

# Analiza szeregów czasowych - raport 2.

Miłosz Kubera(249823)  
Patryk Krukowski(249824)

27 kwietnia 2021

## Spis treści

<b>1</b>	<b>Wstęp i opis eksperymentów</b>	<b>1</b>
<b>2</b>	<b>Estymacja, eliminacja trendu oraz wstępna identyfikacja modelu dla dekady 1960-1969</b>	<b>3</b>
2.1	Wstępne przekształcenie danych . . . . .	4
2.2	Nieparametryczna estymacja i eliminacja trendu . . . . .	4
2.3	Parametryczna estymacja trendu . . . . .	15
<b>3</b>	<b>Estymacja, eliminacja trendu oraz wstępna identyfikacja modelu dla 1965 r.</b>	<b>18</b>
3.1	Wstępne przekształcenie danych . . . . .	18
3.2	Nieparametryczna estymacja i eliminacja trendu . . . . .	19
3.3	Parametryczna estymacja trendu . . . . .	26
<b>4</b>	<b>Wstępna identyfikacja modelu - przypadek lat 1960-1969 oraz 1965 roku.</b>	<b>30</b>

## 1 Wstęp i opis eksperymentów

W sprawozdaniu tym zajmiemy się estymacją oraz eliminacją trendu metodami parametrycznymi oraz nieparametrycznymi (podpunkty (a), (b) i (c)), oraz wstępna identyfikacją modelu (podpunkt (d)) dla zakłócenia losowego na bazie danych dotyczących historycznych notowań indeksu Dow-Jonesa w latach 1960-1969 oraz w krótszym okresie czasowym - wybraliśmy do tego 1965 r. Badana przez nas dekada została wylosowana poprzez funkcję *losuj.dekade* widoczną w kodzie poniżej.

```
losuj.dekade <- function(album)
{
  set.seed(album)
  dekady <- seq(from=30, to=70, by=10)
  sample(dekady, 1)
}

losuj.dekade(249823 + 249824) #nasza dekada to 60

## [1] 60
```

Wczytujemy i przygotowujemy pod analizę odpowiednie dane i dokonujemy wstępnych obserwacji.

```
dane <- read.table("Dow_60-69.html", blank.lines.skip = T, fill = T, sep="\t", header=T, skip=2)
dane <- dane[-nrow(dane),] # usuwamy ostatni wiersz (znacznik </PRE>)
kurs <- dane$closing.values
daty <- as.POSIXlt(paste0("19",as.character(dane$Date)), format = "%Y%m%d")

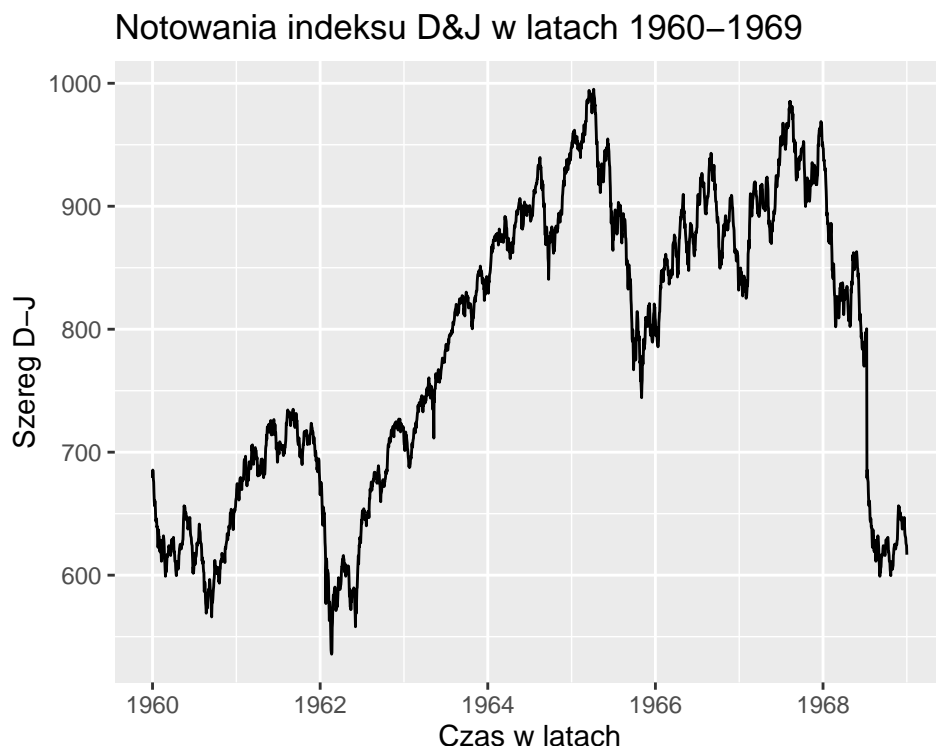
# obiekt klasy 'ts'
szereg.ts <- ts(kurs, start=1960, end=1969, frequency=292)
anyNA(szereg.ts) #Brak NA

## [1] FALSE

length(szereg.ts) #długość szeregu czasowego

## [1] 2629
```

W szczególności zauważmy, że nie mamy do czynienia z obserwacjami zakodowanymi jako NA, długość szeregu czasowego dla pełnej dekady to 2629. Zwizualizujmy teraz dane.



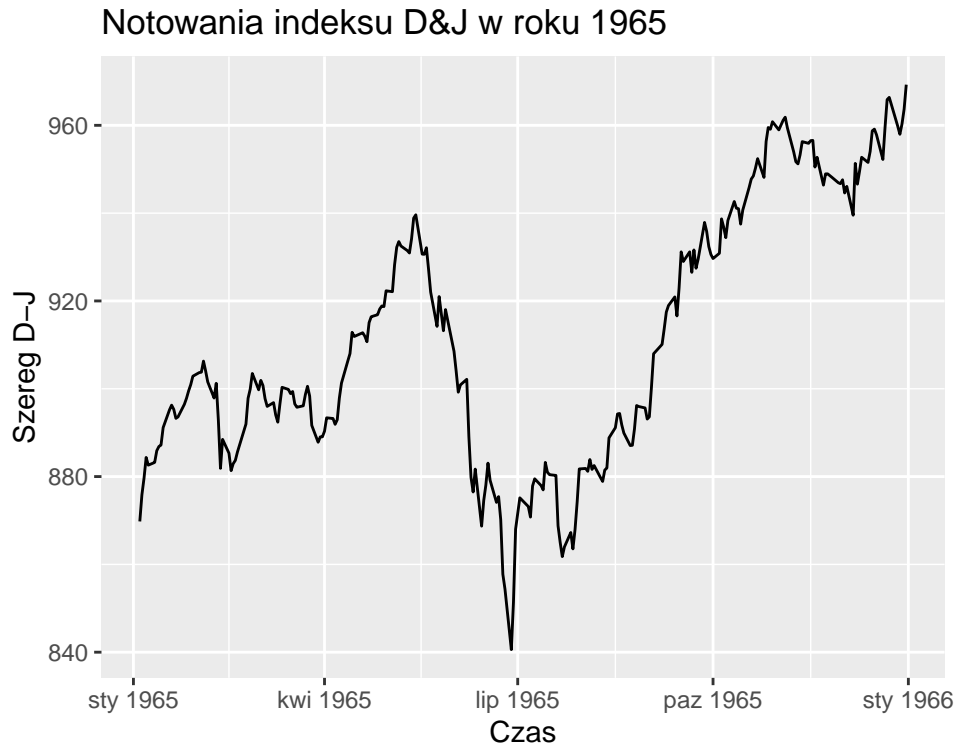
Rysunek 1: Notowania indeksu Dow-Jonesa w latach 1960-1969

Na rysunku (1) możemy zaobserwować, że rozważany szereg czasowy charakteryzuje się dużą **wariancją** i potrzebna nam będzie transformacja potęgowa. Podobny wniosek nasuwa się po spojrzeniu na rysunek (2) przedstawiający wykres szeregu czasowego dla 1965 roku. Ponadto

zakładamy (tak jak zostało napisane w zadaniu), że

$$X_t = f(t) + Z_t,$$

gdzie  $X_t$  to szereg czasowy,  $f(t)$  - deterministyczna składowa trendu,  $Z_t$  - zakłócenia losowe.



Rysunek 2: Notowania indeksu Dow-Jonesa w 1965 roku

Przedstawimy teraz spis najważniejszych narzędzi użytych w analizie.

- Transformacja potęgowa.
- Różnicowanie szeregu czasowego jako metoda eliminacji trendu.
- Metody nieparametryczne estymacji składowej trendu, takie jak wygładzanie metodą symetrycznej średniej ruchomej, wygładzanie eksponencjalne, filtr Spencera, estymatory jądrowe typu Nadaraya-Watsona z różnymi jądrami.
- Metoda estymacji trendu przy pomocy sumy funkcji *naturalnych sześciennych B-sklejanych* (ang. natural cubic B-spline functions), jest to także metoda nieparametryczna.
- Parametryczne metody estymacji trendu, takie jak estymacja przy pomocy skończonej liniowej kombinacji funkcji ciągłych pochodzących z pewnej rodziny, u nas jest to rodzina wielomianów i wielomianów trygonometrycznych.
- Narzędzia użyte w podpunkcie, takie jak test Boxa-Coxa, estymatory funkcji ACF i PACF.

Po wstępnym obejrzeniu danych przystąpmy do analizy.

## 2 Estymacja, eliminacja trendu oraz wstępna identyfikacja modelu dla dekady 1960-1969

Poniżej wykonamy zadania ze sprawozdania, przy czym porównań odpowiednich podejść dokonamy pod koniec sekcji.

### 2.1 Wstępne przekształcenie danych

Naszym zadaniem jest przekształcenie danych tak, aby wariancja była, możliwie jak najbardziej, jednorodna. W tym celu rozważymy różne transformacje Boxa-Coxa, przedstawione na poniższej animacji.

Rysunek 3: Transformacje potęgowe Boxa-Coxa dla szeregu czasowego D-J z lat 1960-1969

Na podstawie powyższej animacji wyciągamy wniosek, że transformacja potęgowa z parametrem  $\lambda = 0$  jest najlepsza w tym sensie, że wariancja tak otrzymanego (zlogarytmowanego) szeregu jest stosunkowo najmniejsza. Przy okazji zauważmy, że szereg początkowy możemy zlogarytmować, ponieważ wszystkie rozważane wartości są dodatnie.

W szczególności warto odnotować, że optymalny parametr  $\lambda$  wybrany przez pakiet R jest równy w przybliżeniu 1.89.

```
BoxCox.lambda(szereg.ts)
```

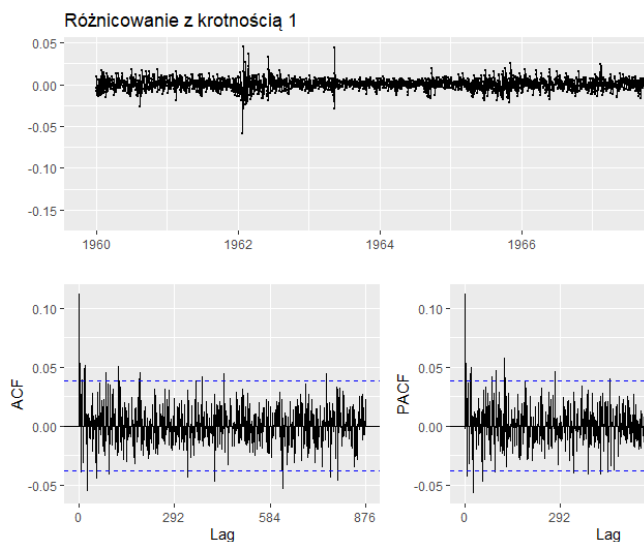
```
## [1] 1.892485
```

**Wniosek 1** *W analizie warto jest osobiście przyjrzeć się wartościom wybranym automatycznie, ponieważ mogą prowadzić one do wzbudzających podejrzenia wyników.*

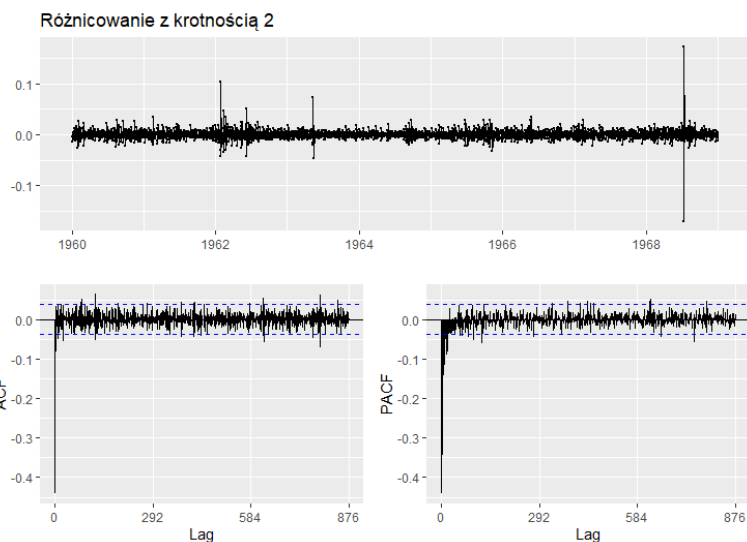
### 2.2 Nieparametryczna estymacja i eliminacja trendu

- (a) Różnicowanie - jest to metoda eliminacji trendu z szeregu czasowego, polegająca na zastąpieniu oryginalnych danych szeregiem różnic, wyznaczonych dla ustalonych opóźnień

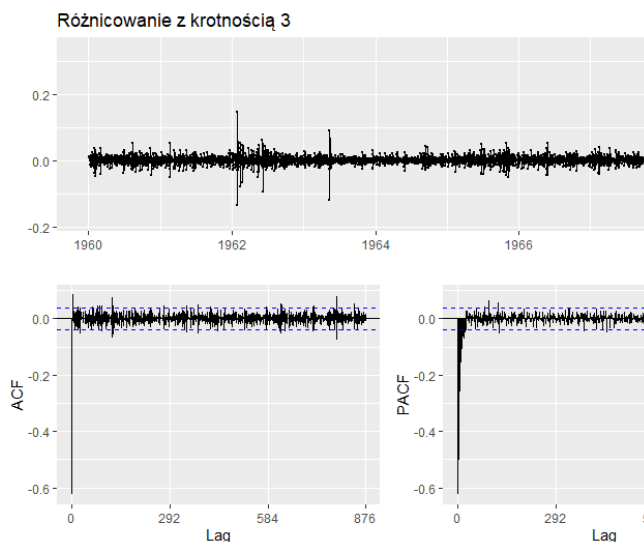
czasowych (u nas opóźnienie jest równe 1). Wybierzmy teraz odpowiednią krotność różnicowania. Z rysunków nr (4), (5), (6) i (7) wnioskujemy, że w zasadzie możemy przyjąć, iż jednokrotne różnicowanie jest wystarczające - nie obserwujemy wówczas cech charakterystycznych dla obecności deterministycznej składowej trendu, jak np. wolne zanikanie estymatora funkcji ACF. To jednak, czy możemy uznać tak oto powstały szereg za realizację białego szumu rozpatrzymy w ostatniej sekcji.



