

Sprawozdanie 1

Mateusz Cieślak, Michał Merta

2025-11-23

Spis treści

1 Zadanie 3 - Testowanie białoszumowości	2
1.1 Krótki opis zagadnienia	2
1.2 Opis eksperymentów	2
1.3 Wyniki	4
1.3.1 Przykłady działania testów białoszumowości	4
1.3.2 Porównanie metody graficznej i formalnych testów statystycznych	10
1.4 Podsumowanie	12
2 Zadanie 4 - Dekompozycja szeregów czasowych - eliminacja trendu i sezonowości	13
2.1 Wstęp	13
2.2 Metody graficzne	13
2.3 Metody dekompozycji	15
2.3.1 Dekompozycja na podstawie ruchomej średniej	15
2.3.2 Dekompozycja na podstawie modelu regresji	17
2.3.3 Dekompozycja STL oparta na metodzie loess	20
2.4 Czy zastosowanie transformacji Boxa-Coxa prowadzi do poprawy jakości dopasowania modeli dekompozycji?	25
2.5 Eliminacja trendu poprzez różnicowanie danych	26
2.5.1 Szereg różnic po użyciu decompose	27
2.5.2 Szereg różnic po użyciu tslm	28
2.5.3 Szereg różnic po użyciu stl	29
2.5.4 Szereg różnic otrzymany po różnicowaniu	30
2.6 Podsumowanie i wnioski końcowe	30

Spis wykresów

1 Wykres szeregu czasowego oraz ACF dla szumu białego z rozkładu normalnego.	4
2 Wykres szeregu czasowego oraz ACF dla szumu białego z rozkładu wykładniczego	5
3 Wykres szeregu czasowego oraz ACF dla próbki modelu MA(1)	6
4 Wykres szeregu czasowego oraz ACF dla próbki błędzenia losowego	7
5 Wykres szeregu czasowego oraz ACF dla próbki szumu białego z nałożonym trendem deterministycznym	8
6 Wykres szeregu czasowego oraz ACF dla próbki szumu białego z nałożoną sezonowością.	9
7 Wykres przedstawiający szereg czasowy bezrobocia	13
8 Wykres sezonowy dla bezrobocia	14
9 Wykres podszeregów dla bezrobocia	14
10 Wykres reszt	15
11 Wykres dekompozycji addytywnej szeregu czasowego	15
12 Dopasowanie addytywnego modelu dekompozycji	16
13 Wykres dekompozycji mnożystywnej szeregu czasowego	16

14	Dopasowanie multiplikatywnego modelu dekompozycji	17
15	Dekompozycja na podstawie modelu regresji	17
16	Dekompozycja na podstawie modelu regresji	18
17	Dekompozycja na podstawie modelu regresji	18
18	Porównanie dopasowań	19
19	Dekompozycja na podstawie modelu regresji dla wyższych stopni	19
20	Dekompozycja na podstawie modelu regresji dla wyższych stopni	20
21	Porównanie dopasowań dla wyższych stopni	20
22	Dekompozycja STL dla domyślnych parametrów	21
23	Dekompozycja STL dla domyślnych parametrów	21
24	Dekompozycja STL dla domyślnych parametrów	22
25	Dekompozycja STL dla s.window=7, t.window - domyślne	22
26	Dekompozycja STL dla s.window=13, t.window - domyślne	23
27	Dekompozycja STL dla s.window=13, t.window=21	23
28	Dekompozycja STL dla s.window=13, t.window=7	24
29	Dekompozycja STL dla s.window=7, t.window=7	24
30	Dekompozycja STL dla s.window - periodic, t.window=7	25
31	Szereg reszt dla metody decompose	27
32	ACF dla metody decompose	27
33	Szereg reszt dla metody tslm	28
34	Szereg reszt dla metody stl	29
35	Szereg reszt po różnicowaniu	30

Spis tabel

1	Decyzje testów biało-szumowości dla realizacji z rozkładu normalnego	4
2	Decyzje testów biało-szumowości dla realizacji szumu białego z rozkładu wykładniczego	5
3	Decyzje testów biało-szumowości dla realizacji modelu MA(1)	6
4	Decyzje testów biało-szumowości dla realizacji błędzenia losowego	7
5	Decyzje testów biało-szumowości dla realizacji szumu białego z trendem deterministycznym	8
6	Decyzje testów biało-szumowości dla realizacji szumu białego z sezonowością	9
7	Proporcja odrzuceń hipotezy H_0 dla białego szumu, w zależności od rozkładu	10
8	Proporcja odrzuceń hipotezy H_0 dla białego szumu o różnej długości szeregu (n)	10
9	Proporcja odrzuceń hipotezy H_0 dla różnych wartości maksymalnego opóźnienia (h_{\max})	11
10	Proporcja odrzuceń hipotezy H_0 dla MA(1)	11
11	Proporcja odrzuceń hipotezy H_0 dla szeregu z trendem deterministycznym	11
12	Proporcja odrzuceń hipotezy H_0 dla błędzenia losowego	11
13	Proporcja odrzuceń hipotezy H_0 dla szeregu z sezonowością	12
14	Proporcja odrzuceń hipotezy H_0 dla mapowania logistycznego	12
15	Porównanie RMSE dla różnych metod dekompozycji przed i po Box-Cox	26

1 Zadanie 3 - Testowanie białoszumowości

1.1 Krótki opis zagadnienia

W tym zadaniu zajmujemy się problemem weryfikacji hipotezy, że obserwowany szereg czasowy jest białym szumem. Biały szum rozumiemy jako sekwencję niezależnych zmiennych losowych. W szczególności, jego autokorelacja teoretyczna spełnia $\rho(h) = 0$ dla $h \neq 0$.

Celem pracy jest porównanie efektywności podejścia graficznego oraz formalnego, w oparciu o symulacje komputerowe. W szczególności zbadamy wpływ:

- długości szeregu czasowego,
- różnych rozkładów generujących białe szumy,
- wyboru maksymalnego opóźnienia h_{\max} ,
- rozważenia szeregu nienależących do klasy i.i.d., takich jak procesy MA(1), błędzenie losowe, szeregi z trendem deterministycznym lub sezonowością.

Wyniki symulacji pozwolą ocenić, w jakich warunkach poszczególne metody testowania białoszumowości są najbardziej wiarygodne i jakie ograniczenia praktyczne wiążą się z ich stosowaniem.

1.2 Opis eksperymentów

W celu porównania metody graficznej weryfikacji hipotezy, że szereg czasowy jest białym szumem z testami formalnymi musimy najpierw dostosować metodę graficzną do wielokrotnych symulacji. W tym celu należy ją nieco sformalizować. W przedstawionej funkcji hipotezę zerową H_0 , że szereg jest białym szumem, odrzucamy, jeśli spełnione jest przynajmniej jedno z dwóch kryteriów:

1. Kryterium pasa ufności dla autokorelacji:

Dla każdej próbki wyznaczamy granice ufności $CI = \pm z_{1-\alpha/2} / \sqrt{n}$, gdzie n jest długością szeregu. Jeśli proporcja wartości próbkowej funkcji autokorelacji (ACF) znajdujących się poza tym pasmem przekracza poziom istotności α , uznajemy to za przesłankę do odrzucenia H_0 .

2. Kryterium reguły 3 sigm:

Wartości ACF, które przekraczają teoretyczny próg $3/\sqrt{n}$, traktujemy jako istotnie odstające. Wystąpienie przynajmniej jednej takiej wartości powoduje odrzucenie H_0 .

Kod zaimplementowanej funkcji znajduje się poniżej:

```
test_white_noise_autocorrelation_graphic <- function(x, alpha = 0.05, hmax = NULL)
{
  n <- length(x)

  #Przygotowanie argumentów dla ACF
  acf_args <- list(x = x, plot = FALSE)
  if (!is.null(hmax)) {
    acf_args$lag.max <- hmax
  }

  #Obliczenie ACF
  acf_obj <- do.call(acf, acf_args)
  acf_values <- acf_obj$acf[-1]

  #Kryterium Pasa Ufności (alpha)
  ci <- qnorm(1 - alpha/2) / sqrt(n)

  outside_ci <- which(abs(acf_values) > ci)
  prop_outside <- length(outside_ci) / length(acf_values)
  reject_ci <- prop_outside > alpha
```

```

#Kryterium Reguły 3 Sigm

outlier_3sigma <- which(abs(acf_values) > 3 / sqrt(n))
reject_3sigma <- length(outlier_3sigma) > 0

#Ogólna Decyzja
reject_H0_graphic <- reject_ci | reject_3sigma

#Zwracamy pojedynczą wartość logiczną
return(list(
  rejected = reject_H0_graphic,
  reject_ci = reject_ci,
  reject_3sigma = reject_3sigma
))
}

```

Oprócz tego używać będziemy wbudowanych funkcji do formalnych testów weryfikujących tą samą hipotezę, są to:

Test Box–Pierce'a, który opiera się na statystyce testowej Q :

$$Q = n \sum_{h=1}^H \hat{\rho}_h^2$$

gdzie:

- n to długość szeregu czasowego.
- $\hat{\rho}_h$ to próbkowy współczynnik autokorelacji dla opóźnienia h .
- H to maksymalne opóźnienie brane pod uwagę w teście.

Test Ljung–Box'a, który jest powszechnie stosowaną modyfikacją testu Box–Pierce'a. Statystyka testowa Q^* jest definiowana jako:

$$Q^* = n(n + 2) \sum_{h=1}^H \frac{\hat{\rho}_h^2}{n - h}$$

W obu testach, decyzja o odrzuceniu hipotezy zerowej jest podejmowana na podstawie wartości p : jeżeli wartość p jest mniejsza niż przyjęty poziom istotności α , odrzucamy H_0 i wnioskujemy, że w szeregu występuje istotna autokorelacja. W symulacjach do wykonania tych testów wykorzystano wbudowaną w R funkcję `Box.test`.

W celu porównania testów, będziemy wielokrotnie generować szeregi czasowe o różnym pochodzeniu oraz sprawdzać decyzję o odrzuceniu hipotezy zerowej dla każdego z nich. Wykorzystamy do tego wartość % odrzuceń dla każdej metody. W szczególności zwróciśmy uwagę na:

- **Różne rozkłady:** Badania obejmą biały szum generowany z różnych rozkładów: (Normalny ($N(0, 1)$), Binarny ($\mathbb{P}(X = 1) = 0.5$) oraz Wykładniczy ($Exp(1)$)).
- **Różną długość szeregów:** Porównane zostaną trzy długości szeregów: $n = 25$, $n = 100$ oraz $n = 500$.
- **Różny wybór maksymalnego opóźnienia h_{\max} :** Chociaż główna analiza zostanie przeprowadzona przy standardowym opóźnieniu, w celu oceny wpływu tego parametru na wyniki, zostaną rozważone też inne warianty maksymalnego opóźnienia.
- **Przykłady szeregów czasowych innych niż szum i.i.d.:** W celu oceny mocy testów (zdolności do odrzucenia fałszywej H_0), uwzględnione zostaną szeregi wykazujące:

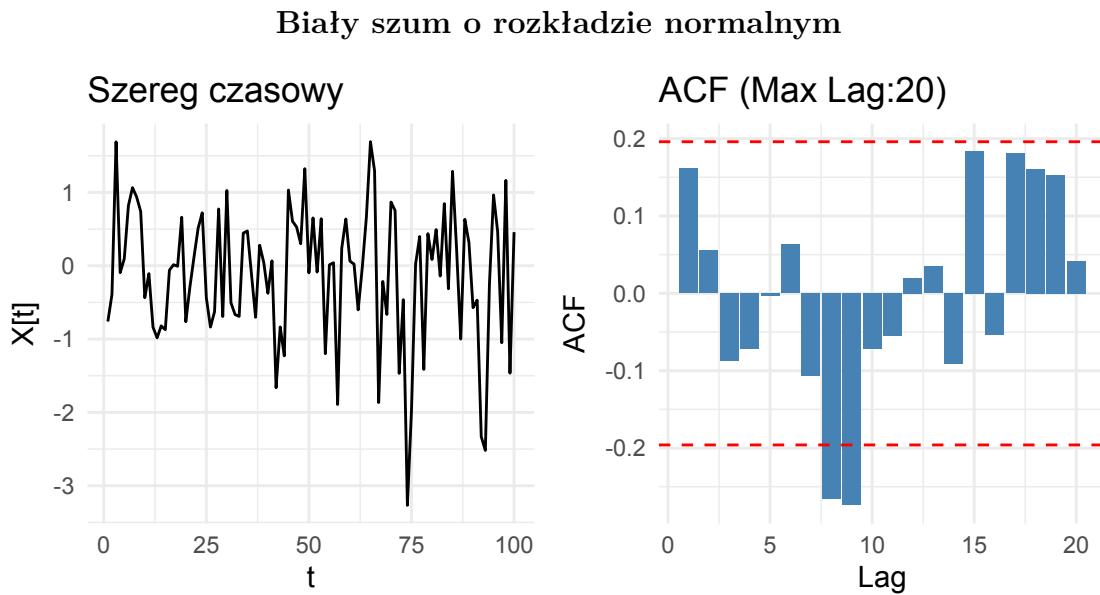
- Niewielką korelację: Model stacjonarny MA(1) ($\theta = 0.5$).
- Niestacjonarność/silną korelację: Błądzenie Losowe.
- Strukturę Deterministyczną: Szum IID z Trendem Liniowym oraz Szum IID z Sezonowością.
- Szereg chaotyczny generowany przez mapowanie logistyczne.

