

Sprawozdanie 2

Mateusz Cieślak, Michał Merta

2026-02-23

Spis treści

1	Zadanie 3 - Dopasowanie modeli autoregresji	2
1.1	Krótki opis zagadnienia	2
1.2	a) Sprawdzenie stacjonarności i transformacje danych	2
1.2.1	Odpowiedni wariant różnicowania	2
1.2.2	Dodatkowa weryfikacja: Próba dwukrotnego różnicowania ($d=2$)	3
1.2.3	Estymacja i eliminacja trendu wielomianowego	5
1.2.4	Sprawdzenie trendu liniowego (Weryfikacja)	6
1.3	b) Wybór optymalnych rzędów modeli autoregresji	6
1.3.1	Analiza funkcji PACF	7
1.3.2	Analiza kryteriów AIC i FPE	8
1.3.3	Automatyczna weryfikacja rzędu p (funkcja ar)	9
1.4	c) Wyznaczenie estymatorów parametrów	9
1.5	d) Asymptotyczne przedziały ufności i weryfikacja istotności	10
1.6	e) Weryfikacja (na podstawie analizy reszt) poprawności dopasowania	11
1.6.1	Analiza wizualna reszt	12
1.6.2	Testy statystyczne (Białoszumowość)	13
1.7	f) Szczegółowa prognoza i porównanie podejść	14
1.7.1	Przygotowanie danych i obliczenia	14
1.7.2	Analiza wizualna prognoz	16
1.8	g) Końcowe wnioski i podsumowanie	20
2	Zadanie 4 - Porównanie dokładności prognoz dla danych euretail	21
2.1	Wstęp	21
2.2	Podział danych na część uczącą i testową	21
2.3	Dopasowanie odpowiednich modeli do danych	21
2.3.1	SARIMA	23
2.3.2	ETS	26
2.3.3	TSLM	29
2.4	Wyznaczenie prognoz dla zbioru testowego na podstawie dopasowanych modeli oraz wybranej metody referencyjnej	30
2.5	Przedstawienie skonstruowanych prognoz na wykresie i porównanie z wartościami rzeczywistymi	32
2.6	Porównanie dokładności prognoz dla zbioru testowego i uczącego	32
2.7	Wnioski	33

Spis wykresów

1	Wykres szeregu czasowego <code>gtemp_both</code>	2
2	Szereg <code>gtemp_both</code> po jednokrotnym zróżnicowaniu z zaznaczonym dryfem (średnią).	3
3	Porównanie ACF dla $d=1$ (lewo) i $d=2$ (prawo).	4
4	Dopasowanie trendu kwadratowego (góra) i reszty z modelu regresji (dół).	5

5	Dopasowanie trendu kwadratowego (górn) i reszty z modelu regresji (dół).	6
6	Funkcja PACF dla danych zróżnicowanych (lewa) i po usunięciu trendu wielomianowego (prawa).	7
7	Przebieg kryteriów AIC i FPE.	8
8	Diagnostyka reszt: ACF (lewa kolumna) i Q-Q Plot (prawa kolumna). Górny wiersz: AR(5) Różnicowanie, Dolny wiersz: AR(1) trend wielomianowy.	12
9	Prognoza ręczna z modelu AR(5) (punktowa).	16
10	Prognoza ARIMA(5,1,0) z rosnącym przedziałem ufności.	17
11	Prognoza deterministyczna: Trend wielomianowy + AR(1).	18
12	Porównanie podejścia stochastycznego i deterministycznego.	19
13	Wykres szeregu czasowego eureka	21
14	Wykres szeregu czasowego dla zbioru uczącego eureka	22
15	Wykres szeregu czasowego dla zbioru uczącego eureka po transformacji Boxa-Coxa	22
16	Analiza autokorelacji szeregu po różnicowaniu zwykłym (d=1) i sezonowym (D=1)	23
17	Analiza reszt i opóźnień w najlepszym modelu SARIMA	26
18	Analiza reszt i opóźnień w najlepszym modelu ETS	27
19	Analiza reszt i opóźnień w modelu Holta-Wintersa	28
20	Analiza reszt i opóźnień w modelu TSLM	29
21	Wykres prognozy naiwnej z sezonowością SNAIVE	30
22	Wykres prognozy modelu SARIMA(0,1,0)(0,1,1)[4]	30
23	Wykres prognozy modelu ETS(A,A,A)	31
24	Wykres prognozy modelu H-W	31
25	Wykres prognozy modelu TSLM	31
26	Porównanie prognoz do wartości rzeczywistych	32

Spis tabel

1	Wyniki testów stacjonarności dla surowego szeregu gtemp_both.	2
2	Wyniki testów stacjonarności dla szeregu zróżnicowanego.	3
3	Wyniki testów stacjonarności dla d=2.	4
4	Wyniki testów stacjonarności dla szeregu po eliminacji trendu wielomianowego.	6
5	Testy stacjonarności dla reszt z trendu liniowego.	6
6	Porównanie estymatorów YW i MLE dla modelu AR(5) (dane zróżnicowane).	10
7	Porównanie estymatorów YW i MLE dla modelu AR(1) (dane po usunięciu trendu wielomianowego).	10
8	Asymptotyczne przedziały ufności dla modelu AR(5) (Dane zróżnicowane).	11
9	Asymptotyczne przedziały ufności dla modelu AR(1) (Dane po usunięciu trendu wielomianowego).	11
10	Wyniki testów diagnostycznych reszt (Ljung-Box i Shapiro-Wilk).	13
11	Porównanie kryterium informacyjnego AICc i BIC dla wybranych modeli SARIMA	24
12	Wpływ uwzględnienia parametru dryfu na jakość modelu SARIMA	25
13	Porównanie kryteriów informacyjnych dla automatycznego i ręcznego modelu ETS	27
14	Błędy dopasowania ETS (zbiór uczący)	27
15	Błędy dopasowania klasycznych metod wygładzania (zbiór uczący)	28
16	Błędy dopasowania na zbiorze uczącym	32
17	Dokładność prognoz na zbiorze testowym (h=12)	32

Wykorzystane biblioteki:

```
library(ggplot2)
library(forecast)
library(knitr)
library(dplyr)
library(fpp2)
library(kableExtra)
library(asts)
```

```
library(stats)
library(tseries)
library(lmtest)
library(patchwork)
library(xtable)
```

1 Zadanie 3 - Dopasowanie modeli autoregresji

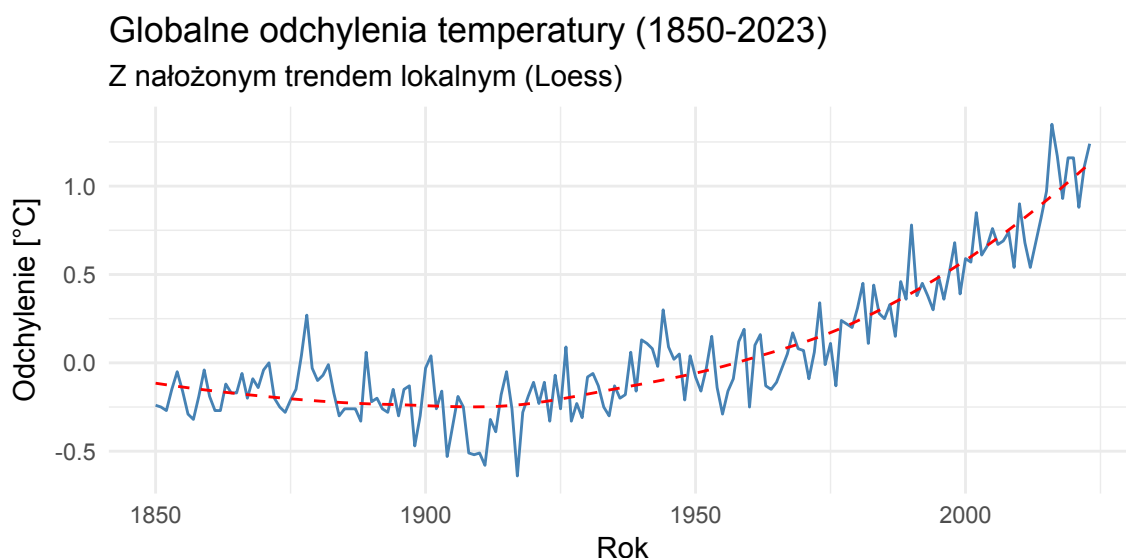
1.1 Krótki opis zagadnienia

W niniejszym zadaniu analizie poddajemy szereg czasowy `gtemp_both`, przedstawiający odchylenia średniej rocznej temperatury globalnej (lądy i oceany) w latach 1850-2023. Celem badania jest przeprowadzenie kompletnej ścieżki modelowania z wykorzystaniem procesów autoregresyjnych $AR(p)$.

Analiza obejmuje weryfikację stacjonarności, zastosowanie odpowiednich transformacji (różnicowanie vs eliminacja trendu wielomianowego), identyfikację rzędu modelu, estymację parametrów oraz diagnostykę reszt i prognozowanie.

1.2 a) Sprawdzenie stacjonarności i transformacje danych

W pierwszej kolejności dokonujemy wizualizacji danych źródłowych w celu wstępnej oceny charakteru szeregu.



Wykres 1: Wykres szeregu czasowego `gtemp_both`

Na Wykresie 1 obserwujemy wyraźny, nieliniowy trend wzrostowy, sugerujący niestacjonarność średniej. Aby zweryfikować tę hipotezę formalnie, przeprowadzamy formalne testy stacjonarności: **Augmented Dickey-Fuller (ADF)** oraz **KPSS**.

Test	Statystyka	p.value	Wniosek
ADF	-0.79	0.96	Nie odrzucamy H_0 (Niestacjonarny)
KPSS	0.72	0.01	Odrzucamy H_0 (Niestacjonarny)

Tabela 1: Wyniki testów stacjonarności dla surowego szeregu `gtemp_both`.

Wyniki w Tabeli 1 jednoznacznie wskazują na niestacjonarność badanego procesu ($p\text{-value} > 0.05$ dla ADF oraz $p\text{-value} < 0.05$ dla KPSS). Konieczne jest zatem zastosowanie transformacji.

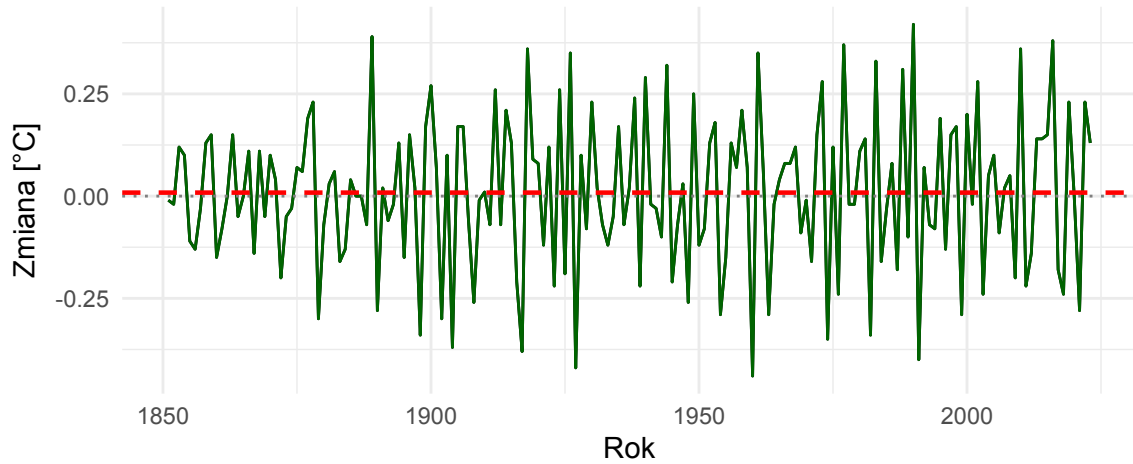
1.2.1 Odpowiedni wariant różnicowania

Pierwszym podejściem jest zastosowanie różnicowania rzędu pierwszego ($d = 1$), co odpowiada eliminacji trendu stochastycznego.

$$Y_t = \nabla X_t = X_t - X_{t-1}$$

Przyrosty roczne temperatury (Zróżnicowany szereg)

Czerwona linia oznacza średni dryf = 0.0086



Wykres 2: Szereg gtemp_both po jednokrotnym zróżnicowaniu z zaznaczonym dryfem (średnią).

Wizualnie szereg przedstawiony na Rysunku 2 wykazuje cechy stacjonarności (oscylacje wokół stałego poziomu). Potwierdzamy to testami statystycznymi.

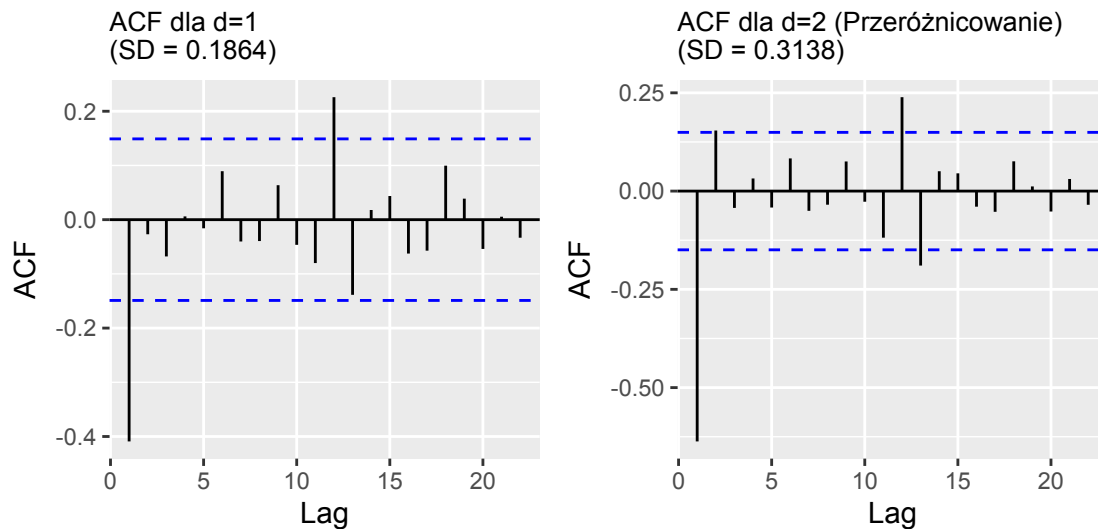
Test	Statystyka	p.value	Wniosek
ADF	-8.16	0.01	Odrzucamy H_0 (Stacjonarny)
KPSS	0.17	0.10	Nie odrzucamy H_0 (Stacjonarny)

Tabela 2: Wyniki testów stacjonarności dla szeregu zróżnicowanego.

Zgodnie z Tabelą 2, jednokrotne różnicowanie skutecznie sprowadziło szereg do stacjonarności.

