie sezonowości kwartalnej oraz zmienności jej wariancji w czasie.

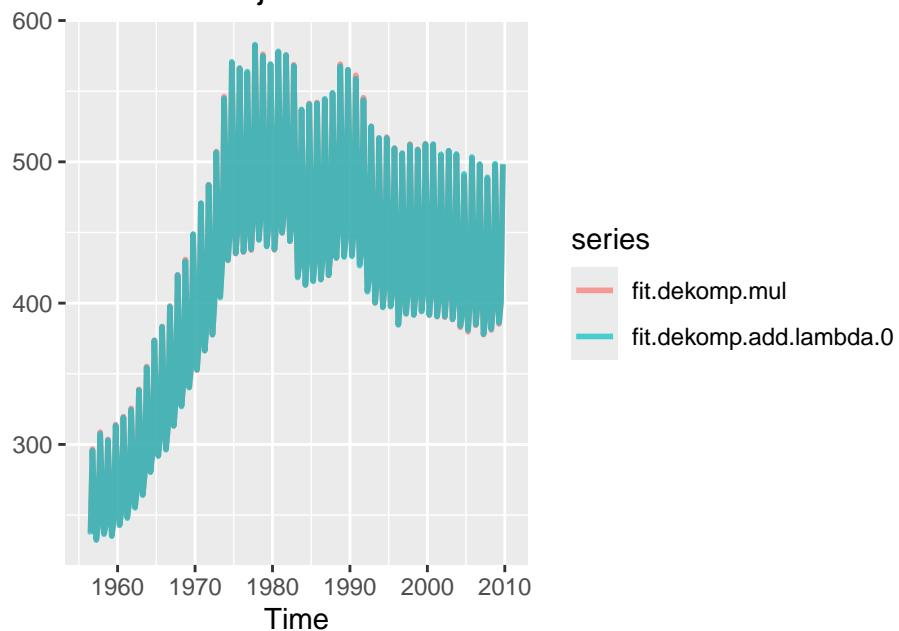
## 2.5 Próba rozstrzygnięcia czy zastosowanie transformacji Boxa-Coxa prowadzi do poprawy jakości dopasowania modeli dekompozycji

### 2.5.1 Dla decompose (używamy transformacji Boxa-Coxa z lambda = 0)



Widać poprawę dopasowania dekompozycji addytywnej po zastosowaniu transformacji boxa-coxa

