

Projekt: Analiza spektralna

Magdalena Smalarz

1. Wstęp

Stopa bezrobocia to wielkość statystyczna opisująca nasilenie zjawiska bezrobocia w danej populacji. Najczęściej definiuje się ją jako stosunek liczby osób bezrobotnych do liczby ludności aktywnej zawodowo (zasobu siły roboczej danej populacji). Tak zdefiniowaną stopę bezrobocia oblicza się różnie, w zależności od przyjętej definicji osoby bezrobotnej.

Najczęściej praktykowane, ale nie do końca prawidłowe, jest podawanie stopy bezrobocia rejestrowanego, do wyliczenia której przyjmuje się, że bezrobotny, to osoba, która jest aktualnie zarejestrowana zgodnie z prawem danego kraju, jako poszukująca pracy. Natomiast Międzynarodowa Organizacja Pracy zaleca wyliczanie stopy bezrobocia zgodnie z następującą definicją: za bezrobotną uważa się osobę, która spełnia równocześnie wszystkie poniższe warunki:

- ukończyła 15 lat;
- aktualnie nie pracuje (nie świadczy pracy w rozumieniu definicji MOP), ani nie odbywa przyuczenia do zawodu z elementami nauki praktycznej;
- poszukiwała aktywnie pracy w tygodniu badania lub w ciągu 6 tygodni poprzedzających badanie;
- jest zdolna do podjęcia pracy w tygodniu badanym lub następnym i wyraża gotowość podjęcia takiej pracy.

W Polsce podaje się najczęściej stopę bezrobocia rejestrowanego, natomiast GUS bada także stopę bezrobocia według definicji MOP (w ramach BAEL – Badania Aktywności Ekonomicznej Ludności, co kwartał na próbie rotacyjnej 20 000 gospodarstw domowych). Stopa bezrobocia rejestrowanego wypada przeciętnie ok. 1,5% poniżej stopy bezrobocia ustalonej w badaniu GUS.

W projekcie postaram się dowiedzieć czy biorąc pod uwagę wielkość stopy bezrobocia rejestrowanego można mówić o dominujących cyklach mających wpływ na jej zmienność. Przyjęty poziom istotności w projekcie: $\alpha = 0,05$.

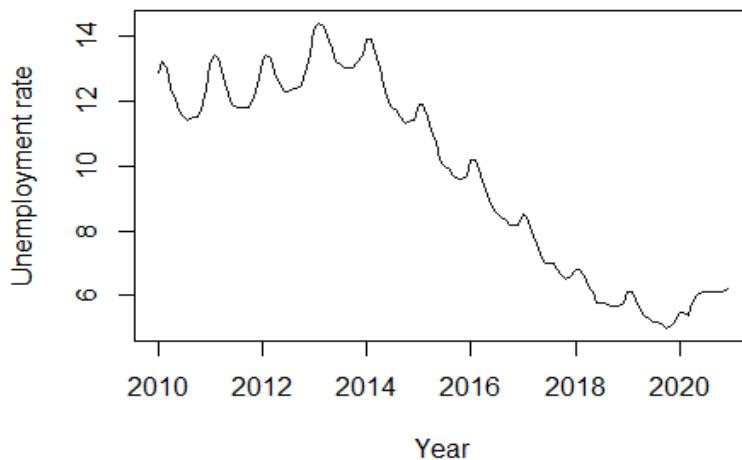
2. Dane

Dane pochodzą z Banku Danych Makroekonomicznych i dotyczą stopy bezrobocia rejestrowanego w Polsce (stan w końcu okresu) wyrażonego w procentach [%] w ujęciu miesięcznym. W projekcie uwzględniono dane z okresu: styczeń 2010 - grudzień 2020. Stopę bezrobocia obliczono przy wykorzystaniu danych m.in. dot. pracujących w rolnictwie indywidualnym (będących składową cywilnej ludności aktywnej zawodowo) wyszacowanych na podstawie wyników Spisów:

- do 1995 r. - Narodowego Spisu Powszechnego 1988,
- od 1996 r. do 2002 r. (w wariantie "a") - Powszechnego Spisu Rolnego 1996,
- od 2002 r. (w wariantie "b") - Narodowego Spisu Powszechnego Ludności i Mieszkań oraz Powszechnego Spisu Rolnego 2002,
- od grudnia 2010 r. - Powszechnego Spisu Rolnego 2010.