1.2.2 Dodatkowa weryfikacja: Próba dwukrotnego różnicowania ($d=2$)

Aby upewnić się, że rząd różnicowania $d = 1$ jest optymalny, przeprowadzamy eksperymentalnie drugie różnicowanie ($\nabla^2 X_t$). Zgodnie z teorią, jeśli szereg został “przeróżnicowany”, powinniśmy zaobserwować wzrost odchylenia standardowego oraz silną, ujemną autokorelację w pierwszym opóźnieniu (ACF).



Wykres 3: Porównanie ACF dla $d=1$ (lewo) i $d=2$ (prawo).

Wnioski z analizy $d=2$:

1. **Wzrost zmienności:** Odchylenie standardowe dla $d = 2$ jest wyższe niż dla $d = 1$ (wzrost z ok. 0.186 do 0.314). Oznacza to, że drugie różnicowanie wprowadziło dodatkowy szum zamiast uprościć strukturę danych.
2. **Sztuczna autokorelacja:** Na prawym wykresie (dla $d = 2$) widoczna jest silniejsza ujemna korelacja dla opóźnienia 1 ($\rho(1) \approx -0.5$). Jest to klasyczny objaw przeróżnicowania, wynikający z wprowadzania sztucznej zależności między kolejnymi błędami.

Powyższe obserwacje ostatecznie potwierdzają, że optymalnym rzędem różnicowania jest $d = 1$.

Weryfikacja stacjonarności po dwukrotnym różnicowaniu

Dla pełnego obrazu przeprowadzono również testy stacjonarności dla szeregu podwójnie zróżnicowanego ($d = 2$).

Test	Statystyka	p.value	Wniosek
ADF	-11.65	0.01	Odrzucamy H_0 (Stacjonarny)
KPSS	0.02	0.10	Nie odrzucamy H_0 (Stacjonarny)

Tabela 3: Wyniki testów stacjonarności dla $d=2$.

Komentarz do wyników: Jak widać w Tabeli 3, testy potwierdzają stacjonarność również dla $d = 2$. Jest to wynik oczekiwany (różnicowanie procesu słabo stacjonarnego zachowuje słabą stacjonarność). Jednakże, sam fakt bycia stacjonarnym nie oznacza, że transformacja jest optymalna. Decydującym kryterium odrzucenia $d = 2$ na rzecz $d = 1$ jest wspomniany wcześniej wzrost wariancji oraz zwiększona ujemna autokorelacja (widoczna na wykresach ACF), świadczące o przeróżnicowaniu.

Jako ostateczne potwierdzenie wyboru rzędu różnicowania d , wykorzystamy funkcję `ndiffs()` z pakietu `forecast`. Funkcja ta algorytmicznie dobiera minimalną liczbę różnicowań wymaganą do osiągnięcia stacjonarności, opierając się na wynikach testów statystycznych (KPSS oraz ADF).

Wniosek i wyniki testów:

Algorytmiczna weryfikacja potwierdza nasze wcześniejsze ustalenia. Test KPSS wskazał jako optymalny rząd różnicowania $d = 1$, natomiast test ADF również zasugerował $d = 1$.

Jest to obiektywne potwierdzenie, że jednokrotne różnicowanie jest optymalne dla analizowanego szeregu. Skuteczność jednokrotnego różnicowania ($d = 1$) można uzasadnić, przyjmując model Błądzenia Losowego z Dryfem jako proces generujący dane. Równanie takiego procesu ma postać:

$$X_t = c + X_{t-1} + Z_t$$

gdzie c jest stałą (dryfem), a Z_t białym szumem.

Po zastosowaniu operatora różnicowania $\nabla X_t = X_t - X_{t-1}$, równanie sprowadza się do postaci:

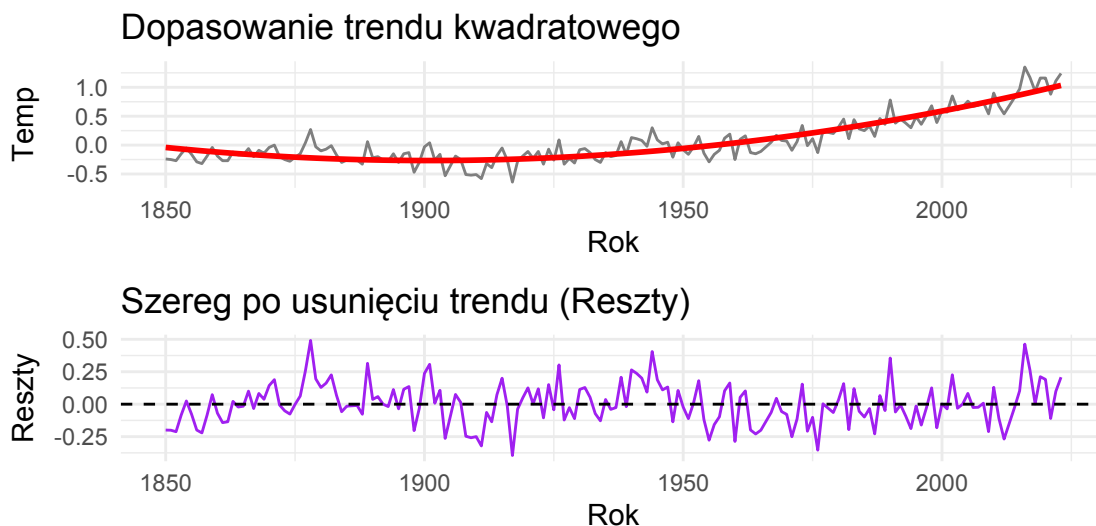
$$\nabla X_t = c + Z_t$$

Otrzymany proces posiada stałą wartość oczekiwaną $E[\nabla X_t] = c$ oraz stałą wariancję $Var(\nabla X_t) = \sigma_Z^2$, co spełnia warunki stacjonarności. Wynik ten sugeruje, że obserwowana nieliniowość ("przyspieszenie") wzrostu temperatury na wykresie pierwotnym może być interpretowana jako efekt kumulacji stochastycznych wstrząsów w procesie z dodatnim dryfem, a niekoniecznie jako deterministyczny trend wielomianowy wyższego rzędu.

1.2.3 Estymacja i eliminacja trendu wielomianowego

Drugim podejściem jest założenie trendu deterministycznego. Ze względu na widoczne przyspieszenie wzrostu temperatury, dopasowano model trendu wielomianowego drugiego stopnia (kwadratowy):

$$X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + Z_t$$



Wykres 4: Dopasowanie trendu kwadratowego (góra) i reszty z modelu regresji (dół).

Reszty modelu (Rysunek 4, panel dolny) wydają się spełniać założenia stacjonarności. Sprawdzimy to formalnymi testami.

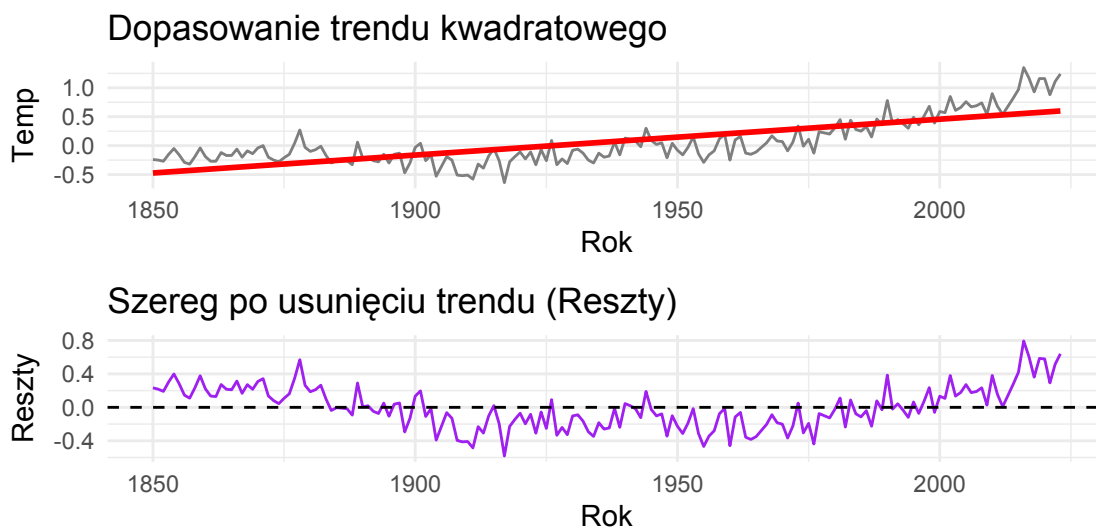
Test	Statystyka	p.value	Wniosek
ADF	-3.68	0.03	Odrzucamy H_0 (Stacjonarny)
KPSS	0.11	0.10	Nie odrzucamy H_0 (Stacjonarny)

Tabela 4: Wyniki testów stacjonarności dla szeregu po eliminacji trendu wielomianowego.

Zgodnie z Tabelą 4, dopasowanie wielomianu kwadratowego skutecznie usunęło trend i doprowadziło szereg do postaci stacjonarnej.

1.2.4 Sprawdzenie trendu liniowego (Weryfikacja)

Dla porównania i weryfikacji charakteru trendu sprawdzimy jeszcze prostszy model, zakładający liniowy wzrost temperatury: $X_t = \beta_0 + \beta_1 t + Z_t$.



Wykres 5: Dopasowanie trendu kwadratowego (góra) i reszty z modelu regresji (dół).

Na Rysunku 5 (panel dolny) widać wyraźnie, że reszty nie oscylują losowo wokół zera, lecz tworzą charakterystyczny kształt litery "U" (są dodatnie na krańcach i ujemne w środku przedziału). Oznacza to, że model liniowy nie wychwytuje przyspieszenia wzrostu temperatury. Potwierdzają to testy stacjonarności reszt w Tabeli 5. Optymalnym modelem będzie zatem uznanie trendu kwadratowego.

Test	Statystyka	p.value	Wniosek
ADF	-0.79	0.96	Nie odrzucamy H_0 (niestacjonarny)
KPSS	0.72	0.01	Odrzucamy H_0 (niestacjonarny)

Tabela 5: Testy stacjonarności dla reszt z trendu liniowego.

1.3 b) Wybór optymalnych rzędów modeli autoregresji

Kolejnym etapem jest identyfikacja rzędu p modelu autoregresyjnego $AR(p)$. W tym celu wykorzystano dwie uzupełniające się metody:

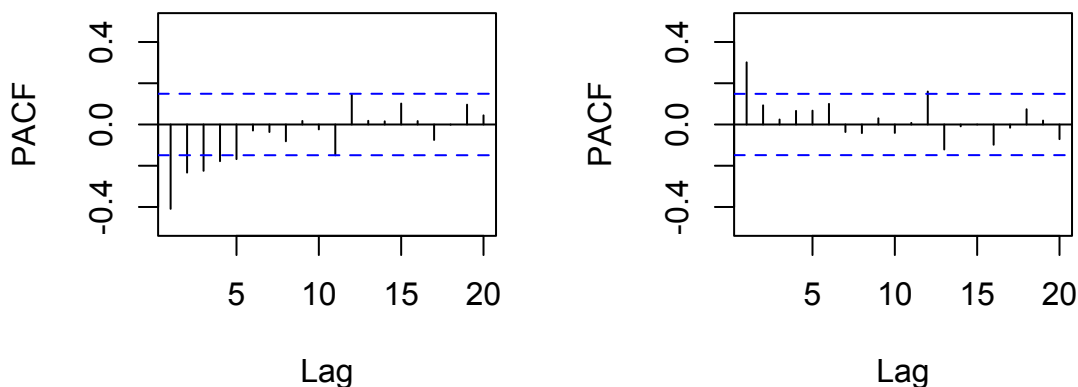
1. **Analiza funkcji cząstkowej autokorelacji (PACF):** Dla procesu $AR(p)$ teoretyczna funkcja PACF powinna "ucinać się" po opóźnieniu p .

2. **Kryteria informacyjne (AIC, FPE):** Szukamy rzędu p , dla którego wartości kryteriów są minimalne.

1.3.1 Analiza funkcji PACF

Poniżej przedstawiono wykresy PACF dla obu badanych szeregów stacjonarnych.

PACF: Szereg zróżnicowany i szereg z usuniętym trendem wiel



Wykres 6: Funkcja PACF dla danych zróżnicowanych (lewa) i po usunięciu trendu wielomianowego (prawa).

Wnioski z analizy PACF (Rysunek 6):

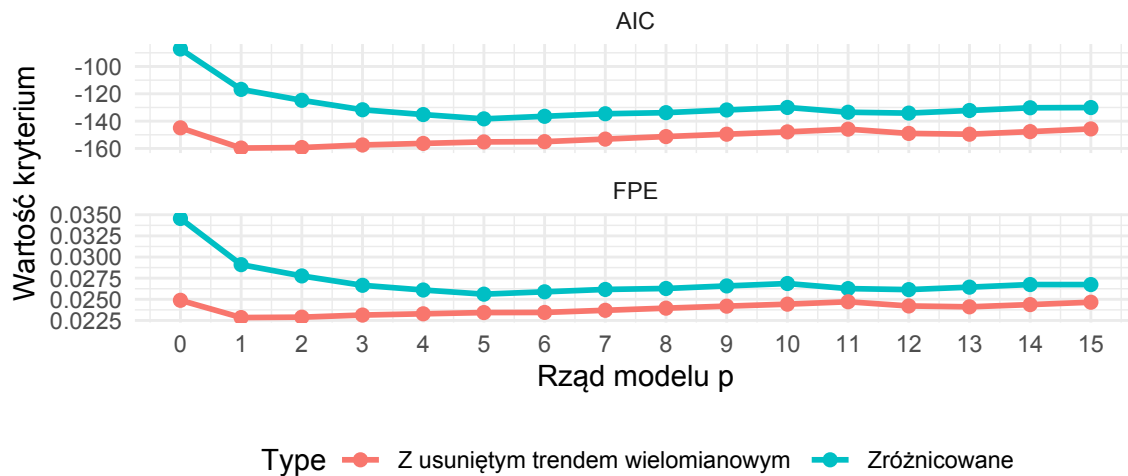
- **Szereg zróżnicowany (lewy):** Obserwujemy istotne statystycznie, ujemne wartości PACF dla opóźnień od 1 do 3, a słupki przy opóźnieniach 4 i 5 znajdują się na granicy istotności. Sugeruje to konieczność zastosowania modelu wyższego rzędu, np. **AR(5)**, **AR(4)** lub **AR(3)**.
- **Szereg z usuniętym trendem wielomianowym (prawy):** Wykres jest bardzo czytelny – występuje jeden wyraźny, dodatni słupek dla opóźnienia $p=1$, po czym funkcja gwałtownie zanika (pozostałe opóźnienia mieszczą się w przedziałach ufności lub wychodzą bardzo nieznacznie). Jest to charakterystyczny obraz dla procesu **AR(1)**.

