

**POLITECHNIKA LUBELSKA**

**WYDZIAŁ PODSTAW TECHNIKI**

*Kierunek: MATEMATYKA*



**Praca inżynierska**

Wykorzystywanie modeli klasy ARFIMA do prognozowania cen  
surowców

*Application of ARFIMA models in forecasting raw materials prices*

*Praca wykonana pod kierunkiem:*  
prof. dr. hab. Edward Kozłowski

*Autor:*  
Agnieszka Wrzos  
nr albumu: 089917

**Lublin 2021**



## Spis treści

<b>Wstęp</b>	5
<b>Rozdział 1. Szeregi czasowe</b>	7
1.1. Podstawowe charakterystyki szeregów czasowych	7
1.2. Współczynnik Hursta	10
1.3. Funkcja gamma	12
1.4. Transformacja danych	14
1.5. Różnice rzędu ułamkowego	15
1.6. Modele stacjonarnych szeregów czasowych	17
1.7. Model ARIMA	21
1.8. Testy pierwiastka jednostkowego	22
1.8.1. Rozszerzony test Dickeya-Fullera	22
1.8.2. Test KPSS	23
1.8.3. Test Phillipsa-Perrona	24
1.9. Model ARFIMA	25
<b>Rozdział 2. Identyfikacja zachowania wybranych cen surowców za pomocą modeli ARFIMA</b>	29
2.1. Identyfikacja i predykcja cen palladu	30
2.1.1. Odczyt danych, analiza stacjonarności	30
2.1.2. Dopasowanie modelu ARFIMA	31
2.1.3. Predykcja dla modelu ARFIMA	34
2.2. Identyfikacja i predykcja cen kakao	35
2.2.1. Odczyt danych, analiza stacjonarności	35
2.2.2. Dopasowanie modelu ARFIMA	37
2.2.3. Predykcja dla modelu ARFIMA	39
2.3. Identyfikacja i predykcja cen oleju opałowego	41
2.3.1. Odczyt danych, analiza stacjonarności	41
2.3.2. Dopasowanie modelu ARFIMA	42
2.3.3. Predykcja dla modelu ARFIMA	45

<b>Podsumowanie i wnioski . . . . .</b>	<b>47</b>
<b>Bibliografia . . . . .</b>	<b>49</b>
<b>Spis rysunków . . . . .</b>	<b>51</b>
<b>Spis tabel . . . . .</b>	<b>53</b>
<b>Załączniki . . . . .</b>	<b>55</b>
<b>Streszczenie (Summary) . . . . .</b>	<b>57</b>

## Wstęp

W obecnych czasach coraz częściej doceniane jest gromadzenie zbiorów danych oraz ich efektywna analiza. Dzieje się to, ze względu na przydatność przy podejmowaniu ważnych decyzji gospodarczych, technologicznych, ekologicznych itp. Wzrasta zapotrzebowanie dotyczące predykcji zachowań pewnych procesów. Rozwój metod analizy danych pozwala na coraz dokładniejsze i bardziej wydajne prognozy. Głównym problemem jest wybór odpowiedniego modelu opisującego zachowanie analizowanego zjawiska tak, by prognoza była jak najbardziej trafna. Istotne także okazuje się uzyskiwanie wiedzy na temat mechanizmów, które miały wpływ na przebieg rozważanego zjawiska. Problematyka wyżej wspomnianych prognoz jest bardzo różnorodna, a przy tym złożona. W praktyce często stosowana jest analiza szeregów czasowych. Szereg czasowy jest to proces stochastyczny, realizację którego obserwujemy w rzeczywistości. W szeregu czasowym identyfikujemy następujące składowe:

- trend - czynnik, który wyraża długookresową tendencję rozwojową;
- sezonowość - czynnik, który wpływa na okresowe wahania badanego zjawiska, w czasie nieprzekraczającym roku;
- cykliczność - czynnik, który odpowiada za długookresowe, rytmiczne wahania wartości szeregu wokół tendencji rozwojowej;
- losowość - czynnik, który reprezentuje wahania o charakterze losowym.

W niektórych przypadkach, bezpośrednia identyfikacja wyżej wymienionych składowych jest utrudniona. W niniejszej pracy zostanie przedstawiona klasa modeli ARIMA uwzględniająca tendencje rozwojowe lub spadkowe oraz rozszerzenie ARFIMA na przypadek modeli z długą pamięcią. Przedstawione zostaną różnice i podobieństwa pomiędzy szeregami czasowymi opisanymi za pomocą modeli ARMA, ARIMA oraz ARFIMA. Zachowania cen niektórych surowców zostaną przedstawione za pomocą modeli ARFIMA. Do części praktycznej pracy zostanie wykorzystane środowisko R, zaś dane pobrane zostaną z serwisu *Yahoo Finance* [37].



## Rozdział 1

### Szeregi czasowe

#### 1.1. Podstawowe charakterystyki szeregów czasowych

W celu zrozumienia dalszych definicji, wprowadzimy najpierw niezbędne oznaczenia. Zbiór liczb naturalnych będziemy oznaczać przez  $\mathbb{N}$ . Zbiory liczb całkowitych i rzeczywistych oznaczamy odpowiednio przy pomocy symboli  $\mathbb{Z}$  i  $\mathbb{R}$ . Przez  $(\Omega, \mathcal{F}, P)$  rozumiemy przestrzeń probabilistyczną. Estymator wartości oczekiwanej oznaczamy przez  $E$ . Ponadto konieczne jest przedstawienie następujących pojęć:

**Definicja 1.1** (Szereg czasowy). *Szeregiem czasowym  $\{x_t\}_{t \in T}$  nazywamy ciąg zmiennych losowych  $x_t : \Omega \rightarrow \mathbb{R}$  dla dowolnego  $t \in T$ . W zależności od potrzeb przyjmujemy  $T = \mathbb{N}$  lub  $T = \mathbb{Z}$ .*

Ważnym podziałem szeregów czasowych jest ich rozróżnienie ze względu na stacjonarność. Stacjonarne szeregi czasowe nie zmieniają swoich własności dynamicznych przy przesunięciu czasowym. Poniżej przedstawione zostaną definicje szeregów stacjonarnych. [16]

**Definicja 1.2** (Szereg ściśle stacjonarny). Dane są dowolne chwile czasowe  $t_1, t_2, \dots, t_m \in \mathbb{N}$ , przesunięcie czasowe  $k \in \mathbb{N}$  oraz  $m \in \mathbb{N}$ , oznaczające liczbę elementów ciągu. Szereg  $\{x_t\}_{t \in T}$  nazywamy *szeregiem ściśle stacjonarnym*, jeżeli  $m$ -elementowe ciągi  $x_{t_1}, x_{t_2}, \dots, x_{t_m}$  oraz  $x_{t_1+k}, x_{t_2+k}, \dots, x_{t_m+k}$  mają identyczne rozkłady łączne.

**Definicja 1.3** (Szereg słabo stacjonarny). Szereg czasowy  $\{x_t\}_{t \in T}$ , którego elementy mają skończone drugie momenty ( $Ex_t^2 < \infty$ , dla  $t \in \mathbb{Z}$ ), nazywamy *słabo stacjonarnym* (stacjonarnym w szerszym sensie), jeżeli spełnione są następujące warunki:

1.  $Ex_t = \mu = \text{const}$ , dla dowolnego  $t \in T$
2. dla każdego  $t \in \mathbb{T}$  i dowolnego  $k \in \mathbb{Z}$  spełniona jest równość

$$\text{cov}(x_t, x_{t+k}) = \text{cov}(x_0, x_k).$$

**Definicja 1.4** (Funkcja autokowariancji). Dla szeregu czasowego  $\{x_t\}_{t \in T}$ , funkcję autokowariancji (ACVF) z przesunięciem czasowym  $k \in \mathbb{N}$  określamy jako

$$\gamma_k = \text{cov}(x_t, x_{t+k}) = E(x_t - \mu)(x_{t+k} - \mu), \quad (1.1)$$

gdzie  $\mu = Ex_t$ , a wielkość  $\sigma^2 = \gamma_0 = E(x_t - \mu)^2$  nazywamy wariancją szeregu. Dla każdego  $k \in \mathbb{N}$  funkcja  $\gamma_k$  spełnia następujące własności:

- $\gamma_0 \geq 0$ ;
- $|\gamma_k| \leq \gamma_0$ ;
- $\gamma_{-k} = \gamma_k$ .

**Definicja 1.5** (Funkcja autokorelacji). Dla szeregu czasowego  $\{x_t\}_{t \in T}$  funkcję autokorelacji (ACF) z opóźnieniem czasowym  $k \in \mathbb{N}$  określamy za pomocą wyrażenia

$$\rho_k = \text{corr}(x_t, x_{t+k}) = \frac{\text{cov}(x_t, x_{t+k})}{\sigma^2} \quad (1.2)$$

gdzie przez  $\sigma^2$  rozumiemy wariancję procesu  $x_t$ .

Funkcja  $\rho_k$  posiada następujące własności:

- $\rho_0 \geq 0$ ;
- $|\rho_k| \leq \rho_0$ ;
- $\rho_{-k} = \rho_k$ .

**Definicja 1.6** (Funkcja autokorelacji cząstkowej). Funkcję autokorelacji cząstkowej (PACF) szeregu  $\{x_t\}_{t \in T}$  z opóźnieniem czasowym  $k \in \mathbb{N}$  definiujemy jako: [3]

$$\alpha_1 = \text{corr}(x_t, x_{t-1}) = \rho_1, \text{ dla } k=1, \quad (1.3)$$

oraz dla opóźnienia czasowego  $k \geq 2$ :

$$\alpha_k = \text{corr}(x_{t+k} - P_{t,k}(x_{t+k}), x_t - P_{t,k}(x_t)), \quad (1.4)$$

gdzie  $P_{t,k}(x)$  oznacza operator ortogonalnej projekcji  $x$  na liniową podprzestrzeń przestrzeni Hilberta, utworzoną przez  $x_{t+1}, \dots, x_{t+k-1}$ .

Procesy z krótką pamięcią cechują się tym, że ich funkcja ACF zanika w tempie wykładniczym. W przypadku występowania długiej pamięci funkcja ACF zanika znacznie wolniej, w tempie hiperbolicznym. Innymi słowy długa pamięć pozwala na uwzględnienie korelacji nawet między stosunkowo odległymi od siebie elementami szeregu. Klasycznymi przykładami szeregów czasowych z brakiem pamięci i długą pamięcią są biały szum i błędzenie losowe.



Niech  $(\Omega, \mathcal{F}, P)$  będzie przestrzenią probabilistyczną, na której definiujemy nie-malejącą rodzinę  $\sigma$ -ciał  $\mathcal{F}_j = \sigma \{ \varepsilon_i : i = 0, 1, \dots, j \}$ ,  $j \in \mathbb{N}$ . Dodatkowo dany jest ciąg  $\{ \varepsilon_t \}_{t \in \mathbb{N}}$  niezależnych zmiennych losowych o rozkładzie normalnym  $N(0, \sigma^2)$ , gdzie:

$$E\varepsilon_t = 0, \text{ dla każdego } t \in \mathbb{N}$$

$$\text{cov}(\varepsilon_t, \varepsilon_s) = \begin{cases} 0, & \text{gdy } t \neq s; \\ \sigma^2, & \text{gdy } t = s. \end{cases}$$

gdzie  $\varepsilon_t \sim N(0, \sigma^2)$  oraz  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$ , przy  $t \neq s$ .

**Definicja 1.7.** Ciąg niezależnych zmiennych losowych  $\{ \varepsilon_t \}_{t \in \mathbb{N}}$  o rozkładzie normalnym  $N(0, \sigma^2)$  nazywamy *białym szumem*.

**Definicja 1.8** (Błądzenie losowe). Dla szeregu czasowego  $\{x_t\}_{t \in \mathbb{N}}$ , proces *błądzenia losowego* definiujemy jako:

$$x_t = x_{t-1} + \varepsilon_t \quad (1.5)$$

Dodatkowo dla  $t \in \mathbb{N}$ :

$$x_t = \sum_{i=1}^t \varepsilon_i. \quad (1.6)$$

Błądzenie losowe zwane także modelem trendu stochastycznego posiada następujące własności:

— wartość oczekiwana

$$Ex_t = \sum_{i=1}^t E\varepsilon_i = 0; \quad (1.7)$$

— wariancja

$$\gamma_0(t) = Ex_t^2 = \sum_{j=1}^t \sigma^2 = t\sigma^2; \quad (1.8)$$

— kowariancja

$$\gamma_k(t) = Ex_t x_{t-k} = (t-k)\sigma^2; \quad (1.9)$$

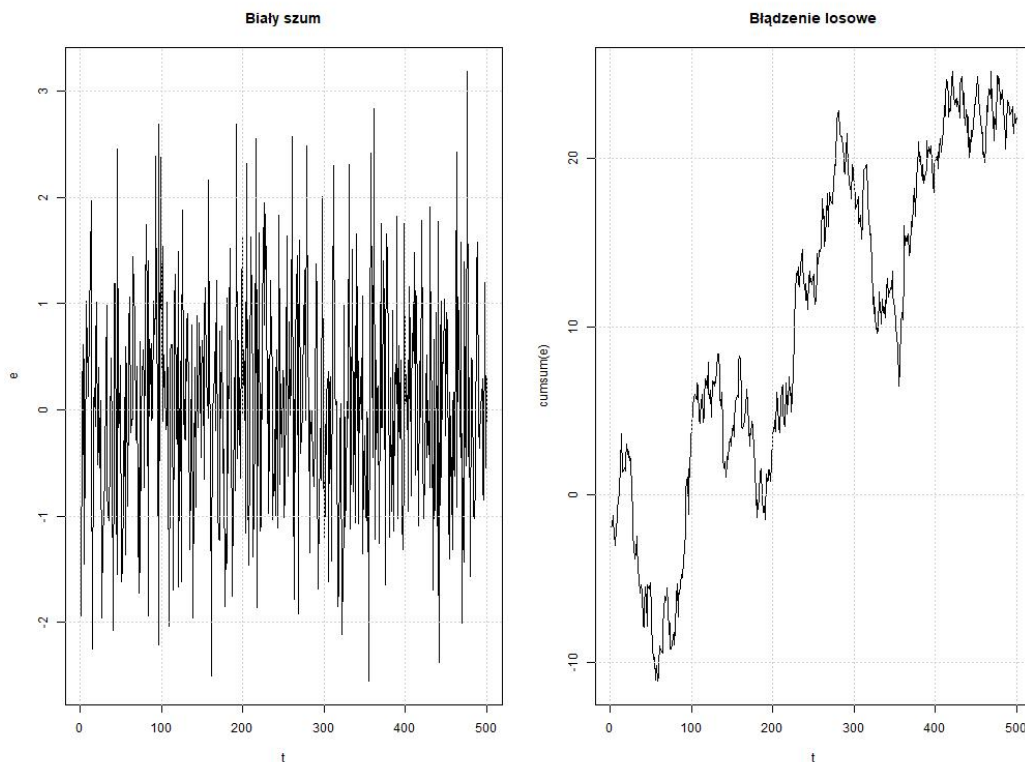
— współczynnik korelacji

$$\rho_k(t) = \frac{\gamma_k(t)}{\sqrt{\gamma_0(t) \gamma_0(t-k)}} = \frac{(t-k)}{\sqrt{t(t-k)}} = \sqrt{1 - \frac{k}{t}}. \quad (1.10)$$

Rysunek 1.1 przedstawia symulację białego szumu oraz błędzenia losowego za pomocą wzoru (1.5).

Dla ustalonego  $k \in \mathbb{N}$ , w szeregu  $\{x_t\}_{t \in \mathbb{N}}$  między elementami odległymi o  $k$  momentów występuje mocna korelacja. Wynika to z równości:

$$\lim_{t \rightarrow \infty} \rho_k(t) = \lim_{t \rightarrow \infty} \sqrt{1 - \frac{k}{t}} = 1.$$



**Rysunek 1.1:** Symulacje: biały szum oraz błądzenie losowe

Dla procesów z krótką pamięcią wyrażenie  $\sum_{k=-\infty}^{\infty} \rho_k$  przyjmuje wartość skończoną, natomiast procesów z długą pamięcią szereg  $\sum_{k=-\infty}^{\infty} \rho_k$  jest rozbieżny, tzn.

$$\lim_{n \rightarrow \infty} \sum_{k=-n}^n \rho_k = \infty \quad (1.11)$$

Rozbieżność ciągu  $\{\rho_k\}_{k \in \mathbb{Z}}$  powoduje istnienie zależności (relacji) pomiędzy elementami szeregu  $\{x_t\}$  przy przesunięciu  $k \in \mathbb{Z}$ . Jednym ze wskaźników, który jest wykorzystywany przy badaniu istnienia długiej pamięci jest współczynnik Hursta.

