

Dopasowanie modelu ARIMA i porównanie dokładności prognoz

Stanisław Olek

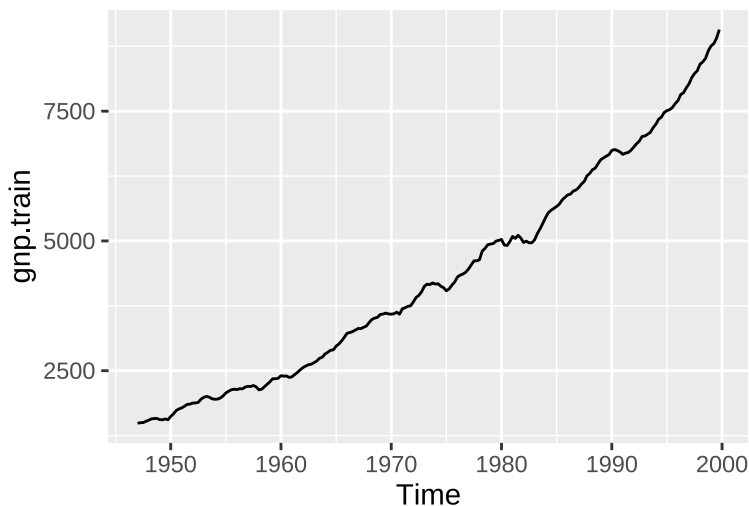
Spis treści

1	Dopasowanie modelu ARIMA do szeregu gnp	1
2	Porównanie dokładności prognoz dla danych euretail	12
2.1	Dopasowanie odpowiednich modeli do danych	12
2.2	Wyznaczenie, przedstawienie na wykresach i porównanie dokładności prognoz dla zbioru testowego	16
2.3	Wnioski	20

1 Dopasowanie modelu ARIMA do szeregu gnp

W tym rozdziale spróbujemy dopasować do danych `gnp` z biblioteki `astsa` odpowiedni model ARIMA, aby jak najlepiej przewidzieć zachowanie szeregu w czasie. Dane te dotyczą dochodu narodowego USA od 1. kwartału 1947 roku do 3. kwartału roku 2002. Najpierw spróbujemy zidentyfikować niezbędne transformacje na tym szeregu, aby przekształcić go do szeregu stacjonarnego.

Przed zastosowaniem niezbędnych transformacji podzielimy zbiór na uczący i testowy. Uczący zbiór zawiera dane do 4. kwartału 1999 roku, a testowy - pozostałe do 3. kwartału 2002 roku. Zobaczmy, jak prezentuje się szereg dla danych treningowych.

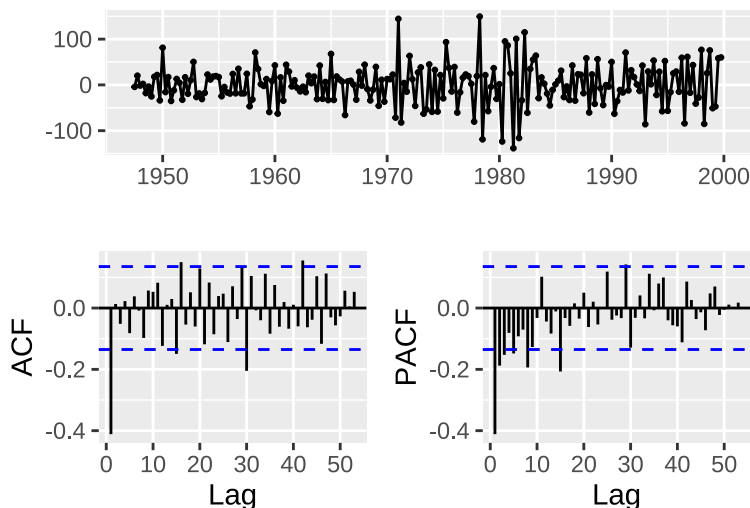


Rysunek 1: Wykres danych `gnp`

Na rysunku 1 widzimy występujący w danych trend, który przypomina trend kwadratowy. Nie widzimy natomiast sezonowości w danych. Stąd możemy przeprowadzić podwójne różnicowanie z opóźnieniem 1, aby ten trend usunąć. To samo mówi nam funkcja `ndiffs`, wg której powinniśmy zastosować 2 różnicowania z opóźnieniem 1, natomiast zgodnie z naszymi przewidywaniami i z funkcją `nsdiffs` nie powinniśmy stosować różnicowania z opóźnieniem sezonowym. Sprawdzimy jeszcze drugie podejście - zastosowanie transformacji

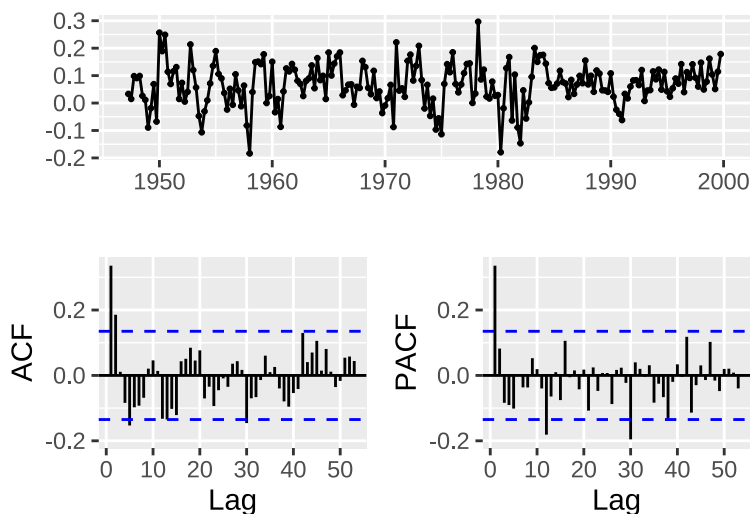
Boxa-Coxa (B-C), a następnie zróżnicowanie z opóźnieniem 1 (zasugerowane przez funkcję `ndiffs`). Jako parametr λ przy transformacji B-C przyjmujemy λ zaproponowane przez funkcję `BoxCox.lambda` metodą największej wiarygodności. Zgodnie z nią λ wynosi 0.25. Ponadto, dla modeli z transformacją B-C oraz różnicowaniem dodamy dryf, ponieważ średnia tego szeregu po transformacji B-C i zróżnicowaniu jest różna od 0 (dokładniej wynosi 0.067).

Najpierw sprawdzimy, jakie p i q dobrać do modeli AR i MA dla obu zastosowanych przekształceń. Za pomocą rysunku 2 zidentyfikujemy modele dla danych po podwójnym zróżnicowaniu.



Rysunek 2: Wykres danych oraz ACF i PACF danych po podwójnym zróżnicowaniu

Szereg możemy nazwać stacjonarnym, gdyż nie widzimy w nim pozostałości po trendzie, a ACF nie zanika ani nie jest okresowa. Na podstawie ACF możemy rozważyć modele MA(30) oraz MA(42) (ostatni raz ACF odstaje poza przedziały ufności dla $h = 42$, jednak jest to mało wyraźne w porównaniu z $h = 30$). Na podstawie PACF możemy rozważyć modele AR(15) oraz AR(29) (ostatni raz PACF odstaje poza przedziały ufności dla $h = 29$, jednak jest to mało wyraźne w porównaniu z $h = 15$). Kolejnym krokiem będzie zidentyfikowanie modeli dla szeregu po transformacji B-C oraz różnicowaniu.



Rysunek 3: Wykres danych oraz ACF i PACF danych po transformacji B-C i zróżnicowaniu

Szereg na rysunku 3 możemy nazwać stacjonarnym, gdyż nie widzimy w nim pozostałości po trendzie, a ACF nie zanika ani nie jest okresowa. Na podstawie funkcji ACF możemy rozważyć modele MA(2) oraz MA(30)

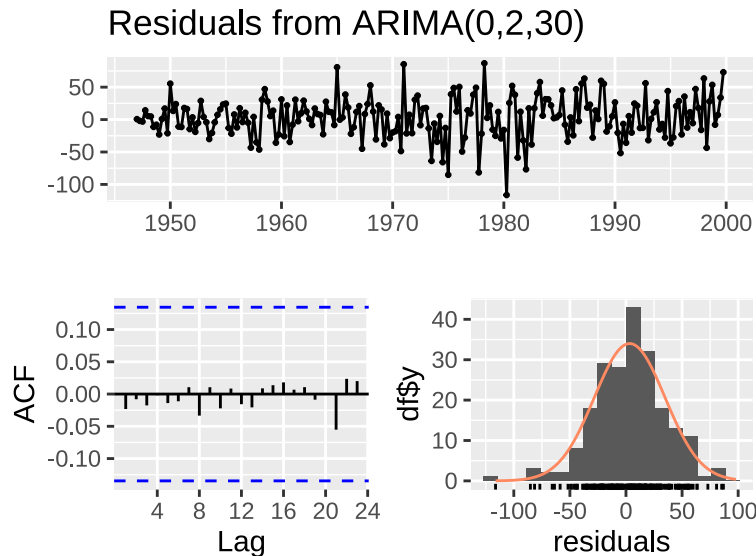
(dla $h = 30$ ACF odstaje po raz ostatni poza przedziały ufności, natomiast naprawdę wyraźne odstawanie widzimy ostatni raz dla $h = 2$). Na podstawie funkcji PACF możemy rozważyć model AR(30) (nie rozważamy dodatkowych modeli, gdyż dla $h = 30$ odstawanie poza przedziały ufności jest wyraźne).