1.3.2 Analiza kryteriów AIC i FPE

W celu precyzyjnego doboru rzędu, obliczono wartości kryteriów AIC oraz FPE dla modeli AR rzędu $p = 0, \dots, 15$.

```
calc_criteria <- function(data, p_max=15) {  
  aic_vals <- numeric(p_max+1)  
  fpe_vals <- numeric(p_max+1)  
  p_vec <- 0:p_max  
  n <- length(data)  
  
  for(i in 1:length(p_vec)) {  
    p <- p_vec[i]  
    fit <- tryCatch(arima(data, order=c(p,0,0), method="ML"), error=function(e) NULL)  
  
    if(!is.null(fit)) {  
      aic_vals[i] <- fit$aic  
      sigma2 <- fit$sigma2  
      fpe_vals[i] <- sigma2 * (n + p) / (n - p)  
    } else {  
      aic_vals[i] <- NA; fpe_vals[i] <- NA  
    }  
  }  
  return(data.frame(p = p_vec, AIC = aic_vals, FPE = fpe_vals))  
}  
  
results_diff <- calc_criteria(diff_data)  
results_det <- calc_criteria(detrended_data)
```

Kryteria AIC i FPE dla rzędów p od 0 do 15



Wykres 7: Przebieg kryteriów AIC i FPE.

Analiza wykresów kryteriów (Rysunek 7):

1. **Dla danych z usuniętym trendem wielomianowym (linia czerwona):** Minimum obu kryteriów jest osiągane bardzo szybko, już dla $p=1$. Dalsze zwiększanie rzędu powoduje wzrost wartości kryterium. Jest to zgodne z obserwacją z wykresu PACF.
2. **Dla danych zróżnicowanych (linia niebieska):** Wartość kryterium spada wyraźnie aż do rzędu $p=5$, gdzie osiąga lokalne minimum. Wskazuje to, że struktura autokorelacji w przyrostach temperatury jest bardziej złożona i wymaga dłuższego “pamiętania” historii.

Decyzja o wyborze modeli: Biorąc pod uwagę zasadę oszczędności oraz zgodność obu metod identyfikacji (PACF i AIC), do dalszej analizy przyjęto:

- Model **AR(5)** dla podejścia zróżnicowanego (ze względu na minimum AIC przy $p=5$ i istotne PACF do lag 5).
- Model **AR(1)** dla podejścia z trendem wielomianowym (ze względu na wyraźne odcięcie PACF i minimum AIC przy $p=1$).

1.3.3 Automatyczna weryfikacja rzędu p (funkcja `ar`)

Aby potwierdzić wnioski wyciągnięte z analizy wykresów PACF oraz przebiegu kryteriów informacyjnych, wykorzystano wbudowaną funkcję `ar()`. Funkcja ta automatycznie dobiera optymalny rząd modelu autoregresyjnego, minimalizując kryterium AIC.

```
auto_diff <- ar(diff_data, order.max=15, method="mle")
auto_det <- ar(detrended_data, order.max=15, method="mle")
```

Komentarz do wyników automatycznych:

Wyniki uzyskane algorytmicznie funkcją `ar()` potwierdzają wcześniejszą analizę wizualną:

1. Dla **szeregu zróżnicowanego** automat wskazał $p = 5$ (co pokrywa się z lokalnym minimum AIC na wykresie).
2. Dla **szeregu z usuniętym trendem wielomianowym** automat wskazał $p = 1$ (co idealnie potwierdza obserwację jednego silnego słupka na wykresie PACF).

Ostatecznie utwierdza to nas w wyborze modeli **AR(5)** (dla podejścia różnicowego) oraz **AR(1)** (dla podejścia z trendem).

1.4 c) Wyznaczenie estymatorów parametrów

Dla zidentyfikowanych w poprzednim punkcie rzędów modeli (**AR(5)** dla szeregu zróżnicowanego oraz **AR(1)** dla szeregu po usunięciu trendu wielomianowego) przeprowadzono estymację parametrów strukturalnych ϕ_i oraz wariancji szumu σ^2 .

Porównano wyniki uzyskane dwiema metodami:

1. **Metoda Yule’a-Walkera (YW):** Metoda momentów, polegająca na rozwiązaniu liniowego układu równań wiążących teoretyczną autokorelację z parametrami modelu. Jest efektywna obliczeniowo, ale traktowana jako estymacja wstępna.
2. **Metoda Największej Wiarygodności (MLE):** Metoda optymalizacyjna, szukająca takich wartości parametrów, które maksymalizują funkcję wiarygodności danych. Jest to estymator asymptotycznie efektywny, preferowany w końcowym modelowaniu.

```
p_diff <- 5
fit_diff_yw <- ar(diff_data, order.max=p_diff, aic=FALSE, method="yw")
fit_diff_mle <- ar(diff_data, order.max=p_diff, aic=FALSE, method="mle")

p_det <- 1
fit_det_yw <- ar(detrended_data, order.max=p_det, aic=FALSE, method="yw")
fit_det_mle <- ar(detrended_data, order.max=p_det, aic=FALSE, method="mle")
```

```

create_comparison_table <- function(fit_yw, fit_mle) {
  params <- length(fit_yw$ar)

  df <- data.frame(
    Parametr = paste0("phi_", 1:params),
    YW = fit_yw$ar,
    MLE = fit_mle$ar,
    Roznica = fit_yw$ar - fit_mle$ar
  )

  df <- rbind(df, data.frame(
    Parametr = "sigma^2",
    YW = fit_yw$var.pred,
    MLE = fit_mle$var.pred,
    Roznica = fit_yw$var.pred - fit_mle$var.pred
  ))

  return(df)
}

```

Parametr	YW	MLE	Roznica
phi_1	-0.62668	-0.63119	0.00451
phi_2	-0.46256	-0.46989	0.00733
phi_3	-0.39216	-0.39398	0.00182
phi_4	-0.27803	-0.28121	0.00318
phi_5	-0.16784	-0.17191	0.00407
sigma^2	0.02519	0.02417	0.00102

Tabela 6: Porównanie estymatorów YW i MLE dla modelu AR(5) (dane zróżnicowane).

Parametr	YW	MLE	Roznica
phi_1	0.30132	0.30542	-0.00409
sigma^2	0.02289	0.02258	0.00031

Tabela 7: Porównanie estymatorów YW i MLE dla modelu AR(1) (dane po usunięciu trendu wielomianowego).

Komentarz do wyników estymacji: Analiza tabel 6 oraz 7 wykazuje, że różnice pomiędzy estymatorami uzyskanymi metodą Yule’a-Walkera a metodą MLE są marginalne (rzędu 10^{-3}). Wynika to z dużej liczby próby ($N > 170$), dla której estymatory te są praktycznie asymptotycznie równoważne.

Stabilność wyników potwierdza poprawność specyfikacji modeli. Do dalszych etapów analizy (konstrukcja przedziałów ufności, weryfikacja hipotez i prognozowanie) wykorzystane zostaną parametry oszacowane metodą **MLE**, jako posiadające lepsze własności w skończonych próbach.

1.5 d) Asymptotyczne przedziały ufności i weryfikacja istotności

Zgodnie z teorią estymacji procesów autoregresyjnych, estymatory największej wiarygodności (MLE) w dużych próbach mają rozkład asymptotycznie normalny. Pozwala to na wyznaczenie błędów standardowych (SE) oraz konstrukcję przedziałów ufności dla parametrów ϕ_i według wzoru:

$$\hat{\phi}_i \pm u_{1-\alpha/2} \cdot SE(\hat{\phi}_i)$$

gdzie $u_{1-\alpha/2} \approx 1.96$ dla poziomu istotności $\alpha = 0.05$.

Na tej podstawie zweryfikowano hipotezę zerową $H_0 : \phi_i = 0$ (o nieistotności parametru). Jeśli zero nie należy do wyznaczonego przedziału ufności, hipotezę odrzucamy (parametr jest istotny).

```
analize_significance <- function(model_fit) {
  params <- model_fit$ar

  cov_mat <- model_fit$asy.var.coef
  se <- sqrt(diag(cov_mat))

  z_crit <- qnorm(0.975) # ok. 1.96
  lower <- params - z_crit * se
  upper <- params + z_crit * se

  is_significant <- ifelse(lower > 0 | upper < 0, "Tak (*)", "Nie")

  df <- data.frame(
    Parametr = paste0("phi_", 1:length(params)),
    Estymata = params,
    SE = se,
    CI_Dolny = lower,
    CI_Gorny = upper,
    Istotny = is_significant
  )
  return(df)
}
```

Parametr	Estymata	SE	CI_Dolny	CI_Gorny	Istotny
phi.1	-0.6312	0.0747	-0.7776	-0.4847	Tak (*)
phi.2	-0.4699	0.0860	-0.6384	-0.3013	Tak (*)
phi.3	-0.3940	0.0880	-0.5664	-0.2215	Tak (*)
phi.4	-0.2812	0.0860	-0.4498	-0.1127	Tak (*)
phi.5	-0.1719	0.0747	-0.3184	-0.0255	Tak (*)

Tabela 8: Asymptotyczne przedziały ufności dla modelu AR(5) (Dane zróżnicowane).

Parametr	Estymata	SE	CI_Dolny	CI_Gorny	Istotny
phi.1	0.3054	0.0722	0.1639	0.4470	Tak (*)

Tabela 9: Asymptotyczne przedziały ufności dla modelu AR(1) (Dane po usunięciu trendu wielomianowego).

Wnioski z weryfikacji hipotez:

1. **Dla modelu AR(5) (Tabela 8):** Analiza przedziałów ufności pokazuje, że wszystkie opóźnienia są statystycznie istotne.
2. **Dla modelu AR(1) (Tabela 9):** Parametr ϕ_1 jest wysoce istotny statystycznie. Potwierdza to występowanie silnej autokorelacji w resztach po usunięciu trendu wielomianowego, którą model AR(1) skutecznie opisuje.

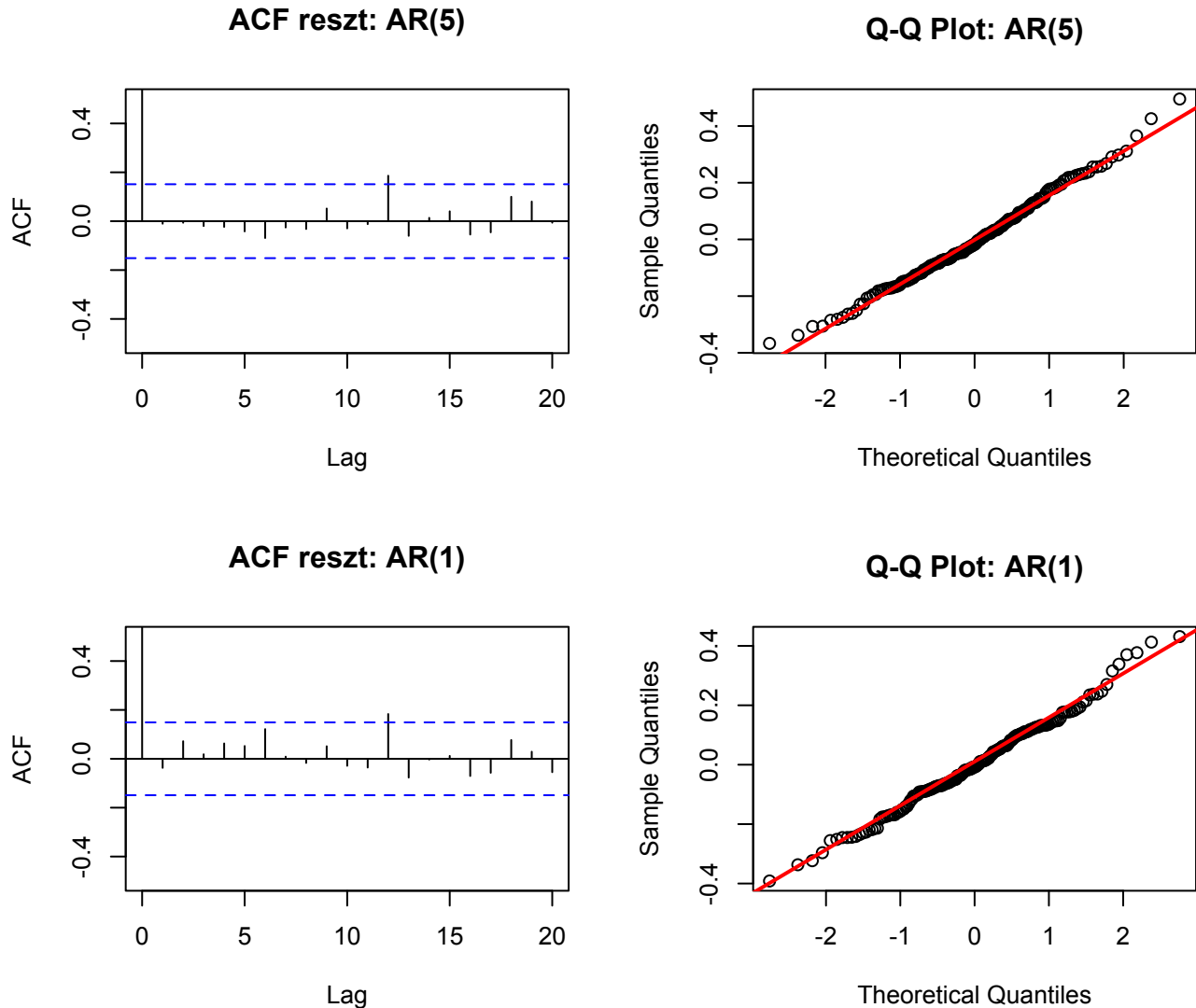
1.6 e) Weryfikacja (na podstawie analizy reszt) poprawności dopasowania

Ostatnim, kluczowym etapem modelowania jest diagnostyka reszt. Prawidłowo zbudowany model powinien pozostawić reszty będące realizacją procesu **Białego Szumu**. Oznacza to, że reszty powinny być nieskorelowane oraz posiadać rozkład zbliżony do normalnego.

W celu weryfikacji modeli **AR(5)** oraz **AR(1)**, wyestymowanych metodą MLE, przeprowadzono analizę wizualną oraz formalne testy statystyczne.

1.6.1 Analiza wizualna reszt

Sporządzono wykresy funkcji autokorelacji (ACF) reszt, aby sprawdzić, czy nie pozostały w nich istotne zależności, oraz wykresy kwantylowe (Q-Q Plot) w celu oceny normalności rozkładu.



Wykres 8: Diagnostyka reszt: ACF (lewa kolumna) i Q-Q Plot (prawa kolumna). Górny wiersz: AR(5) Różnicowanie, Dolny wiersz: AR(1) trend wielomianowy.