## 1.2. Współczynnik Hursta

Zapoczątkowane przez Grangera i Joyeux w 1980 roku, procesy z długą pamięcią, zostały docenione ze względu na przydatność w modelowaniu długoterminowych prognoz. Oficjalnie mówimy o latach osiemdziesiątych, jednak pierwsze artykuły pojawiały się już w latach pięćdziesiątych w takich dziedzinach jak hydrologia, czy klimatologia. Najbardziej znanym przykładem jest artykuł hydrologa Harolda Edwina Hursta (1951). Zauważył on okresowo pojawiające się wylewy Nilu. Analizując dane z poprzednich lat, opracował metodę pozwalającą na rozwiązanie jego problemu, pokazując przy tym, że jest to proces losowy, którego elementy nie są niezależne [18].

Za pomocą współczynnika Hursta możemy określić wpływ procesów z długą pamięcią na realizację szeregu czasowego. Pozwala na oszacowanie poziomu stochastyczności w szeregu czasowym i oznaczany jest przez symbol  $H$ . Jedną z metod obliczania współczynnika Hursta jest metoda analizy przeskalowanego zakresu (metoda analizy R/S), który wyraża się wzorem: [33], [21]

$$E(R/S)_n \equiv cn^H, \quad (1.12)$$

gdzie  $c$  jest pewną dodatnią stałą, a  $H$  jest wykładnikiem Hursta. Aby wyznaczyć  $H$ , należy dla różnych  $n$ -elementowych podziałów realizacji szeregu  $\{x_t\}_{1 \leq t \leq N}$  obliczyć  $E(R/S)_n$ , a następnie, przy pomocy regresji liniowej, oszacować parametry strukturalne  $c$  i  $H$  [15], [5] równania

$$\ln E(R/S)_n = \ln c + H \ln n. \quad (1.13)$$

Szeregi czasowe dzielimy ze względu na wartość współczynnika Hursta:

- $0 < H < 0.5$  - szereg posiada tendencję do częstej zmiany kierunku (proces stacjonarny z długą pamięcią);
- $H = 0.5$  - biały szum;
- $0.5 < H < 1$  - szereg persystentny z długą pamięcią.

Niech  $\{x_1, x_2, \dots, x_N\}$  będzie realizacją szeregu czasowego. Algorytm wyznaczania współczynnika Hursta przebiega następująco: [35]

1. Dzielimy ciąg  $\{x_t\}_{t \in N}$  na  $m$  części złożonych z  $n$  elementów,

$$m = \left\lceil \frac{N}{n} \right\rceil, \quad (1.14)$$

gdzie:  $\lceil \cdot \rceil$  oznacza część całkowitą. Jeśli iloraz  $\frac{N}{n}$  nie jest liczbą całkowitą, to  $mn < N$ . W tym przypadku, w dalszej części algorytmu użyjemy wartości  $y_{ij} = x_{(i-1) \cdot n + j}$ .

2. Definiujemy sumę częściową  $z_{ij}$  dla  $i = 1, 2, \dots, m$  i  $j = 1, 2, \dots, n$ :

$$z_{ij} = y_{ij} - \bar{y}_i, \quad (1.15)$$

gdzie:  $y_{ij}$  oznacza  $j$ -tą wartość w  $i$ -tym przedziale. Średnia wartość dla  $i$ -tego podciągu wynosi:

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{ij}. \quad (1.16)$$

3. Dla  $i$ -tego przedziału, wyznaczamy ciąg sum częściowych  $\{q_{ij}\}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$  postaci:

$$q_{ij} = \sum_{j=1}^i z_{ij} \quad (1.17)$$

dla  $i = 1, 2, \dots, m$  oraz  $j = 1, 2, \dots, n$ .

4. Zakres  $i$ -tego przedziału definiujemy rozstęp jako:

$$R_{in} = \max_{1 \leq j \leq n} (q_{ij}) - \min_{1 \leq j \leq n} (q_{ij}) \quad (1.18)$$

5. Obliczamy rozstęp przeskalowany dla każdego skumulowanego szeregu czasowego, tzn. dzielimy rozstęp przez odchylenie standardowe tego szeregu:

$$\alpha_{in} = \frac{R_i}{S_i}, \quad (1.19)$$

gdzie  $S_i = \sqrt{\frac{1}{n} \sum_{j=1}^m z_{ij}^2}$ ,  $1 \leq i \leq m$

6. Ostatecznie obliczamy wartość średnią wskaźnika R/S dla podciągów o długości  $n$

$$\left(\frac{R}{S}\right)_n = \frac{1}{m} \sum_{l=1}^m \alpha_{ln} \quad (1.20)$$

Powyższą procedurę przeprowadza się dla różnych podziałów realizacji szeregu czasowego  $\{x_t\}_{1 \leq t \leq N}$ , gdzie  $10 \leq n \leq \frac{N}{2}$ . Tym sposobem otrzymujemy zależność wielkości R/S od długości  $n$  elementów w każdej części podciągu.

7. Zależność pomiędzy  $R/S$  a długością podziału  $n$  jest dana wzorem

$$\left(\frac{R}{S}\right)_n = cn^H \quad (1.21)$$

gdzie  $H$  jest wykładnikiem Hursta,  $c$  jest stałą. Wartość wykładnika Hursta wyznaczamy z dopasowania równania liniowego:

$$\ln((R/S)_t) = \ln c + H \ln t \quad (1.22)$$

W tym celu stosujemy metodę najmniejszych kwadratów. Modyfikacja metody  $R/S$  została zaproponowana przez Lo [22]

### 1.3. Funkcja gamma

Zgodnie z definicją Eulera rozważmy iloczyn postaci [19]

$$\Gamma(z) = \frac{1}{z} \prod_{n=1}^{\infty} \frac{\left(1 + \frac{1}{n}\right)^z}{1 + \frac{z}{n}} \quad (1.23)$$

dla  $z \in \mathbb{R} \setminus \{0, -1, -2, \dots\}$ .

Niech

$$\Gamma^m(z) = \frac{1}{z} \frac{\left(1 + \frac{1}{1}\right)^z \left(1 + \frac{1}{2}\right)^z \dots \left(1 + \frac{1}{m}\right)^z}{\left(1 + z\right) \left(1 + \frac{z}{2}\right) \dots \left(1 + \frac{z}{m}\right)} = \frac{1}{z} \cdot \frac{\left(\frac{2}{1}\right)^z \left(\frac{3}{2}\right)^z \dots \left(\frac{m+1}{m}\right)^z}{(z+1)(z+2) \dots (z+m)} \cdot m! =$$

$$= \frac{m!(m+1)^z}{z(z+1)(z+1)\dots(z+m)} = \left(\frac{m+1}{m}\right)^z \cdot \frac{m!m^z}{z(z+1)(z+2)\dots(z+m)}.$$

Dla każdego ustalonego  $z \in \mathbb{R} \setminus \{0, -1, -2, \dots\}$

$$\begin{aligned}\Gamma(z) &= \lim_{m \rightarrow \infty} \Gamma^m(z) = \lim_{m \rightarrow \infty} \left(\frac{m+1}{m}\right)^z \cdot \frac{m!m^z}{z(z+1)(z+2)\dots(z+m)} = \\ &= \lim_{m \rightarrow \infty} \frac{m!m^z}{z(z+1)(z+2)\dots(z+m)}\end{aligned}$$

Oszacujemy iloraz

$$\frac{\Gamma(z+1)}{\Gamma(z)} = \lim_{m \rightarrow \infty} \frac{\Gamma^m(z+1)}{\Gamma^m(z)} = \lim_{m \rightarrow \infty} \frac{\frac{m!m^{z+1}}{(z+1)(z+2)\dots(z+m+1)}}{\frac{m!m^z}{z(z+1)(z+2)\dots(z+m)}} = \lim_{m \rightarrow \infty} \frac{mz}{z+m+1} = z$$

Wobec powyższego otrzymujemy zależność

$$\Gamma(z+1) = z \cdot \Gamma(z), \text{ dla } z \in \mathbb{R} \setminus \{0, -1, -2, \dots\}. \quad (1.24)$$

Ponadto dla  $z \in \mathbb{N}$ :

$$\Gamma(z+1) = z \cdot \Gamma(z) = z(z-1)! = z! \quad (1.25)$$

Funkcja gamma zaproponowana przez D. Bernoulliego rozszerza pojęcie silni na liczby rzeczywiste i zespolone. W przypadku, gdy część rzeczywista liczby zespolonej  $z$  jest dodatnia, funkcja gamma ma postać:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt. \quad (1.26)$$

Jest to całka niewłaściwa znana jako *całka Eulera drugiego rodzaju*. Za pomocą całkowania przez części otrzymujemy

$$\Gamma(z+1) = \int_0^\infty t^z e^{-t} dt = -t^z e^{-t} \Big|_0^\infty + \int_0^\infty z t^{z-1} e^{-t} dt = 0 - 0 + z \int_0^\infty t^{z-1} e^{-t} dt = z \cdot \Gamma(z) \quad (1.27)$$

gdzie  $\lim_{t \rightarrow \infty} \frac{t^z}{e^t} = 0$ .

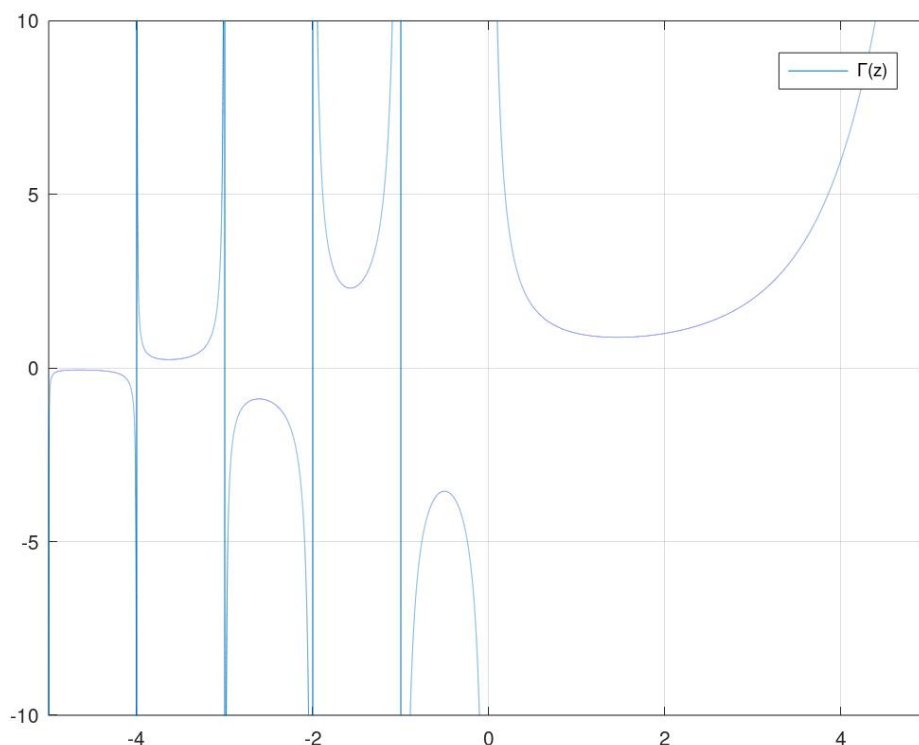
Dla  $z = 0$  mamy

$$\Gamma(1) = \int_0^\infty t^0 e^{-t} dt = \int_0^\infty e^{-t} dt = -e^{-t} \Big|_0^\infty = 0 + 1 = 1 \quad (1.28)$$

Korzystając z własności funkcji gamma *symbol Newtona* szacujemy jako

$$\binom{k}{j} = \frac{\Gamma(k+1)}{\Gamma(j+1)\Gamma(k-j+1)}.$$

Wartości funkcji gamma dla  $z \in \mathbb{R}$  przedstawia rysunek 1.2.

**Rysunek 1.2:** Wykres funkcji gamma*Źródło:* Opracowanie własne

## 1.4. Transformacja danych

Głównym celem transformacji danych (np. realizacji szeregu czasowego) jest modyfikacja do rozkładu normalnego. Popularną metodą jest transformacja Boxa-Coxa. [4] Najczęściej tę transformację stosujemy do danych o rozkładzie asymetrycznym (lewostronnie lub prawostronnie skośnych lub też „uciętych” rozkładów normalnych). Dla realizacji szeregu  $\{x_t\}_{1 \leq t \leq N}$  stosujemy transformację  $f_\lambda$  Boxa-Coxa postaci:

$$f_\lambda(x_t) = \begin{cases} \lambda^{-1}(x_t^\lambda - 1), & \text{gdy } \forall t \ x_t \geq 0 \text{ oraz } \lambda > 0, \\ \ln x_t, & \text{gdy } \forall t \ x_t > 0 \text{ oraz } \lambda = 0. \end{cases}$$

Przy podejmowaniu decyzji o wspomnianych przekształceniach, warto zwrócić uwagę na założenie o dodatniości danych. W przypadku, gdy realizacja szeregu  $\{x_t\}_{1 \leq t \leq N}$  zawiera wartości niedodatnie, to transformację Boxa-Coxa stosujemy dla ciągu

$$\{x_t + C\}_{1 \leq t \leq N},$$

gdzie  $C > -\min_{1 \leq t \leq N} x_t$ . Należy jednak przesunąć dokonywać ostrożnie, gdyż w niektórych przypadkach może to spowodować spore odchylenia. Warto także zwrócić uwagę na to, czy transformując dane nie bagatelizujemy problemu wartości odstających.

## 1.5. Różnice rzędu ułamkowego

Przy identyfikacji procesów niestacjonarnych oraz procesów z długą pamięcią wykorzystujemy technikę różnicowania. Różnica ta jest jednym ze sposobów sprowadzenia szeregu do szeregu spełniającego warunek stacjonarności. Intuicyjnie poprzez różnicowanie rozumiemy zamianę szeregu czasowego zawierającego wartości, na szereg różnic pomiędzy odpowiednimi punktami w czasie. Do modelowania części odpowiadających za autoregresję i ruchomą średnią w szeregach czasowych  $\{x_t\}_{t \in \mathbb{N}}$  wykorzystujemy operatory opóźnienia i różnicowy.

**Definicja 1.9** (Operator opóźnienia). Przez *operator opóźnienia*  $B$  rozumiemy operator, którego aplikacja na danym elemencie powoduje uzyskanie elementu go poprzedzającego [25], tzn.

$$B^k x_t = x_{t-k}, \text{ dla } k \in \mathbb{N} \quad (1.29)$$

**Definicja 1.10** (Operator różnicowy). Dla  $k \in \mathbb{N}$  będącego rzędem różnicowania szeregu, *operator różnicowania* definiujemy w sposób następujący: [14]

$$\Delta^k x_t = (1 - B)^k x_t. \quad (1.30)$$

W szczególności, gdy  $k = 1$ :

$$\Delta x_t = x_t - x_{t-1} = (1 - B)x_t, \quad (1.31)$$

W praktyce do osiągnięcia stacjonarności zazwyczaj niezbędne jest wielokrotne różnicowanie szeregu  $\{x_t\}_{t \in \mathbb{N}}$ . Oznacza to, że powyższy proces powtarzamy, aż uda nam się wyeliminować niestacjonarność. Dzięki metodzie różnicowania, możemy zbadać rodzaj występującego trendu w badanym szeregu. Gdy pojawia się konieczność, aby rozpatrzeć różnice różnic, przydatny okazuje się parametr  $d$ , oznaczający *stopień zintegrowania szeregu*. Szereg niestacjonarny, który można sprowadzić do szeregu stacjonarnego, poprzez zastosowanie operatora różnicowego  $d$  razy, nazywmy *szeregiem zintegrowanym stopnia  $d$* , oraz oznaczamy  $\{x_t\}_{t \in \mathbb{N}} \sim I(d)$ .

**Twierdzenie 1.1.** Dla operatora różnicowego rzędu  $k \in \mathbb{N}$ , postaci (1.30)–(1.31) spełniona jest równość

$$\Delta^k x_t = (1 - B)^k x_t = \sum_{j=0}^k (-1)^j \binom{k}{j} B^j x_t \quad (1.32)$$

gdzie  $\binom{k}{j}$  jest symbolem Newtona,  $0 \leq j \leq k$ ,  $0! = 1$ .

Zgodnie z [14], [5] oraz [34] wprowadzimy definicje operatora całkowania (integracji) oraz operatorów różnicowego i całkowego rzędu ułamkowego, które są niezbędne przy modelowaniu szeregów czasowych.

