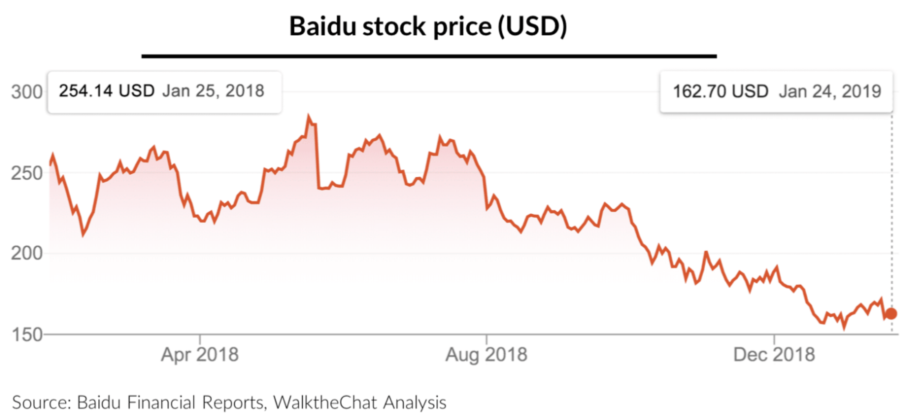
Metody statystyczne w analizie szeregów czasowych (time series)

Szymon Frelich, Bartosz Włodarski

1 Definicja szeregu czasowego

Szeregi czasowe (time series) to realizacja procesu stochastycznego, którego dziedziną jest czas. Inaczej mówiąc jest to ciąg informacji uporządkowany w czasie.



Rysunek 1: Przykład szeregu czasowego - indeks giełdowy

Ze względu na odstęp czasowymiędzy kolejnymi elementami szeregi dzielimy na:

- regularne - stałe odstępy między elementami
- rozmyte - nie regularne odstępy

1.1 Notacja

Do oznaczania ciągów czasowych stosowane są różne notacje. Często ciąg czasowy X ndeksowany liczbami naturalnymi zapisuje się jako:

$$X = X_1, X_2, X_3, ... \tag{1}$$

Można się również spotkać z zapisem

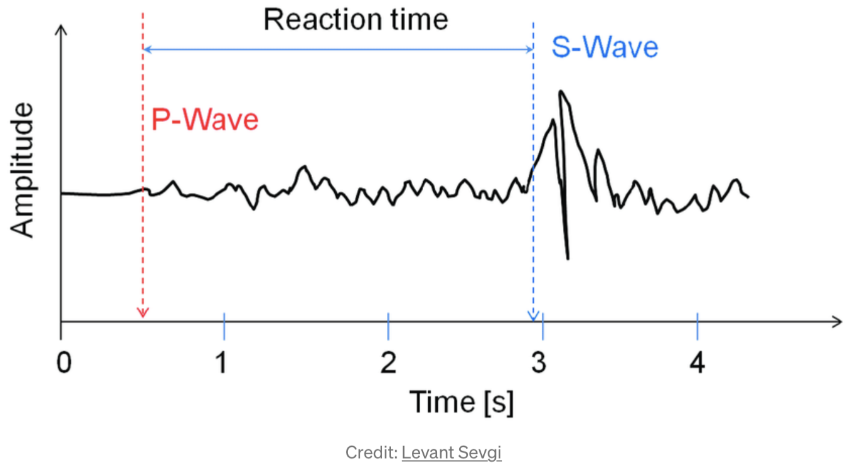
$$Y = Y_t : t \in T \tag{2}$$

gdzie T to zbiór indeksujący

1.2 Przykłady szeregów czasowych

Szeregi czasowe są wykorzystywane m.in. w

- przetwarzaniu sygnałów
- matematyce finansowej
- ekonometrii
- prognozowaniu pogody
- przewidywaniu trzęsień ziemi



Rysunek 2: Szereg czasowy pokazujący przebiegu trzęsienia ziemi

2 Analiza szeregów czasowych

2.1 Cele analizy

Do głównych celów analizy należą:

- prognozowanie (przewidywanie przyszłych wartości szeregu czasowego)
- wykrywanie natury zjawiska, którego reprezentacją jest szereg czasowy

Aby zrealizować powyższe cele należy zidentyfikować i opisać elementy szeregu czasowego.