Wnioski wizualne (Rysunek 8):

1. **Autokorelacja (ACF):** Dla obu modeli słupki autokorelacji mieszczą się niemal idealnie w granicach przedziałów ufności (oczywiście poza słupkiem dla argumentu zero). Świadczy to o tym, że oba modele skutecznie “wyczyściły” szereg z zależności czasowych.
2. **Normalność (Q-Q Plot):** Punkty na wykresach kwantylowych układają się dosyć dokładnie wzdłuż czerwonej linii teoretycznej, nawet na krańcach rozkładu. Sugeruje to, że rozkład reszt jest bardzo zbliżony do normalnego, co jest sytuacją bardzo pożądaną.

1.6.2 Testy statystyczne (Białoszumowość)

Weryfikację wizualną uzupełniono testami statystycznymi:

- **Test Ljung-Boxa:** H_0 : Reszty są nieskorelowane.
- **Test Shapiro-Wilka:** H_0 : Reszty mają rozkład normalny.

```
run_residual_tests <- function(residuals, model_name) {  
  lb_test <- Box.test(residuals, lag = 10, type = "Ljung-Box")  
  sw_test <- shapiro.test(residuals)  
  
  data.frame(  
    Model = model_name,  
    `Ljung-Box (stat)` = lb_test$statistic,  
    `Ljung-Box (p-val)` = lb_test$p.value,  
    `Shapiro-Wilk (p-val)` = sw_test$p.value  
  )  
}  
  
diag_results <- rbind(  
  run_residual_tests(resid_diff, "AR(5) - Zróźnicowane"),  
  run_residual_tests(resid_det, "AR(1) - Z usuniętym trendem wielomianowym")  
)  
  
xtab_diag <- xtable(diag_results,  
  caption = "Wyniki testów diagnostycznych reszt (Ljung-Box i Shapiro-Wilk).",  
  label = "tab:diag-tests", digits = 4)  
print(xtab_diag, include.rownames=FALSE, table.placement="H", comment=FALSE)
```

Model	Ljung.Box..stat.	Ljung.Box..p.val.	Shapiro.Wilk..p.val.
AR(5) - Zróźnicowane	2.2718	0.9938	0.4895
AR(1) - Z usuniętym trendem wielomianowym	5.8056	0.8313	0.5047

Tabela 10: Wyniki testów diagnostycznych reszt (Ljung-Box i Shapiro-Wilk).

Interpretacja wyników testów (Tabela 10):

1. **Brak autokorelacji (Ljung-Box):** Zarówno dla modelu AR(5) ($p \approx 0.99$), jak i AR(1) ($p \approx 0.83$), wartości p-value są bardzo wysokie (znacznie powyżej poziomu istotności $\alpha = 0.05$).
 - **Wniosek:** Nie ma podstaw do odrzucenia hipotezy zerowej. Reszty są w pełni **nieskorelowane**. Oznacza to, że oba modele poprawnie opisały strukturę zależności czasowej w danych.
2. **Normalność rozkładu (Shapiro-Wilk):** W badanym przypadku testy **nie odrzucają hipotezy o normalności**.
 - Dla AR(5): $p \approx 0.49 > 0.05$.
 - Dla AR(1): $p \approx 0.50 > 0.05$.
 - **Wniosek:** Reszty mają rozkład zbliżony do normalnego.

Podsumowanie diagnostyki: Oba modele przeszły weryfikację wzorowo. Spełniają one rygorystyczne założenia białoszumowości (brak korelacji oraz gaussowski charakter reszt). Ze względu na nieco wyższe p-value w teście Ljung-Boxa, model **AR(5)** dla danych zróźnicowanych wydaje się minimalnie lepiej dopasowany, ale oba są w pełni poprawne statystycznie.

1.7 f) Szczegółowa prognoza i porównanie podejść

W tej sekcji przeprowadzono prognozę na horyzont $h = 10$ lat (2024–2033). Aby dokładnie przeanalizować różnice metodyczne, rozbito proces na cztery osobne wizualizacje, obejmujące zarówno podejście “ręczne” (rekonstrukcja zmiennej), jak i automatyczne (ARIMA), a także porównanie z modelem deterministycznym.

1.7.1 Przygotowanie danych i obliczenia

W pierwszej kolejności wykonano obliczenia prognoz. Dla modelu stochastycznego manualnego oraz modelu deterministycznego wyznaczamy jedynie prognozy punktowe (bez przedziałów ufności), aby skupić się na porównaniu ścieżek trendu.

```
library(ggplot2)
library(forecast)
h <- 10
future_years <- seq(end(gtemp_both)[1] + 1, length.out = h)

pred_diff_manual <- predict(fit_diff_mle, n.ahead = h)

last_val <- as.numeric(tail(gtemp_both, 1))
forecast_manual_stoch <- last_val + cumsum(pred_diff_manual$pred)

model_arima_auto <- Arima(gtemp_both, order=c(5,1,0), include.drift=TRUE)
fcast_auto <- forecast(model_arima_auto, h=h, level=95)

pred_resid <- predict(fit_det_mle, n.ahead = h)
coef_trend <- coef(model_trend)

t_start <- time(gtemp_both)[length(gtemp_both)]
t_seq <- seq(t_start + 1, length.out = h)
trend_vals <- coef_trend[1] + coef_trend[2]*t_seq + coef_trend[3]*t_seq^2

forecast_manual_det <- trend_vals + pred_resid$pred

df_hist <- data.frame(
  Rok = as.numeric(time(gtemp_both)),
  Wartosc = as.numeric(gtemp_both),
  Typ = "Historia"
)
df_hist <- subset(df_hist, Rok >= 2005)

df_p1 <- data.frame(
  Rok = future_years,
  Wartosc = as.numeric(forecast_manual_stoch)
)

df_p2 <- data.frame(
  Rok = future_years,
  Wartosc = as.numeric(fcast_auto$mean),
  Lower = as.numeric(fcast_auto$lower),
  Upper = as.numeric(fcast_auto$upper)
)

df_p3 <- data.frame(
  Rok = future_years,
```



```
Wartosc = as.numeric(forecast_manual_det)
)
```

1.7.2 Analiza wizualna prognoz

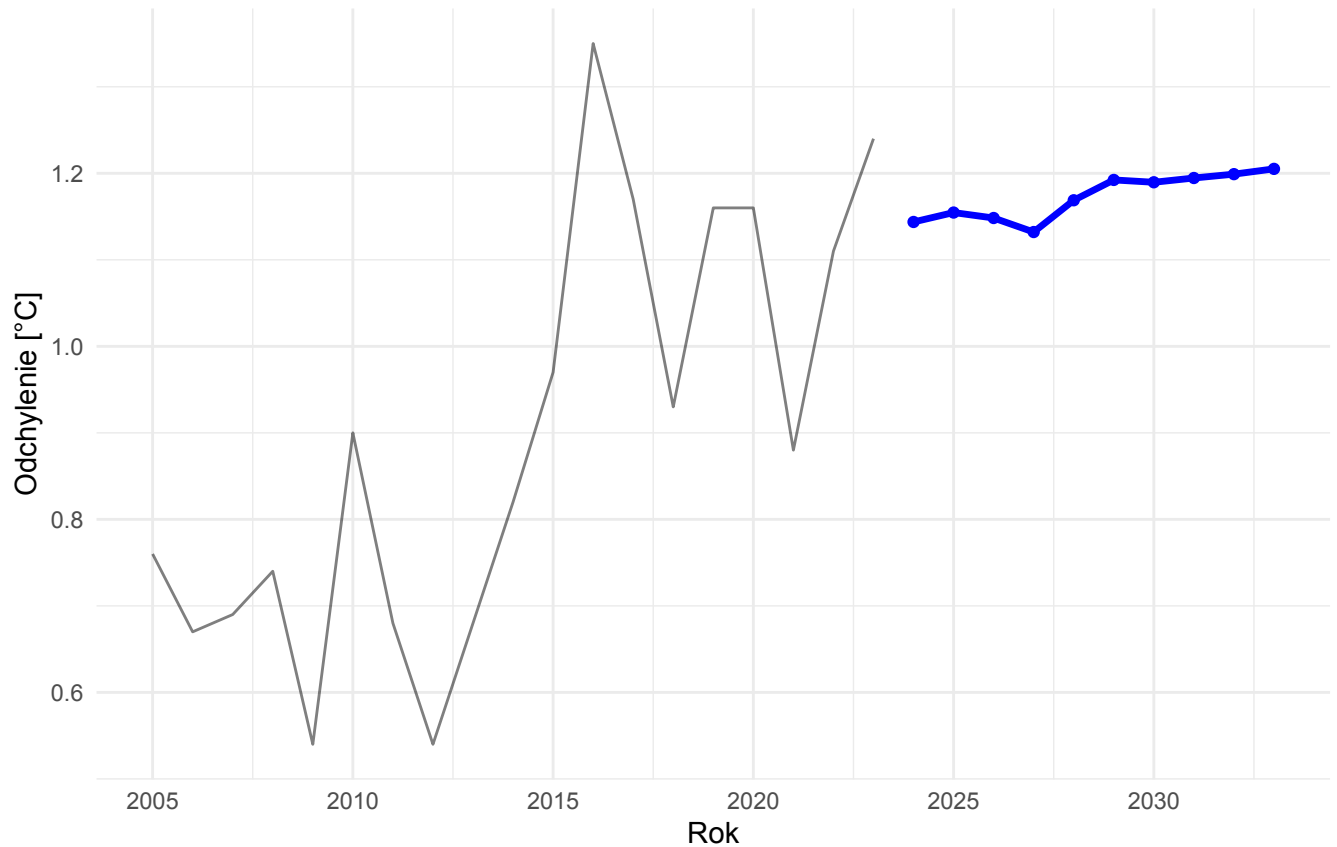
Poniżej przedstawiono cztery ujęcia prognozowanego zjawiska.

Wariant 1: Podejście stochastyczne (Obliczenia ręczne)

Pierwszy wykres prezentuje “czystą” ścieżkę prognozy uzyskaną poprzez zsumowanie prognozowanych przyrostów z modelu AR(5). Metoda ta pokazuje najbardziej prawdopodobny scenariusz wynikający z dynamiki procesu, startując od ostatniej obserwacji.

1. Model różnicowy (Obliczenia ręczne)

Rekonstrukcja zmiennej (prognoza punktowa)

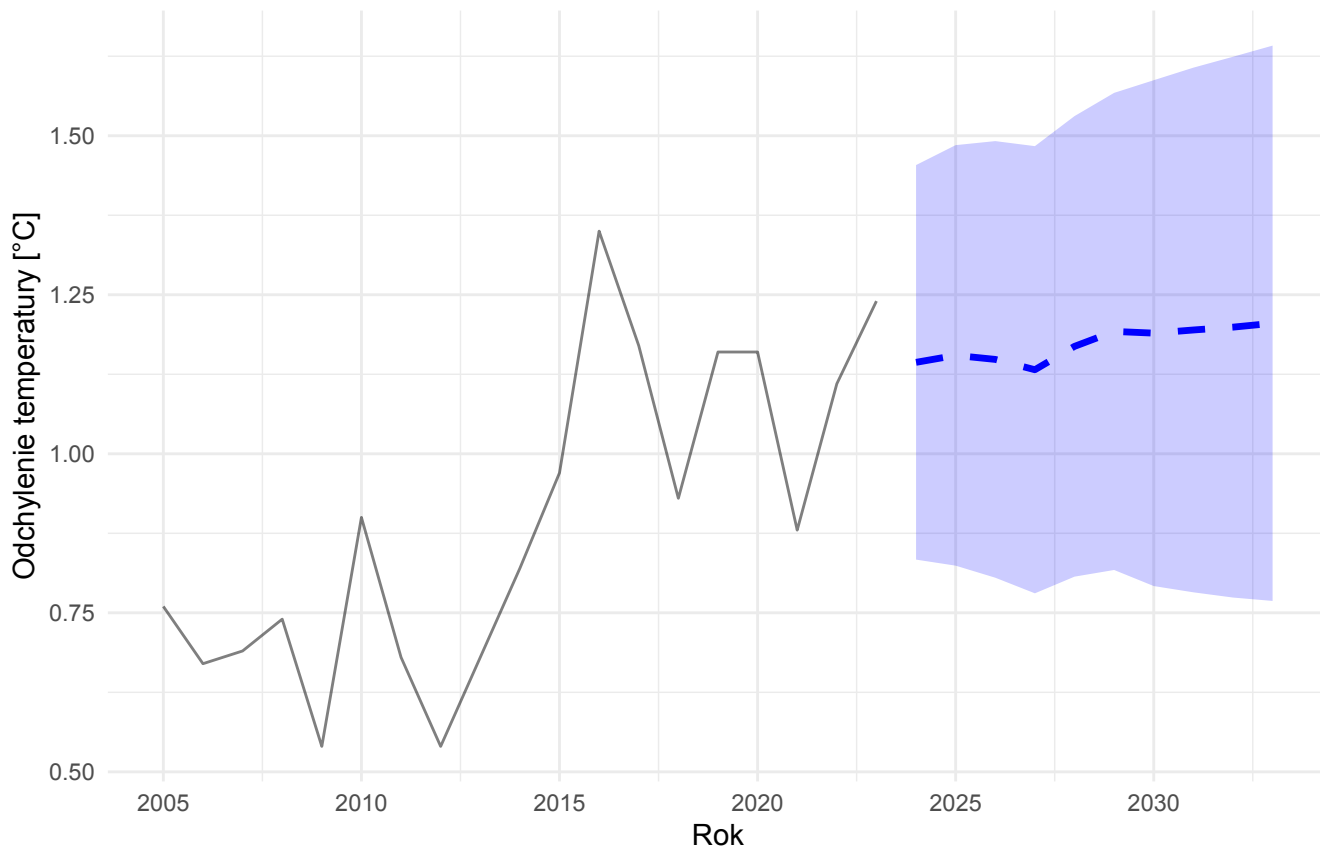


Wykres 9: Prognoza ręczna z modelu AR(5) (punktowa).

Wariant 2: Podejście stochastyczne (ARIMA z przedziałami ufności) W tym ujęciu wykorzystano funkcję `forecast` dla modelu ARIMA(5,1,0). Kluczowym elementem jest tutaj “rozszerzający się” przedział ufności. Wynika on z faktu, że dla procesu zintegrowanego wariancja błędu prognozy rośnie liniowo wraz z horyzontem czasu. Ten wykres prezentujemy z przedziałami ufności jako punkt odniesienia dla niepewności prognozy.