**Definicja 1.11** (Operator całkowania (integracji)). Operatorem odwrotnym do operatora różnicowania jest *operator całkowania (integracji)* rzędu  $k \in \mathbb{N} \setminus \{0\}$ , który definiujemy jako:

$$(1 - B)^{-k} = \sum_{j=0}^{\infty} \binom{k+j-1}{j} B^j = \sum_{j=0}^{\infty} \frac{\Gamma(k+j)}{\Gamma(j+1)\Gamma(k)} B^j \quad (1.33)$$

*Wniosek 1.1.* Proces błędzenia losowego (1.5) możemy przedstawić w postaci

$$(1 - B)x_t = \varepsilon_t$$

oraz

$$x_t = (1 - B)^{-1} \varepsilon_t = (1 + B + B^2 + \dots) \varepsilon_t.$$

**Definicja 1.12** (Operator różnicowy rzędu ułamkowego). Dla szeregu  $\{x_t\}_{t \in \mathbb{Z}}$ , przez operator różnicowy rzędu ułamkowego  $k \in \mathbb{R}$  rozumiemy postaci:

$$(1 - B)^k x_t = \sum_{j=0}^{\infty} (-1)^j \binom{k}{j} B^j x_t. \quad (1.34)$$

Operator różnicowy rzędu ułamkowego wyprowadzimy poprzez analogię (zgodnie z [15], [34], [11], [12]). Korzystając z własności funkcji  $\Gamma(z)$ , wzór 1.34 możemy przedstawić w postaci:

$$\begin{aligned} (1 - B)^k x_t &= \sum_{j=0}^{\infty} (-1)^j \frac{k(k-1) \dots (k-j+1)}{j!} B^j x_t \\ &= \sum_{j=0}^{\infty} \frac{(j-k-1)(j-k-2) \dots (-k)}{j!} B^j x_t \\ &= \sum_{j=0}^{\infty} \frac{\Gamma(j-k)}{\Gamma(-k)\Gamma(j+1)} B^j x_t \end{aligned} \quad (1.35)$$

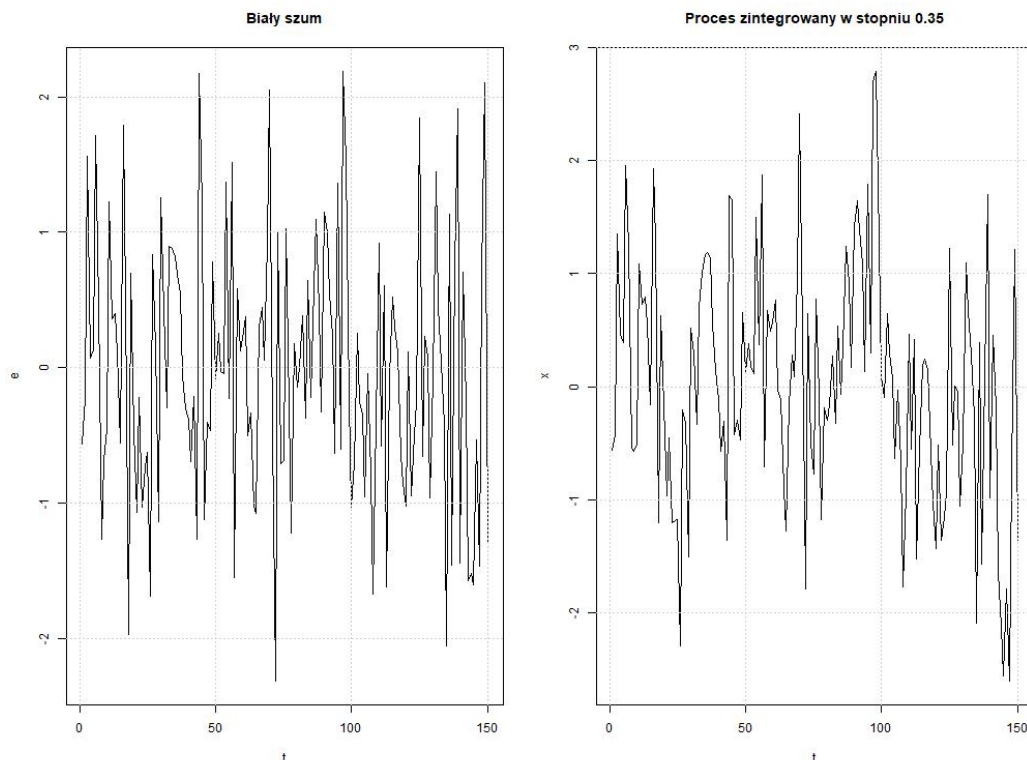
**Definicja 1.13** (Operator całkowania rzędu ułamkowego). Operator całkowania (integracji) rzędu  $k \in \mathbb{R}$  ma postać:

$$(1 - B)^{-k} x_t = \sum_{j=0}^{\infty} \binom{k}{j} B^j x_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+k)}{\Gamma(k)\Gamma(j+1)} B^j x_t. \quad (1.36)$$



Operatory całkowania i różnicowania rzędu ułamkowego wykorzystujemy do modelowania i identyfikacji szeregów z długą pamięcią. Rysunek 1.3 przedstawia symulację białego szumu  $\{\varepsilon_t\}_{1 \leq t \leq 150}$  oraz wykres procesu zintegrowanego  $\{x_t\}_{1 \leq t \leq 150}$  rzędu 0.35 postaci

$$(1 - B)^{-0.35} x_t = \varepsilon_t \quad (1.37)$$



**Rysunek 1.3:** Symulacja białego szumu oraz wykres integracji rzędu 0.35 białego szumu

## 1.6. Modele stacjonarnych szeregów czasowych

Stacjonarne szeregi czasowe najczęściej modelujemy za pomocą autoregresji i średniej ruchomej. W modelu autoregresji elementy szeregu czasowego są objaśniane za pomocą poprzednich wartości tego szeregu. W modelu średniej ruchomej elementy szeregu są określone jako kombinacje liniowe zaburzeń zewnętrznych opóźnionych w czasie. Połączeniem tych dwóch modeli jest model ARMA (Autoregressive and Moving Average), najczęściej wykorzystywany do prognoz krótkoterminowych. Wspomniane modele cechują się krótką pamięcią, czyli im bardziej oddalone są od siebie elementy szeregu, tym słabsza jest korelacja między nimi.

Niech  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  będzie ciągiem niezależnych zmiennych losowych o rozkładzie  $N(0, \sigma^2)$ , określonym na przestrzeni probabilistycznej  $(\Omega, \mathcal{F}, P)$ . Modelem autoregresji  $AR(p)$ ,  $p \in \mathbb{N}$  nazywamy proces, w którym każdy element jest liniową kombinacją poprzednich. Elementy szeregu  $AR(p)$ ,  $p \in \mathbb{N}$  określamy za pomocą wzoru: [16]

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + \varepsilon_t = \sum_{i=1}^p a_i x_{t-i} + \varepsilon_t, \quad (1.38)$$

gdzie  $t \in \mathbb{N}$ , a współczynniki  $a_1, a_2, \dots$  są parametrami modelu. Za pomocą operatora przesunięcia  $B$ , równanie 1.38 możemy przedstawić w postaci

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) x_t = \varepsilon_t, \quad (1.39)$$

lub

$$\Phi_p(B) x_t = \varepsilon_t, \quad (1.40)$$

gdzie  $\Phi_p(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$  jest operatorem odpowiadającym za autoregresję.

Najprostszym modelem autoregresji jest  $AR(1)$ , gdzie elementy szeregu określamy jako  $(1 - aB)x_t = \varepsilon_t$ . Własności szeregu  $AR(1)$ :

— średnia

$$Ex_t = 0; \quad (1.41)$$

— wariancja

$$\gamma_0 = E\varepsilon_t^2 = \frac{\sigma^2}{1 - a^2}; \quad (1.42)$$

— kowariancja

$$\gamma_k = cov(\varepsilon_t, \varepsilon_{t-k}) = \frac{\sigma^2 a^k}{1 - a^2}; \quad (1.43)$$

— współczynnik korelacji

$$\rho_k = \frac{\gamma_k}{\gamma_0} = a^k, \quad (1.44)$$

dla przesunięcia czasowego  $k \in \mathbb{N}$ .

*Wniosek 1.2.* Zauważamy, że jeśli dla wyrażenia 1.42 wartość  $|a|$  jest bliska jedności, to wariancja elementów szeregu  $\{x_t\}_{t \in \mathbb{N}}$  jest znacznie większa od wariancji zmiennych losowych  $\varepsilon_t$ ,  $t \in \mathbb{N}$ . Zatem, w przypadku silnej korelacji między sąsiednimi elementami szeregu  $\{x_t\}_{t \in \mathbb{N}}$ , to niewielkie wahania szeregu  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  spowodują duże wahania szeregu  $\{x_t\}_{t \in \mathbb{N}}$ .

*Wniosek 1.3.* Z wzoru 1.44 wynika, że wraz ze wzrostem przesunięcia czasowego  $k \in \mathbb{N}$  maleje wykładniczo stopień zależności pomiędzy elementami szeregu  $\{x_t\}_{t \in \mathbb{N}}$ .

Własności szeregu AR(p):

— wartość średnia elementów szeregu

$$Ex_t = 0; \quad (1.45)$$

— wariancja

$$\gamma_0 = cov(\varepsilon_t, \varepsilon_t) = a_1\gamma_1 + a_2\gamma_2 + \dots + a_p\gamma_p + \sigma^2; \quad (1.46)$$

— kowariancja

$$\gamma_k = cov(\varepsilon_t, \varepsilon_{t-k}) = a_1\gamma_{k-1} + a_2\gamma_{k-2} + \dots + a_p\gamma_{k-p}; \quad (1.47)$$

— współczynnik korelacji

$$\rho_k = \frac{\gamma_k}{\gamma_0} = a_1\rho_{k-1} + a_2\rho_{k-2} + \dots + a_p\rho_{k-p}, \quad (1.48)$$

dla  $k > p$ .

Elementy szeregu ruchomej średniej MA(q),  $q \in \mathbb{N}$  określamy za pomocą równania: [3]

$$x_t = \varepsilon_t - b_1\varepsilon_{t-1} - b_2\varepsilon_{t-2} - \dots - b_q\varepsilon_{t-q}. \quad (1.49)$$

Za pomocą operatora przesunięcia  $B$  wzór 1.49 możemy przedstawić w postaci:

$$x_t = \Theta_q(B)\varepsilon_t \quad (1.50)$$

gdzie  $t \in \mathbb{N}$ , współczynniki  $b_1, b_2, \dots$  są parametrami modelu, zaś  $\Theta_q(B) = 1 - b_1B - b_2B^2 - \dots - b_qB^q$ .

Proces średniej ruchomej MA(q) zakłada, że wartość szeregu zależy od zaburzeń z przeszłości do chwili obecnej. Parametr  $q$  określa rząd procesu MA, czyli ile z wcześniejszych zaburzeń ma wpływ na wartość szeregu.

Własności modelu średniej ruchomej MA(q):

— wartość średnia elementów szeregu:

$$Ex_t = 0; \quad (1.51)$$

— wariancja

$$\gamma_0 = cov(\varepsilon_t, \varepsilon_t) = \sigma^2(1 + b_1^2 + \dots + b_q^2); \quad (1.52)$$

— kowariancja

$$\gamma_k = cov(\varepsilon_t, \varepsilon_{t-k}) = \begin{cases} \sigma^2(1 + b_1^2 + \dots + b_q^2), & \text{gdzie } k = 0; \\ \sigma^2(-b_k + b_1b_{k+1} + \dots + b_{q-k}b_q), & \text{gdzie } k = 1, 2, \dots, q; \\ 0, & \text{gdzie } k > q; \end{cases} \quad (1.53)$$

— współczynnik korelacji

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1, & \text{gdy } k = 0; \\ \frac{-b_k + b_1 b_{k+1} + \dots + b_{q-k} b_q}{1 + b_1^2 + \dots + b_q^2}, & \text{gdy } k = 1, 2, \dots, q; \\ 0, & \text{gdy } k > q. \end{cases} \quad (1.54)$$

Aby opisać szereg czasowy, w którym występują jednocześnie własności autoregresyjne i własności związane ze średnią ruchomą, posługujemy się definicją modelu ARMA. ([3], [32],[16], [25]) Stanowi on uogólnienie modeli AR i MA. Jest to model liniowy, w którym wartości zmiennej są objaśniane poprzez elementy szeregu opóźnione w czasie, zaś zaburzenie zewnętrzne jest kombinacją liniową składników losowych. Dla *modelu autoregresji i ruchomej średniej*  $ARMA(p, q)$ ,  $p \geq 1$  i  $q \geq 1$ , elementy szeregu są określone za pomocą równania:

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \varepsilon_t - \sum_{i=1}^q b_i \varepsilon_{t-i}, \quad (1.55)$$

gdzie  $t \in \mathbb{N}$ , współczynniki  $a_1, a_2, \dots, a_p$  oraz  $b_1, b_2, \dots, b_q$  są parametrami modelu. Za pomocą operatora przesunięcia  $B$  wzór 1.55 możemy przedstawić w postaci:

$$\Phi_p(B)x_t = \Theta_q(B)\varepsilon_t, \quad (1.56)$$

gdzie  $\Phi_p(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$  oraz  $\Theta_q(B) = 1 - b_1 B - b_2 B^2 - \dots - b_q B^q$ .

Własności modelu  $ARMA(p, q)$ :

— wartość średnia elementów szeregu

$$Ex_t = 0; \quad (1.57)$$

— kowariancja

$$\gamma_k = a_1 \gamma_{k-1} + \dots + a_p \gamma_{k-p}, \text{ dla } k \geq q + 1 \quad (1.58)$$

— współczynnik korelacji

$$\rho_k = a_1 + \rho_{k-1} + \dots + a_p \rho_{k-p}, \text{ dla } k \geq q + 1. \quad (1.59)$$

Zgodnie z oznaczeniami  $AR(p) = ARMA(p, 0)$  oraz  $MA(q) = ARMA(0, q)$ .

Dla szeregów AR, MA i ARMA sprawdzamy własności stacjonarności i odwracalności. Dla modelu  $AR(p)$  oraz  $ARMA(p, q)$ ,  $p \geq 1$  sprawdzamy spełnienie własności stacjonarności, natomiast dla  $MA(q)$  oraz  $ARMA(p, q)$ ,  $q \geq 1$  spełnienie własności odwracalności.

— proces  $AR(p)$  spełnia warunek stacjonarności, wtedy i tylko wtedy, gdy pierwiastki równania charakterystycznego

$$(1 - a_1 z - a_2 z^2 - \dots - a_p z^p) = 0 \quad (1.60)$$

znajdują się poza okręgiem jednostkowym, tzn.  $|z_i| > 1$  dla  $1 \leq i \leq p$ .

- proces MA(q) spełnia warunek odwracalności, wtedy i tylko wtedy, gdy pierwiastki równania charakterystycznego

$$(1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p) = 0 \quad (1.61)$$

znajdują się poza okręgiem jednostkowym, tzn.  $|z_i| > 1$  dla  $1 \leq i \leq p$ .

- model ARMA(p,q) spełnia warunki stacjonarności i odwracalności jednocześnie, wtedy i tylko wtedy, gdy dla równań (1.60, 1.61) moduły wszystkich pierwiastków są większe od jedności.

Własności stacjonarności i odwracalności są szczegółowo przedstawione w pracy [3].

## 1.7. Model ARIMA

W praktyce często pojawiają się przypadki procesów niestacjonarnych. Przydatne wtedy okazuje się skorzystanie z modelu ARMA rozszerzonego o operację różnicowania. Modele te zapoczątkowane zostały przez Boxa i Jenkinsa(1976) i nazywane są zintegrowanymi modelami autoregresyjnymi ze średnią ruchomą (autoregressive integrated moving average - ARIMA(p,d,q))[3]. Są narzędziem wykorzystywanym do modelowania szeregów stacjonarnych lub sprowadzania szeregów niestacjonarnych do stacjonarnych. Modele ARIMA poza własnościami AR i MA, wykorzystują technikę różnicowania, która pozwala to na eliminację trendu i niestacjonarności. Model ARIMA(p,d,q) opisujemy przy pomocy następujących składników: ([4], [13], [32])

- autoregresja AR rzędu  $p \in \mathbb{N}$ ;
- średnia ruchoma MA rzędu  $q \in \mathbb{N}$ ;
- integracja I stopnia  $d \in \mathbb{N}$ .