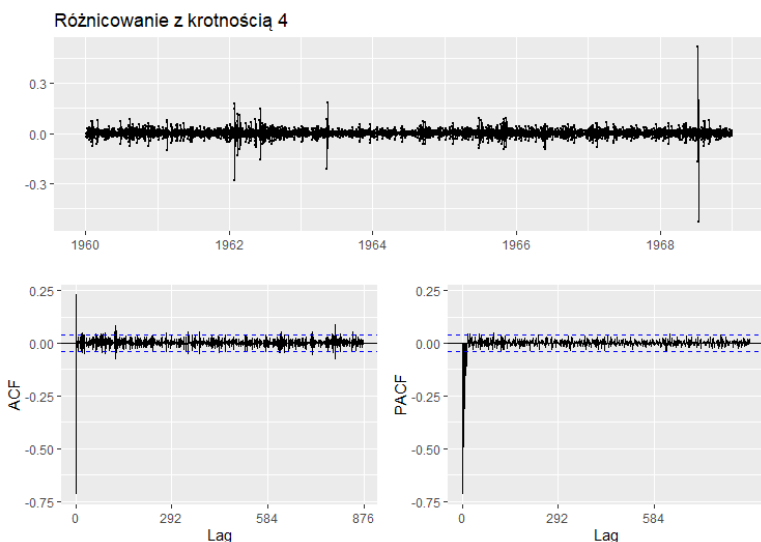
Rysunek 4: Efekt jednokrotnego różnicowania



Rysunek 5: Efekt dwukrotnego różnicowania



Rysunek 6: Efekt potrójnego różnicowania



Rysunek 7: Efekt poczwórnego różnicowania

- (b) Wygładzanie poprzez symetryczną ruchomą średnią - polega na symetrycznym uśrednianiu  $2q + 1$  wartości szeregu, gdzie  $q$  jest rzędem metody, kontrolującym stopień wygładzenia. Estymator trendu jest postaci

$$\hat{f}(t) = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}.$$

W przeciwieństwie do metody z poprzedniego podpunktu, w tym wypadku dostajemy nieparametryczny estymator trendu. W podpunkcie (a) jedynie usunęliśmy takowy trend z szeregu czasowego. Zastanówmy się, jaki rząd wygładzania powinniśmy dobrać. W tym celu przeanalizujemy poniższą pomocniczą animację.

Rysunek 8: Wybór rzędu wygładzenia szeregu czasowego D-J z lat 1960-1969 metodą symetrycznej ruchomej średniej

Z animacji odczytujemy, że symetryczna ruchoma średnia rzędu 31 w zadowalającym stopniu wygładza szereg, zachowując jednocześnie informację o lokalnych wahaniami szeregu. Zwizualizujemy przyjęty estymator trendu.

**Notka 1** Musimy uważać na parametryzację przyjętą w funkcji `ma` z pakietu `forecast`, ponieważ rzędem w tejże funkcji jest liczba  $2q + 1 = 31$ , więc nasz parametr  $q$  jest równy 15.



Rysunek 9: Estymacja trendu metodą symetrycznej ruchomej średniej rzędu 31 dla szeregu D-J dla lat 1960-1969

- (c) Filtr Spencera - jest to przykład ważonej ruchomej średniej (w poprzednim przypadku wszystkie wagi były równe  $2q + 1$ ), dla której rząd  $q = 7$  oraz wagi są określone następująco

$$[w_0, w_1, \dots, w_7] = [74, 67, 46, 21, 3, -5, -6, -3].$$

Dużą zaletą filtru Spencera jest następujący fakt.

**Fakt 1** *Filtr Spencera przepuszcza bez zniekształceń (o ile występują) trendy wielomianowe do stopnia trzeciego, eliminując jednocześnie skutecznie losowe fluktuacje.*

Natomiast za estymator trendu przyjmujemy

$$\hat{f}(t) = \sum_{j=-7}^7 w_j \cdot X_{t-j},$$

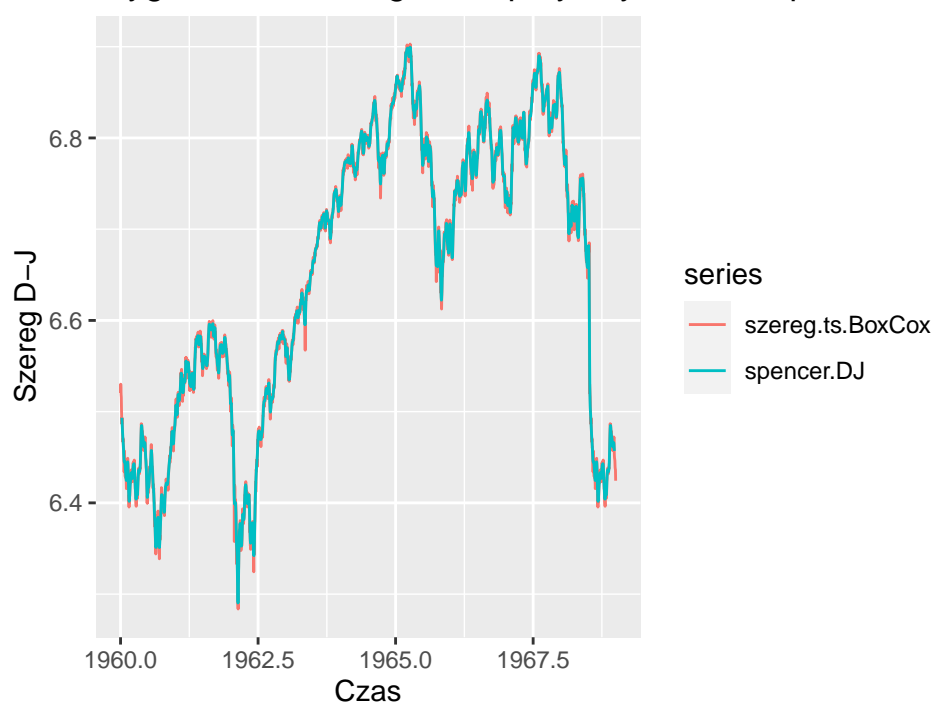
gdzie wagi  $w_j$  są określone jak wyżej.

```
spencer.DJ <- spencer(szereg.ts.BoxCox) #estymowany trend

# Uwaga: konieczna jest konwersja na obiekt 'ts'
spencer.DJ <- ts(spencer.DJ, frequency=292, start=start(szereg.ts.BoxCox),
                 end=end(szereg.ts.BoxCox))

autoplot(ts.union(szereg.ts.BoxCox, spencer.DJ), lwd=0.5, main='Wygląd szereg D-J przy użyciu filtru Spencera') +
  labs(x='Czas', y='Szereg D-J')
```

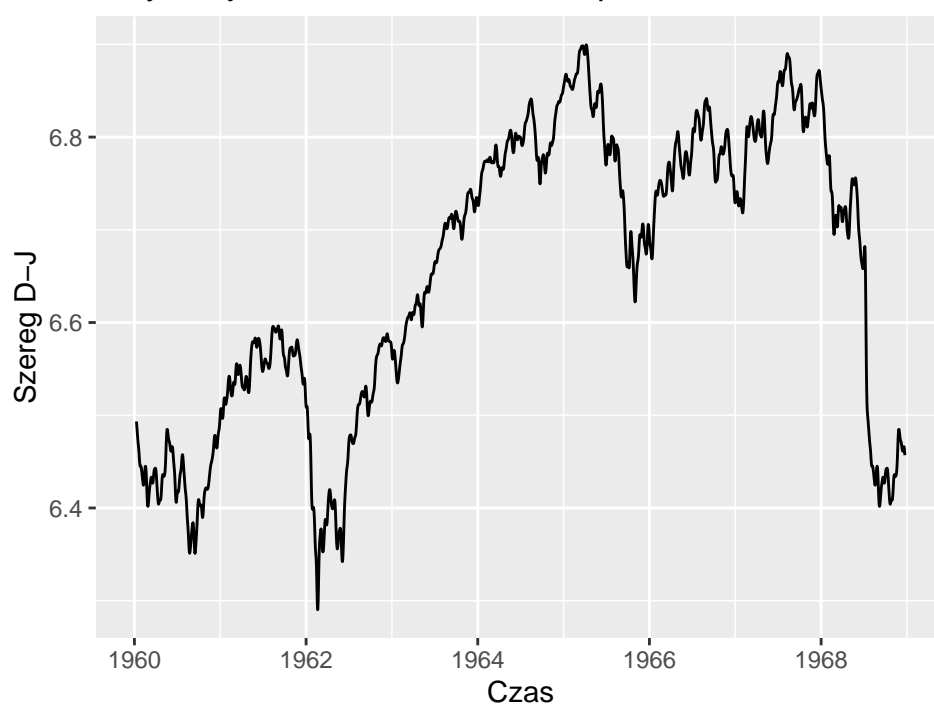
### Wygladzanie szeregu D-J przy użyciu filtru Spencera



Rysunek 10: Estymacja trendu przy użyciu filtru Spencera dla szeregu D-J lat 1960-1969

Jako że szeregi nachodzą na siebie, wykonajmy jeszcze jeden oddzielny rysunek.

### Estymacja trendu metoda filtru Spencera



Rysunek 11: Estymator trendu uzyskany przy użyciu filtru Spencera dla szeregu D-J z lat 1960-1969

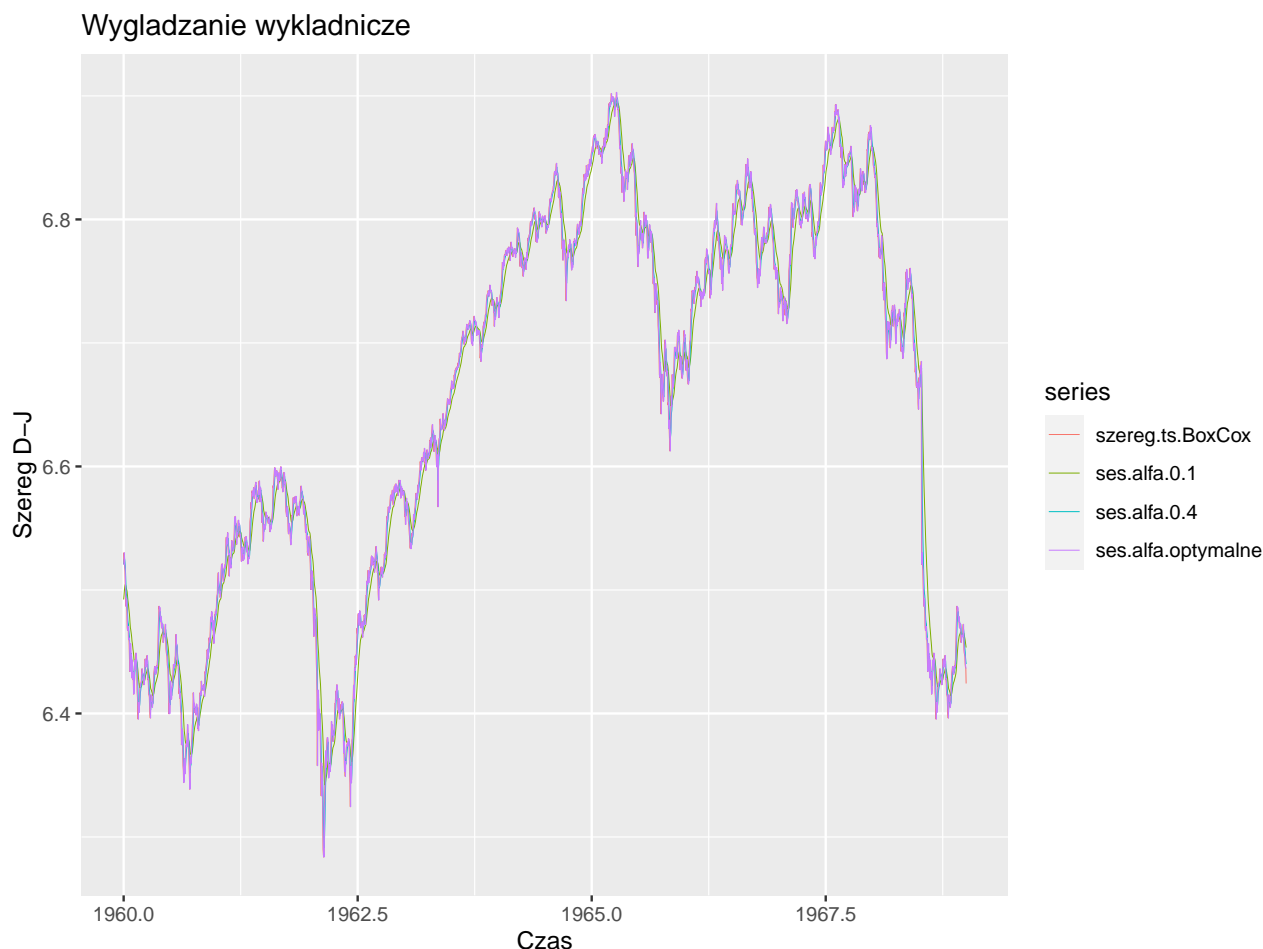


Efekt nie jest zbyt zadowalający, ponieważ obserwujemy zbyt małe wygładzenie trendu w miejscach, w których następuje dynamiczny charakter zmian szeregu czasowego.

**Wniosek 2** *Filtr Spencera nie działa najlepiej w przypadku estymacji trendu o dynamicznej strukturze zmian.*

- (d) Proste wygładzanie wykładnicze (SES) - zdecydowaliśmy się użyć SES z tego względu, że w zachowaniu badanego szeregu czasowego nie obserwujemy wyraźnych trendów długoterminowych oraz założyliśmy model, w którym nie występują efekty sezonowe. Algorytm SES przyporządkowuje wykładniczo malejące wagi postaci  $(1 - \alpha)^k$  do obserwacji z poprzednich okresów, przy czym  $\alpha$  jest parametrem kontrolującym wygładzanie.