Musimy dopasować modele także na podstawie kryteriów informacyjnych AIC, AICC oraz BIC. Zrobimy to dla szeregu bez i z transformacją B-C za pomocą funkcji `auto.arima`.

Zgodnie ze wszystkimi kryteriami informacyjnymi dla szeregu z transformacją B-C wybieramy model $\text{ARIMA}(1, 1, 0)(0, 0, 0)_4$, natomiast na podstawie kryteriów AIC i AICC dla szeregu bez transformacji B-C wybieramy model $\text{ARIMA}(0, 2, 3)(0, 0, 0)_4$, a zgodnie z kryterium BIC dla tego szeregu wybieramy model $\text{ARIMA}(1, 2, 1)(0, 0, 0)_4$.

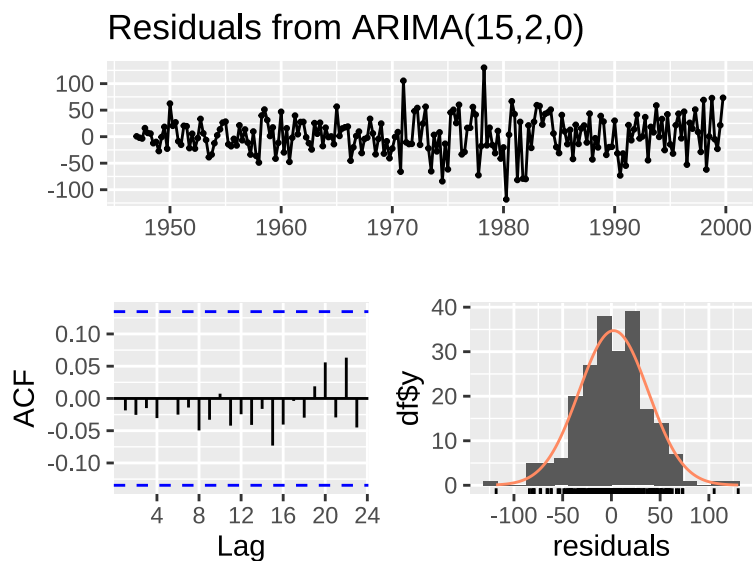
W związku z tym rozważanymi modelami będą: $\text{ARIMA}(0, 2, 30)$, $\text{ARIMA}(0, 2, 42)$, $\text{ARIMA}(15, 2, 0)$, $\text{ARIMA}(29, 2, 0)$, $\text{ARIMA}(0, 1, 2)$ z transformacją B-C, $\text{ARIMA}(0, 1, 30)$ z transformacją B-C, $\text{ARIMA}(30, 1, 0)$ z transformacją B-C (dla tych modeli nie uwzględniamy czynnika sezonowego - stąd skrócony zapis), $\text{ARIMA}(1, 1, 0)(0, 0, 0)_4$, $\text{ARIMA}(0, 2, 3)(0, 0, 0)_4$, $\text{ARIMA}(1, 2, 1)(0, 0, 0)_4$.

Należy sprawdzić, czy modele te mają losowe reszty - inaczej będziemy sprawdzać, czy reszty w tych modelach są białym szumem. Dodatkowo, sprawdzimy, w których modelach reszty pochodzą z rozkładu normalnego. Losowość reszt sprawdzimy za pomocą testu Ljung-Boxa (L-B) z funkcji `checkresiduals`, a normalność reszt za pomocą testu Shapiro-Wilka (S-W) z funkcji `shapiro.test`. Przyjmujemy poziom istotności $\alpha = 0.05$.



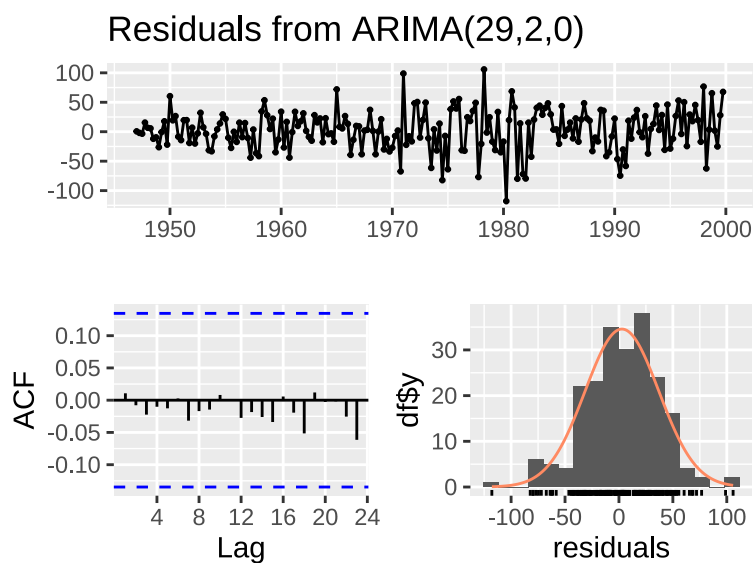
Rysunek 4: Analiza reszt modelu $\text{ARIMA}(0,2,30)$

W przypadku modelu $\text{ARIMA}(0, 2, 30)$ reszty są losowe i normalne, co pokazuje wykres ACF i histogram na rysunku 4 oraz co potwierdzają testy L-B i S-W.



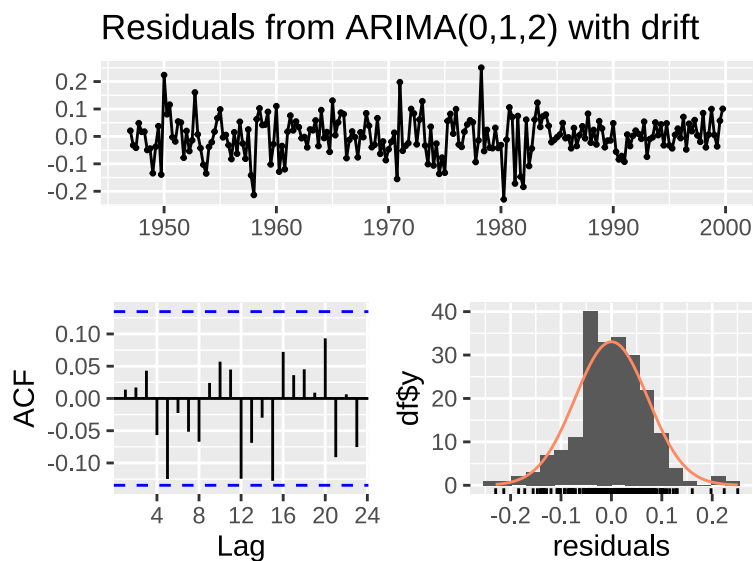
Rysunek 5: Analiza reszt modelu ARIMA(15, 2, 0)

W przypadku modelu ARIMA(15, 2, 0) reszty są losowe i normalne, co pokazuje wykres ACF i histogram na rysunku 5 oraz co potwierdzają testy L-B i S-W.