2. Model różnicowy (Automat ARIMA)

Przedziały ufności uwzględniające niestacjonarność procesu

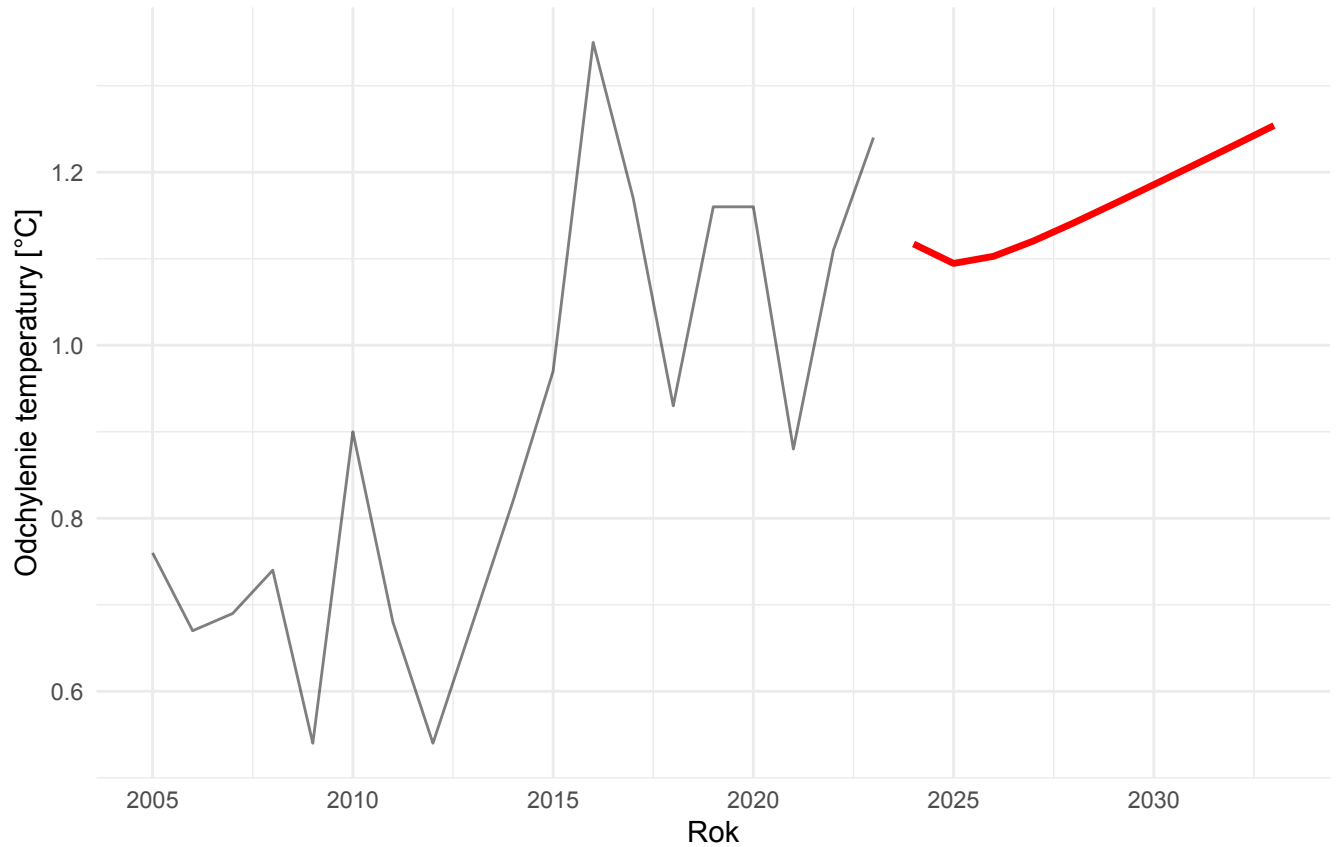


Wykres 10: Prognoza ARIMA(5,1,0) z rosnącym przedziałem ufności.

Wariant 3: Podejście deterministyczne (Trend + AR) Wykres trzeci przedstawia prognozę opartą na ekstrapolacji trendu wielomianowego (paraboli).

3. Model Trendu + AR(1) (Obliczenia ręczne)

Ekstrapolacja trendu kwadratowego skorygowana o prognozę reszt

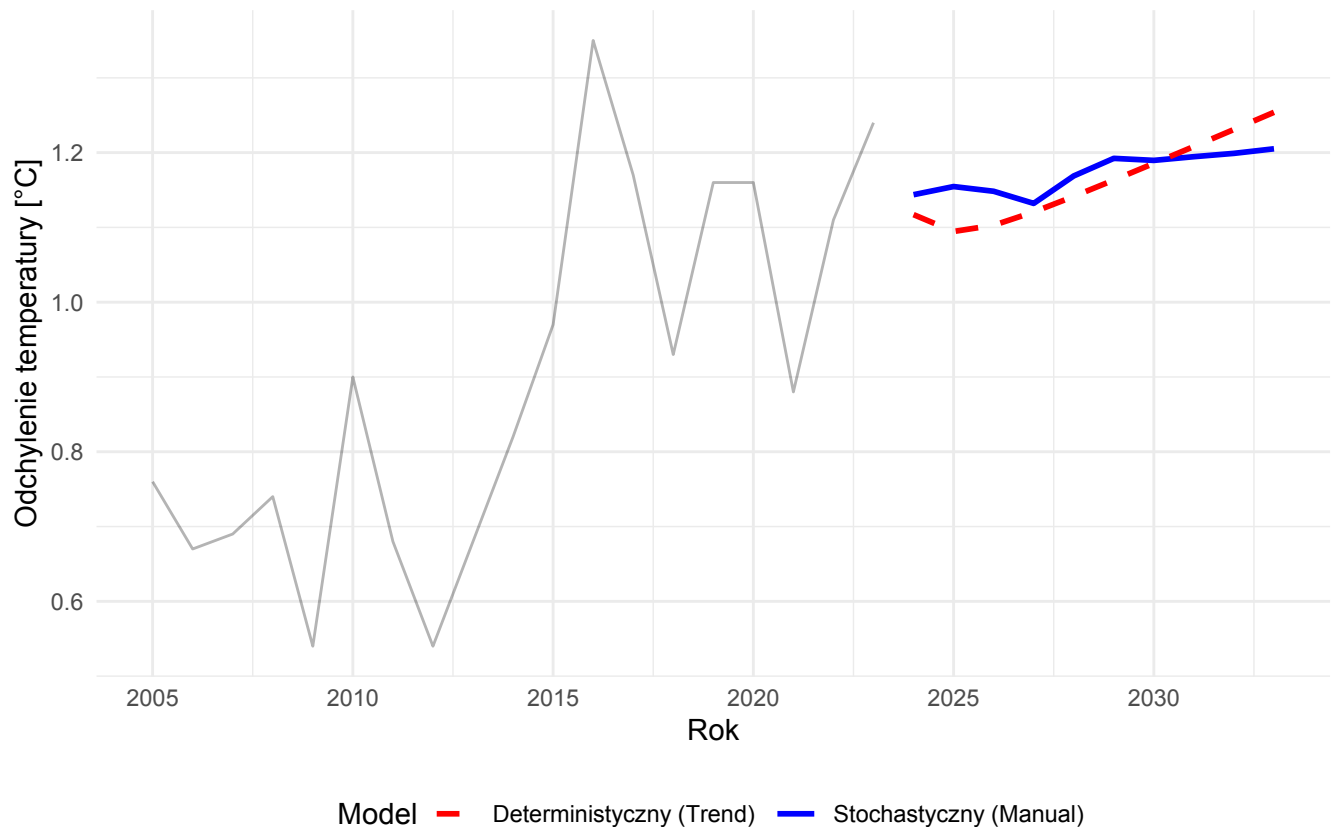


Wykres 11: Prognoza deterministyczna: Trend wielomianowy + AR(1).

Wariant 4: Bezpośrednie porównanie Zestawienie obu podejść (manualnego stochastycznego i deterministycznego) w formie czystych linii prognozy pozwala na wyraźne porównanie ich dynamiki. Model deterministyczny (czerwony) przewiduje szybszy, nieliniowy wzrost temperatur (“ucieczkę” w górę) ze względu na kwadratową naturę trendu. Model stochastyczny (niebieski) jest bardziej liniowy w swojej strukturze.

4. Porównanie: Manual Stochastyczny vs Deterministyczny

Niebieski: Różnicowanie (AR5) | Czerwony: Trend kwadratowy + AR(1)



Wykres 12: Porównanie podejścia stochastycznego i deterministycznego.

1.8 g) Końcowe wnioski i podsumowanie

Przeprowadzona analiza szeregu czasowego globalnych anomalii temperatury (zbiór `gtemp_both`) pozwoliła na porównanie dwóch odmiennych podejść do modelowania procesów niestacjonarnych: podejścia stochastycznego (różnicowanie) oraz deterministycznego (usunięcie trendu wielomianowego).

Porównanie dopasowania i diagnostyki:

1. **Stacjonarność:** Obie metody skutecznie sprowadziły szereg do postaci stacjonarnej. Testy ADF i KPSS potwierdziły, że zarówno jednokrotne różnicowanie ($d = 1$), jak i usunięcie trendu kwadratowego wyeliminowało niestacjonarność w średniej.
2. **Struktura modelu:**
 - Dla danych zróżnicowanych zidentyfikowano model **AR(5)**. Jest to model bardziej złożony, co wynika z konieczności opisanie dynamiki krótkoterminowych zmian.
 - Dla danych po usunięciu trendu kwadratowego najlepszy okazał się model **AR(1)**, co sugeruje, że większość “informacji” o procesie została zawarta w deterministycznym trendzie wielomianowym.
3. **Diagnostyka reszt:** Oba modele przeszły pomyślnie testy białoszumowości (Ljung-Box). Co istotne, w obu przypadkach reszty wykazywały rozkład zbliżony do normalnego, co jest wynikiem bardzo dobrym jak na dane środowiskowe.

Wnioski z prognoz (2024–2033):

1. **Kierunek zmian:** Niezależnie od przyjętej metodyki, oba modele prognozują dalszy, wyraźny wzrost średniej temperatury globalnej w najbliższej dekadzie.
2. **Różnice w scenariuszach:**
 - Podejście **deterministyczne (Trend kwadratowy)** generuje prognozy wyższe (“pesymistyczne”), zakładając stałe przyspieszenie ocieplenia wynikające z natury funkcji kwadratowej.
 - Podejście **stochastyczne (Różnicowanie/ARIMA)** jest bardziej “ostrożne” i adaptacyjne. Jego prognoza w większym stopniu zależy od ostatniej dynamiki przyrostów, a przedziały ufności słusznie rozszerzają się wraz z horyzontem prognozy, odzwierciedlając rosnącą niepewność procesów skumulowanych.

Rekomendacja: Biorąc pod uwagę naturę zjawisk klimatycznych (które rzadko podążają idealnie po sztywnych krzywych matematycznych w długim okresie) oraz poprawność diagnostyczną, model **stochastyczny (ARIMA/Różnicowanie)** wydaje się bezpieczniejszym narzędziem do prognozowania krótkoterminowego, gdyż lepiej oddaje losową naturę zmienności klimatu i nie narzuca naiwnego kształtu wzrostu “w nieskończoność”, jak robi to wielomian. Niemniej jednak, zgodność obu prognoz co do kierunku zmian jest silnym sygnałem potwierdzającym trend ocieplenia.

2 Zadanie 4 - Porównanie dokładności prognoz dla danych euretail

2.1 Wstęp

Cel zadania: porównanie dokładności prognoz skonstruowanych na bazie modeli (S)ARIMA, modeli dekompozycji oraz algorytmów wygładzania wykładniczego dla danych `euretail` (pakiet `fpp2`) zawierających kwartalne wartości indeksu handlu detalicznego (ang. retail trade index) dla strefy euro.

2.2 Podział danych na część uczącą i testową

Zacznijmy od sprawdzenia zakresu naszych danych:

```
start(euretail)
```

```
## [1] 1996      1
```

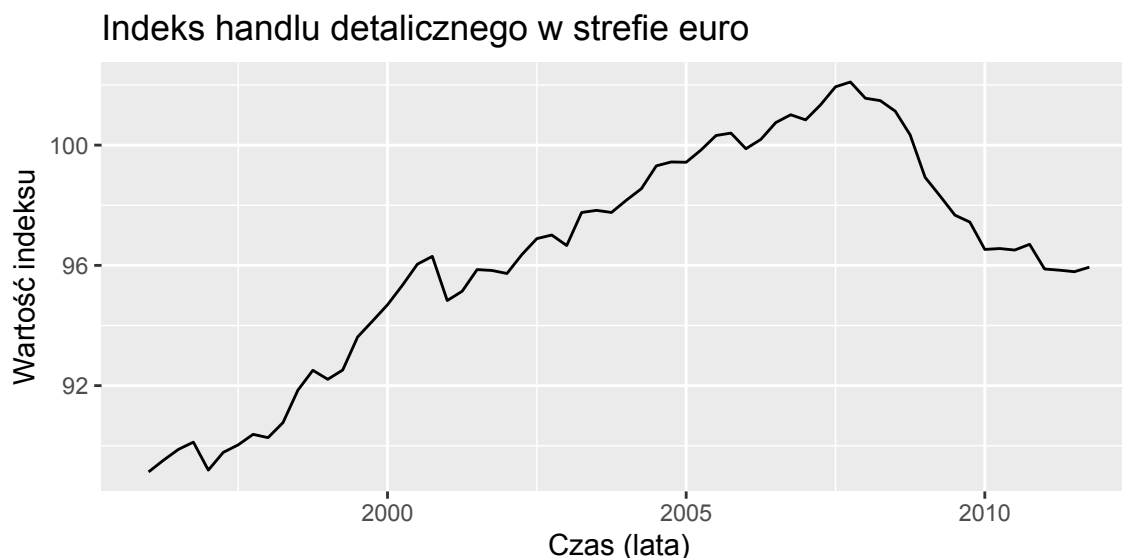
```
end(euretail)
```

```
## [1] 2011      4
```

```
length(euretail)
```

```
## [1] 64
```

Spójrzmy na wykres:



Wykres 13: Wykres szeregu czasowego euretail

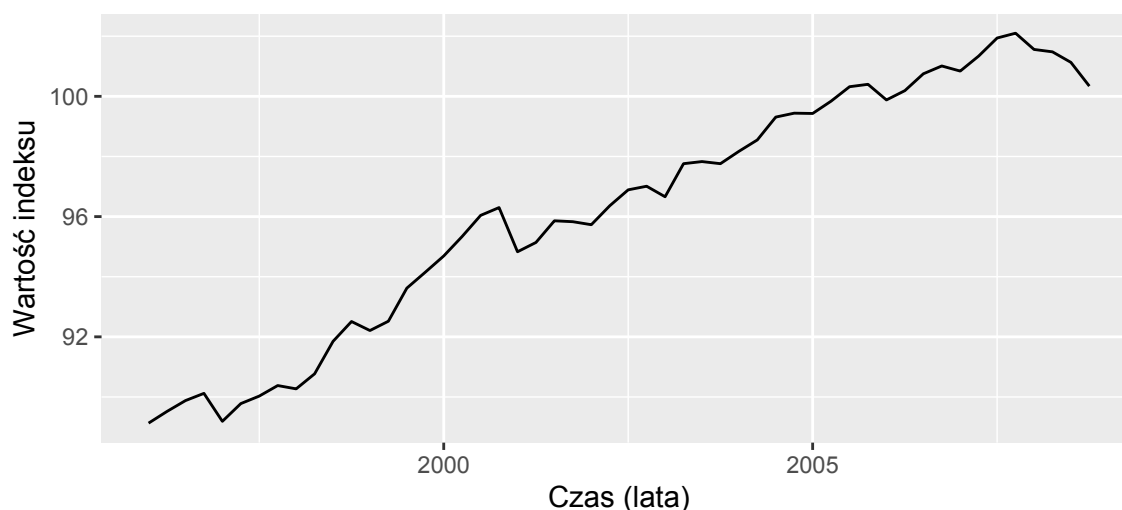
Możemy zauważyć, że po roku 2008 występuje gwałtowne załamanie trendu wzrostowego.