Poniżej przedstawiono kilka losowo выбрanych danych oraz wykres danych.

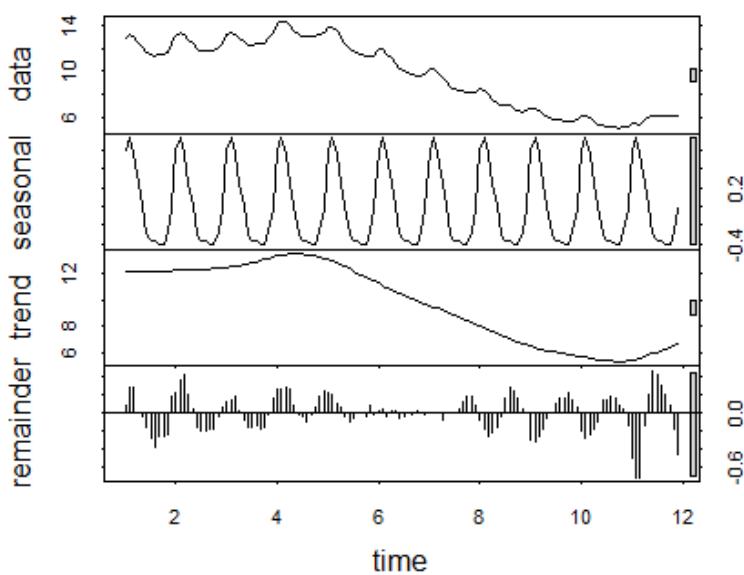
```
## [1] 11.5 13.2 5.7 5.5 11.8 6.8 9.9 6.8 9.7 5.8
```



3. Dekompozycja szeregu czasowego

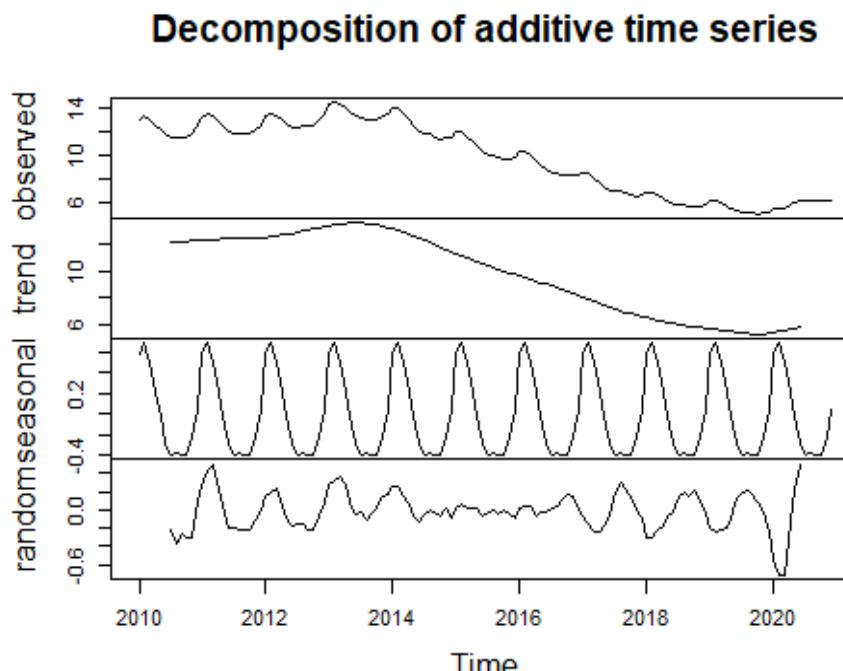
3.1. Metoda LOESS

Dekompozycja szeregu czasowego przy użyciu metody LOESS (STL) to algorytm, który został opracowany, w celu podzielenia szeregu czasowego na trzy komponenty: trend, sezonowość i resztę. STL jest dostępny w R za pośrednictwem funkcji `stl()`.



