

# Szeregi czasowe - Projekt

## Analiza konsumpcji i wydatków na komunikację w UK

Patryk Gronkiewicz 164157

### Spis treści

<b>1</b>	<b>Użyte dane</b>	<b>1</b>
<b>2</b>	<b>Analiza</b>	<b>2</b>
2.1	Główne cechy danych . . . . .	2
2.2	Dekompozycja . . . . .	7
2.3	Eliminacja trendu i sezonowości . . . . .	9
2.4	Prognozowanie naiwne . . . . .	11
2.5	Wyznaczenie współczynników dla modelu AR . . . . .	11
2.6	Wyznaczenie współczynników dla modelu MA . . . . .	11
2.7	Wyznaczenie optymalnych modeli . . . . .	11
<b>3</b>	<b>Wnioski</b>	<b>11</b>

## 1 Użyte dane

W projekcie użyto danych na temat:

1. Całkowitego kosztu konsumpcji w gospodarstwach domowych UK jako szereg z sezonowością (zawiera także trend).
2. Wydatków na komunikację w UK jako szereg z trendem.

Dane te pochodzą z ONS (odpowiednik GUS-u). W analizie zostanie pominięty okres od 2019Q4 jako anomalia ze względu na pandemię.

Już na oficjalnej stronie można zauważyć, że dane te w obu przypadkach zawierają wyraźny trend wzrostowy, natomiast jedynie całkowity koszt konsumpcji ma wyraźną sezonowość z peakiem w czwartym kwartale każdego roku.

Szereg zawierający dane nt. komunikacji odnosi się do wydatków na usługi pocztowe oraz telefon i fax (z uwzględnieniem sprzętu, jak i usług).

W danych dotyczących wydatków Brytyjczyków uwzględnione zostały wydatki w gospodarstwach zarówno rezydentów i nierezydentów (osób posiadających brytyjski paszport lub nie - jest to koncept inny od obywateli państwa)

W obu przypadkach dane opublikowane zostały 31.03.2021 roku z danymi za 2020Q4, więc można zauważyć, że dostępne są z kwartalnym opóźnieniem.

Analiza tych szeregów pozwala na lepsze planowanie wydatków, nawet na poziomie pojedynczego gospodarstwa ze względu na możliwość uwzględnienia wzrostu cen czy inflacji stylu życia. Analiza wydatków na komunikację jest także dobrym wskaźnikiem do przekazania jak bardzo “zdalne” społeczeństwo jest. W społeczeństwie, w którym małe grupy ludzi dzielą znaczne odległości wydatki na takie usługi będą wyższe ze względu na częstość wykorzystania takich możliwości.

Do ich obróbki zostały użyte biblioteki zaimportowane poniżej

```
library(forecast)
```

## 2 Analiza

### 2.1 Główne cechy danych

Na początku dane zostały załadowane z plików CSV. W nie interesują nas niektóre z linii widocznych w pliku (linie 1-44 ze względu na metadane i dane roczne, a nie kwartalne).

```
wydatki <- ts(read.csv('wydatki.csv',
                      skip=43,
                      col.names = c("q", "v"))$v,
             start = c(1985, 01),
             frequency = 4)
komunikacja <- ts(read.csv('komunikacja.csv',
                          skip=43,
                          col.names = c("q", "v"))$v,
                 start = c(1985, 01),
                 frequency = 4)
wydatki <- window(wydatki, start = start(wydatki), end = c(2019, 04))
komunikacja <- window(komunikacja, start = start(komunikacja), end = c(2019, 04))
```

Na początku zostały przedstawione dane na kilku wykresach.

```
par(mfrow=c(2,1), mar=c(2,4,2,2))
plot(wydatki)
plot(komunikacja)
```

Na wykresie 1 bardzo wyraźnie widać sezonowość w postaci “ząbków” dla wydatków ogólnych, czego na pierwszy rzut oka nie można stwierdzić o wydatkach na komunikację. Oba szeregi zawierają wyraźny trend.

```
par(mfrow = c(1,2), mar=c(5,3,4,1))
monthplot(wydatki, ylab = NA, xlab = "wydatki")
monthplot(komunikacja, ylab = NA, xlab = "komunikacja")
```

Na wykresie 2 widać wyraźnie trendy wzrostowe między odpowiadającymi kwartałami, więc zależność została zachowana w przypadku obu szeregów. Jak łatwo zauważyć dla wydatków nie występuje “ząbkowanie” na poszczególnych wykresach, dlatego można wnioskować, że ich sezonowość to pewna wielokrotność 4. Może to wynikać z wyższych kosztów w kwartale 4 ze względu na ogrzewanie i droższą żywność ze względu na zwiększony import w miesiącach jesienno-zimowych. Załamanie w wydatkach wynika z kryzysu w 2007-2009 roku spowodowanym załamaniem rynku kredytów hipotecznych wysokiego ryzyka.

```
par(mfrow = c(1,2))
boxplot(wydatki, xlab="wydatki")
boxplot(komunikacja, xlab="komunikacja")
```

Wykresy na rysunku 3 nie zostały pokazane na jednej osi ze względu na bardzo rozbieżne wartości między szeregami, przez co dane nt. komunikacji nie były czytelne. Jak można zauważyć dużo dłuższe linie błędów są w górę w przypadku wydatków i w dół dla komunikacji.

```
lag.plot(wydatki, lags = 4, labels = F)
```