Dane `euretail` są typu kwartalnego, zatem możemy je podzielić na podstawie liczby kwartałów. Zbiór testowy powinien stanowić około 20% całkowitej długości. W naszym przypadku 20% to 12.8 kwartałów. Niech zatem naszym zbiorem testowym będzie ostatnie 12 kwartałów, czyli 3 lata.

2.3 Dopasowanie odpowiednich modeli do danych

Przyjrzyjmy się wykresowi dla zbioru uczącego.

Zbiór treningowy szeregu euretail

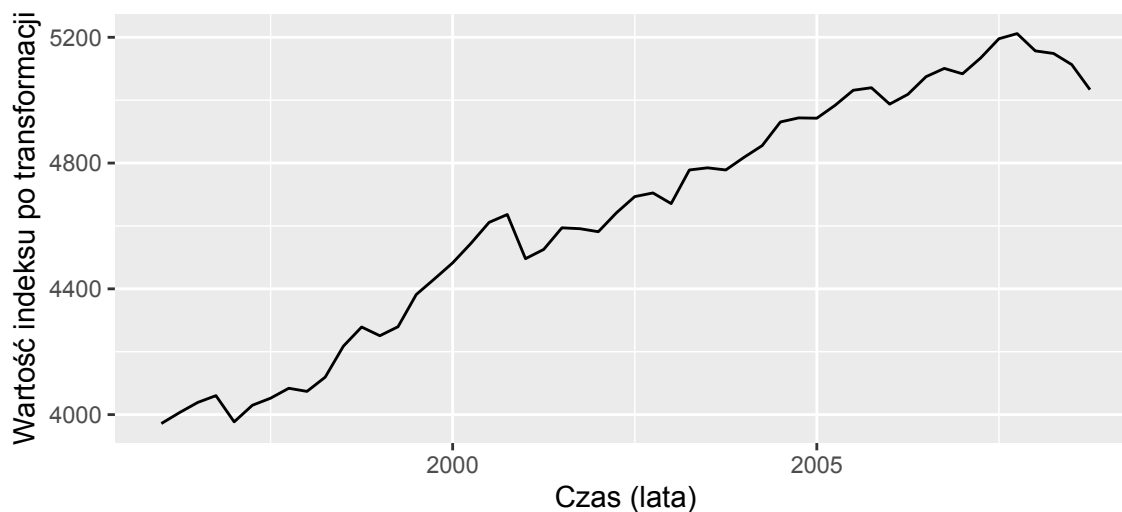


Wykres 14: Wykres szeregu czasowego dla zbioru uczącego euretail

Warto zauważyć, że przyjęty podział danych sprawił, iż zbiór uczący kończy się tuż przed wyraźnym załamaniem trendu wzrostowego (widocznym na wykresie 13). Modele klasy SARIMA oraz ETS opierają się na ekstrapolacji dotychczasowych tendencji, można zatem postawić hipotezę, że wyznaczone prognozy będą obarczone znacznym błędem wynikającym z braku informacji o zmianie strukturalnej trendu w danych treningowych.

W celu stabilizacji wariancji szeregu czasowego zastosujemy transformację Boxa-Coxa. Optymalna wartość parametru λ została wyznaczona metodą największej wiarygodności i wyniosła $\lambda = 2$.

Zbiór treningowy szeregu euretail po transformacji Boxa-Coxa



Wykres 15: Wykres szeregu czasowego dla zbioru uczącego euretail po transformacji Boxa-Coxa

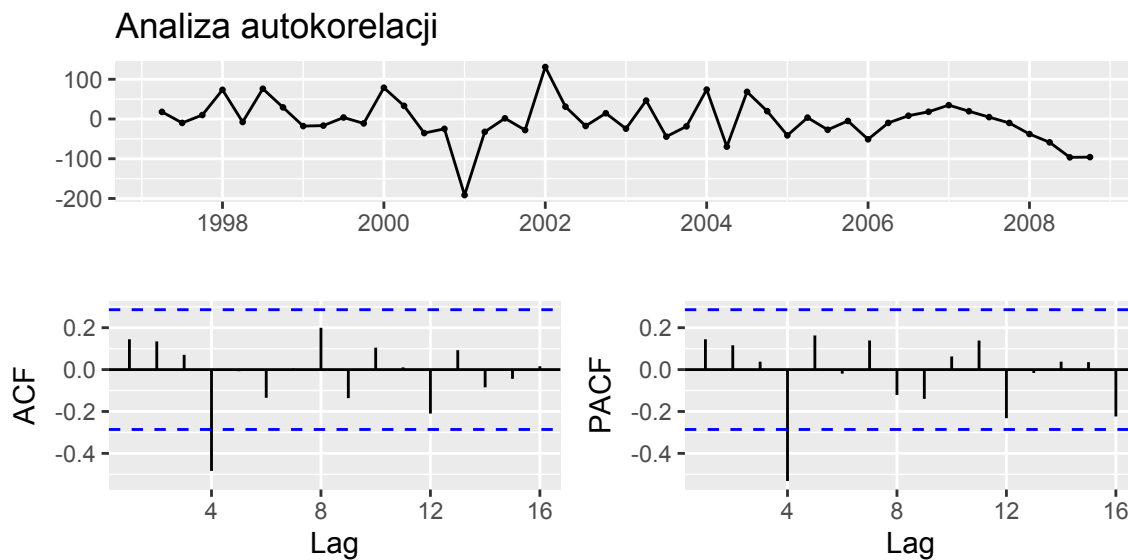
Po transformacji Boxa-Coxa wizualnie niewiele się zmieniło. Jednakże jej zastosowanie jest uzasadnione z punktu widzenia statystycznego. Pozwala ona na lepsze spełnienie założenia o homoskedastyczności i normalności składnika losowego w dalszych etapach modelowania, co przekłada się na bardziej rzetelną estymację przedziałów ufności prognoz.

Teraz wykorzystamy funkcje `ndiffs` i `nsdiffs` aby sprawdzić jakie różnicowania wykonać.

```
d<-ndiffs(train.retail.bc)
```

```
D<-nsdiffs(train.retail.bc)
```

Otrzymaliśmy wartości $d = 1$ i $D = 1$. Zatem wykonamy jednokrotne różnicowanie zwykłe i jednokrotne różnicowanie sezonowe (z opóźnieniem 4, ze względu na kwartalny charakter danych). Warto również podkreślić, że dodatnia wartość D daje nam sygnał, że w danych występuje sezonowość. Z tego powodu skupimy się na modelach SARIMA.



Wykres 16: Analiza autokorelacji szeregu po różnicowaniu zwykłym ($d=1$) i sezonowym ($D=1$)

Wykres 16 pozwala ocenić skuteczność przeprowadzonego różnicowania. Brak wyraźnych trendów na wykresie czasowym oraz szybkie wygasanie funkcji ACF i PACF sugerują, że szereg stał się stacjonarny i jest gotowy do estymacji parametrów modelu. Widzimy jedynie istotne słupki dla opóźnienia 4, zatem mamy potwierdzenie, że w danych występuje sezonowość i użycie SARIMA jest uzasadnione.

Warto również sprawdzić średnią, która wskaże nam czy należy uwzględnić dryf.

```
mean(train.diff)
```

```
## [1] -3.904187
```

Fakt, że średnia ta wyraźnie odbiega od zera, jest przesłanką do uwzględnienia stałej (dryfu) w modelu SARIMA.

2.3.1 SARIMA

2.3.1.1 Wybór parametrów Model SARIMA opisujemy siedmioma parametrami (p, d, q) (P, D, Q) [s].

Częstotliwość sezonowa [s] - określa liczbę obserwacji, po których cykl sezonowy się powtarza. Dla euretail:

- $s = 4$.

Część trendu (p, d, q) - co dzieje się między sąsiednimi kwartałami (opóźnienia 1-3):

- d (różnicowanie): U nas $d = 1$, bo wykonaliśmy jednokrotne różnicowanie zwykłe.

- p (autoregresja): decyduje o niej zachowanie wykresu PACF dla niskich opóźnień (1-3). U nas od początku są poniżej poziomu istotności, stąd $p = 0$.
- q (średnia ruchoma): decyduje o niej zachowanie wykresu ACF dla niskich opóźnień (1-3). Tutaj ponownie $q = 0$, gdyż początkowe słupki nie są istotne.

Część sezonowa (P , D , Q) – co dzieje się między tymi samymi kwartałami rok do roku (opóźnienia 4, 8, 12):

- D (różnicowanie sezonowe): Wybieramy $D = 1$, bo wykonaliśmy jednokrotne różnicowanie sezonowe.
- P (sezonowa autoregresja): Wyraźny pojedynczy słupek w $t = 4$ na wykresie PACF sugeruje $P = 1$. Oprócz tego przetestujemy wariant $P = 0$.
- Q (sezonowa średnia ruchoma): Na wykresie ACF wystaje duży słupek dla opóźnienia 4. Sprawdźmy więc $Q = 1$ jak i $Q = 0$.

Porównamy następujące modele:

- Pierwszy model $SARIMA(0,1,0)(0,1,1)[4]$
- Drugi model $SARIMA(0,1,0)(1,1,0)[4]$
- Trzeci model $SARIMA(0,1,0)(1,1,1)[4]$
- Czwarty model $SARIMA$ - automatyczny

2.3.1.2 Kryteria oceny skuteczności Do oceny skuteczności modelu wspomozemy się dwoma kryteriami:

- AICc (Akaike Information Criterion - corrected) - Wskazuje model, który daje najlepsze prognozy nawet jeśli model jest złożony.
- BIC (Bayesian Information Criterion) Zazwyczaj wybiera modele prostsze (oszczędniejsze) niż AICc.

W przypadku obu - mniejszy wynik to lepsza skuteczność.

Tabela 11: Porównanie kryterium informacyjnego AICc i BIC dla wybranych modeli SARIMA

Specyfikacja modelu	Wartość AICc	Wartość BIC
$SARIMA(0,1,0)(0,1,1)$	489.74	493.17
$SARIMA(0,1,0)(1,1,0)$	495.45	498.88
$SARIMA(0,1,0)(1,1,1)$	491.57	496.56
Auto-ARIMA	503.28	512.46

Pod względem AICc mamy dwie podobne niskie wartości dla modelu pierwszego o parametrach $P=0$, $Q=1$ i trzeciego $P=1$, $Q=1$. Jednakże BIC, zgodnie z oczekiwaniami, wskazuje na wybór prostszego modelu pierwszego o parametrach $P=0$ i $Q=1$.

2.3.1.3 Sprawdzenie istotności współczynników Sprawdźmy, czy współczynnik jest istotny.

```
lmtest::coeftest(fit1)
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## sma1 -0.87912    0.18852 -4.6634 3.111e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Otrzymany wynik (***) testu istotności potwierdza, że współczynnik sezonowej średniej ruchomej $Q = 1$ (oznaczony jako `sma1`) jest istotny statystycznie ($p < 0,001$), jeśli za poziom istotności przyjmujemy $\alpha = 0.05$.

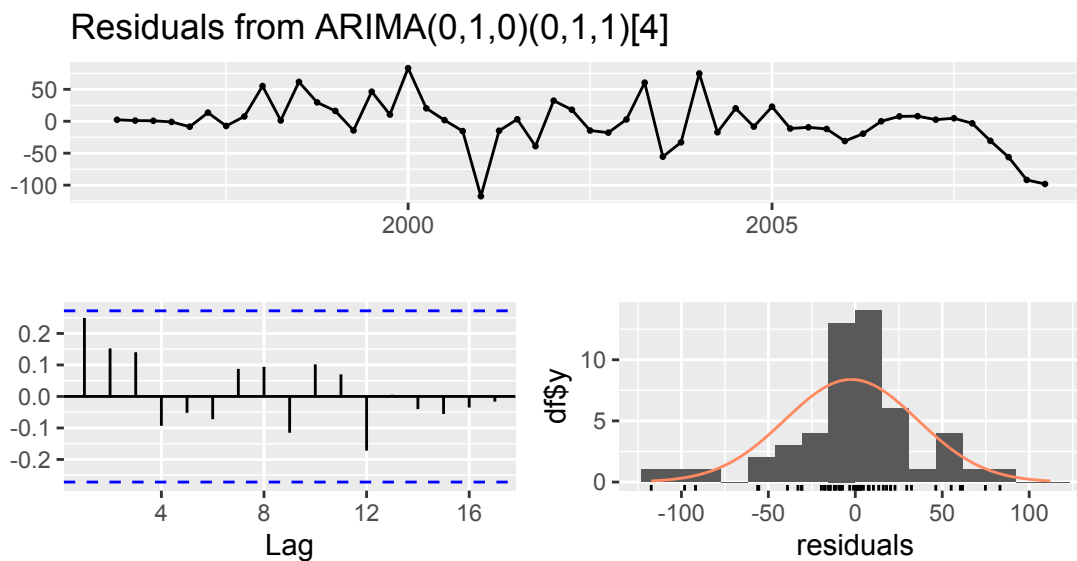
2.3.1.4 Dodanie dryfu

Tabela 12: Wpływ uwzględnienia parametru dryfu na jakość modelu SARIMA

Specyfikacja	Wartość AICc	Wartość BIC
SARIMA(0,1,0)(0,1,1)	489.74	493.17
SARIMA(0,1,0)(0,1,1) z dryfem	489.74	493.17

Warto zauważyć, że próba włączenia dryfu do modelu fit1 nie zmieniła jego parametrów. Jest to zgodne z dokumentacją funkcji Arima(), która stwierdza: ‘if there is more than one difference taken, no constant is included [...] as otherwise quadratic and higher order polynomial trends would be induced’. Ponieważ w analizowanym modelu zastosowano $d=1$ i $D=1$, algorytm słusznie pominął stałą, chroniąc prognozę przed niestabilnym trendem kwadratowym.

2.3.1.5 Wykres reszt i test Ljung-Boxa



Wykres 17: Analiza reszt i opóźnień w najlepszym modelu SARIMA

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)(0,1,1)[4]
## Q* = 7.8865, df = 7, p-value = 0.3427
##
## Model df: 1.   Total lags used: 8
```

Jak widzimy na wykresie 17 wszystkie słupki opóźnień mieszczą się w przedziale ufności. Ponadto histogram przypomina kształt dzwonu, co sugeruje, że założenie o normalności rozkładu składnika losowego jest w przybliżeniu spełnione. Jednakże pojedyncze kreski daleko od centrum (np. te w okolicach -100) potwierdzają obecność nietypowych zdarzeń w danych.

