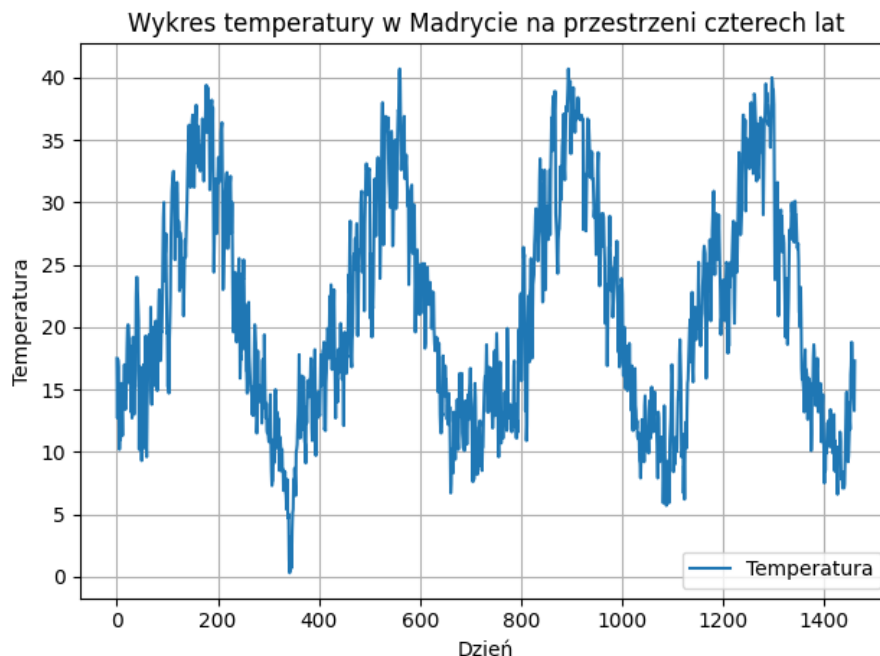


Aleksandra Górską, 268734,
Marcin Pałyga, 268768

Komputerowa Analiza Szeregów czasowych - Raport I

1. Wstęp i opis danych

Tematem naszego raportu będzie analiza danych pogodowych przy pomocy modelu ARMA. Wybrane dane dotyczą wysokości temperatury w Madrycie na przestrzeni czterech lat, a dokładnie od 01.02.2020 r. do 01.02.2024 r. Próbką jest długości 1462 i zawiera dane z każdego dnia naszego okresu. Dane pochodzą ze strony: ["https://meteostat.net/en/place/es/madrid?s=08222t=2020-02-01/2024-02-01"](https://meteostat.net/en/place/es/madrid?s=08222t=2020-02-01/2024-02-01). Na rysunku 1. możemy zobaczyć wykres temperatury na przestrzeni czterech lat. Analiza danych została wykonana za pomocą języka Python i następujących bibliotek: numpy, scipy, matplotlib, pandas, statsmodels oraz sklearn.



Rysunek 1: Wykres temperatury

2. Dekompozycja szeregu czasowego

Przed wykonaniem dekompozycji Walda sprawdziliśmy jak wyglądają wykresy ACF (funkcja autokorelacji) oraz PACF (funkcja częściowej autokorelacji). Empiryczne wartości ACF i PACF możemy wyznaczyć za pomocą następujących wzorów:

— Próbkowa funkcja autokowariancji (ACVF):

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{x})(X_t - \bar{x})$$

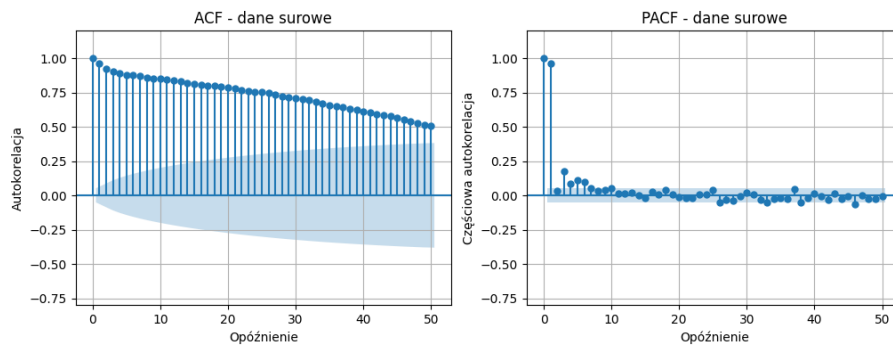
— Próbkowa funkcja autokorelacji (ACF):