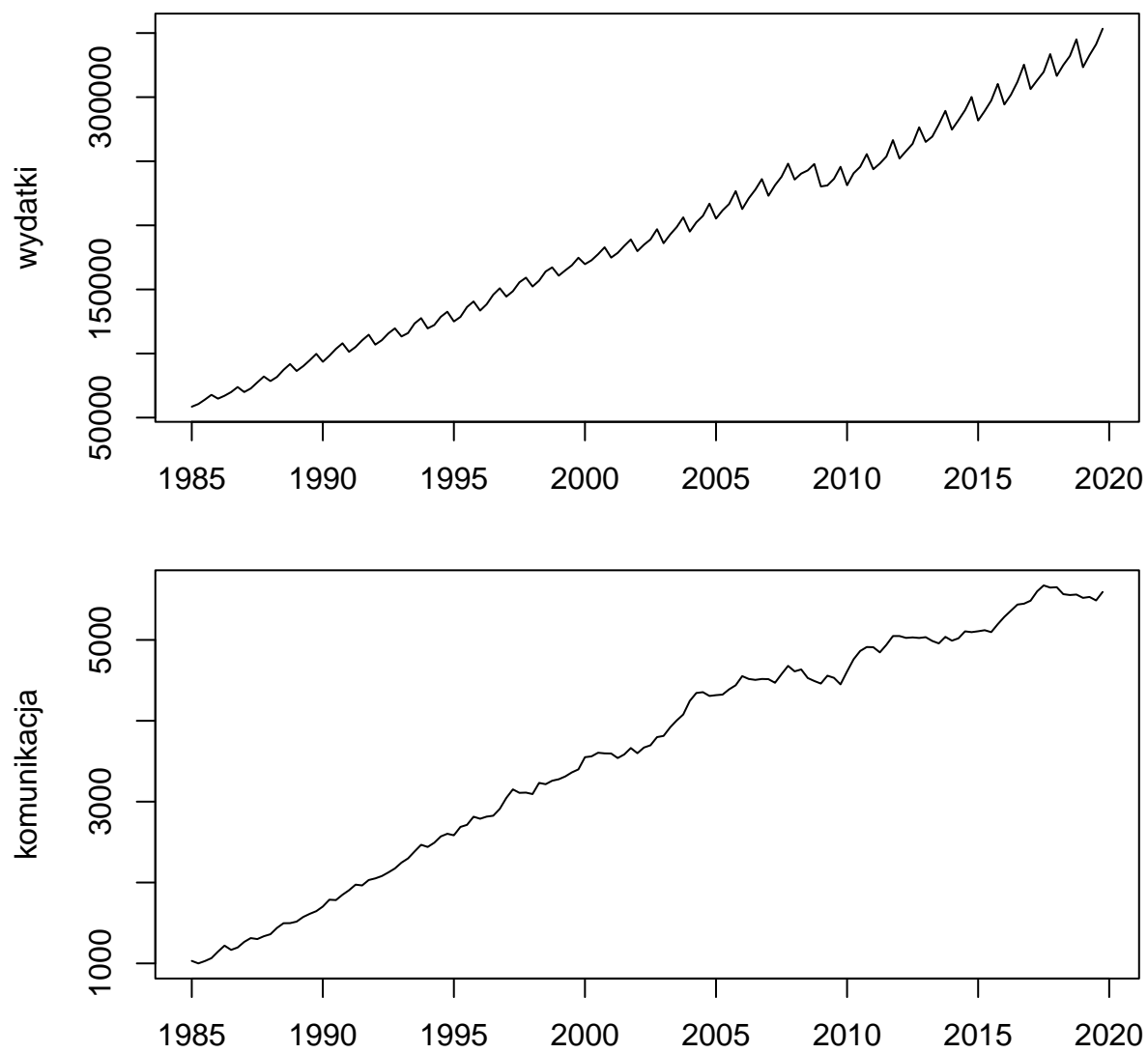
```
lag.plot(komunikacja, lags = 4, labels = F)
```

Jak widać na wykresach z rysunku 4 najbardziej skoncentrowane wartości są dla lag=4, czego można się było spodziewać przy analizie rysunku 2.

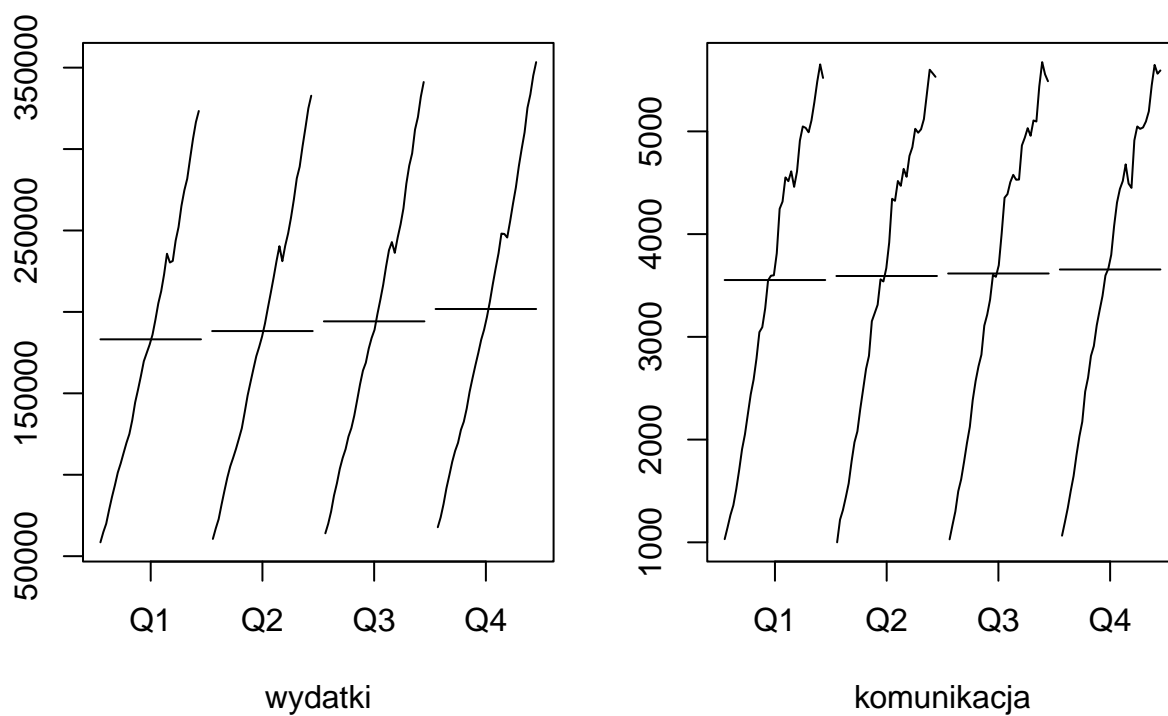
```
tsdisplay(wydatki)
```

```
tsdisplay(komunikacja)
```