Symulacje zostaną przeprowadzone z przyjętym poziomem istotności $\alpha = 0.05$.

1.3 Wyniki

1.3.1 Przykłady działania testów białośzumowości

W tej części zadania przystępujemy do praktycznej weryfikacji hipotezy o białośzumowości na podstawie asymptotycznych własności autokorelacji na kilku przykładach. Celem jest wizualna ilustracja działania testu graficznego oraz porównanie jego mechanizmu z wynikami formalnych testów: Boxa–Pierce'a i Ljunga–Boxa. W metodzie graficznej sprawdzamy, czy nie jest zbyt dużo obserwacji wychodzących poza przyjęte normy oraz czy nie ma żadnej obserwacji istotnie wychodzącej.



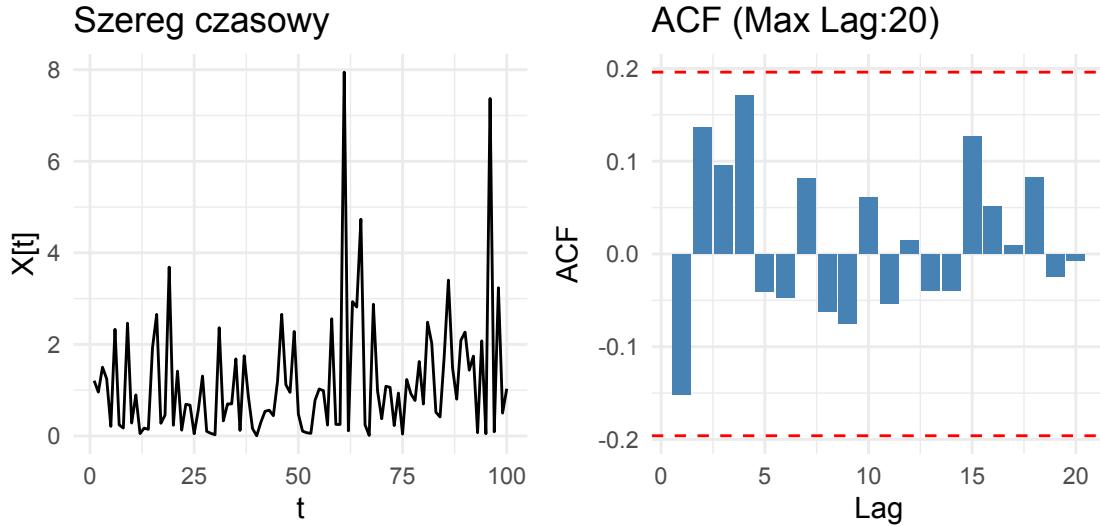
Wykres 1: Wykres szeregu czasowego oraz ACF dla szumu białego z rozkładu normalnego.

Test	H0 odrzucona
Graficzny	TRUE
Box-Pierce	FALSE
Ljung-Box	FALSE

Tabela 1: Decyzje testów biało-szumowości dla realizacji z rozkładu normalnego

Na podstawie Rysunku 1, widać, że aż dwie wartości ACF są poza ustalonymi przedziałami. Z racji, że jest to aż 10% danych ACF, odrzucamy hipotezę zerową. W tabeli 1 widzimy, że sformalizowana wersja metody graficznej również odrzuca Hipotezę zerową, natomiast formalne testy statystyczne słusznie nie podejmują decyzji o jej odrzuceniu, zatem są odporne na tego typu fluktuacje.

Biały szum o rozkładzie wykładniczym



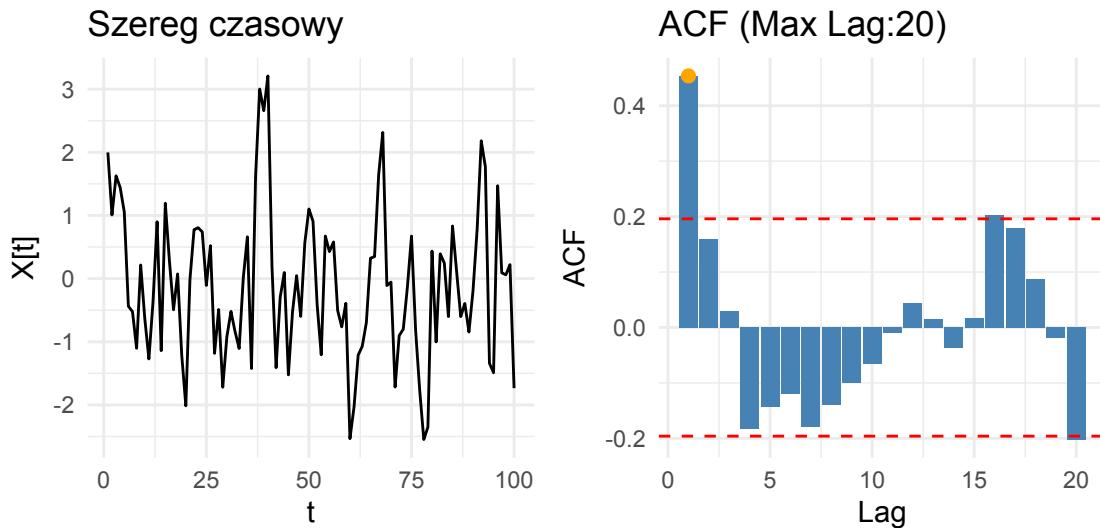
Wykres 2: Wykres szeregu czasowego oraz ACF dla szumu białego z rozkładu wykładniczego

Test	H0 odrzucona
Graficzny	FALSE
Box-Pierce	FALSE
Ljung-Box	FALSE

Tabela 2: Decyzje testów biało-szumowości dla realizacji szumu białego z rozkładu wykładniczego

Na podstawie Rysunku 2, widać, że żadna wartość ACF nie jest poza ustalonymi przedziałami, zatem nie mamy podstaw do odrzucenia hipotezy zerowej. W tabeli 2 widzimy, że sformalizowana wersja metody graficznej oraz formalne metody statystyczne również jej nie odrzucają.

Model MA(1)



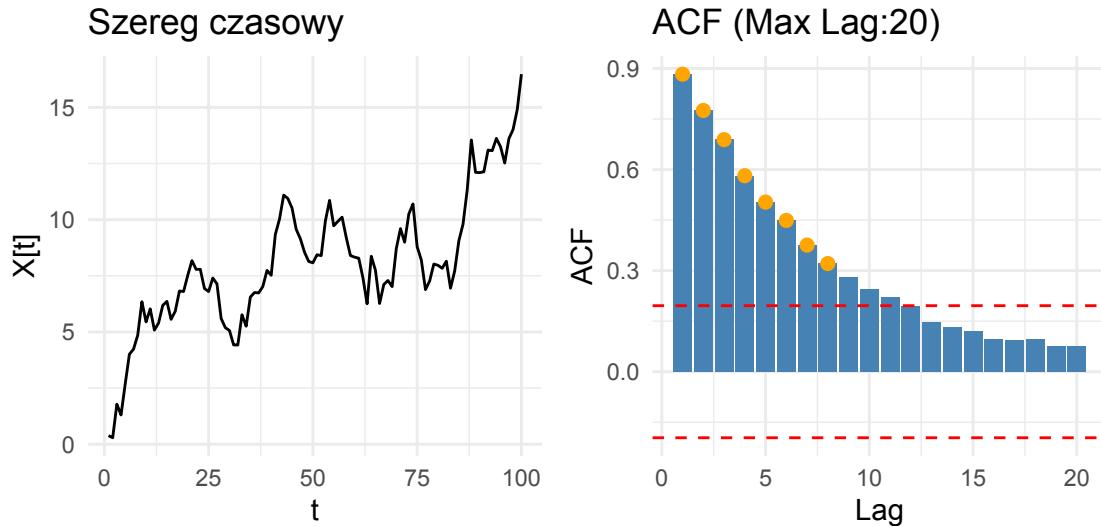
Wykres 3: Wykres szeregu czasowego oraz ACF dla próbki modelu MA(1)

Test	H0 odrzucona
Graficzny	TRUE
Box-Pierce	TRUE
Ljung-Box	TRUE

Tabela 3: Decyzje testów biało-szumowości dla realizacji modelu MA(1)

Na podstawie Rysunku 3, widać, że jedna wartość ACF jest poza ustalonymi przedziałami, ale z racji, że jest ona “istotnie” wychodząca, odrzucamy hipotezę zerową. W tabeli 3 widzimy, że sformalizowana wersja metody graficznej oraz formalne metody również odrzucają hipotezę H0.

Błędzenie losowe

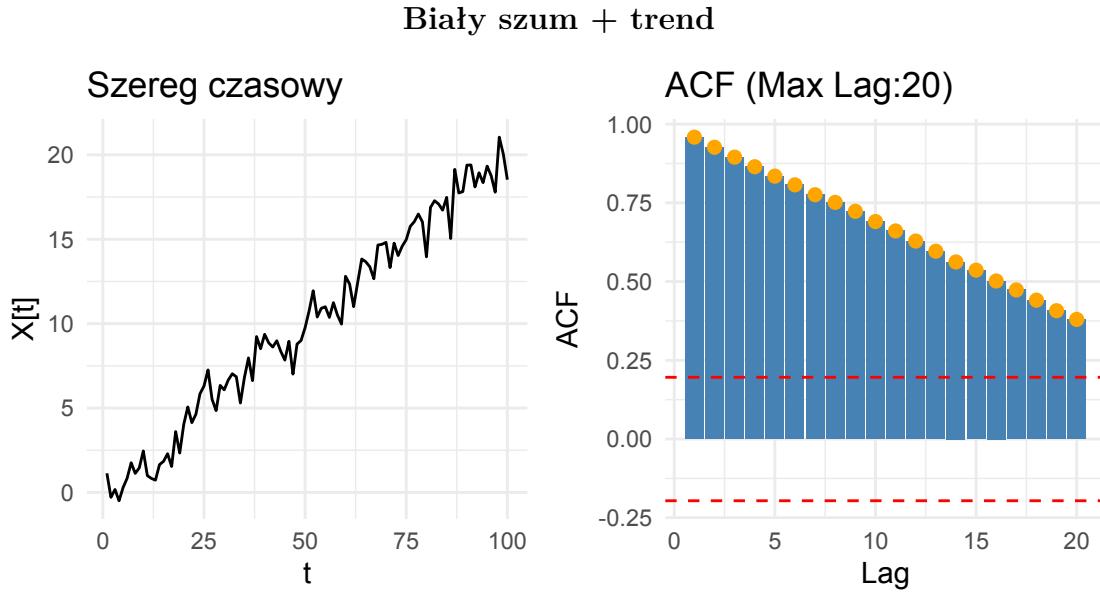


Wykres 4: Wykres szeregu czasowego oraz ACF dla próbki błędzenia losowego

Test	H0 odrzucona
Graficzny	TRUE
Box-Pierce	TRUE
Ljung-Box	TRUE

Tabela 4: Decyzje testów biało-szumowości dla realizacji błędzenia losowego

Na podstawie Rysunku 4, widać, że większość wartości ACF jest poza ustalonymi przedziałami, ponadto kilka z nich jest “istotnie” wychodzących, z tego powodu z pewnością odrzucamy zatem hipotezę zerową. W tabeli 4 widzimy, że sformalizowana wersja metody graficznej oraz formalne metody statystyczne również podejmują decyzję o odrzuceniu hipotezy H0.