Elementy szeregu ARIMA(p,d,q), gdzie  $p, d, q \in \mathbb{N}$ , są określone za pomocą równania:

$$\Delta^d x_t = \sum_{i=1}^p \alpha_i \Delta^d x_{t-i} + \varepsilon_t - \sum_{j=1}^q b_j \varepsilon_{t-j} \quad (1.62)$$

$t \in \mathbb{N}$ ,  $\Delta^d$  jest operatorem różnicowania,

$$\Delta x_t = x_t - x_{t-1} = (1 - B)x_t \quad (1.63)$$

$$\Delta^{d+1} x_t = \Delta(\Delta^d x_t) = (1 - B)^{d+1} x_t \quad (1.64)$$

Równanie 1.62 możemy również zapisać w postaci

$$\Phi_p(B) (1 - B)^d x_t = \Theta_q(B) \varepsilon_t, \quad (1.65)$$

gdzie  $\Phi_p(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ ,  $\Theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ .

**Twierdzenie 1.2.** Szereg  $\{x_t\}_{t \in \mathbb{N}}$  jest szeregiem  $ARIMA(p, d, q)$ , wtedy i tylko wtedy, gdy szereg  $\{\Delta^d x_t\}_{t \in \mathbb{N}}$  jest szeregiem  $ARMA(p, q)$ .

Dobór modelu  $ARIMA(p, d, q)$  jest związany z obecnością parametru  $d \in \mathbb{N}$ , który nazywany jest stopniem integracji. Reprezentuje on liczbę koniecznych do wykonania operacji różnicowania, aby szereg sprowadzić do szeregu stacjonarnego. Ze względu na często występującą niestacjonarność szeregu, modele  $ARIMA$  mają znacznie szersze zastosowanie w praktyce, niż wcześniej omówione modele  $AR$ ,  $MA$ ,  $ARMA$ . [25]

## 1.8. Testy pierwiastka jednostkowego

Istnieją różne czynniki wskazujące na niestacjonarność szeregu. Wpływ tych czynników powoduje, że średnia i wariancja ewoluują w czasie. W przypadku szeregów stacjonarnych wartości te utrzymują się na stałym poziomie. Również czynnik sezonowości, powoduje, że własność stacjonarności nie jest spełniona. Do weryfikacji stacjonarności szeregu wykorzystujemy testy pierwiastka jednostkowego [25]. Poniżej przedstawione zostaną klasyczne testy dotyczące weryfikacji stacjonarności w szeregach czasowych.

### 1.8.1. Rozszerzony test Dickeya-Fullera

W klasycznym teście Dickey-Fullera (DF) elementy szeregu czasowego modelujemy za pomocą modelu  $AR(1)$ , natomiast własność stacjonarności weryfikujemy analizując parametr strukturalny modelu  $AR(1)$  [7]. Rozszerzony test Dickeya-Fullera (Augmented Dickey-Fuller test) uwzględnia dodatkowe elementy, które pozwalają na redukcję autokorelacji w szeregu  $\{x_t\}_{t \in \mathbb{N}}$ . Elementy szeregu  $\{x_t\}_{t \in \mathbb{N}}$  określamy wprowadzając dodatkowo do  $AR(1)$ , zmienne objaśniające jako zmienne objaśniane opóźnione w czasie, mianowicie: [16]

$$\Delta x_t = \kappa x_{t-1} + \sum_{i=1}^k \alpha_i \Delta x_{t-i} + \varepsilon_t \quad (1.66)$$

gdzie  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  jest ciągiem niezależnych zmiennych losowych o rozkładzie  $N(0, \sigma^2)$ .

Moc rozszerzonego testu Dickeya-Fullera (Augmented Dickey-Fuller test) w przeciwieństwie do testu DF nie spada w przypadku autokorelacji pomiędzy zaburzeniami. W celu wyeliminowania autokorelacji w szeregu  $\{x_t\}_{t \in \mathbb{N}}$ , dobieramy odpowiednie przesunięcie  $k$ , opisujące rząd autoregresji, przy założeniu:

$$k < \left\lceil 4 \left( \frac{N}{100} \right)^{0.25} \right\rceil \quad (1.67)$$

gdzie  $\lceil \cdot \rceil$  oznacza część całkowitą.

Na poziomie istotności  $0 < \alpha < 1$  tworzymy hipotezę roboczą:

—  $H_0 : \kappa = 0$ , szereg jest niestacjonarny

oraz hipotezę alternatywną:

—  $H_1 : \kappa \in (-2, 0)$ , szereg jest stacjonarny

Statystyka testowa

$$DF = \frac{\hat{\kappa}}{S(\kappa)} \quad (1.68)$$

gdzie  $\hat{\kappa}$  jest estymatorem parametru  $\kappa$ , natomiast  $S(\kappa)$  oznacza odchylenie standardowe tego parametru, ma rozkład Dickey-Fullera. Wartość rozkładu można znaleźć w [16],[10], [1] .

### 1.8.2. Test KPSS

Weryfikację niestacjonarności szeregu czasowego również przeprowadzamy za pomocą testu Kwiatkowskiego-Phillipsa-Schmidta-Shina (KPSS), gdzie hipoteza zerowa zakłada brak istnienia pierwiastka jednostkowego. Elementy szeregu  $\{x_t\}_{t \in \mathbb{N}}$  przedstawiamy za pomocą równania:

$$x_t = \beta t + r_t + \varepsilon_t \quad (1.69)$$

gdzie  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  jest ciągiem niezależnych zmiennych losowych o rozkładzie  $N(0, \sigma_u^2)$ , zaś  $\{r_t\}_{t \in \mathbb{N}}$  jest szeregiem opisujących błądzenie losowe  $r_t = r_{t-1} + u_t$ , gdzie przez  $\{u_t\}_{t \in \mathbb{N}}$  rozumiemy ciąg niezależnych zmiennych losowych o rozkładzie  $N(0, \sigma^2)$ . Zależności w teście KPSS określone są następująco: [28]

- jeżeli  $\sigma_u = 0$ , to wartości procesu  $\{r_t\}_{t \in \mathbb{N}}$  są stałe w czasie ( $r_t = \text{const}$ , dla każdego  $t \in \mathbb{N}$ ), to  $r_t$  jest stałą w modelu 1.69, zatem elementy szeregu  $\{x_t\}_{t \in \mathbb{N}}$  są opisane poprzez sumę stałej  $r$  i trendu deterministycznego  $r + \beta t$  oraz składnika losowego  $\varepsilon_t$ . Wobec powyższego elementy szeregu  $\{x_t\}_{t \in \mathbb{N}}$  są zintegrowane w stopniu zero ( $\{x_t\}_{t \in \mathbb{N}} \sim I(0)$ ) ;
- jeśli  $\sigma_u^2 > 0$ , to jeden ze składników równania 1.69 jest procesem błądzenia losowego  $\{r_t\}_{t \in \mathbb{N}}$ , co implikuje niestacjonarność szeregu  $\{x_t\}_{t \in \mathbb{N}}$  jako suma elementów procesu niestacjonarnego  $\{r_t\}_{t \in \mathbb{N}}$  oraz stacjonarnego  $\{\varepsilon_t\}_{t \in \mathbb{N}}$ ;

Test Kwiatkowskiego-Phillipsa-Schmidta-Shina przeprowadzamy w sposób następujący. Na poziomie istotności  $0 < \alpha < 1$  tworzymy hipotezę roboczą

—  $H_0 : \sigma_u^2 = 0$  - stacjonarność szeregu czasowego  $\{x_t\}_{t \in \mathbb{N}}$  ( $\{x_t\}_{t \in \mathbb{N}} \sim I(0)$  )

wobec hipotezy alternatywnej

—  $H_1 : \sigma_u^2 > 0$  - niestacjonarność szeregu czasowego  $\{x_t\}_{t \in \mathbb{N}}$  ( $\{x_t\}_{t \in \mathbb{N}} \sim I(0), d \geq 1$ )

Dobieramy odpowiednie  $k$ , opisujące rząd autoregresji, przy założeniu:

$$k < \left\lceil 12 \left( \frac{N}{100} \right)^{0.25} \right\rceil \quad (1.70)$$

gdzie  $\lceil \cdot \rceil$  oznacza część całkowitą. Statystyka testowa dana jest wzorem:

$$\eta = \frac{\sum_{t=1}^n S_t^2}{n^2 S^2(k)} \quad (1.71)$$

gdzie  $S_t = \sum_{i=1}^t \varepsilon_i$ ,  $S^2(k) = \frac{1}{n} \left( \sum_{i=1}^n \varepsilon_i^2 + 2 \sum_{s=1}^k w(s, k) \sum_{t=s+1}^n \varepsilon_t \varepsilon_{t-s} \right)$ , dla rzędu opóźnienia  $k$  oraz wag  $w(s, k) = 1 - \frac{s}{k+1}$ . Wartości krytyczne dla statystyki testowej 1.71 można znaleźć np. w [17], [16].

### 1.8.3. Test Phillipsa-Perrona

Założenia dla testu Phillipsa-Perrona (PP) są identyczne jak w przypadku testu ADF. W przypadku hipotezy zerowej zakładamy niestacjonarność szeregu  $\{x_t\}_{t \in \mathbb{N}}$ , czyli istnienie pierwiastka jednostkowego, natomiast hipoteza alternatywna zakłada, że szereg jest stacjonarny. Dla testu Phillipsa-Perrona odpowiada modelowi AR(1) dla szeregu  $\{x_t\}$ .

$$\Delta x_t = \varphi + \beta x_{t-1} + \varepsilon_t \quad (1.72)$$

Test PP redukuje serial correlation i heteroskedastyczność w  $\varepsilon_t$  poprzez modyfikację testów statystycznych  $t_\beta$  i  $T\hat{\beta}$ . Wspomniane zmodyfikowane statystyki oznaczamy poprzez  $Z_t$  oraz  $Z_\beta$  i opisujemy przez równania [36]

$$Z_t = \left( \frac{\hat{\sigma}^2}{\hat{\lambda}^2} \right)^{1/2} \cdot t_{\beta=0} - \frac{1}{2} \left( \frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \cdot \left( \frac{T \cdot SE(\hat{\beta})}{\hat{\sigma}^2} \right) \quad (1.73)$$

$$Z_\beta = T\hat{\beta} - \frac{1}{2} \frac{T^2 \cdot SE(\hat{\beta})}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2) \quad (1.74)$$

Przez  $\hat{\sigma}^2$  oraz  $\hat{\lambda}^2$  rozumiemy estymatory parametrów odpowiednio  $\sigma^2$  i  $\lambda^2$  wyrażonych wzorami:

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E[\varepsilon_t^2]$$

$$\lambda^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^T E[T^{-1} S_T^2]$$

gdzie  $S_T = \sum_{t=1}^T \varepsilon_t$ . Przy założeniu hipotezy zerowej, że  $\beta = 0$ , statystyki  $Z_t$ ,  $Z_\beta$  testu PP mają taki sam rozkład jak statystyka testowa  $DF$  (1.68).



## 1.9. Model ARFIMA

Popularną klasą modeli z długą pamięcią są ułamkowo zintegrowane procesy autokoregresji i średniej ruchomej (ARFIMA). Przedstawione przez Granger i Joyeux(1980), Hosking(1981) zdobyły popularność ze względu na możliwość zbadania długookresowych zależności między obserwacjami. Modele tej klasy są naturalnym rozwinięciem modeli ARIMA. Czynniki odpowiadające za ruchomą średnią i autokoregresję są identyczne jak dla modeli ARIMA, natomiast stopień integracji nie należy do zbioru liczb rzeczywistych.

**Definicja 1.14** (Model ARFIMA). Model ARFIMA( $p, d, q$ ), gdzie  $p \geq 1$ ,  $d \in \mathbb{R}$  i  $q \geq 1$ , jest opisany za pomocą wyrażenia: [20]

$$\Phi_p(B)\Delta^d x_t = \Theta_q(B)\varepsilon_t, \quad (1.75)$$

lub

$$\Phi_p(B)(1 - B)^d x_t = \Theta_q(B)\varepsilon_t, \quad (1.76)$$

gdzie  $t \in \mathbb{N}$ ,  $B$  oznacza operator opóźnień,  $\Delta^d$  jest operatorem różnicowania,  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  oznacza biały szum. Operatory odpowiadające za autokoregresję i ruchomą średnią:

$$\Phi_p(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \quad (1.77)$$

$$\Theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (1.78)$$

są wielomianami opóźnień odpowiedniego stopnia  $p$  i  $q$ , których pierwiastki leżą na zewnątrz koła jednostkowego.

Z równania 1.76 prostym przekształceniu otrzymujemy równanie opisujące integrację białego szumu:

$$\Phi_p(B) = \Theta_q(B)(1 - B)^{-d}\varepsilon_t, \quad (1.79)$$

Ponadto symbol  $(1 - B)^{-d}$  jest zdefiniowany jako

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} B^j = \sum_{j=0}^{\infty} \eta_j B^j \quad (1.80)$$

gdzie  $\Gamma$  jest funkcją gamma Eulera

Przeanalizujemy teraz własności parametru  $d$  ze względu na równania 1.76 i 1.79. Jeśli parametr  $d$  jest bliski wartości  $\frac{1}{2}$ , to wszystkie współczynniki  $\eta_j$  są dodatnie i zbiegają do zera, wraz ze zwiększaniem  $d$ . Z równania 1.79, wnioskujemy, że w modelu ARFIMA(0,  $d$ , 0), obserwacja  $x_t$  zależy nie tylko od obecnej wartości białego szumu  $\{\varepsilon_t\}_{t \in \mathbb{N}}$ , ale także od wszystkich poprzednich wartości. Wynika z tego, że w szeregach ARFIMA(0,  $d$ ,  $q$ ) z  $d \approx \frac{1}{2}$  nawet obserwacje będące odległe od siebie, są znacząco

skorelowane między sobą. Wskazuje to na występowanie zjawiska z długą pamięcią w modelach ARFIMA [5]. Podczas modelowania metodą ARFIMA, poszukujemy takiego parametru  $d \in (-0.5, 0.5)$  różnicującego szereg tak, by uchwycić zjawisko długiej pamięci. Istnienie niezerowej wartości parametru  $d$  implikuje występowanie zjawiska długiej pamięci, zaś im bardziej przybliży się on do  $|0.5|$ , tym większa jest siła danego zjawiska [26], [21]. Parametr  $d = H - \frac{1}{2}$  liniowo zależy od wartości współczynnika Hursta.

**Twierdzenie 1.3.** *Niech  $\{x_t\}_{t \in \mathbb{N}}$  będzie procesem klasy ARFIMA*

$$x_t = (1 - B)^d \varepsilon_t \quad (1.81)$$

gdzie  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  jest ciągiem niezależnych zmiennych losowych o rozkładzie  $N(0, \sigma^2)$  oraz  $\sum_{k=0}^{\infty} \frac{|\Gamma(k+d)|}{|\Gamma(d)| \cdot k!} < \infty$ . Wtedy [34]

1. Jeśli  $d < 0.5$ , to szereg  $x_t$  jest stacjonarny,
2. Jeśli  $d > -0.5$ , to szereg  $x_t$  jest odwracalny,
3. Jeśli  $d \in (-0.5, 0.5)$ , to funkcja ACF jest wyrażona jako:

$$\rho_k = \frac{\Gamma(-d+1)\Gamma(k+d)}{\Gamma(d)\Gamma(k-d+1)}, \quad k \in \mathbb{N}, \quad (1.82)$$

gdzie  $\rho_k \simeq k^{2d-1}$ , przy  $k \rightarrow \infty$ . Dodatkowo,  $\sigma^2 = \frac{\Gamma(-2d+1)}{[\Gamma(-d+1)]^2}$ .

4. Jeśli  $0 < d < 0.5$  to szereg  $x_t$  posiada własność długiej pamięci.

**Twierdzenie 1.4.** *Niech  $\{x_t\}_{t \in \mathbb{N}}$  będzie procesem klasy ARFIMA (zdefiniowanym w równaniu 1.76). Wtedy: [34]*

1. Jeśli  $d < 0.5$  i wszystkie pierwiastki  $\varphi_i = 0$  (dla  $i = 1, 2, \dots, p$ ) znajdują się poza okręgiem jednostkowym, to szereg  $x_t$  jest stacjonarny,
2. Jeśli  $d > 0.5$  i wszystkie pierwiastki  $\theta_i = 0$  (dla  $i = 1, 2, \dots, p$ ) znajdują się poza okręgiem jednostkowym, to szereg  $x_t$  jest odwracalny,
3. Jeśli  $d \in (-0.5, 0.5)$  oraz  $d \neq 0$ , to funkcja ACF spełnia:

$$\rho_k \simeq k^{2d-1}, \quad \text{przy } k \rightarrow \infty. \quad (1.83)$$

Korzystając z własności funkcji gamma, (zgodnie z [6]) możemy opisać własności modelu ARFIMA(p,d,q):

— autokowariancja

$$\gamma_k = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \cdot \frac{\Gamma(k+d)}{\Gamma(1+k-d)} \quad (1.84)$$

— autokorelacja

$$\rho_k = \frac{\Gamma(1-d)}{\Gamma(d)} \cdot \frac{\Gamma(k+d)}{\Gamma(1+k-d)} \quad (1.85)$$

Dla przesunięcia czasowego  $k \in \mathbb{N} \setminus \{0\}$  i parametru  $d \in (-0.5, 0.5)$ .