2.2 Główne elementy szeregu czasowego

Szeregi czasowe można charakteryzować za pomocą wielu elementów, jednak na potrzeby naszej analizy tego problemu skupimy się na poniższych trzech:

- stacjonarność (stationarity)
- autokorelacja (autocorelation)
- sezonowość (seasonality)

2.3 Stacjonarność

Stacjonarność należy do najważniejszych charakterystyk szeregów czasowych. Określa ona czy szereg posiada stałą średnią i wariancję oraz że jego kowariancja jest niezależna w czasie.

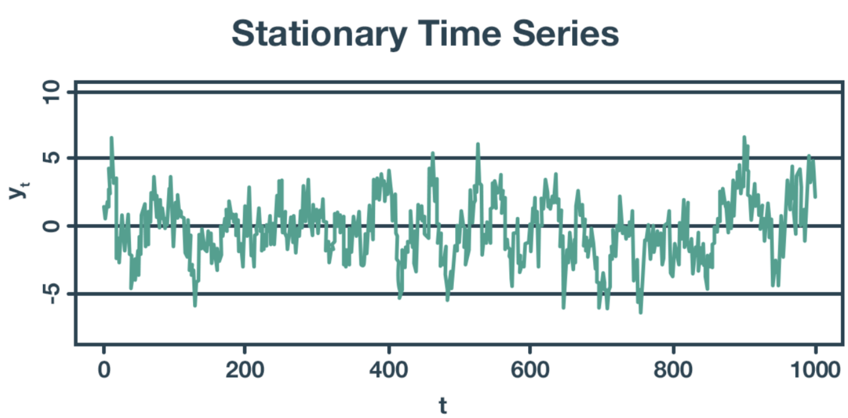


Fig.1: Stationary Time Series

Rysunek 3: Przykład stacjonarnego szeregu czasowego

Powyższy rysunek przedstawia szereg czasowy, którego średnia i wariancja nie zmienia się w czasie. Stacjonarne szeregi czasowe są porządane podczas procesu modelowania. Mimo iż nie wszyskie szeregi czasowe są stacjonarne jest możliwe uczynienie ich stacjonarnymi za pomocą różnych przekształceń.

2.4 Testowanie stacjonarności

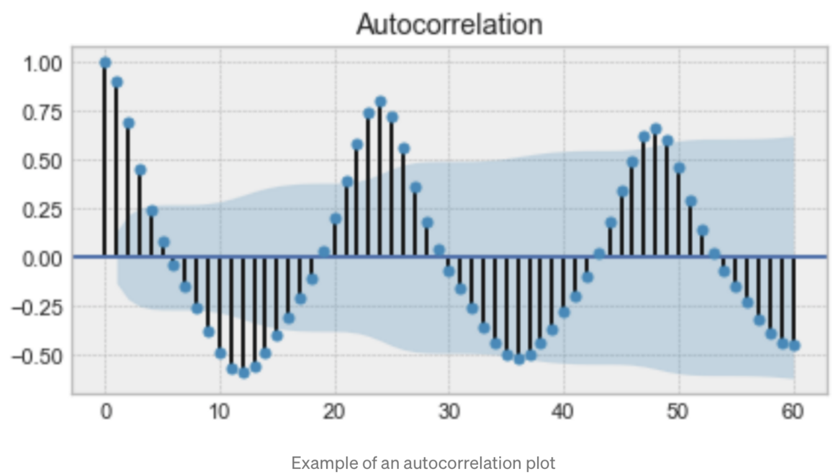
Do testowania stacjonarności można wykorzystać test Dickeya-Fuller'a, który został opracowany w 1979 roku przez D.A.Dickeya i W.A. Fullera.
Bazowy model ma postać:

$$y_t =_{t-1} + u_t \tag{3}$$

gdzie y_t to zmienna objaśniana, t indeks czasowy, ρ współczynnik, a u_t błąd oszaczowania (biały szum). Hipotezą zerową jest to, że występuje pierwiastek jednostkowy w modelu. Gdy $p > 0$, wtedy szereg jest stacjonarny, w przeciwnym przypadku odrzucamy hipotezę zerową i uznajemy szereg za stacjonarny.

2.5 Autokorelacja

Potocznie mówi się, że autokorelacja pokazuje podobieństwo między obserwacjami jako funkcja odstępu czasu między nimi.