2.3.2 ETS

W tej części analizujemy modele klasy ETS (Error, Trend, Seasonal). Zamiast polegać wyłącznie na algorytmie, przetestujemy dwa podejścia:

- automatyczny dobór struktury
- model zadeklarowany ręcznie na podstawie charakterystyki danych euretail

Wyberzemy model ETS(A,A,A), czyli klasyczną metodę Holta-Wintersa. Jest to model, który najlepiej pasuje do naszych danych, bo rozkłada szereg na trzy logiczne części, z których każda jest traktowana jako stała wartość (addytywnie):

- Pierwsze A (Error): Przyjmujemy addytywny rozkład błędów, co oznacza, że losowe wahania wokół trendu mają stałą rozpiętość.
- Drugie A (Trend): Zakładamy trend liniowy – handel detaliczny w naszym zbiorze treningowym rośnie w miarę stałym tempie.
- Trzecie A (Seasonality): Wybieramy addytywną sezonowość, ponieważ po zastosowaniu transformacji Boxa-Coxa wahania kwartalne zostały wyrównane i mają podobną skalę w całym okresie.

```
fit_ets_auto <- ets(train.retail, lambda = lambda.bc)

fit_ets_manual <- ets(train.retail, model = "AAA", lambda = lambda.bc)
```

Spójrzmy na wartości kryteriów informacyjnych dla wybranych przez nas modeli.

Tabela 13: Porównanie kryteriów informacyjnych dla automatycznego i ręcznego modelu ETS

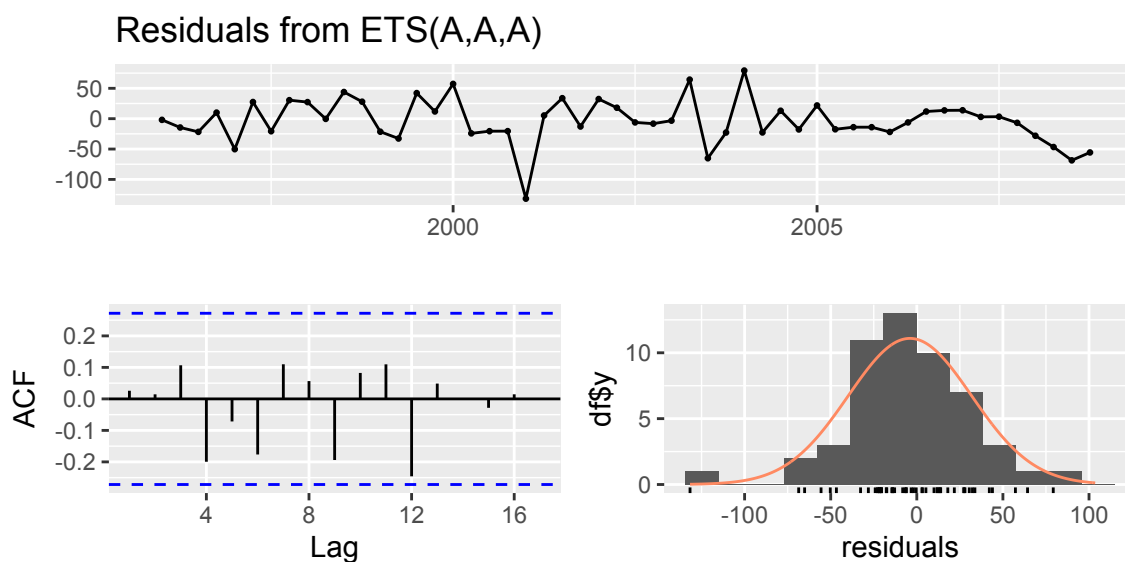
Model	AICc	BIC
ETS Auto: ETS(A,A,A)	599.67	612.95
ETS(A,A,A)	599.67	612.95

Warto zauważyć, że algorytm automatycznego doboru modelu `ets()` wskazał na strukturę ETS(A,A,A) jako optymalną, co jest w pełni zgodne z modelem zadeklarowanym ręcznie. Identyczne wartości kryteriów AICc i BIC dla obu podejść potwierdzają, że addytywna metoda Holta-Wintersa najlepiej oddaje strukturę szeregu euretail po transformacji Boxa-Coxa.

Spójrzmy również na miary dokładności prognoz.

Tabela 14: Błędy dopasowania ETS (zbiór uczący)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ETS Auto	-0.039	0.372	0.278	-0.039	0.291	0.254	0.013
ETS(A,A,A)	-0.039	0.372	0.278	-0.039	0.291	0.254	0.013



Wykres 18: Analiza reszt i opóźnień w najlepszym modelu ETS

```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(A,A,A)
## Q* = 6.1942, df = 8, p-value = 0.6255
##
## Model df: 0.   Total lags used: 8
```

Analiza diagnostyczna 18 modelu ETS(A,A,A) potwierdza jego wysoką jakość. Wykres ACF nie wykazuje istotnych korelacji (wszystkie słupki mieszczą się w przedziale ufności), a wynik testu Ljung-Boxa z wartością

p-value = 0.6255 (powyżej $\alpha = 0.05$) pozwala przyjąć hipotezę o braku autokorelacji reszt. Oznacza to, że model poprawnie wydobyl sygnał z danych, a pozostałe błędy mają charakter białego szumu.

W celu rozszerzenia analizy poza automatyczne algorytmy, przyjrzymy się klasycznym metodom wygładzania wykładniczego: metodzie Holta oraz metodzie Holta-Wintersa w wariancie addytywnym. Pozwala to na bezpośrednią kontrolę nad strukturą trendu i sezonowości, niezależnie od statystycznej optymalizacji stosowanej w funkcji `ets()`.

Pomijamy wariant multiplikatywny metody Holta-Wintersa, ponieważ transformacja Boxa-Coxa ustabilizowała wariancję szeregu, więc sezonowość ma charakter addytywny.

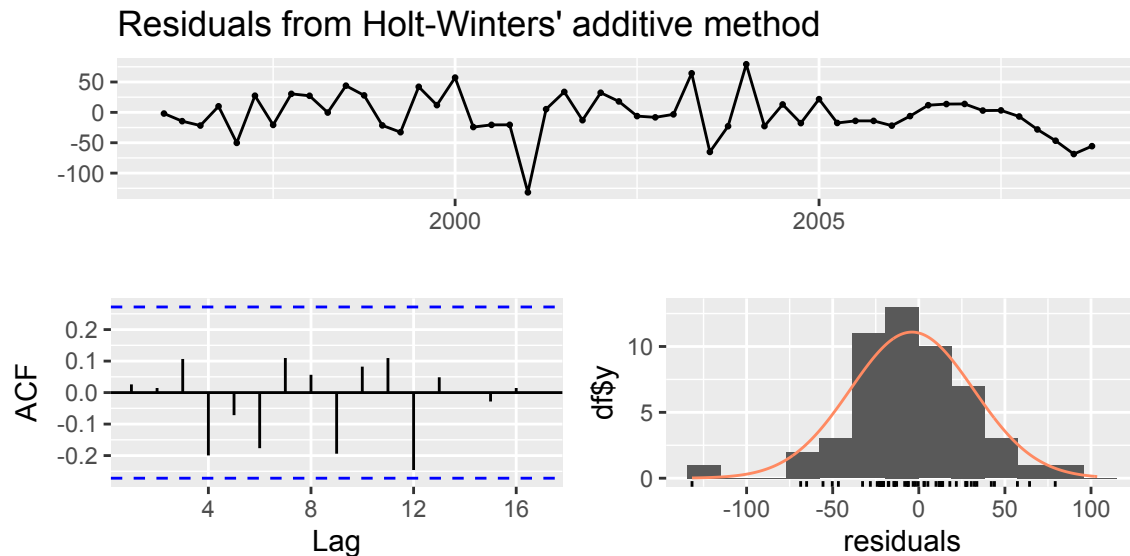
Tabela 15: Błędy dopasowania klasycznych metod wygładzania (zbiór uczący)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Metoda Holta (Trend)	-0.046	0.498	0.372	-0.046	0.387	0.339	0.060
Metoda Holta-Wintersa (Trend+Sezon)	-0.039	0.372	0.278	-0.039	0.291	0.254	0.013

Jak widzimy w tabeli 15, pominięcie komponentu sezonowego w metodzie Holta skutkuje znacznym wzrostem błędów dopasowania (RMSE wzrasta z 0.37 do 0.50). Potwierdza to wcześniejsze przypuszczenia o silnej sezonowości szeregu *eurotail* i zasadności stosowania wyłącznie metod sezonowych w dalszej części prognozowania.

Ponadto analiza miar dokładności wykazała, że modele `ETS(A,A,A)` oraz `Holt-Winters` generują identyczne wyniki. Świadczy to o ich matematycznej równoważności w kontekście analizowanego szeregu *eurotail*. Spodziewamy się zatem, że obie metody wygenerują identyczne prognozy punktowe oraz przedziały ufności, co ostatecznie potwierdzi, że model `ETS` jest statystycznym uogólnieniem klasycznego algorytmu `Holta-Wintersa`.

Spojrzymy jeszcze na wykres reszt dla metody `Holta-Wintersa`.



Wykres 19: Analiza reszt i opóźnień w modelu `Holta-Wintersa`

```
##
##  Ljung-Box test
##
## data:  Residuals from Holt-Winters' additive method
## Q* = 6.1939, df = 8, p-value = 0.6255
##
```

```
## Model df: 0.    Total lags used: 8
```

Widzimy, że wykresy reszt dla Holta-Wintersa są identyczne do wykresów reszt dla modelu ETS(A,A,A).

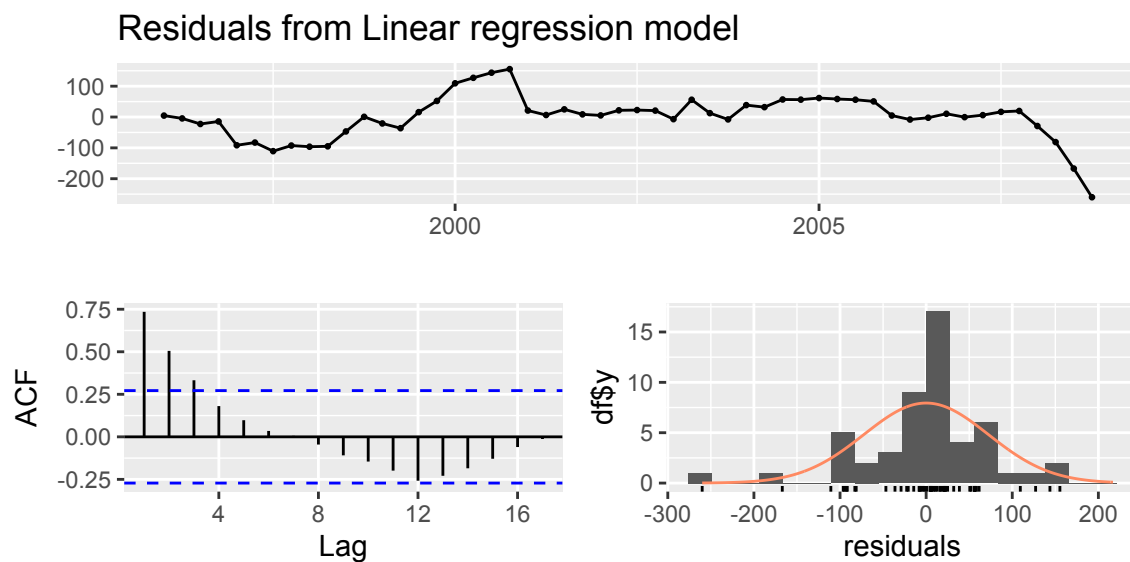
2.3.3 TSLM

W tej części zbudujemy model TSLM, które traktuje szereg czasowy jako sumę trendu i sezonowości. Jest to klasyczne podejście do dekompozycji szeregu.

```
fit_tslm2 <- tslm(train.retail.bc ~ trend + season)
fc_tslm2.bc <- forecast::forecast(fit_tslm2, h = 12)

fc_tslm2 <- fc_tslm2.bc
fc_tslm2$mean <- InvBoxCox(fc_tslm2.bc$mean, lambda = lambda.bc)
fc_tslm2$lower <- InvBoxCox(fc_tslm2.bc$lower, lambda = lambda.bc)
fc_tslm2$upper <- InvBoxCox(fc_tslm2.bc$upper, lambda = lambda.bc)
fc_tslm2$x <- train.retail
```

Model TSLM został dopasowany do danych przetransformowanych parametrem λ . Uwzględnienie komponentu `season` pozwala na odwzorowanie kwartalnych wahań indeksu, których nie byłby w stanie wychwycić prosty trend liniowy. Prognozy zostały następnie sprowadzone do oryginalnych jednostek za pomocą funkcji `InvBoxCox`.



Wykres 20: Analiza reszt i opóźnień w modelu TSLM

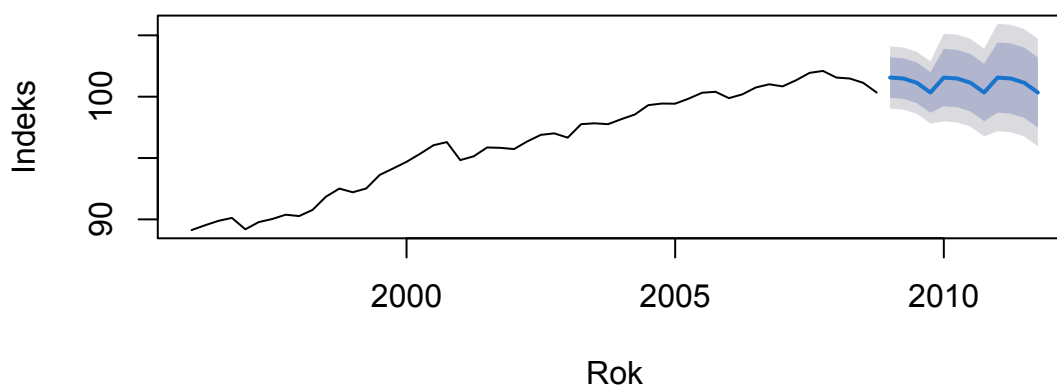
```
##
## Breusch-Godfrey test for serial correlation of order up to 8
##
## data: Residuals from Linear regression model
## LM test = 42.075, df = 8, p-value = 1.311e-06
```

Analiza reszt modelu TSLM wykazała bardzo niską wartość p-value ($p < 0.001$), co pozwala na odrzucenie hipotezy o braku autokorelacji. Wykres ACF wyraźnie pokazuje istotne statystycznie korelacje dla niskich opóźnień. Oznacza to, że model TSLM nie zdołał wydobyć wszystkich informacji z szeregu czasowego, co tłumaczy jego niską precyzję prognostyczną na tle modeli klasy ETS i SARIMA. Model ten zbyt upraszcza dynamikę indeksu euretail, traktując go jedynie jako sztywną kombinację trendu i sezonowości.