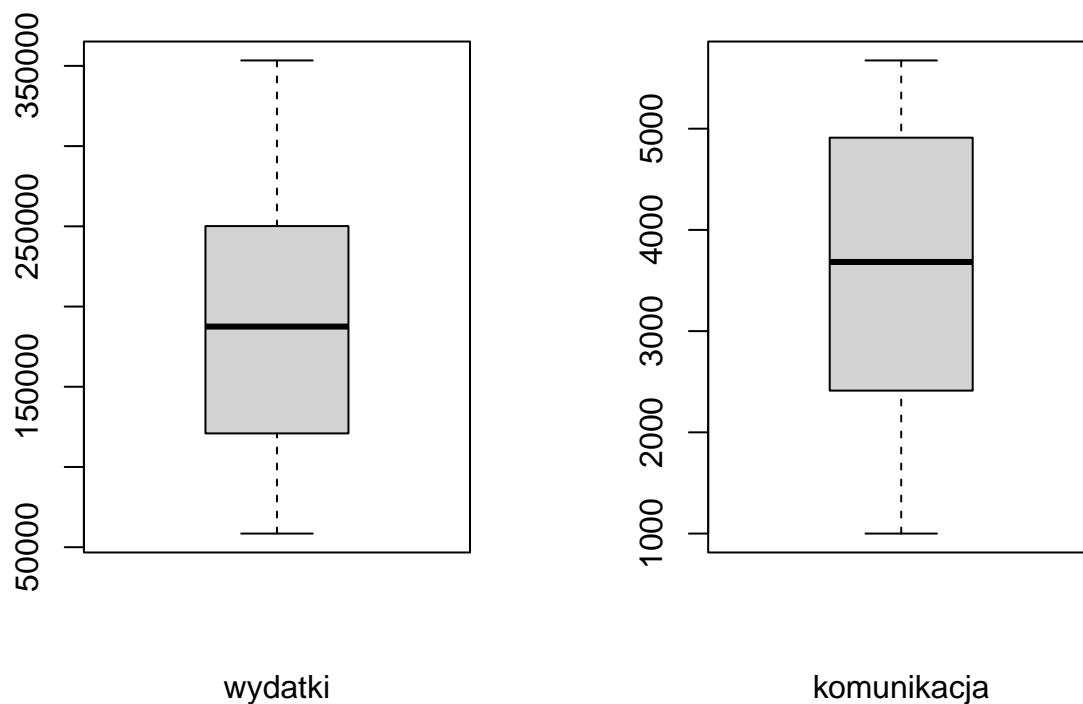
Na wykresach generowanych przez funkcję `tsdisplay` (rysunki 6 i 7) bardzo dobrze widać, że dla ogólnych wydatków największa korelacja jest z rocznym opóźnieniem. Korelacja z opóźnieniem dwuletnim jest na



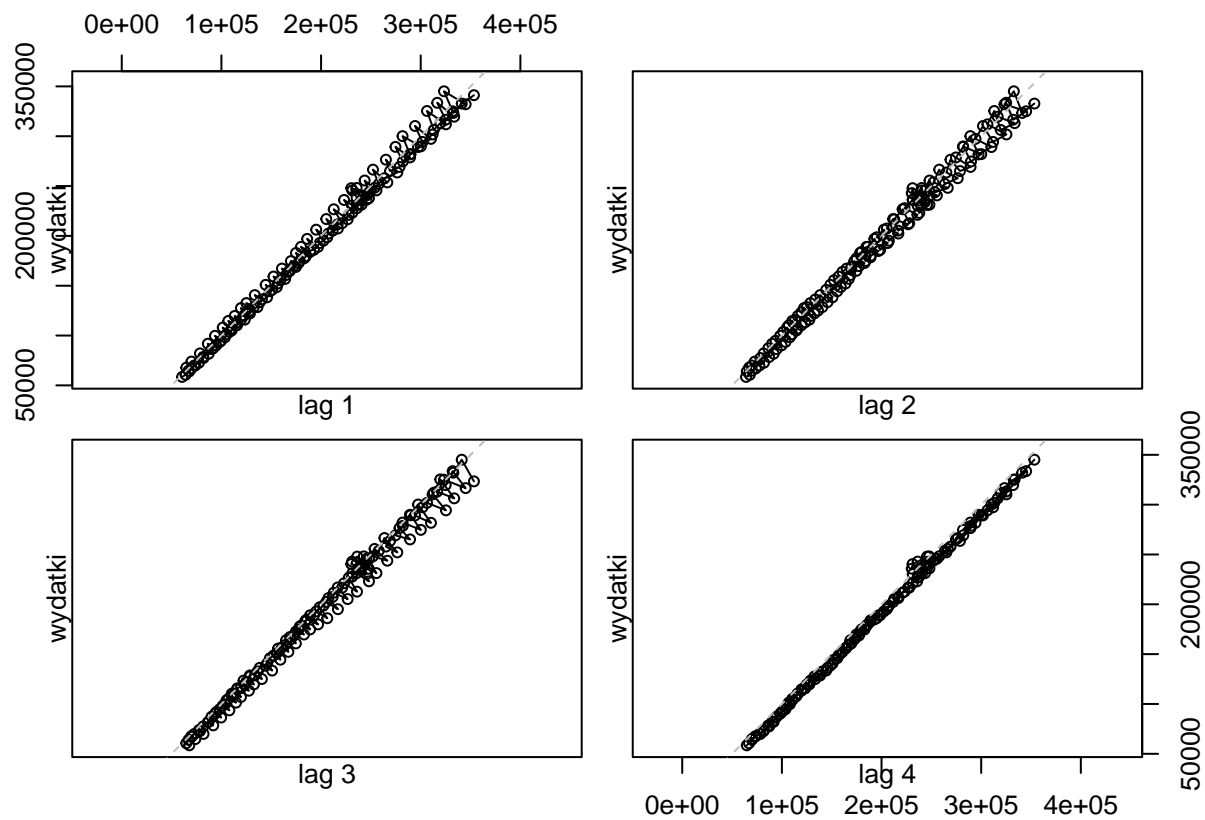
Rysunek 1: Wydatki w czasie przed jakąkolwiek obróbką