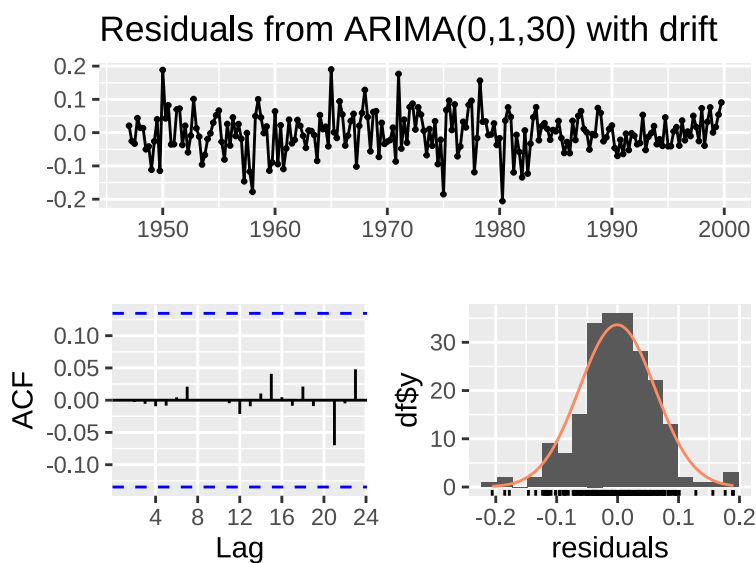
Rysunek 6: Analiza reszt modelu ARIMA(29, 2, 0)

W przypadku modelu ARIMA(29, 2, 0) reszty są losowe i normalne, co pokazuje wykres ACF i histogram na rysunku 6 oraz co potwierdzają testy L-B i S-W.



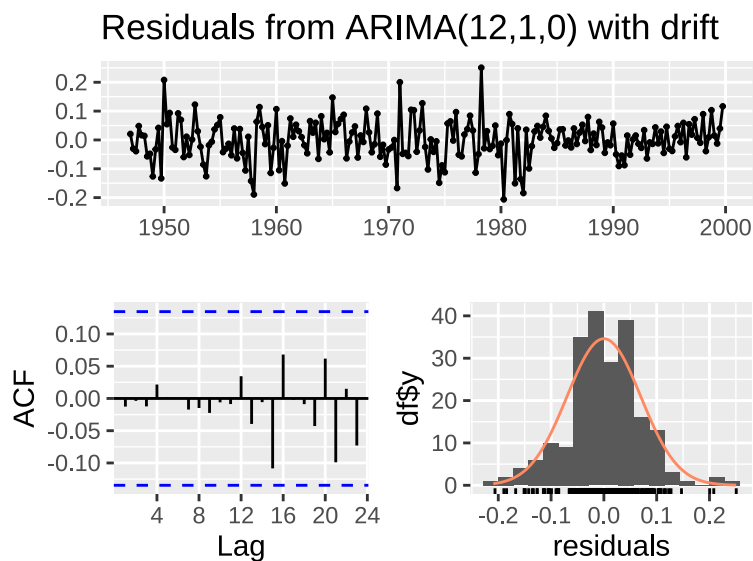
Rysunek 7: Analiza reszt modelu ARIMA(0, 1, 2)

W przypadku modelu ARIMA(0, 1, 2) z transformacją B-C reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 7 oraz co potwierdzają testy L-B i S-W.



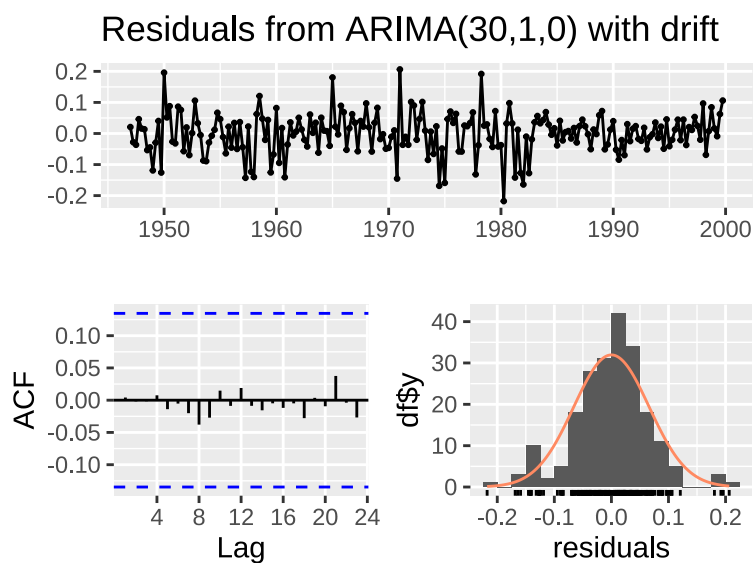
Rysunek 8: Analiza reszt modelu ARIMA(0, 1, 30)

W przypadku modelu ARIMA(0, 1, 30) z transformacją B-C reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 8 oraz co potwierdzają testy L-B i S-W.



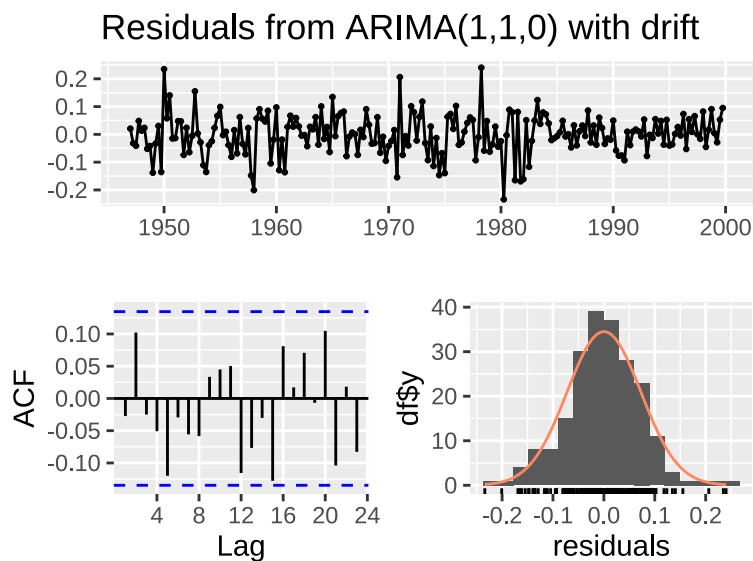
Rysunek 9: Analiza reszt modelu ARIMA(12, 1, 0)

W przypadku modelu ARIMA(12, 1, 0) z transformacją B-C reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 9 oraz co potwierdzają testy L-B i S-W.



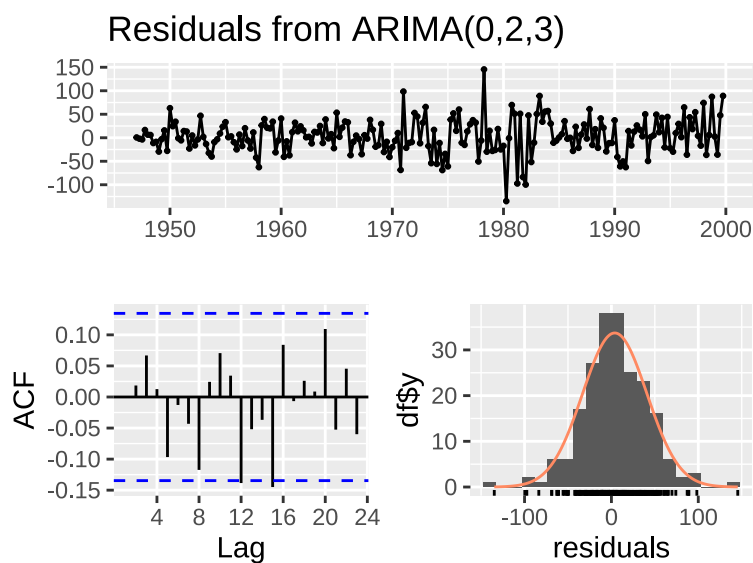
Rysunek 10: Analiza reszt modelu ARIMA(30, 1, 0)

W przypadku modelu ARIMA(30, 1, 0) z transformacją B-C reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 10 oraz co potwierdzają testy L-B i S-W.



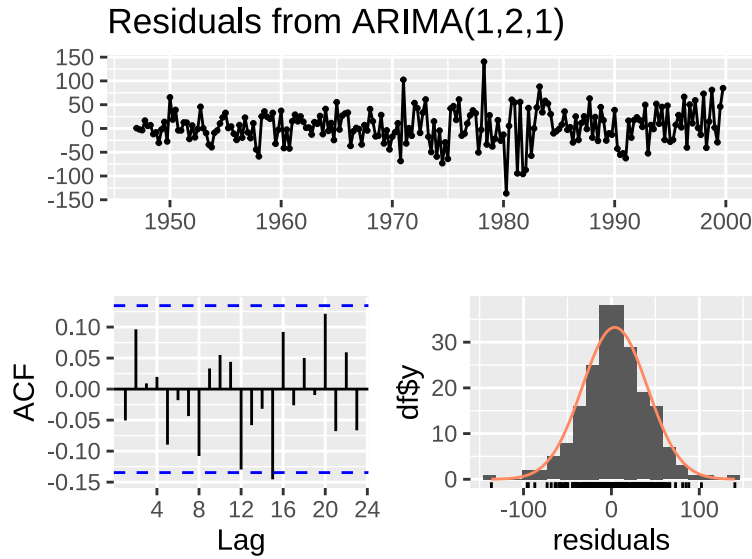
Rysunek 11: Analiza reszt modelu $ARIMA(1, 1, 0)(0, 0, 0)_4$

W przypadku modelu $ARIMA(1, 1, 0)(0, 0, 0)_4$ z transformacją B-C reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 11 oraz co potwierdzają testy L-B i S-W.