2.4 Wyznaczenie prognoz dla zbioru testowego na podstawie dopasowanych modeli oraz wybranej metody referencyjnej

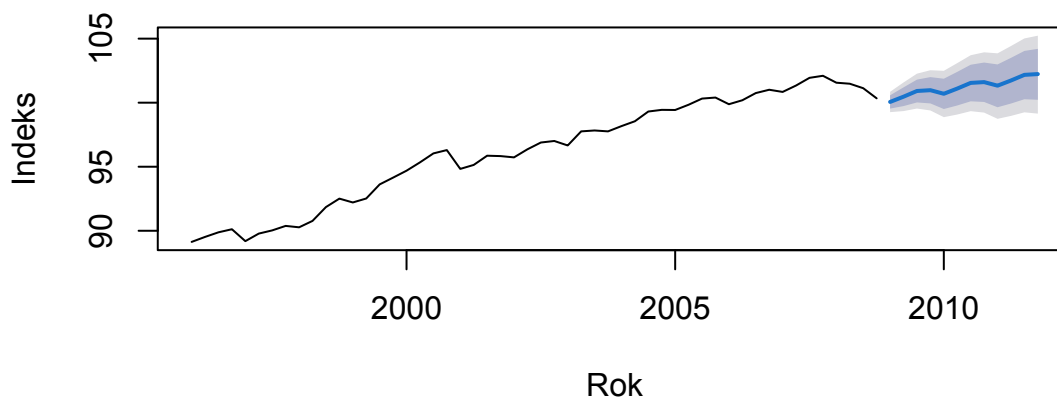
W niniejszym punkcie zajmiemy się przedstawieniem prognoz dla wybranych przez nas modeli i jednej metody naiwnej. W przypadku danych z silną sezonowością, najrozsądniejszym wyborem jest metoda sezonowa naiwna (*snaive*). Zakłada ona, że prognoza na dany kwartał będzie identyczna jak wartość w tym samym kwartale poprzedniego roku.

Prognoza SNAIVE (Benchmark)



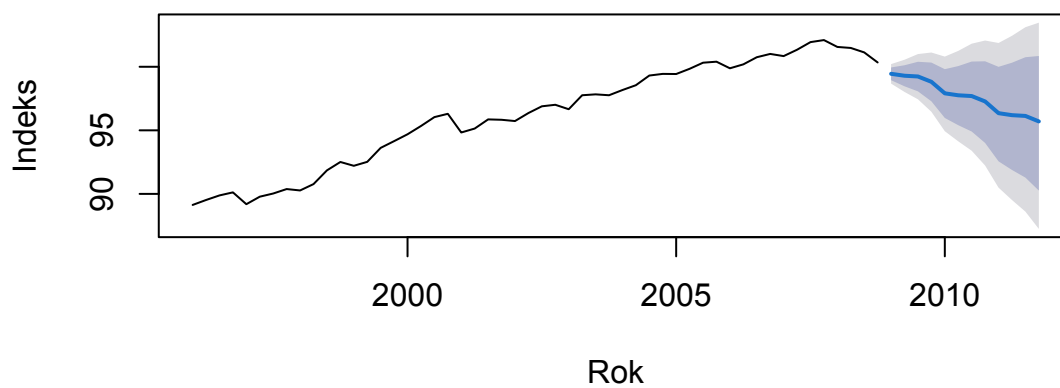
Wykres 21: Wykres prognozy naiwnej z sezonowością SNAIVE

Prognoza SARIMA



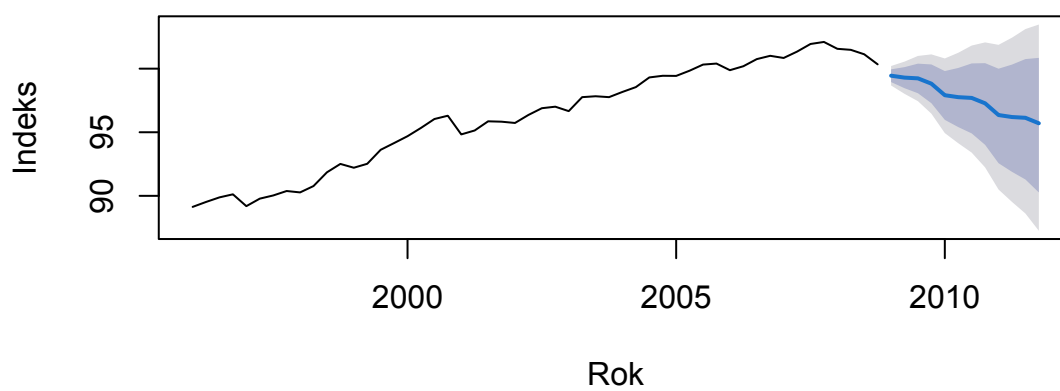
Wykres 22: Wykres prognozy modelu SARIMA(0,1,0)(0,1,1)[4]

Prognoza ETS



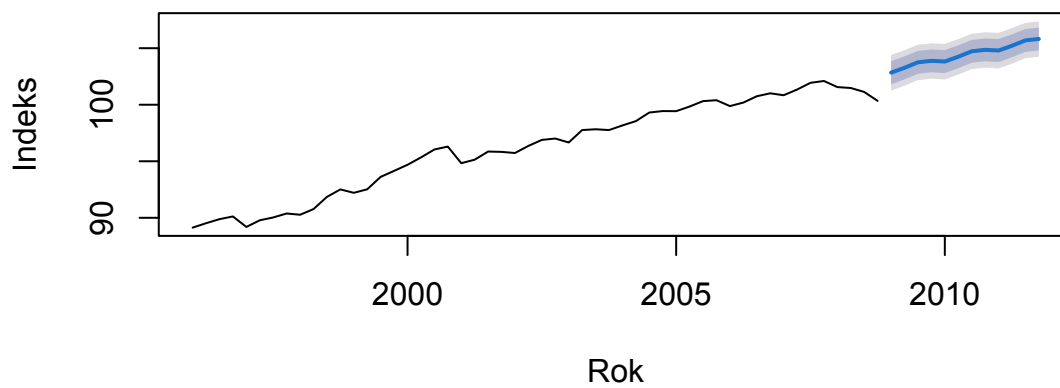
Wykres 23: Wykres prognozy modelu ETS(A,A,A)

Prognoza Holt-Winters



Wykres 24: Wykres prognozy modelu H-W

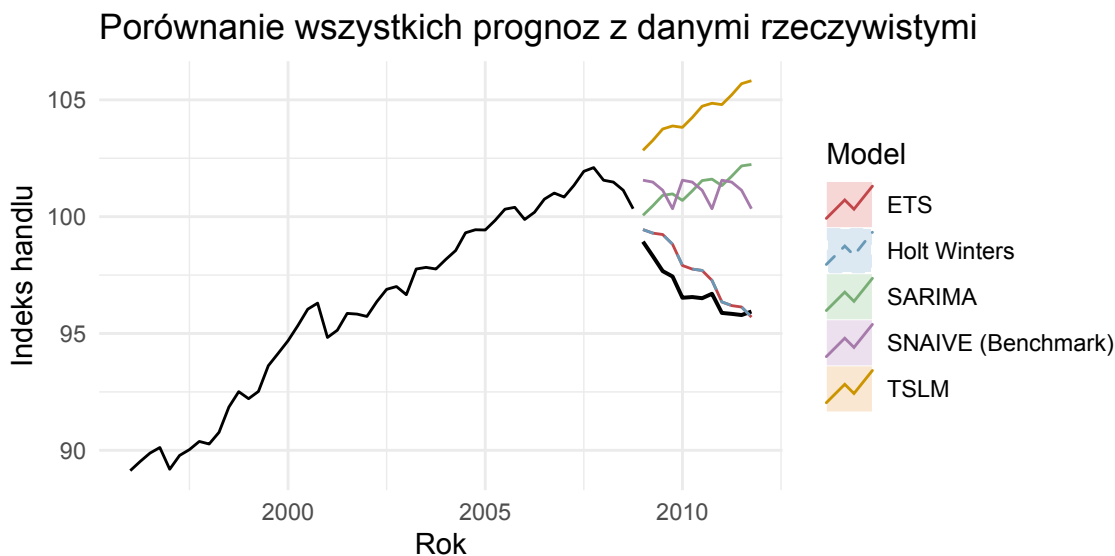
Prognoza TSLM



Wykres 25: Wykres prognozy modelu TSLM

2.5 Przedstawienie skonstruowanych prognoz na wykresie i porównanie z wartościami rzeczywistymi

Spójrzmy jak wygenerowane prognozy sprawdzają się z rzeczywistymi danymi.



Wykres 26: Porównanie prognoz do wartości rzeczywistych

Mimo przeprowadzenia rygorystycznej procedury doboru parametrów i pozytywnej diagnostyki reszt, model SARIMA okazał się mniej odporny na zmianę strukturalną trendu (kryzys finansowy) niż model ETS i równoważny mu Holt-Winters. Wynika to z charakterystyki wygładzania wykładniczego, które szybciej adaptuje się do lokalnych zmian poziomu szeregu. Co więcej nawet metoda naiwna zdaje się lepiej prognozować spadek niż SARIMA. Podejściem, które poradziło sobie zdecydowanie najgorzej, jest model TSLM – ze względu na swoją sztywną strukturę liniową nie był on w stanie wychwycić gwałtownego załamania trendu wzrostowego.

2.6 Porównanie dokładności prognoz dla zbioru testowego i uczącego

Przyjrzyjmy się miarom dokładności prognoz dla zbioru uczącego i testowego.

Tabela 16: Błędy dopasowania na zbiorze uczącym

	RMSE	MAE	MAPE	MASE
SARIMA	0.39	0.27	0.28	0.24
ETS	0.37	0.28	0.29	0.25
Holt-Winters	0.37	0.28	0.29	0.25
TSLM	0.74	0.51	0.53	0.47
SNAIVE	1.29	1.10	1.14	1.00

Tabela 17: Dokładność prognoz na zbiorze testowym ($h=12$)

	RMSE	MAE	MAPE	MASE
SARIMA	4.66	4.39	4.55	4.00
ETS	0.97	0.85	0.87	0.77

Holt-Winters	0.97	0.85	0.88	0.77
TSLM	7.79	7.57	7.83	6.89
SNAIVE	4.41	4.29	4.44	3.90

2.7 Wnioski

Na podstawie przeprowadzonej analizy oraz uzyskanych miar dokładności prognoz dla zbioru uczącego i testowego możemy wyciągnąć następujące wnioski:

- Zwycięzca analizy: Za optymalny model dla szeregu euretail uznajemy $ETS(A,A,A)$, który jest statystycznym odpowiednikiem addytywnej metody Holta-Wintersa. Model ten uzyskał najniższe błędy na zbiorze testowym ($MAPE = 0,88\%$, $RMSE = 0,97$), co oznacza, że średnio mylił się o mniej niż 1% wartości indeksu. Model ten wykazuje się największą stabilnością – różnica między błędami na zbiorze uczącym ($MAPE = 0,29\%$) a testowym ($MAPE = 0,87-0,88\%$) jest relatywnie niewielka. Oznacza to, że model nie uległ przeuczeniu i zachował wysoką zdolność generalizacji.
- Adaptacyjność ponad złożoność i sztywność: Modele wygładzania wykładniczego (ETS) okazały się znacznie lepsze od modeli regresyjnych (TSLM) i autoregresyjnych (SARIMA). Wynika to z faktu, że algorytm ETS nadaje większe wagi nowszym obserwacjom, przez co szybciej “zauważył” załamanie trendu wzrostowego po 2008 roku i dostosował do niego prognozę.
- Problemy modelu SARIMA: Mimo poprawnej specyfikacji i przejścia testów diagnostycznych reszt, model $SARIMA(0,1,0)(0,1,1)[4]$ zbyt mocno zasugerował się historycznymi wzrostami. W efekcie jego prognoza była zbyt optymistyczna względem rzeczywistych danych, co skutkowało dużym błędem $MAPE$ na poziomie 4,55%. Analiza porównawcza błędów również ujawnia istotny problem. Choć na zbiorze uczącym uzyskał on bardzo dobre wyniki ($MAPE = 0,28\%$, poziom zbliżony do ETS), to na zbiorze testowym błąd ten drastycznie wzrósł do poziomu 4,55%. Sugeruje to, że model zbyt sztywno dopasował się do historycznego trendu wzrostowego (overfitting).
- Nieadekwatność modelu TSLM: Model regresji liniowej okazał się najmniej użyteczny. Już na etapie diagnostyki test Breuscha-Godfrey wykazał silną autokorelację reszt ($p < 0,001$), co oznacza, że model nie wyciągnął wszystkich informacji z danych. Sztywne założenie o liniowym trendzie doprowadziło do całkowitego rozminięcia się prognozy z rzeczywistością ($MAPE = 7,83\%$).
- Rola benchmarku: Warto zauważyć, że w warunkach gwałtownej zmiany trendu, prosta metoda sezonowa naiwna (SNAIVE) poradziła sobie lepiej niż zaawansowany TSLM czy SARIMA. Potwierdza to, że przy dużej niestabilności szeregu, proste metody bazujące na ostatnim znanym cyklu sezonowym mogą być bezpieczniejszym punktem odniesienia. Jednakże nie oznacza to, że prognozy metody naiwnej były skuteczne.

Podsumowanie: Dla rozważanego szeregu czasowego najbardziej adekwatne jest podejście oparte na wygładzaniu wykładniczym (ETS). Pozwala ono na zachowanie balansu między stabilną sezonowością a elastycznym reagowaniem na nagłe zmiany poziomu indeksu, co w przypadku danych gospodarczych o dużej zmienności jest kluczowe. Model ten jako jedyny skutecznie połączył bardzo dobre dopasowanie do danych historycznych z realną zdolnością prognozowania w warunkach zmiany trendu, unikając przy tym negatywnych skutków przeuczenia.