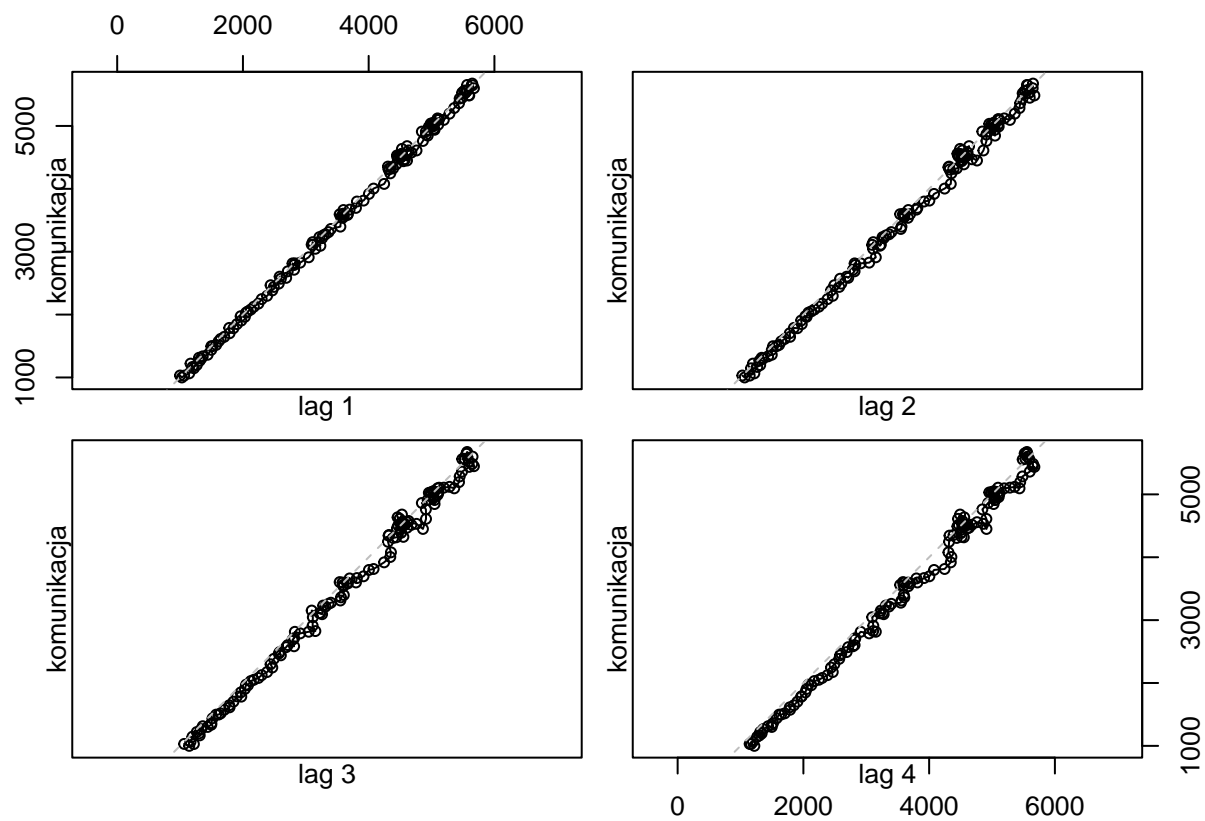
Rysunek 2: Wykresy `monthplot` jak widać mogą dotyczyć nie tylko miesięcy, ale innych okresów w roku, takich jak kwartały. Dla czytelności wykresy zostały podpisane na osi poziomej, jednak kwota wydatków jest na osi pionowej



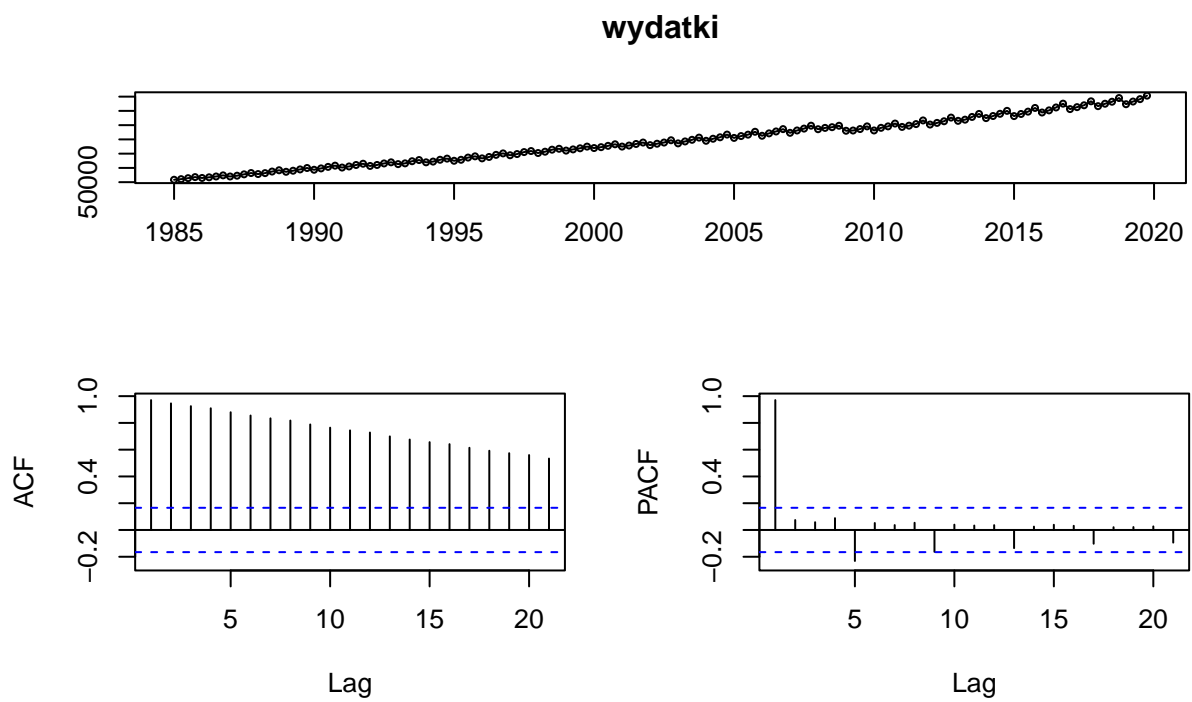
Rysunek 3: Wydatki w gospodarstwach domowych ogółem oraz na komunikację.



