

Analiza szeregów czasowych

Time Series Analysis

część 1



Jerzy Stefanowski
Politechnika Poznańska

Projekt eksploracji danych
Poznań 2020

Plan

1. Uwagi wstępne (szeregi, przykłady, prognozowanie,...)
2. Cel analizy szeregów czasowych
3. Struktura szeregów czasowych
(trend/składowa stała, wahania sezonowe, wahania cykliczne)
4. Podstawowe modele matematyczne
 - addytywne
 - multiplikatywne
5. Szereg czasowy bez trendu
6. Szereg czasowy z trendem (wyodrębnianie trendu)
7. Analiza sezonowości
8. Zagadnienia prognozowania

Wprowadzenie

Time-Series – dane zmieniające się wraz z upływem czasu;
dane zawierające serie (szeregi) wartości / wielkości
zmieniających się w czasie

Szereg czasowy – ciąg obserwacji pewnego zjawiska w
kolejnych jednostkach czasu [def. statystyczna]

y_t ; gdzie $t = 1, 2, \dots$

Także – y_t realizacja pewnego procesu stochastycznego $\{Y_t\}$

Ogólna postać szeregu czasowego

...Czas (t)	Zjawisko (y_t)
t_1	y_1
t_2	y_2
...	...
t_n	y_n

Wprowadzenie

Szeregi czasowe – jednowymiarowe i wielowymiarowe

Typowe cele:

Analiza dynamicznych własności szeregu

Identyfikacja struktury szeregu czasowego

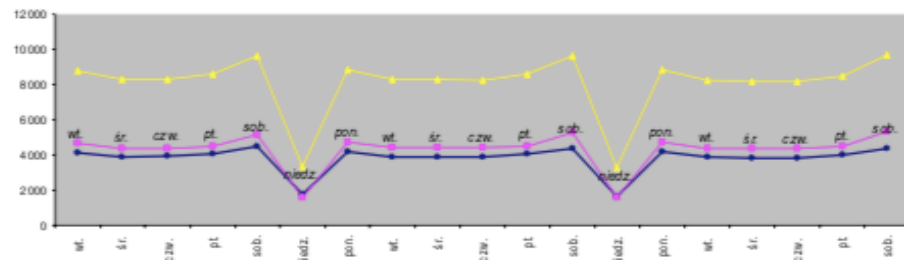
Prognozowanie (przewidywanie nowych wartości)

Także – monitorowania i sterowanie procesami

Jednowymiarowe:

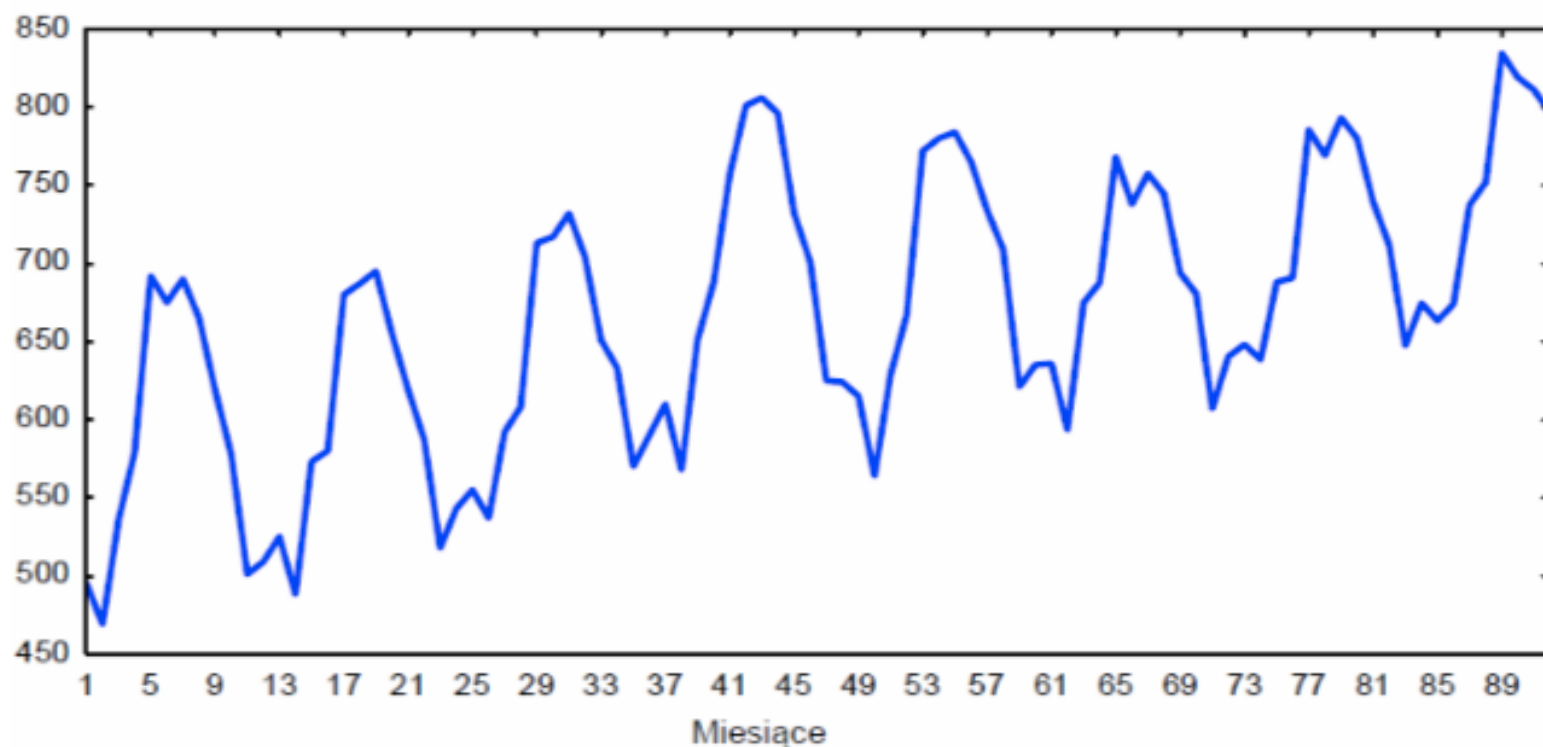
- Wnioskowanie na temat dynamiki zmiennej y jest przeprowadzane jedynie na podstawie przeszłych i bieżących obserwacji dla tej zmiennej.
- Kształt wykresu niesie informacje

Sprzedaż artykułu w dniach 12-30.11.2013 r.



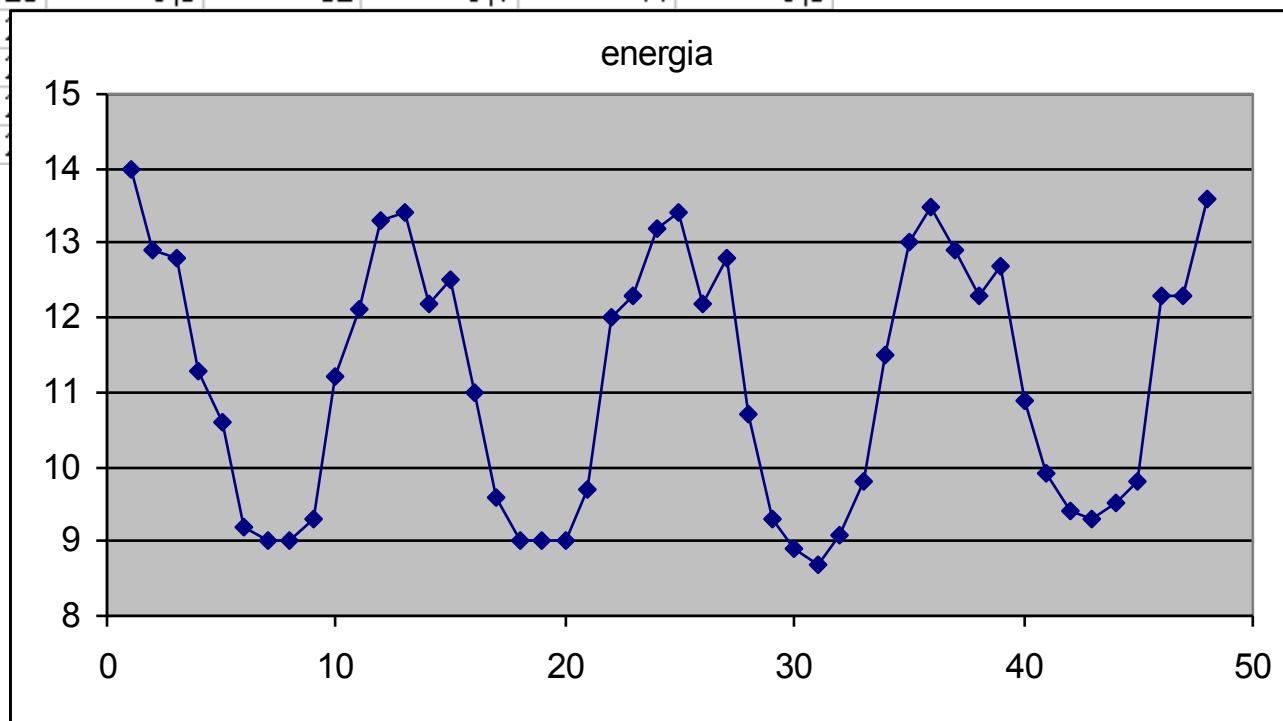
Przykłady

Wielkość skupu mleka w Polsce w latach 2002-2009
(w mln l; dane miesięczne)