Rysunek 12: Analiza reszt modelu $ARIMA(0, 2, 3)(0, 0, 0)_4$

W przypadku modelu $ARIMA(0, 2, 3)(0, 0, 0)_4$ reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 12 oraz co potwierdzają testy L-B i S-W.



Rysunek 13: Analiza reszt modelu $ARIMA(1, 2, 1)(0, 0, 0)_4$

W przypadku modelu $ARIMA(1, 2, 1)(0, 0, 0)_4$ reszty są losowe, ale nie są normalne, co pokazuje wykres ACF i histogram na rysunku 13 oraz co potwierdzają testy L-B i S-W.

Uwaga! 1 Model $ARIMA(0, 2, 4)_2$ nie przeszedł testu białoszumowości reszt, dlatego nie uwzględniliśmy go tutaj i nie będziemy go już wcale uwzględniać.

Uwaga! 2 W dalszej analizie nie będziemy też uwzględniać modeli AR i MA uzyskanych przez podwójne różnicowanie, ponieważ dają o wiele gorsze prognozy niż modele po transformacji B-C i zróżnicowaniu.

Możemy zbadać jakość dopasowania modeli w oparciu o kryteria informacyjne. Musimy jednak pamiętać, że modele, w których nie uwzględniliśmy transformacji B-C będą miały inne wartości kryteriów niż modele, w których uwzględniliśmy tę transformację. W tabeli 1 zostały przedstawione wartości poszczególnych kryteriów informacyjnych dla analizowanych modeli.

Modele	AIC	AICC	BIC
ARIMA(0,1,2) z transf. B-C	-501.80	-501.61	-488.39
ARIMA(0,1,30) z transf. B-C	-492.68	-480.82	-385.42
ARIMA(12,1,0) z transf. B-C	-495.92	-493.77	-448.99
ARIMA(30,1,0) z transf. B-C	-480.65	-468.78	-373.39
ARIMA(1, 1, 0)(0, 0, 0) ₄ z transf. B-C	-501.89	-501.77	-491.83
ARIMA(0, 2, 3)(0, 0, 0) ₄ bez transf. B-C	2125.68	2125.87	2139.06
ARIMA(1, 2, 1)(0, 0, 0) ₄ bez transf. B-C	2124.73	2124.85	2134.77

Tabela 1: Tabela porównująca wartości kryteriów informacyjnych modeli

Co ciekawe, model wybrany za pomocą kryteriów AIC i AICC bez transformacji B-C ma wyższe wartości tych kryteriów niż model dopasowany za pomocą BIC bez transformacji B-C. Wynika to z tego, że funkcja `auto.arima` ma domyślnie ustawiony argument `stepwise=TRUE`, przez co pomija przy wyborze niektóre modele ARIMA, które mogą mieć niższe wartości AIC lub AICC. W tym przypadku tak jest, więc model $ARIMA(0, 2, 3)(0, 0, 0)_4$ bez transformacji B-C pominiemy z dalszych rozważań. Tak samo pominiemy model $ARIMA(30, 1, 0)$ z transformacją B-C, ponieważ ma on najwyższe wartości wszystkich kryteriów informacyjnych spośród wszystkich modeli, u których została zastosowana transformacja B-C.

Po porównaniu kryteriów informacyjnych zobaczymy, które współczynniki w pozostałych modelach są istotne. Po sprawdzeniu istotności współczynników możemy sprawdzić, czy po usunięciu nieistotnych współczynników

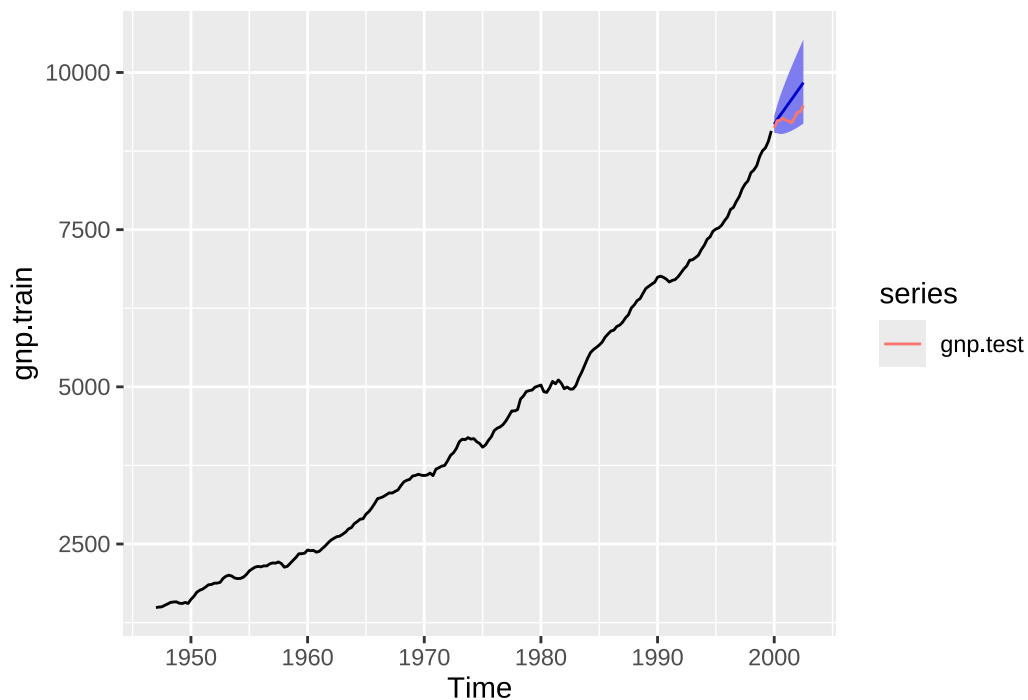
reszty dalej są losowe i mają/nie mają rozkładu normalnego oraz czy wartości kryteriów informacyjnych nie zmieniły się istotnie.

- Istotne współczynniki w modelu ARIMA(0, 1, 2) z transf. B-C: 1, 2, 3 (dryf też jest istotny). Nie usunęliśmy żadnego współczynnika z modelu, więc jest to dalej ten sam model.
- Istotne współczynniki w modelu ARIMA(0, 1, 30) z transf. B-C: 1, 2, 5, 7, 8, 15, 23, 26, 29, 31 (dryf też jest istotny). Reszty w tym modelu nie są już losowe i dalej nie mają rozkładu normalnego. Wartości kryteriów AIC, AICC i BIC wynoszą odpowiednio -518.9496543, -517.6230211, -482.0792148, a więc mniej niż w oryginalnym modelu.
- Istotne współczynniki w modelu ARIMA(12, 1, 0) z transf. B-C: 1, 12, 13 (dryf też jest istotny). Reszty w tym modelu nie są już losowe i dalej nie mają rozkładu normalnego. Wartości kryteriów AIC, AICC i BIC wynoszą odpowiednio -504.7353894, -504.5412147, -491.3279569, a więc mniej niż w oryginalnym modelu.
- Istotne współczynniki w modelu ARIMA(1, 1, 0)(0, 0, 0)₄ z transf. B-C: 1, 2 (dryf też jest istotny). Nie usunęliśmy żadnego współczynnika z modelu, więc jest to dalej ten sam model.
- Istotne współczynniki w modelu ARIMA(1, 2, 1)(0, 0, 0)₄ z transf. B-C: 1, 2. Nie usunęliśmy żadnego współczynnika z modelu, więc jest to dalej ten sam model.

Widzimy, że modele z małą ilością współczynników pozostały bez zmian, natomiast te, które miały więcej współczynników nie przeszły testów losowości, przez co nie będziemy ich uwzględniać w dalszej analizie.