## Dekompozycja multiplikatywna z i addytywna z transformacją Boxa–Coxa

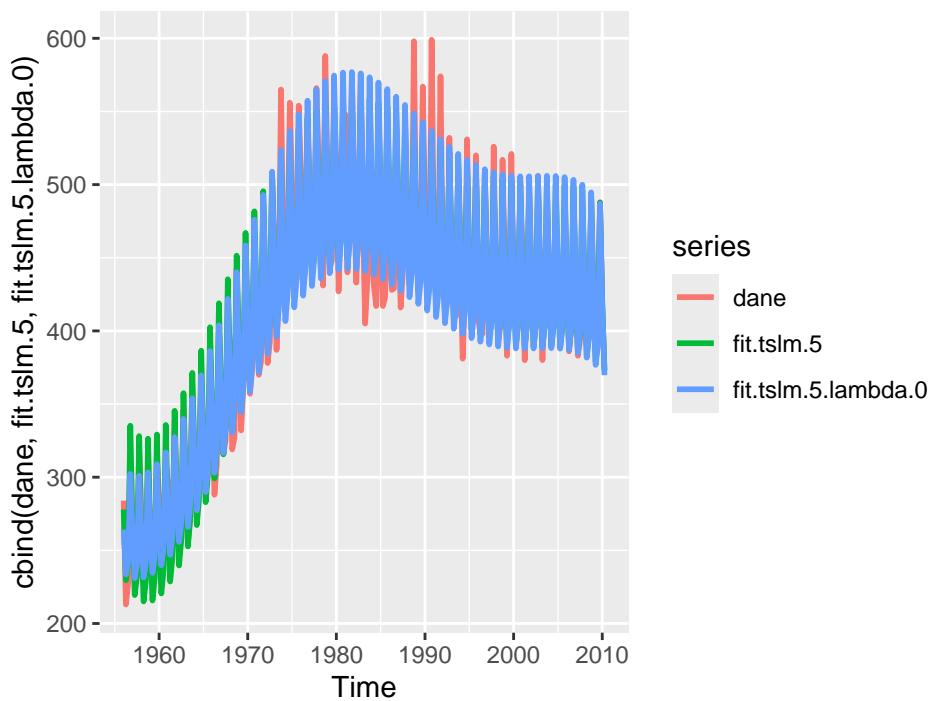
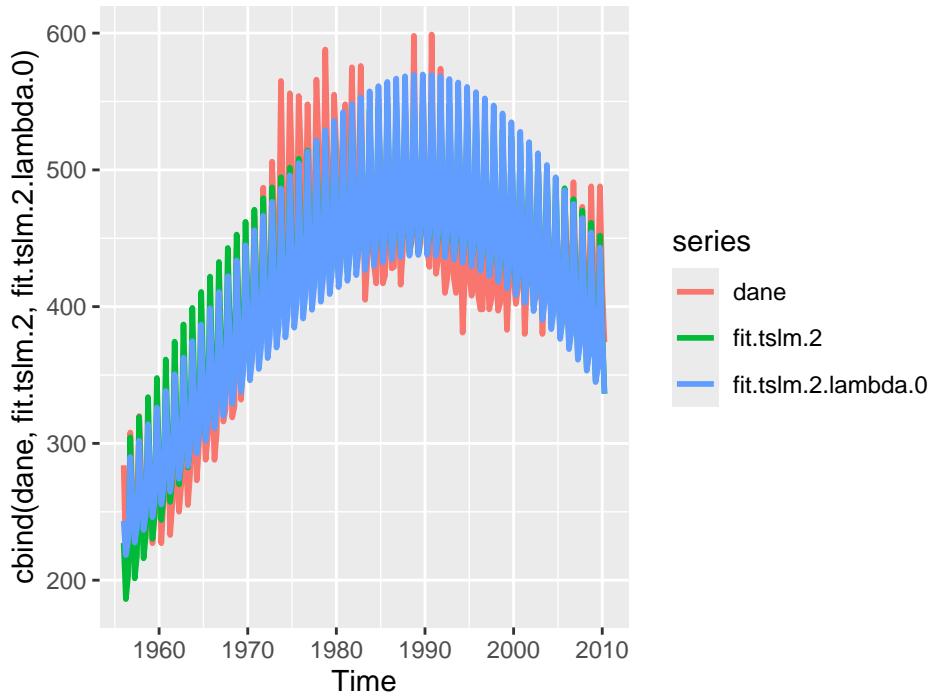


Zastosowanie transformacji Boxa–Coxa wyraźnie poprawia działanie dekompozycji addytywnej. W praktyce dekompozycja multiplikatywna prowadzi niemal do identycznych rezultatów jak sekwencja:

- 1) transformacja Boxa–Coxa z  $\lambda = 0$  (czyli logarytmowanie),
- 2) dekompozycja addytywna,
- 3) odwrotna transformacja Boxa–Coxa.

Oznacza to, że multiplikatywną strukturę sezonowości można skutecznie sprowadzić do postaci addytywnej za pomocą odpowiedniej transformacji.

### 2.5.2 Dla tslm (używamy transformacji Boxa-Coxa z lambda = 0)



Porównanie modeli regresyjnych bez transformacji i po zastosowaniu transformacji Boxa–

Coxa wskazuje, że wpływ tej transformacji jest silnie uzależniony od jakości samego modelu.