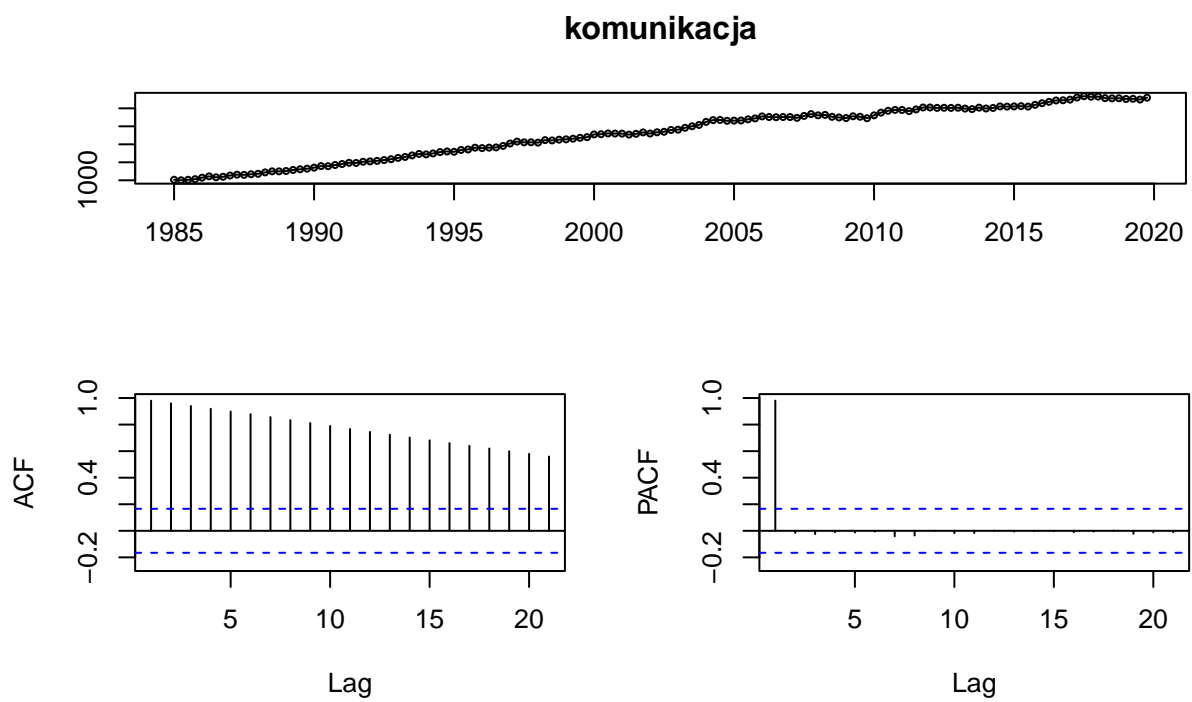
Rysunek 4: Lag plot wydatków



Rysunek 5: Lag plot komunikacji



Rysunek 6: Wykresy z autokorelacją dla wydatków



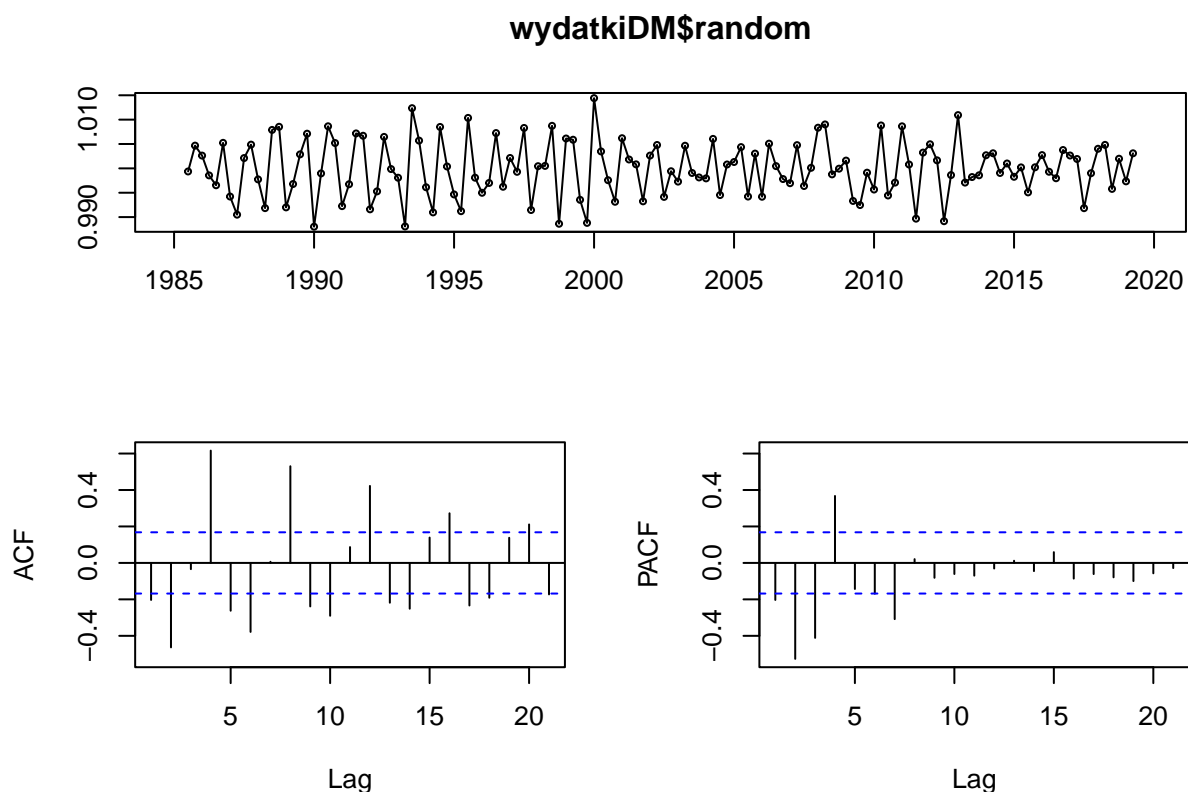
Rysunek 7: Wykresy z autokorelacją dla komunikacji

granicy istotności, więc w naszych analizach ją pominiemy. Zgodnie z przewidywaniami dla wydatków na komunikację nie istnieje żadna istotna sezonowość.