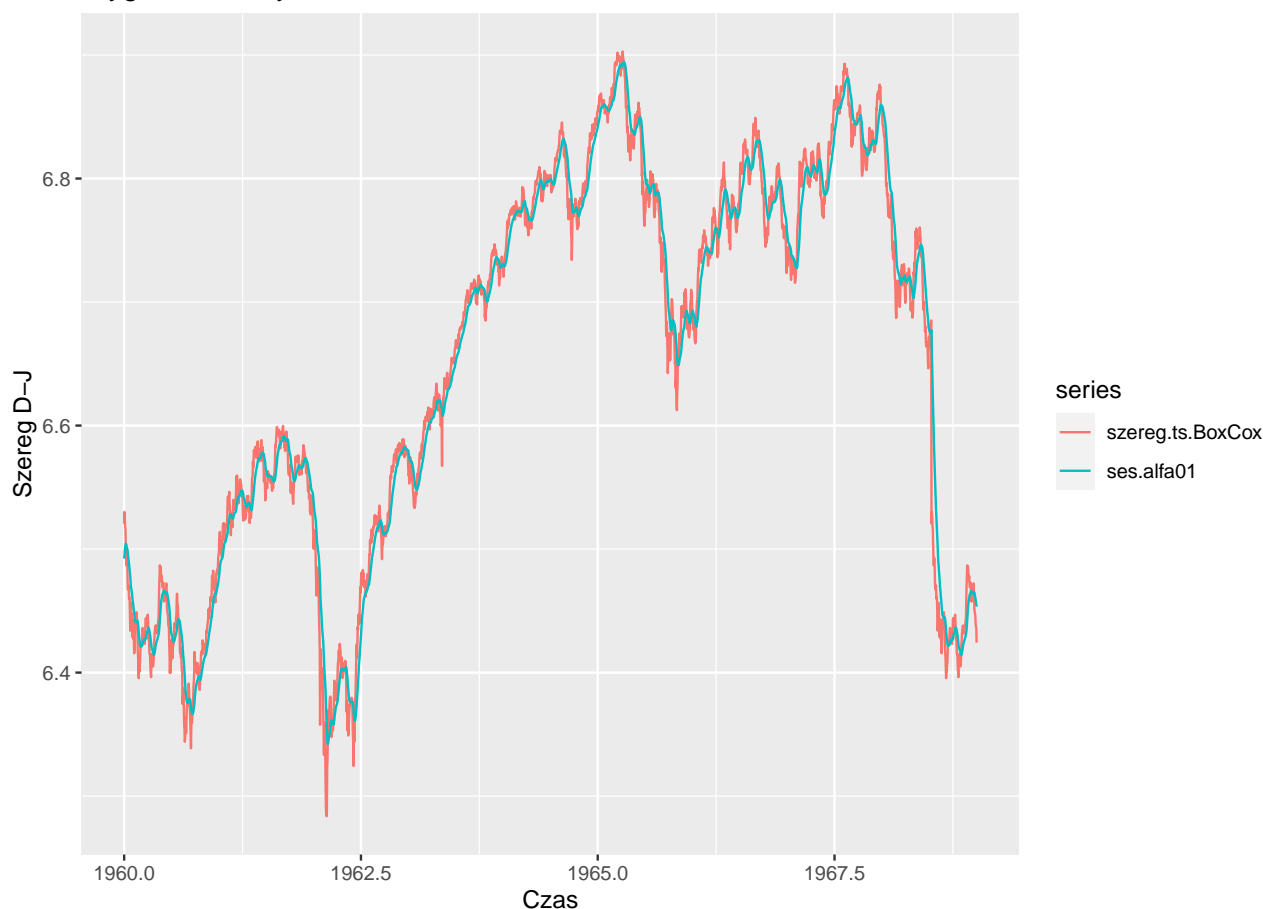
```
ses.alfa01 <- ses(szereg.ts.BoxCox, alpha=0.1)$fitted
ses.alfa04 <- ses(szereg.ts.BoxCox, alpha=0.4)$fitted
ses.alfa.optim <- ses(szereg.ts.BoxCox)$fitted
dane.ses <- ts.union(szereg.ts.BoxCox, ses.alfa.0.1=ses.alfa01, ses.alfa.0.4=ses.alfa04, ses.alfa.optimalne=ses.alfa.optim)
autoplot(dane.ses, main="Wygładzanie wykładnicze", lwd=0.1) +
  labs(x='Czas', y='Szereg D-J')
```



Rysunek 12: Estymacja trendu przy użyciu SES dla szeregu D-J z lat 1960-1969

Rysunek nr 12 jest raczej mało czytelny z uwagi na szeroki okres czasu, w którym estymujemy trend. Weźmy zatem pod uwagę  $\alpha = 0.01$ .

#### Wyglądanie wykładnicze



Rysunek 13: Estymacja trendu przy użyciu SES dla  $\alpha = 0.1$

Rysunek nr 13 sugeruje, że estymacja trendu przy pomocy SES dla parametru  $\alpha = 0.01$  radzi sobie całkiem dobrze, gdyż także zachowuje wiele informacji na temat lokalnych wahań badanego szeregu czasowego.

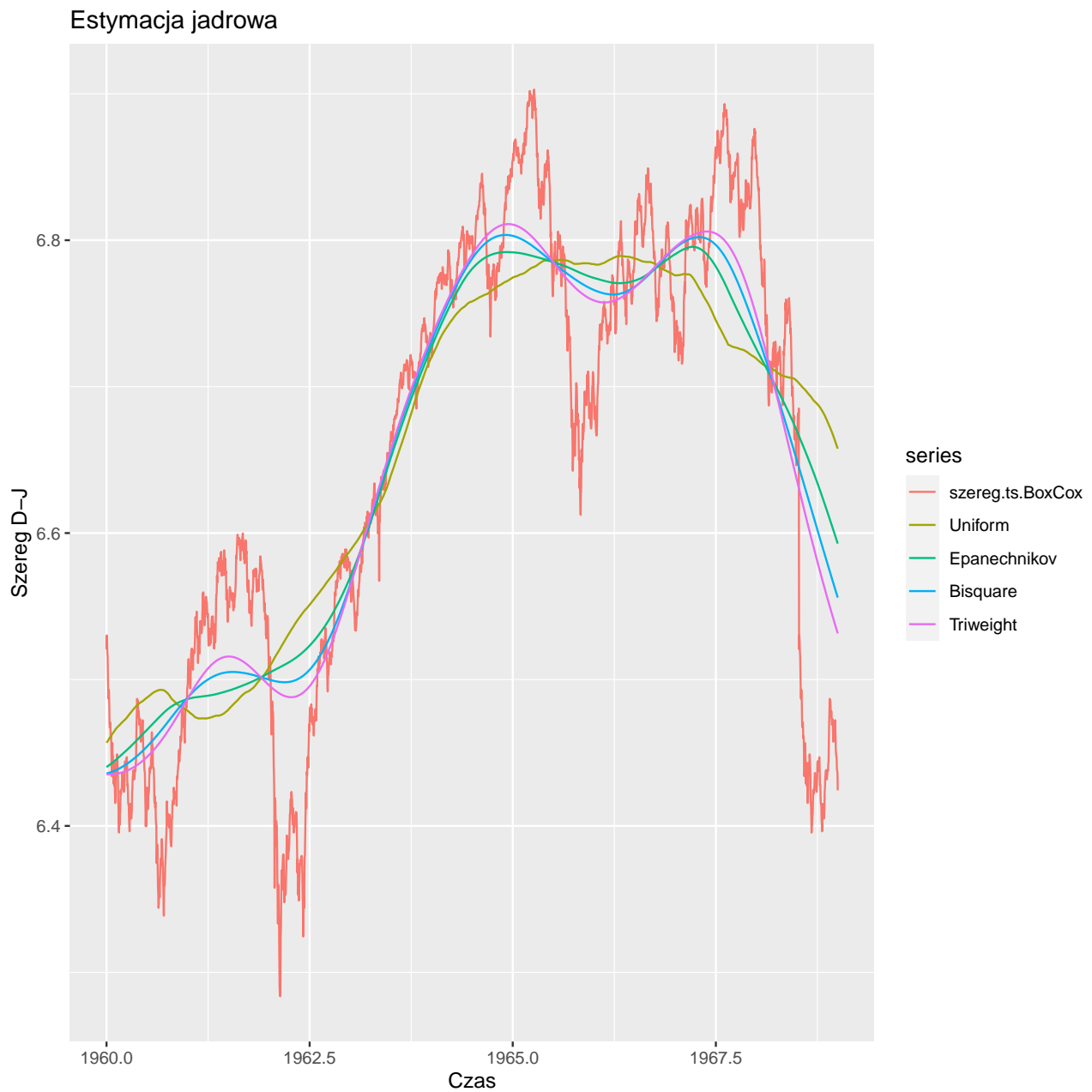
(e) Estymator jądrowy typu Nadaraya-Watsona - jest to estymator postaci

$$\hat{f}(t) = \hat{f}\left(\frac{t}{n}\right), t = 1, 2, \dots, n,$$

gdzie  $\hat{f}\left(\frac{t}{n}\right)$  jest jądrowym estymatorem funkcji regresji. Przeanalizujemy teraz, jaki estymator typu Nadaraya-Watsona dobierzemy do estymacji trendu, uwzględniając w wyborze różne jądra oraz różne szerokości okien  $h$ . W eksperymentach użyjemy jądra jednostajnego, Epanechnikova oraz typu bisquare i triweight, korzystając z funkcji *knsmooth* z paczki *smooths*.

- Bandwidth=0.15

```
autoplot(szeregi_trend_kernel_0.15, lwd=.5, main='Estymacja jądrowa') +  
  labs(x='Czas', y='Szereg D-J')
```

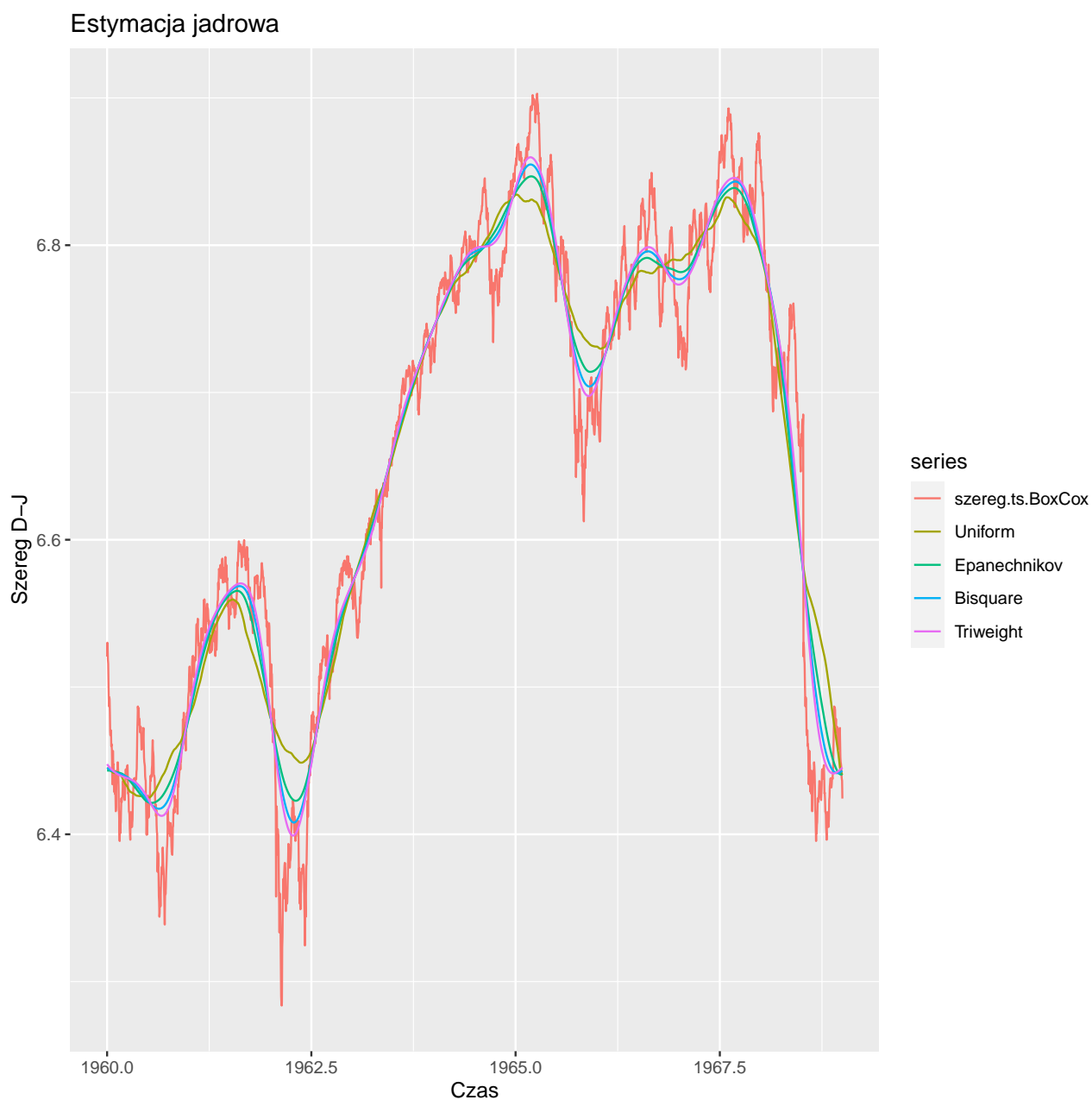


Rysunek 14: Estymacja jądrowa typu Nadaraya-Watsona trendu dla szeregu D-J z lat 1960-1969, bandwidth=0.15

Z rysunku nr 14 wynika, że powinniśmy nieco lepiej kontrolować szerokość okna.