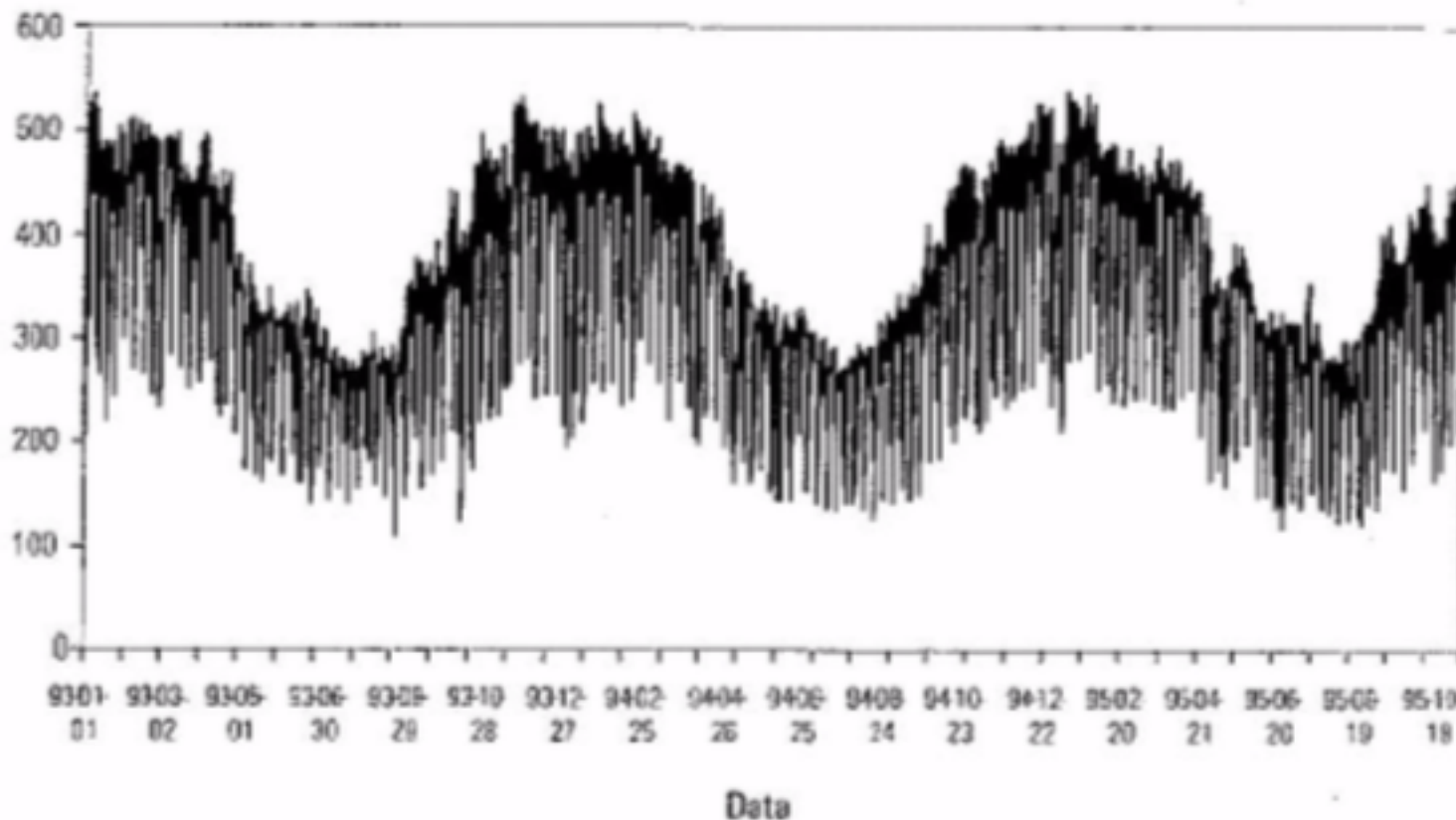
wielkość produkcji energii elektrycznej w kolejnych miesiącach lat 1991-94

1991		1992		1993		1994	
czas	energia	czas	energia	czas	energia	czas	energia
1	14	13	13,4	25	13,4	37	12,9
2	12,9	14	12,2	26	12,2	38	12,3
3	12,8	15	12,5	27	12,8	39	12,7
4	11,3	16	11,0	28	10,7	40	10,9
5	10,6	17	9,6	29	9,3	41	9,9
6	9,2	18	9,0	30	8,9	42	9,4
7	9	19	9,0	31	8,7	43	9,3
8	9	20	9,0	32	9,1	44	9,5
9	9,3						
10	11,2						
11	12,1						
12	13,3						



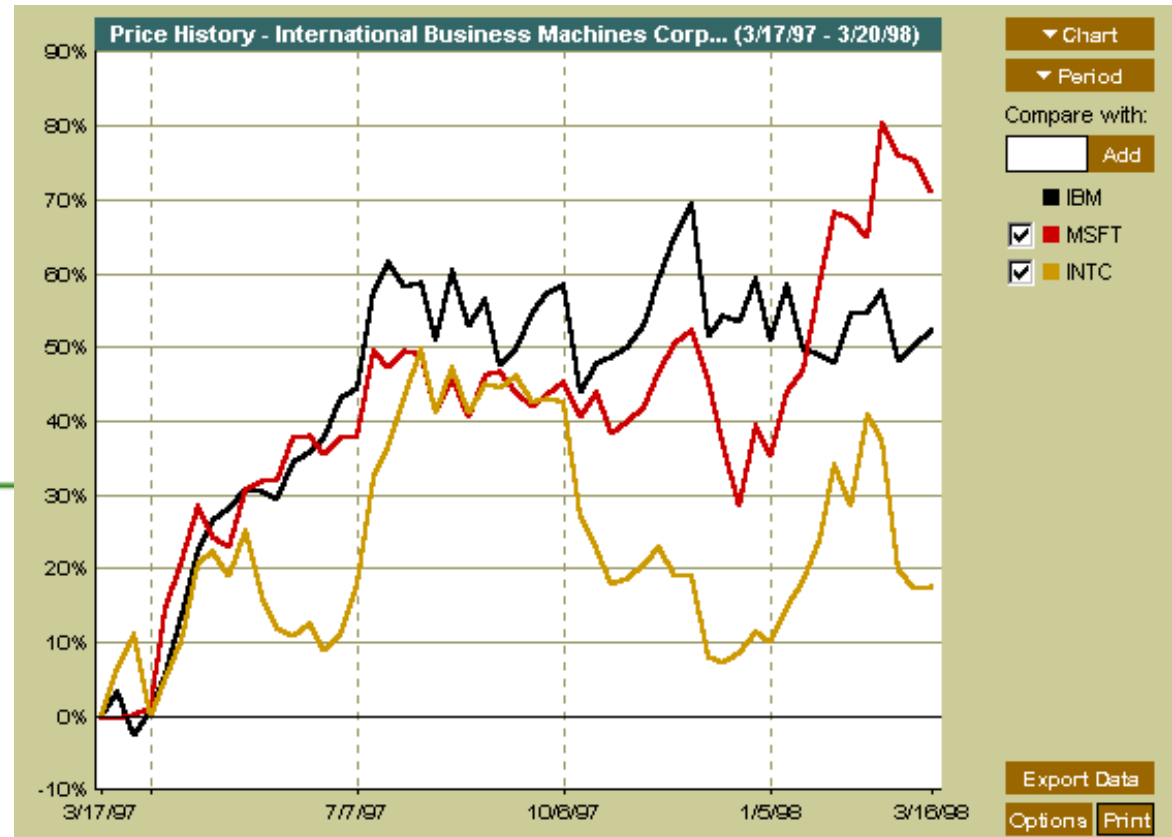
Energia w Łodzi

Zużycie energii elektrycznej w Łodzi w latach 1993-95
(w MW; dane dzienne)



Dane giełdowe

Date	Stock	Price\$
June 11, 93	IBM	98.5
June 11, 93	MSFT	78.0
June 11, 93	INTC	76.5
June 12, 93	IBM	99.5
June 12, 93	MFST	80.0
June 12, 93	INTC	77.0
June 13, 93	IBM	98.0
:	:	:
:	:	:



Zastosowania

- Analiza danych giełdowych i finansowych
- Modelowanie zjawisk ekonomicznych
- Opracowywanie danych GUS (spójrz *Roczniki Statystyczne* lub *Biuletyny Statystyczne*).
- Wspomaganie decyzji w zarządzaniu przedsiębiorstwami, w szczególności tworzenie **prognozy sprzedaży** a także analiza **dynamiki** procesów produkcyjnych, zaopatrzenia, zapasów, finansów, siły roboczej.
- Analiza danych diagnostycznych i prognozy postępowania w medycynie
- Diagnostyka techniczna stanu urządzeń
- Analiza wyników eksperymentów naukowych.
- ...