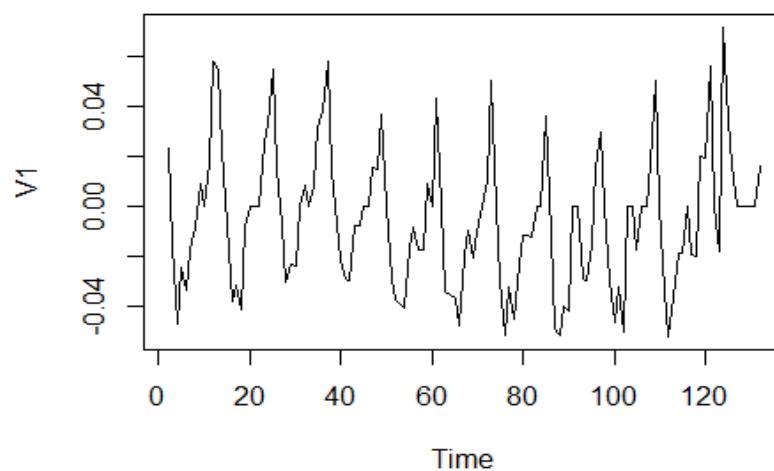
3.2. Średnie ruchome

Do wyznaczenia elementów, z jakich składa się badany szereg czasowy wykorzystano także średnie ruchome oraz dostępną w pakiecie RStudio funkcję `decompose()`.



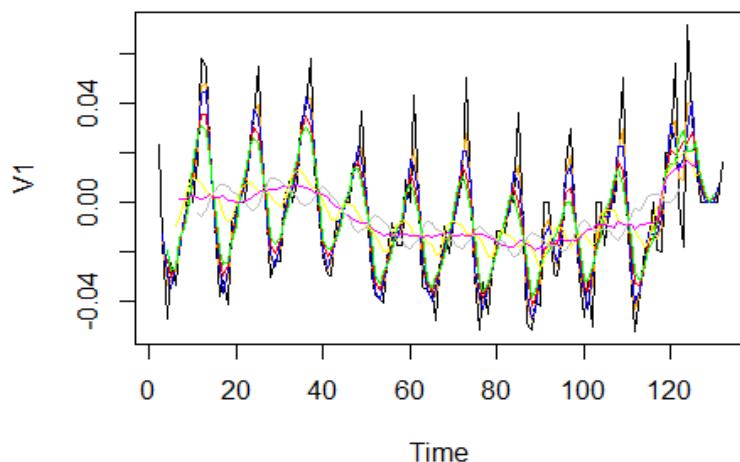
3.3. Usunięcie trendu z danych

Aby uniknąć występowania wolnych częstotliwości zaburzających interpretację otrzymanego wyniku analizy spektralnej, usunięto trend z danych. Wykorzystano funkcję `diff()` pakietu RStudio.



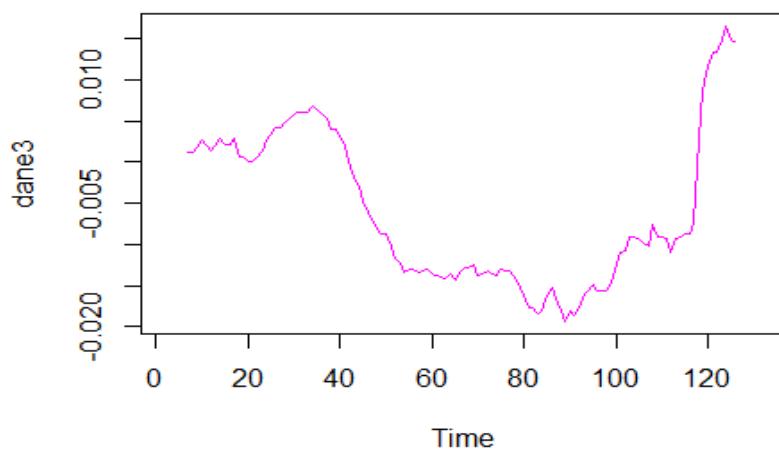
3.4. Określenie rzędu okresowości

Wykorzystując usunięcie trendu z danych i wcześniej wspomiane średnie ruchome, określono rzяд okresowości badanych danych. Poniżej zostały zestawione wykresy wygładzone średnimi ruchomymi rządu: 2,3,4,5,10,12 i 20.