- Bandwidth=0.05

```
autoplot(szeregi_trend_kernel_0.05, lwd=.5, main='Estymacja jądrowa') +  
  labs(x='Czas', y='Szereg D-J')
```

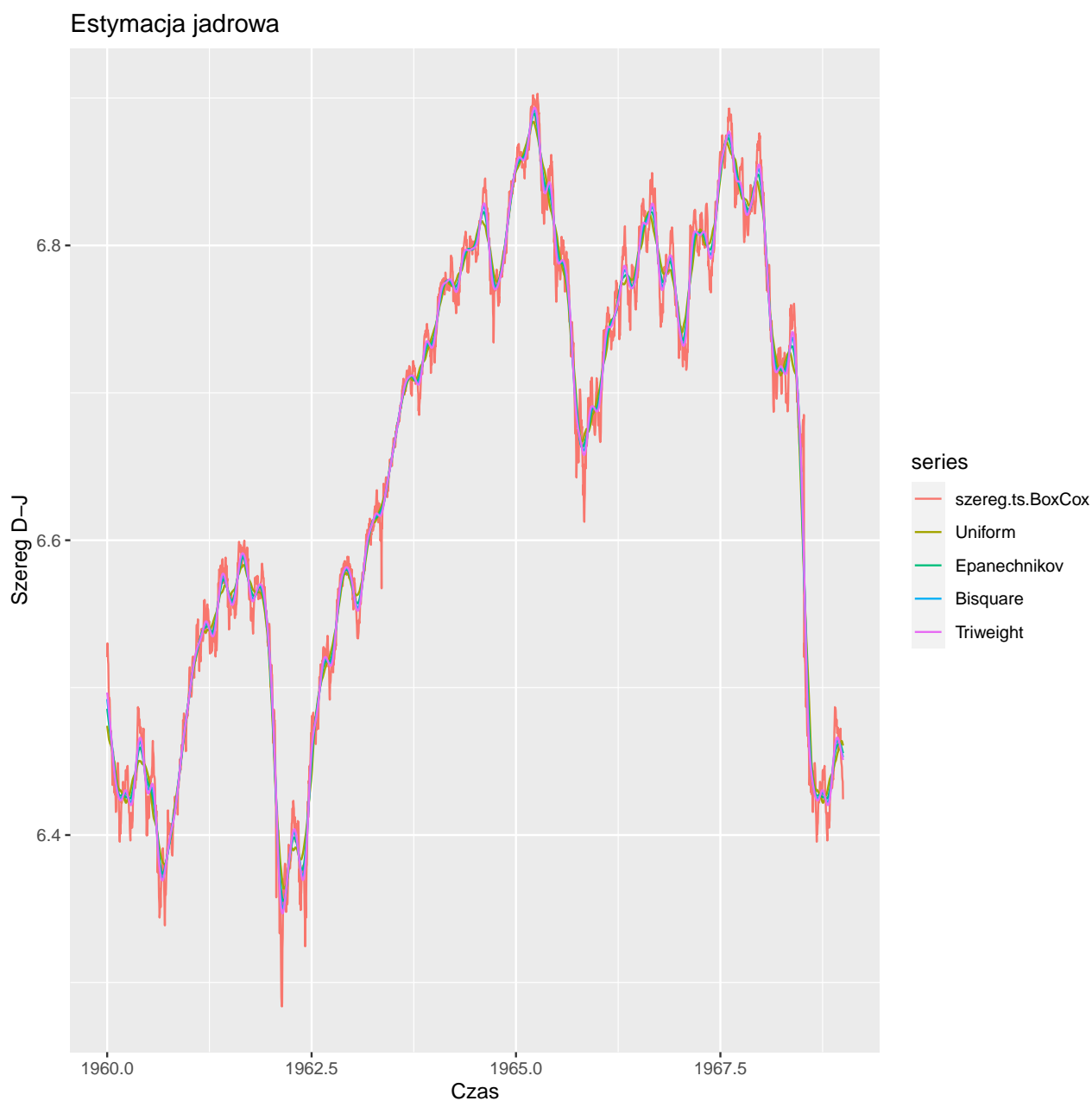


Rysunek 15: Estymacja jądrowa typu Nadaraya-Watsona trendu dla szeregu D-J z lat 1960-1969, bandwidth=0.05

Z rysunku nr 15 wynika, że estymacja jądrowa wciąż nie oddaje dobrze zmienności szeregu czasowego.

- Bandwidth=0.01

```
autoplot(szeregi_trend_kernel_0.01, lwd=.5, main='Estymacja jądrowa') +
  labs(x='Czas', y='Szereg D-J')
```



Rysunek 16: Estymacja jądrowa typu Nadaraya-Watsona trendu dla szeregu D-J z lat 1960-1969, bandwidth=0.01

Za estymator jądrowy trendu typu Nadaraya-Watsona przyjmujemy zatem estymator z jądrem typu Triweight i szerokością okna równą 0.05. Możemy także zaobserwować następujący wniosek.

**Wniosek 3** *Badane przez nas jądra dają podobny rezultat w nieparametrycznej metodzie estymacji jądrowej. Na dokładność estymacji większy wpływ ma wybór szerokości okna.*

- (f) Metoda estymacji przy użyciu naturalnych sześciennych funkcji B-sklejanych - przyjmujemy, że trend jest postaci

$$\hat{f}_t = \sum_{i=0}^q \hat{\beta}_i B_{i,n}(t),$$

gdzie współczynniki  $\hat{\beta}_0, \dots, \hat{\beta}_q$  są wyliczone metodą najmniejszych kwadratów, a funkcje  $B_{i,n}(t)$  są tak zwanymi funkcjami bazowymi B-sklejanymi zdefiniowanymi rekurencyjnie jako

$$\begin{aligned} B_{i,0}(t) &:= \mathbb{1}_{\{t_i \leq t \leq t_{i+1}\}}(t) \\ B_{i,n}(t) &:= \frac{t - t_i}{t_{i+n} - t_i} B_{i,n-1}(t) + \frac{t_{i+n+1} - t}{t_{i+n+1} - t_{i+1}} B_{i+1,n-1}(t) \end{aligned} \quad (1)$$

, gdzie punkty  $t_i$  są tzw. węzłami (posortowanymi niemalejąco), a parametr  $q$  jest rzędem metody. Zwizualizujmy tę metodę estymacji i sprawdźmy, jak działa.

```
oopt = ani.options(interval=0.1)

for (p in seq(10,70,by=10)) {
  trend.funkcja.sklejana <- tslm(szereg.ts.BoxCox~ns(trend, p))$fitted
  modele <- ts.union(trend.funkcja.sklejana, szereg.ts.BoxCox)
  g <- autoplot(modele, lwd=1) +
    ggtitle(paste0("Natural cubic B-spline functions: order = ", p)) +
    labs(x='Czas', y='Szereg D-J')
  plot(g)
  ani.pause()
}
```

Rysunek 17: Wybór rzędu wygładzenia szeregu czasowego D-J z lat 1960-1969 dla naturalnych sześciennych funkcji B-sklejanych

```
ani.options(oopt)
```

Dla  $q = 30$  metoda dobrze wygładziła szereg i zadziałała bardzo dobrze pomimo dynamicznego charakteru szeregu czasowego Dow-Jones.

**Notka 2** W metodzie tej w prosty sposób możemy uwzględnić sezonowość, dodając tzw. *seasonal factors* oznaczone jako  $s_t$  (a właściwie to ich estymatory) w następujący sposób

$$\hat{m}_t + \hat{s}_t = \sum_{i=0}^q \hat{\beta}_i B_{i,n}(t)$$

Porównajmy teraz ze sobą przeprowadzone symulacje (poza różnicowaniem, gdyż jest to jedynie metoda eliminacji trendu z szeregu czasowego). Naszym zdaniem najsłabiej prezentuje się estymacja trendu filtrem Spencera, ponieważ metoda ta stara się tak bardzo dopasować do szeregu, że nie wygładza go wystarczająco dobrze. Pozostałe metody prezentują się całkiem dobrze i ciężko jest nam wyłonić jednoznacznego zwycięzcę w konkursie na najlepszy estymator trendu. Natomiast cechą charakterystyczną dla metod nieparametrycznych jest fakt, że potrafią w rozsądny sposób opisać dynamiczny charakter zmian nieregularnych (ale bez sezonowości) szeregów czasowych.

**Wniosek 4** *Metody nieparametryczne potrafią dobrze dopasować się do danych, zachowując wiele cennych informacji o lokalnych zmianach szeregu czasowego, czyli miejscach, w których obserwujemy "piki".*

## 2.3 Parametryczna estymacja trendu

W tej podsekcji zajmiemy się parametryczną estymacją trendu. Podejście takie, jak zobaczymy, z jednej strony jest mniej elastyczne, ale z drugiej pozwala nam explicite znaleźć analityczny wzór na postać trendu, co jest szczególnie istotne przy ewentualnym zagadnieniu dotyczącym prognozowania szeregów czasowych. W symulacjach rozpatrzymy estymatory trendu jako skończone kombinacje ciągłych funkcji pochodzących z rodziny wielomianów oraz wielomianów trygonometrycznych.

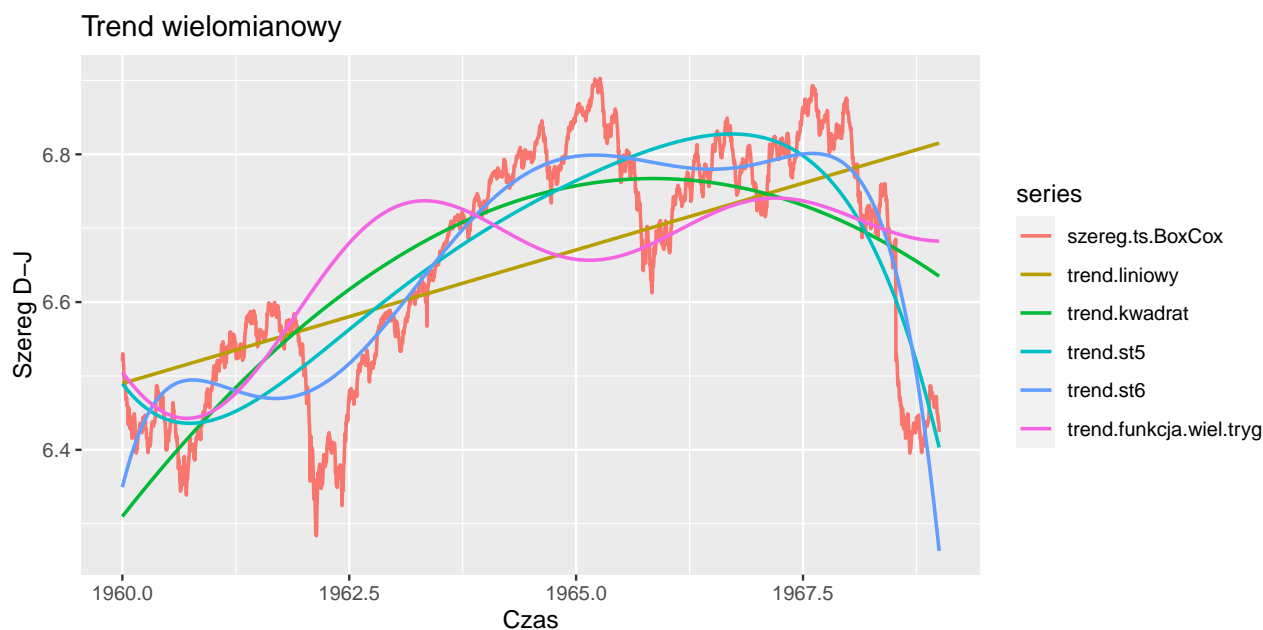
Przystąpmy do skonstruowania odpowiednich modeli, tzn. stwórzmy model liniowy, kwadratowy, model odpowiadający wielomianowi stopnia piątego, szóstego oraz model odpowiadający wielomianowi trygonometrycznemu w sposób przedstawiony w poniższym kodzie.

```
trend.liniowy <- tslm(szereg.ts.BoxCox~trend)$fitted
trend.kwadrat <- tslm(szereg.ts.BoxCox~I(trend^2)+trend)$fitted
trend.st5 <- tslm(szereg.ts.BoxCox~poly(trend, degree=5, raw=TRUE))$fitted
trend.st6 <- tslm(szereg.ts.BoxCox~poly(trend, degree=6, raw=TRUE))$fitted

wielomian.tryg <- function(x,n,s) {
  x <- ts(x)
  y <- 0
  for (k in 1:n) {
    y <- y + sin(k*x/s) + cos(k*x/s)
  }
  return(y)
}

trend.funkcja.wiel.tryg <- tslm(szereg.ts.BoxCox~I(wielomian.tryg(trend,4, 800)))$fitted
```

Zobaczmy teraz, jak sprawdzają się poszczególne modele w praktyce.



Rysunek 18: Estymacja parametryczna trendu dla szeregu czasowego D-J dla lat 1960-1969

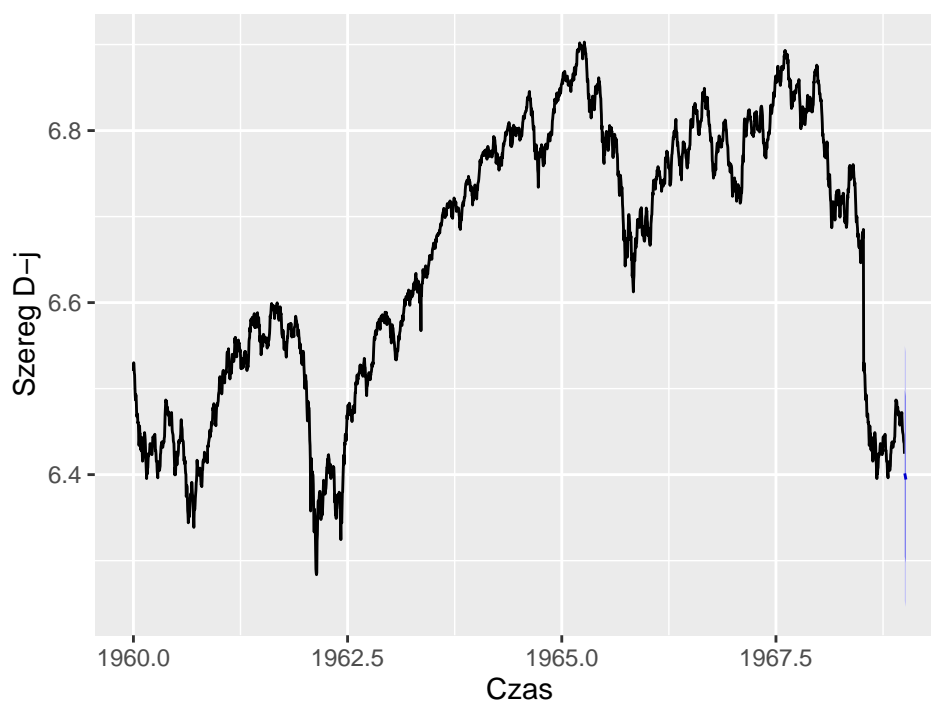
Wielomiany trygonometrycznie nie są odpowiednim modelem do estymacji parametrycznej trendu. W zasadzie nie powinno nas to dziwić - w końcu nie obserwujemy sezonowości w badanym szeregu czasowym. Skupmy się zatem na wielomianach. Który z modeli wybrać? Najlepiej estymujące trend są wielomiany stopnia piątego i szóstego. Wielomiany te dają jednak na tyle podobne rezultaty, że lepszą taktyką będzie przyjęcie tego prostszego modelu - wielomianu stopnia piątego. Sprawdźmy jeszcze, czy nie mamy do czynienia z **overfittingiem**. W tym celu wyznaczmy odpowiednie prognozy z horyzontem prognozy równym 5.

```
tslm.1 <- tslm(szereg.ts.BoxCox~poly(trend, degree=5, raw=TRUE))
tslm.2 <- tslm(szereg.ts.BoxCox~poly(trend, degree=6, raw=TRUE))

# wyznaczamy prognozy na podstawie dopasowanego trendu wielomianowego
prognozy.1 <- forecast::forecast(tslm.1, h=5)
prognozy.2 <- forecast::forecast(tslm.2, h=5)
```

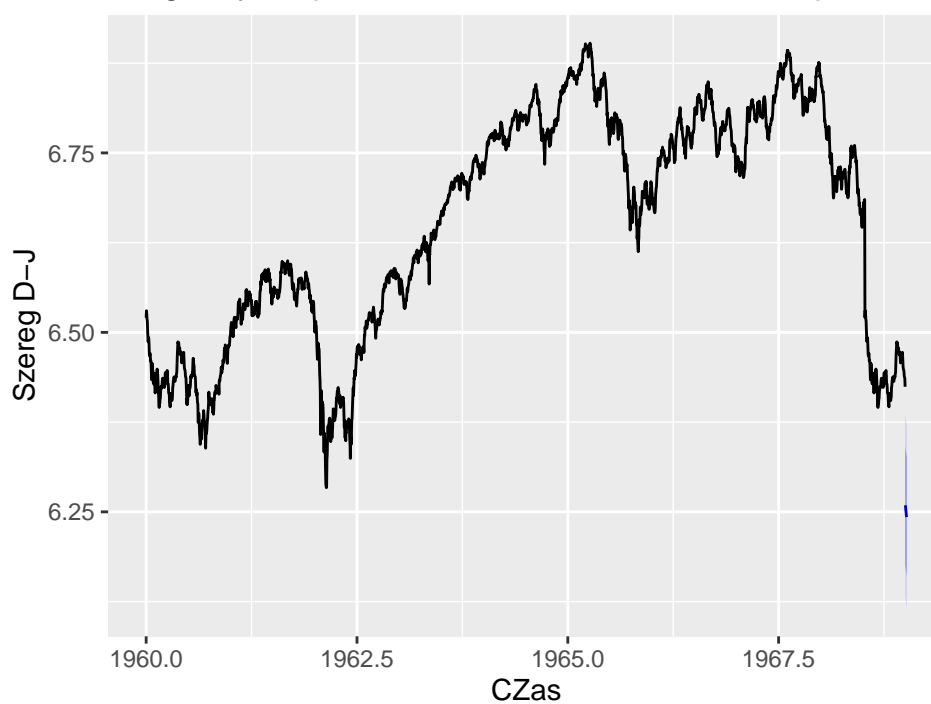


Prognozy na podstawie trendu wielomianu stopnia 5.