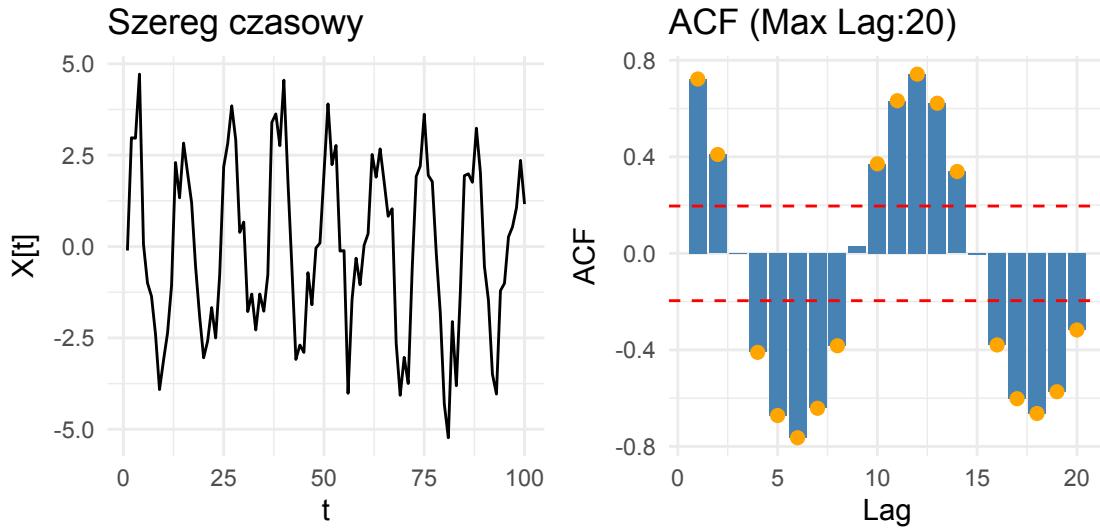
Wykres 5: Wykres szeregu czasowego oraz ACF dla próbki szumu białego z nałożonym trendem deterministycznym

Test	H0 odrzucona
Graficzny	TRUE
Box-Pierce	TRUE
Ljung-Box	TRUE

Tabela 5: Decyzje testów biało-szumowości dla realizacji szumu białego z trendem deterministycznym

Na podstawie Rysunku 5, widać, że wszystkie wartości ACF są poza ustalonymi przedziałami, ponadto wszystkie są “istotnie” wychodzące, zatem z pewnością odrzucamy zatem hipotezę zerową. W tabeli 5 widzimy, że sformalizowana wersja metody graficznej oraz formalne metody również podejmują decyzję o odrzuceniu hipotezy H0.

Biały szum + Sezonowość



Wykres 6: Wykres szeregu czasowego oraz ACF dla próbki szumu białego z nałożoną sezonowością.

Test	H0 odrzucona
Graficzny	TRUE
Box-Pierce	TRUE
Ljung-Box	TRUE

Tabela 6: Decyzje testów biało-szumowości dla realizacji szumu białego z sezonowością

Na podstawie Rysunku 6, widać, że duża część wartości ACF jest poza ustalonymi przedziałami, zatem odrzucamy hipotezę zerową. W tabeli 6 widzimy, że sformalizowana wersja metody graficznej oraz formalne metody również podejmują decyzję o odrzuceniu hipotezy H0.

1.3.2 Porównanie metody graficznej i formalnych testów statystycznych

W tej części zadania zajmiemy się porównaniem metod testowania białoszumowości. W tym celu przeprowadzimy wielokrotne symulacje dla różnych przypadków pochodzenia danych oraz dla różnych parametrów. Dla odpowiedniej analizy każda symulacja zostanie przeprowadzona 1000 razy.

Różne rozkłady

W pierwszej kolejności sprawdzimy jak zachowują się testy dla białego szumu w zależności od rozkładu, z którego pochodzi. W wszystkich przypadkach przyjmujemy $n=100$ oraz nie ustalamy h_{\max} (pozwalamy na ustalenie go wbudowanym funkcjom).

Test	Normalny	Wykładniczy	Binarny
Graficzny	0.273	0.263	0.273
Box-Pierce	0.047	0.031	0.054
Ljung-Box	0.049	0.033	0.057

Tabela 7: Proporcja odrzuceń hipotezy H_0 dla białego szumu, w zależności od rozkładu

Jak widzimy w tabeli 7, Metoda graficzna niezależnie od rozkładu bardzo często odrzuca hipotezę zerową w około 26% przypadków (pomimo jej prawdziwości). Znacznie lepiej radzą sobie z tym testy formalne, które odrzucają ją tylko w okolicach poziomu istotności, tzn 5%.

Różne n

Teraz sprawdzimy, jak reagują testy na różne długości szeregów. W każdym z przypadków rozważymy biały szum pochodzący z rozkładu normalnego oraz domyślne h_{\max} .

Test	n=20 (Krótki)	n=100 (Średni)	n=500 (Długi)
Graficzny	0.145	0.178	0.325
Box-Pierce	0.034	0.038	0.045
Ljung-Box	0.053	0.040	0.046

Tabela 8: Proporcja odrzuceń hipotezy H_0 dla białego szumu o różnej długości szeregu (n)

Jak widzimy, dla dużych n test graficzny coraz częściej niesłusznie odrzuca hipotezę H_0 . Jest to prawdopodobnie spowodowane faktem, że dla dużych n jest coraz to więcej wartości ACF do analizy, przez co częściej któraś z nich może przez przypadek istotnie wykroczyć poza dopuszczany przedział. Natomiast dwa formalne testy statystyczne wykazują się dużą stabilnością dla dowolnych n .

Różne h_{\max}

Teraz przetestujemy, jak zachowują się funkcje dla różnych dobranych maksymalnych wartości h_{\max} . W każdym z przypadków analizować będziemy szereg długości $n = 100$ typu biały szum o rozkładzie normalnym. Sprawdzimy:

- domyślne h_{\max} (czyli $10 \cdot \log_{10} n$),
- $h_{\max} = 25$ (czyli $n/4$),
- $h_{\max} = 10$.

Test	Domyślne	$h_{\max} = 25$	$h_{\max} = 10$
Graficzny	0.170	0.212	0.362
Box-Pierce	0.038	0.027	0.046
Ljung-Box	0.041	0.084	0.069

Tabela 9: Proporcja odrzuceń hipotezy H_0 dla różnych wartości maksymalnego opóźnienia (h_{\max})

Jak widać w tabeli 9 test graficzny radzi sobie najlepiej dla domyślnego h_{\max} . Natomiast testy formalne lekko tracą stabilność dla h_{\max} innych niż domyślne (wyniki odbiegają od poziomu istotności).

Szeregi inne niż biały szum

Model MA(1)

Test	Odrzucenia	Proporcja_Odrzuceń
Graficzny	938	0.938
Box-Pierce	987	0.987
Ljung-Box	988	0.988

Tabela 10: Proporcja odrzuceń hipotezy H_0 dla MA(1).

jak widać w Tabeli 10 Formalne metody prawie zawsze wykrywają brak białego szumu, pomimo wysokiej autokorelacji tylko dla pierwszego ACF. Metoda graficzna również jest cechującą się natomiast nieco niższą skutecznością.

Szum losowy + trend

Test	Odrzucenia	Proporcja_Odrzuceń
Graficzny	1000	1.000
Box-Pierce	1000	1.000
Ljung-Box	1000	1.000

Tabela 11: Proporcja odrzuceń hipotezy H_0 dla szeregu z trendem deterministycznym

Jak widać w Tabeli 11 każda z metod z 100% skutecznością odrzuca hipotezę zerową, co jest logiczne w przypadku występującego istotnego trendu.

Błądzenie losowe

Test	Odrzucenia	Proporcja_Odrzuceń
Graficzny	1000	1.000
Box-Pierce	1000	1.000
Ljung-Box	1000	1.000

Tabela 12: Proporcja odrzuceń hipotezy H_0 dla błądzenia losowego

Jak widać w Tabeli 12 ponownie każda z metod z 100% skutecznością odrzuca hipotezę zerową. W przypadku błądzenia losowego autokorelacja jest wysoka i wynik ten jest zgodny z teorią.

Szum losowy + sezonowość

Test	Odrzucenia	Proporcja_Odrzuceń
Graficzny	1000	1.000
Box-Pierce	1000	1.000
Ljung-Box	1000	1.000

Tabela 13: Proporcja odrzuceń hipotezy H_0 dla szeregu z sezonowością

Ponownie w Tabeli 13 zauważamy 100% skuteczność każdej z metod. Sezonowość zgodnie z teorią mocno wpływa na wartości ACF, co powoduje słuszne odrzucanie hipotezy zerowej.

mapowanie logistyczne

Test	Odrzucenia	Proporcja_Odrzuceń
Graficzny	196	0.196
Box-Pierce	58	0.058
Ljung-Box	63	0.063

Tabela 14: Proporcja odrzuceń hipotezy H_0 dla mapowania logistycznego

W tym przypadku dane pochodzą z rozkładu w teorii deterministycznego, jednak bardzo mocno przypominającego biały szum. Z tego powodu wyniki w Tabeli 14 są podobne do wyników z zwykłego białego szumu.

1.4 Podsumowanie

Na podstawie powyższej analizy można śmiało wysunąć wniosek, że formalne testy statystyczne są znacznie skuteczniejsze w przypadku testowania białoszumowości, niż metoda graficzna. Dodatkowo można wysnuć wnioski, że:

- Testy formalne odrzucają hipotezę zerową w przybliżeniu zgodnym z założonym poziomem istotności α , podczas gdy metoda graficzna jest subiektywna i nie gwarantuje kontrolowanego błędu pierwszego rodzaju.
- Formalne testy pozostają stabilne niezależnie od długości szeregu czasowego, natomiast metoda graficzna staje się nadmiernie czuła przy dużych próbach, co może prowadzić do błędnych wniosków.

2 Zadanie 4 - Dekompozycja szeregów czasowych - eliminacja trendu i sezonowości

2.1 Wstęp

Cel zadania: Zastosowanie i porównanie wybranych metod dekompozycji, w kontekście ich zastosowania do eliminacji trendu i sezonowości. W niniejszym sprawozdaniu będziemy przeprowadzać analizę sezonowości szeregu **bezrobocie** z biblioteki **TSAFBook**.

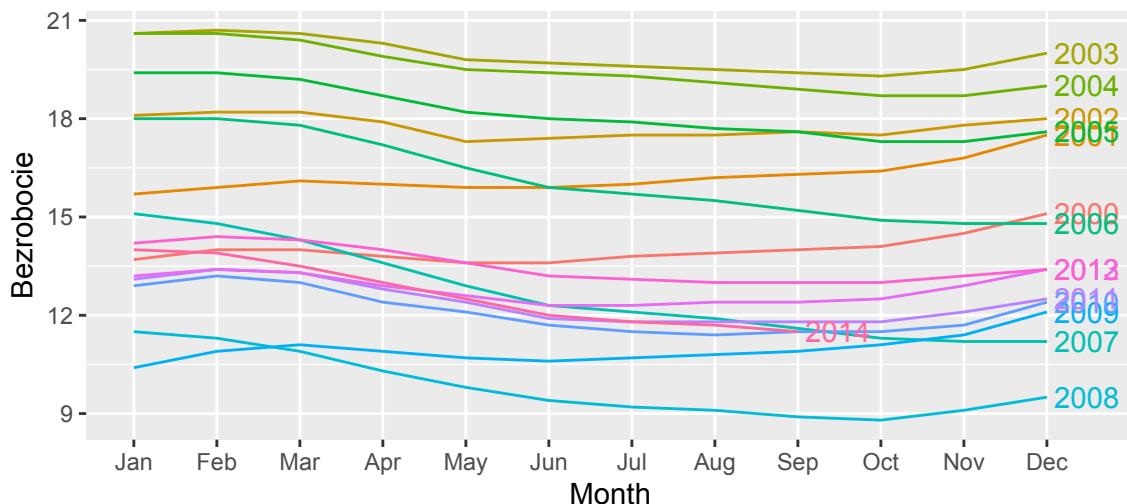
2.2 Metody graficzne



Wykres 7: Wykres przedstawiający szereg czasowy bezrobocia

Na wykresie 7 nie obserwujemy jednolitego trendu. W pierwszych latach poziom bezrobocia rośnie, następnie znacznie maleje, a w końcowym okresie ponownie rośnie. Oznacza to, że szereg wykazuje zmienny, nieliniowy charakter.

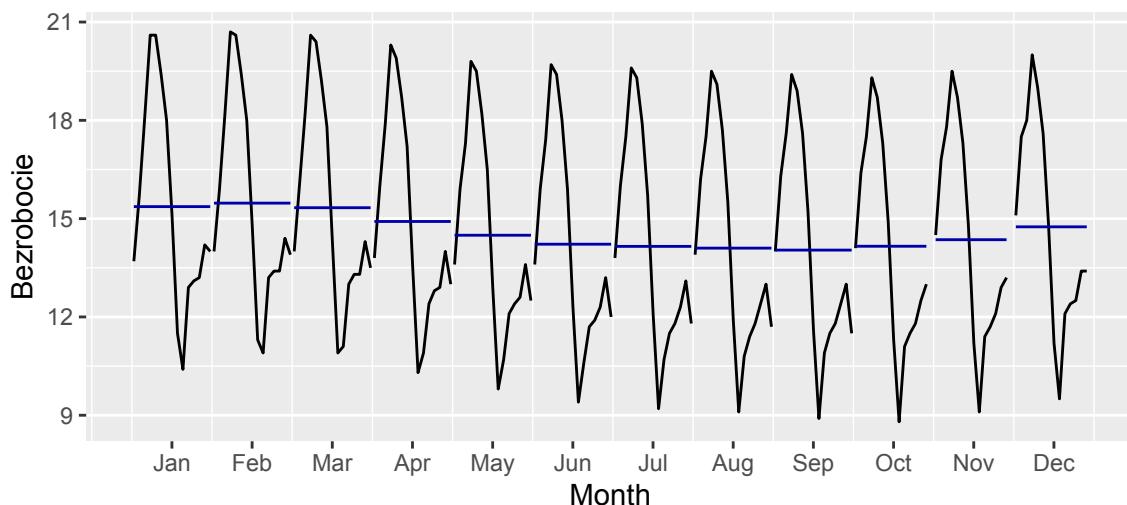
Wykres sezonowy dla bezrobocia



Wykres 8: Wykres sezonowy dla bezrobocia

Wykres 8 pozwala zauważyć, że każdego roku bezrobocie zmienia się bardzo podobnie, jest wyższe od grudnia do marca i szczególnie niższe od czerwca do października. Jest to zgodne z cyklem zapotrzebowania na pracę w sektorach silnie zależnych od pory roku, takich jak rolnictwo, ogrodnictwo czy turystyka, które zimą ograniczają działalność.