Legenda: linia pomarańczowa - rząd 2; linia niebieska - rzząd 3; linia czerwona - rzząd 4; linia zielona - rzząd 5; linia żółta - rzząd 10; linia różowa - rzząd 12; linia szara - rzząd 20.

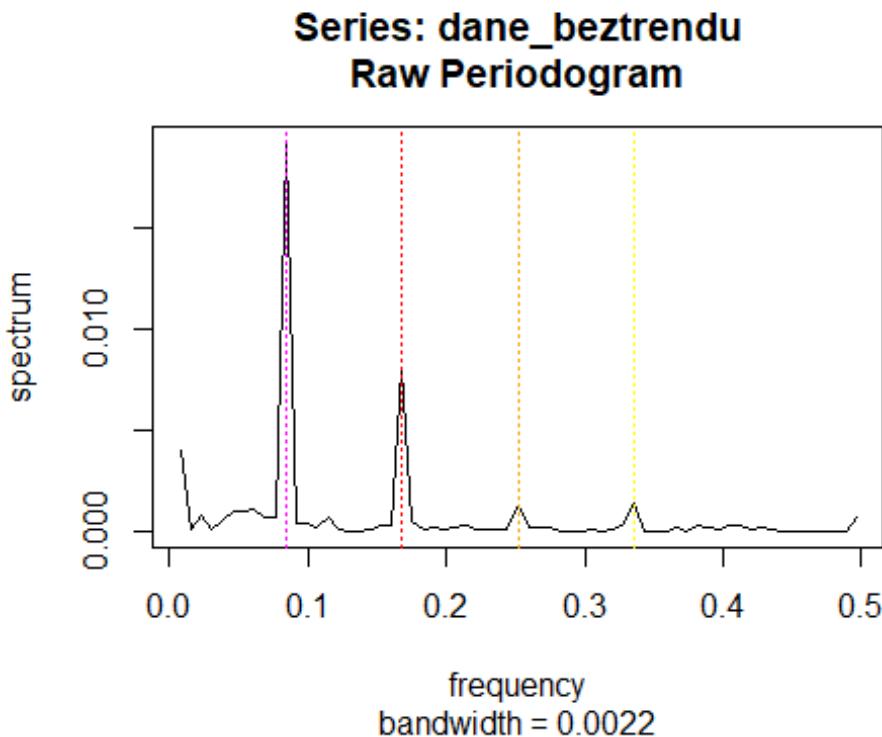
Długość okresu (rzяд okresowości) występujący w danych jest równy 12, ponieważ widać, że większy ani mniejszy rzząd średniej ruchomej nie poprawia wygładzenia wykresu, wręcz powoduje kolejne wahania. Poniżej został zaprezentowany wygładzony sygnał średnią ruchomą rzędu 12.



4. Periodogramy

W celu określenia długości cykli mających dominujący wpływ na zmienność stopy bezrobocia rejestrowanego w Polsce na przestrzeni lat 2010-2020 wyznaczono różne rodzaje periodogramów. Do ich wykreślenia użyto danych pozbawionych trendu.