Identyfikacji i estymacji parametrów modelu ARFIMA(p,d,q), dokonujemy korzystając z poniższego schematu: [29]

1. Estymujemy parametr  $d$  modelu ARIMA(p,d,q), oznaczając estymator poprzez  $\hat{d}$ .
2. Obliczamy  $u_t = (1 - B)^{\hat{d}} x_t$ .
3. Identyfikujemy i estymujemy parametry  $\varphi_1, \dots, \varphi_p$  dla operatora autoregresji  $\phi_p(B)$  oraz  $\theta_1, \dots, \theta_q$  dla operatora ruchomej średniej  $\Theta_q(B)$  modelu ARMA(p,q), gdzie  $\phi_p(B)u_t = \theta_q(B)z_t$ , korzystając z kryterium Akaikego.
4. Obliczamy  $z_t = \frac{\varphi_p(B)}{\theta_q(B)} u_t$ .
5. Stosując algorytm R/S wyznaczamy stopień integracji szeregu  $\{z_t\}_{t \in \mathbb{N}}$ , gdzie  $z_t = (1 - B)^{-\tilde{d}} \varepsilon_t$ .

Jako stopień integracji w modelu ARFIMA(p,d,q) przyjmujemy  $d = \hat{d} + \tilde{d}$



## Rozdział 2

### Identyfikacja zachowania wybranych cen surowców za pomocą modeli ARFIMA

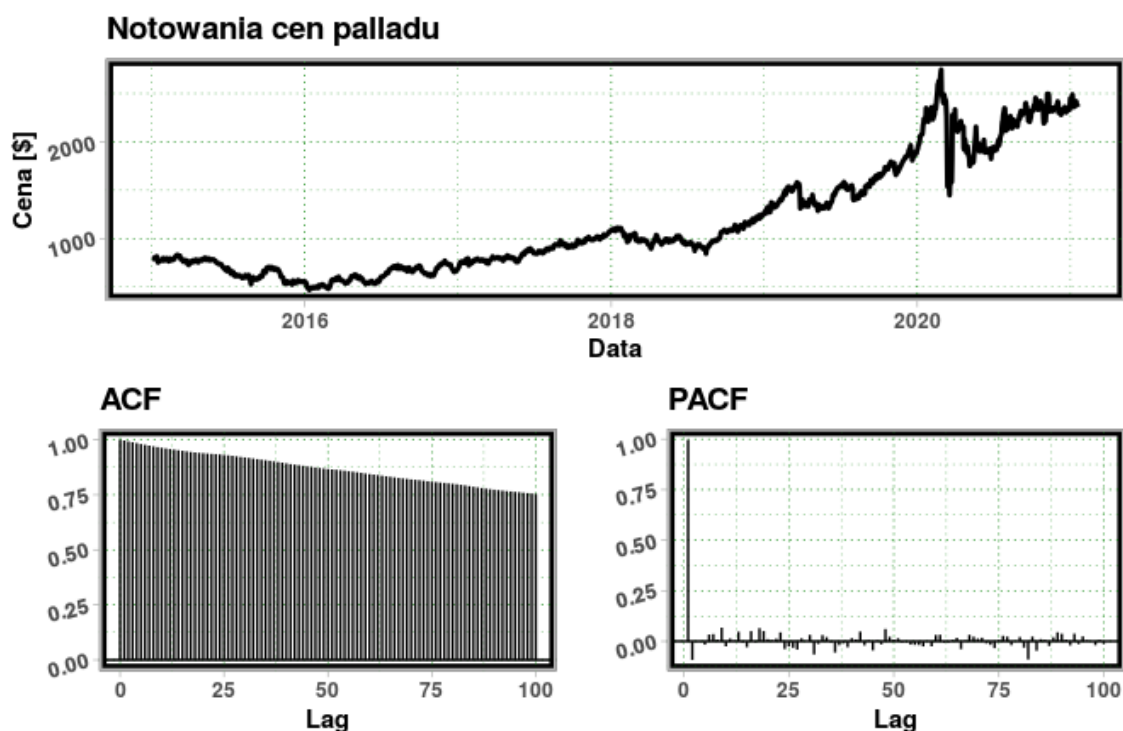
Dane wykorzystane w pracy zostały wczytane z serwisu *Yahoo Finance*[37] przy użyciu funkcji *get.hist.quote()* z biblioteki *tseries*. Wykresy zostały stworzone przy użyciu funkcji *ggplot* z biblioteki *ggplot2*. Konkretnie surowce zostały wybrane na podstawie analizy współczynnika Hursta, który odpowiada za istnienie bądź brak długiej pamięci szeregu. Estymację wykładnika Hursta przeprowadzono wykorzystując zaimplementowaną funkcję *hurstexp* z biblioteki *pracma*. Wartość z przedziału  $(0.5, 1)$  wskazuje na przypadek z długą pamięcią. Modele klasy ARFIMA zostały wykorzystane do modelowania i predykcji wybranych cen surowców.

W celu zbadania związków pomiędzy wartościami szeregu posłużymy się funkcją autokorelacji oraz funkcją autokorelacji cząstkowej. Skorzystamy z zaimplementowanych funkcji *acf* i *pacf* z biblioteki *stats*. Wyniki przedstawimy w formie wykresów, z których odczytamy wartości współczynnika korelacji dla różnych przesunięć pomiędzy elementami szeregu. Następnie dokonamy analizy stacjonarności szeregu. Jest to niezbędne do oceny wpływu długookresowego zachowania szeregu w przeszłości na jego przebieg w przyszłości. [25]. Testy pierwiastka jednostkowego ADF, KPSS i PP, przeprowadzimy korzystając z wbudowanych w bibliotece *tseries* funkcji odpowiednio *adf.test*, *kpss.test* oraz *pp.test*. Za pomocą funkcji *arfima()* z biblioteki *forecast*. Testy normalności reszt przeprowadzono korzystając z testów Shapiro-Wilka, Lillieforsa, Andersona-Darlinga oraz Cramera von Misesa. Użyto przy tym zaimplementowanych funkcji odpowiednio *shapiro.test()*, *lillie.test()*, *ad.test()*, *cvm.test()* z biblioteki *nor-test*. Dodatkowo za pomocą testu Ljunga-Boxa zbadano niezależność reszt.

## 2.1. Identyfikacja i predykcja cen palladu

### 2.1.1. Odczyt danych, analiza stacjonarności

Z serwisu *Yahoo Finance*[37] zostały pobrane dane od 2015-01-02 do 2021-01-29 dotyczące cen palladu (Palladium). W celu zbadania jakości powstałego modelu, szereg został podzielony na zbiór uczący i testowy. Obserwacje od 2015-05-01 do 2021-01-19 (1503 notowania) zostały wykorzystane do uczenia modelu. Dane od 2021-01-20 do 2021-01-29 (6 notowań) posłużyły do zbadania jakości modelu, czyli porównania czy przewidywane wartości są bliskie wartościom rzeczywistym. Na podstawie notowań dokonano identyfikacji zachowania cen palladu. Pierwszym etapem analizy jest zbadanie zależności między wartościami szeregu. Rysunek 2.1 przedstawia notowania cen palladu oraz wartości funkcji autokorelacji oraz korelacji cząstkowej.



Rysunek 2.1: Notowania cen palladu oraz wykresy ACF i PACF

Z wykresu funkcji autokorelacji możemy odczytać, że wartości współczynnika korelacji  $\rho_\tau$  maleją hiperbolicznie wraz ze wzrostem przesunięcia  $\tau$ . Na wykresie widzimy duże wartości funkcji ACF, co może wskazywać na występowanie trendu, bądź integracji w szeregu czasowym. Na wykresie PACF dostrzegamy, że dla niezerowych przesunięć między elementami szeregu wartości oscylują blisko zera. Zarówno funkcja ACF jak i PACF wskazują na występowanie długiej pamięci w szeregu. Dodatkowym potwierdzeniem jest wysoka wartość współczynnika Hursta wynosi 0.8807604.

Do weryfikacji stacjonarności szeregu cen palladu zastosowano testy pierwiasta jednostkowego. Przyjęto poziom istotności równy  $\alpha = 0.05$ . Wartość p.value dla testu KPSS wynosi  $p = 0.01$ , więc hipotezę zerową, która zakłada spełnienie stacjonarności szeregu należy odrzucić na korzyść hipotezy alternatywnej. Dla testów ADF oraz PP otrzymaliśmy odpowiednio  $p = 0.3076$  i  $p = 0.09577$ , zatem nie mamy podstaw do odrzucenia hipotezy zerowej, która zakłada niestacjonarność szeregu. Wartości statystyk oraz p.value dla poszczególnych testów przedstawia tabela 2.1

**Tabela 2.1:** Wyniki testów badania stacjonarności szeregu cen palladu

Test	Statystyka	p.val
ADF	-2.6396	0.3076
KPSS	16.117	0.01
PP	-18.39	0.09577

### 2.1.2. Dopasowanie modelu ARFIMA

Do oszacowania parametrów wykorzystano funkcję *arfima()* z biblioteki *forecast*. [8] Logarytm z funkcji wiarygodności wynosi  $-7475.7509858$ , natomiast tabela 2.2 przedstawia wartości parametrów strukturalnych oraz wyniki testów istotności tych parametrów.

**Tabela 2.2:** Wartości parametrów dla modelu ARFIMA

	Estimate	Std. Error	z.value	Pr(> z )
d	0.499407	0.000836	597.556557	0.000e+00
ma.ma1	-0.725706	0.025230	-28.763176	6.198e-182
ma.ma2	-0.501037	0.030075	-16.659703	2.573e-62
ma.ma3	-0.414031	0.024749	-16.729485	7.993e-63
ma.ma4	-0.416607	0.028931	-14.400023	5.173e-47
ma.ma5	-0.222724	0.029359	-7.586149	3.296e-14

Z tabeli 2.2 odczytujemy, że na poziomie istotności  $\alpha = 0.05$  hipotezy robocze dotyczące braku istotności parametrów strukturalnych należy odrzucić na korzyść hipotez alternatywnych, zatem stopień integracji oraz współczynniki ruchomej śred-

niej w modelu ARFIMA istotnie różnią się od zera. Zachowania cen palladu możemy przedstawić za pomocą modelu ARFIMA(0,0.499,5) postaci

$$(1 - B)^{0.499}x_t = (1 + 0.726B + 0.501B^2 + 0.414B^3 + 0.417B^4 + 0.223B^5)\varepsilon_t, \quad (2.1)$$

gdzie odchylenie standardowe  $\sigma$  szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  wynosi 34.9. Rysunek 2.2 przedstawia wartości rzeczywiste szeregu  $\{x_t\}_{1 \leq t \leq n}$  oraz wartości dopasowane szeregu  $\{\hat{x}_t\}_{1 \leq t \leq n}$  oszacowane na podstawie modelu ARFIMA (2.1).

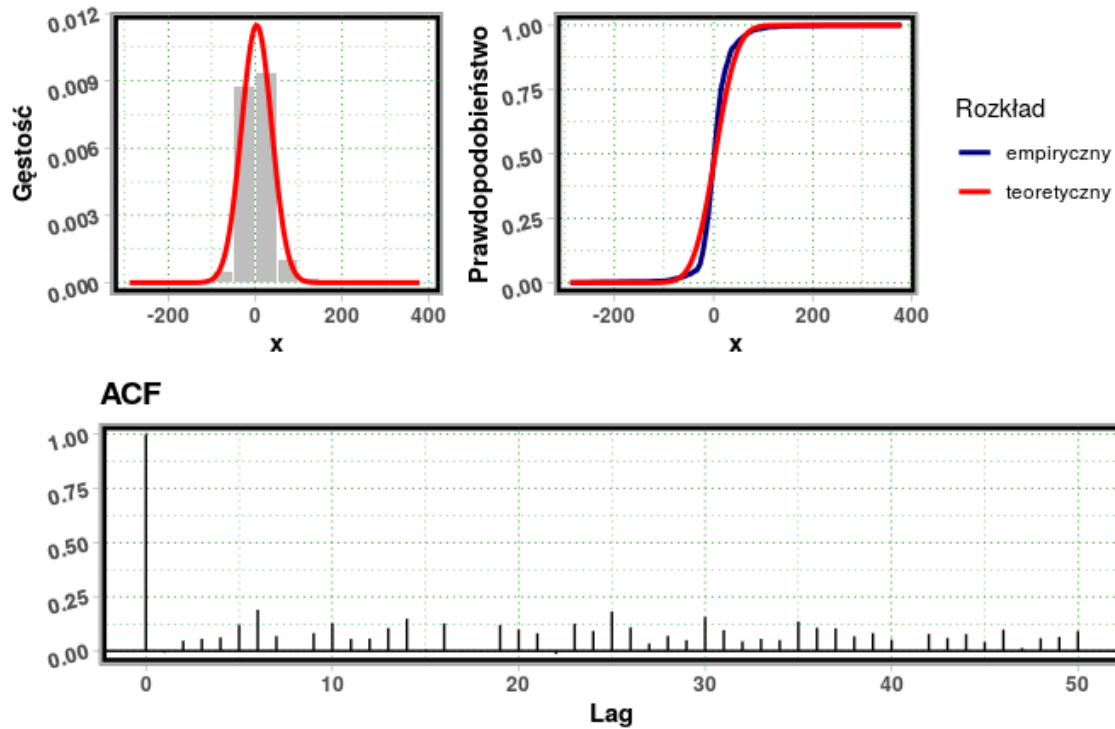


**Rysunek 2.2:** Wartości rzeczywiste cen palladu oraz dopasowane za pomocą modelu ARFIMA

Poniżej przedstawiona zostanie analiza reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$ , gdzie  $\varepsilon_t = x_t - \hat{x}_t$  dla  $1 \leq t \leq n$ . Rysunek 2.3 przedstawia histogram wraz z estymatorem jądrowym funkcji gęstości, rozkład empiryczny i teoretyczny  $N(0, \sigma^2)$ , dodatkowo wartości funkcji autokorelacji.

Do badania normalności rozkładu reszt zastosowano testy: Shapiro-Wilka, Lillieforsa, Andersona-Darlinga oraz Cramera von Misesa. Hipoteza robocza zakłada że elementy szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  mają rozkład normalny ([28], [31]). Wartości statystyk oraz wartości p.value przedstawia tabela 2.3.





**Rysunek 2.3:** Rozkład szeregu reszt oraz wartości funkcji autokorelacji

**Tabela 2.3:** Wyniki testów badania normalności rozkładu reszt

Test	Statystyka	p.val
Shapiro-Wilka	0.8168646	2.662e-38
Lillieforsa	0.1223802	2.206e-61
Andersona-Darlinga	52.5395886	3.700e-24
Cramera von Misesa	9.2126024	7.370e-10

Z przeprowadzonych analiz wynika, że na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą normalności rozkładu szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  należy odrzucić na korzyść hipotezy alternatywnej, tzn. elementy szeregu reszt nie spełniają założenia o normalności rozkładu. Rysunek 2.3 również zawiera wykres funkcji autokorelacji szeregu reszt. Wartości bezwzględne funkcji autokorelacji ( $\rho_\tau^\varepsilon = \text{corr}(\varepsilon_t, \varepsilon_{t+\tau})$ ), dla  $1 \leq \tau \leq 50$  nie przekraczają wartości 0.25. Na wykresie nie widać wyraźnych związków pomiędzy elementami szereg reszt, co nie oznacza jednak, że one nie występują. Dlatego niezbędne jest przeprowadzenie testu Ljunga-Boxa, służącego do badania losowości szeregów czasowych. Za pomocą testu Ljung-Box dokonano analizy niezależności elementów szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$ .

Dla przesunięć  $1 \leq \tau \leq 10$  weryfikowano hipotezę roboczą, która zakłada niezależność reszt. Wartości współczynnika korelacji  $\rho_\tau^\varepsilon$ , wartości statystyki Ljunga-Boxa

$$\chi_{(\tau)}^2 = n(n+2) \sum_{j=1}^{\tau} \frac{(\rho_j^\varepsilon)^2}{n-j}, \quad (2.2)$$

która ma rozkład  $\chi^2$  o  $\tau$  stopniach swobody oraz wartości p.value testu Ljunga-Boxa przedstawia tabela 2.4.