## 2.2 Dekompozycja

Wykorzystana została dekompozycja na podstawie modelu regresji liniowej - na pierwszy rzut oka wygląda na adekwatny dla tych szeregów.

```
wydatkiDM <- decompose(wydatki, type = "multiplicative")
tsdisplay(wydatkiDM$random)
```



Rysunek 8: Reszty dla szeregu po dekompozycji multiplikatywnej trendu z wydatków

Jak widać na rysunku 8 czysta dekompozycja nie była w stanie sobie poradzić z tym szeregiem. Nadal bardzo widoczna jest autokorelacja dla  $\text{lag} = 4n, n \in \mathbb{N}$

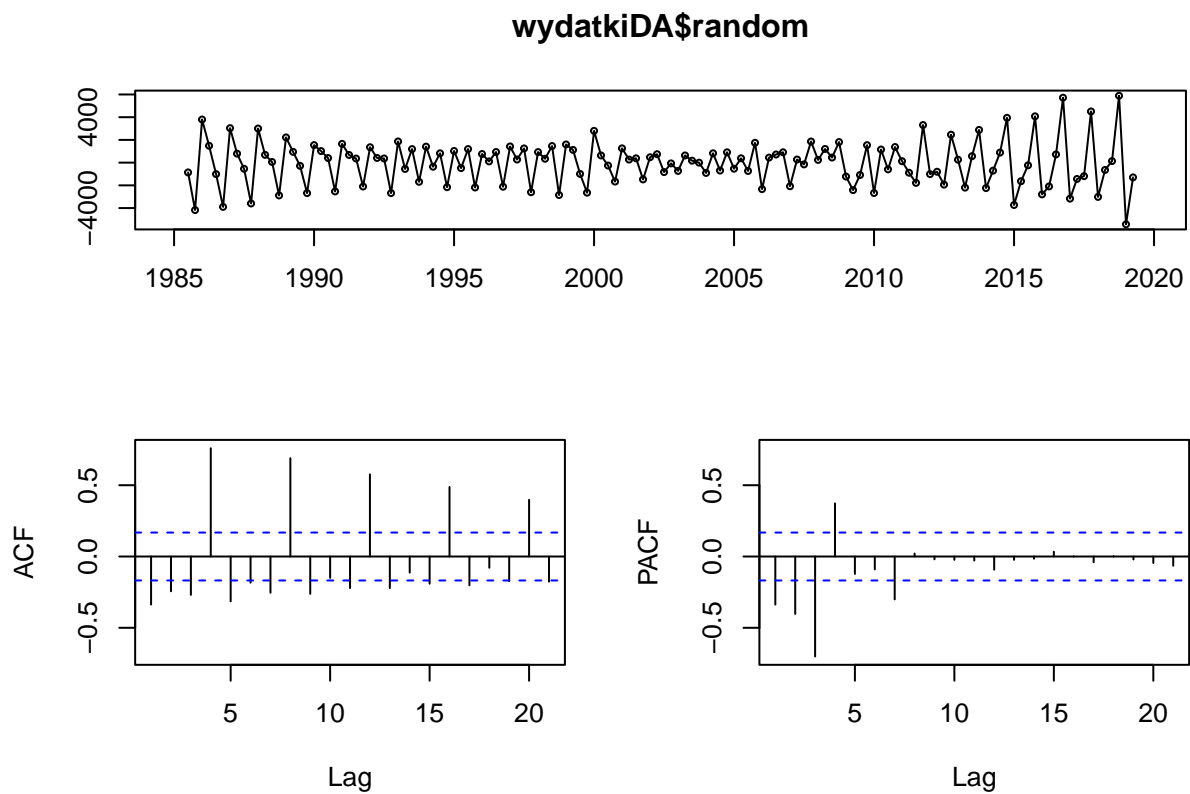
```
wydatkiDA <- decompose(wydatki, type = "additive")
tsdisplay(wydatkiDA$random)
```

W tym wypadku model multiplikatywny z rysunku 8 wydaje się lepszym rozwiązaniem ze względu na charakter danych - zmiany takie jak inflacja nakładają się mnożąc zmiany (np. coś przy inflacji na poziomie 5% po dwóch latach będzie kosztować  $1,05^2x$ , a nie  $(1,05 + 1,05)x$ ).

