

Magdalena Sudół, Jagoda Płócienniczak

Analiza danych rzeczywistych przy pomocy modelu ARMA - temperatura w Bostonie w latach 2013–2023.

24.01.2025 r.

1. Wstęp

1.1. Cel pracy

Celem niniejszego raportu jest opracowanie modelu ARMA, który najlepiej opisuje zmiany temperatury w Bostonie w latach 2013–2023. Ponadto, w pracy zostanie przeprowadzona ocena dopasowania modelu oraz szczegółowa analiza wyników uzyskanych w trakcie modelowania.

1.2. Model ARMA

Model ARMA, czyli model autoregresyjny średniej ruchomej, jest modelem, który ma szerokie zastosowanie w analizie szeregów czasowych. Szereg czasowy jest szeregiem ARMA(p,q), jeżeli jest stacjonarny w słabym sensie oraz dla każdego t spełnia następujące równanie:

$$X_t - \Phi_1 X_{t-1} - \dots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \dots + \Theta_q Z_{t-q},$$

gdzie: $Z_t \sim WN(0, \sigma^2)$, WN - biały szum (ciąg nieskorelowanych zmiennych losowych o tym samym rozkładzie).

1.2.1. Szereg stacjonarny w słabym sensie

Szereg czasowy X_t jest szeregiem stacjonarnym w słabym sensie, jeżeli spełnia następujące warunki:

1. $\mathbb{E}X_t = \text{const}$, $\forall t \in \mathbb{Z}$,
2. $\gamma(t, s) = \gamma(t - s, 0) = \gamma(t - s)$,

gdzie γ jest funkcją autokowariancji:

$$\gamma(t, s) = \gamma(t, t + h) = \text{cov}(X_t, X_{t+h}) = \mathbb{E}[X_t X_{t+h}] - \mathbb{E}[X_t] \cdot \mathbb{E}[X_{t+h}].$$

1.3. Charakterystyka danych

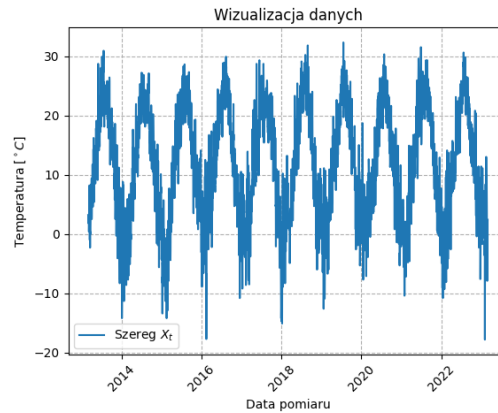
Dane wykorzystane do analizy zostały pobrane z platformy [Kaggle](https://www.kaggle.com/datasets/swaroopmeher/boston-weather-2013-2023)¹. Zbiór ten został przygotowany na podstawie informacji dostarczanych codziennie przez oficjalne stacje meteorologiczne, które monitorują i rejestrują warunki pogodowe. Ze zbioru danych wyodrębniono do analizy kolumny "time" oraz "tavg", w których zostały zapisane informacje dotyczące odpowiednio daty pomiaru oraz średniej temperatury powietrza danego dnia w stopniach Celsjusza. W raporcie będą brane pod uwagę dane zebrane w okresie od 01.03.2013 do 01.03.2023 (łącznie 3653 obserwacje).

1.4. Wizualizacja danych

Poniższy wykres przedstawia zależność temperatury od czasu - jest to wizualizacja badanego szeregu czasowego X_t . Widoczna jest wyraźna sezonowość, zatem aby umożliwić dopasowanie do

¹ <https://www.kaggle.com/datasets/swaroopmeher/boston-weather-2013-2023>

danych modelu ARMA konieczne będzie przeprowadzenie dekompozycji Wolda. Dodatkowo, nie zaobserwowano obecności wartości odstających, które mogłyby wpływać na jakość dopasowania modelu.



Rysunek 1: Zależność temperatury od dnia przeprowadzenia pomiaru

2. Przygotowanie danych do analizy

2.1. Jakość danych

W analizowanym zbiorze danych zidentyfikowano pojedynczą brakującą wartość temperatury. W celu jej uzupełnienia zastosowano interpolację poprzez obliczenie średniej temperatury z dnia poprzedzającego oraz dnia następującego po brakującym pomiarze. Pozostałe dane są kompletne i obejmują pełny okres od 2013 do 2023 roku. Dodatkowo zidentyfikowano najniższą oraz najwyższą zanotowaną temperaturę – wynosiły one odpowiednio $-17,8^{\circ}\text{C}$ oraz $32,3^{\circ}\text{C}$.

2.2. Dekompozycja szeregu czasowego

Dekompozycja szeregu czasowego polega na wyodrębnieniu kluczowych składowych, takich jak trend oraz sezonowość, w celu uzyskania resztowego procesu losowego o właściwościach stacjonarnych. Pamiętajmy, że stacjonarność szeregu jest niezbędna do prawidłowego dopasowania modelu ARMA.

2.2.1. Wykres ACF oraz PACF dla surowych danych

Zgodnie z wcześniejszymi założeniami, dane mogą zawierać komponent sezonowy. Aby zweryfikować tę hipotezę, należy przyjrzeć się funkcjom autokorelacji (ACF) oraz częściowej autokorelacji (PACF).

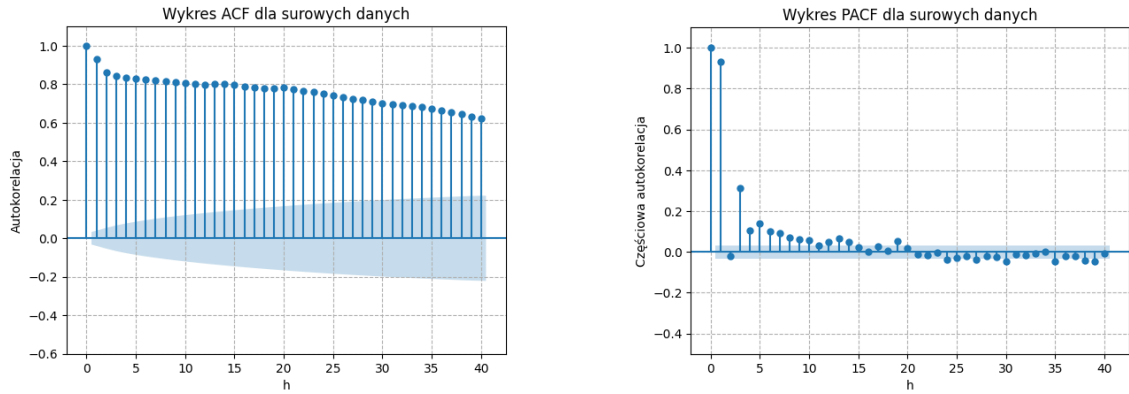
Wartości funkcji ACF i PACF można obliczyć przy użyciu poniższych wzorów, gdzie funkcję autokorelacji wyznaczamy na podstawie definicji funkcji autokowariancji.