Cel analizy szeregów czasowych

- Zbudowanie modelu pewnego zjawiska/procesu w oparciu o obserwowane zmiany w czasie mierzalnych wielkości opisujących ten proces.
- Ogólne założenie: obserwowany przebieg składa się z:
 - **Części systematycznej** (trend, składowa stała, wahania sezonowe i cykliczne) – w oparciu, o które buduje się model.
 - **Części przypadkowej** (szumu, wahań przypadkowych).
- Wymienione składniki – czynniki determinujące rozważane zjawisko;
- W analizie szeregów dąży się do ich wyodrębnienia i pomiaru – **dekompozycja szeregu czasowego**.
- Przy użyciu otrzymanego modelu można dokonywać **predykcji** (prognozowania) przebiegu szeregu lub jego składowych w nowych momentach czasu

Typowe metody

- Modele naiwne
- Średnie ruchome
- Wygładzanie wykładnicze
- Analiza harmoniczna
- Klasyczna dekompozycja
- Modele z trendem i wahaniami sezonowymi
- ARIMA
(autoregresyjny zintegrowany proces średniej ruchomej)
- Modele ze zmiennymi objaśniającymi
- Modele ze zmiennymi opóźnionymi w czasie
- (...)

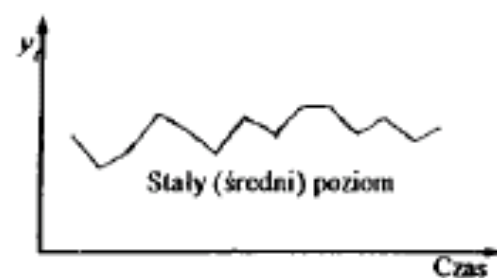
Podstawowa struktura szeregów czasowych

- **Stały** (przeciętny) poziom zmiennej.
- **Trend** (tendencja rozwojowa) – reprezentuje ogólny kierunek rozwoju zjawiska (systematyczne zmiany, jakim podlega zjawisko); rozróżnia się, np., trend liniowy lub nieliniowy.
- **Składowa okresowa** (wahania okresowe / regularne odchylenia od tendencji rozwojowej) – składnik powtarzający się cyklicznie.
- **Szum** (zakłócenia, wahania przypadkowe).

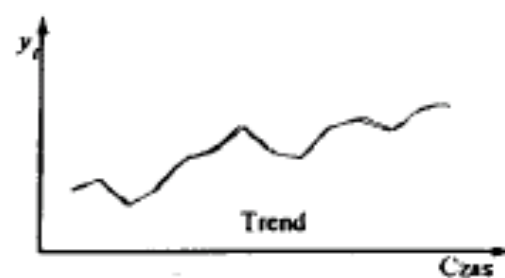
Składowa okresowa

- **cyklicznych** - długookresowe, rytmiczne wahania (cykl koniunkturalny gospodarki, cykl rozwoju populacji nabywców danego produktu, itp.),
- **sezonowych** – krótkookresowe do 1 roku, odzwierciedlają wpływ zachowań wynikający z „kalendarza” (np. rytm pracy w skali tygodnia, dnia, pory roku, świąt, ...)

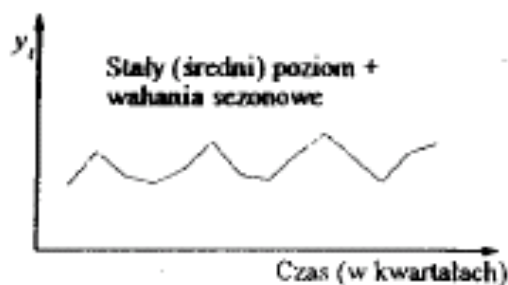
a)



b)



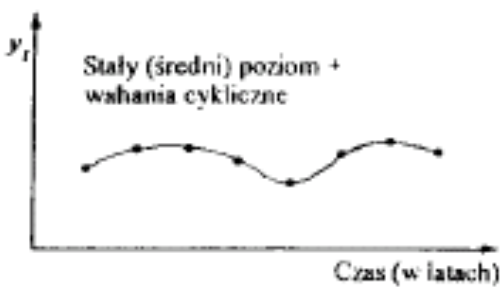
c)



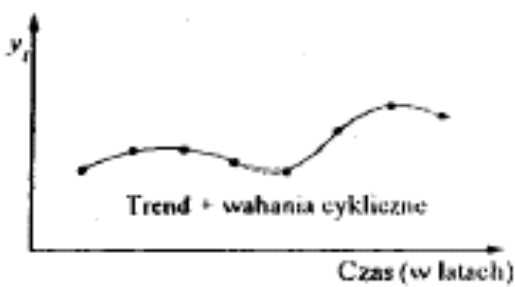
d)



e)



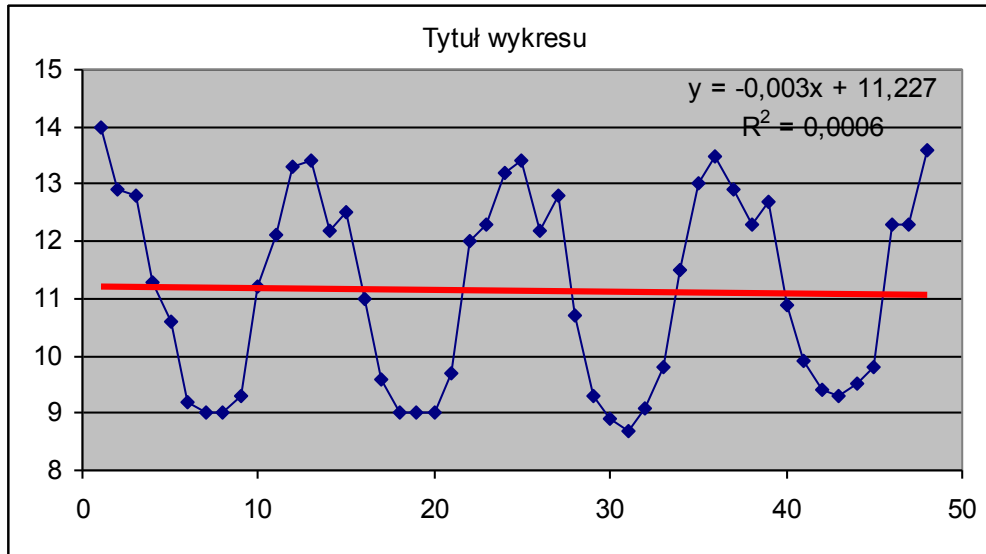
f)



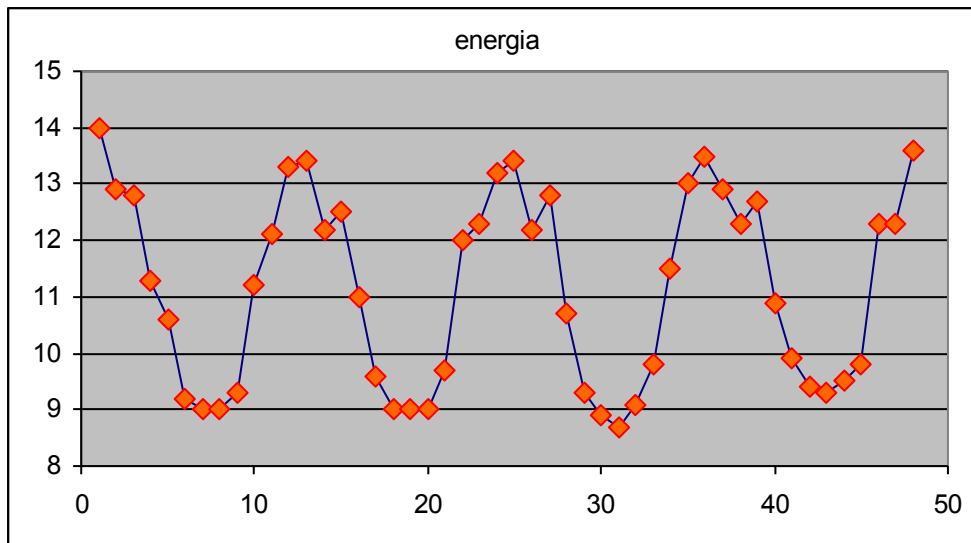
Rys. 4.2. Szeregi czasowe z różnymi rodzajami składowych