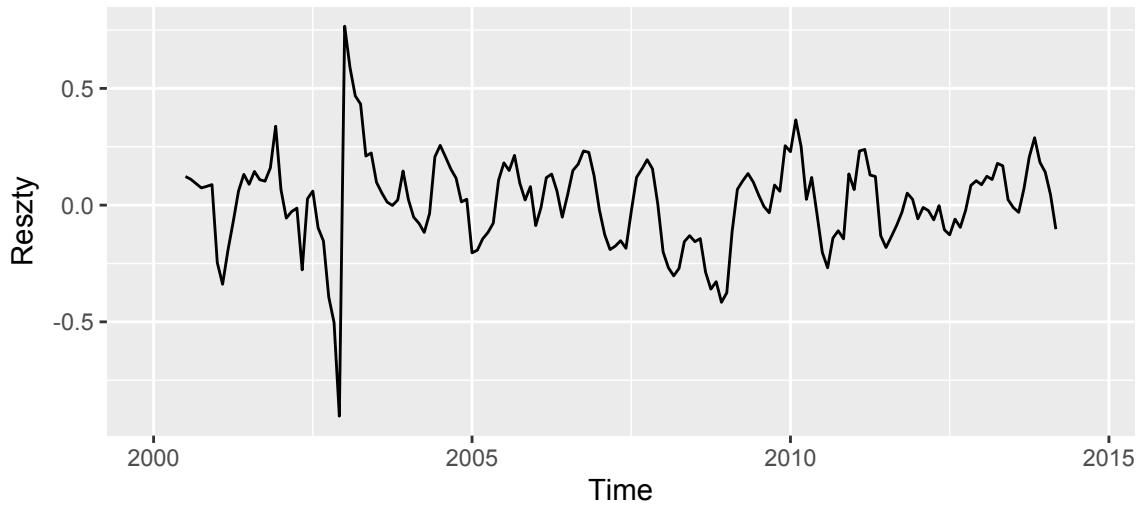
Wykres podszeregow (subseries) dla bezrobocia



Wykres 9: Wykres podszeregow dla bezrobocia

Na 9 trend w obrębie każdego miesiąca był podobny. Ponownie widzimy, że w okresie zimowym bezrobocie było wyższe.

Reszty po dekompozycji

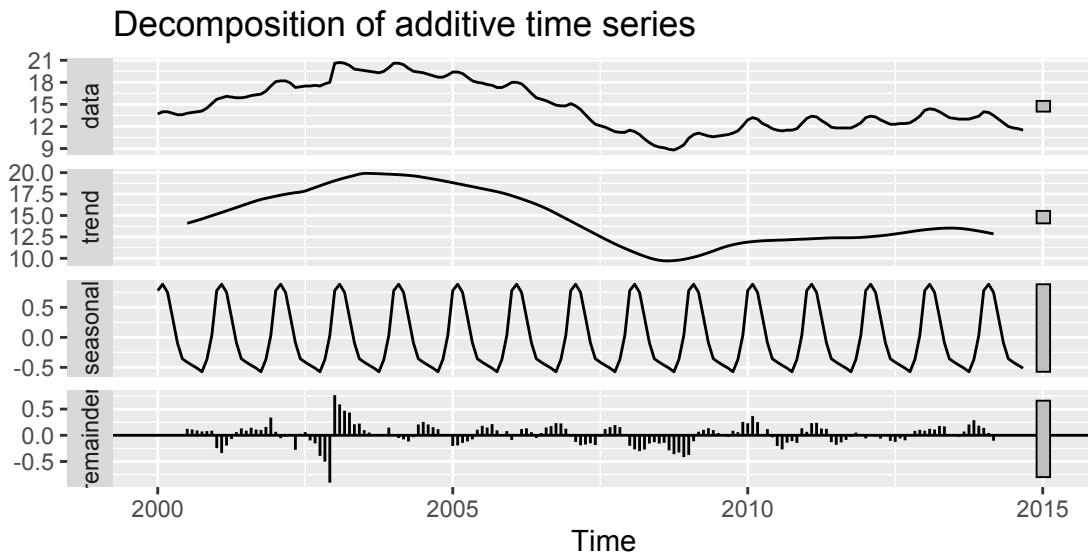


Wykres 10: Wykres reszt

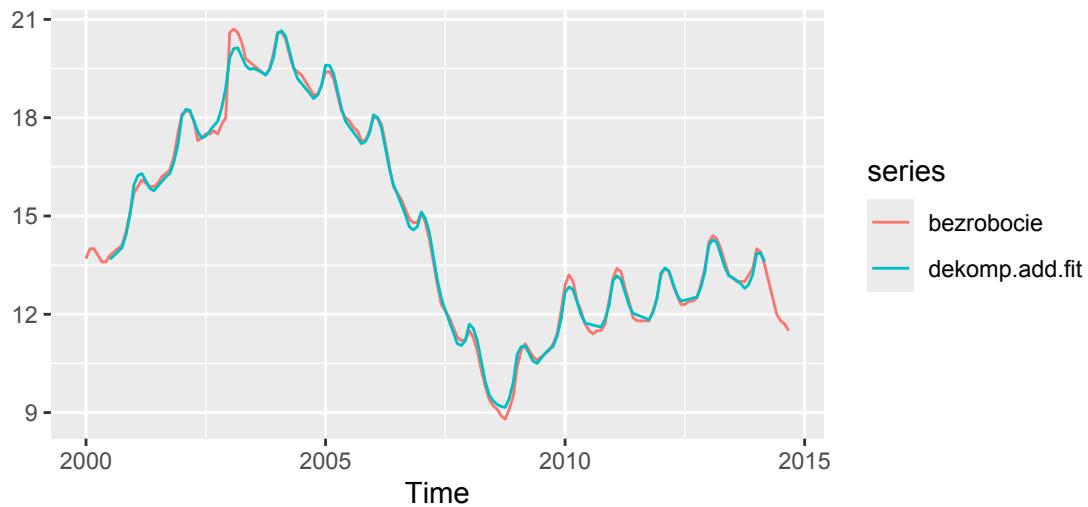
Na wykresie reszt 10 amplituda wahań przez cały okres utrzymuje się na stabilnym poziomie około ± 0.25 , co wskazuje na jednorodność wariancji. W jednym krótkim fragmencie szeregu widoczny jest gwałtowny wzrost rozrzutu (od -0.75 do 0.75), który ma charakter pojedynczej anomalii, a nie trwałej zmiany wariancji.

2.3 Metody dekompozycji

2.3.1 Dekompozycja na podstawie ruchomej średniej

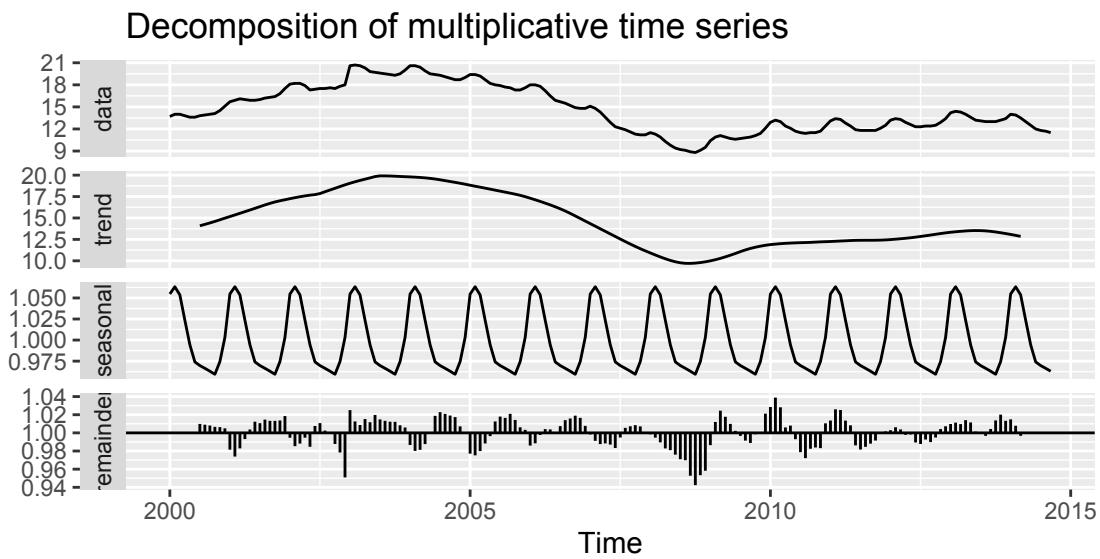


Wykres 11: Wykres dekompozycji addytywnej szeregu czasowego

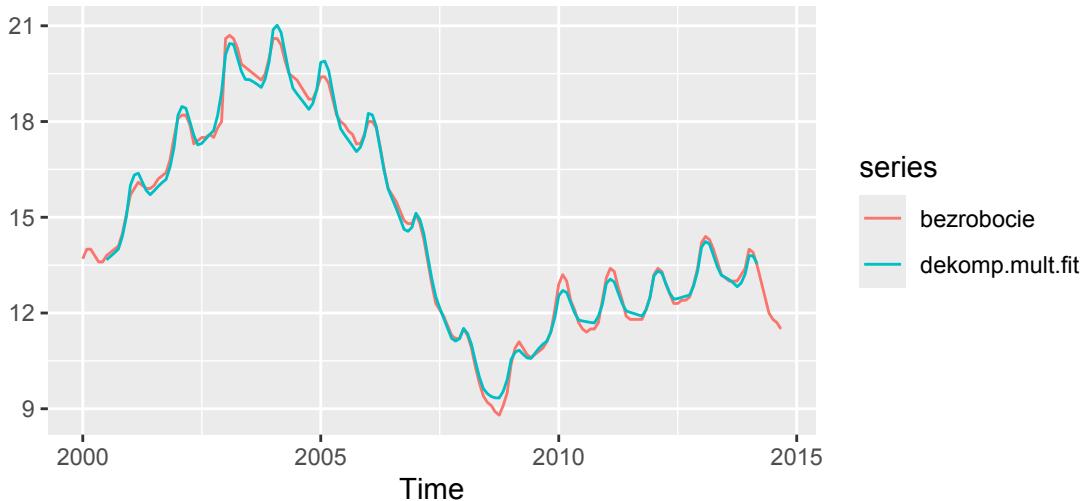


Wykres 12: Dopasowanie addytywnego modelu dekompozycji

Dekompozycja na podstawie ruchomej średniej w wersji addytywnej dopasowuje się bardzo dobrze do oryginalnego szeregu.



Wykres 13: Wykres dekompozycji multiplikatywnej szeregu czasowego

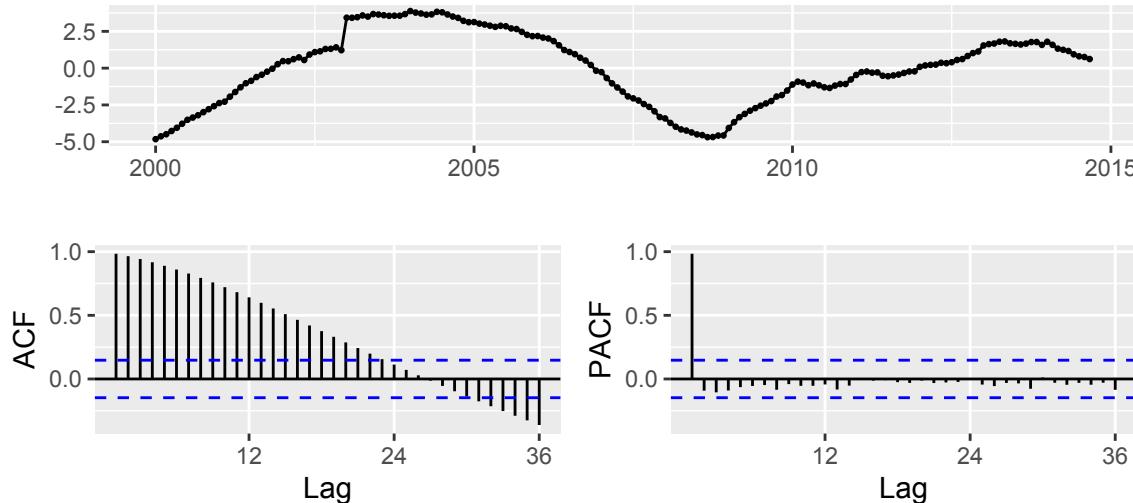


Wykres 14: Dopasowanie multiplikatywnego modelu dekompozycji

Dekompozycja na podstawie ruchomej średniej w wersji multiplikatywnej może lepiej oddawać dynamikę, jeśli amplituda sezonowa rośnie lub maleje. W tym przypadku amplituda jest stała, więc ten model nie poprawia dopasowania. Co więcej, dopasowanie nawet trochę się pogorszyło, widać to w szczególności gdy porównamy 11 i 13, ponieważ na 13 amplituda reszt się zwiększyła i widać, że ma wyraźną strukturę.