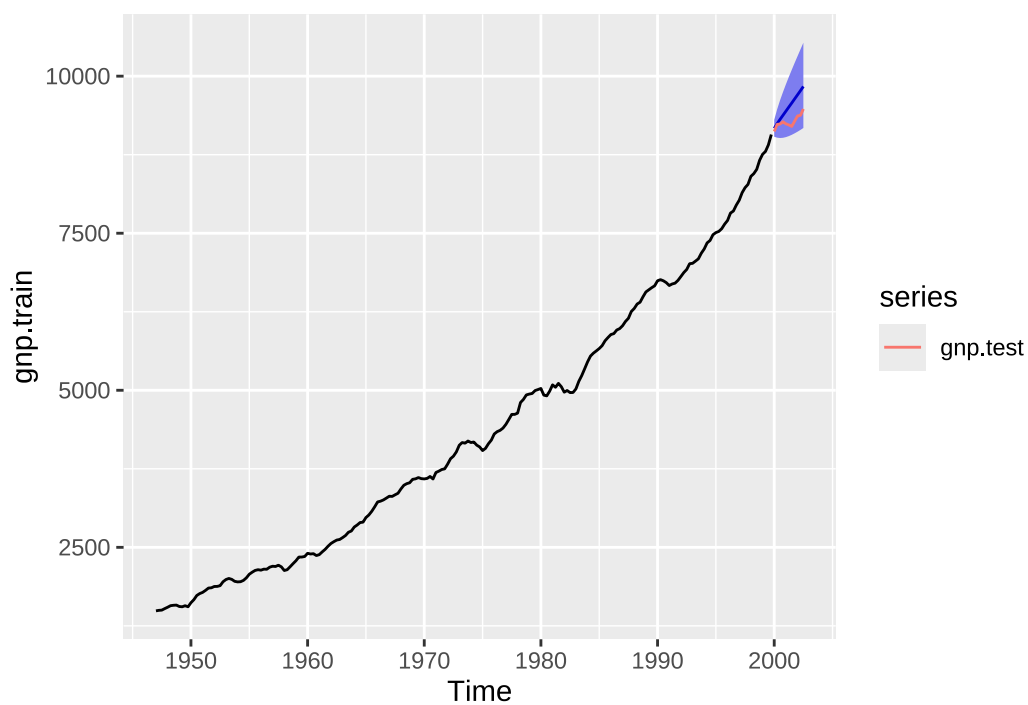
Pozostały modele ARIMA(0, 1, 2), ARIMA(1, 1, 0)(0, 0, 0)₄ oraz ARIMA(1, 2, 1)(0, 0, 0)₄. Możemy stwierdzić, że model ARIMA(1, 1, 0)(0, 0, 0)₄ jest lepszy niż ARIMA(0, 1, 2), gdyż ma mniejsze wartości kryteriów informacyjnych oraz ma mniej współczynników. Ciężko natomiast porównać model ARIMA(1, 2, 1)(0, 0, 0)₄ z pozostałymi modelami, gdyż w tym modelu nie zastosowaliśmy transformacji B-C, przez co skala kryteriów informacyjnych oraz np. wariancji jest zupełnie inna. Czymś, co odróżnia ten model od pozostałych jest normalność reszt. Pod tym względem może być lepszy od dwóch pozostałych modeli. Aby to rozstrzygnąć, sprawdzimy prognozy na zbiorze testowym oraz błędy predykcji.

Na rysunku 14 pokazana jest prognoza na zbiorze testowym modelu ARIMA(0, 1, 2) z zaznaczonymi realizacjami przedziałów ufności na poziomie ufności 0.95.



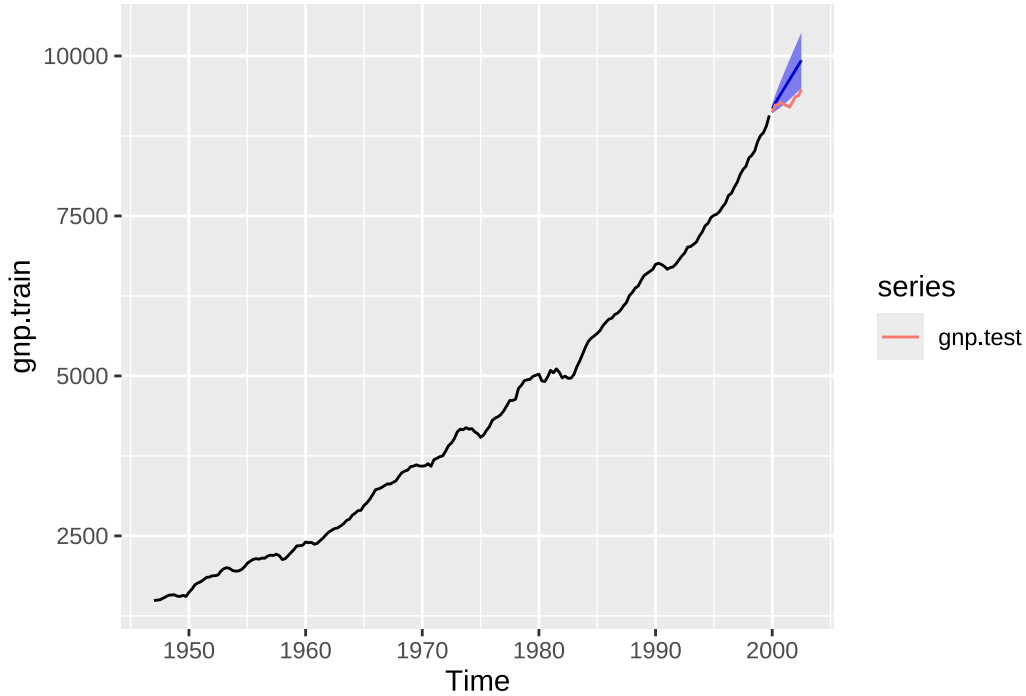
Rysunek 14: Prognoza modelu $ARIMA(0, 1, 2)$ vs faktyczne dane

Widzimy, że model nie do końca dobrze przewidział zachowanie danych w późniejszym okresie. Predykcja modelu odbiega od danych, ale całe dane znajdują się w przedziałach ufności. Sprawdźmy, jak prezentuje się prognoza drugiego modelu uwzględniającego transformację B-C.



Rysunek 15: Prognoza modelu $ARIMA(1, 1, 0)(0, 0, 0)_4$ vs faktyczne dane

Na rysunku 15 widzimy bardzo podobną predykcję modelu $ARIMA(1, 1, 0)(0, 0, 0)_4$ do predykcji modelu $ARIMA(0, 1, 2)$. Modelowi $ARIMA(1, 1, 0)(0, 0, 0)_4$ nie udało się uchwycić całkowitej zmienności danych, ale wszystkie obserwacje ze zbioru testowego mieszczą się w przedziałach ufności.



Rysunek 16: Prognoza modelu $ARIMA(1, 2, 1)(0, 0, 0)_4$ vs faktyczne dane

Na rysunku 16 zauważymy, że model $ARIMA(1, 2, 1)(0, 0, 0)_4$ nie poradził sobie z prognozą danych. Dane nie mieszczą się w większości w przedziałach ufności i znacząco odbiegają od linii obrazującej oryginalne dane.

Możemy dokładniej porównać błędy popełnione przez wybrane modele za pomocą RMSE (Root Mean Square Error), MAE (Mean Absolute Error) oraz MASE (Mean Absolute Scaled Error). Wartości te otrzymamy dzięki funkcji `accuracy` z pakietu `forecast`. W tabeli 2 zostały przedstawione poszczególne błędy dla każdego modelu.

Modele	RMSE	MAE	MASE
$ARIMA(0,1,2)$ z transf. B-C	272.09	234.15	1.51
$ARIMA(1,1,0)$ z transf. B-C	269.62	230.60	1.48
$ARIMA(1, 2, 1)(0, 0, 0)_4$ bez transf. B-C	327.80	282.05	1.81

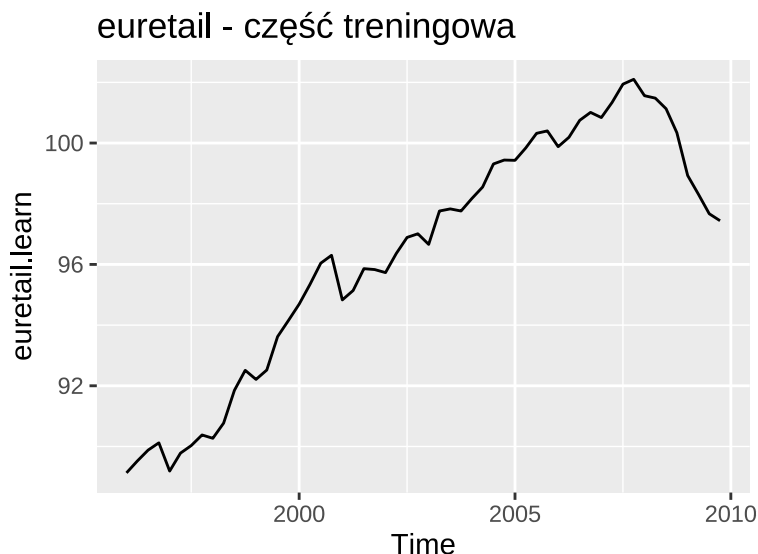
Tabela 2: Tabela błędów predykcji dla uzyskanych modeli

Widzimy, że największy błąd predykcji ma w każdym przypadku model $ARIMA(1, 2, 1)(0, 0, 0)_4$. Najmniejszy błąd dla każdej z metod ma natomiast model $ARIMA(1, 1, 0)(0, 0, 0)_4$. Jest to także model z najmniejszą ilością współczynników oraz najniższymi kryteriami informacyjnymi z tych trzech. W związku z tym możemy sformułować następujący wniosek.