Rysunek 19: Prognozy na podstawie trendu wielomianowego stopnia piątego dla szeregu czasowego D-J dla lat 1960-1969

Prognozy na podstawie trendu wielomianu stopnia 6.



Rysunek 20: Prognozy na podstawie trendu wielomianowego stopnia szóstego dla szeregu czasowego D-J dla lat 1960-1969

Jak widzimy, zdecydowanie lepszym posunięciem jest przyjęcie wielomianu stopnia piątego jako parametryczny estymator trendu także dlatego, że prognozy przewidywane przez model z wielomianem 6. stopnia są w zasadzie nie do przyjęcia. Taka sytuacja pokazuje też, że nie warto mierzyć dokładności modelu przy pomocy jednej miary, np. przy pomocy  $R^2$ , które byłoby generalnie nie mniejsze, niż  $R^2$  dla wielomianu stopnia piątego.

**Wniosek 5** *W przypadku parametrycznej estymacji trendu lepiej jest przyjąć prostszy model, dający podobne lub nieco gorsze rezultaty, aby uniknąć overfittingu.*

### 3 Estymacja, eliminacja trendu oraz wstępna identyfikacja modelu dla 1965 r.

W sekcji tej, tak jak i w tytule, zajmiemy się estymacją, eliminacją oraz identyfikacją modelu dla szeregu czasowego notowań indeksu Dow-Jonesa w 1965 roku.

#### 3.1 Wstępne przekształcenie danych

Z rysunku nr 2 widzimy, że potrzebna jest nam operacja ujednolicenia wariancji. Uzyskamy ją poprzez transformację potęgową Boxa-Coxa, przy czym rozważymy wybór różnych parametrów przy pomocy animacji. Możemy ją bez problemów zastosować, gdyż wartości szeregu czasowego są dodatnie.

Rysunek 21: Wybór rzędu wygładzenia szeregu czasowego D-J z lat 1960-1969 dla naturalnych sześciennych funkcji B-sklejanych

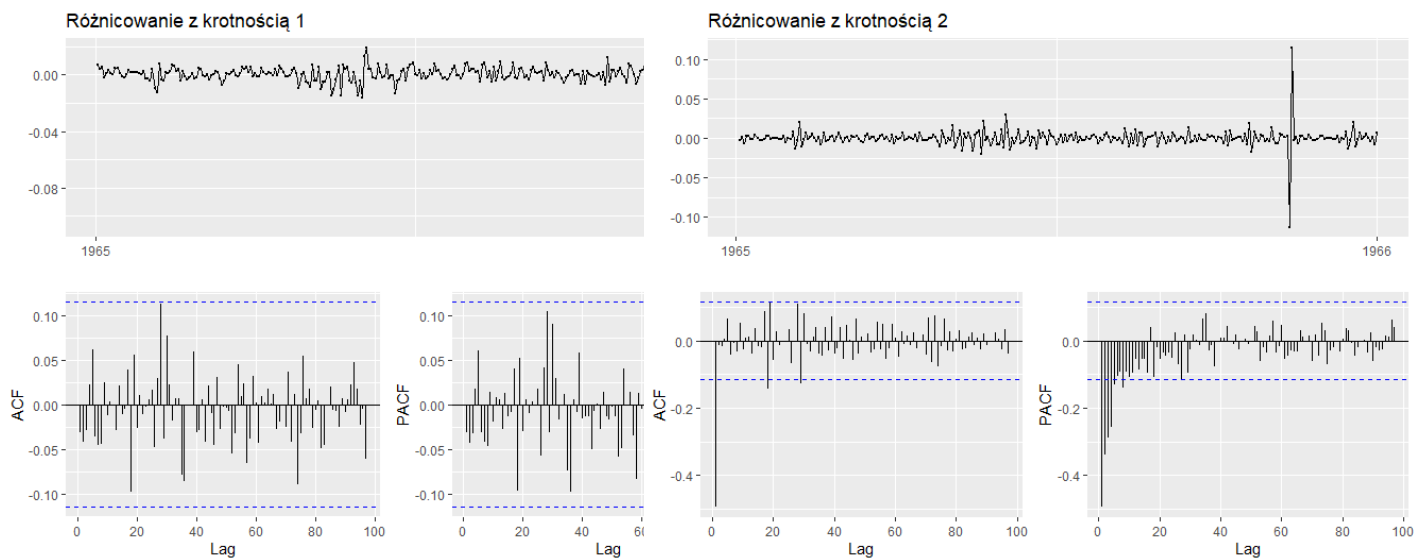
Znów najlepszym wyborem będzie przyjęcie parametru  $\lambda = 0$ . W szczególności zauważmy, że  $\lambda$  przyjęta automatycznie przez R wynosi 1, co potwierdza wniosek wynięty przez nas w poprzedniej sekcji. Szereg czasowych po zlogarytmowaniu wygląda o wiele lepiej.

```
BoxCox.lambda(szereg.ts.1965)
```

```
## [1] 1
```

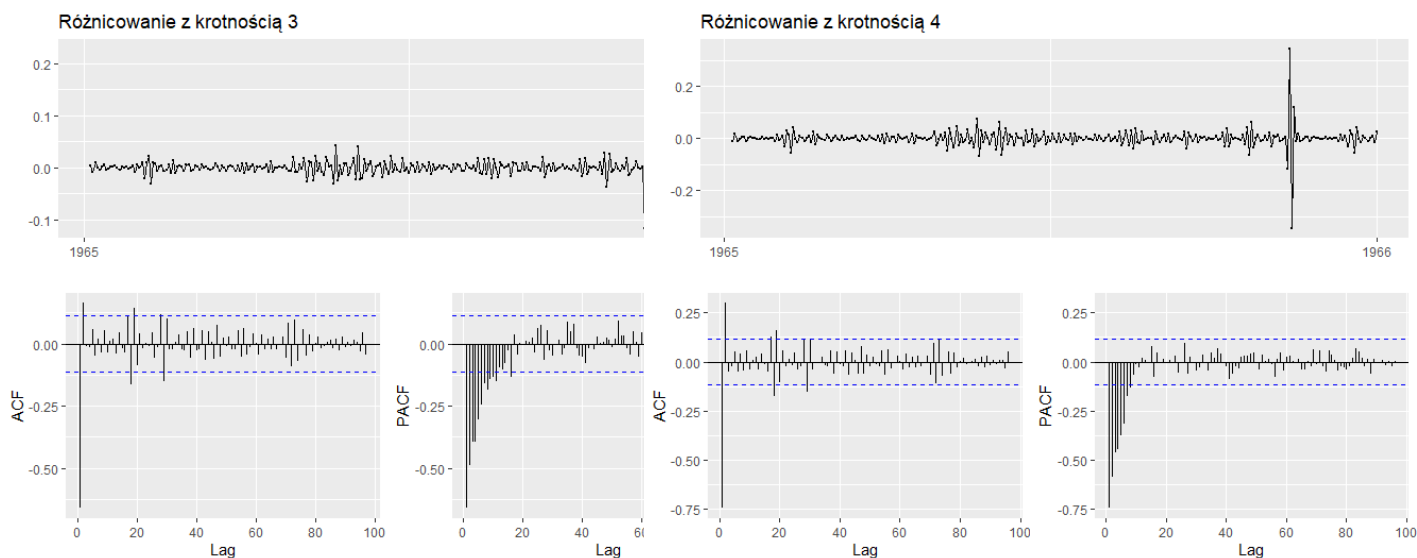
### **3.2 Nieparametryczna estymacja i eliminacja trendu**

- (a) Różnicowanie - zacznijmy od narysowania zróżnicowanych szeregów.



Rysunek 22: Efekt jednokrotnego różnicowania dla szeregu D-J w 1965 r.

Rysunek 23: Efekt dwukrotnego różnicowania dla szeregu D-J w 1965 r.



Rysunek 24: Efekt potrójnego różnicowania dla szeregu D-J w 1965 r.

Rysunek 25: Efekt poczwórnego różnicowania dla szeregu D-J w 1965 r.

Tak jak wcześniej, jednokrotne różnicowanie jest już satysfakcjonujące dla nas. Więcej szczegółów dotyczących otrzymanego wyniku na temat tego, jak bardzo "jednokrotne różnicowanie jest satysfakcjonujące" omówimy podsekcji dotyczącej identyfikacji modelu.

- (b) Wyglądanie poprzez symetryczną ruchomą średnią - w metodzie tej przyjęliśmy rząd metody  $(2q + 1)$  równy 25.

Rysunek 26: Wybór rzędu wygładzenia szeregu czasowego D-J z 1965 roku metodą symetrycznej ruchomej średniej

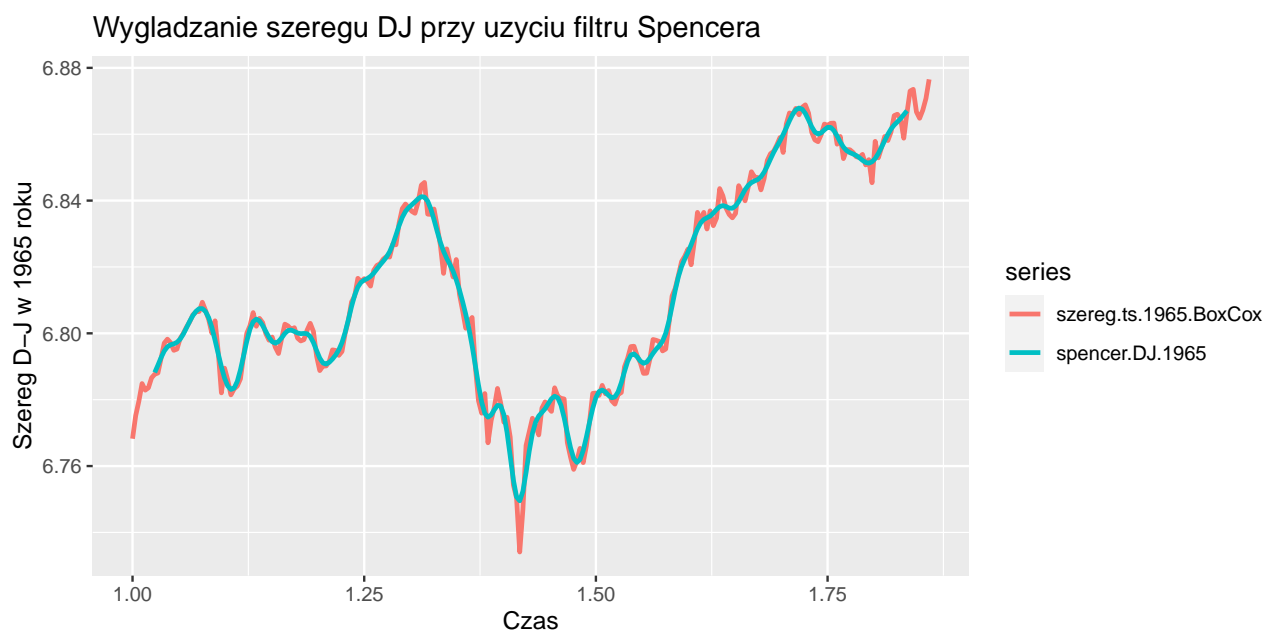
Na animacji na rysunku nr 26 możemy dopatrzeć się "uciętych" wartości po obu stronach szeregu. Na animacji z rysunku nr 8 jest to także widoczne, aczkolwiek nie aż tak widoczne z powodu o wiele dłuższego szeregu czasowego. Ucięte wartości wynikają z samej konstrukcji metody symetrycznej średniej ruchomej.