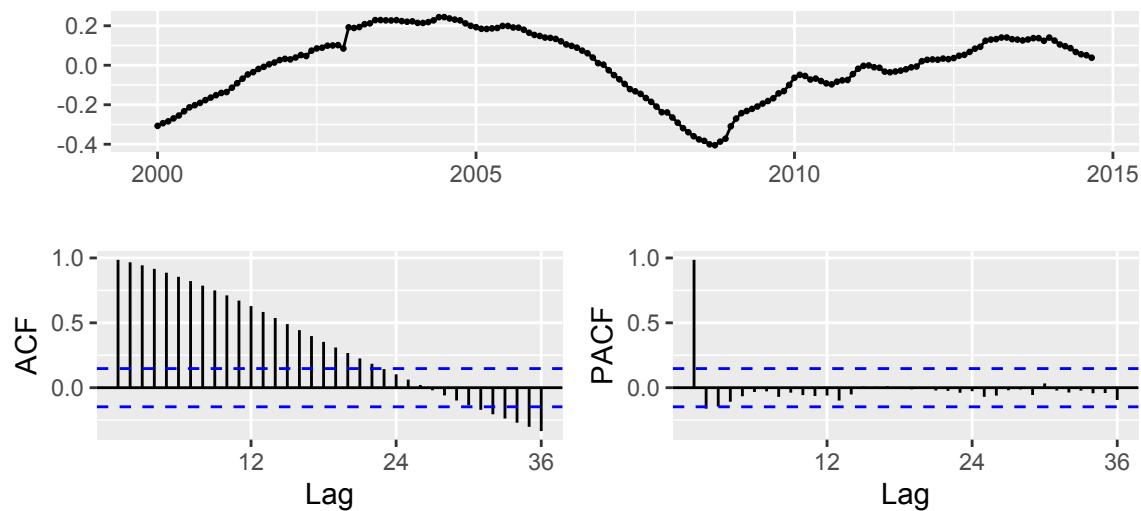
2.3.2 Dekompozycja na podstawie modelu regresji

Reszty losowe dla trendu liniowego



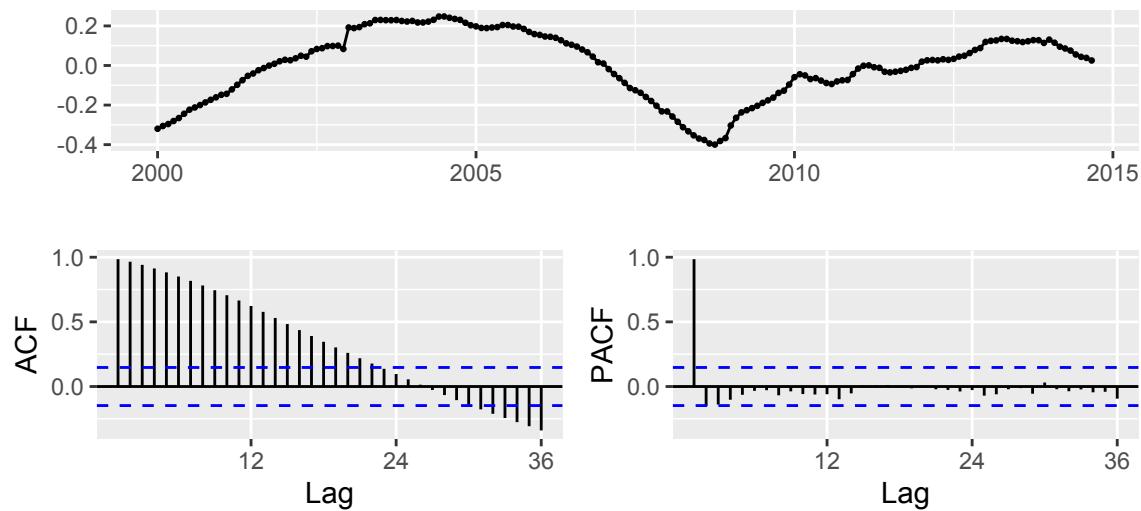
Wykres 15: Dekompozycja na podstawie modelu regresji

Reszty losowe dla zlogarytmowanego trendu liniowego



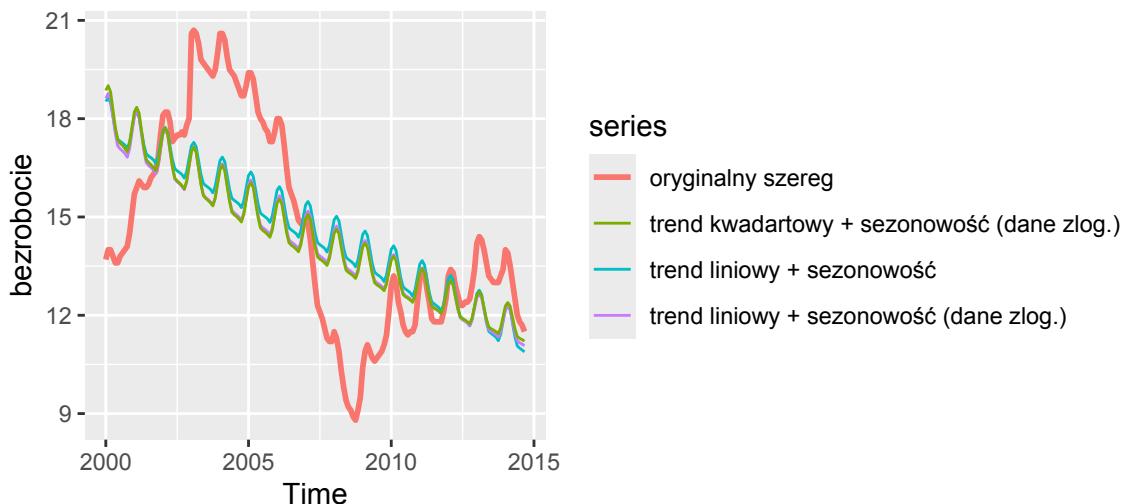
Wykres 16: Dekompozycja na podstawie modelu regresji

Reszty losowe dla zlogarytmowanego trendu kwadratowego



Wykres 17: Dekompozycja na podstawie modelu regresji

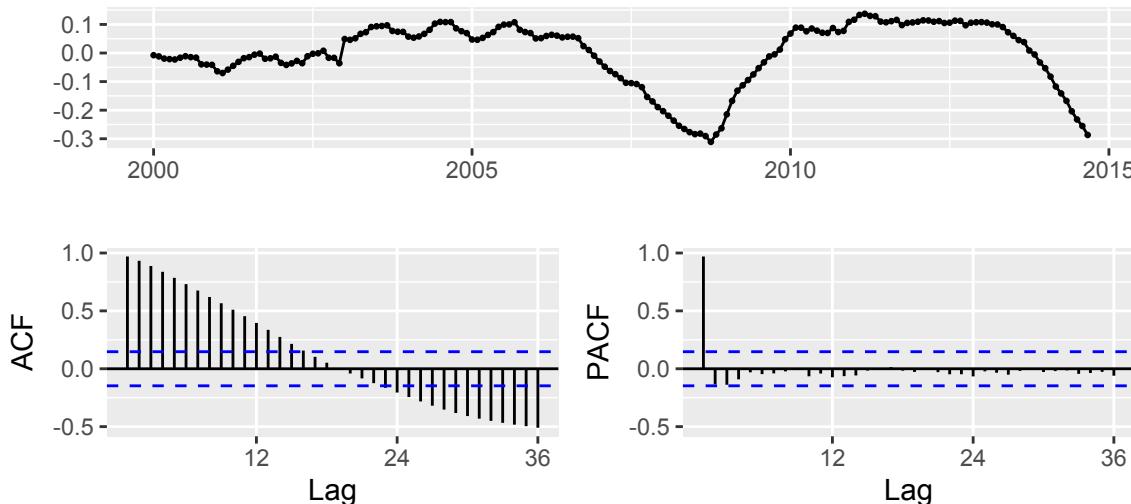
Porównanie modeli dekompozycji



Wykres 18: Porównanie dopasowań

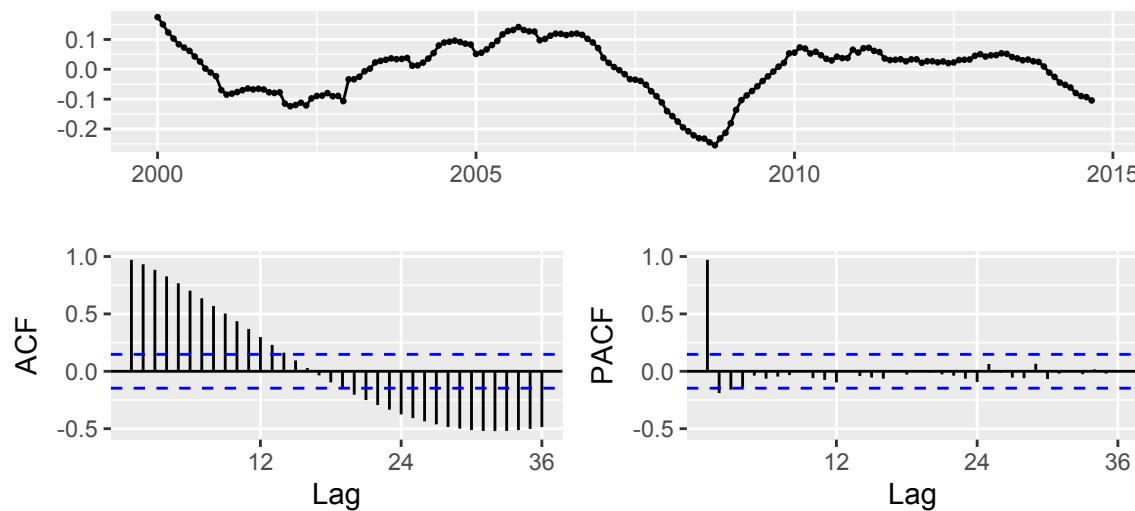
Na wykresie 18 widać, że trend liniowy dobrze oddaje ogólny spadek bezrobocia w czasie, ale nie odwzorowuje lokalnych wzrostów i spadków. Wszystkie modele dobrze uwzględniają sezonowość, widać, że jest stabilna i powtarzalna niezależnie od tego czy użyto transformacji logarytmicznej. Dodanie trendu kwadratowego nie daje znacząco lepszego dopasowania.

Reszty losowe dla zlogarytmowanego trendu sześciennego



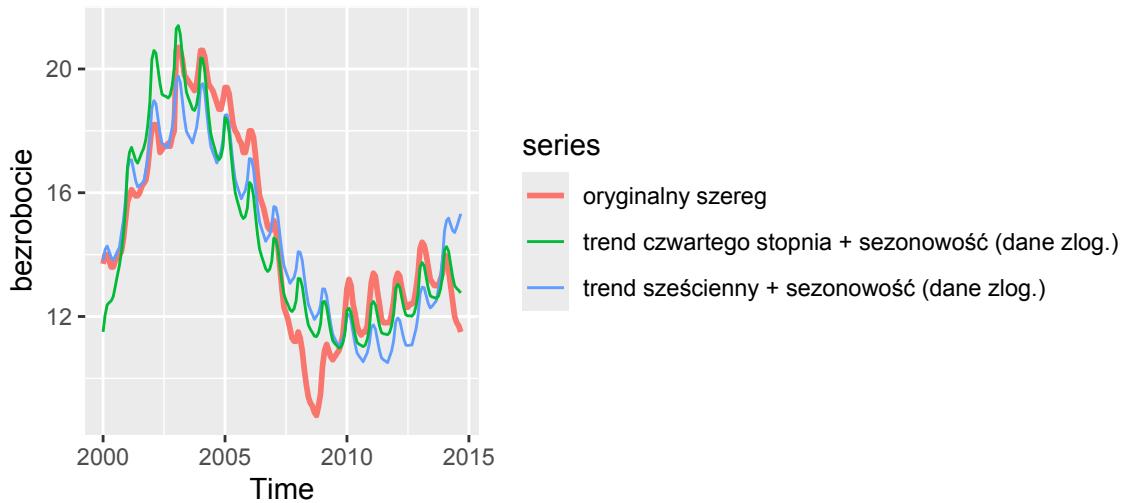
Wykres 19: Dekompozycja na podstawie modelu regresji dla wyższych stopni

Reszty losowe dla zlogarytmowanego trendu czwartego stopnia



Wykres 20: Dekompozycja na podstawie modelu regresji dla wyższych stopni

Porównanie modeli dekompozycji



Wykres 21: Porównanie dopasowań dla wyższych stopni

Używając wielomianów stopnia 3 i 4, możemy znacznie lepiej dopasować modele do trendu, linie zaczynają przypominać krzywą.