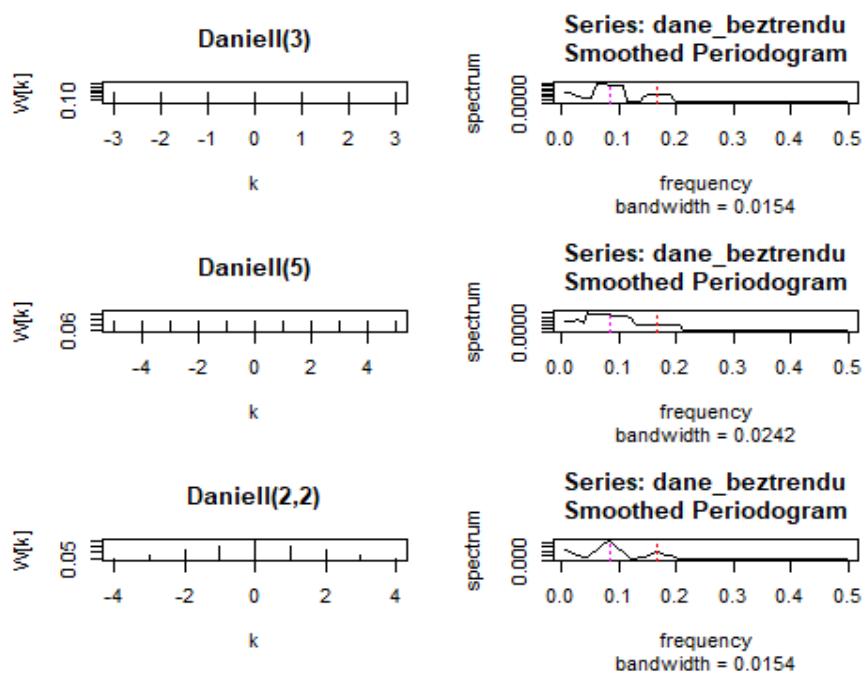
4.1. Periodogram naiwny



Legenda: linia różowa - częstotliwość $1/12$; linia czerwona - częstotliwość $1/6$; linia pomarańczonwa - częstotliwość $1/4$; linia żółta - częstotliwość $1/3$.

Na periodogramie widać jeden dominujący pik przy częstotliwości $1/12$. Analizowane dane są w ujęciu miesięcznym co oznacza, że w danych występuje cykl roczny. Oprócz tego, pojawiają się mniejsze piki przy częstotliwościach bliskich $1/6$, $1/4$ i $1/3$, co wskazuje na występowanie cykli półrocznych, czteromiesięcznych i kwartalnych.

4.2. Periodogram wygładzony oknem Daniella

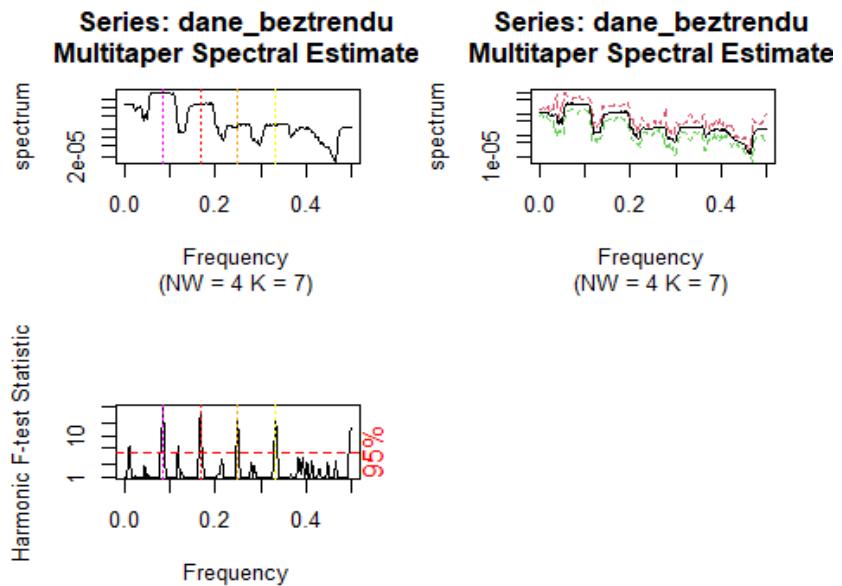


Legenda: linia różowa - częstotliwość $1/12$; linia czerwona - częstotliwość $1/6$.

Po wygładzeniu periodogramu oknem Daniella $m=c(2,2)$ można dostrzec dwa piki przy częstotliwościach $1/12$ i $1/6$. Przy oknach o $m=3$ i $m=5$ występuje sytuacja trudna do interpretacji, ponieważ periodogram stał się bardziej płaski i każda z częstotliwości w takim przypadku byłaby równie istotna.