**Wniosek 6** *Powinniśmy dokładnie rozważyć użycie symetrycznej średniej ruchomej do estymacji trendu w przypadku szeregów czasowych o małej długości.*

(c) Filtr Spencera - wiuzalizujemy metodę

```
spencer.DJ.1965 <- spencer(szereg.ts.1965.BoxCox) #estymowany trend
# Uwaga: konieczna jest konwersja na obiekt 'ts'
spencer.DJ.1965 <- ts(spencer.DJ.1965, frequency=292, start=start(szereg.ts.1965),
                      end=end(szereg.ts.1965))

autoplot(ts.union(szereg.ts.1965.BoxCox, spencer.DJ.1965), lwd=1,
         main='Wygładzanie szeregu DJ przy użyciu filtru Spencera') +
  labs(y='Szereg D-J w 1965 roku', x='Czas')
```



Rysunek 27: Estymator trendu uzyskany przy użyciu filtra Spencera dla szeregu D-J dla 1965 roku

**Wniosek 7** *Estymacja trendu poprzez filtr Spencera jest o wiele dokładniejsza (dobre dopasowanie do szeregu + wygładzenie) dla szeregów czasowych o mniejszej długości.*

- (d) Proste wygładzanie wykładnicze (SES) - zwiualizujemy metodę i dobierzmy odpowiedni parametr. Wybraliśmy algorytm SES z tych samych względów, co w poprzedniej sekcji.

```
ses.alfa01.1965 <- ses(szereg.ts.1965.BoxCox, alpha=0.1)$fitted
ses.alfa02.1965 <- ses(szereg.ts.1965.BoxCox, alpha=0.2)$fitted
ses.alfa04.1965 <- ses(szereg.ts.1965.BoxCox, alpha=0.4)$fitted
ses.alfa.optim.1965 <- ses(szereg.ts.1965.BoxCox)$fitted
dane.ses.1965 <- ts.union(szereg.ts.1965.BoxCox, ses.alfa.0.1=ses.alfa01.1965,
                        ses.alfa.0.4=ses.alfa04.1965,
                        ses.alfa.optymalne=ses.alfa.optim.1965,
                        ses.alfa.0.2=ses.alfa02.1965)

autoplot(dane.ses.1965, main="Wygładzanie wykładnicze", lwd=.8) +
  labs(x='Czas', 'Szereg D-J')
```

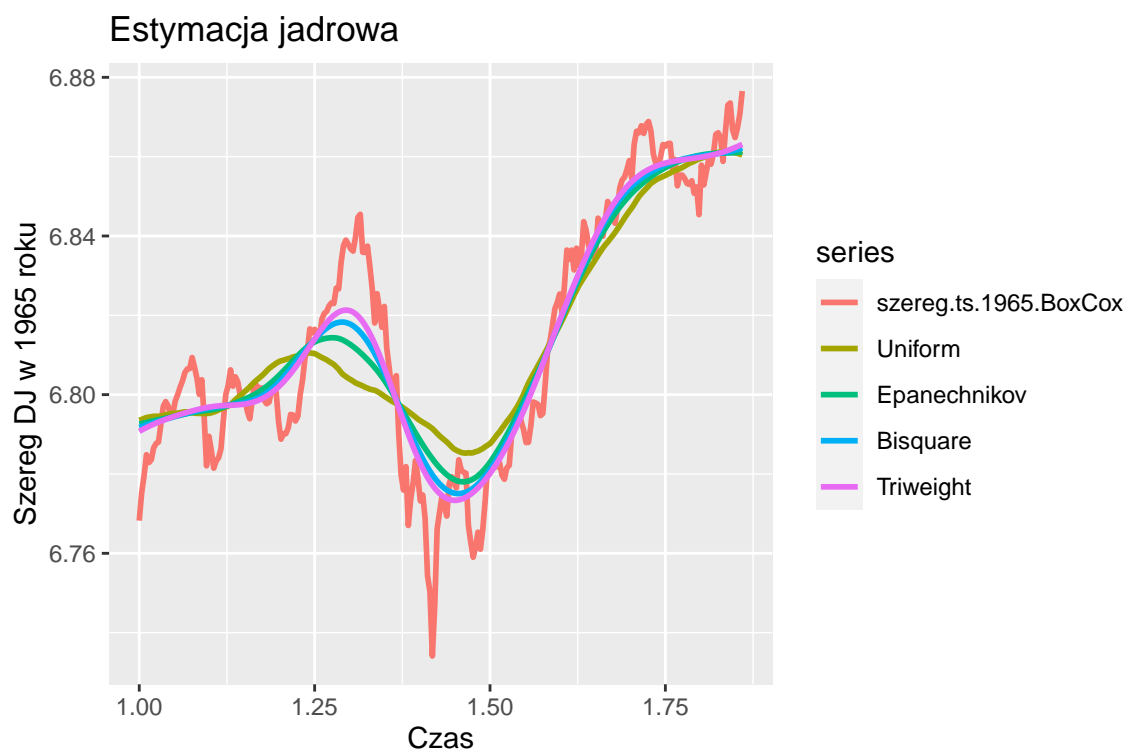


Rysunek 28: Estymacja trendu przy użyciu SES dla szeregu D-J z lat 1965 roku

Z dużym powodzeniem możemy przyjąć  $\alpha = 0.2$ . Z pewnością metoda ta poradziła sobie lepiej, niż estymacja metodą symetrycznej średniej ruchomej, w szczególności "na brzegach" szeregu.

- (e) Jądrowy estymator typu Nadaraya-Watsona - wizualizujemy tę metodę estymacji trendu i, podobnie jak wcześniej, rozpatrujemy jądro jednostajne, Epanechnikova, Bisquare oraz Triweight przy różnych szerokościach okna  $h$ .

- $h = 0.15$

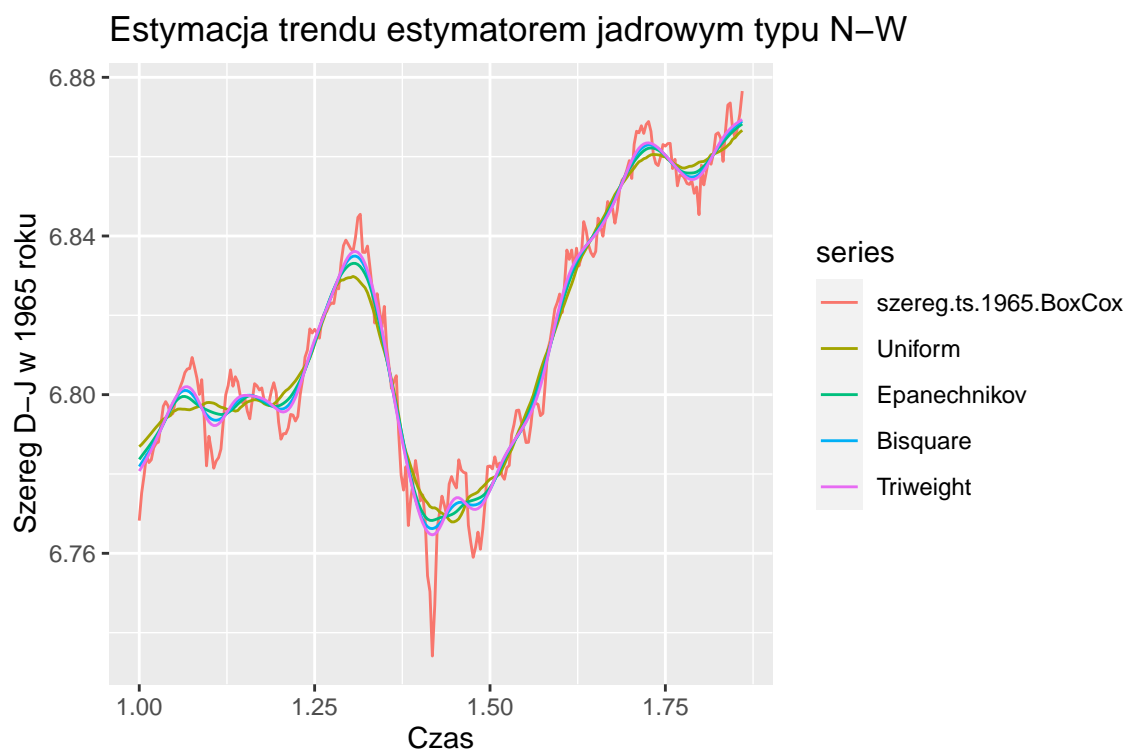


Rysunek 29: Estymacja jądrowa typu Nadaraya-Watsona trendu dla szeregu D-J z 1965 roku,  $\text{bandwidth}=0.15$

Z rysunku nr 29 wynika, że musimy lepiej dobrać szerokość okna.

- $h = 0.05$





Rysunek 30: Estymacja jądrowa typu Nadaraya-Watsona trendu dla szeregu D-J z 1965 roku,  $\text{bandwidth}=0.05$

Za estymator trendu w tym wypadku przyjmujemy estymator jądrowy typu Nadaraya-Watsona z jądrem typu Triweight i  $h = 0.05$ . Możemy także zauważyć następujący wniosek.

**Wniosek 8** *Zarówno w przypadku dłuższego, jak i krótszego szeregu czasowego, estymacja jądrowa typu Nadaraya-Watsona z różnymi przykładami jąder prowadzi do podobnych rezultatów. Jest to szczególnie widoczne przy małym parametrze bandwidth.*

- (f) Metoda estymacji przy użyciu naturalnych sześciennych funkcji B-sklejanych - wizualizujemy metodę i wybierzmy odpowiedni rząd  $q$  metody.

Rysunek 31: Wybór rzędu wygładzenia szeregu czasowego D-J z 1965 roku dla naturalnych sześciennych funkcji B-sklejanych

Znów najlepszym estymatorem będzie naszym zdaniem rząd  $q = 30$ .

**Wniosek 9** *Naturalne sześciennne funkcje B-sklejane są dobrymi estymatorami trendu dla krótkich i dłuższych szeregów czasowych.*

Podsumowując, to jaki nieparametryczny estymator trendu wybierzemy zależy od długości szeregu oraz dynamiki zmian, np. jeśli szereg czasowy będzie miał małą długość, to warto przemyśleć dobranie symetrycznej średniej ruchomej jako estymator trendu. Może to spowodować sporo problemów przy uzyskaniu prognozy szeregu, która będzie w jakiś sposób zależeć od uciętych przez średnią ruchomą obserwacji na "brzegach" szeregu. Teraz na odwrót, filtr Spencera powinien dobrze wygładzić krótszy szereg, natomiast może mieć problemy z wygładzeniem dłuższego szeregu, szczególnie jeśli występuje tam zauważalny dynamiczny charakter zmian. Najbardziej elastycznymi dla nas metodami nieparametrycznymi estymacji trendu jest metoda funkcji sklejanych oraz metoda estymacji jądrowej typu Nadaraya-Watsona. Proste wygładzanie wykładnicze także jest sensowne w użyciu, ale musimy pilnować wówczas, czy w badanym szeregu czasowym nie występują przypadkiem tendencje długoterminowe (lub efekty sezonowe - to nas nie dotyczy z uwagi na dobrany model w raporcie).

### 3.3 Parametryczna estymacja trendu

W podsekcji tej sprawdzimy, jak poradzą sobie sposoby estymacji parametrycznej dla 1965 roku użyte dla dekady lat 1960-1969. Tym razem przetestujemy modele oparte na wielomianach stopnia trzeciego, piątego, szóstego, siódmego oraz modelu opartego na wielomianie trygonometrycznym.

```
trend.cubic.1965 <- tslm(szereg.ts.1965.BoxCox~I(trend^3))$fitted
trend.st5.1965 <- tslm(szereg.ts.1965.BoxCox~poly(trend, degree=5,
                                                    raw=TRUE))$fitted
```

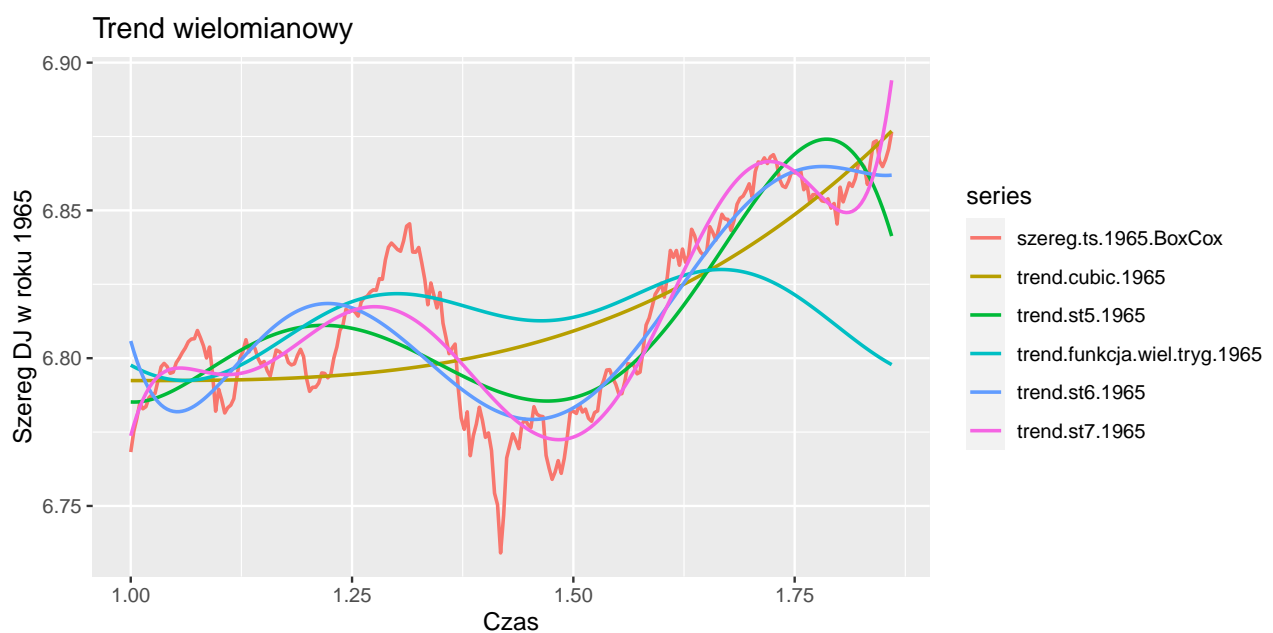
```

trend.st6.1965 <- tslm(szereg.ts.1965.BoxCox~poly(trend, degree=6,
                                                    raw=TRUE))$fitted

trend.st7.1965 <- tslm(szereg.ts.1965.BoxCox~poly(trend, degree=7,
                                                    raw=TRUE))$fitted

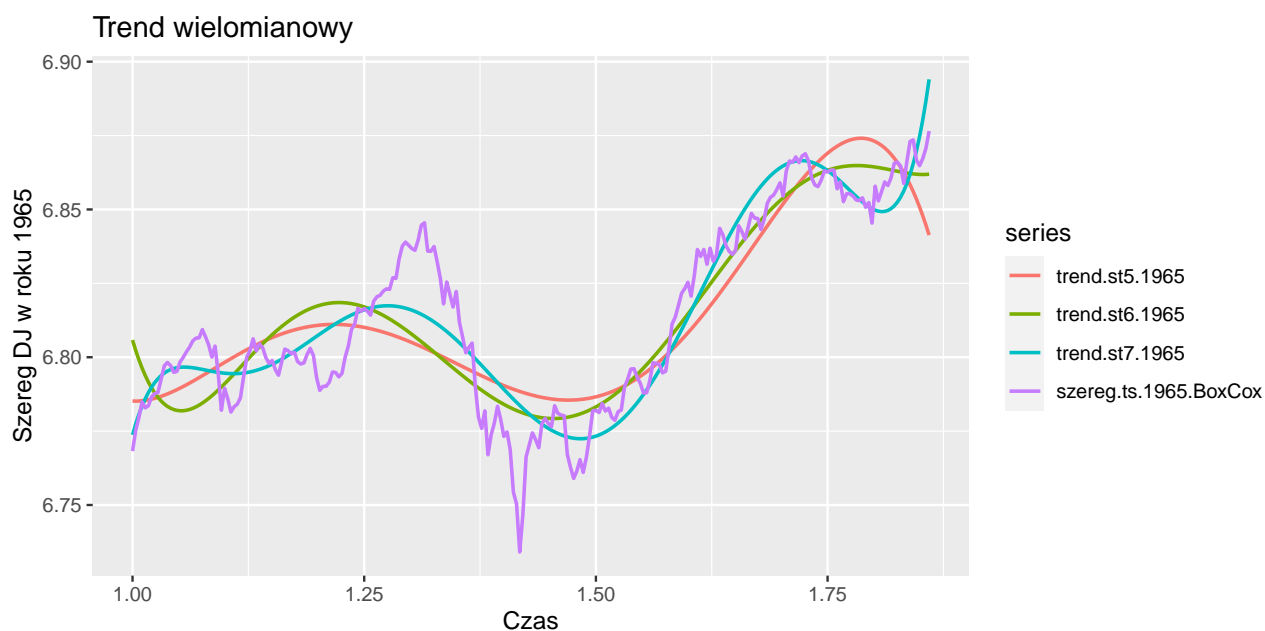
trend.funkcja.wiel.tryg.1965 <- tslm(szereg.ts.1965.BoxCox~I(wielomian.tryg(trend,2, 40)
ted
dane.trend.1965 <- ts.union(szereg.ts.1965.BoxCox, trend.cubic.1965, trend.st5.1965,
                           trend.funkcja.wiel.tryg.1965, trend.st6.1965, trend.st7.1965)
autoplot(dane.trend.1965, lwd=.75, main="Trend wielomianowy") +
  labs(x='Czas', y='Szereg DJ w roku 1965')

```



Rysunek 32: Estymacja parametryczna trendu dla szeregu czasowego D-J dla 1965 roku

Najlepiej radzą sobie wielomiany stopnia piątego, szóstego i siódmego. Wielomiany trygonometryczne znów zawodzą z powodów opisanych w sekcji dotyczącej metod parametrycznej estymacji trendu dla dekady (sekcja 2.3). Zrobmy jeszcze jeden, nieco bardziej przejrzysty, rysunek.