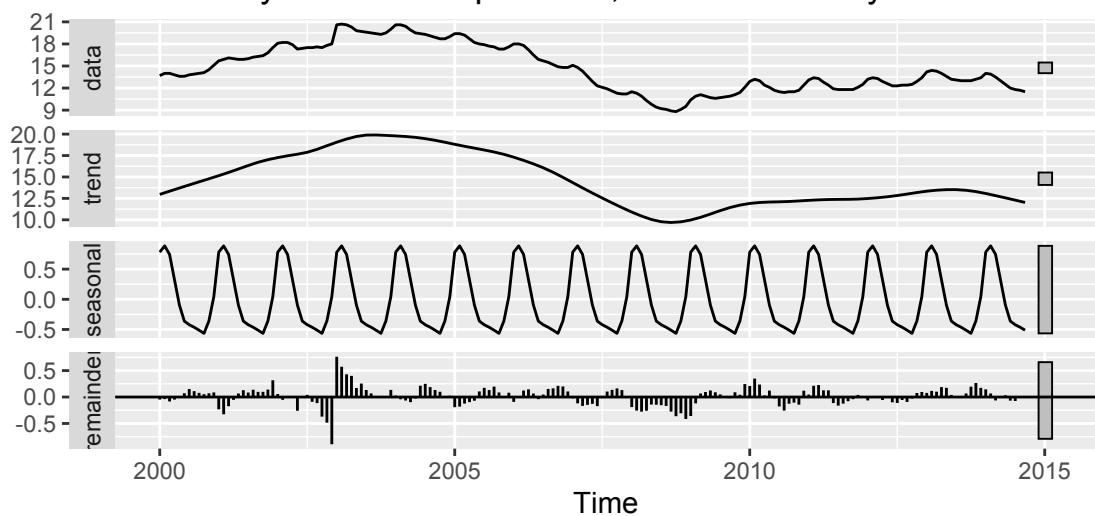
W tym momencie warto zauważać, że w każdym z modeli, niezależnie od stopnia wielomianu, wykresy ACF 17 i 20 przedstawiają bardzo dużą autokorelację. Oznacza to, że wielomian jest nieodpowiednim narzędziem do tych danych, bo nie modeluje ich autokorelacji.

2.3.3 Dekompozycja STL oparta na metodzie loess

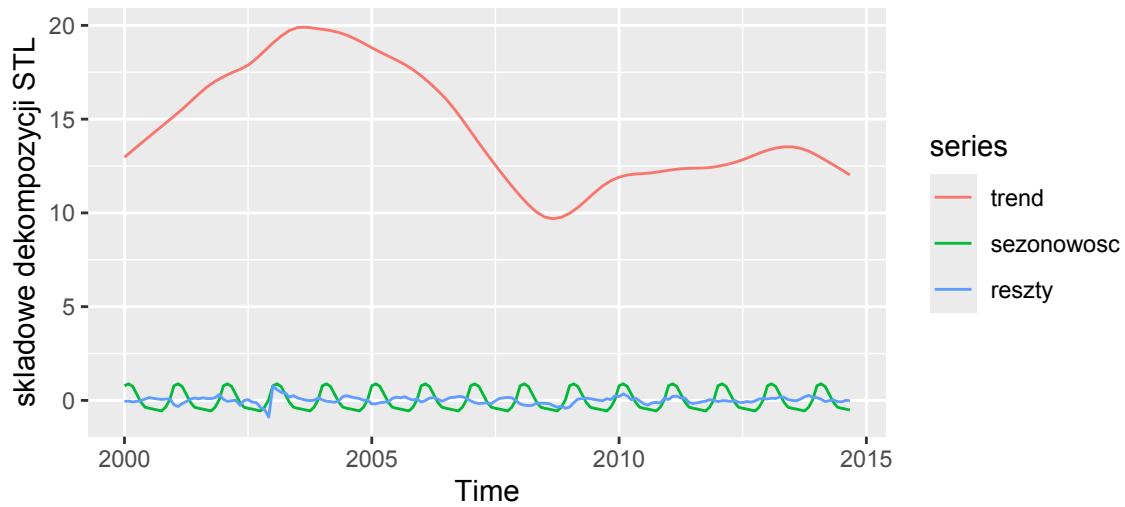
W tej metodzie możemy eksperymentować z parametrami `s.window` i `t.window` tak aby reszty były jak najmniejsze.

Najpierw dla domyślnych parametrów.

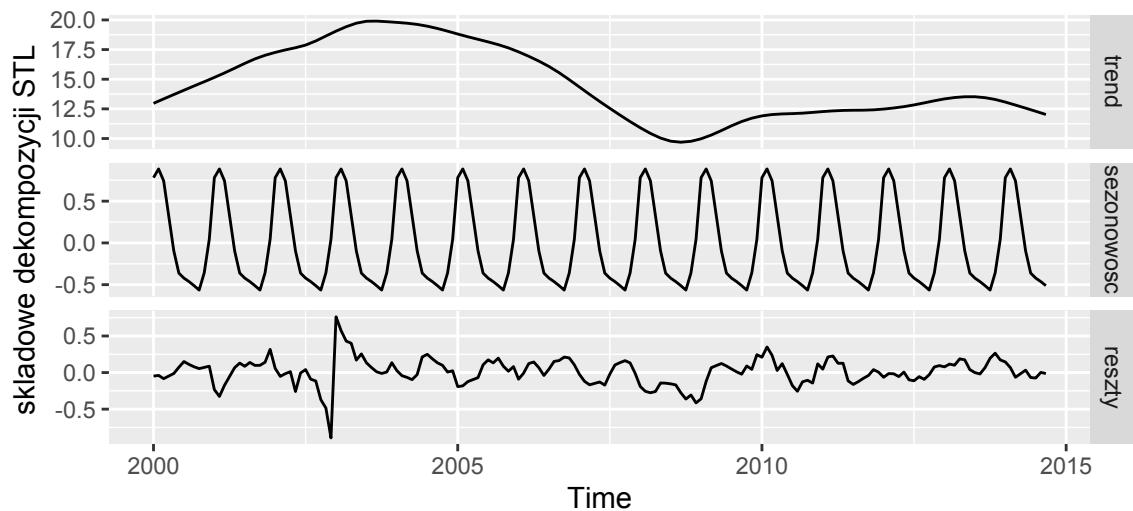
Parametry: s.window="periodic", t.window - domyślne



Wykres 22: Dekompozycja STL dla domyślnych parametrów

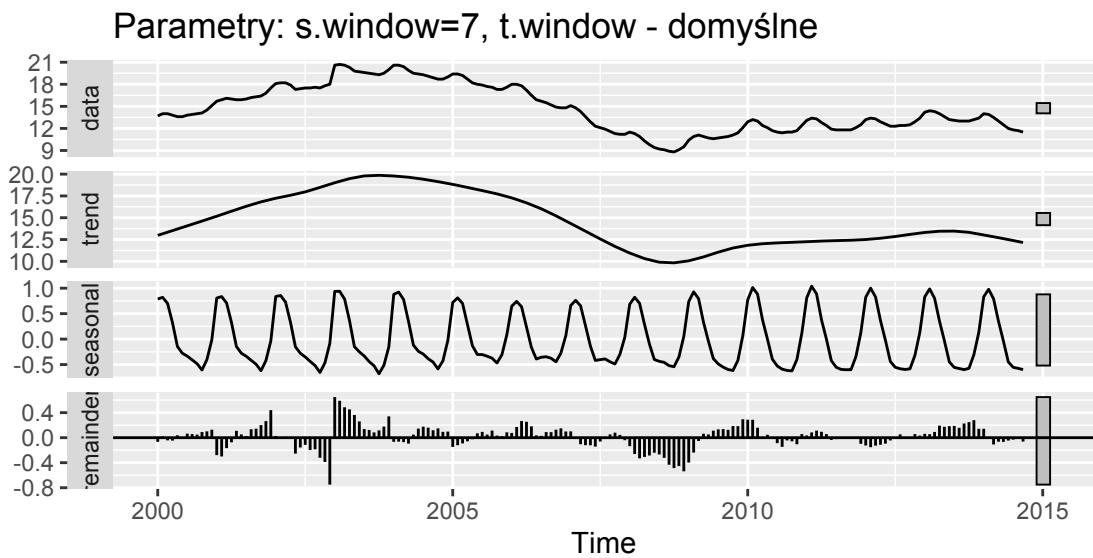


Wykres 23: Dekompozycja STL dla domyślnych parametrów



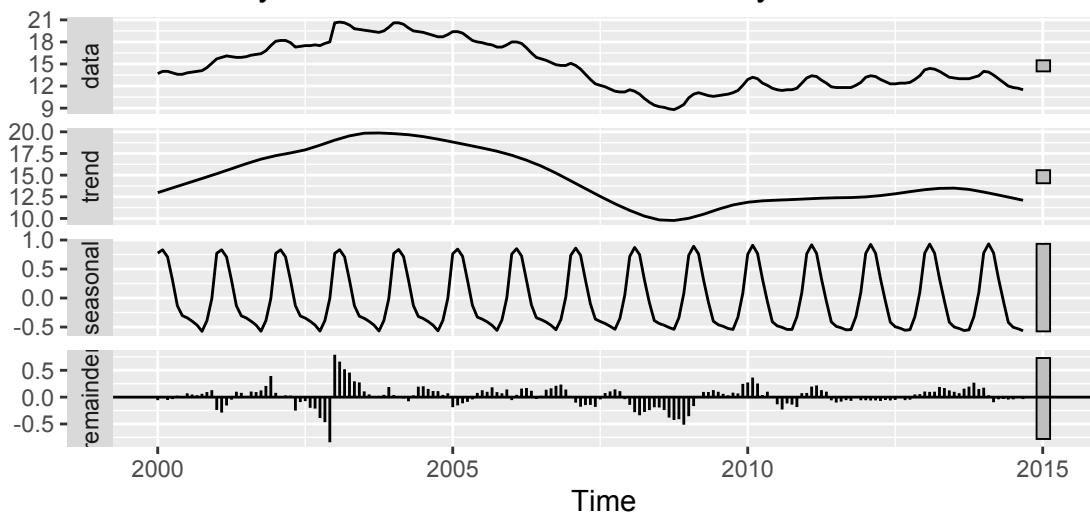
Wykres 24: Dekompozycja STL dla domyślnych parametrów

Teraz zobaczymy jak wygląda dopasowanie gdy zmienimy `s.window` i `t.window`.



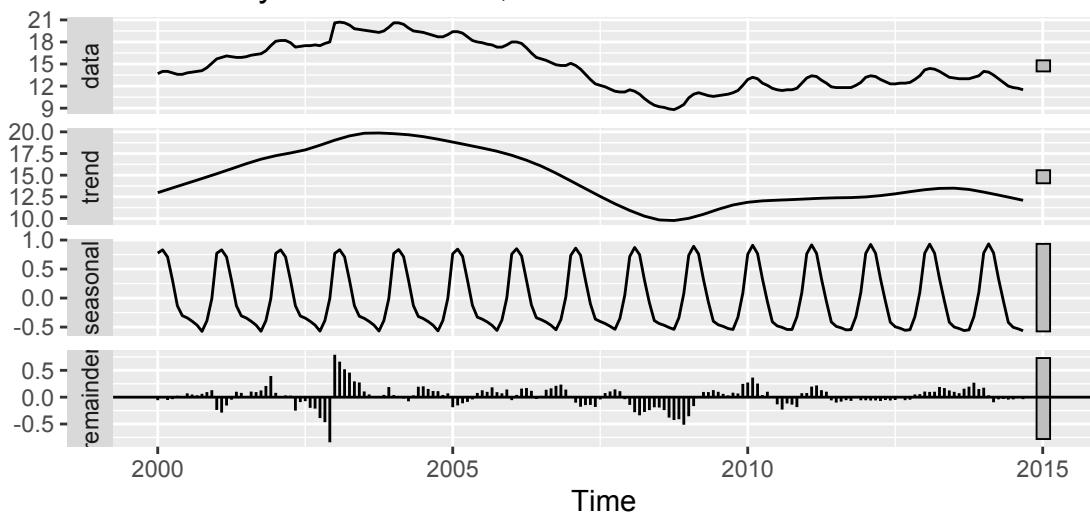
Wykres 25: Dekompozycja STL dla `s.window=7, t.window - domyślne`