4.3. Periodogram multitaper

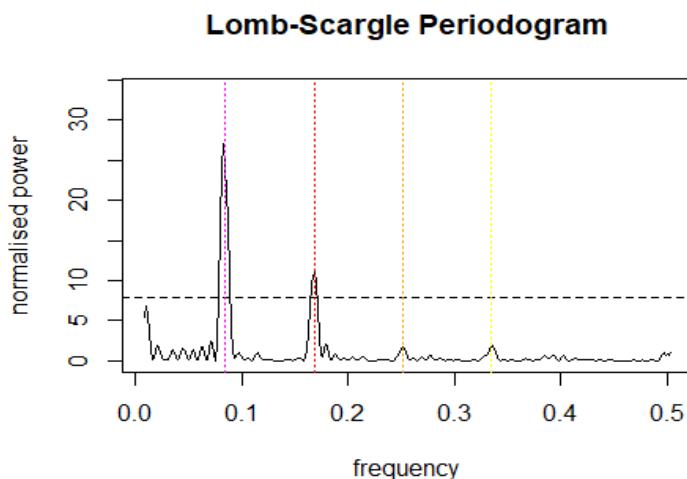
Wykresy przedstawiają kolejno: periodogram wykonany metodą multitaper, ten sam periodogram lecz z przedziałami ufności na przyjętym poziomie istotności oraz wykres statystki F.



Legenda: linia różowa - częstotliwość 1/12; linia czerwona - częstotliwość 1/6; linia pomarańczonwa - częstotliwość 1/4; linia żółta - częstotliwość 1/3.

Periodogram wykonany metodą multitaper pokazuje ogólne występowanie dominujących częstotliwości, które można odczytać z wykresu statystyki F. Ponownie zostały wskazane częstotliwości 1/12, 1/6, 1/4 i 1/3. Wszystkie z nich są istotne statystycznie na poziomie istotności $\alpha=0.05$. Pomiędzy nimi można zaobserwować również inne mniejsze piki, na granicy istotności zupełnie bądź nieistotne.

4.4. Periodogram Lomb-Scargle

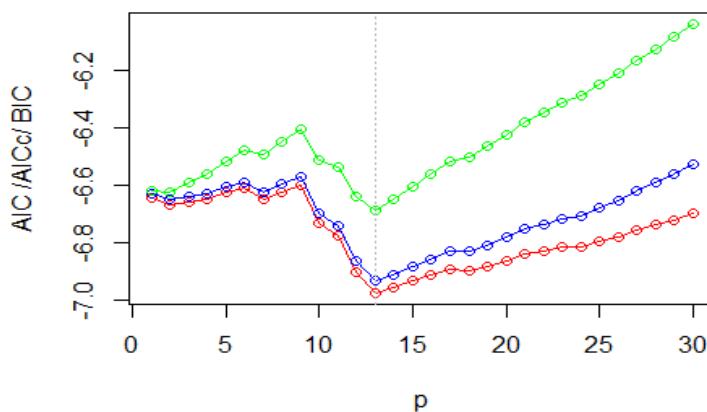


Legenda: linia różowa - częstotliwość 1/12; linia czerwona - częstotliwość 1/6; linia pomarańczonwa - częstotliwość 1/4; linia żółta - częstotliwość 1/3; przerywana czarna linia - poziom ufności.

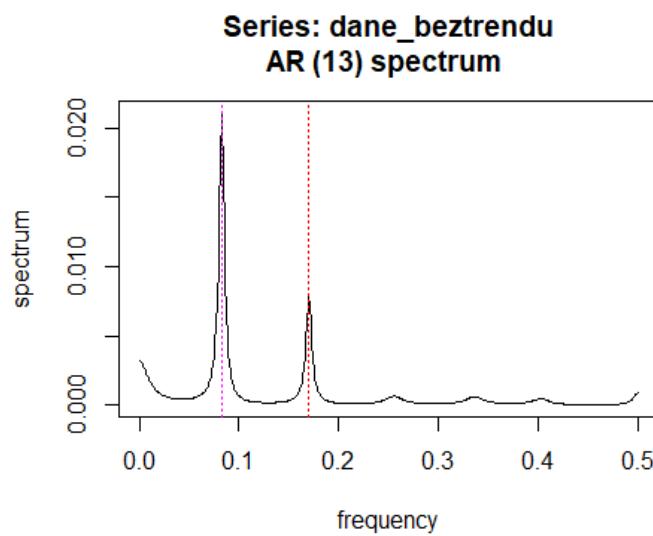
Periodogram LS wskazał dwa duże, istotne statystycznie piki przy częstotliwościach $1/12$ i $1/6$. Na periodogramie występują także mniejsze piki przy częstotliwościach $1/4$ i $1/3$, nieistotne statystycznie.