Wniosek: 1 *Najlepszym modelem ARIMA do prognozy szeregu gnp jest model $ARIMA(1, 1, 0)(0, 0, 0)_4$ z wcześniejszym zastosowaniem transformacji Boxa-Coxa z parametrem $\lambda = 0.25$.*

2 Porównanie dokładności prognoz dla danych euretail

W tej sekcji porównamy dokładności prognoz dla danych **euretail**, które przedstawiają kwartalne wartości indeksu handlu detalicznego dla strefy euro w latach 1996-2011. Prognozy skonstruujemy na bazie modeli: ARIMA, dekompozycji oraz algorytmów wygładzania wykładniczego.



Rysunek 17: Wykres dla danych euretail - część treningowa

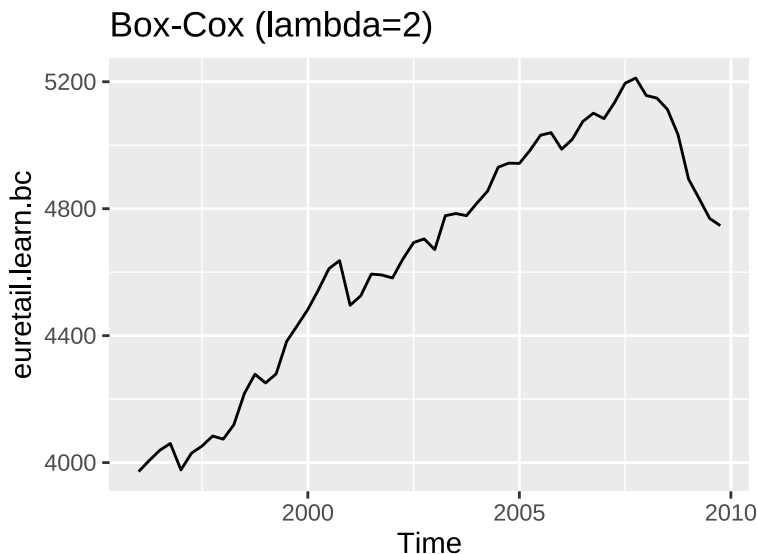
Na Rysunku 17 przedstawiono dane **euretail** po podziale na część treningową i testową. Część treningowa obejmuje dane do końca 2009 roku, natomiast część testowa to dane od początku 2010 roku.

2.1 Dopasowanie odpowiednich modeli do danych

W tej części zajmiemy się identyfikacją odpowiednich modeli do danych **euretail**. W tym celu rozważymy modele $AR(p)$, $MA(q)$, ARIMA, modele dekompozycji oraz modele oparte na algorytmach wygładzania wykładniczego.

2.1.1 Identyfikacja modeli AR(p), MA(q), ARIMA

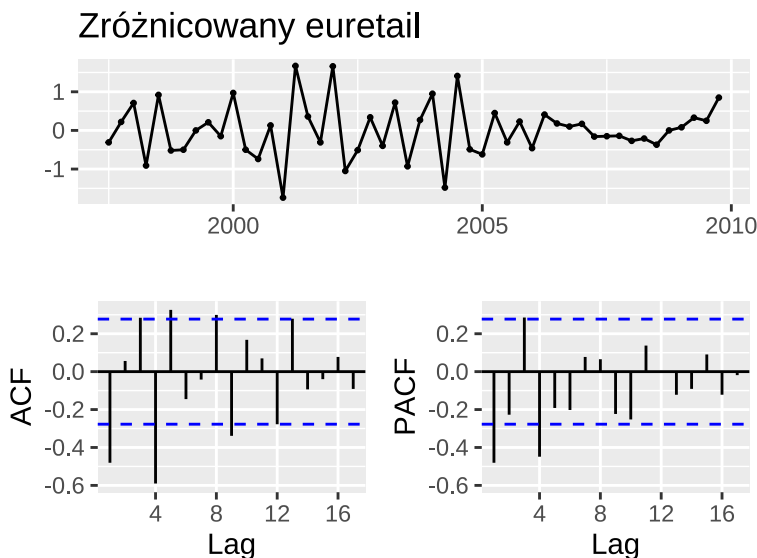
1. Transformacja Boxa-Coxa



Rysunek 18: Dane euretail po transformacji Boxa-Coxa z automatycznie dobranym parametrem $\lambda = 2$

Jak widać (Rysunek 18) dane po zastosowaniu transformacji Boxa-Coxa z parametrem $\lambda = 2$ zasadniczo nie zmieniły się, także w dalszej części analizy nie będziemy ich rozważać (inne wartości parametru λ również nie przynoszą poprawy).

2. Różnicowanie



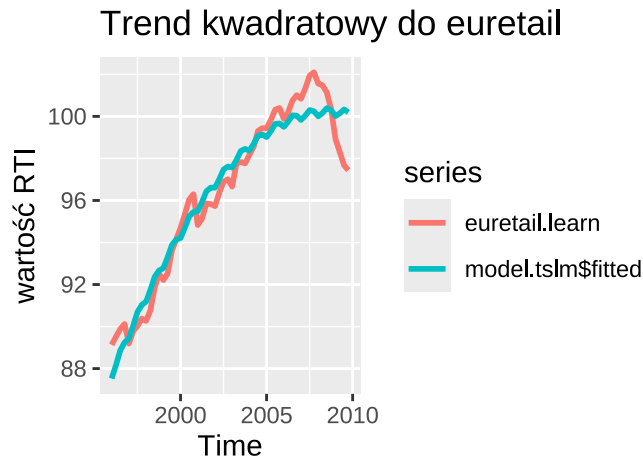
Rysunek 19: Dane euretail po jednokrotnym zróżnicowaniu z $\text{lag} = 4$ i dwukrotnym zróżnicowaniu z $\text{lag} = 1$

Jak można zauważyć (Rysunek 19) operacje: jednokrotnego różnicowania z $\text{lag} = 4$ (w celu usunięcia sezonowości) oraz dwukrotnego różnicowania z $\text{lag} = 1$ (w celu usunięcia trendu) zdaje się, że pozwoliły na uzyskanie szeregu stacjonarnego dla danych `euretail`. Wykresy ACF i PACF (Rysunek 19) sugerują, że

potencjalnymi rzędami modelu dla $AR(p)$ może być $p=4$, natomiast dla $MA(q)$: $q=4$ i $q=9$, także w naszej analizie rozważymy 4 modele:

- $ARIMA(4,2,0)(0,1,0)$
- $ARIMA(0,2,4)(0,1,0)$
- $ARIMA(0,2,9)(0,1,0)$
- $ARIMA(0,1,3)(0,1,1)$ (model automatyczny)

2.1.2 Identyfikacja modeli dekompozycji



Rysunek 20: Wykres euretail z dopasowanym trendem kwadratowym

Na rysunku 20 przedstawiono dopasowanie modelu dekompozycji z trendem kwadratowym do danych *euretail*, który wydaje się być dopasowany nie najgorzej do danych, także w dalszej części analizy będziemy go rozważać.