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad \hat{\rho}(0) = 1$$

gdzie:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad -n < h < n$$

Na rysunku 2. przedstawiliśmy jak zachowują się te funkcje dla opóźnień w zakresie od 0 do 50. Po wykresie ACF możemy stwierdzić niestacjonarność naszego szeregu. Chcąc to potwierdzić wykorzystaliśmy test ADF (Augmented Dickey-Fuller Test), którego wynik p-value na wysokości 0.096 potwierdził nasze stwierdzenie.



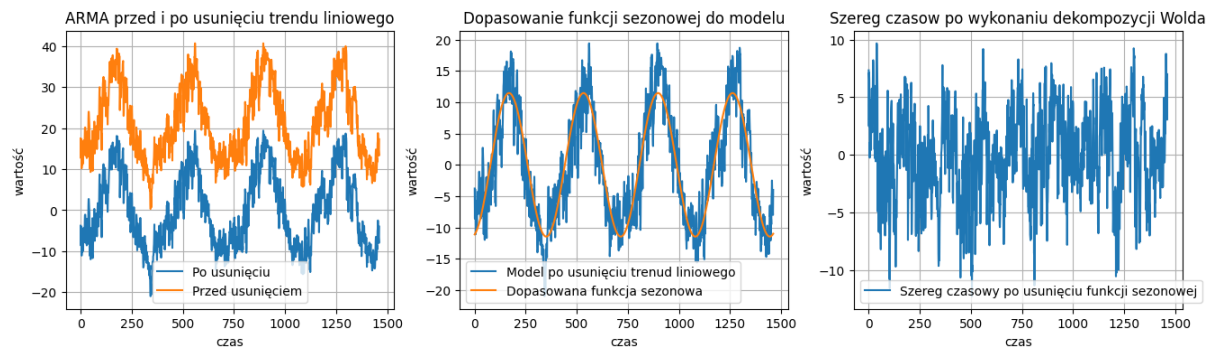
Rysunek 2: ACF oraz PACF dla danych surowych

Następnie przeszliśmy do wykonania dekompozycji Walda szeregu czasowego korzystając ze wzoru:

$$Y_t = s(t) + m(t) + X_t$$

gdzie: $m(t)$, $s(t)$, X_t - trend, funkcja okresowa, szereg stacjonarny w słabym sensie

Na rysunku 3. możemy zobaczyć trzy etapy dekompozycji Walda. Najpierw dopasowaliśmy i odjęliśmy od początkowego szeregu trend. Następnie dopasowaliśmy funkcję okresową odpowiadającą za sezonowość danych. Do danych dopasowaliśmy funkcję sinus z odpowiednimi parametrami amplitudy, częstotliwości i przesunięcia. Po odjęciu jej od wcześniej uzyskanego szeregu powinniśmy dostać szereg stacjonarny co możemy ponownie sprawdzić testem ADF. Test zwrócił p-value równe 0.0 z czego wnioskujemy, że szereg uzyskany z dekompozycji jest stacjonarny.



Rysunek 3: ACF oraz PACF dla danych surowych

3. Dobranie rzędu modelu ARMA

Aby dobrać rząd modelu ARMA skorzystaliśmy z następujących kryteriów informacyjnych: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) oraz Hannan–Quinn Information Criterion (HQIC). W każdym z nich wybieramy takie p oraz q dla modelu ARMA, aby wynik tych kryteriów był jak najmniejszy. Wyniki każdego z kryteriów jednoznacznie wskazują na dobranie p i q o wartości odpowiednio 2 i 2. Pokazują to rysunki 4. 5. oraz 6., które przedstawiają posortowane (od najmniejszego) wartości odpowiednich kryteriów.