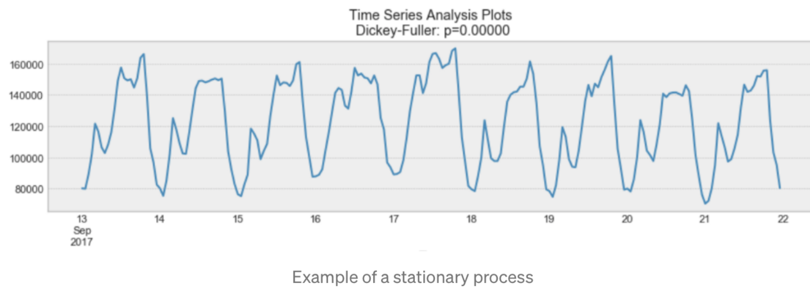
Rysunek 4: Przykład zautokorelowanego szeregu czasowego

Na powyższym szeregu doskonale widać, że wartości między którymi jest odpowiedni odstęp (stały) są silnie skorelowane.

2.6 Sezonowość

Sezonowość szeregu czasowego objawia się w okresowych wahaniach jego wartości. Do szeregów, na których silnie widoczna jest sezonowości należą:

- Szereg pokazujący wartości temperatury na przestrzeni lat (będą występować sezonowe minima i maksima)
- Szereg pokazujący zmiany w zapotrzebowaniu energii w ciągu dnia (wzrosty w dzień, spadki w nocy)
- Szereg pokazujący przychody sklepów internetowych (sezonowy wzrosty w okresie świątecznym)



Rysunek 5: Przykład sezonowości w szeregu czasowym - konumpcja energii na przestrzeni dni

Na powyższym wykresie widać, że spadki i wzroty konsumpcji występują w regularnych porach w ciągu dnia. Wzrosty występują do wieczora, natomiast najniższe zużycie ma miejsce na początku i pod koniec dnia.

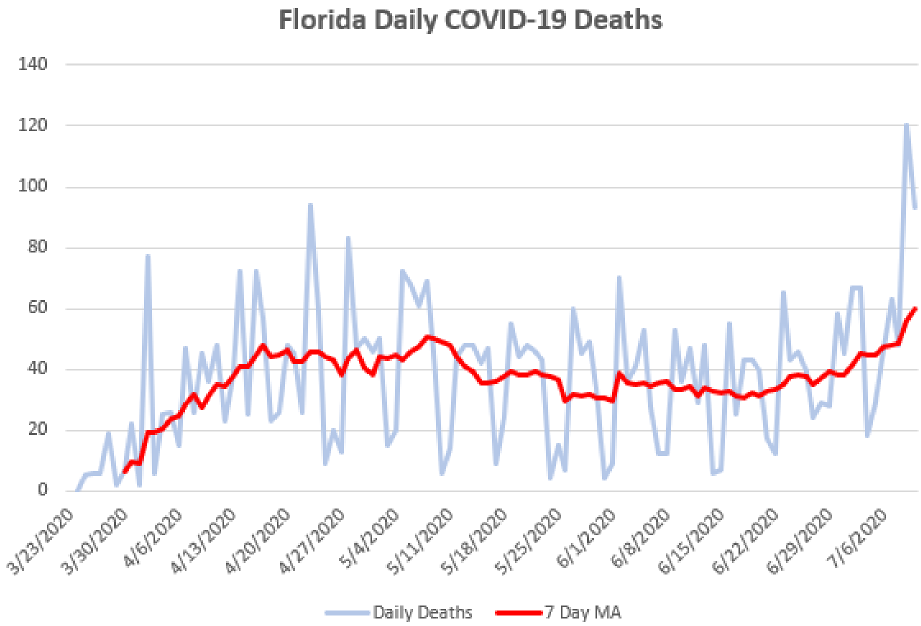
3 Modelowanie szeregów czasowych

Istnieje wiele sposobów na modelowanie szeregów czasowych aby wykonywać przewidywania. Tutaj przybliżymy:

- średnią ruchomą
- wygładzenie wykładnicze
- zintegrowany model autoregresyjny ze średnią ruchomą (ARIMA)
- analizę opóźnień
- analizę pojedynczego widma Fouriera
- wyrównanie wykładnicze
- dekompozycję sezonową

3.1 Średnia ruchoma

Jest to podejście najbardziej naiwne. Ten model przewiduje, że następna obserwacja będzie średnią z wszystkich wcześniejszych obserwacji. Model ten może być przydatny do wygładzenia szeregu czasowego, oraz do zauważenia ogólnych tendencji danych.