Parametry: s.window=13, t.window - domyślne



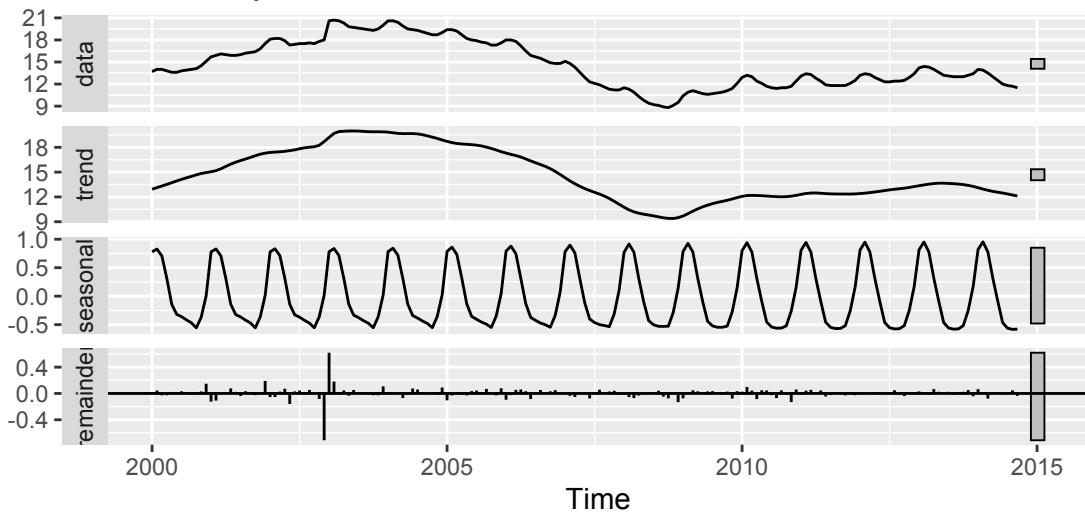
Wykres 26: Dekompozycja STL dla s.window=13, t.window - domyślne

Parametry: s.window=13, t.window=21



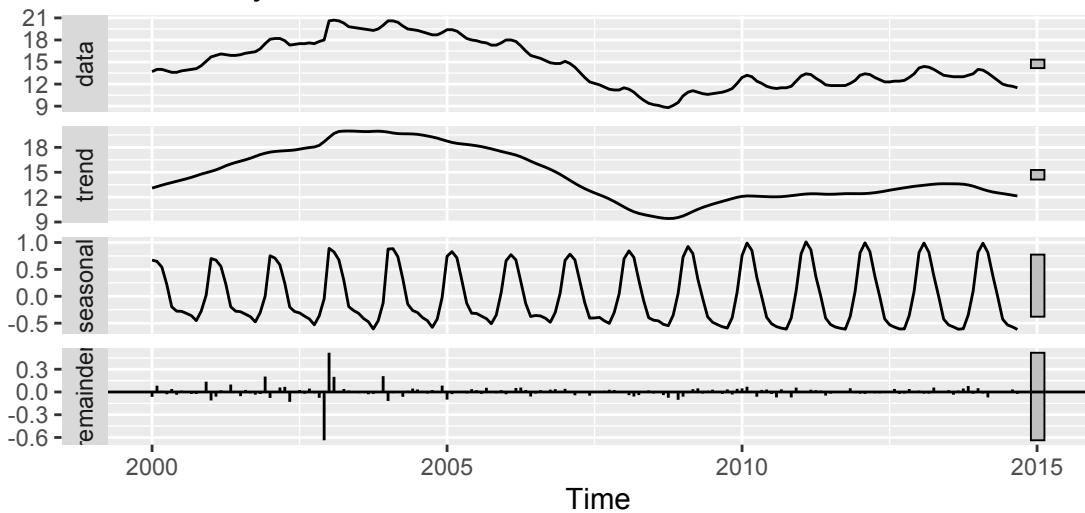
Wykres 27: Dekompozycja STL dla s.window=13, t.window=21

Parametry: s.window=13, t.window=7



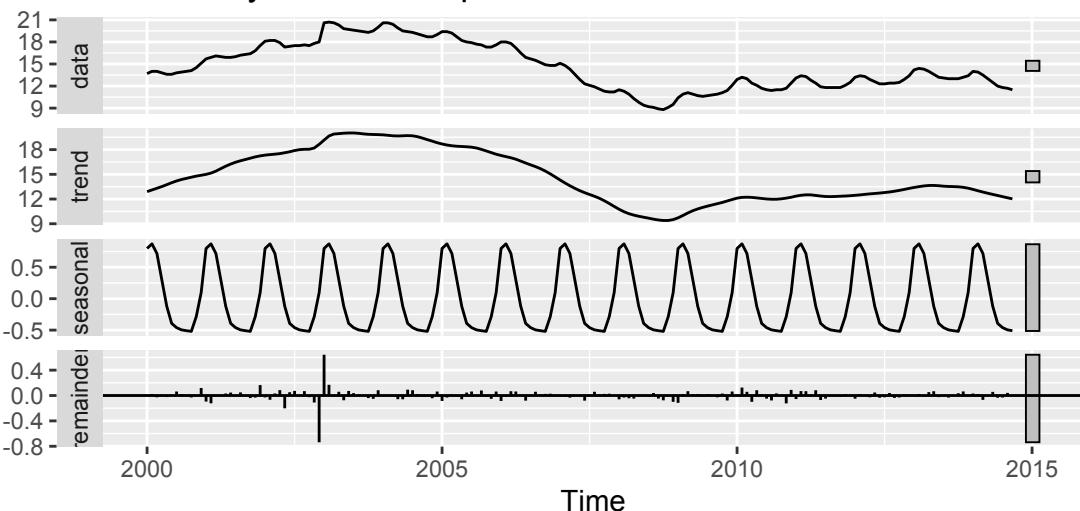
Wykres 28: Dekompozycja STL dla s.window=13, t.window=7

Parametry: s.window=7, t.window=7



Wykres 29: Dekompozycja STL dla s.window=7, t.window=7

Parametry: s.window - periodic, t.window=7



Wykres 30: Dekompozycja STL dla s.window - periodic, t.window=7

Dzięki wykresom 28, 29, 30 możemy dostrzec, że parametr `t.window` skutecznie wpływa na zmniejszenie reszt. Nieważne jak ustawiliśmy `s.window`, dla `t.window = 7` reszty były najmniejsze, czyli zarówno trend jak i sezonowość zostały najlepiej dopasowane. Możemy z tego wynieść wniosek, że dla dobrego dopasowania STL najważniejsze było odpowiednie dopasowanie do szybko zmieniającego się trendu. Natomiast sezonowość jest stabilna i łatwa do oddzielenia w przypadku tych danych, więc `s.window` nie znacząco wpływa na oddzielenie reszt.

2.4 Czy zastosowanie transformacji Boxa-Coxa prowadzi do poprawy jakości dopasowania modeli dekompozycji?

Do porównania “przed i po” transformacji Boxa-Coxa wybieramy `decompose` w wersji addytywnej, bo to on ma większy sens dla tych danych. Z modeli regresji wybieramy wariant liniowy + sezonowość, ponieważ jest prosty i przynajmniej dobrze wskazuje trend. Natomiast spośród STL będziemy porównywać jeden z lepiej dopasowanych wariantów o parametrach `s.window = 13, t.window = 7`.

Najpierw wyznaczmy optymalną wartość parametru λ metodą guerrero, która stosuje podział szeregu na podokresy i szuka λ minimalizującego współczynnik zmienności w podokresach.

```
# wybór lambdy
lambda_hat <- BoxCox.lambda(y, method="guerrero")
# transformacja
y_bc <- BoxCox(y, lambda_hat)
```

Aby móc właściwie porównać dane odwracamy transformację Boxa-Coxa i przywracamy dane do oryginalnej skali.

```
# 1. decompose
fit.dekomp.add <- dekomp.add$trend + dekomp.add$seasonal
fit.dekomp.add.bc <- InvBoxCox(dekomp.add.bc$trend + dekomp.add.bc$seasonal, lambda_hat)

# 2. TSLM
fit.ts1 <- fitted(ts1)
fit.ts1.bc <- InvBoxCox(fitted(ts1.bc), lambda_hat)

# 3. STL
```

```

fit.stl.4 <- trendcycle(dekomp.stl.4) + seasonal(dekomp.stl.4)
fit.stl.4.bc <- InvBoxCox(trendcycle(dekomp.stl.4.bc) + seasonal(dekomp.stl.4.bc), lambda_hat)

```

Aby sprawdzić dopasowanie obliczamy pierwastek z średniego błędu kwadratowego. Pokażemy nam to jak średnio duże są błędy dopasowania modelu do danych.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - (\hat{\text{trend}}_t + \hat{\text{sezonowość}}_t))^2}$$

gdzie:

- y_t – obserwowana wartość szeregu w czasie t
- $\hat{\text{trend}}_t$ – wartość trendu w czasie t wyznaczona przez model
- $\hat{\text{sezonowość}}_t$ – wartość składowej sezonowej w czasie t wyznaczona przez model
- n – liczba obserwacji

```

rmse <- function(actual, fitted){
  sqrt(mean((actual - fitted)^2, na.rm = TRUE))
}

```

Tabela 15: Porównanie RMSE dla różnych metod dekompozycji przed i po Box-Cox

Metoda	RMSE_oryginal	RMSE_BoxCox
Decompose	0.1968756	0.1968791
TSLM	2.4477829	2.4477272
STL	0.0912944	0.0873081

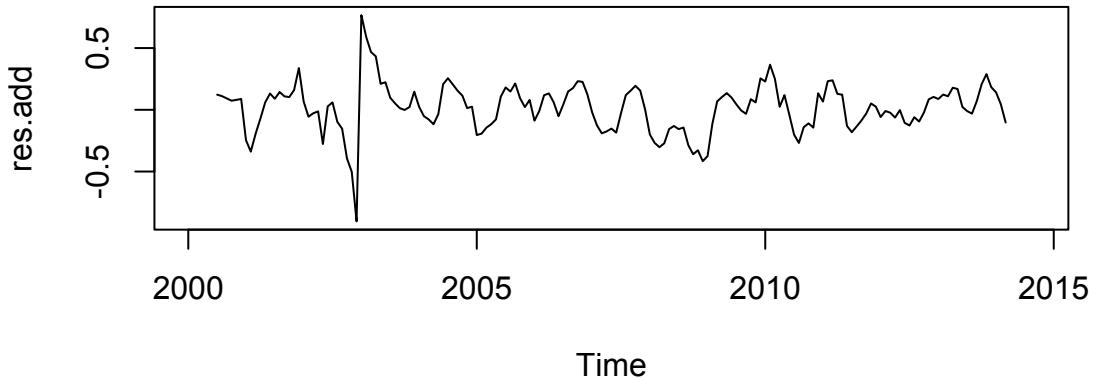
Addytywny wariant decompose nie poprawił swojego dopasowania po transformacji Boxa-Coxa, jest lepszy niż `tslm`, ale gorszy niż `stl`. Brak zmian mamy również w przypadku `tslm` liniowego, z tą różnicą, że dopasowanie było bardzo złe i pozostało bardzo złe. Natomiast w przypadku `stl` udało się poprawić wynik.

2.5 Eliminacja trendu poprzez różnicowanie danych

W tym podpunkcie przyjrzymy się szeregom reszt otrzymanym po eliminacji trendu i sezonowości. Porównamy działanie metod `decompose` `tslm` i `stl` do różnicowania.

2.5.1 Szereg różnic po użyciu decompose

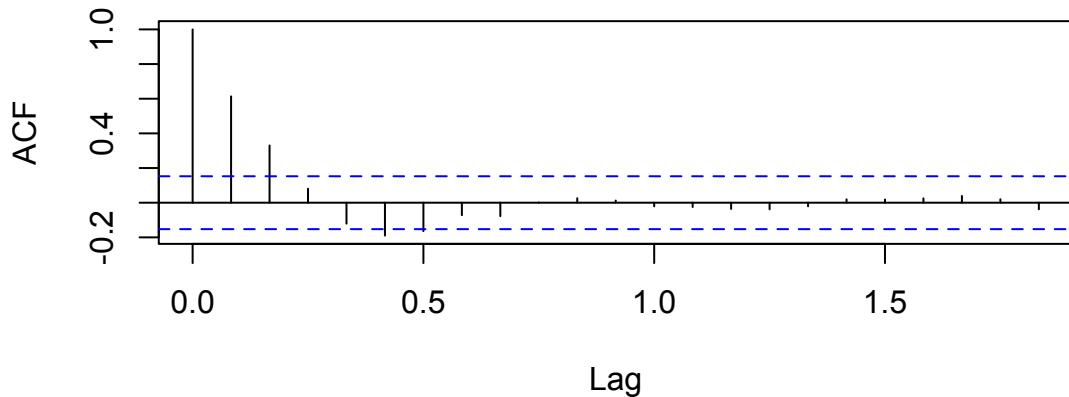
Reszty: `decompose()` additive



Wykres 31: Szereg reszt dla metody decompose

Po zastosowaniu decompose, szereg reszt 31 nie zdradza trendu ani wzorców sezonowych. Wygląda jak losowy szum, z dużą, lecz stabilną amplitudą na całym przedziale, oprócz jednego miejsca między rokiem 2000 i 2005.