	p	q	AIC	BIC	HQIC
8	2	2	6665.815249	6697.540613	6677.649276
5	1	2	6674.147562	6700.585365	6684.009251
7	2	1	6682.336903	6708.774706	6692.198592

Rysunek 4: Akaike Information Criterion

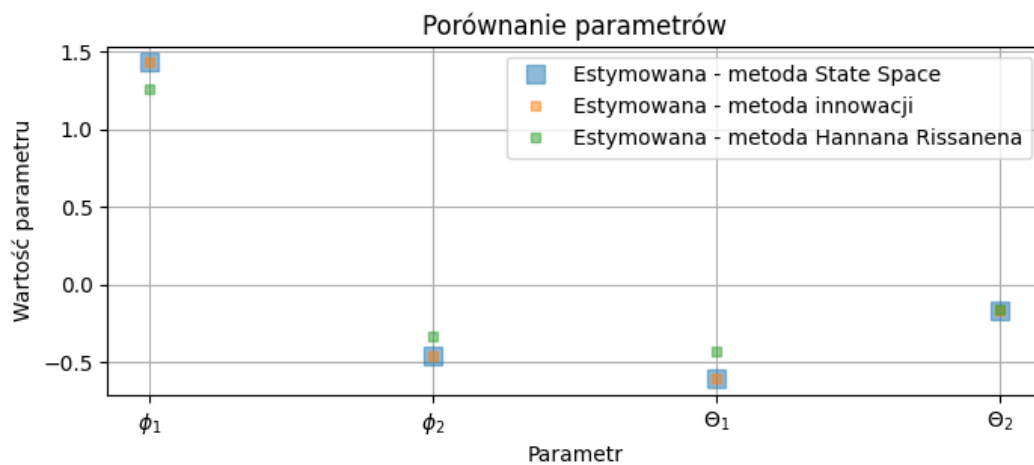
	p	q	AIC	BIC	HQIC
8	2	2	6665.815249	6697.540613	6677.649276
5	1	2	6674.147562	6700.585365	6684.009251
3	1	0	6689.078522	6704.941204	6694.995536

Rysunek 5: Bayesian Information Criterion

p	q		AIC	BIC	HQIC
8	2	2	6665.815249	6697.540613	6677.649276
5	1	2	6674.147562	6700.585365	6684.009251
7	2	1	6682.336903	6708.774706	6692.198592

Rysunek 6: Hannan–Quinn Information Criterion

Następnie przeszliśmy do estymacji parametrów modelu. Tutaj także skorzystaliśmy, z trzech metod estymacji, mianowicie: SSM (State Space Model), metoda Hannana Rissanena oraz metoda innowacji. Metoda innowacji oraz SSM dały bardzo zbliżone wyniki dlatego zdecydowaliśmy się skorzystać z wyników uzyskanych przez te metody. Na rysunku 7. możemy zobaczyć porównanie wyników uzyskanych przez każdą z metod.

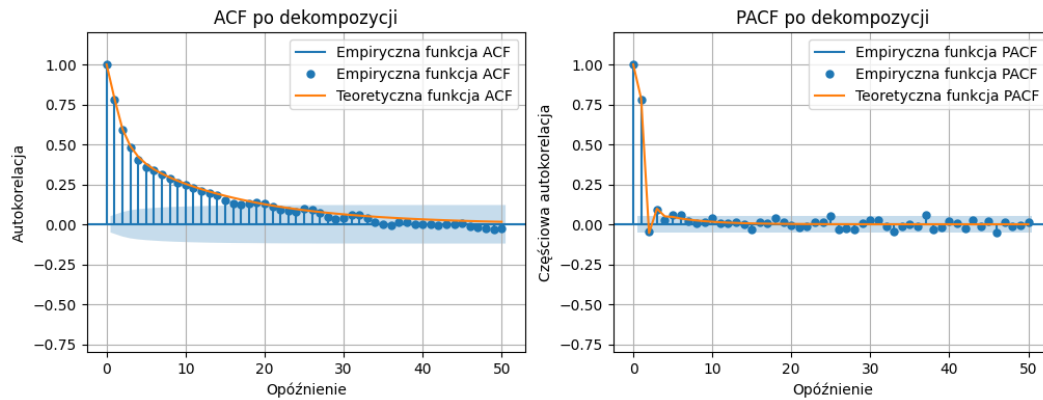


Rysunek 7: Estymacja parametrów modelu

Parametry uzyskane przez metodę SSM to 1.43, -0.46, -0.61, -0.17 odpowiednio dla $\Phi_1, \Phi_2, \Theta_1, \Theta_2$ i takich użyjemy do wyznaczenia teoretycznych wartości funkcji ACF oraz PACF.