Rysunek 33: Estymacja parametryczna trendu wielomianami stopnia piątego, szóstego oraz siódmego dla szeregu czasowego D-J dla 1965 roku

Początkowo wydawałoby się, że najlepszym modelem jest wielomian siódmego stopnia. Sprawdźmy prognozy dla horyzontu  $h = 5$  dla wielomianów z rysunku nr 33, czy nie mamy do czynienia z **overfittingiem**.

```
tslm.1.1965 <- tslm(szereg.ts.1965.BoxCox~poly(trend, degree=5, raw=TRUE))
tslm.2.1965 <-tslm(szereg.ts.1965.BoxCox~poly(trend, degree=6, raw=TRUE))
tslm.3.1965 <-tslm(szereg.ts.1965.BoxCox~poly(trend, degree=7, raw=TRUE))

# wyznaczamy prognozy na podstawie dopasowanego trendu wielomianowego
prognozy.1.1965 <- forecast::forecast(tslm.1.1965, h=5)
prognozy.2.1965 <- forecast::forecast(tslm.2.1965, h=5)
prognozy.3.1965 <- forecast::forecast(tslm.3.1965, h=5)
```

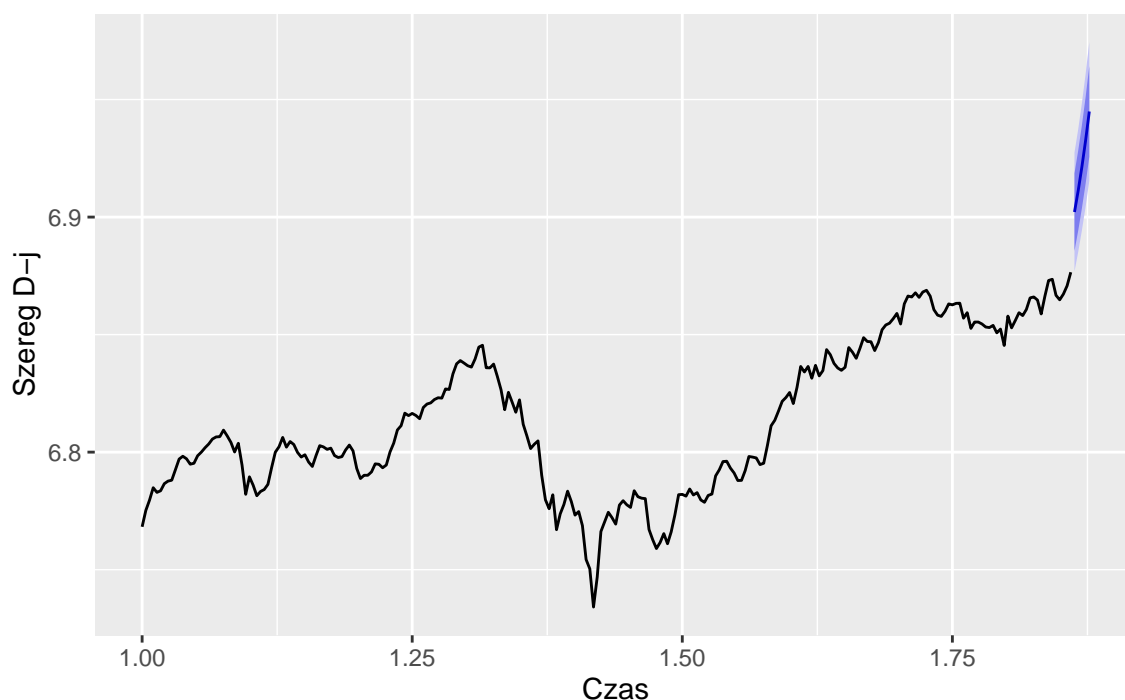


Rysunek 34: Prognozy na podstawie trendu wielomianu stopnia 5. dla szeregu czasowego D-J dla 1956 roku



Rysunek 35: Prognozy na podstawie trendu wielomianu stopnia 6. dla szeregu czasowego D-J dla 1956 roku

Prognozy na podstawie trendu wielomianu stopnia 7.



Rysunek 36: Prognozy na podstawie trendu wielomianu stopnia 7. dla szeregu czasowego D-J dla 1956 roku

Analizując rysunki nr 34, 35 i 36 dochodzimy do wniosku, że zdecydowanie powinniśmy odrzucić model estymacji trendu metodą, w której wykorzystujemy wielomian stopnia siódmego. Przyjmujemy zatem za estymator trendu wielomian stopnia szóstego. Porównując ten wynik symulacji z wynikiem uzyskanym w 2.3 dochodzimy do poniższych wniosków.

**Wniosek 10** *Nie można dopasować wielomianowego modelu estymacji trendu tak, aby był on dobry globalnie (w sekcji 2.3 uzyskaliśmy inny wynik symulacji).*

**Wniosek 11** *Parametryczny model estymatora trendu dopasowuje się lepiej w przypadku, gdy mamy do czynienia z szeregiem czasowym potencjalnie małej długości.*

Podobnie, jak w sekcji 2.3, parametryczny estymator nie jest tak bardzo elastyczny, jak estymatory nieparametryczne, ale za to otrzymujemy analityczny wzór estymatora trendu, co może być szczególnie użyteczne przy prognozowaniu szeregów czasowych. W związku z tym, powinniśmy zawsze dobierać metodę estymacji trendu w zależności od naszych potrzeb, oczekiwań oraz danych.

## 4 Wstępna identyfikacja modelu - przypadek lat 1960-1969 oraz 1965 roku.

W podpunkcie (d) sprawozdania nr 2 naszym zadaniem jest wyznaczenie procesu  $\hat{Z}_t = X_t - \hat{f}(t)$  dla wyznaczonych w podpunktach (b) i (c) estymatorów trendu dla dekady oraz pojedynczego roku. Za  $X_t$  przyjmujemy nasz bazowy szereg po transformacji logarytmicznej, natomiast za  $\hat{f}(t)$  przyjmujemy:

1. Dla dekady, nieparametryczne:

- Wygładzanie ruchomą średnią z parametrem rzędu równym 31
- Filtr spencera, frequency=292
- Wygładzanie wykładnicze z parametrem  $\alpha = 0.1$
- estymator jądrowy Naradaya-Watsona z parametrami  $\mu=3$ ,  $b=0.01$
- Naturalne sześciennie B-sklejane funkcje,  $q=30$

2. Dla dekady, parametryczne:

- wielomian stopnia 5

3. Dla roku 1965, nieparametryczne:

- Wygładzanie ruchomą średnią z parametrem  $q=25$
- Filtr Spencera, frequency=292
- Wygładzanie wykładnicze z parametrem  $\alpha = 0.2$
- Estymator jądrowy Naradaya-Watsona z parametrami  $\mu=3$ ,  $b=0.05$
- Naturalne sześciennie B-sklejane funkcje,  $q=30$

4. Dla roku 1965, parametryczne:

- Wielomian stopnia 6

Dodatkowo, w metodach nieparametrycznych  $Z_t$  przyjmujemy różnicowany szereg, zarówno dla dekady jak i dla roku 1965.

W celu określenia czy otrzymany szereg  $Z_t$  jest stacjonarny posłużymy się testami na białoszumowość, ponieważ jest to dobry wyznacznik stacjonarności. Zastosujemy test Ljunga-Boxa a także zaimplementowane w sprawozdaniu pierwszym: test graficzny oraz test Ljunga-Boxa z dostosowanym parametrem  $h$ .

Otrzymujemy wyniki tych testów w poniższych tabelach:

	graphic	ljung.box	adjusted.h
ma	FALSE	FALSE	FALSE
spencer	FALSE	FALSE	FALSE
exp	FALSE	FALSE	FALSE
NW	FALSE	FALSE	FALSE
BSklej	FALSE	FALSE	FALSE
Diff	FALSE	TRUE	FALSE

Tabela 1: Nieparametryczne, dekada

Widzimy, że prawie wszystkie testy białoszumowości nie odrzucają hipotezy zerowej o białoszumowości. Moglibyśmy zatem przyjąć, że nie są to szeregi stacjonarne. Wyjątkiem jest test Ljunga-Boxa z parametrem  $lag = \lfloor \frac{n}{4} \rfloor$  dla szeregu  $\hat{Z}_t$  będącego szeregiem różnicowanym dla całej dekady.

Przyjrzyjmy się jednak jeszcze wykresom ACF:

Skróty:

	graphic	ljung.box	adjusted.h
lin	FALSE	FALSE	FALSE
quad	FALSE	FALSE	FALSE
X5deg	FALSE	FALSE	FALSE
trig	FALSE	FALSE	FALSE

Tabela 2: Parametryczne, dekada

	graphic	ljung.box	adjusted.h
ma	FALSE	FALSE	FALSE
Spencer	FALSE	FALSE	FALSE
exp	FALSE	FALSE	FALSE
NW	FALSE	FALSE	FALSE
BSklej	FALSE	FALSE	FALSE
Diff	TRUE	TRUE	TRUE

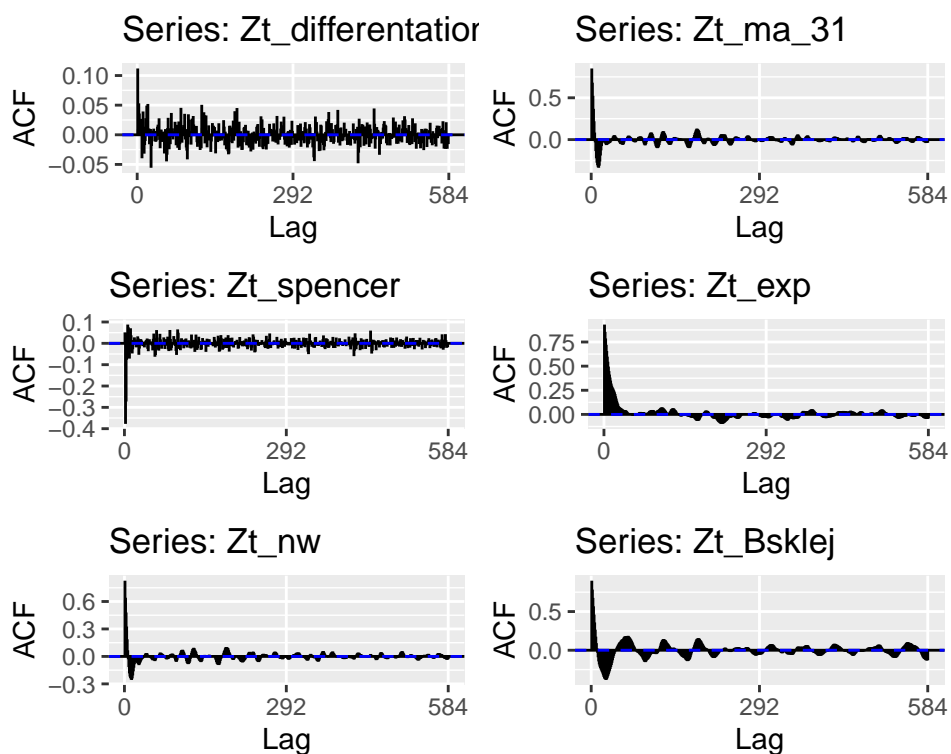
Tabela 3: Nieparametryczne, 1965

- ma - symetryczna ruchoma średnia
- spencer - filtr Spencera
- exp - SES
- Nw - estymator jądrowy typu Nadaraya-Watsona
- Diff - różnicowanie
- lin - wielomian liniowy
- quad - wielomian kwadratowy
- X5deg - wielomian stopnia piątego dla dekady
- trig - wielomian trygonometryczny
- BSklej - naturalna sześcienna funkcja B-sklejana
- cubic - wielomian stopnia trzeciego
- st5 - wielomian stopnia piątego, 1965 r.
- st6 - wielomian stopnia szóstego
- st7 - wielomian stopnia siódmego

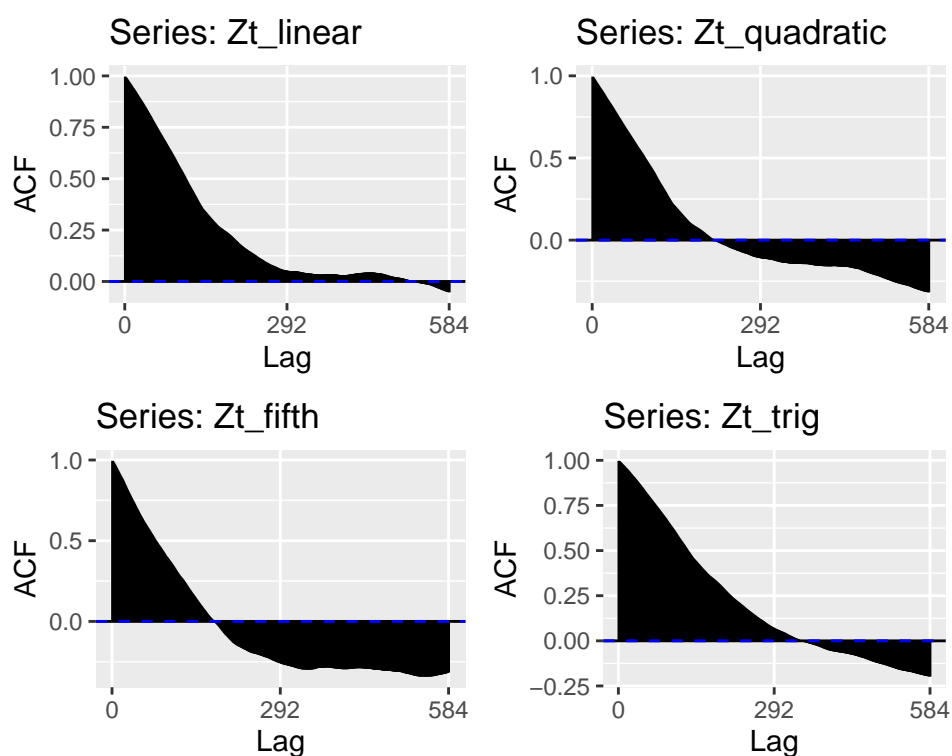


	graphic	ljung.box	adjusted.h
cubic	FALSE	FALSE	FALSE
st5	FALSE	FALSE	FALSE
st6	FALSE	FALSE	FALSE
st7	FALSE	FALSE	FALSE
trig	FALSE	FALSE	FALSE