Dekompozycja szeregu czasowego

-przykład produkcji energii



Trend - Tendencja rozwojowa



Wahania okresowe (roczne)

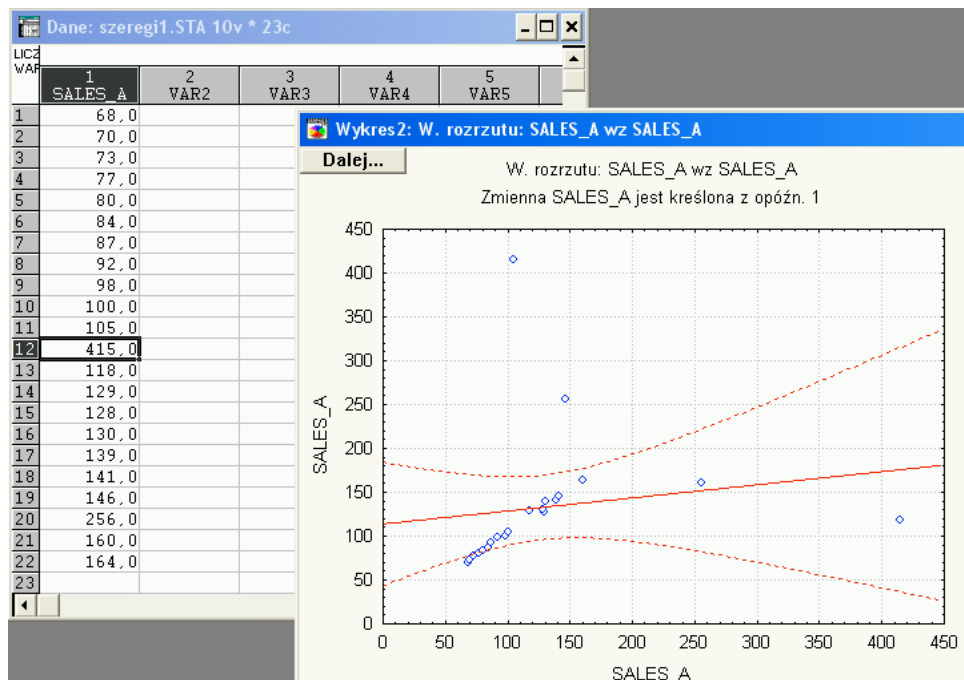
Założenia odnośnie danych

- Jednostki czasu użyte do pomiarów powinny być równe (szereg okresów).
- Dane historyczne powinny dotyczyć okresu, w którym parametry modelu opisującego proces były (choćby w przybliżeniu) stałe.
- „Im więcej danych, tym lepiej”,
- Problemy z uwzględnianiem „kalendarza”, np.
 - standaryzacja długości miesiąca,
 - standaryzacja długości tygodnia,
 - uwzględnianie dni świątecznych.
- Występowanie obserwacji odstających i brakujących:
 - brak pomiaru lub dzień świąteczny,
 - silny wpływ rzadko występującego czynnika (np. jednorazowa realizacja bardzo dużego zamówienia, awaria urządzenia produkcyjnego).

Identyfikacja obserwacji odstających

Dla stacjonarnego szeregu czasowego:

- Oblicz pierwszy (Q_1) i trzeci (Q_3) kwartył z szeregu y_t
- $R_Q = Q_3 - Q_1$; rozstęp międzykwartyłowy
- Obserwacje odstające spoza przedziałów ($Q_1 - 3R_Q$; $Q_3 + 3R_Q$)
- Przykład (Dittmann) sprzedaży miesięcznej produktu A



Uwagi ogólne

- **Dekompozycja szeregu** to proces wyodrębnienia poszczególnych składowych danego szeregu czasowego.
- Identyfikację poszczególnych składowych szeregu czasowego zmiennej umożliwia:
 - **Ocena wzrokowa** sporządzonego wykresu,
 - **Analiza autokorelacji**,
 - **Odpowiednie testy** (np. Danielsa, Bartletta lub von Neumanna).
- Dla wielu szeregów czasowych wystarczająco adekwatne mogą się okazać modele ujmujące tylko niektóre składowe szeregu czasowego

Prognozowanie - Metody naiwne

- Obserwacja z wybranego okresu historycznego stanowi najlepszą prognozę dla przyszłej wartości, tj. "jutro będzie tak jak dziś".
- W najprostszej postaci prognoza naiwna dla okresu t wynosi tyle samo co obserwacja w poprzednim okresie, tj. $t-1$
- Model naiwny można rozbudować tak, żeby był bardziej odpowiedni dla danych z trendem lub sezonowością:
 - Sezonowość można uwzględnić poprzez odwołanie się do obserwacji sprzed okresu (np. sprzed roku)
 - Trend opisany zostaje przez dodanie do poprzedniej obserwacji różnicy między dwoma poprzednimi obserwacjami

Podstawowe modele szeregów

W ogólności przyjmuje się addytywną lub multiplikatywną formę modelu.

- **Model addytywny:**

$$y_t = f(t) + g(t) + h(t) + \zeta_t$$

- **Model multiplikatywny:**

$$y_t = f(t) \cdot g(t) \cdot h(t) \cdot \zeta_t$$

- gdzie $f(t)$ – funkcja trendu,
 $g(t)$ – funkcja czasu charakteryzująca wahania sezonowe,
 $h(t)$ – funkcja czasu charakteryzująca wahania cykliczne,
 ζ_t – składnik losowy.

- Stosuje się także modele mieszane.

Metody wyrównania mechaniczne

- Wyrównanie szeregu czasowego pozwala na wyeliminowanie z szeregu wahań przypadkowych, a przy odpowiednim postępowaniu także wahań okresowych.
- Inna nazwa → wygładzanie szeregów czasowych (ang. smoothing)
- Często stosowane:
 - Średnie ruchome (kroczące)
 - zcentrowane i niezcentrowane
 - Wygładzanie wykładnicze

Średnie ruchome zwykłe

- Oblicza się najczęściej z nieparzystej liczby sąsiadujących ze sobą wyrazów szeregu, tak aby uzyskany wynik móc przyporządkować wartości t znajdującej się w środku uwzględnionego w obliczeniach przedziału czasowego:

$$\bar{y}_t = \frac{1}{2q+1} \sum_{r=-q}^q y_{t+r} \quad (t = q, q+1, \dots, n-q)$$

- gdzie $2q+1$ to liczba wyrazów szeregu uwzględnianych przy obliczaniu średniej ruchomej, przy czym q jest ustalony liczbą naturalną
- Dla parzystej liczby sąsiadujących ze sobą wyrazów szeregu, uwzględnia się połowę wartości pierwszego wyrazu z danego okna, następnie wszystkie pozostałe wyrazy składające się na pełne okno oraz połowy wartości pierwszego wyrazu z następnego okna:

$$\bar{y}_t = \frac{1}{2q} \left[\frac{1}{2} y_{t-q} + \sum_{r=-q+1}^{q-1} y_{t+r} + \frac{1}{2} y_{t+q} \right]$$

Średnie zcentrowane

- **zcentrowana** – średnia arytmetyczna n pomiarów wokół punktu t (czyli $y_{t-n/2}, \dots, y_{t+n/2}$)
- średnia ruchoma **ważona** – preferencja dla aktualnych wartości (większe wagi)
- Zamiast średniej arytmetycznej stosuje się także medianę.