Rysunek 6: Model średniej ruchomej wraz z danymi

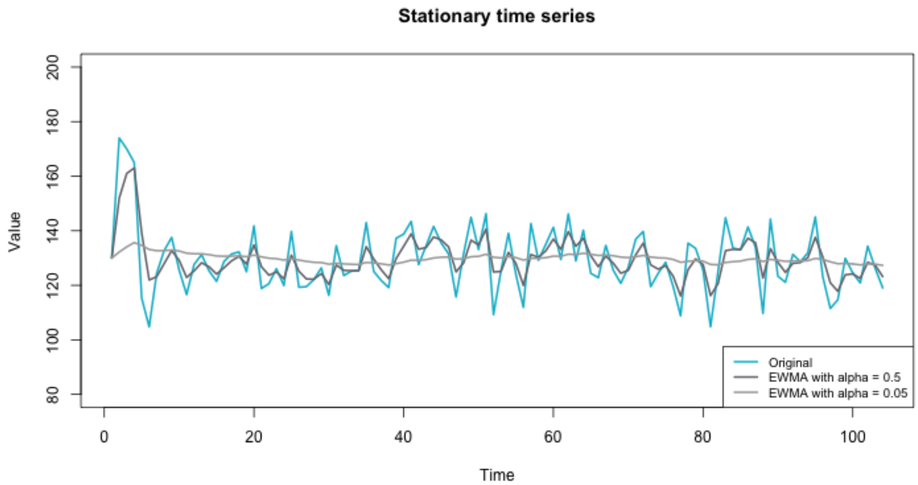
3.2 Wygładzenie wykładnicze

Polega ono na podobnym podejściu jak średnia ruchoma, ale tym razem każda obserwacja ma wagę, która jest tym mniejsza im dalej jest oddalona od czasu, który w tym momencie analizujemy. Wygładzenie wykładnicze możemy wyrazić wzorem:

$$y = \alpha x_t + (1 - \alpha) y_{t-1}, t > 0$$

Rysunek 7: Wzór na wygładzenie wykładnicze

Gdzie alfa jest czynnikiem wygładzającym z zakresu (0,1) i od niego zależy jak bardzo maleje waga dla przeszłych obserwacji.

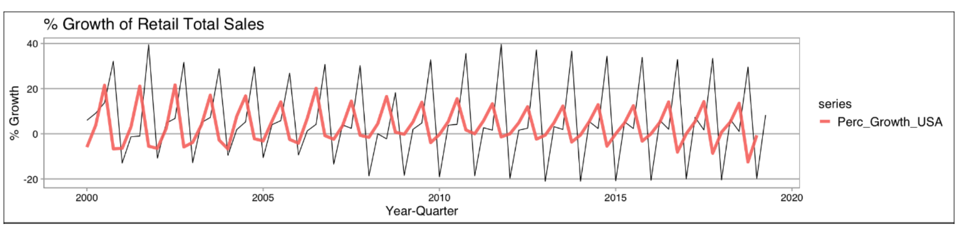


Rysunek 8: Model wygładzenia wykładniczego dla dwóch różnych wartości alfa

3.3 Zintegrowany model autoregresyjny ze średnią ruchomą (ARIMA)

Ten model powstał z kombinacji prostszych modelów:

- Zaczynamy od stworzenia modelu autoregresyjnego
- Następnie dodajemy model średniej ruchomej
- Na koniec dodajemy jeszcze sezonowość



Rysunek 9: Zintegrowany model autoregresyjny ze średnią ruchomą

3.4 Analiza z uwzględnieniem opóźnień

Jest to specjalistyczna technika wykorzystywana gdy zmienne mogą zależeć od siebie z pewnym opóźnieniem w czasie. Na przykład weźmy pod uwagę ilość osób zaszczepionych na COVID-19 względem ilości nowych przypadków. Opóźnienieniami w tym przypadku są zdobywanie odporności przez osoby zaszczepione jak i okres pomiędzy zakażeniem a wystąpieniem pierwszych objawów u osoby chorej. Takie problemy często występują w ekonometrii gdzie rynek/giełda reaguję z pewnym opóźnieniem na wydarzenia na świecie. A więc oczekujemy czasowo opóźnionej korelacji między danymi.