**Tabela 2.4:** Wartości współczynnika korelacji, wartości statystyki  $\chi_{(\tau)}^2$  oraz p.value dla testu Ljunga-Boxa

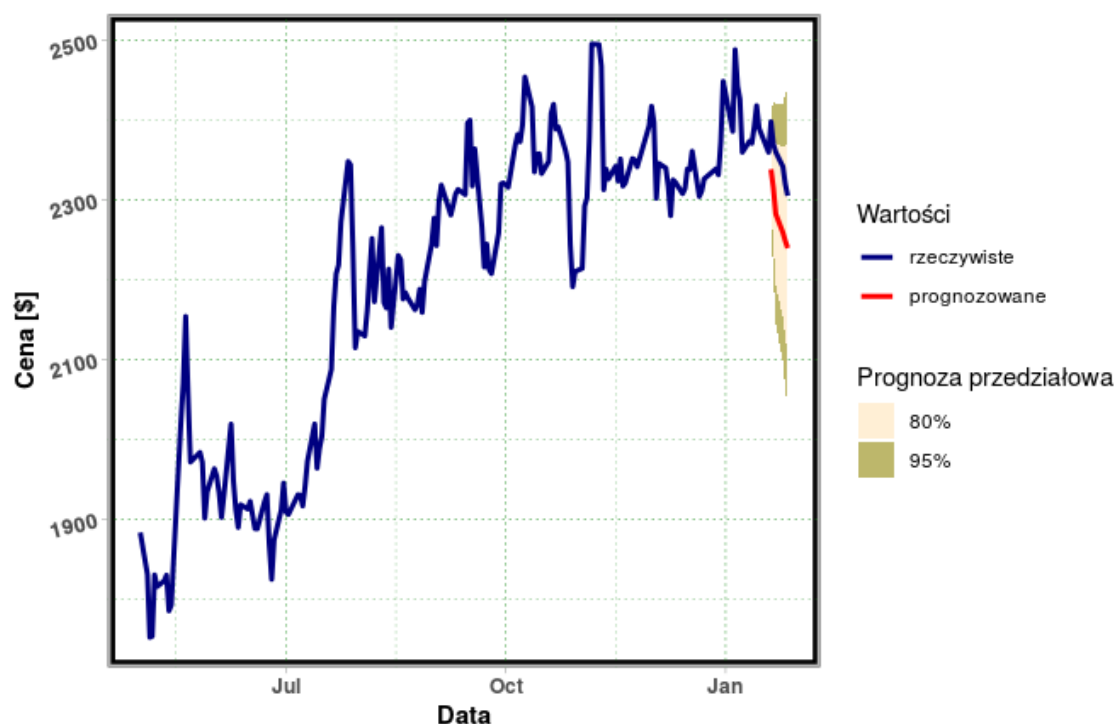
$\tau$	$\rho_\tau^\varepsilon$	$\chi_{(\tau)}^2$	p.val
1	-0.0085346	0.1096965	0.7404899
2	0.0468446	3.4166990	0.1811646
3	0.0554492	8.0532437	0.0449240
4	0.0622202	13.8951782	0.0076372
5	0.1212801	36.1059192	0.0000009
6	0.1899685	90.6360650	0.0000000
7	0.0682357	97.6762985	0.0000000
8	-0.0010871	97.6780865	0.0000000
9	0.0818426	107.8196311	0.0000000
10	0.1281892	132.7161612	0.0000000

Również na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą niezależności reszt należy odrzucić na korzyść hipotezy alternatywnej, zatem elementy szeregu reszt są skorelowane.

### 2.1.3. Predykcja dla modelu ARFIMA

Rysunek 2.4 przedstawia wartości notowań cen palladu od 2020-05-01 do 2021-01-29 - kolor granatowy. W celu weryfikacji prognoz również na wykresie zaznaczono wartości rzeczywiste (kolor granatowy) oraz prognozowane (kolor czerwony) w okresie od 2021-01-20 do 2021-01-29. Dodatkowo wyznaczono przedziały ufności dla poziomu 80% oraz 95%.

Na rysunku 2.4 widać, że wartości prognozowane notowań mieszczą się w przedziale ufności dla poziomu 95% . Tabela 2.5 przedstawia granice przedziałów ufności oraz wartości rzeczywiste i prognozowane dla notowań od 2021-01-20 do 2021-01-29.



Rysunek 2.4: Notowania cen palladu

Tabela 2.5: Przedziały ufności, wartości rzeczywiste i prognozowane cen palladu

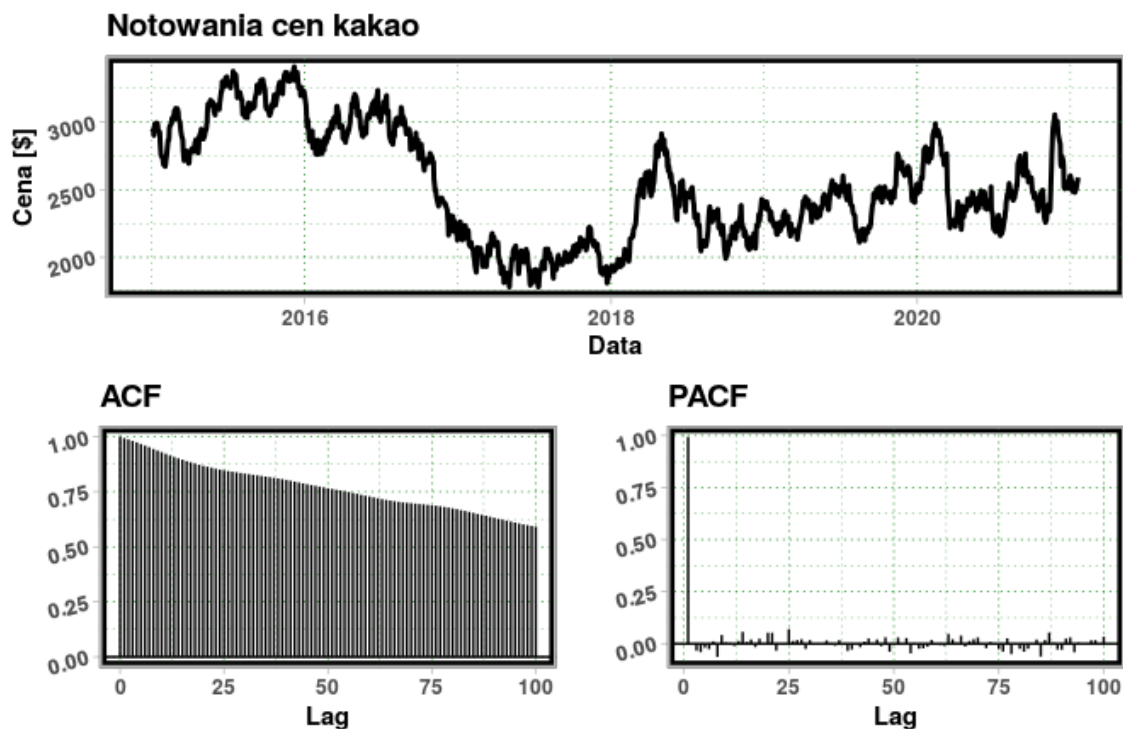
Data	l80	l95	u80	u95	Rzeczywiste	Prognozowane
2021-01-20	2293.607	2269.937	2383.035	2406.705	2397.9	2338.321
2021-01-21	2244.296	2206.863	2385.719	2423.151	2369.8	2315.007
2021-01-22	2192.588	2145.052	2372.184	2419.721	2359.9	2282.386
2021-01-25	2153.241	2097.275	2364.684	2420.650	2341.6	2258.963
2021-01-26	2127.671	2063.687	2369.410	2433.394	2319.7	2248.540
2021-01-27	2107.755	2037.886	2371.726	2441.595	2305.2	2239.740

## 2.2. Identyfikacja i predykcja cen kakao

### 2.2.1. Odczyt danych, analiza stacjonarność

Z serwisu *Yahoo Finance*[37] zostały pobrane dane od 2015-01-02 do 2021-01-27 dotyczące cen kakao (Cocoa). W celu zbadania jakości powstałego modelu, szereg został podzielony na zbiór uczący i testowy. Obserwacje od 2015-05-01 do 2021-01-19 (1510 notowań) zostały wykorzystane do uczenia modelu. Dane od 2021-01-20 do 2021-01-29 (8 notowań) posłużyły do zbadania jakości modelu, czyli porównania czy

przewidywane wartości są bliskie wartościom rzeczywistym. Na podstawie notowań dokonano identyfikacji zachowania cen kakao. Pierwszym etapem analizy jest zbadanie zależności między wartościami szeregu. Rysunek 2.5 przedstawia notowania cen kakao oraz wartości funkcji autokorelacji oraz korelacji cząstkowej.



**Rysunek 2.5:** Notowania cen kakao oraz wykresy ACF i PACF

Z wykresu funkcji autokorelacji możemy odczytać, że wartości współczynnika korelacji  $\rho_\tau$  maleją hiperbolicznie wraz ze wzrostem przesunięcia  $\tau$ . Na wykresie widzimy duże wartości funkcji ACF, co może wskazywać na występowanie trendu, bądź integracji w szeregu czasowym. Na wykresie PACF dostrzegamy, że dla niezerowych przesunięć między elementami szeregu wartości oscylują blisko zera. Zarówno funkcja ACF jak i PACF wskazują na występowanie długiej pamięci w szeregu. Dodatkowym potwierdzeniem jest wysoka wartość współczynnika Hursta wynosi 0.8701133.

Do weryfikacji stacjonarności szeregu cen palladu zastosowano testy pierwiasta jednostkowego. Przyjęto poziom istotności równy  $\alpha = 0.05$ . Wartość p.value dla testu KPSS wynosi  $p = 0.01$ , więc hipotezę zerową, która zakłada spełnienie stacjonarności szeregu należy odrzucić na korzyść hipotezy alternatywnej. Dla testów ADF oraz PP otrzymaliśmy odpowiednio 0.3518 i 0.4225, zatem nie mamy podstawy do odrzucenia hipotezy zerowej, która zakłada niestacjonarność szeregu. Wartości statystyk oraz p.value dla poszczególnych testów przedstawia tabela 2.6

**Tabela 2.6:** Wyniki testów badania stacjonarności szeregu cen kakao

Test	Statystyka	p.val
ADF	-2.5352	0.3518
KPSS	6.4071	0.01
PP	-12.32	0.4225

### 2.2.2. Dopasowanie modelu ARFIMA

Do oszacowania parametrów wykorzystano funkcję *arfima()* z biblioteki *forecast*. Logarytm z funkcji wiarygodności wynosi  $-7851.9403201$ . W tabeli 2.7 przedstawione zostały wartości parametrów strukturalnych oraz wyniki testów istotności tych parametrów.

**Tabela 2.7:** Wartości parametrów dla modelu ARFIMA

	Estimate	Std. Error	z.value	Pr(> z )
d	0.019316	0.021098	0.915548	3.599e-01
ar.ar1	0.992653	0.003321	298.869333	0.000e+00

Z tabeli 2.7 widzimy, że na poziomie istotności  $\alpha = 0.05$  nie ma podstaw do odrzucenia hipotezy roboczej, a współczynnik integracji nieistotnie różni się od zera, natomiast hipotezę roboczą o braku istotności współczynnika autoregresji należy odrzucić na korzyść hipotezy alternatywnej. Zachowania cen kakao przedstawiamy za pomocą modelu ARFIMA(1, 0.019, 0) postaci

$$(1 + 0.993B) \cdot (1 - B)^{0.019} x_t = \varepsilon_t \quad (2.3)$$

gdzie odchylenie standardowe  $\sigma$  szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  wynosi 43.9. Rysunek 2.6 przedstawia wartości rzeczywiste szeregu  $\{x_t\}_{1 \leq t \leq n}$  oraz wartości dopasowane szeregu  $\{\hat{x}_t\}_{1 \leq t \leq n}$  oszacowane na podstawie modelu ARFIMA (2.3).

Poniżej przedstawiona zostanie analiza reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$ , gdzie  $\varepsilon_t = x_t - \hat{x}_t$  dla  $1 \leq t \leq n$ . Rysunek 2.7 przedstawia histogram wraz z estymatorem jądrowym funkcji gęstości, rozkład empiryczny i teoretyczny  $N(0, \sigma^2)$ , dodatkowo wartości funkcji autokorelacji.



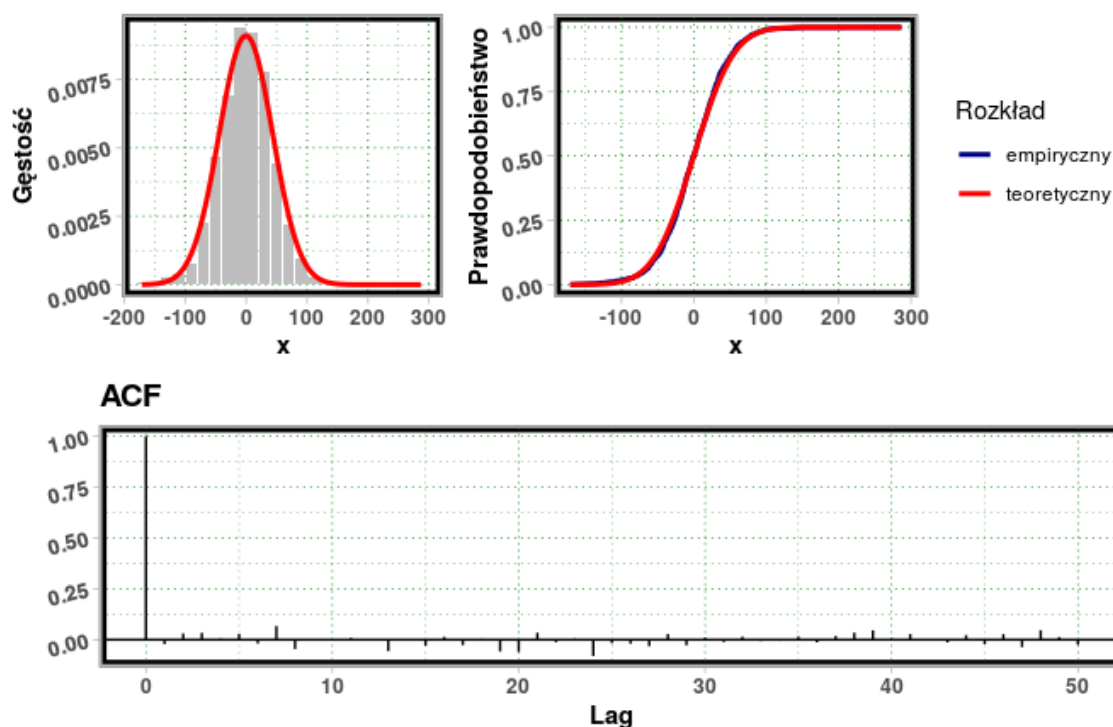
**Rysunek 2.6:** Wartości rzeczywiste cen kakao oraz dopasowane za pomocą modelu ARFIMA

Do badania normalności rozkładu reszt zastosowano testy: Shapiro-Wilka, Lillieforsa, Andersona-Darlinga oraz Cramera von Misesa. Hipoteza robocza zakłada że elementy szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  mają rozkład normalny. [28], [31] wartości statystyk oraz wartości p.value przedstawia tabela 2.8.