Przykład

średnia ruchoma 3-okresowa $\bar{y}_t = (y_{t-1} + y_t + y_{t+1})/3$

średnia ważona – wagi (1,4,1); np. $\frac{1 \times 3 + 4 \times 7 + 1 \times 2}{1 + 4 + 1} = 5.5$

Dane oryginalne	3	7	2	0	4	5	9	7	2
Średnia ruchoma		4	3	2	3	6	7	6	
Średnia ważona		5.5	2.5	1	3.5	5.5	8	6.5	

Przykład ilustracyjny

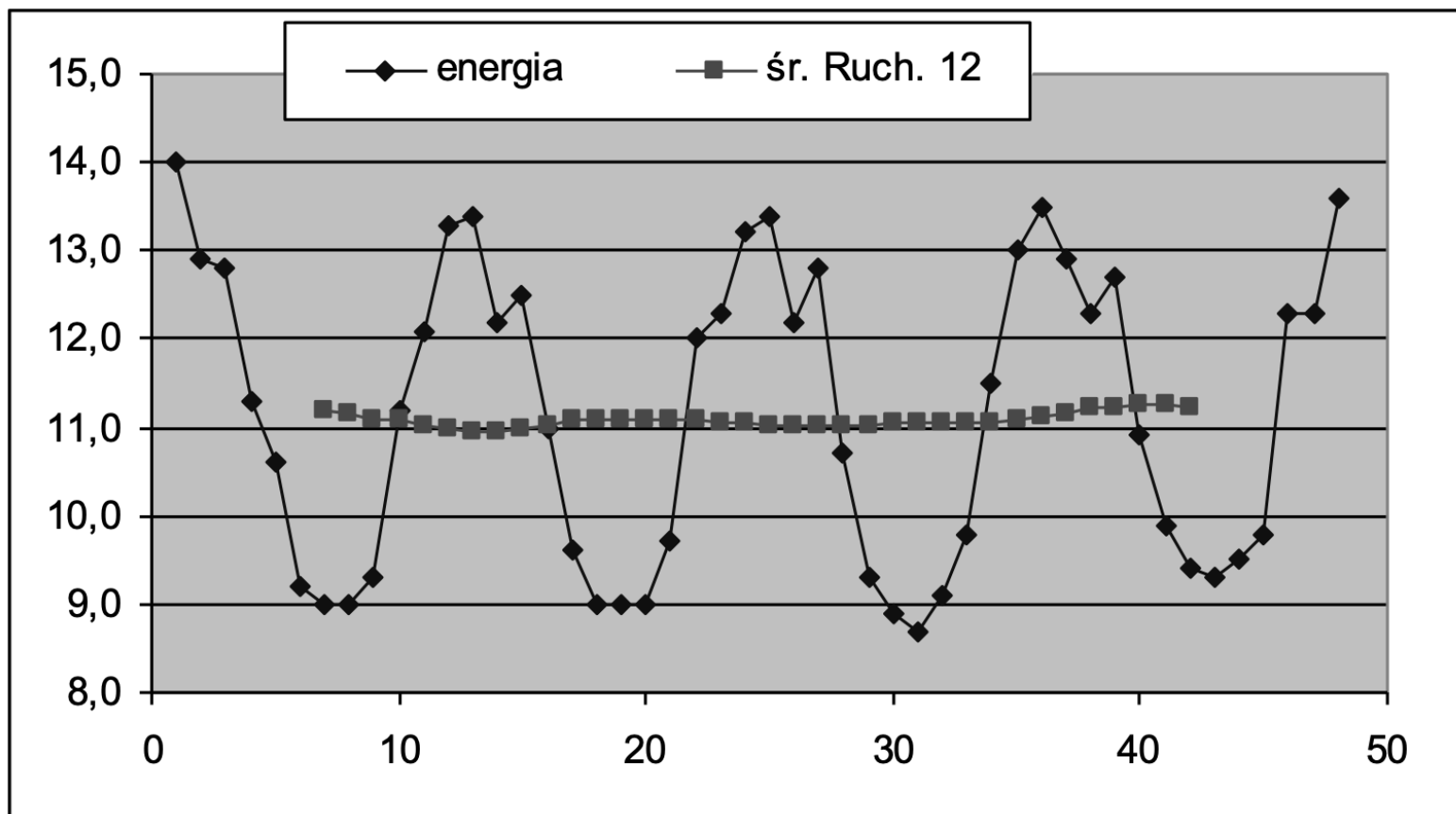
za „Seasonal Adjustment” Christian Harhoff (materiał Statistics Denmark)

Moving Averages in the additive model

Year	Quarter	Time Series X	Moving Average G
1	1	5,0	...
	2	6,0	...
	3	6,5	6,26
	4	6,3	6,86
2	1	7,5	7,39
	2	8,3	7,81
	3	8,4	8,15
	4	7,8	8,44
3	1	8,7	8,74
	2	9,4	9,13
	3	9,7	...
	4	9,6	...

$$G_{1,3} = \frac{\frac{1}{2}5,0 + 6,0 + 6,5 + 6,3 + \frac{1}{2}7,5}{4} = 6,26$$

Średnia – wygładzanie / trend



Średnie ruchome

- Stosuj, gdy zaobserwowany w okresie badawczym poziom wartości zmiennej prognozowanej jest względnie stały, z pewnymi niewielkimi odchyleniami przypadkowymi (silna stacjonarność)
- **Idea:** wartość zmiennej prognozowanej jest średnią ruchomą z k ostatnich wartości tej zmiennej

$$y_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i$$

- k – stała wygładzania (przyjmuje się tę, dla której wartość średniego błędu ex-post prognoz wygasłych jest najmniejsza)

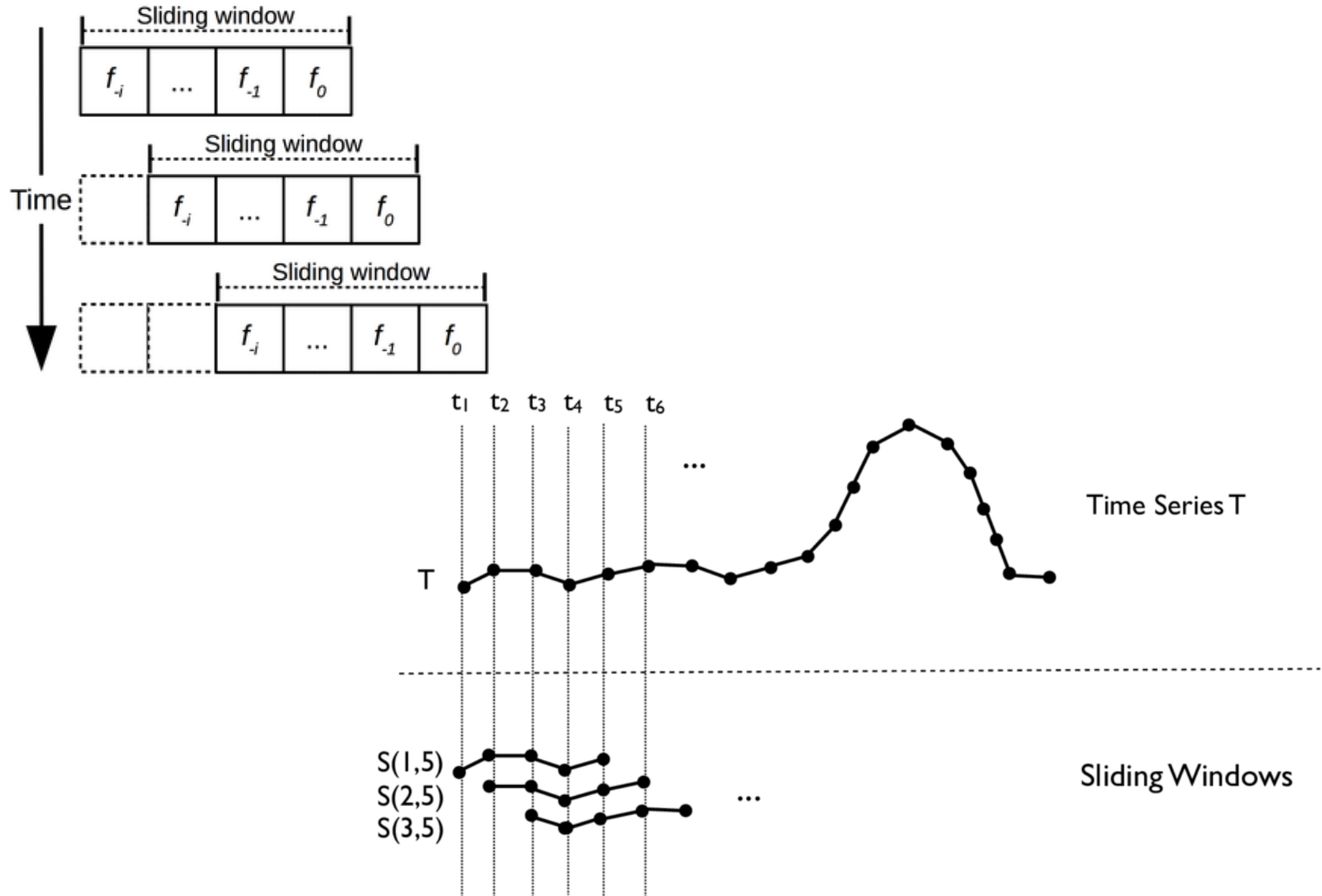
Szereg czasowy sprzedanych samochodów w pewnym salonie

Tydzień (t)	Sprzedaż w szt. y_t
1	15
2	17
3	12
4	16
5	15
6	11
7	18
8	17
9	13
10	16