Ogólny model:

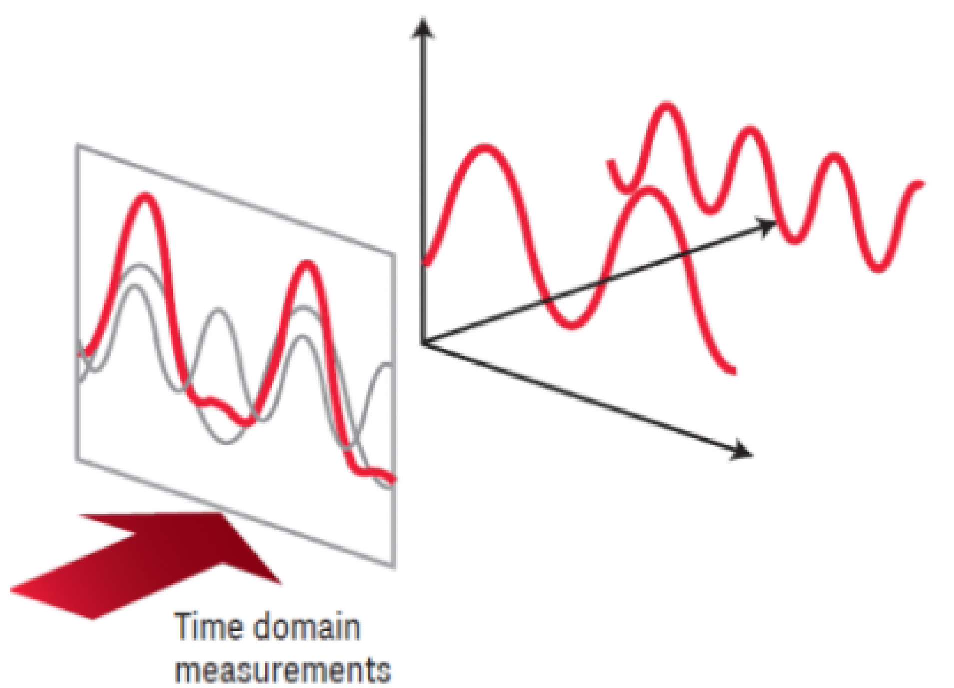
- opisujemy związek za pomocą prostej zależności liniowej

$$Y_t = \sum \beta_i * x_{t-i} \tag{4}$$

- w równaniu tym, wartość zmiennej Y w czasie t jest wyrażona jako funkcja liniowa x mierzonego w czasach poprzedzających t
- Wagi beta to parametry strukturalne modelu

3.5 Analiza z pojedynczego widma (Fouriera)

Dzięki tej analizie wykonujemy dekompozycję złożonego szeregu czasowego na kilka funkcji podstawowych sinusoidalnych. Stosując tą analizę możemy szereg, który początkowo wygląda na losowy rozbić na szereg funkcji podstawowych, które łatwiej analizować.



Rysunek 10: Szereg czasowy zdekomponowany na dwie podstawowe funkcje sinusoidalne

3.6 Wrównanie wykładnicze

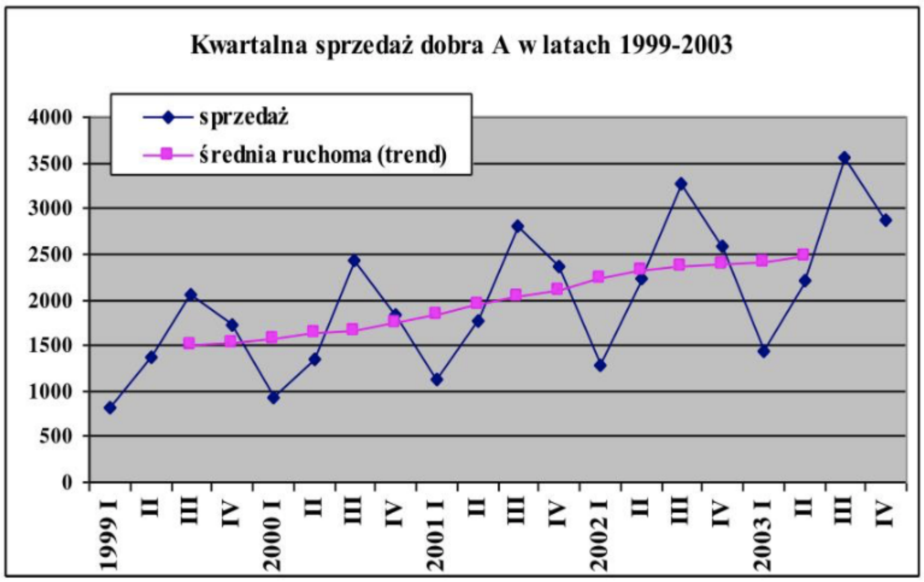
Do skorzystania z tej metody musimy przyjąć model, w którym każda obserwacja w szeregu czasowym ma postać:

$$X_t = b + \epsilon_t \tag{5}$$

Gdzie b to pewna stała a ϵ_t to składnik losowy. Stała b jest dość stabilna w czasie, co pozwala nam na przewidzenie jej wartości poprzez obliczenie średniej ruchomej między bieżącą a poprzedzającą obserwacją. Wyrównanie wykładnicze działa w ten sposób, że istotność dawnej obserwacji spada w czasie w sposób wykładniczy. Wzór takiego szacowania miałby postać:

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1} \tag{6}$$

gdzie X_t to zaobserwowana wartość szeregu a S_t wartości wygładzone



Rysunek 11: Przykład wygładzania szeregu czasowego

3.7 Dekompozycja sezonowa (Census I)

Dekompozycja sezonowa zakłada, że szereg czasowy składa się z 4 składników:

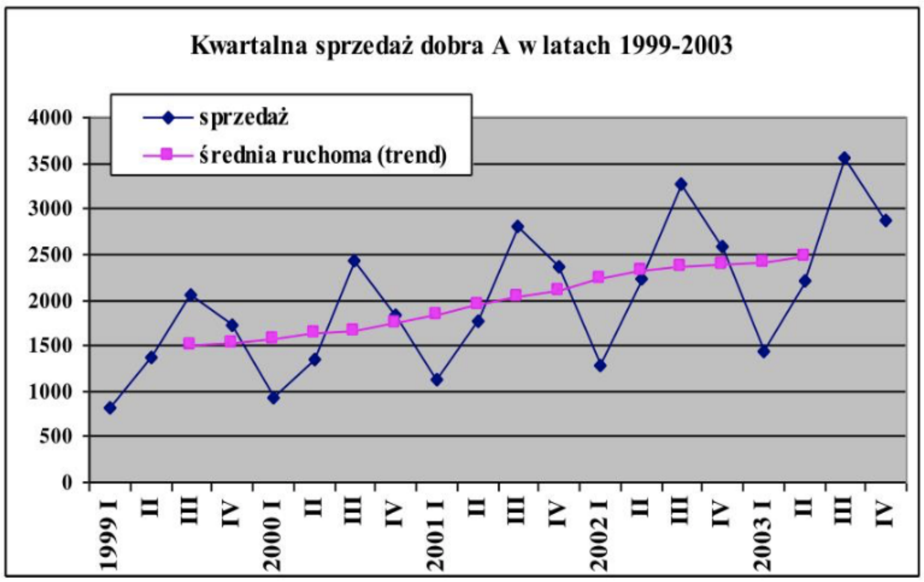
- składnika sezonowego S_t
- składnika trendu T_t
- składnika cyklicznego C_t
- składnika losowego błędu I_t

Róznica między składnikiem cyklicznym i sezonowym polega na tym,że składnik sezonowy pojawia się w regularnych odstępach, z kolei cykle charakteryzują się dłuższym czasem trwania oraz zmiennością czasu trwania cyklu. Zależność funkcyjna między trendem a cyklicznością mogą zostać przedstawione za pomocą dwóch modelów

- Model addytywny: $X_t = TC_t + S_t + I_t$
- Model multiplikatywny: $X_T = T_t * C_t * S_t * I_t$

Podczas dekompozycji sezonowej można korzystać z:

- średniej ruchomej o długości okresu wahań
- Ilorazów(dla modelu multiplikatywnego) i róznic(model addytywny) szeregu i średniej ruchomej
- Wskaźnik sezonowy będący średnią(model addytywny) lub średnią środkową(model multiplikatywny) dla każdego punktu w sezonie



Rysunek 12: Przykład wygładzania szeregu czasowego

4 Bibliografia

- https://en.wikipedia.org/wiki/Time-series
- https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775
- https://www.statsoft.pl/textbook/stathome-stat.html
- https://pl.wikipedia.org/wiki/Szereg-czasowy