a) próbkowa funkcja autokowariancji (ACVF):

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+h} - \bar{x})(x_t - \bar{x}), \quad \text{gdzie: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

b) próbkowa funkcja autokorelacji (ACF):

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$



Rysunek 2: Funkcje ACF i PACF

Na wykresie funkcji autokorelacji (ACF) można zauważyć stopniowy, ale wciąż powolny spadek wartości autokorelacji w miarę wzrostu opóźnienia (h). Taki kształt wykresu jest charakterystyczny dla danych, które mogą wykazywać trend lub sezonowość, ponieważ obserwacje w szeregu czasowym są silnie skorelowane z przeszłymi wartościami. W przypadku tego typu zachowania danych, przed przeprowadzeniem dalszej analizy, konieczne będzie przeprowadzenie transformacji szeregu w celu uzyskania stacjonarności.

2.2.2. Dekompozycja

Zgodnie ze wzorem dekompozycji, mamy:

$$X(t) = m(t) + s(t) + Y(t)$$

gdzie:

$m(t)$ - trend (funkcja deterministyczna)

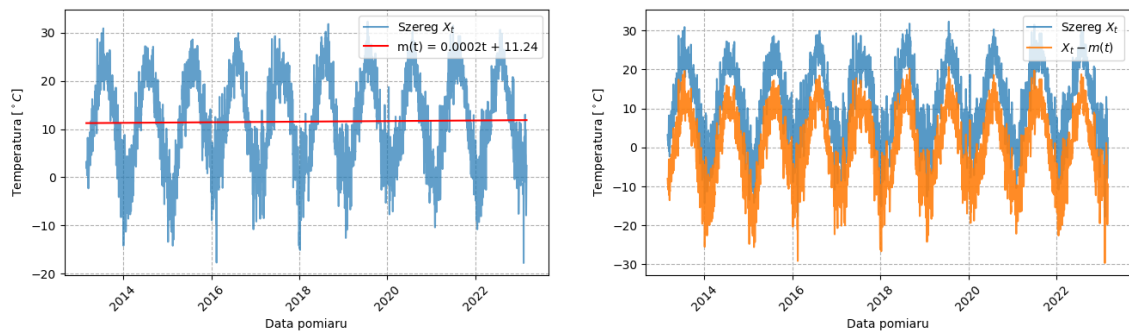
$s(t)$ - komponent sezonowy (funkcja deterministyczna)

$Y(t)$ - komponent opisujący (stacjonarny) proces losowy, często nazywany szumem

Dekompozycję możemy podzielić na trzy główne etapy:

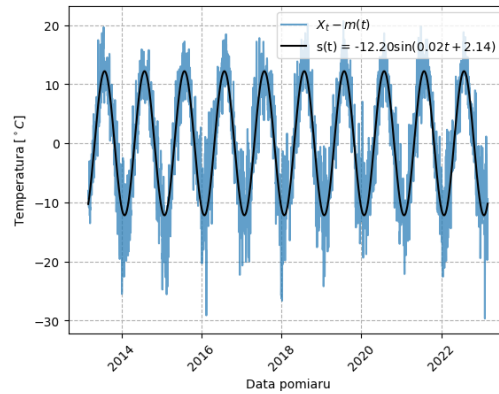
1. Wyodrębnienie i usunięcie trendu $m(t)$ z oryginalnych danych (w tym przypadku: prosta regresja liniowa - dopasowanie prostej do szeregu czasowego X_t).

Na prawym wykresie, oryginalny szereg X_t oznaczono kolorem niebieskim, a wartości szeregu po usunięciu trendu $m(t)$ przedstawiono kolorem pomarańczowym.



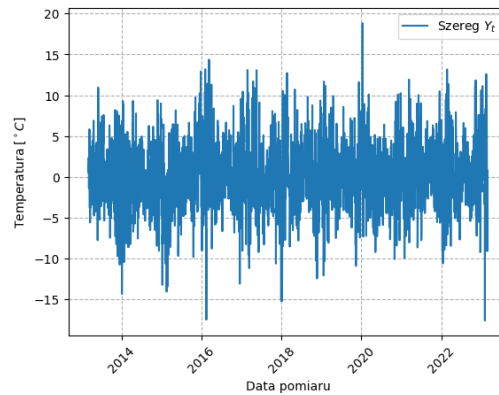
Rysunek 3: Składowa trendu $m(t)$ przed i po usunięciu z szeregu X_t

2. Identyfikacja komponentu sezonowego $s(t)$ na danych pozbawionych trendu (w tym przypadku: dopasowanie odpowiednich parametrów funkcji sinusoidalnej $s(t) = a \cdot \sin(bt + c)$).



Rysunek 4: Komponent sezonowy $s(t)$

3. Wyznaczenie komponentu reszty Y_t jako różnicy między oryginalnym szeregiem a składowymi trendu i sezonowości.

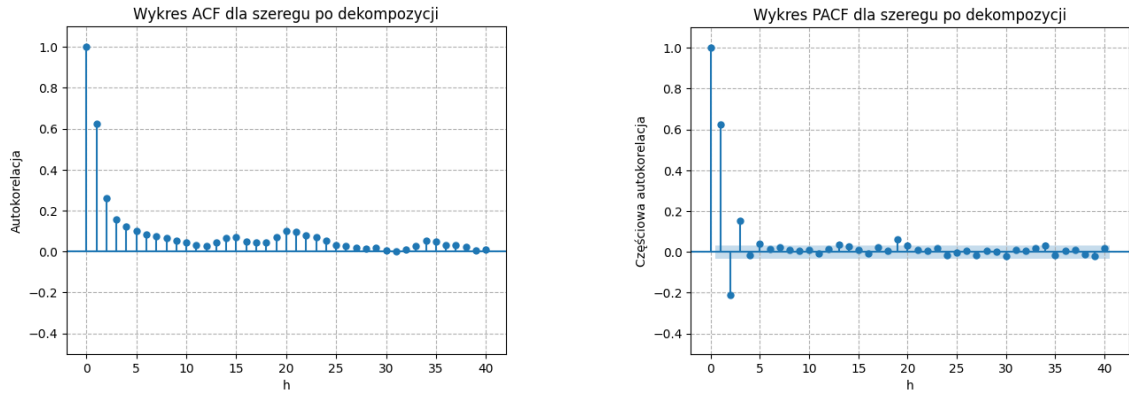


Rysunek 5: Szereg Y_t

Powyższy wykres znacznie bardziej przypomina szereg stacjonarny, w którym nie można zaobserwować wyraźnych oznak sezonowości ani trendu, co sugeruje, że proces dekompozycji przebiegł pomyślnie.