- **Model `tslm2` vs `tslm2 + Box-Cox`**

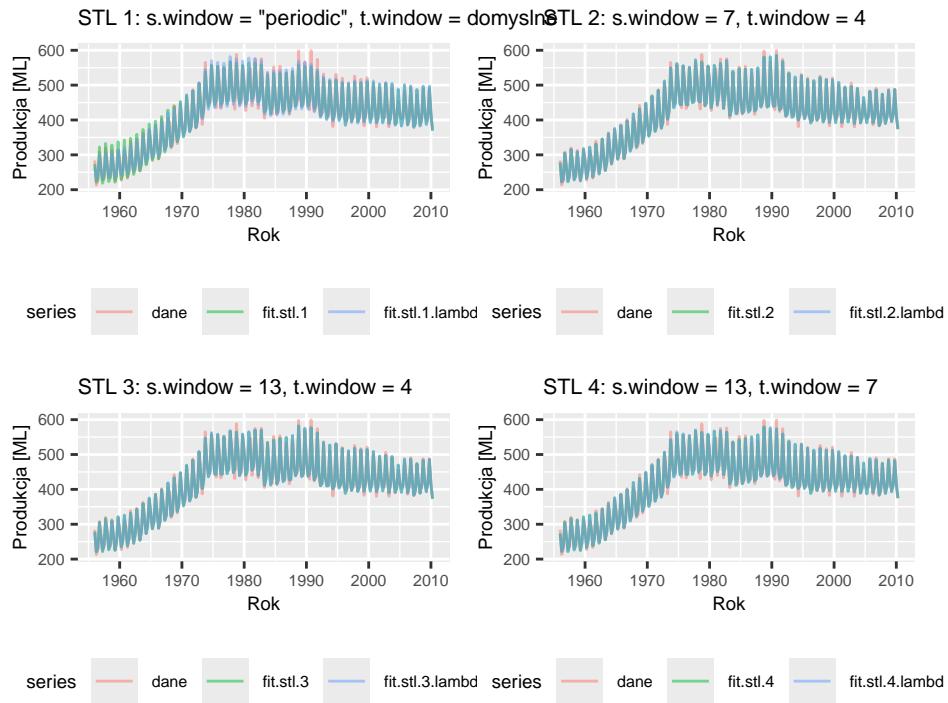
Wnioski są analogiczne jak dla `tslm1`: dodanie samej sezonowości nie wystarcza, aby uchwycić pełną dynamikę szeregu, a Box–Cox nie wpływa istotnie na poprawę dopasowania.

- **Model `tslm5` vs `tslm5 + Box-Cox`**

Dopiero w bardziej złożonym modelu, uwzględniającym zarówno sezonowość, jak i wielomianowy trend, transformacja Boxa–Coxa przynosi zauważalną poprawę. Model po transformacji lepiej odwzorowuje ogólny trend oraz cechy zmienności szeregu.

Podsumowując, transformacja Boxa–Coxa przynosi korzyści głównie w modelach, w których trend i sezonowość zostały już poprawnie zidentyfikowane. W prostszych modelach, niewystarczająco opisujących strukturę szeregu, jej wpływ jest minimalny