- Jakiej sprzedaży samochodów możemy oczekiwać w kolejnych tygodniach, np. w 11 tygodniu?
- Średnia arytmetyczna** wyznaczona ze wszystkich elementów szeregu - szt. Inne podejście - **metoda średniej ruchomej**

Czteroelementowa $= \frac{15+17+12+16}{4} = 15 \rightarrow$ błąd prognozy $y_t - y_t^*$

Okna przesuwne (sliding windows)



Skumulowane vs. przesuwne okna

Landmark vs. sliding windows

Sliding Window

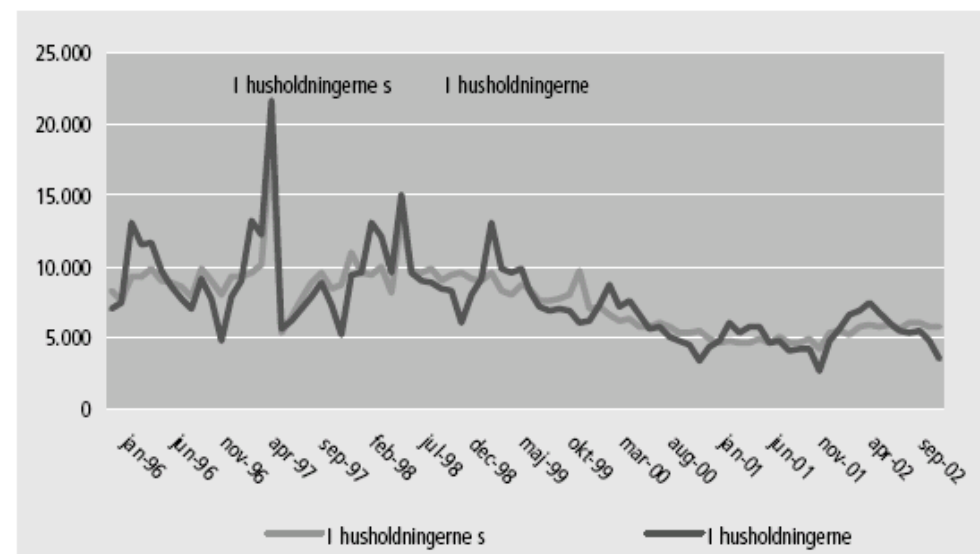
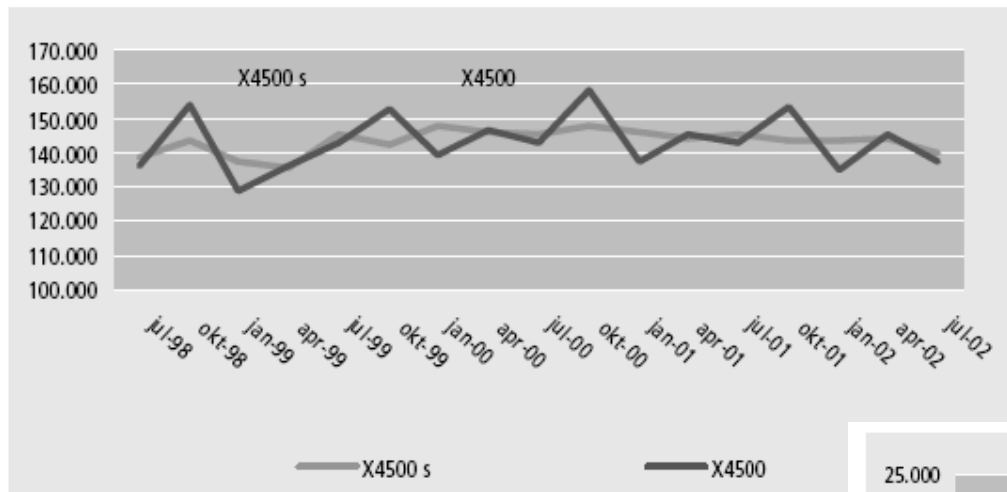


Expanding Window



Prognozowanie

- Wybór i konstrukcja modelu o dobrych własnościach predykcji przyszłych wartości zmiennej y_t .
- Ocena – minimalizacja błędu prognozy *ex post*



Błędy prognozy

1. Mean Squared Error(MSE) :

$$MSE = \sum_{j=1}^N (observation_j - prediction_j)^2 / N$$

2. Root Mean Squared Error($RMSE$) :

$$RMSE = \sqrt{\sum_{j=1}^N (observation_j - prediction_j)^2 / N}$$

3. Normalized Mean Squared Error($NMSE$) :

$$\sum_{j=1}^N (observation_j - prediction_j)^2 / \sum_{j=1}^N (observation_j - mean)^2$$

4. Mean Absolute Percentage Error($MAPE$) :

$$MAPE = 100 \times \frac{\sum_{i=0}^N (|forecasted_i - actual_i|) / actual_i}{N}$$

5. Unbiased Absolute Percentage Error ($UAPE$) aka Symmetric Mean Absolute Percentage Error($SMAPE$):

$$UAPE = \frac{100}{N} \times \sum_{i=0}^N \frac{|forecasted_i - actual_i|}{(actual_i + forecasted_i)/2}$$

6. Median Absolute Percentage Error ($MdAPE$): This measure is almost similar to the $MAPE$, but in $MAPE$ mean is used for summarization whereas in $MdAPE$ median is used for summarization across series.

W literaturze spotkasz także inne propozycje

Przykład sprzedaży samochodów
- szereg czasowy bez trendu.

Średnia ruchoma i błędy prognozy

t	Y_t	Średnia ruchoma cztero- elementowa	Błąd prognozy	(Błąd) ²	Średnia ruchoma pięcio- elementowa	Błąd prognozy	(Błąd) ²
1	2	3	4	5	6	7	8
1	15	-	-	-	-	-	-
2	17	-	-	-	-	-	-
3	12	-	-	-	-	-	-
4	16	-	-	-	-	-	-
5	15	15,0	0	0	0	-	-
6	11	15,0	-4,0	16	15,0	-4	16,00
7	18	13,5	4,5	20,25	14,2	3,8	14,44
8	17	15,0	2,0	4,0	14,4	2,6	6,67
9	13	15,2	-2,2	4,84	15,4	-2,4	5,76
10	16	14,75	1,25	1,563	14,8	1,2	1,44
11	15	16		$\Sigma = 46,653$	15		$\Sigma = 44,4$

Źródło: Na podstawie danych z przykładu 5.1.

Średnia cztero-elementowa = ?

Prognozowanie ze średnimi

- Poziom zmiennej względnie stały bez wahań okresowych
- Dla wyraźnej liniowej tendencji rozwojowej, do konstrukcji prognozy stosuj model podwójnej średniej ruchomej.
- Wygładzony (średnią ruchomą) szereg wartości zmiennej prognozowanej poddaje się powtórnemu wygładzeniu metodą średniej ruchomej.

Gdy zmienna charakteryzuje się wyraźnym trendem :

Należy zastosować metodę podwójnej średniej ruchomej



Wstępne wygładzenie danych – obliczenie średnich ruchomych (zwykłych dla k nieparzystego i scentrowanych dla k parzystego)



Obliczenie prognozy na podstawie wartości wygładzonych, zgodnie z formułą średniej prostej lub ważonej

Uwagi o dobrze stałej k w średnich

- Wraz ze wzrostem wartości k rośnie efekt wyrównywania.
 - Średnia ruchoma wyznaczona z większej liczby wyrazów będzie silniej wygładzała szereg, lecz jednocześnie będzie wolniej reagowała na zmiany poziomu prognozowanej zmiennej.
- Wyznaczona z mniejszej liczby wyrazów będzie szybciej odzwierciedlała aktualne zmiany zachodzące w wartościach prognozowanej zmiennej, lecz większy wpływ będą wywierały na nią wahania przypadkowe (mniejszy będzie efekt wygładzania szeregu)
- Uczenie się wartości k na historycznych danych, np. okna przesuwne (an. sliding windows)