2.2.3. Wykres ACF oraz PACF dla uzyskanego szeregu

Wykresy funkcji ACF i PACF po dekompozycji wykazują istotne zmiany w strukturze danych. Wykres ACF wykazuje wyraźnie szybsze zbieganie wartości autokorelacji do zera, z wyższymi wartościami tylko dla kilku pierwszych opóźnień.



Rysunek 6: Funkcje ACF i PACF

2.2.4. Test ADF dla sprawdzenia stacjonarności szeregu po dekompozycji

Po dekompozycji sprawdzimy, czy otrzymany szereg jest stacjonarny. W tym celu wprowadzimy pojęcia stacjonarności szeregu oraz testu ADF.

a) Stacjonarność szeregu

Szereg ARMA(p, q) jest stacjonarny, jeżeli wszystkie pierwiastki wielomianu:

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

znajdują się poza okręgiem jednostkowym.

b) Test ADF

Rozszerzony test Dickeya-Fullera (Augmented Dickey-Fuller Test; ADF) jest testem statystycznym, używanym do sprawdzenia stacjonarności danego szeregu czasowego. Będziemy weryfikować następujące hipotezy:

H_0 : Szereg nie jest stacjonarny.

H_1 : Szereg jest stacjonarny.

Statystyka testowa jest obliczana na podstawie modelu:

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^p \delta_i \Delta X_{t-i} + \epsilon_i$$

gdzie:

X_t jest badanym szeregiem czasowym,

$\Delta X_t = X_t - X_{t-1}$,

α – wyraz wolny,

β – współczynnik trendu,

γ – kluczowy parametr testu,

p – parametr p w modelu AR(p),

δ_i – współczynniki dla dodatkowych różnic szeregu,

ϵ_i – składnik losowy.

W dalszej części należy obliczyć wartość parametru γ . Jeżeli $\gamma < 0$, oznacza to stacjonarność szeregu.

Dla szeregu po dekompozycji otrzymano wartość -10,159. Oznacza to, że odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną (otrzymany szereg Y_t jest stacjonarny).

3. Modelowanie danych przy pomocy ARMA

Skupmy się teraz na wyznaczeniu rzędu dopasowywanego modelu oraz estymacji jego parametrów.

3.1. Dobranie rzędu modelu (kryteria informacyjne)

Aby dobrać odpowiednie rzędy p i q modelu, rozważamy kilka par parametrów i dla każdej z nich obliczamy wartość danego kryterium informacyjnego. Najlepszym wyborem parametrów będą te, które prowadzą do uzyskania najmniejszej wartości kryterium. W tym celu skorzystamy z trzech różnych kryteriów informacyjnych.

3.1.1. Kryterium informacyjne Akaikego (Akaike Information Criterion, AIC)

Wartość zwracaną przez kryterium informacyjne Akaikego dla parametrów p i q oblicza się według wzoru:

$$AIC(p, q) = 2(p + q) - 2\ln(L),$$

gdzie: L - maksymalna wartość funkcji wiarygodności.

3.1.2. Bayesowskie kryterium informacyjne (Bayesian Information Criterion, BIC)

Wartość zwracaną przez bayesowskie kryterium informacyjne oblicza się według wzoru:

$$BIC = k\ln(n) - 2\ln(L),$$

gdzie:

- k - liczba parametrów,
- n - liczba obserwacji,
- L - maksymalna wartość funkcji wiarygodności.

Kryterium to, uwzględniając liczbę obserwacji, kładzie nacisk na prostotę modelu.

3.1.3. Kryterium informacyjne Hannana-Quinna (Hannan-Quinn Information Criterion, HQIC)

Kryterium informacyjne Hannana-Quinna jest określane wzorem:

$$HQIC = -2\ln(L) + 2k\ln(\ln(n)),$$

gdzie:

- k - liczba parametrów,
- n - liczba obserwacji,
- L - maksymalna wartość funkcji wiarygodności

Kryterium to dobrze sprawdza się w przypadku dużych zbiorów danych.

3.1.4. Wyniki

Po obliczeniu wartości wszystkich trzech kryteriów informacyjnych dla każdej pary parametrów, uzyskano jednoznaczne wyniki. Wszystkie kryteria wskazały, że najlepszym modelem do opisu danych jest model ARMA(2,2). W tabeli poniżej znajdują się dokładne wyniki otrzymane dla każdego kryterium informacyjnego. Najmniejsze wartości zaznaczono kolorem czerwonym.

p	q	AIC	BIC	HQIC
2	2	18637,794062	18675,013886	18651,048506
1	2	18655,183531	18686,200051	18666,228901
2	1	18659,323432	18690,339952	18670,368802
1	1	18661,407956	18686,221172	18670,244252
0	2	18702,924067	18727,737283	18711,760363

Tabela 1: Wartości kryteriów informacyjnych dla różnych par parametrów p i q

3.2. Estymacja parametrów modelu

W tej sekcji dokonano estymacji wartości współczynników modelu: Φ i Θ . Zostały do tego wykorzystane trzy metody.

3.2.1. Metoda przestrzeni stanów (statespace)

Metoda przestrzeni stanów polega na przedstawieniu równania ARMA jako modelu przestrzeni stanów, a następnie estymacji ukrytych stanów i maksymalizacji funkcji wiarygodności modelu w odniesieniu do obserwacji, co prowadzi do numerycznego wyznaczenia optymalnych wartości współczynników.

3.2.2. Metoda największej wiarygodności estymacji innowacji (Maximum Likelihood Estimation of Innovations)

Metoda największej wiarygodności estymacji innowacji to metoda, której celem jest oszacowanie parametrów modelu, które maksymalizują funkcję wiarygodności, z wykorzystaniem innowacji (reszt) modelu. Innowacje są błędami modelu, które nie są możliwe do przewidzenia na podstawie poprzednich obserwacji. Funkcja wiarygodności jest tworzona na podstawie założeń o rozkładzie innowacji. Parametry modelu ARMA są wyznaczone w taki sposób, aby zminimalizować różnicę między wartościami obserwowanymi a przewidywanymi przy uwzględnieniu modelu innowacji.