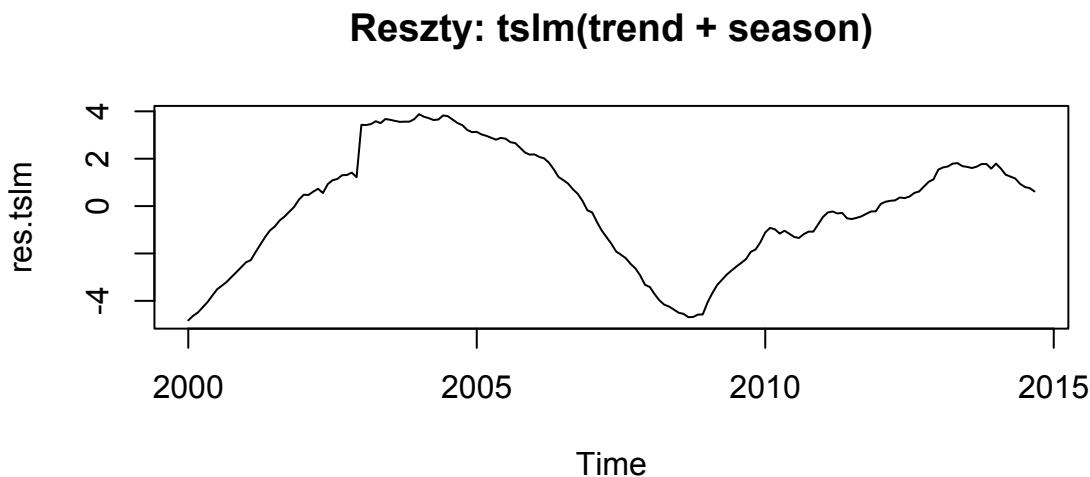
ACF – `decompose()`



Wykres 32: ACF dla metody decompose

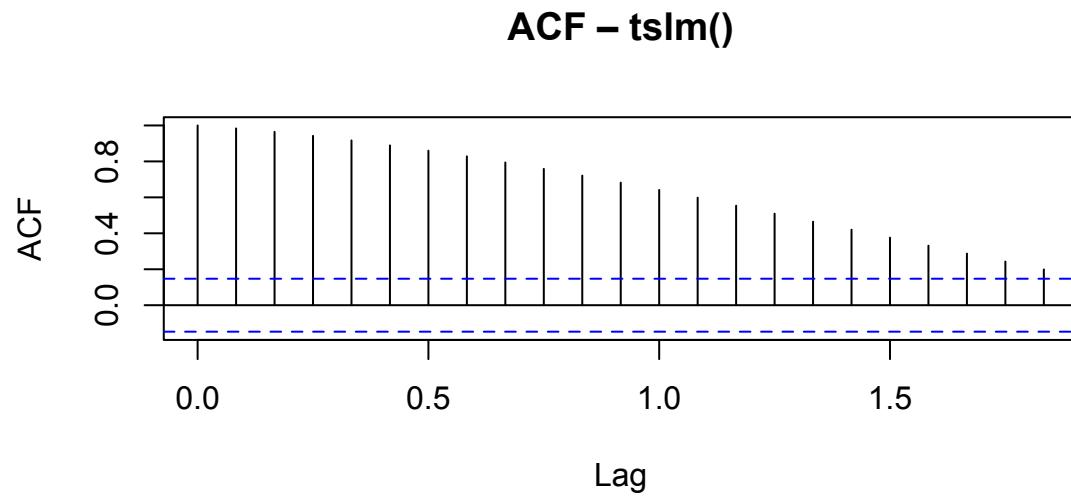
Na 32 widzimy, że dopiero od czwartego opóźnienia ACF schodzi poniżej poziomu istotności, a na stałe utrzymuje się poniżej dopiero od ósmego opóźnienia. Oznacza to, że trend i sezonowość mogły nie zostać do końca usunięte.

2.5.2 Szereg różnic po użyciu tslm



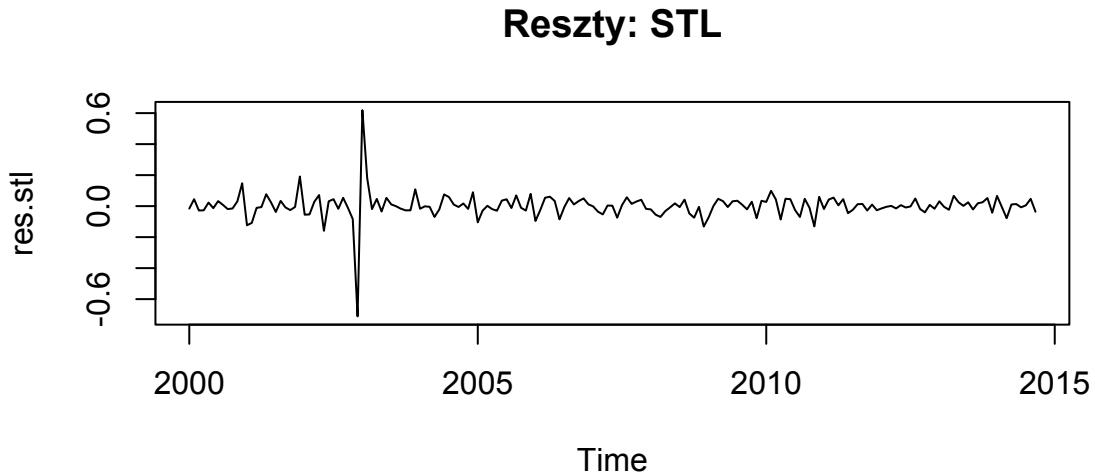
Wykres 33: Szereg reszt dla metody tslm

Na wykresie 33 widzimy wyraźny trend i sezonowość, tak jak zauważyliśmy wcześniej, dekompozycja oparta na podstawie modelu regresji nie nadaje się do wykorzystania z tymi danymi.



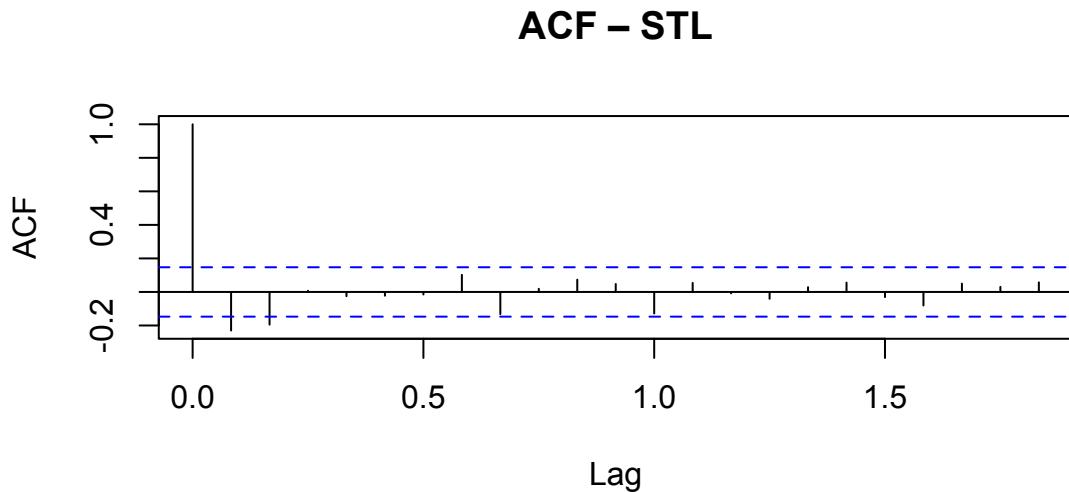
Wszystkie opóźnienia znajdują się powyżej poziomu istotności.

2.5.3 Szereg różnic po użyciu stl



Wykres 34: Szereg reszt dla metody stl

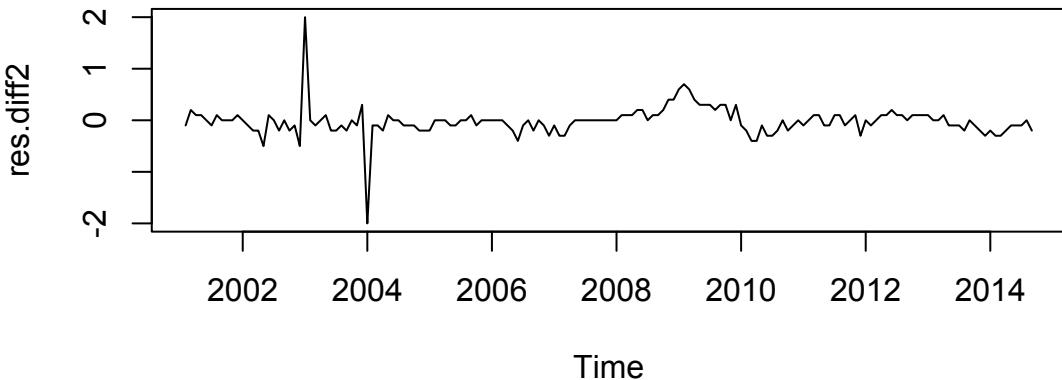
Jak możemy zauważyć na 34 poza jednym skokiem wariancja jest stała i bardzo mała, wygląda jak szum wokół zera. Nie widać trendu ani sezonowości.



ACF bardzo szybko maleje, już od czwartego opóźnienia spada poniżej poziomu istotności. Jest to bardzo dobry znak jeśli chodzi o stacjonarność. Spośród metod dekompozycji metoda `stl` dała najlepszy wynik.

2.5.4 Szereg różnic otrzymany po różnicowaniu

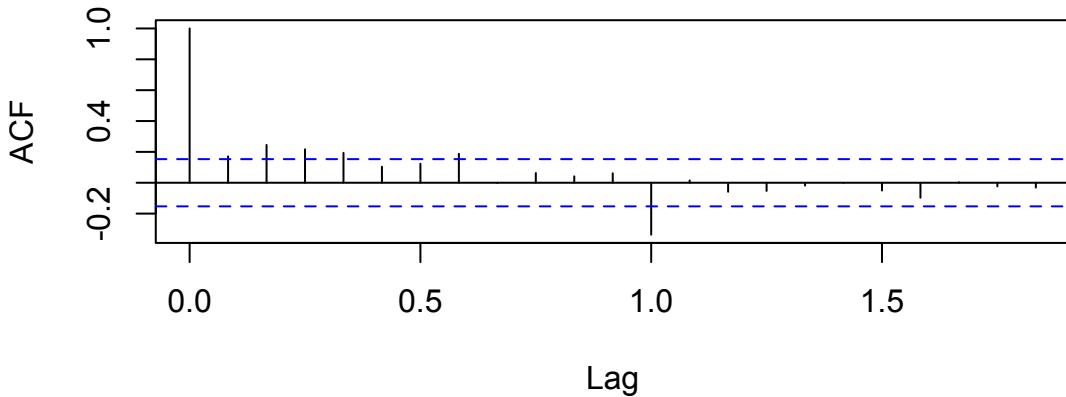
Reszty: różnicowanie



Wykres 35: Szereg reszt po różnicowaniu

Wykres reszt 35 prezentuje się dość obiecująco, ale pojawiają się dwa skoki w okolicach roku 2004 i co gorsza fragment między 2008 i 2010 nie wygląda jak szum losowy tylko jak trend rosnący, a następnie malejący.

ACF – różnicowanie



ACF bardzo szybko maleje, ale do 7 opóźnienia, większość opóźnień jest powyżej poziomu istotności co nie jest dobrym znakiem.

2.6 Podsumowanie i wnioski końcowe

Przeprowadzona analiza szeregu czasowego pozwoliła na porównanie skuteczności różnych metod dekompozycji oraz ocenę stacjonarności uzyskanych reszt. Główne wnioski przedstawiają się następująco:

1. Efektywność modeli dekompozycji:

- Najwyższą jakość dopasowania uzyskała **metoda STL** ($RMSE = 0.0913$). Dobór parametrów `s.window` oraz `t.window = 7` pozwolił na precyzyjne oddzielenie nieliniowego trendu od składowej sezonowej.
- **Klasyczna dekompozycja addytywna** wykazuje poprawne, choć mniej precyzyjne dopasowanie ($RMSE = 0.1969$).
- **Model regresyjny (TSLM)** okazał się całkowicie nieadekwatny ($RMSE = 2.4478$), co potwierdza, że sztywna struktura liniowa nie jest w stanie opisać dynamiki analizowanego zjawiska.

2. Wpływ transformacji Boxa-Coxa:

- Zastosowanie transformacji z parametrem λ wyznaczonym metodą Guerrero przyniosło istotną poprawę jedynie w przypadku **metody STL** (spadek RMSE do poziomu **0.0873**). W pozostałych metodach wpływ transformacji na błędy dopasowania był marginalny.

3. Ocena stacjonarności reszt:

- Za szereg zbliżony do **stacjonarnego (szum biały)** można uznać jedynie reszty otrzymane metodą **STL**. Potwierdza to wykres ACF, który najszybciej wygasza, schodząc poniżej poziomu istotności już przy 4. opóźnieniu.
- Reszty po **różnicowaniu** oraz z metody **decompose** wykazują niestabilność wariancji oraz powracające autokorelacje, co sugeruje, że część struktury trendu lub sezonowości nie została w pełni wyeliminowana.
- Reszty z modelu **TSLM** jednoznacznie nie są stacjonarne – wykazują silny trend resztowy.

Wniosek końcowy: Metoda **STL z transformacją Boxa-Coxa** jest optymalnym podejściem dla badanego szeregu, zapewniając najlepszą separację sygnału od szumu oraz generując składnik losowy o pożądanych właściwościach statystycznych.