4.5. Periodogram autoregresyjny

Wykres przedstawia zmodyfikowane kryteria informacyjne: AIC (czerwony), AICc (niebieski) oraz BIC (zielony).



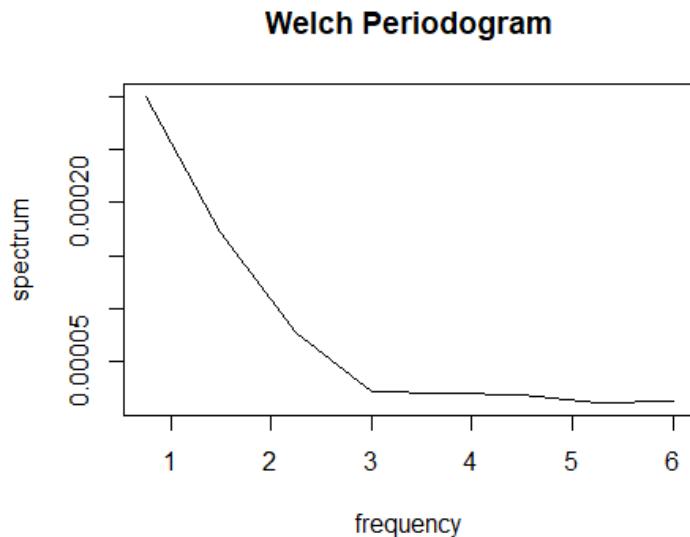
Na podstawie osiąganych przez wszystkie kryteria minimum można określić rzяд opóźnień równy 13 (szara linia). Znając tę wielkość wykreślono periodogram autoregresyjny.



Legenda: linia różowa - częstotliwość $1/12$; linia czerwona - częstotliwość $1/6$.

Periodogram wskazuje na obecność cykli rocznych i półrocznych.

4.6. Periodogram Welch'a



W przypadku analizowanych danych periodogram Welch'a wskazuje na występowanie wolnych częstotliwości, co sugeruje dalsze występowanie trendu w danych. Podobnie do pozostałych, wskazuje większy obszar częstotliwości w granicach niskich częstotliwości, gdzie pozostałe periodogramy posiadają piki. Nie wskazuje natomiast żadnych szczególnych częstotliwości.

5. Podsumowanie

Cykl/Periodogram	Naiwny	Okno Daniella	Multitaper	Lomb-Scargle	Autoregresyjny	Welcha
roczny	+	+	+	+	+	-
polroczny	+	+	+	+	+	-
czteromiesieczny	-	-	+	-	-	-
kwartalny	-	-	+	-	-	-

Nie wszystkie metody tworzenia periodogramów wskazały te same cykle. Każdy z pozostałych periodogramów wskazał na obecność cyklu rocznego i półrocznego w analizowanym zestawie danych. Jeden z nich (multitaper) podał także informacje o istotnym istnieniu cyklu kwartalnego i czteromiesięcznego. Periodogram Welch'a okazał się najgorszy ze wszystkich i nie poradził sobie z badanymi danymi.

Podsumowując, wśród analizowanych danych dotyczących stopy bezrobocia rejestrowanego w Polsce na przestrzeni ostatnich 10 lat można zauważać regularnie występujący cykl wpływający na zmienność badanej wielkości powtarzający się co rok oraz co pół roku.

Bibliografia

1. Dane ze strony: r-bloggers.com/2013/01/seasonal-trend-decomposition-in-r/ [Dostęp: 10.02.2021].