3.2.3. Metoda Hannan-Rissanen

Metoda Hannan-Rissanen polega na wyestymowaniu dla danych współczynników modelu autoregresyjnego AR(p), gdzie p jest dobierane przy pomocy kryteriów informacyjnych. Następnie estymuje się współczynniki modelu MA(q). Na samym końcu następuje korekta, aby jak najdokładniej dopasować współczynniki modelu ARMA(p, q).

3.2.4. Wyniki

	statespace	innovations_mle	Hannan-Rissanen
Φ_1	1,253302	1,253313	1,084936
Φ_2	-0,287910	-0,287929	-0,223862
Θ_1	-0,465374	-0,465383	-0,295183
Θ_2	-0,414144	-0,414129	-0,343620

Tabela 2: Wartości parametrów dla różnych metod estymacji

Otrzymane wyniki zamieszczono w tabeli powyżej. Na jej podstawie zobaczyć można, że metody statespace oraz innovations_mle dały zbliżone wyniki. Metoda Hannan-Rissanen zwróciła wyniki różniące się od pozostałych o około 0,1-0,2.

Ponieważ wyniki uzyskane dwiema pierwszymi metodami są niemal identyczne, uznamy je za najbardziej dokładne. Różnice między metodami statespace i innovations_mle pojawiają się dopiero na piątym miejscu po przecinku, dlatego możemy bez obaw przyjąć parametry oszacowane za pomocą metody statespace jako reprezentatywne dla badanego modelu.

Ostatecznie do danych został dopasowany następujący model ARMA:

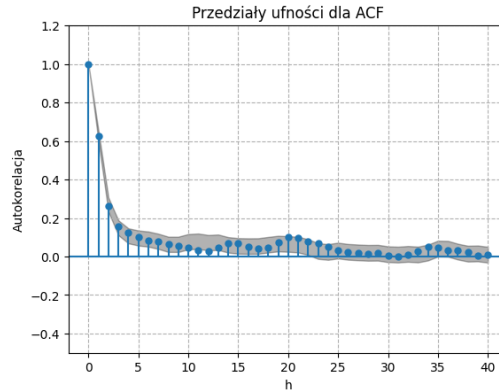
$$X_t - 1,253302X_{t-1} + 0,287910X_{t-2} = Z_t - 0,465374Z_{t-1} - 0,414144Z_{t-2}$$

4. Ocena dopasowania modelu

W następnym kroku należy się zastanowić, czy model faktycznie został dobrze dopasowany. W tym celu sprawdzimy, czy otrzymane dla naszych danych wartości funkcji autokorelacji oraz częściowej autokorelacji znajdują się w odpowiednim przedziale ufności dla szeregu ARMA(2,2) oraz porównamy wybrane linie kwantylowe szeregu ARMA(2,2) z trajektorią naszych danych.

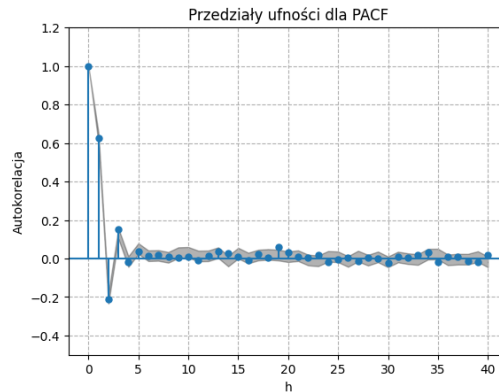
4.1. Przedziały ufności dla ACF/PACF

Na pierwszym wykresie (*Rysunek 7.*) przedstawiono funkcję autokorelacji szeregu Y_t wraz z 95% przedziałami ufności dla modelu ARMA(2,2). Oznacza to, że wartości empirycznej autokorelacji powinny z prawdopodobieństwem 95% mieścić się w wyznaczonym przedziale. Wyniki te są zgodne z oczekiwaniami – zdecydowana większość wartości funkcji autokorelacji szeregu Y_t znajduje się wewnątrz przedziału ufności. Nieliczne punkty wychodzące poza ten obszar pozostają jednak blisko jego granic, co pozwala uznać szereg Y_t za dobrze opisany modelem ARMA(2,2).



Rysunek 7: Wykres ACF wraz z przedziałami ufności

Kolejny wykres (*Rysunek 8.*) przedstawia funkcję częściowej autokorelacji szeregu Y_t wraz z przedziałami ufności 95% dla szeregu ARMA(2,2). Również w tym przypadku znaczna większość wartości funkcji częściowej autokorelacji znajduje się na obszarze wyznaczonym przez przedział ufności 95%. Nieliczne wartości znajdują się poza przedziałem ufności 95%. Analiza wykresu funkcji częściowej autokorelacji szeregu Y_t pozwala na wyciągnięcie wniosku, że szereg ten jest szeregiem ARMA(2,2), zgodnie z obliczeniami i założeniami z poprzedniego podpunktu.

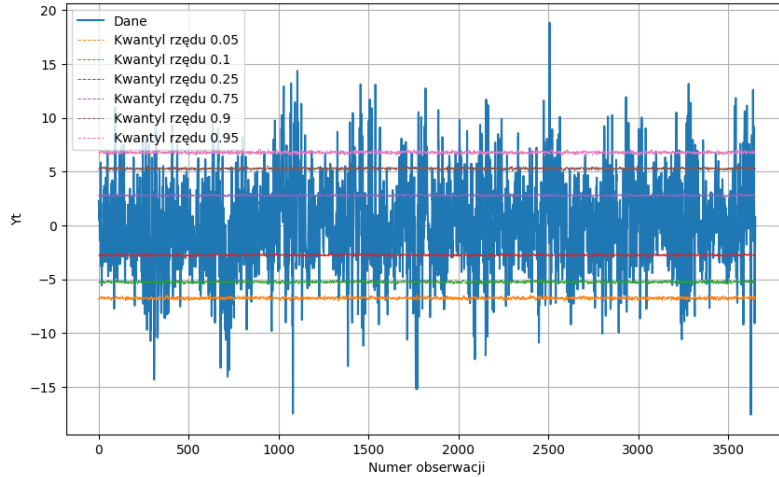


Rysunek 8: Wykres PACF wraz z przedziałami ufności

4.2. Porównanie trajektorii z liniami kwantylowymi

Aby dodatkowo zweryfikować, czy szereg Y_t jest szeregiem ARMA(2,2), można wykorzystać linie kwantylowe. Są to linie, które wyznaczają wartości kwantyli danego rzędu. Rząd kwantyli wyznacza wartość, poniżej której znajduje się dany procent obserwacji (np. linia kwantylowa rzędu 0,05 wyznacza wartość, od której mniejsze jest 5% obserwacji z szeregu). Na poniższym rysunku (*Rysunek 9.*) zaznaczono linie kwantylowe dla rzędu 0,05; 0,1; 0,25; 0,75; 0,9 oraz 0,95. Analiza wykresu trajektorii szeregu Y_t z naniesionymi liniami kwantylowymi dla szeregu ARMA(2,2) potwierdza, że otrzymany szereg jest szeregiem ARMA(2,2): powyżej linii kwantylowej wyznaczającej kwantyl