### 2.5.3 Dla stl (używamy transformacji Boxa-Coxa z lambda = 0)

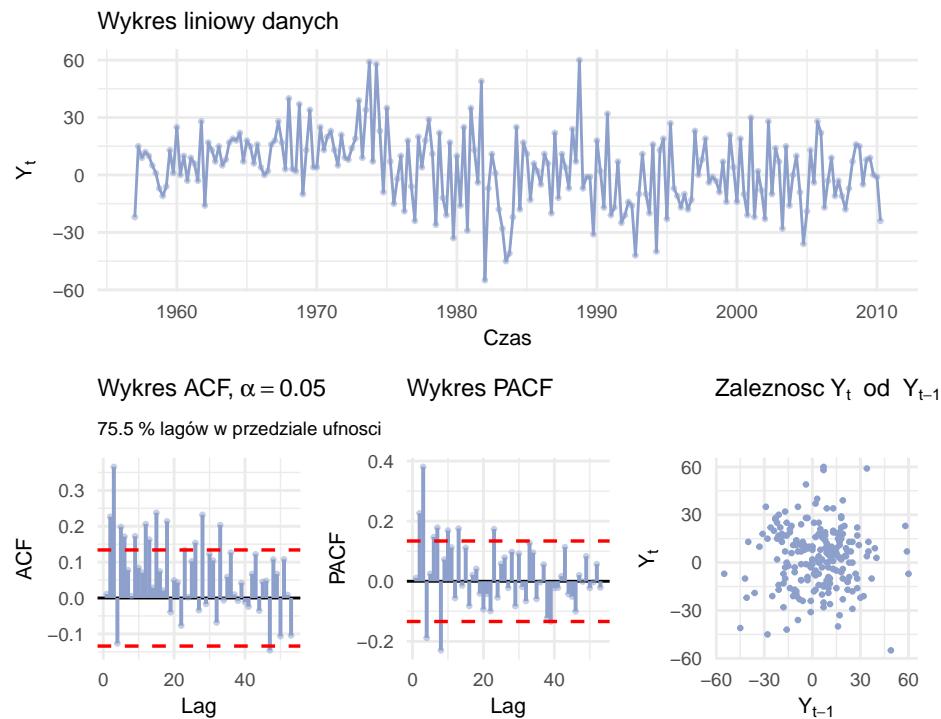


Widać poprawę dopasowania modeli poprzez zastosowanie transformacji Boxa-Coxa

Najlepiej dopasowane po transformacji są modele z s.window równym 13 lub 7 oraz t.window równym 4

## 2.6 Porównanie wyników eliminacji trendu i sezonowości na podstawie rozważanych metod dekompozycji z wynikami uzyskanymi poprzez odpowiednie różnicowanie danych

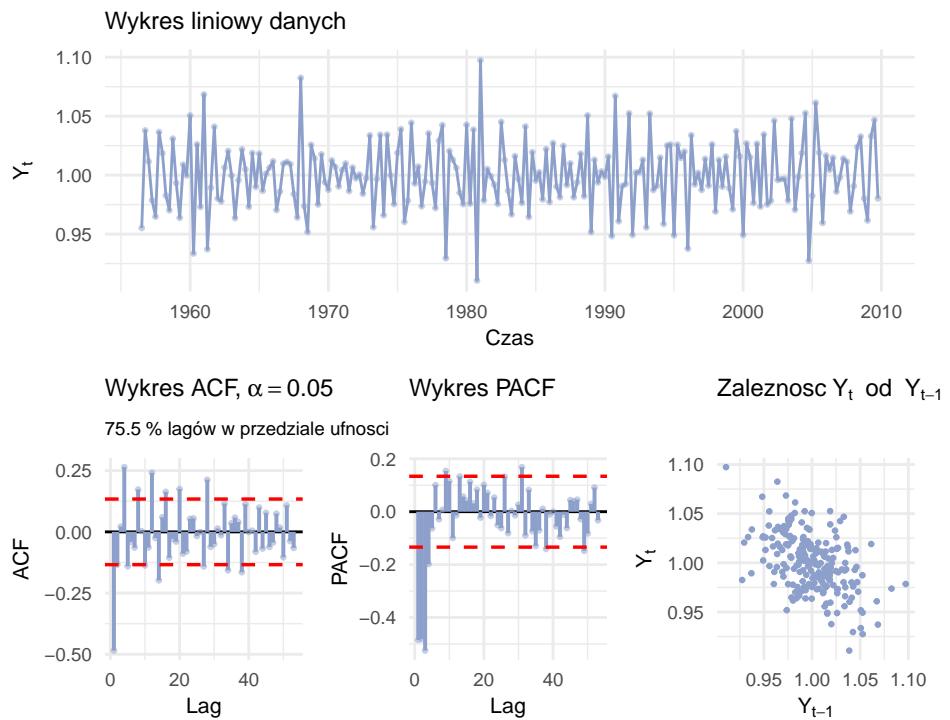
### 2.6.1 Dane zróżnicowane co 4



Na podstawie wykresu szeregu po różnicowaniu o 4 (usunięciu sezonowości) oraz odpowiadających mu funkcji ACF i PACF można stwierdzić, że:

- w przekształconym szeregu **nie obserwuje się już wyraźnego trendu ani regularnej sezonowości**, a wahania oscylują wokół stałej średniej
- około 75,5% wartości ACF** znajduje się w przedziale ufności – część lagów wciąż przekracza granice, ale nie tworzy wyraźnego, systematycznego wzorca, nie ma jednorodności wariancji

Podsumowując, szereg po różnicowaniu o 4 można uznać za **bliski stacjonarności**:

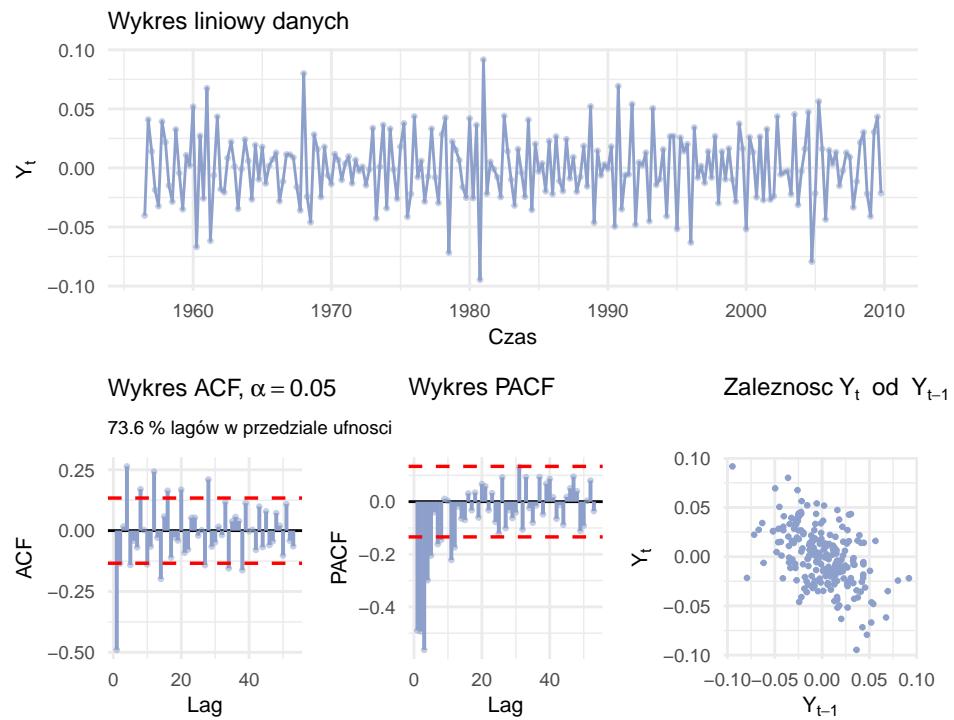


Na podstawie wykresu reszt po dekompozycji mnożnikowej oraz funkcji ACF i PACF można stwierdzić, że:

- reszty oscylują wokół stałej średniej, bez wyraźnego trendu ani sezonowości, co przemawia za **w przybliżeniu stacjonarnym charakterem** szeregu
- około 75% wartości ACF leży w przedziałach ufności, nie występują regularne wzorce oraz zachowana jest jednorodność wariancji. - jest to wyraźna poprawa względem danych pierwotnych, ale nadal obserwujemy kilka istotnych lagów, co oznacza utrzymującą się **słabą autokorelację**

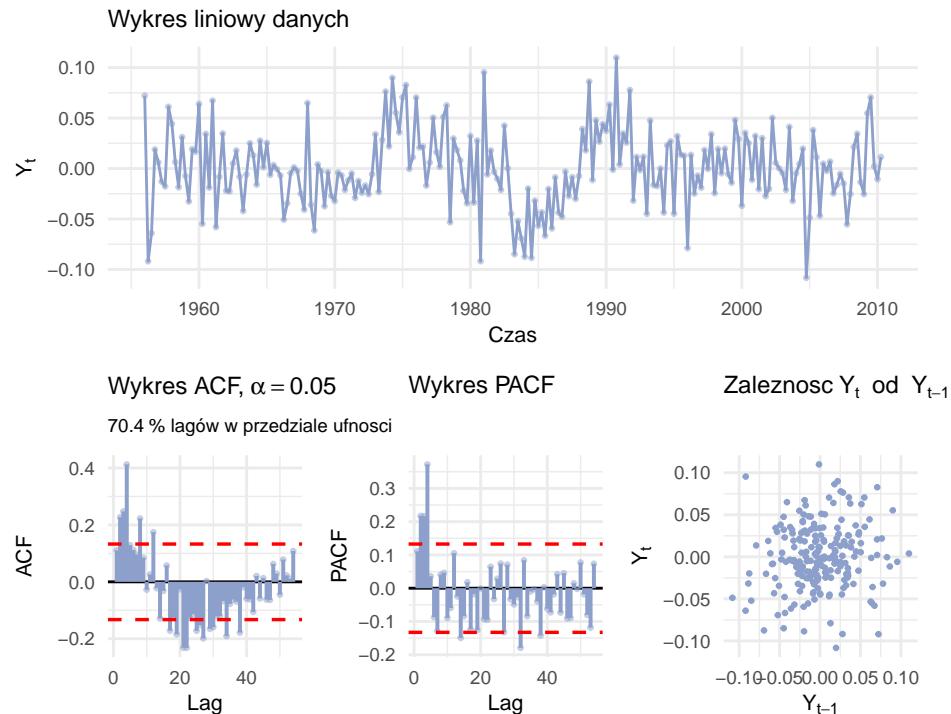
Wniosek: **reszty można uznać za w przybliżeniu stacjonarne**, ale nie spełniają one w pełni założeń białego szumu