2.1.3 Identyfikacja modeli algorytmów wygładzania wykładniczego

Rozważymy 4 modele wygładzania wykładniczego:

- model prosty (SES)
- model Holta z tłumieniem
- model addytywny Holta-Wintersa
- model automatyczny ETS

2.1.4 Diagnostyka modeli

Diagnostykę modeli przeprowadzamy w oparciu o test Ljungu-Boxa (weryfikacja białoszumowości reszt), test Shapiro Wilka (weryfikacja normalności reszt) oraz istotność współczynników modelu. Formalnie odpowiednie hipotezy to:

- H_0 : Reszty modelu są białym szumem
- H_1 : Reszty modelu nie są białym szumem

oraz

- H_0 : Reszty modelu pochodzą z rozkładu normalnego
- H_1 : Reszty modelu nie pochodzą z rozkładu normalnego

Tabela 3: Wyniki testów diagnostycznych dla modeli ARIMA

Model	p-value (L-B)	decyzja (L-B)	p-value (S-W)	decyzja (S-W)	istotność współczynników
ARIMA(4,2,0)(0,1,0)	0.0103	-	0.7419	+	częściowo istotne
ARIMA(0,2,4)(0,1,0)	0.0450	-	0.2533	+	częściowo istotne
ARIMA(0,2,9)(0,1,0)	0.1555	+	0.8249	+	częściowo istotne
ARIMA(0,1,3)(0,1,1)	0.9450	+	0.6604	+	wszystkie istotne

Tabela 4: Wyniki testów diagnostycznych dla modelu dekompozycji tslm

Model	p-value (L-B)	decyzja (L-B)	p-value (S-W)	decyzja (S-W)	istotność współczynników
trend kwadr. + sezon	0.0000	-	0.0532	+	częściowo istotne

Tabela 5: Wyniki testów diagnostycznych dla modeli wygładzania wykładniczego

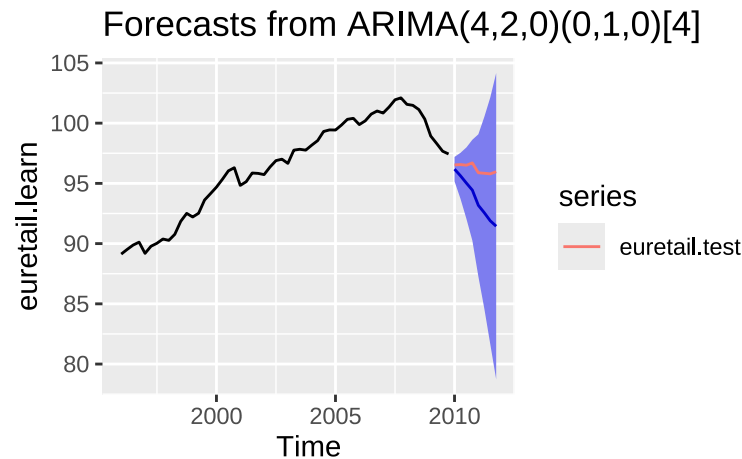
Model	p-value (L-B)	decyzja (L-B)	p-value (S-W)	decyzja (S-W)
SES	0.0004	-	0.0105	-
Holt (tłumiony)	0.0000	-	0.0091	-
Holt-Winters (addytywny)	0.5761	+	0.4161	+
ETS (automatyczny)	0.7380	+	0.2732	+

Wyniki przeprowadzonych testów diagnostycznych, przedstawione w tabelach 3, 4 oraz 5, pozwoliły na wyłonienie trzech faworytów:

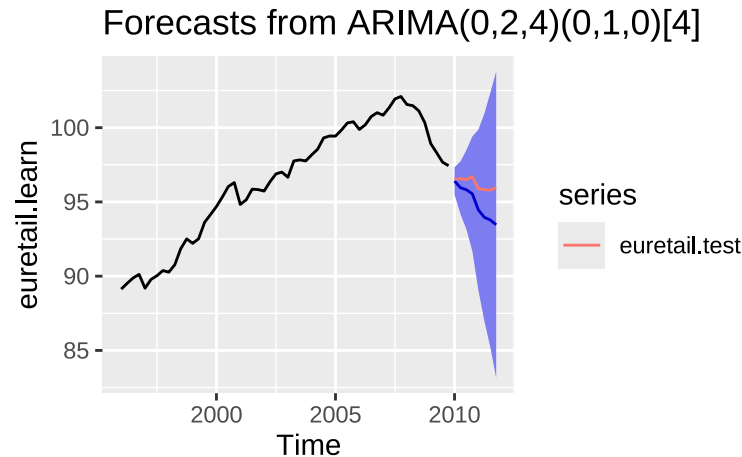
- **ARIMA(0,1,3)(0,1,1)**
- **addytywny model Holta-Wintersa**
- **automatyczny model ETS**

2.2 Wyznaczenie, przedstawienie na wykresach i porównanie dokładności prognoz dla zbioru testowego

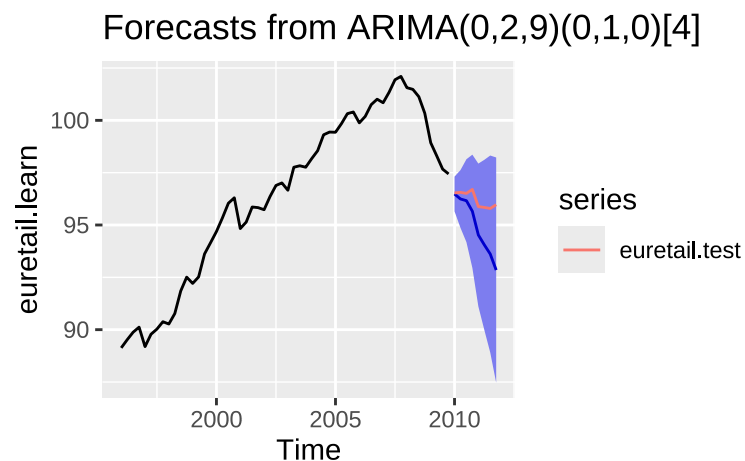
W tej części zajmiemy się wyznaczeniem prognoz dla zbioru testowego na podstawie dopasowanych modeli oraz sezonowej metodzie naiwnej (metoda referencyjna).



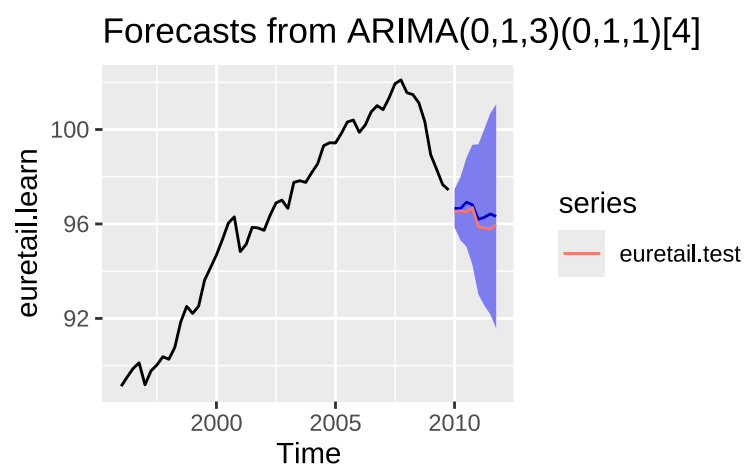
Rysunek 21: Prognoza dla modelu AR(4) na zbiorze testowym euretail



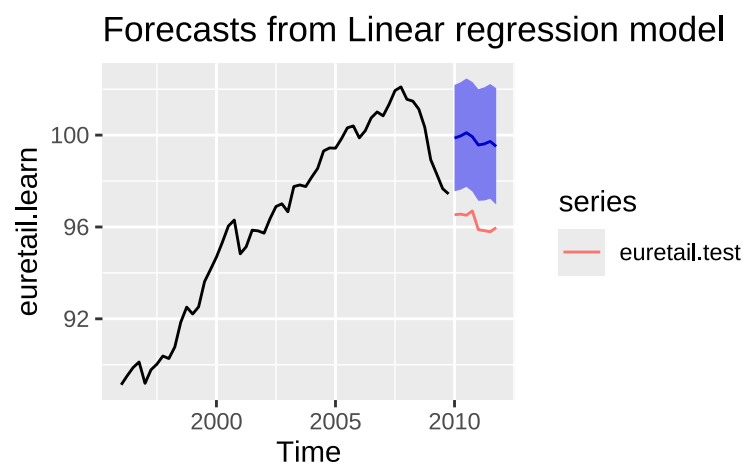
Rysunek 22: Prognoza dla modelu MA(4) na zbiorze testowym euretail