4. Ocena dopasowania modelu

Aby ocenić dopasowanie modelu ARMA do naszych danych porównamy teoretyczne oraz empiryczne wartości funkcji ACF oraz PACF. Na rysunku 8. widzimy, że wartości empiryczne mają bardzo zbliżone wartości do funkcji teoretycznej stąd zakładamy, że rząd modelu jest dobrze dopasowany.



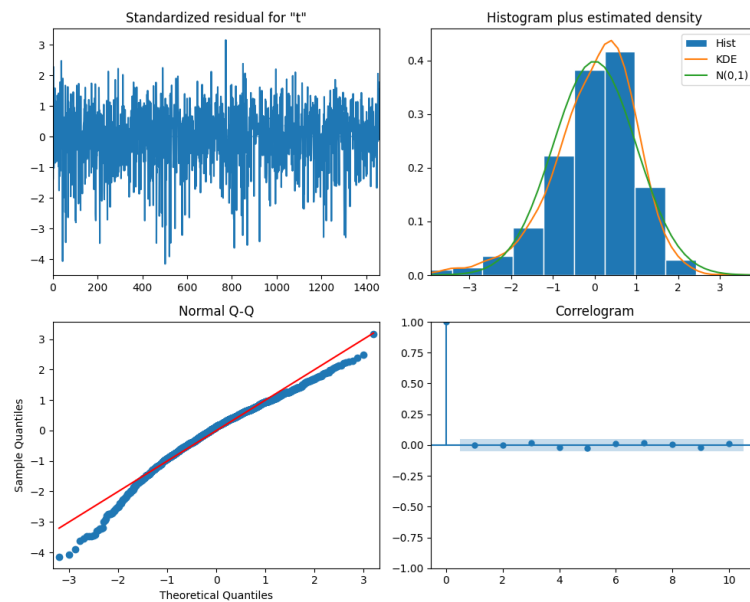
Rysunek 8: Teoretyczne i empiryczne wartości ACF oraz PACF

5. Weryfikacja założeń dotyczących szumu

Założenia dotyczące szumu są następujące:

- Średnia równa 0
- Stała wariancja
- Niezależność danych
- Normalność rozkładu

Na rysunku 9. widzimy wykresy weryfikujące założenia dotyczące średniej, wariancji oraz normalności rozkładu. Widzimy że rozkład przypomina normalny jednak testy Shapiro-Wilka oraz Jarque-Bera odrzucają taką hipotezę. Średnia danych jest zbliżona do 0 i wynosi 0.02, jednak nie jest to wystarczająco bliska wartość, co potwierdza p-value t testu na poziomie 0.99. Do sprawdzenia założenia dotyczącego wariancji użyliśmy Arch testu (Autoregressive conditional heteroskedasticity test), który zwrócił wartość p-value na poziomie 0.0 co oznacza, że wariancja nie jest stała. W ostateczności, aby sprawdzić czy dane są niezależne używamy testu Ljunga-Boxa. Wartości równe 0.0, wskazują na zależność danych.



Rysunek 9: Wykresy weryfikujące założenia dotyczące szumu

6. Podsumowanie

Porównanie wartości teoretycznych i empirycznych ACF oraz PACF wskazują na odpowiednie dobranie parametrów do modelu ARMA jednak założenia dotyczące szumu nie zostały spełnione. Może to świadczyć o dalszej zależności danych po wykonaniu dekompozycji Walda. Możliwe jest, że wykonanie transformacji Boxa-Coxa czy skorzystanie z różnicowania dałoby lepsze wyniki.