rzędu 0,95 znajduje się bardzo niewiele obserwacji w porównaniu z rozmiarem próbki, co może wskazywać na to, że faktycznie jedynie 5% obserwacji znajduje się powyżej wyznaczonej wartości. Również linie kwantylowe dla pozostałych rzędów kwantyli wydają się potwierdzać te wnioski, zatem ostatecznie przyjmujemy, że model ARMA(2,2) został dobrze dopasowany do danych.

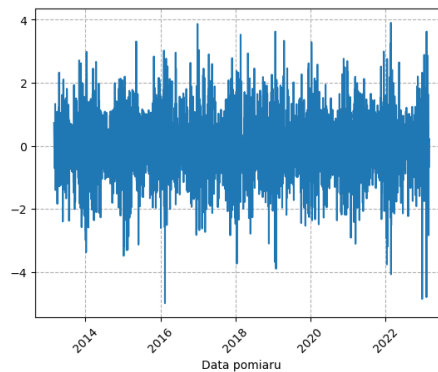


Rysunek 9: Trajektoria szeregu Y_t wraz z liniami kwantylowymi dla szeregu ARMA(2,2)

5. Weryfikacja założeń dotyczących szumu

Analiza szumu w kontekście szeregów czasowych polega na badaniu reszt modelu, czyli różnicy między rzeczywistymi obserwacjami a wartościami przewidywanymi przez model. Celem tej analizy jest sprawdzenie, czy reszty (szum) są losowe i nie zawierają ukrytych wzorców, które mogłyby sugerować niedopasowanie modelu lub obecność dodatkowych zależności w danych.

Zgodnie z wynikami przedstawionymi na poniższym wykresie, nie występują żadne wyraźne wzorce.



Rysunek 10: Wartości resztowe

Aby jednak zyskać pewność przeprowadzimy dodatkową analizę. Szum szeregu czasowego powinien spełniać założenia przedstawione w sekcjach 5.1 - 5.3. Dodatkowo, istnieje możliwość weryfikacji dodatkowego założenia dotyczącego pochodzenia szumu z rozkładu normalnego $N(0, \sigma^2)$, co omawiane jest w sekcji 5.4.

5.1. Średnia szumu $\mu = 0$

Reszty modelu powinny oscylować wokół zera, co wskazuje na poprawne usunięcie wszelkich systematycznych wzorców w danych, takich jak trend czy sezonowość. Zweryfikujmy zatem powyższe założenie.

5.1.1. Test t-Studenta (dla jednej próby)

Test t-Studenta dla jednej próby sprawdza, czy średnia z próby różni się istotnie od założonej wartości teoretycznej populacji. Test ten zakłada, że dane pochodzą z rozkładu normalnego.

Test polega na weryfikacji dwóch hipotez:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

gdzie: H_0 - hipoteza zerowa, H_1 - hipoteza alternatywna

Statystyka testowa t obliczana jest na podstawie wzoru:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

gdzie:

t - statystyka testu, $t \sim t_{n-1}$

\bar{x} - średnia z próby

μ_0 - teoretyczna (oczekiwana) średnia populacji

S - odchylenie standardowe z próby

n - długość próby

- Dla analizowanych danych przyjęto hipotezę zerową: $H_0 : \mu = 0$, którą testowano w odniesieniu do hipotezy alternatywnej: $H_1 : \mu \neq 0$. Poziom istotności testu ustalono na $\alpha = 0,05$.
Otrzymano następujące wyniki:

Statystyka	p-wartość
-0,012	0,991

Tabela 3: Wyniki t-testu

P-wartość wynosząca 0,991 ($> \alpha = 0,05$) wskazuje, że nie mamy podstaw do odrzucenia hipotezy zerowej. W związku z tym, możemy stwierdzić, że założenie dotyczące średniej jest prawdziwe (istotnie $\mu = 0$).

5.2. Stała i skończona wariancja

W tej sekcji przeprowadzimy testy statystyczne w celu weryfikacji założeń dotyczących jednorodności wariancji. Założenie to dotyczy sprawdzenia, czy zmienność (wariancja) jest jednorodna w różnych grupach danych. Będziemy weryfikować następujące hipotezy:

H_0 : Wszystkie grupy mają tę samą wariancję

H_1 : Wariancje są heterogeniczne (niejednorodne)

5.2.1. Test Levene'a

Test Levene'a jest jednym z najczęściej stosowanych testów do oceny jednorodności wariancji w grupach. Jest szczególnie przydatny w sytuacjach, gdzie rozkład danych może odbiegać od normalnego, ponieważ jest mniej wrażliwy na odchylenia od normalności niż na przykład test Bartlett'a. Test Levene'a jest oparty na porównaniu średnich różnic między obserwacjami a średnią grupy w ramach różnych grup.

Statystyka testowa W obliczana jest na podstawie wzoru:

$$W = \frac{(N - k)}{(k - 1)} \cdot \frac{\sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2}$$

gdzie:

W - statystyka testu,

N - liczba obserwacji

k - liczba grup

N_i - liczba obserwacji w i -tej grupie

\bar{Z}_i - średnia wartość Z_{ij} w i -tej grupie

\bar{Z} - średnia wartość Z_{ij}

5.2.2. Test Arch

Test ARCH polega na regresji kwadratów reszt na ich opóźnione wartości.