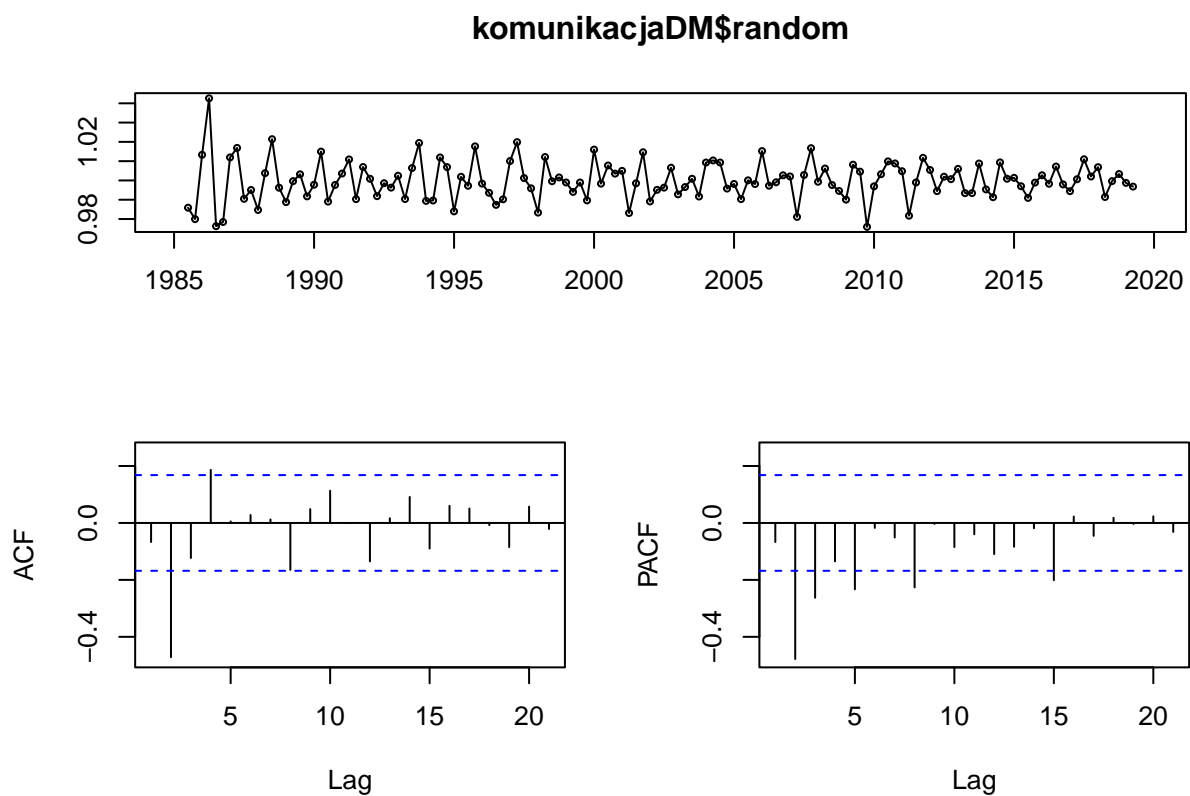
Analogicznie postąpiono dla danych nt. wydatków na komunikację.

```
komunikacjaDM <- decompose(komunikacja, type = "multiplicative")
tsdisplay(komunikacjaDM$random)
```

```
komunikacjaDA <- decompose(komunikacja, type = "additive")
tsdisplay(komunikacjaDA$random)
```



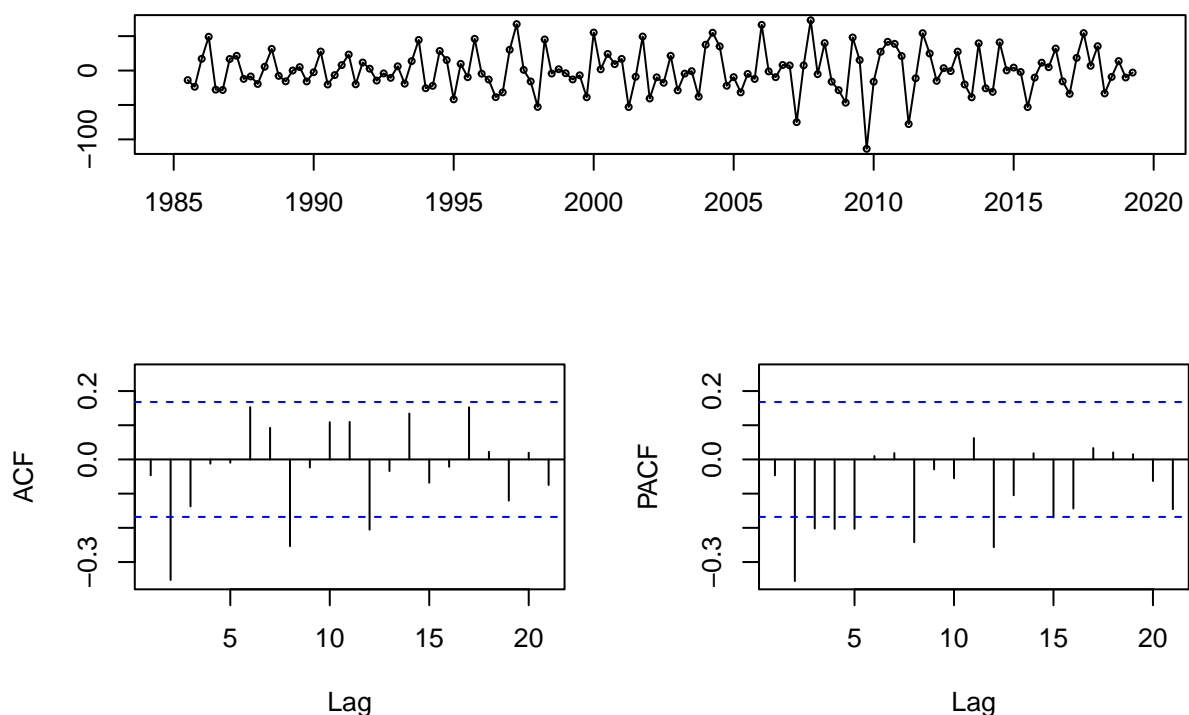
Rysunek 9: Reszty dla szeregu po dekompozycji addytywnej trendu z wydatków



Rysunek 10: Reszty dla szeregu po dekompozycji multiplikatywnej trendu z komunikacji