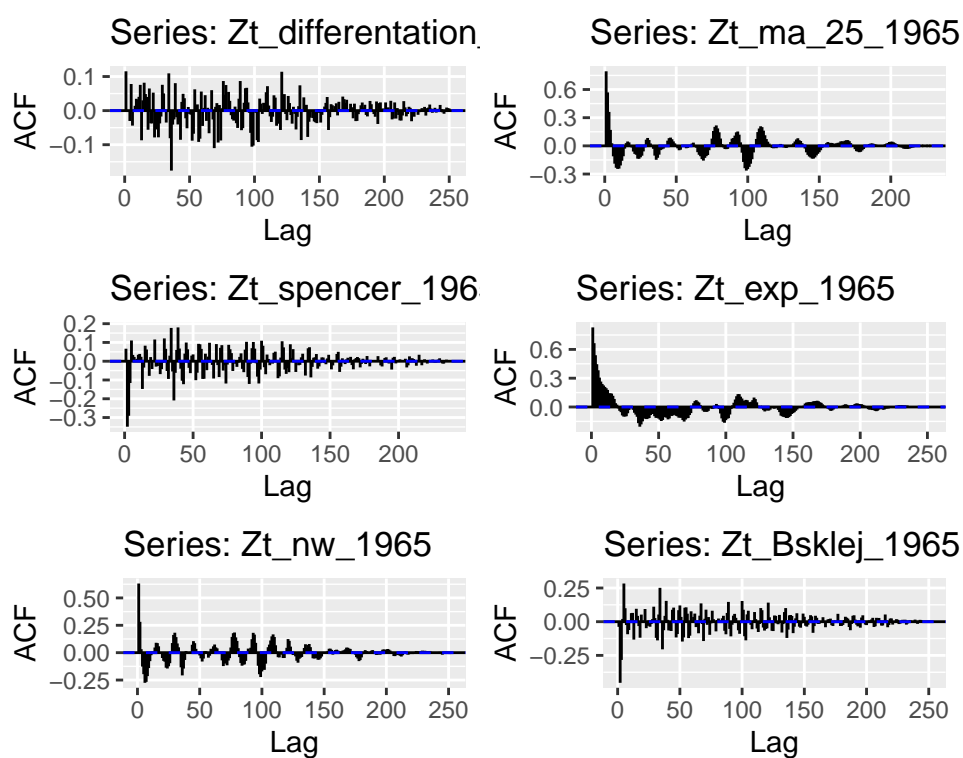
Tabela 4: Parametryczne, 1965



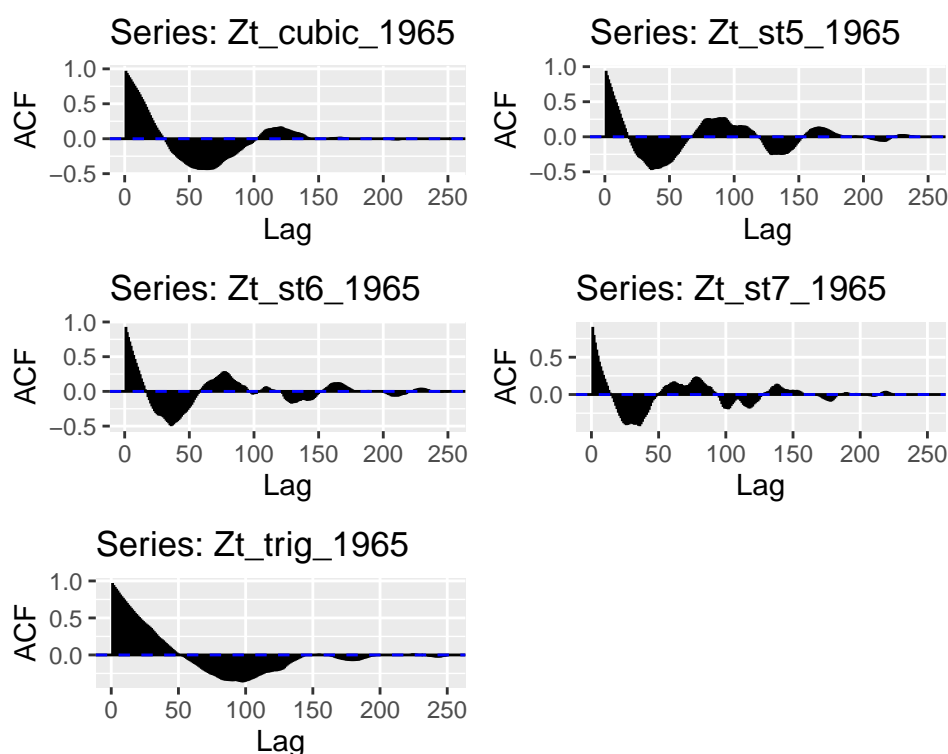
Rysunek 37: Wykresy ACF dla Zt uzyskanych z nieparametrycznych trendów, cała dekada



Rysunek 38: Wykresy ACF dla  $Z_t$  uzyskanych z parametrycznych trendów, cała dekada



Rysunek 39: Wykresy ACF dla  $Z_t$  uzyskanych z nieparametrycznych trendów, rok 1965

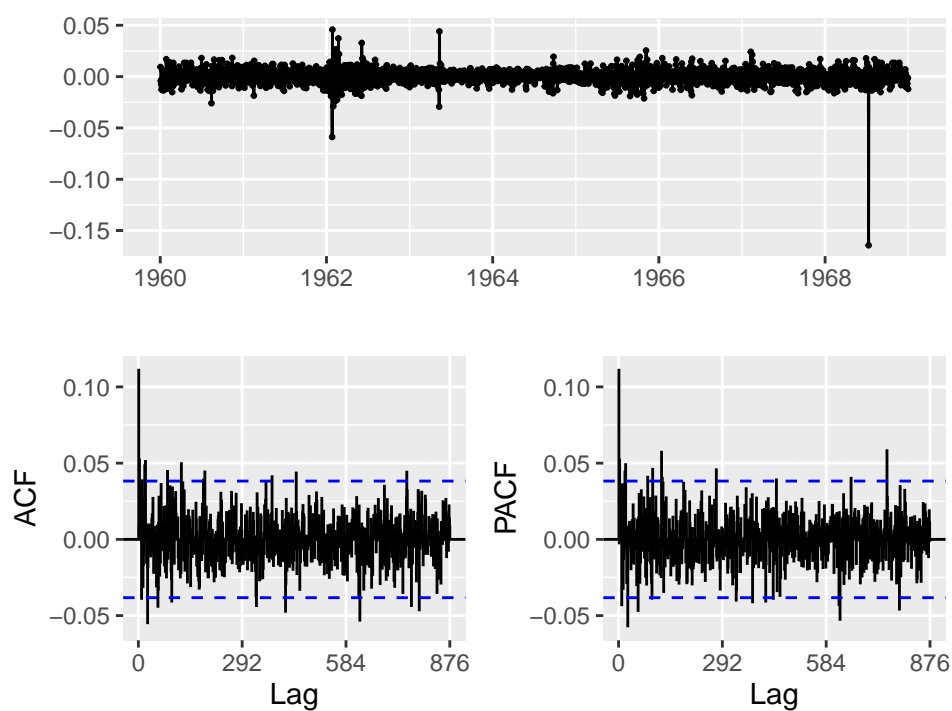


Rysunek 40: Wykresy ACF dla Zt uzyskanych z parametrycznych trendów, rok 1965

Istotnie, większość z tych szeregów bez zastanowienia możemy odrzucić, jeśli chodzi o stacjonarność (mimo, że na tych figurach są dosyć małe. Chcieliśmy zaoszczędzić trochę na miejscu) ponieważ widoczna jest zależność od czasu. Warto jednak przyrzeć się szeregom:

- Zt-differentiation dla całej dekady
- Zt-spencer dla całej dekady
- Zt-differentiation dla roku 1965

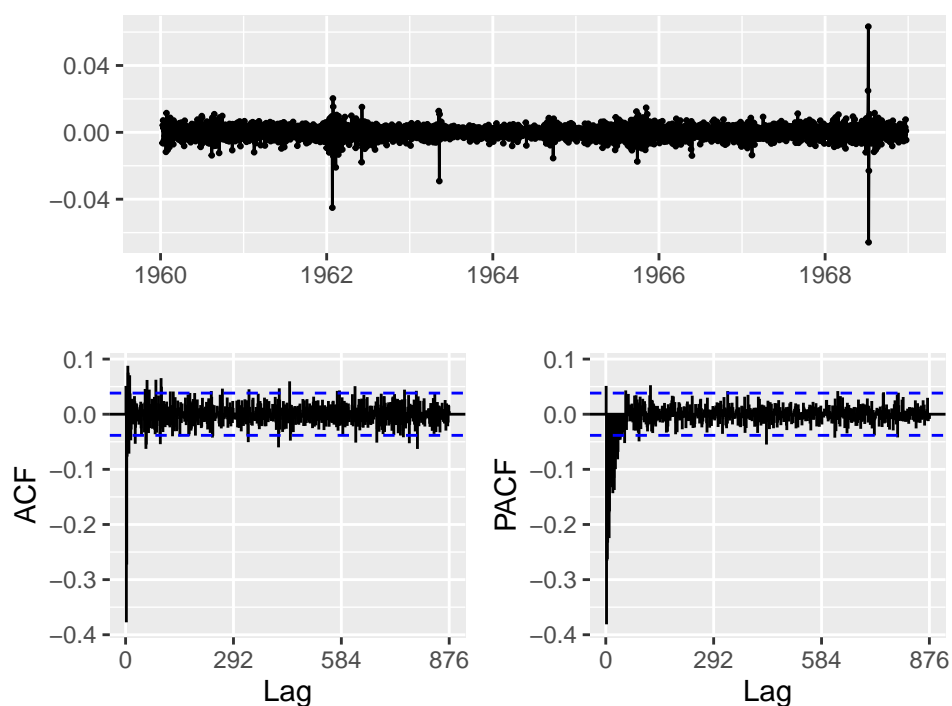
Szereg różnicowy dla całej dekady w szczególności warto rozważyć, ze względu na to, że jest jedynym, który miał (jakikolwiek) test odrzucający hipotezę zerową o białoszumowości. Spójrzmy zatem na wykresy tych szeregów a także ich ACF oraz PACF:



Rysunek 41: Szereg różnicowany dla całej dekady

Istotnym faktem jest, że zarówno ACF jak i PACF przyjmują wysoką wartość dla pierwszego Lagu. Nie widzimy jednak stopniowego spadku wartości do zera wraz ze zwiększaniem Lagów, zatem powinniśmy powiedzieć, że szereg nie jest stacjonarny. Nie możemy zatem dopasowywać modelu stacjonarnego.

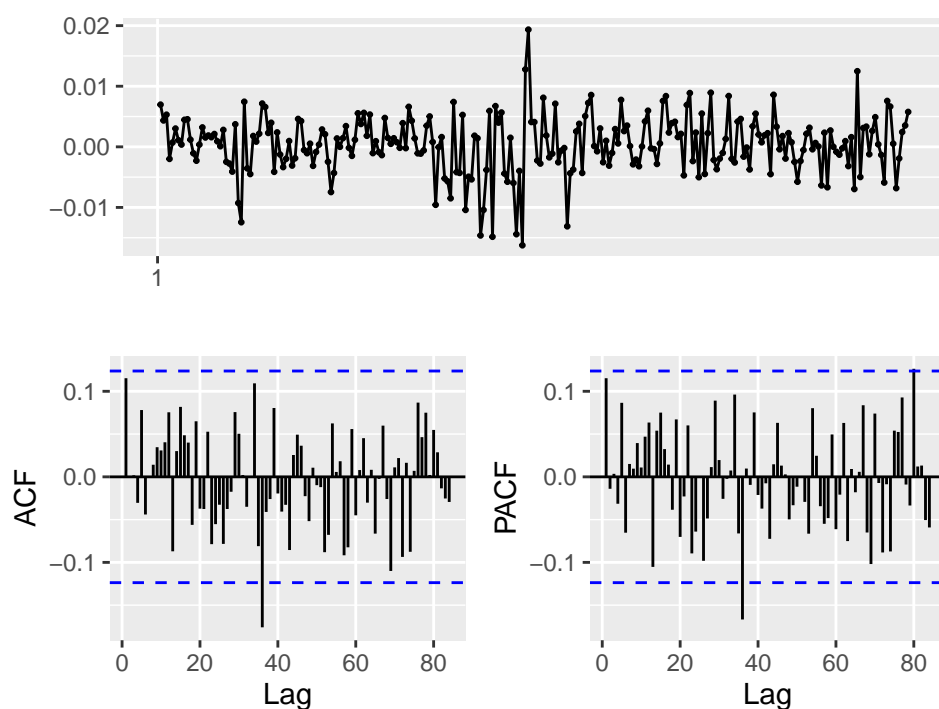
Przejdźmy do szeregu  $\hat{Z}_t$  dla całej dekady uzyskanego poprzez odjęcie trendu estymowanego Filtrem Spencera:



Rysunek 42: Zt dla Filtru Spencera dla całej dekady

Dla tego szeregu przede wszystkim rzuca się w oczy, że PACF osiąga bardzo niskie wartości dla pierwszych kilkudziesięciu lagów, natomiast ACF na początku jest (mniej więcej) naprzemiennie, przyjmując na początku spore wartości. Gdybyśmy mieli widoczne malenie do zera tych wartości moglibyśmy zasugerować model Autoregresyjny (jednak nie ma takiego spadku). Również uważamy, że jest to szereg niestacjonarny.

Na koniec przyjrzyjmy się szeregowi różnicowanemu dla roku 1965:



Rysunek 43: Szereg różnicowany dla roku 1965

Ten przykład jest nieco bardziej interesujący. Oczywiście mamy jedynie 25 lagów, jednak patrząc na ACF pierwsza wartość jest istotna, a następnie wartości (w większości) znajdują się w przedziale ufności, na dodatek mają tendencję naprzemienną. Skupiając się na PACF, pierwsze 4 wartości są dosyć duże (Co do wartości bezwzględnej), oraz w szczególności widzimy, że stopniowo maleją. Ponieważ ACF ma dużą wartość tylko dla pierwszego lagu, oraz PACF maleje możemy rozważyć model MA(1).