Statystyka testowa jest opisana wzorem:

$$Q = nR^2$$

gdzie:

R^2 - to współczynnik determinacji (R-squared) z regresji kwadratów reszt

n - liczba obserwacji

- Dla analizowanych danych przyjęto hipotezę zerową H_0 , którą testowano w odniesieniu do hipotezy alternatywnej H_1 . Poziom istotności testu ustalono na $\alpha = 0,05$.
Otrzymano następujące wyniki:

	Statystyka	p-wartość
Levene	1,026	0,311
Arch	149,74	0

Tabela 4: Wyniku testu Levene'a i Arch

Otrzymane wyniki testów są niejednoznaczne – test Levene'a wskazuje na jednorodność wariancji, podczas gdy Arch sugeruje jej brak. Aby uzyskać bardziej rzetelną ocenę, przeprowadzono dodatkowy test statystyczny.

5.2.3. Test Browna-Forsythe'a

Test Browna-Forsythe'a stanowi modyfikację testu Levene'a i charakteryzuje się większą odpornością na odchylenia od normalności danych. Zamiast średniej wykorzystuje medianę jako miarę centralną, co czyni go mniej wrażliwym na obecność wartości odstających.

Statystyka testowa F obliczana jest na podstawie wzoru:

$$F = \frac{(N - p)}{(p - 1)} \cdot \frac{\sum_{j=1}^p n_j (\tilde{Z}_j - \bar{\tilde{Z}})^2}{\sum_{j=1}^p \sum_{i=1}^{n_j} (Z_{ij} - \tilde{Z}_j)^2}$$

gdzie:

W - statystyka testu,

N - liczba obserwacji

p - liczba grup

n_j - liczba obserwacji w j -tej grupie

\tilde{Z}_j - mediana wartości Z_{ij} w j -tej grupie

$\bar{\tilde{Z}}$ - mediana Z_{ij}

- Dla analizowanych danych przyjęto hipotezę zerową H_0 , którą testowano w odniesieniu do hipotezy alternatywnej H_1 . Poziom istotności testu ustalono na $\alpha = 0,05$.

Otrzymano następujące wyniki:

P-wartość wynosząca 0,16 ($> \alpha = 0,05$) wskazuje, że nie mamy podstaw do odrzucenia hipotezy zerowej.

Ponieważ dwa z trzech testów wskazują na homogeniczność wariancji szumu, możemy przyjąć, że wariancja jest stała w analizowanym szeregu czasowym.

Statystyka	p-wartość
1,85	0,16

Tabela 5: Wyniki testu Browna-Forsythe'a

5.3. Niezależność reszt

W tej części przetestujemy niezależność wartości resztowych.

5.3.1. Test Ljunga-Boxa

Test Ljung-Boxa to test statystyczny, który służy do sprawdzenia, czy w danych szeregach czasowych występuje autokorelacja. Test polega na weryfikacji dwóch hipotez:

H_0 : Dane są nieskorelowane

H_1 : Dane są skorelowane

Statystyka testowa Q obliczana jest na podstawie wzoru:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{p}_k^2}{n-k}$$

gdzie:

Q - statystyka testu

n - wielkość próbki

\hat{p}_k - autokorelacja k -tego rzędu w próbce

h - liczba testowanych autokorelacji

→ Dla analizowanych danych przyjęto hipotezę zerową H_0 , którą testowano w odniesieniu do hipotezy alternatywnej H_1 . Poziom istotności testu ustalono na $\alpha = 0,05$.

Otrzymano następujące wyniki:

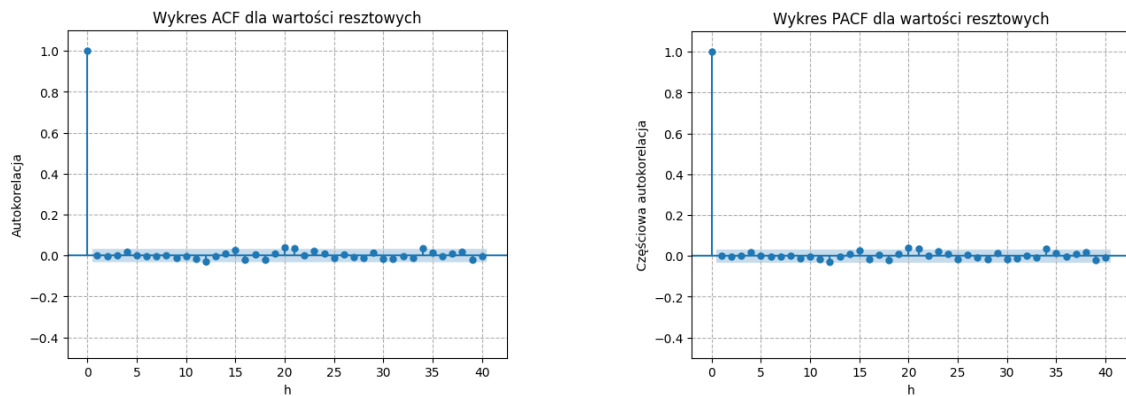
Statystyka	p-wartość
1.792	0.998

Tabela 6: Wyniki testu Ljunga-Boxa

P-wartość wynosząca 0,998 ($> \alpha = 0,05$) wskazuje, że nie ma podstaw do odrzucenia hipotezy zerowej. Oznacza to, że zgodnie z założeniem dane są nieskorelowane.

5.3.2. Wykres ACF/PACF dla wartości resztowych

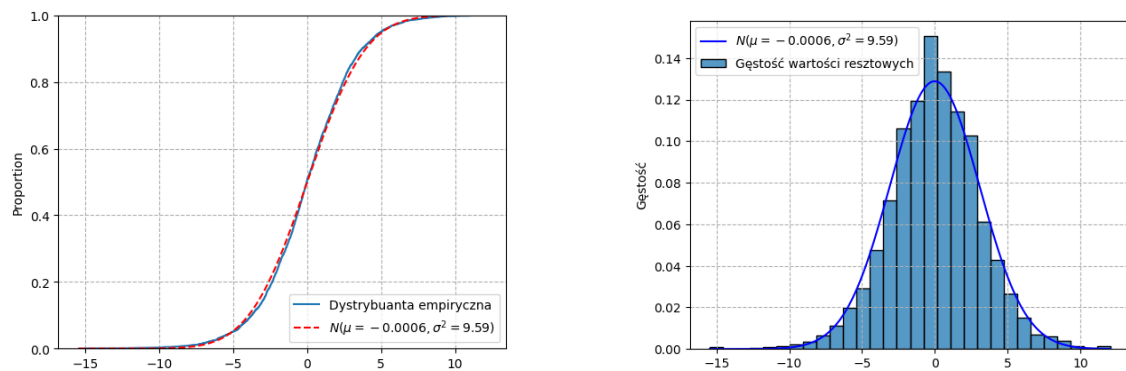
Na wykresie ACF i PACF wartości resztowych widoczna jest większa wartość autokorelacji tylko dla pierwszego opóźnienia ($\text{lag} = 1$), po czym wartości autokorelacji dla kolejnych opóźnień pozostają bliskie zeru i mieszczą się w granicach przedziału ufności. Taki przebieg wykresu sugeruje, że model dobrze wychwytuje zależności w danych, a reszty są losowe i niezależne po uwzględnieniu pierwszego opóźnienia.