Rysunek 23: Prognoza dla modelu MA(9) na zbiorze testowym euretail

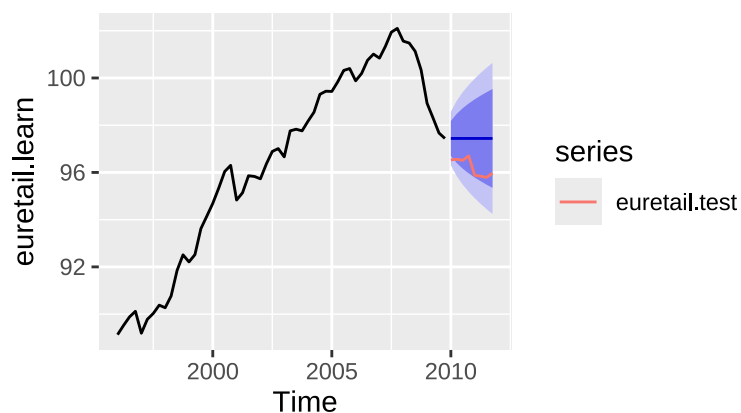


Rysunek 24: Prognoza dla modelu ARIMA(0,1,3)(0,1,1) na zbiorze testowym euretail



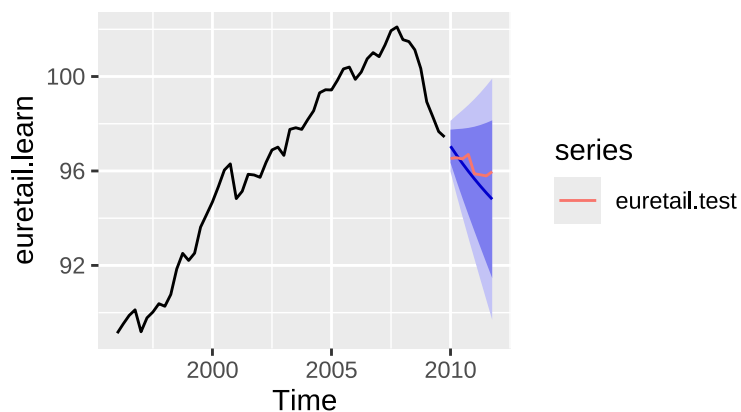
Rysunek 25: Prognoza dla modelu dekompozycji TSLM na zbiorze testowym euretail

Forecasts from Simple exponential smooth



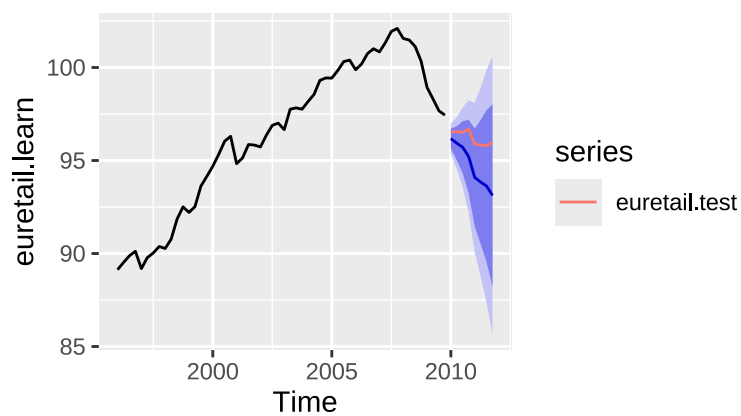
Rysunek 26: Prognoza dla modelu SES na zbiorze testowym euretail

Forecasts from Damped Holt's method

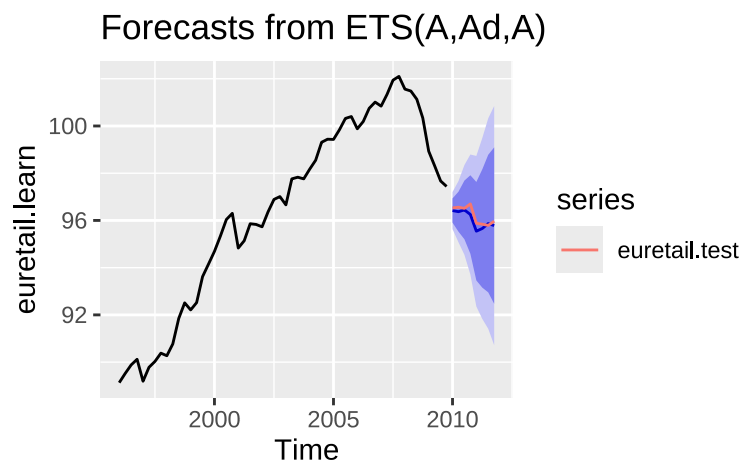


Rysunek 27: Prognoza dla modelu Holta na zbiorze testowym euretail

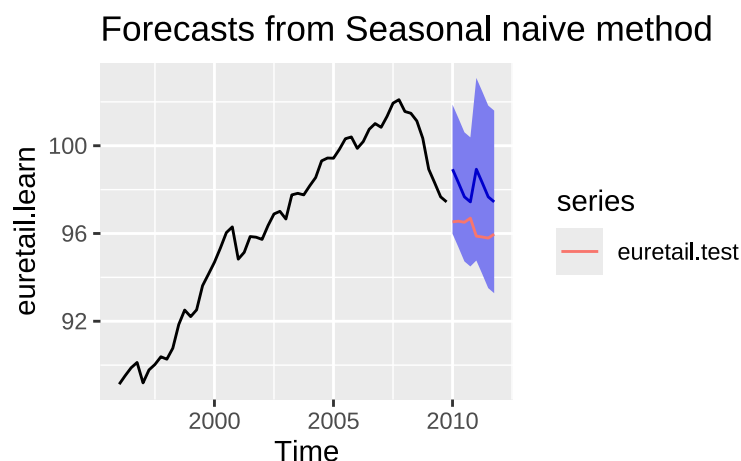
Forecasts from Holt-Winters' additive method



Rysunek 28: Prognoza dla modelu addytywnego Holta-Wintersa na zbiorze testowym euretail



Rysunek 29: Prognoza dla modelu automatycznego ETS na zbiorze testowym euretail



Rysunek 30: Prognoza dla sezonowej metody naiwnej na zbiorze testowym euretail

Tabela 6: Porównanie dokładności prognoz na zbiorze treningowym i testowym

Model	RMSE (trening)	MAE (trening)	RMSE (test)	MAE (test)
ARIMA(4,2,0)(0,1,0)	0.4674	0.3672	2.7840	2.4282
ARIMA(0,2,4)(0,1,0)	0.4191	0.3259	1.5007	1.2971
ARIMA(0,2,9)(0,1,0)	0.3522	0.2725	1.6136	1.2747
ARIMA(0,1,3)(0,1,1)	0.3769	0.2892	0.3588	0.3113
tslm	0.9814	0.8026	3.5707	3.5639
SES	0.5666	0.4599	1.2695	1.2175
Holt (tłumiony)	0.5225	0.3916	0.6087	0.5127
Holt-Winters (addytywny)	0.3847	0.2936	1.7085	1.5084
ETS (automatyczny)	0.3665	0.2728	0.2346	0.1995
Sezonowa metoda naiwna	1.5013	1.2473	1.9942	1.8650

2.3 Wnioski

Modele, które nie uwzględniają sezonowości, takie jak SES (Rysunek 26) i model Holta (Rysunek 27), generują prognozy całkowicie odbiegające od rzeczywistych wartości. Podobnie, ręcznie dobrane modele ARIMA (Rysunki: 21, 22, 23) oraz model dekompozycji tslm (Rysunek 25) próbują naśladować dynamikę szeregu, jednak ich prognozy znacząco odbiegają od danych testowych, co potwierdzają wysokie wartości błędów w Tabeli 6.

Zupełnie inaczej (w większości) prezentują się modele, które pomyślnie przeszły wcześniejszą diagnostykę. Prognozy z modeli **ARIMA(0,1,3)(0,1,1)** (Rysunek 24) oraz **automatycznego ETS** (Rysunek 29) bardzo dobrze pokrywają się z rzeczywistymi danymi ze zbioru testowego, trafnie oddając zarówno trend, jak i sezonowość. Ciekawy jest przypadek modelu addytywnego Holta-Wintersa, który mimo że przeszedł testy diagnostyczne, jego prognozy okazały się zaskakująco słabe, co pokazuje że pomyślna diagnostyka jest warunkiem koniecznym, ale nie wystarczającym.

Ostateczną weryfikację jakości modeli dostarcza porównanie miar błędów prognoz, zebrane w Tabeli 6. Warto przy tym zauważyć, że prawie wszystkie modele osiągnęły dość niskie poziomy błędów na zbiorze treningowym, jednak to co wyróżnia najlepsze modele, to fakt, że osiągają one niższe poziomy błędów na zbiorze testowym, natomiast słabe modele przeciwnie. Przykładowo, model ARIMA(0,2,9)(0,1,0) choć miał najniższy błąd dopasowania do danych treningowych, to osiągnął znacznie gorsze wyniki na zbiorze testowym, co wskazuje na jego przeuczenie i ogólnie słabą zdolność do generalizacji.

Dane dość jednoznacznie wskazują, że model **automatyczny ETS** jest modelem optymalnym. Uzyskał on najniższe wartości $RMSE = 0.2346$, jak i $MAE = 0.1995$ na zbiorze testowym. Bardzo dobre wyniki uzyskał również model ARIMA(0,1,3)(0,1,1), który jest najlepszym kandydatem z rodziny modeli ARIMA, jednak jego błędy są o ponad 50% wyższe niż w przypadku modelu ETS.