**Tabela 2.8:** Wyniki testów badania normalności rozkładu reszt

Test	Statystyka	p.val
Shapiro-Wilka	0.9889779	2.838e-09
Lillieforsa	0.0251947	2.580e-02
Andersona-Darlinga	1.4652550	8.800e-04
Cramera von Misesa	0.1988733	5.554e-03

Z przeprowadzonych analiz wynika, że na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą normalności rozkładu szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  należy odrzucić na korzyść hipotezy alternatywnej, tzn. elementy szeregu reszt nie spełniają założenia o normalności rozkładu. Rysunek 2.7 również zawiera wykres funkcji autokorelacji szere-



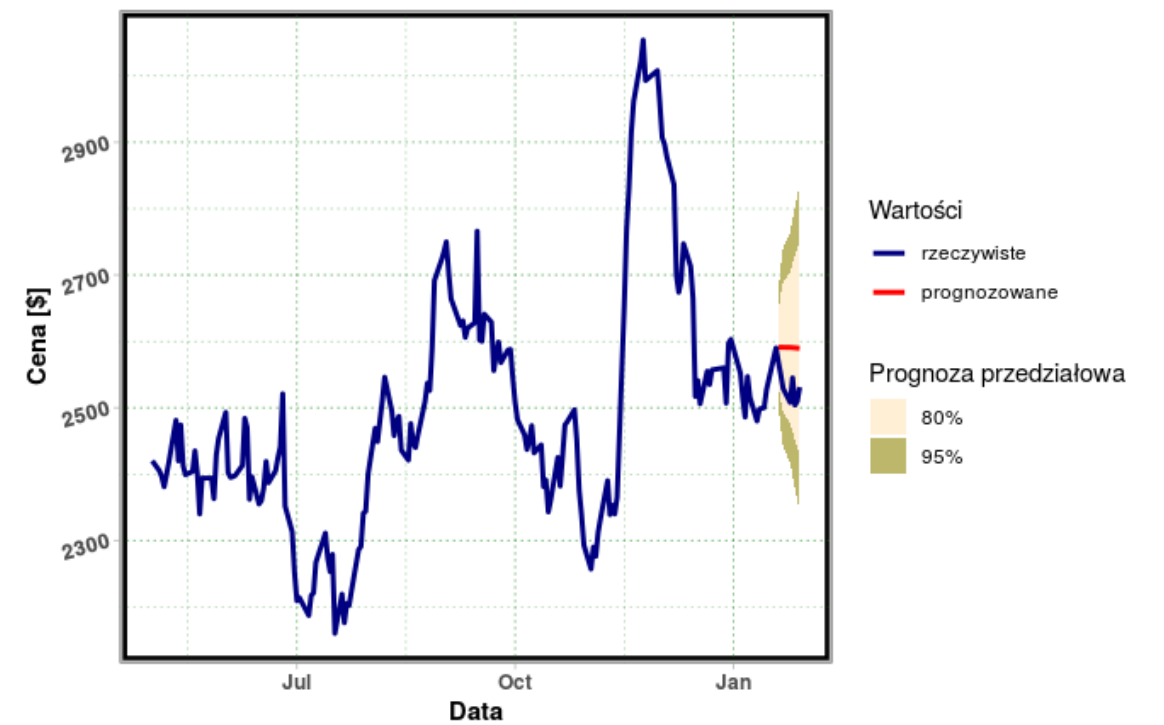
**Rysunek 2.7:** Rozkład szeregu reszt oraz wartości funkcji autokorelacji

gu reszt. Wartości bezwzględne funkcji autokorelacji  $\rho_\tau^\varepsilon = \text{corr}(\varepsilon_t, \varepsilon_{t+\tau})$ , dla  $1 \leq \tau \leq 50$  nie przekraczają wartości 0.1. Na wykresie nie widać związków pomiędzy autokorelacjami reszt, co nie oznacza jednak, że one nie występują. Dlatego niezbędne jest przeprowadzenie testu Ljunga-Boxa, służącego do badania losowości szeregów czasowych.

Za pomocą testu Ljung-Box dokonano analizy niezależności reszt  $\{\varepsilon\}_{1 \leq t \leq n}$ . Dla przesunięć  $1 \leq \tau \leq 10$  weryfikowano hipotezę roboczą, która zakłada niezależność reszt. Wartości współczynnika korelacji  $\rho_\tau^\varepsilon$ , wartości statystyki Ljunga-Boxa (2.2) oraz wartości p.value testu Ljunga-Boxa przedstawia tabela 2.4. Również na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą niezależności reszt należy odrzucić na korzyść hipotezy alternatywnej, zatem elementy szeregu reszt są skorelowane.

### 2.2.3. Predykcja dla modelu ARFIMA

Rysunek 2.8 przedstawia wartości notowań cen kakao od 2020-05-01 do 2021-01-29 - kolor granatowy. W celu weryfikacji prognoz również na wykresie zaznaczono wartości rzeczywiste (kolor granatowy) oraz prognozowane (kolor czerwony). Dodatkowo wyznaczono przedziały ufności dla poziomu 80% oraz 95%.



Rysunek 2.8: Notowania cen kakao

Na rysunku 2.8 widać, że wartości prognozowane notowań mieszczą się w przedziale ufności dla poziomu 80%. Tabela 2.9 przedstawia granice przedziałów ufności oraz wartości rzeczywiste i prognozowane dla notowań od 2021-01-20 do 2021-01-29.

Tabela 2.9: Przedziały ufności, wartości rzeczywiste i prognozowane cen kakao

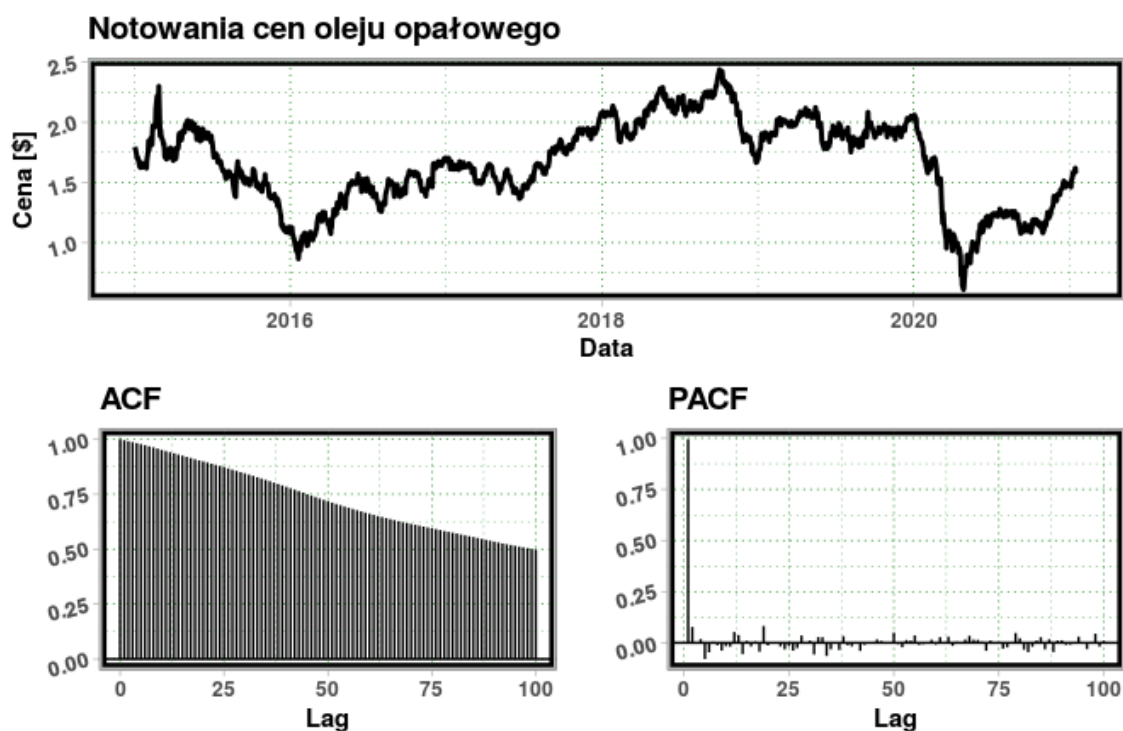
Data	l80	l95	u80	u95	Rzeczywiste	Prognozowane
2021-01-20	2293.607	2269.937	2383.035	2406.705	2397.9	2338.321
2021-01-21	2244.296	2206.863	2385.719	2423.151	2369.8	2315.007
2021-01-22	2192.588	2145.052	2372.184	2419.721	2359.9	2282.386
2021-01-25	2153.241	2097.275	2364.684	2420.650	2341.6	2258.963
2021-01-26	2127.671	2063.687	2369.410	2433.394	2319.7	2248.540
2021-01-27	2107.755	2037.886	2371.726	2441.595	2305.2	2239.740



## 2.3. Identyfikacja i predykcja cen oleju opałowego

### 2.3.1. Odczyt danych, analiza stacjonarności

Z serwisu *Yahoo Finance*[37] zostały pobrane dane od 2015-01-02 do 2021-01-29 dotyczące cen oleju opałowego (Heating Oil). W celu zbadania jakości powstałego modelu, szereg został podzielony na zbiór uczący i testowy. Obserwacje od 2015-05-01 do 2021-01-19 (1510 notowań) zostały wykorzystane do uczenia modelu. Dane od 2021-01-20 do 2021-01-29 (8 notowań) posłużyły do zbadania jakości modelu, czyli porównania czy przewidywane wartości są bliskie wartościom rzeczywistym. Na podstawie notowań dokonano identyfikacji zachowania cen oleju opałowego. Pierwszym etapem analizy jest zbadanie zależności między wartościami szeregu. Na podstawie notowań dokonano identyfikacji zachowania cen oleju opałowego (Heating Oil). Rysunek 2.9 przedstawia notowania cen palladu oraz wartości funkcji autokorelacji oraz korelacji cząstkowej.



**Rysunek 2.9:** Notowania cen palladu oraz wykresy ACF i PACF

Z wykresu funkcji autokorelacji możemy odczytać, że wartości współczynnika korelacji  $\rho_\tau$  maleją hiperbolicznie wraz ze wzrostem przesunięcia  $\tau$ . Na wykresie widzimy duże wartości funkcji ACF, co może wskazywać na występowanie trendu, bądź integracji w szeregu czasowym. Na wykresie PACF dostrzegamy, że dla niezerowych przesunięć między elementami szeregu wartości oscylują blisko zera. Zarówno funkcja

ACF jak i PACF wskazują na występowanie długiej pamięci w szeregu. Dodatkowym potwierdzeniem jest wysoka wartość współczynnika Hursta wynosi 0.8650253.

Do weryfikacji stacjonarności szeregu cen oleju opałowego zastosowano testy pierwiastka jednostkowego. Przyjęto poziom istotności równy  $\alpha = 0.05$ . Wartość p.value dla testu KPSS wynosi  $p = 0.01$ , więc hipotezę zerową, która zakłada spełnienie stacjonarności szeregu należy odrzucić na korzyść hipotezy alternatywnej. Dla testów ADF oraz PP otrzymaliśmy odpowiednio  $p = 0.5827$  i  $p = 0.7049$ , zatem nie mamy podstawy do odrzucenia hipotezy zerowej, która zakłada niestacjonarność szeregu. Wartości statystyk oraz p.value dla poszczególnych testów przedstawia tabela 2.10

**Tabela 2.10:** Wyniki testów badania stacjonarności reszt

Test	Statystyka	p.val
ADF	-1.9897	0.5827
KPSS	2.3688	0.01
PP	-7.2575	0.7049

### 2.3.2. Dopasowanie modelu ARFIMA

Do oszacowania parametrów wykorzystano funkcję *arfima()* z biblioteki *forecast*. [8] Logarytm z funkcji wiarygodności wynosi 2820.4364628, natomiast tabela 2.11 przedstawia wartości parametrów strukturalnych oraz wyniki testów istotności tych parametrów.

**Tabela 2.11:** Wartości parametrów dla modelu ARFIMA

	Estimate	Std. Error	z.value	Pr(> z )
d	0.499113	0.001248	399.900467	0.000e+00
ma.ma1	-0.513813	0.019852	-25.881869	1.066e-147
ma.ma2	-0.429278	0.022059	-19.460567	2.371e-84
ma.ma3	-0.325525	0.020885	-15.586495	8.993e-55
ma.ma4	-0.314377	0.022292	-14.102920	3.644e-45
ma.ma5	-0.195279	0.022983	-8.496574	1.953e-17

Z tabeli 2.11 odczytujemy, że na poziomie istotności  $\alpha = 0.05$  hipotezy robo-

cze dotyczące braku istotności parametrów strukturalnych należy odrzucić na korzyść hipotez alternatywnych, zatem stopień integracji oraz współczynniki ruchomej średniej w modelu ARFIMA istotnie różnią się od zera. Zachowania cen oleju opałowego przedstawiamy za pomocą modelu ARFIMA(0,0.499,5) postaci

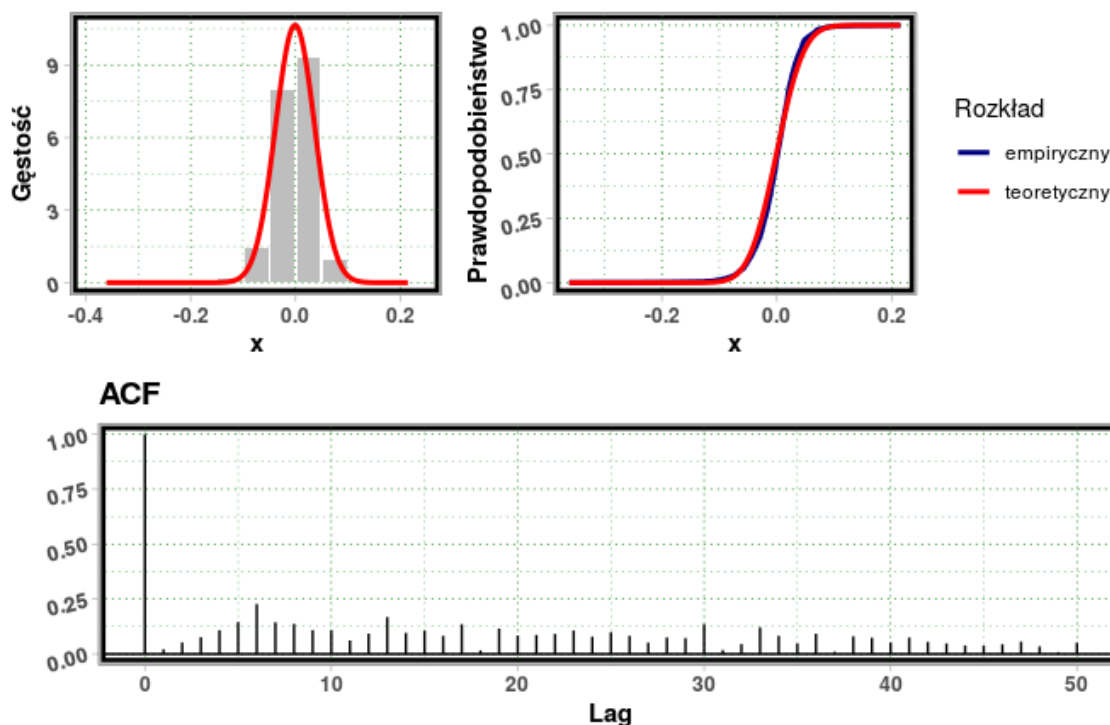
$$(1 - B)^{0.499}x_t = (1 + 0.514B + 0.429B^2 + 0.326B^3 + 0.314B^4 + 0.195B^5)\varepsilon_t \quad (2.4)$$

gdzie odchylenie standardowe  $\sigma$  szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  wynosi 0.0373. Rysunek 2.10 przedstawia wartości rzeczywiste szeregu  $\{x_t\}_{1 \leq t \leq n}$  oraz wartości dopasowane szeregu  $\{\hat{x}_t\}_{1 \leq t \leq n}$  oszacowane na podstawie modelu ARFIMA (2.4).



**Rysunek 2.10:** Wartości rzeczywiste cen oleju opałowego oraz dopasowanie za pomocą modelu ARFIMA

Poniżej przedstawiona zostanie analiza reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$ , gdzie  $\varepsilon_t = x_t - \hat{x}_t$  dla  $1 \leq t \leq n$ . Rysunek 2.11 przedstawia histogram wraz z estymatorem jądrowym funkcji gęstości, rozkład empiryczny i teoretyczny  $N(0, \sigma^2)$ , dodatkowo wartości funkcji autokorelacji.



**Rysunek 2.11:** Rozkład szeregu reszt oraz wartości funkcji autokorelacji

Do badania normalności rozkładu reszt zastosowano testy: Shapiro-Wilka, Lillieforsa, Andersona-Darlinga oraz Cramera von Misesa. Hipoteza robocza zakłada że elementy szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  mają rozkład normalny [28], [31]. Wartości statystyk oraz wartości p.value przedstawia tabela 2.13.