Rysunek 11: Funkcje ACF i PACF dla wartości resztowych

5.4. Normalność rozkładu szumu

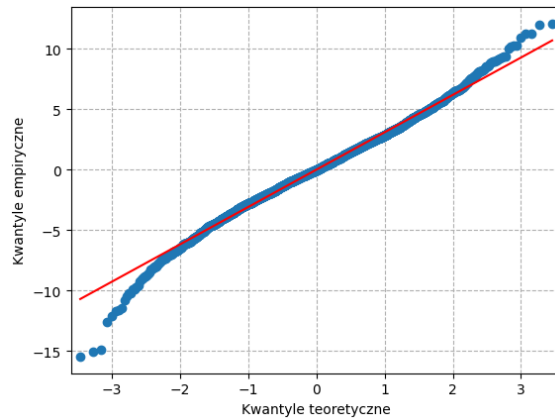
5.4.1. Dystrybuanta i gęstość



Rysunek 12: Funkcja dystrybuanty i histogram gęstości wartości resztowych

Na podstawie przedstawionych wykresów można stwierdzić, że w pierwszym etapie analizy, zarówno teoretyczne wartości dystrybuanty, jak i gęstości dość dobrze dopasowują się do wartości empirycznych obliczonych na podstawie reszt modelu ARMA(2,2). Użyto skrótu $N(\mu, \sigma^2)$, który oznacza teoretyczną dystrybuantę w przypadku lewego wykresu, a w drugim funkcję gęstości rozkładu normalnego o zadanych parametrach μ i σ^2 .

5.4.2. Wykres kwantylowy



Rysunek 13: Wykres kwantylowy wartości resztowych

Powyższy wykres porównuje kwantyle teoretyczne rozkładu normalnego z kwantylami empirycznymi reszt modelu. Linia czerwona wskazuje idealne dopasowanie – punkty blisko niej sugerują zgodność z rozkładem normalnym.

Większość punktów układa się wzdłuż linii, jednak odchylenia w obszarach krańcowych (ogony) mogą wskazywać, że reszty nie są idealnie normalne, a jedynie zbliżone do tego rozkładu.

W celu dokładniejszej oceny zostaną przeprowadzone testy statystyczne weryfikujące normalność.

5.4.3. Testy na normalność

Aby zbadać normalność rozkładu posłużymy się 2 testami statystycznymi, gdzie oba polegają na weryfikacji następujących hipotez:

H_0 : Próba pochodzi z populacji o rozkładzie normalnym

H_1 : Próba nie pochodzi z populacji o rozkładzie normalnym

a) Test Shapiro-Wilka

Test Shapiro-Wilka opiera się na analizie odchylen wartości obserwowanych od wartości oczekiwanych przy założeniu normalności. Jest szczególnie zalecany do małych i średnich próbek (poniżej 2000 obserwacji).

Statystyka testowa W obliczana jest na podstawie wzoru:

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

gdzie:

W - statystyka testu,

a_i - stałe wyliczane na podstawie uporządkowanych wartości próby x_i (tzw. współczynniki Shapiro-Wilka),

x_i - uporządkowane dane próbki (od najmniejszej do największej),

\bar{x} - średnia próbki,

n - liczba obserwacji w próbie.

b) Test Jarque-Bera

Test Jarque-Bera opiera się na dwóch miarach: skośności i kurtozie danych. Test sprawdza, czy wspomniane miary są zgodne z wartościami oczekiwanymi dla rozkładu normalnego (skośność = 0, kurtaza = 3).

Statystyka testowa JB obliczana jest na podstawie wzoru:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right)$$

gdzie:

JB - statystyka testu,

$$S - \text{skośność}, S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

$$K - \text{kurtoza}, K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

→ Dla analizowanych danych przyjęto hipotezę zerową H_0 , którą testowano w odniesieniu do hipotezy alternatywnej H_1 . Poziom istotności testu ustalono na $\alpha = 0,05$.

Otrzymano następujące wyniki:

	Statystyka	p-wartość
Shapiro-Wilk	0,99	0
Jarque-Bera	191,64	0

Tabela 7: Wyniki testów statystycznych na normalność rozkładu

Otrzymane wyniki testów na normalność rozkładu wskazują, że zarówno test Shapiro-Wilka, jak i test Jarque-Bera dają p-wartość równą 0. Oznacza to, że istnieją silne podstawy do odrzucenia hipotezy zerowej. W konsekwencji możemy stwierdzić, że wartości resztowe nie pochodzą z populacji o rozkładzie normalnym.

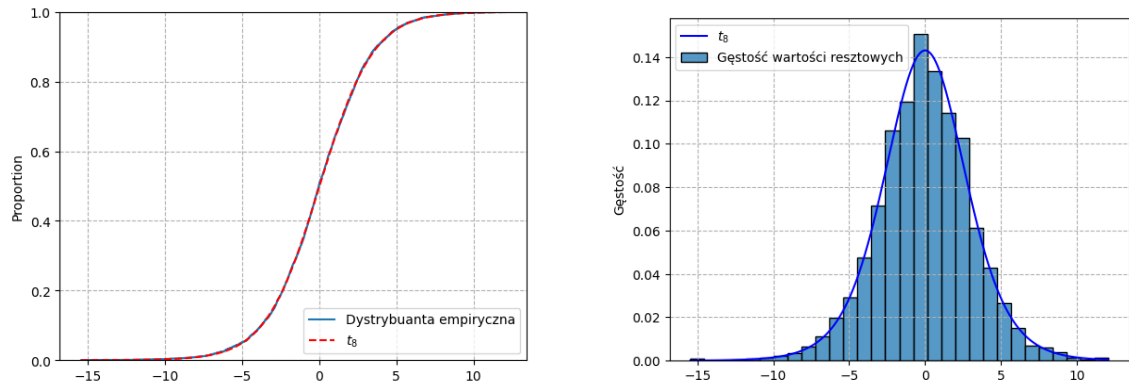
6. Weryfikacja dopasowania wartości resztowych do rozkładu t-Studenta

Ze względu na odrzucenie hipotezy o normalności reszt przeprowadzono dalszą analizę w celu znalezienia bardziej odpowiedniego modelu rozkładu. Postanowiono sprawdzić rozkład t-Studenta, który lepiej uwzględni potencjalne wartości odstające i charakteryzuje się grubszymi ogonami niż rozkład normalny.