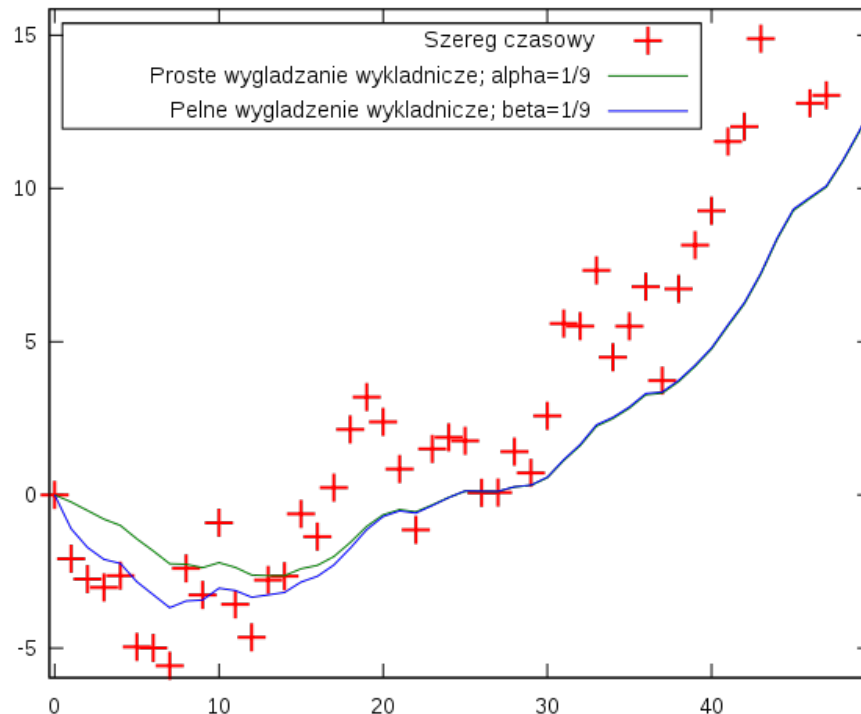
Kilka uwag o średnich ruchomych (kroczących)

Średnie ruchome

- Wygładzają dane (Smooths the data)
- Mogę wyeliminować wahania (sezonowe) i nieregularności
- Pomija się część danych (początkowych)
- W pewnym stopniu wrażliwe na obserwacje nietypowe.

Wyglądanie wykładnicze

- Przydatne do prognozowania szeregów nie mających wyraźnego trendu i wahań sezonowych - gdy są tylko wahania losowe. Wyglądanie przez wpływ ostatnich wartości szeregu na prognozę, z nieliniową redukcją wpływu bardziej odległych obserwacji.
- Największą wagą nadana jest bieżącej obserwacji i mniejsza waga poprzedniej. Wagi zmniejszają się geometrycznie w miarę cofania się w czasie.



Podstawowe wygładzanie wykładnicze

- Zastosowanie → zmienne, których wartości podlegają częstym, „gwałtownym” i raczej przypadkowym wahaniom
- Podstawowa metoda

$$y_t^* = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot y_{t-1}^*, \quad \text{gdzie } \alpha \in (0,1] .$$

Można inaczej zapisać

$$q_{t-1} = y_{t-1} - y_{t-1}^*$$

Wtedy wzór ma postać

$$y_t^* = y_{t-1}^* + \alpha \cdot q_{t-1}$$

α - parametr wygładzania, tym bliższa 1, im zmienna częściej i gwałtowniej się zmienia.

α na ogół dobierana eksperymentalnie na podstawie oceny błędu prognozy

Pierwsza prognoza równa wartości rzeczywistej y

Wygładzenia wykładnicze $\alpha=0,2$ oraz $0,8$

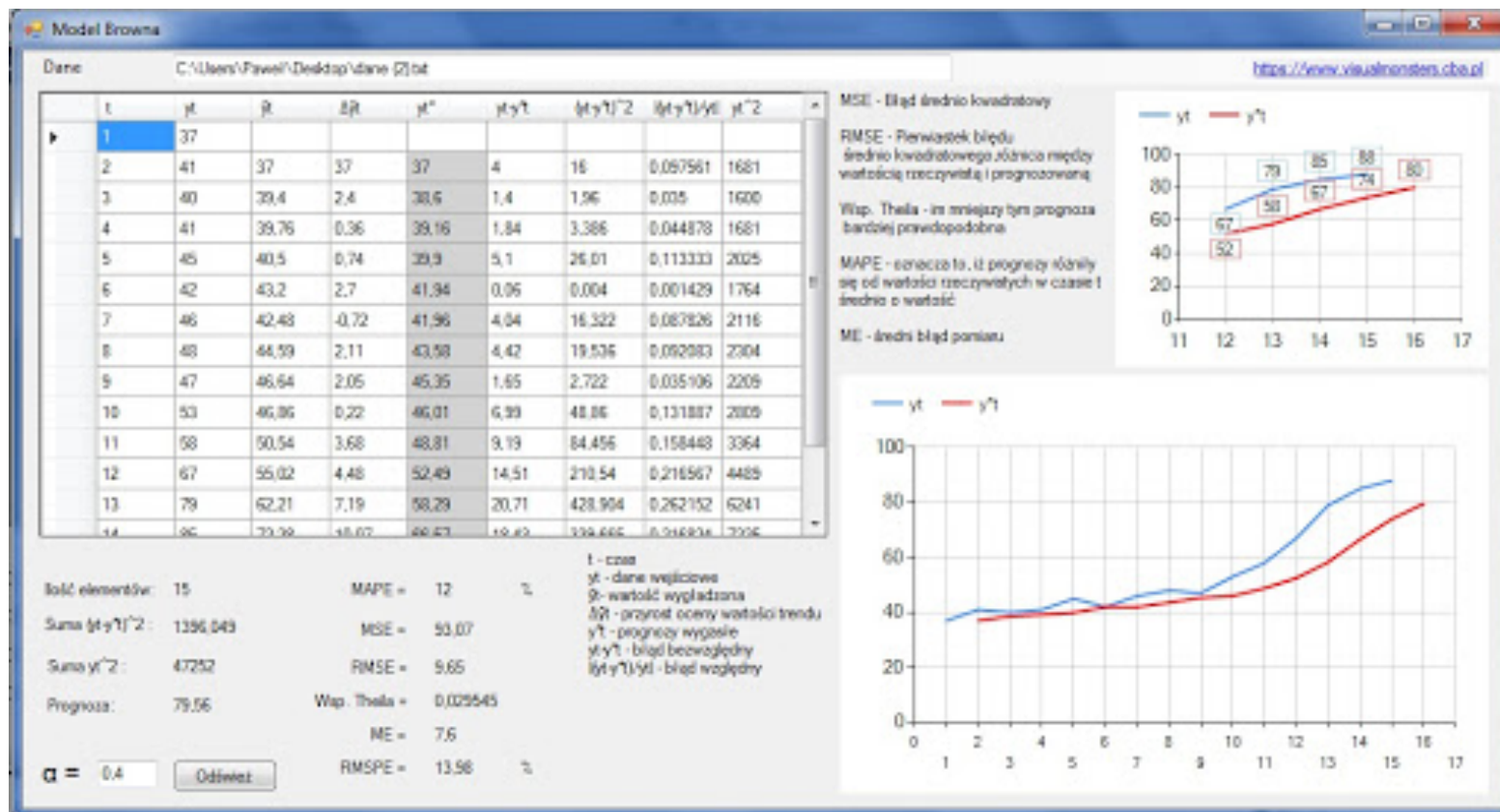
Tabela 5.3
Wyniki metody wyrównywania wykładniczego dla danych z przykładu 5.1

t	Y_t	F_t $\alpha = 0,2$	$Y_t - F_t$ $\alpha = 0,2$	$(Y_t - F_t)^2$ $\alpha = 0,2$	F_t $\alpha = 0,8$	$Y_t - F_t$ $\alpha = 0,8$	$(Y_t - F_t)^2$ $\alpha = 0,8$
1	2	3	4	5	6	7	8
1	15	15,0	0,0	0	15,0	0	0
2	17	15,0	2,0	4,0	15,0	2,0	4,0
3	12	15,4	-3,4	11,56	16,6	-4,6	21,16
4	16	14,7	1,3	1,69	12,9	3,1	9,61
5	15	15,0	0,0	0	15,4	-0,4	0,16
6	11	15,0	-4,0	16,0	15,1	-4,1	16,81
7	18	14,2	3,8	14,44	11,8	6,2	38,44
8	17	15,0	2,0	4,0	16,0	0,2	0,04
9	13	15,4	-2,4	5,76	17,0	-4,0	16,0
10	16	15,0	1,0	1,0	13,8	2,2	4,84
11		15,2		58,45	15,6		111,04

Źródło: Obliczenia własne na podstawie danych z przykładu 5.1.

Który parametr wygładzania α wybrać?

Przykład użycia prostego wygładzania Browna



Rodzaje wygładzania wykładniczego

- Prosty model wygładzania wykładniczego Browna - stosujemy w przypadku występowania prawie stałego poziomu zmiennej prognozowanej oraz niewielkich wahań przypadkowych.
- Podwójne wyrównywanie (np. Holta)- nadaje się do prognozowania, jeżeli w szeregu czasowym obok wahań przypadkowych występuje wyraźna tendencja wzrostowa lub spadkowa;
- Potrójne wyrównanie wykładnicze (np. Wintersa)- jest stosowany, gdy w szeregu czasowym występuje trend, sezonowość i wahania przypadkowe.