**Tabela 2.12:** Wyniki testów badania normalności rozkładu reszt

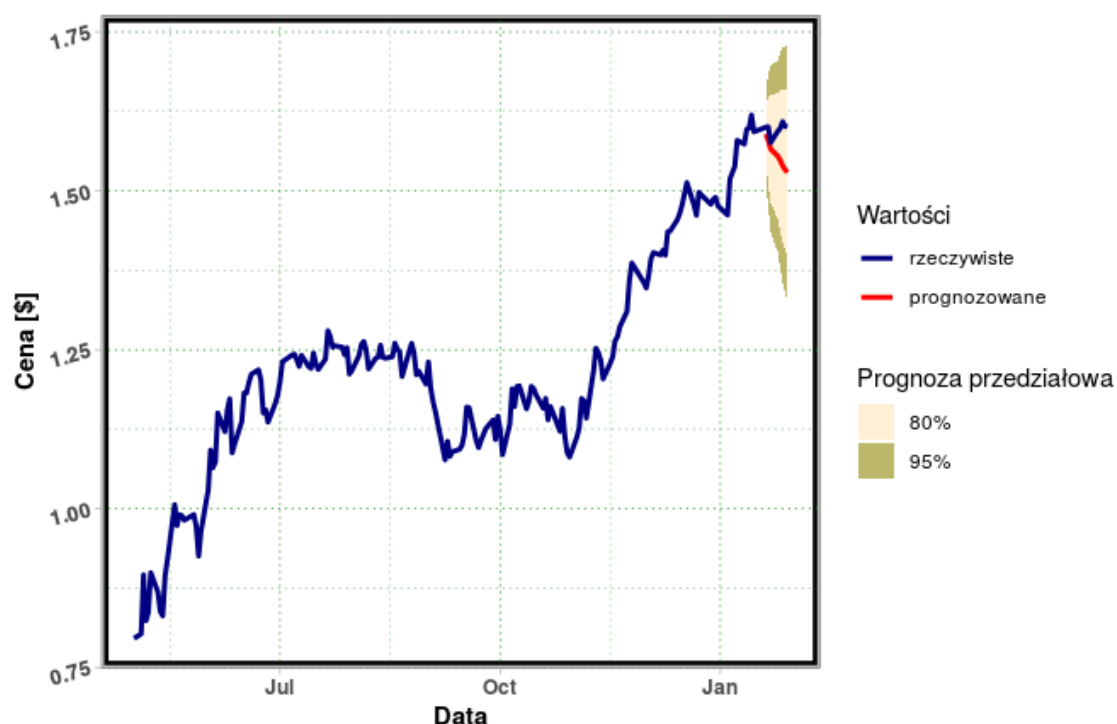
Test	Statystyka	p.val
Shapiro-Wilka	0.9423978	8.555e-24
Lillieforsa	0.0549442	1.669e-11
Andersona-Darlinga	9.8611319	7.698e-24
Cramera von Misesa	1.6488603	7.370e-10

Z przeprowadzonych analiz wynika, że na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą normalności rozkładu szeregu reszt  $\{\varepsilon_t\}_{1 \leq t \leq n}$  należy odrzucić na korzyść hipotezy alternatywnej, tzn. elementy szeregu reszt nie spełniają założenia o normalności rozkładu. Rysunek 2.11 również zawiera wykres funkcji autokorelacji szeregu reszt. Wartości bezwzględne funkcji autokorelacji  $\rho_\tau^\varepsilon = \text{corr}(\varepsilon_t, \varepsilon_{t+\tau})$ , dla  $1 \leq \tau \leq 50$

nie przekraczają wartości 0.25. Na wykresie nie widać wyraźnych związków pomiędzy elementami szeregu reszt, co nie oznacza jednak, że one nie występują. Dlatego niezbędne jest przeprowadzenie testu Ljunga-Boxa, służącego do badania losowości szeregów czasowych. Za pomocą testu Ljung-Box dokonano analizy niezależności reszt. Dla przesunięć  $1 \leq \tau \leq 10$  weryfikowano hipotezę roboczą, która zakłada niezależność reszt. Wartości współczynnika korelacji  $\rho_\tau^\varepsilon$ , wartości statystyki Ljunga-Boxa 2.2 oraz wartości p.value przedstawia tabela 2.4. Również na poziomie istotności  $\alpha = 0.05$  hipotezę roboczą dotyczącą niezależności reszt należy odrzucić na korzyść hipotezy alternatywnej, zatem elementy szeregu reszt są skorelowane.

### 2.3.3. Predykcja dla modelu ARFIMA

Rysunek przedstawia wartości notowań cen oleju opałowego (Heating Oil) od 2020-05-01 do 2021-01-20 - kolor granatowy. W celu weryfikacji prognoz również na wykresie zaznaczono wartości rzeczywiste (kolor granatowy) oraz prognozowane (kolor czerwony). Dodatkowo wyznaczono przedziały ufności dla poziomu 80% oraz 95%.



Rysunek 2.12: Notowania cen oleju opałowego

W celu zbadania jakości powstałego modelu, szereg został podzielony na zbiór uczący i testowy. Obserwacje od 2015-05-01 do 2021-01-19 zostały wykorzystane do uczenia modelu. Dane od 2021-01-20 do 2021-01-29 posłużyły do zbadania jakości modelu, czyli porównania czy przewidywane wartości są bliskie wartościom rzeczywistym.

Na rysunku 2.12 widać, że wartości prognozowane notowań mieszczą się w przedziale ufności dla poziomu 80% . Tabela 2.13 przedstawia granice przedziałów ufności oraz wartości rzeczywiste i prognozowane dla notowań od 2020-01-20 do 2021-01-29.

**Tabela 2.13:** Wartości współczynnika korelacji, wartości statystyki Ljunga-Boxa

Data	l80	l95	u80	u95	Rzeczywiste	Prognozowane
2021-01-20	1.541514	1.516112	1.637487	1.662890	1.6004	1.589501
2021-01-21	1.510905	1.474747	1.647511	1.683669	1.6006	1.579208
2021-01-22	1.480605	1.435525	1.650921	1.696001	1.5760	1.565763
2021-01-25	1.455779	1.403481	1.653364	1.705662	1.5939	1.554572
2021-01-26	1.436783	1.377841	1.659471	1.718412	1.5984	1.548127
2021-01-27	1.418287	1.354173	1.660516	1.724630	1.6089	1.539401
2021-01-28	1.406718	1.339578	1.660376	1.727516	1.6017	1.533547
2021-01-29	1.397814	1.328428	1.659958	1.729344	1.6004	1.528886

## Podsumowanie i wnioski

W pracy przedstawiono modele szeregów czasowych oraz modele z długą pamięcią. W rozdziale pierwszym omówiono podstawowe własności szeregów czasowych, testy pierwiastka jednostkowego, własności funkcji gamma oraz technikę różnicowania ułamkowego. W rozdziale drugim dokonano identyfikacji zachowania cen surowców przy użyciu modeli ARFIMA. Celem pracy była ocena przydatności modeli z długą pamięcią. Analizując wartości współczynnika Hursta wybrano do analizy takie surowce jak pallad, kakao i olej opałowy. Z wykresów funkcji ACF i PACF zauważyć można znaczące związki pomiędzy elementami szeregu, co jest niezbędne do użycia modeli z długą pamięcią. Stosując testy pierwiastka jednostkowego w każdym z przypadków otrzymano wyniki świadczące o niestacjonarności szeregu.

Dzięki zaimplementowanej funkcji *arfima()* w języku R dokonano oszacowania parametrów, a następnie przedstawiono modele opisujące zachowania cen odpowiednich surowców. Dla palladu, kakao i oleju opałowego otrzymano odpowiednio modele  $ARFIMA(0; 0.499; 5)$ ,  $ARFIMA(1; 0.019; 0)$  oraz  $ARFIMA(0; 0.499; 5)$ . Za pomocą testów Shapiro-Wilka, Lillieforsa, Andersona-Darlinga, Cramera von Misesa otrzymano brak normalności rozkładu reszt. Dodatkowo na podstawie testu Ljunga-Boxa odrzucono hipotezę o niezależności reszt w tych modelach.

Szeregi czasowe opisujące ceny surowców są złożone i podlegają wpływom wielu czynników. Dzięki wykorzystaniu danych z pięciu lat i wykorzystaniu zjawiska długiej pamięci otrzymaliśmy zadowalającą dokładność prognoz. Analizując wykresy porównania wartości rzeczywistych i prognozowanych możemy dostrzec, że modele były skuteczne w wysokich przedziałach ufności. Oczywiście dokładność otrzymanych wyników można zwiększyć przy użyciu wiedzy ekonomicznej i gospodarczej, jednak analiza szeregów czasowych pozwala na wstępne dostarczenie informacji o prawidłowościach występujących w danych. Stanowi to odpowiednią podstawę do dalszych analiz.





## Bibliografia

- [1] Alabi N., Are O., Taofik A., *Statistically Significant Relationships between Returns on FTSE 100, S and P 500 Market Indexes and Macroeconomic Variables with Emphasis on Unconventional Monetary Policy*
- [2] Beran J., *Statistics for Long-Memory Processes*,
- [3] Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M., *Time Series Analysis, Forecasting and Control*,
- [4] Brockwell P.J., Davis R.A., *Time series: Theory and Methods*, Springer – Verlag 1991r.
- [5] Burnecki K., Sikora G., *Identification and validation of stable ARFIMA processes with application to UMTS data*, Chaos, Solitons and Fractals, Elsevier (2017)
- [6] Contreras-Reyes J. E., Palma W., *Statistical Analysis of Autoregressive Fractionally Integrated Moving Average Models*,
- [7] Dickey D.A., Fuller W.A., *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*, Journal of the American Statistical Association, Volume 74, Issue 366 (Jun., 1979)
- [8] Fisher R.A., *On the Mathematical Foundations of Theoretical Statistics* Philosophical Transactions of the Royal Society A. 222 (594–604): 309–368.
- [9] Francq C., Makarova S., Zakoïan J.M., *A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test*, Econometrics, MDPI, 2007, 142 (1), pp.312. [ff10.1016/j.jeconom.2007.04.003](https://doi.org/10.1016/j.jeconom.2007.04.003) [ffhal-00501792f](https://doi.org/10.1016/j.jeconom.2007.04.003)
- [10] Fuller W. A. , *Introduction to Statistical Time Series*, New York: John Wiley and Sons (1976),
- [11] Geweke J., Porter-Hudak S., *The estimation and application of long memory time series models*, Journal of time series analysis (1983)
- [12] Granger C.W.J., Joyeux R., *An introduction to long-memory time series models and fractional differencing*, Journal of time series analysis (1980)
- [13] Hamilton J., *Time Series Analysis*, 1994
- [14] Hosking J. R. M., *Fractional Differencing*, Biometrika, Vol. 68, No.1 (Apr., 1981), 165-176
- [15] Hosking J. R. M., *Modeling persistence in hydrological time series using fractional differencing*, Water resources research (1984),

- [16] Kozłowski E., *Analiza i identyfikacja szeregów czasowych*, Lublin 2015
- [17] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y., *Testing the null hypothesis of stationarity against the alternative of a unit root*, Journal of Econometrics 54 (1992) 159-178. North-Holland
- [18] Kwiatkowski J., *Procesy z długą pamięcią i modele ARFIMA*,
- [19] Levy H., Lessman F., *Równania różnicowe skończone*, Warszawa 1966
- [20] Li W.K., McLeod A. I., *Fractional Time Series Modelling*,
- [21] Liu K., Chen Y., Zhang X., *An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs*,
- [22] Lo A.W., *Long-term memory in stock market prices*, Econometrica 1991, vol. 59 1279-1313.
- [23] Maddala G., Kim I., *Unit Roots, Cointegration and Structural Change*, CUP, Cambridge 1998
- [24] Newey W.K., West K.D., *A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*, Princeton University (1986)
- [25] Nielsen A., *Szeregi czasowe: Praktyczna analiza i predykcja z wykorzystaniem statystyki i uczenia maszynowego*, Helion 2020
- [26] Omekara C.O., Okereke O.E., Ukaegwu L.U., *Forecasting Liquidity Ratio of Commercial Banks in Nigeria*,
- [27] Phillips P. C. B., Perron P., *Testing for a unit root in time series regression*, Biometrika (1988), Vol. 75, 2, pp.335-46
- [28] Razali N.M., Wah Y.B., *Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests*, Journal of Statistical Modeling and Analytics (2011), Vol. 2 No. 1, 21-33
- [29] Reisen V., Abraham B., Lopes S., *Estimation of Parameters in ARFIMA Processes: A Simulation Study*,
- [30] Schwert G.W., *Tests for Unit Roots: A Monte Carlo Investigation*,
- [31] Shapiro S.S., Wilk M.B., *An Analysis of Variance Test for Normality*, Biometrika, Vol. 52, No. 3/4. (Dec., 1965), s. 591-611
- [32] Shumway R.H., Stoffer D.S., *Time Series Analysis and Its Applications: With R Examples*,
- [33] Weron A., Weron R., *Inżynieria finansowa*, Warszawa 1998
- [34] Woodward W.A., Gray H.L., Elliott A.C., *Applied Time Series Analysis With R*, CRC press (2017)
- [35] Zeug-Żebro K., *Badanie wpływu redukcji szumu na identyfikację dynamiki chaotycznej na przykładzie finansowych szeregów czasowych*,
- [36] Zivot E., Wang J., *Modeling Financial Time Series with S-PLUS®*  
Strony internetowe:
- [37] <https://finance.yahoo.com/commodities>

## Spis rysunków

1.1	Symulacje: biały szum oraz błędzenie losowe . . . . .	10
1.2	Wykres funkcji gamma . . . . .	14
1.3	Symulacja białego szumu oraz wykres integracji rzędu 0.35 białego szumu . . . .	17
2.1	Notowania cen palladu oraz wykresy ACF i PACF . . . . .	30
2.2	Wartości rzeczywiste cen palladu oraz dopasowane za pomocą modelu ARFIMA	32
2.3	Rozkład szeregu reszt oraz wartości funkcji autokorelacji . . . . .	33
2.4	Notowania cen palladu . . . . .	35
2.5	Notowania cen kakao oraz wykresy ACF i PACF . . . . .	36
2.6	Wartości rzeczywiste cen kakao oraz dopasowane za pomocą modelu ARFIMA .	38
2.7	Rozkład szeregu reszt oraz wartości funkcji autokorelacji . . . . .	39
2.8	Notowania cen kakao . . . . .	40
2.9	Notowania cen palladu oraz wykresy ACF i PACF . . . . .	41
2.10	Wartości rzeczywiste cen oleju opałowego oraz dopasowanie za pomocą modelu ARFIMA . . . . .	43
2.11	Rozkład szeregu reszt oraz wartości funkcji autokorelacji . . . . .	44
2.12	Notowania cen oleju opałowego . . . . .	45



## Spis tabel

2.1	Wyniki testów badania stacjonarności szeregu cen palladu . . . . .	31
2.2	Wartości parametrów dla modelu ARFIMA . . . . .	31
2.3	Wyniki testów badania normalności rozkładu reszt . . . . .	33
2.4	Wartości współczynnika korelacji, wartości statystyki $\chi^2_{(\tau)}$ oraz p.value dla testu Ljunga-Boxa . . . . .	34
2.5	Przedziały ufności, wartości rzeczywiste i prognozowane cen palladu . . . . .	35
2.6	Wyniki testów badania stacjonarności szeregu cen kakao . . . . .	37
2.7	Wartości parametrów dla modelu ARFIMA . . . . .	37
2.8	Wyniki testów badania normalności rozkładu reszt . . . . .	38
2.9	Przedziały ufności, wartości rzeczywiste i prognozowane cen kakao . . . . .	40
2.10	Wyniki testów badania stacjonarności reszt . . . . .	42
2.11	Wartości parametrów dla modelu ARFIMA . . . . .	42
2.12	Wyniki testów badania normalności rozkładu reszt . . . . .	44
2.13	Wartości współczynnika korelacji, wartości statystyki Ljunga-Boxa . . . . .	46



## **Załączniki**

1. Oświadczenie o oryginalności pracy i możliwości jej wykorzystania.
2. Opinia promotora na temat oryginalności pracy oraz w sprawie dopuszczenia do obrony pracy dyplomowej.
3. Potwierdzenie analizy antyplagiatowej.
4. Płyta CD z niniejszą pracą w wersji elektronicznej.





## Streszczenie (Summary)

### Wykorzystywanie modeli klasy ARFIMA do prognozowania cen surowców

Prognozowanie jest ważnym narzędziem ze względu na szerokie zastosowania. Praca zawiera metody analizy szeregów czasowych i głównie szeregów z długą pamięcią. Szeregi z długą pamięcią identyfikowane są za pomocą modeli klasy ARFIMA, które są uogólnieniem modeli ARIMA poprzez dopuszczenie ułamkowych wartości integracji. Modele klasy ARFIMA są zastosowane do predykcji cen palladu, kakao oraz oleju opałowego. Szeregi czasowe cen surowców są przeanalizowane przy użyciu zaimplementowanych funkcji środowiska R.

### *Application of ARFIMA models in forecasting raw materials prices*

*Forecasting is an important tool because of its wide range of applications. This study discusses time series analysis methods and focuses on those with long-memory process. The long-memory time series are identified by ARFIMA models, which are generalisation of ARIMA models by allowing for fractional degrees of integration. ARFIMA models are applied for predicting palladium, cocoa and heating oil prices. To predict raw materials prices, time series are estimated and analysed by using implemented statistical functions for R.*