## 2.6.2 Reszty z dekompozycji addytywnej z transformacją boxa-coxa



Wnioski analogiczne jak do poprzedniego wykresu.

### 2.6.3 Reszty z regresji wielomianem 5 stopnia i transformacją boxa-coxa

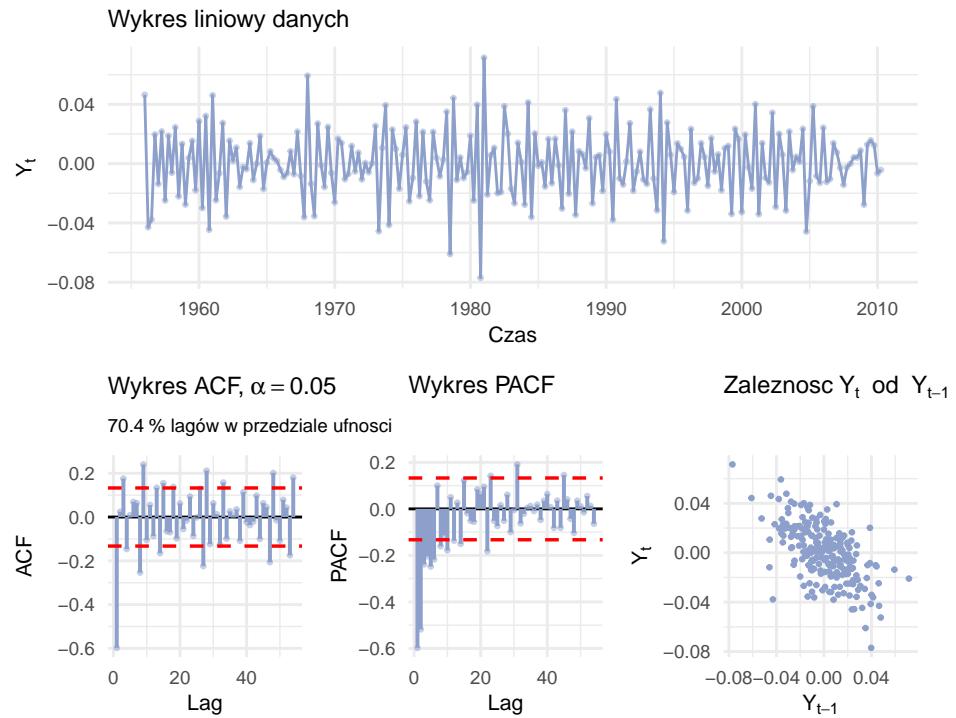


Na podstawie wykresu reszt oraz odpowiadających im funkcji ACF i PACF można stwierdzić, że:

- na wykresie liniowym reszty oscylują wokół zera, bez wyraźnego trendu ani sezonowości – w tym sensie szereg ma charakter zbliżony do stacjonarnego
- jednak tylko ok. 70% wartości ACF mieści się w przedziałach ufności, ale jest niejednorodność wariancji, a kilka lagów jest wyraźnie odstających

W konsekwencji **nie można jednoznacznie uznać reszt za biały szum** – szereg nie zawiera już trendu ani silnej sezonowości, ale zachowana jest niewielka autokorelacja.

#### 2.6.4 Reszty z stl plus Box-Cox



Pomimo tylko 70% danych w przedziale ufności, reszty oscylują wokół 0. Nie widać żadnej sezonowości ACF i jest tylko jedna obserwacja znacznie wychodząca poza przedział ufności, mamy również jednorodność wariancji, co skłania do skategoryzowania reszty jako szeregu stacjonarnego.