## komunikacjaDA\$random



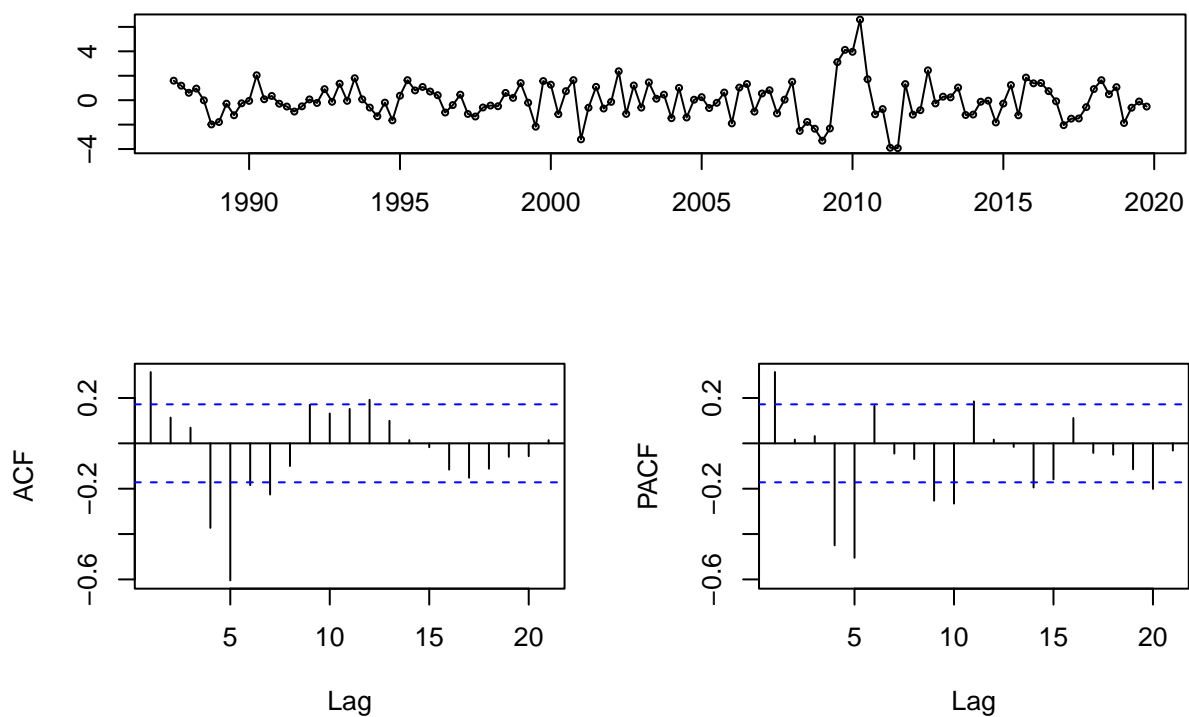
Jak widać na rysunkach ?? 10 i ?? dekompozycja dla komunikacji dała od razu dużo lepsze efekty niż w przypadku wydatków ogólnych. Wynika to z faktu, że ten szereg nie zawiera żadnych istotnych autokorelacji. Nie zmieniło się natomiast nic w kwestii doboru metody - lepszym rozwiązaniem w tym wypadku jest dekompozycja multiplikatywna.

### 2.3 Eliminacja trendu i sezonowości

Następnie usunięto trend przez transformację Boxa-Coxa z automatyczną lambdą oraz różnicowanie z opóźnieniem równym 1 i 8 dla komunikacji i równym kolejno 4, 1 i 5 dla wydatków.

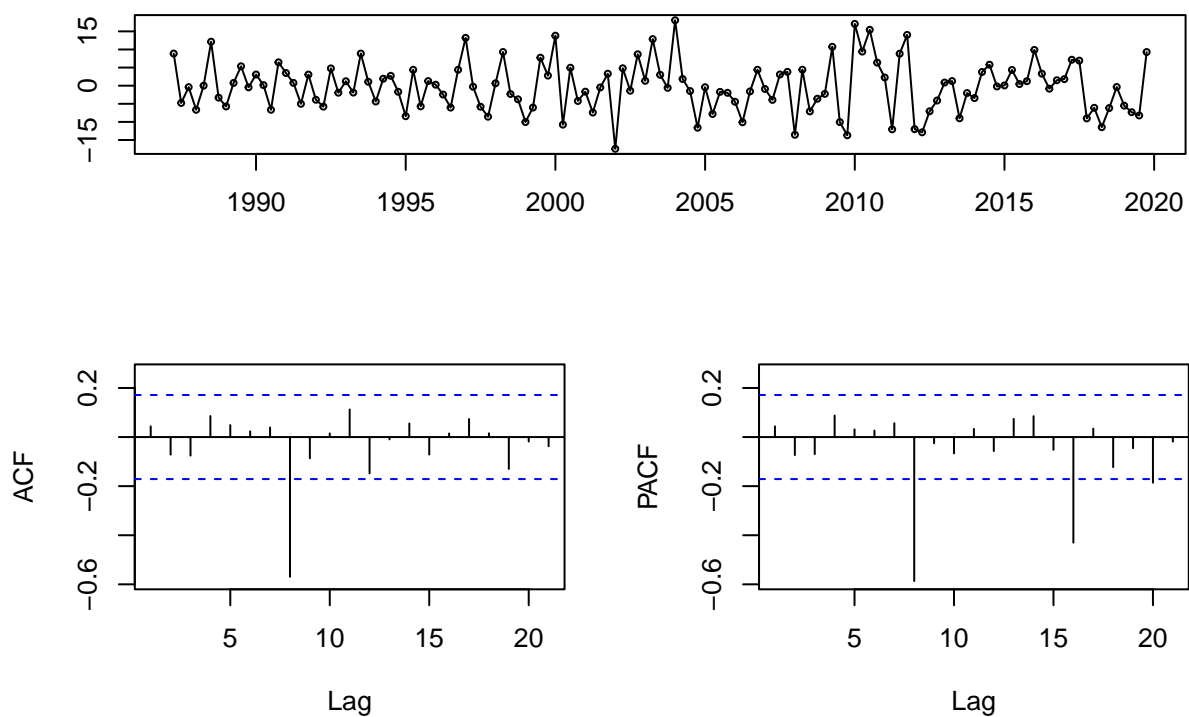
```
wydatkiL <- BoxCox(wydatki, BoxCox.lambda(wydatki))
wydatkiL.4.1.5 <- diff(diff(diff(wydatkiL, lag=4)), lag=5)
tsdisplay(wydatkiL.4.1.5)
```

### wydatkiL.4.1.5



```
komunikacjaL <- BoxCox(komunikacja, BoxCox.lambda(komunikacja))
komunikacjaL.1.8 <- diff(diff(komunikacjaL), lag=8)
tsdisplay(komunikacjaL.1.8)
```

### komunikacjaL.1.8



- 2.4 Prognozowanie naiwne
  - 2.4.1 Porównanie modeli prognozowania
- 2.5 Wyznaczenie współczynników dla modelu AR
- 2.6 Wyznaczenie współczynników dla modelu MA
- 2.7 Wyznaczenie optymalnych modeli
- 3 Wnioski