Dopasowując rozkład t-Studenta do reszt, przyjęto model z 8 stopniami swobody.

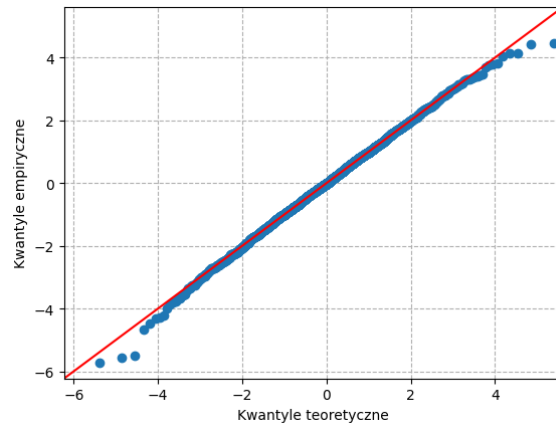
6.1. Dystrybuanta i gęstość

Na podstawie przedstawionych wykresów można zauważyć, że zarówno funkcja dystrybuanty, jak i funkcja gęstości dla rozkładu t-Studenta wykazują lepsze dopasowanie do wartości empirycznych niż w sekcji 5.4.1, gdzie badano rozkład normalny. W celu zachowania czytelności wykresów, skrótowno przyjęto oznaczenie t_8 , które na lewym wykresie reprezentuje teoretyczną funkcję dystrybuanty rozkładu t-Studenta z 8 stopniami swobody, natomiast na prawym wykresie oznacza funkcję gęstości tego samego rozkładu.



Rysunek 14: Funkcja dystrybuanty i histogram gęstości wartości resztowych

6.2. Wykres kwantylowy



Rysunek 15: Wykres kwantylowy wartości resztowych

Wykres kwantylowy dla rozkładu t-Studenta pokazuje znacznie lepsze dopasowanie kwantyli empirycznych reszt modelu do kwantyli teoretycznych w porównaniu z rozkładem normalnym. Chociaż w obszarach krańcowych nadal widoczne są niewielkie odstępstwa, ich zakres jest zdecydowanie mniejszy, co sugeruje, że wartości resztowe mogą być zgodne z rozkładem t-Studenta.

Aby potwierdzić te obserwacje, zostaną przeprowadzone testy statystyczne, które zweryfikują, czy próba pochodzi z populacji o rozkładzie t-Studenta.

6.3. Testy statystyczne

Aby zbadać czy próba pochodzi z rozkładu t-Studenta posłużymy się dwoma testami statystycznymi, gdzie oba polegają na weryfikacji następujących hipotez:

H_0 : Próba pochodzi z populacji o zadanym rozkładzie

H_1 : Próba nie pochodzi z populacji o zadanym rozkładzie

6.3.1. Test Kołmogorowa - Smirnova

Test Kołmogorowa - Smirnova to test nieparametryczny, który sprawdza, czy dane pochodzą z zadanego rozkładu. Test mierzy największą odległość pionową między empiryczną funkcją rozkładu próbki a funkcją rozkładu odniesienia.

Statystyka testowa D obliczana jest na podstawie wzoru:

$$D = \max_x |F_n(x) - F(x)|$$

gdzie:

D - statystyka testu,

$F_n(x)$ - empiryczna funkcja dystrybuanty,

$F(x)$ - teoretyczna funkcja dystrybuanty,

6.3.2. Test Craméra-von Misesa

Podobnie jak test Kołmogorowa-Smirnowa, porównuje on funkcję dystrybuanty empirycznej z funkcją dystrybuanty teoretycznej. Jednak w odróżnieniu od testu KS, test Craméra-von Misesa bierze pod uwagę całą funkcję dystrybuanty i mierzy całkowitą różnicę pomiędzy nimi.

Statystyka testowa W^2 obliczana jest na podstawie wzoru:

$$W^2 = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x)$$

gdzie:

W^2 - statystyka testu,

$F_n(x)$ - empiryczna funkcja dystrybuanty,

$F(x)$ - teoretyczna funkcja dystrybuanty

→ Dla analizowanych danych poziom istotności testu ustalono na $\alpha = 0,05$ oraz przyjęto hipotezy:

H_0 : Próba pochodzi z populacji o rozkładzie t-Studenta

H_1 : Próba nie pochodzi z populacji o rozkładzie t-Studenta.

Otrzymano następujące wyniki:

	Statystyka	p-wartość
Kołmogorow-Smirnov	0,0104	0,82
Cramér-von Mises	0,043	0,92

Tabela 8: Wyniki testów statystycznych na rozkład t-Studenta

Wyniki obu testów jednoznacznie wskazują, że analizowane przez nas wartości resztowe pochodzą z rozkładu t-Studenta.

7. Podsumowanie

Niniejszy raport dotyczył analizy danych rzeczywistych przy pomocy modelu ARMA. Dane opisujące średnią dzienną temperaturę powietrza w Bostonie zostały sprawdzone pod względem brakujących wartości, odczytano największą i najmniejszą wartość, a następnie przeprowadzono dekompozycję – usunięto z danych trend oraz sezonowość, a na końcu sprawdzono stacjonarność otrzymanego szeregu. W dalszej kolejności do szeregu dobrano rząd p i q z użyciem kryteriów informacyjnych i wyestymowano parametry modelu za pomocą 3 metod, porównano wyniki oraz oceniono dopasowanie modelu z użyciem przedziałów ufności oraz linii kwantylowych. Ocena wykazała, że szereg został poprawnie zakwalifikowany jako szereg ARMA(2,2). Na koniec zweryfikowano założenia dotyczące szumu. Wykazano, że średnia szumu jest równa 0, wariancja jest stała, a wartości resztowe są niezależne. Przeprowadzono również test Ljungu-Boxa dla autokorelacji, który wykazał, że reszty są losowe i niezależne po uwzględnieniu pierwszego opóźnienia. Testy na normalność rozkładu wykazały, że wartości resztowe nie pochodzą z populacji o rozkładzie normalnym. Po przeprowadzeniu dalszych testów, wywnioskowano, że wartości resztowe pochodzą z rozkładu t-Studenta.