Szeregi z wyraźniejszą tendencją

- Nadal model wykładniczy → lecz dla szeregu z wyraźną tendencją rozwojową i wahaniami przypadkowymi (bez sezonowych) pomyśl o **modelu Holta**
- W ogólności wartość prognozowana w chwili $t+1$ składa się z dwóch składników F przybliżającego poziom zmiennej oraz T oceniającego jej przyrost (trend):

$$y_{t+1}^* = F_t + T_t$$

$$F_t = \alpha \cdot y_t + (1 - \alpha) \cdot (F_{t-1} + T_{t-1})$$

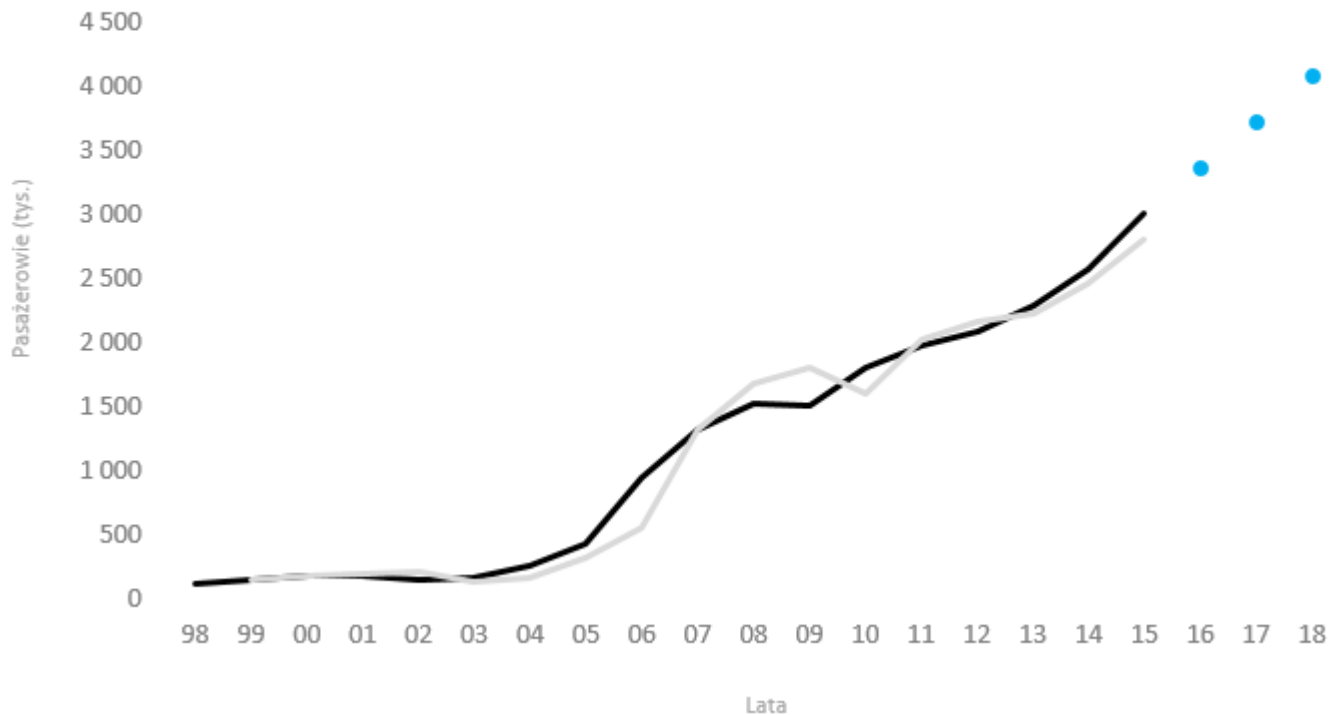
$$T_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot T_{t-1}$$

- jako wartości początkowe można przyjąć $F_1 = y_1$ oraz $T_1 = y_2 - y_1$
- F odpowiada wyrazowi stałemu, a T_1 nachyleniu w funkcji liniowej trendu, gdyby ją wyznaczać

Przykład użycia wygładzania Holta

Zagraniczny regularny pasażerski ruch lotniczy na lotnisku im. Lecha Wałęsy

Wartości rzeczywiste, Wartości wygładzone, [Prognoza](#)



Dostępność w Statistica

Sezonowe i niesezone wyrównywanie wykładnicze [?] [X]

Zabezpiecz Zmienna Długa nazwa zmiennej (szeregu)

L	SZEREG_G	Miesięczna liczba pasażerów (w tysiącach)

[Zakończ]

Liczba zapamiętanych transform. dla 1 zmiennej: 5

Model

SŁADNIK SEZONOWY: opóźnienie= 12

Brak: Addytywny: Multiplikatywny:

Bez trendu: ☐ pojedyn. ☐ ☐ ☐

Trend liniowy: ☐ Holta ☐ ☐ ☒ Wintersa

Wykładniczy: ☐ ☐ ☐ ☐

Trend gasnący: ☐ ☐ ☐ ☐

Alfa: .100 Delta: .100 Gamma: .100 F_i: .100

☐ Wart. początkowa: 0. ☐ Trend początkowy: 0.

☐ Weź wskaźniki sezonowości ze zmiennej: brak

☒ Wykonaj sumaryczny wykres dla każdego wyrównywania

☒ Dodaj prognozy i błędy do obszaru rob. Prognozuj 10 obs.

[?] Sieciowe poszukiwanie najlepszych parametrów [1]

[?] Automatyczne poszukiwanie najlepszych parametrów [3]

[Zapisz zmienną] [Usuń]

Przeglądaj i kreśl zmienne

☐ Wyświetl/kreśl podzbiór obserwacji [?] [Opcje]

[Przeglądaj podświetloną zmienną] [Kreśl]

[Przeglądaj wiele zmiennych] [Kreśl]

[Kreśl 2 listy zmiennych z różnymi skalami]

Autokorelacje

[Autokorelacje] Alfa (podświetl.): .050

☒ Błędy standardowe białego szumu

[Autokor. cząstkowe] N opóźnień: 15

[Histogram] [Statystyki opisowe]

[Wykres normalny] [Bez trendu] [Półnormal.]

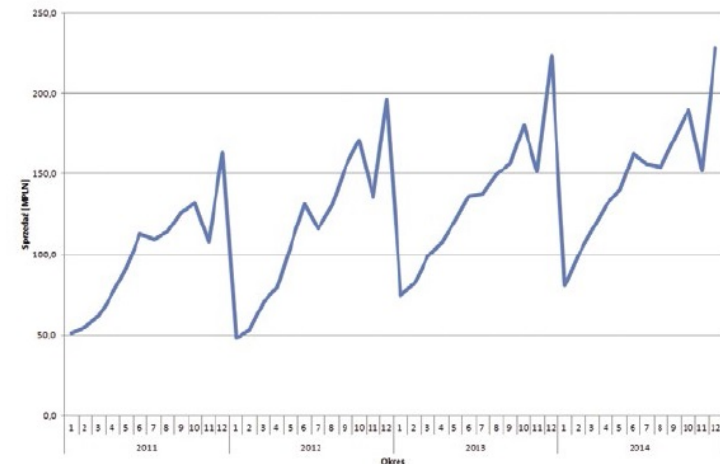
[?] [Inne przekształcenia i wykresy]

Czynnik okresowości

- W przebiegu szeregu czasowego dostrzega się pewne wahania, powtarzające się w tych samych mniej więcej rozmiarach (bezwzględnych lub względnych), co jakiś w przybliżeniu stały odstęp czasu.
- **Cel:** wyodrębnienie składowej okresowej (sezonowej, cyklicznej).
- Różne metody:
 - okres zmian znany – klasyczna dekompozycja szeregu czasowego (np. metoda wskaźników sezonowych, analiza harmoniczna, modele wygładzania wykładniczego np. Wintersa),
 - okres zmian nieznany – modele z opóźnionymi wartościami zmiennej prognozowanej (np. modele ARIMA, analiza autokorelacji)

Rys. 2.

Wykres sprzedaży produktu X rok po roku



Prognoza dla szeregu czasowego metodą Wintersa

- Uogólnienie modelu Holta dla szeregu z wahaniami sezonowymi.
- Oprócz poziomu zmiennej (F), współczynnika trendu (T) wygładzaniu podlega składki sezonowy (S)

$$y_{t+1}^* = F_t + T_t + S_{t+1-r}$$

$$F_t = \alpha \cdot [y_t - S_{t-r}] + (1 - \alpha) \cdot (F_{t-1} + T_{t-1})$$

$$T_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot T_{t-1}$$

$$S_t = \gamma \cdot (y_t - F_t) + (1 - \gamma) \cdot S_{t-r}$$

- r – długość cyklu sezonowości; wartości początkowe $S_1..S_r$ wyznaczamy odejmując od wartości y_i średnią r pierwszych obserwacji

- Trudności z doбором α, β, γ

Prognozowanie w modelu Wintersa

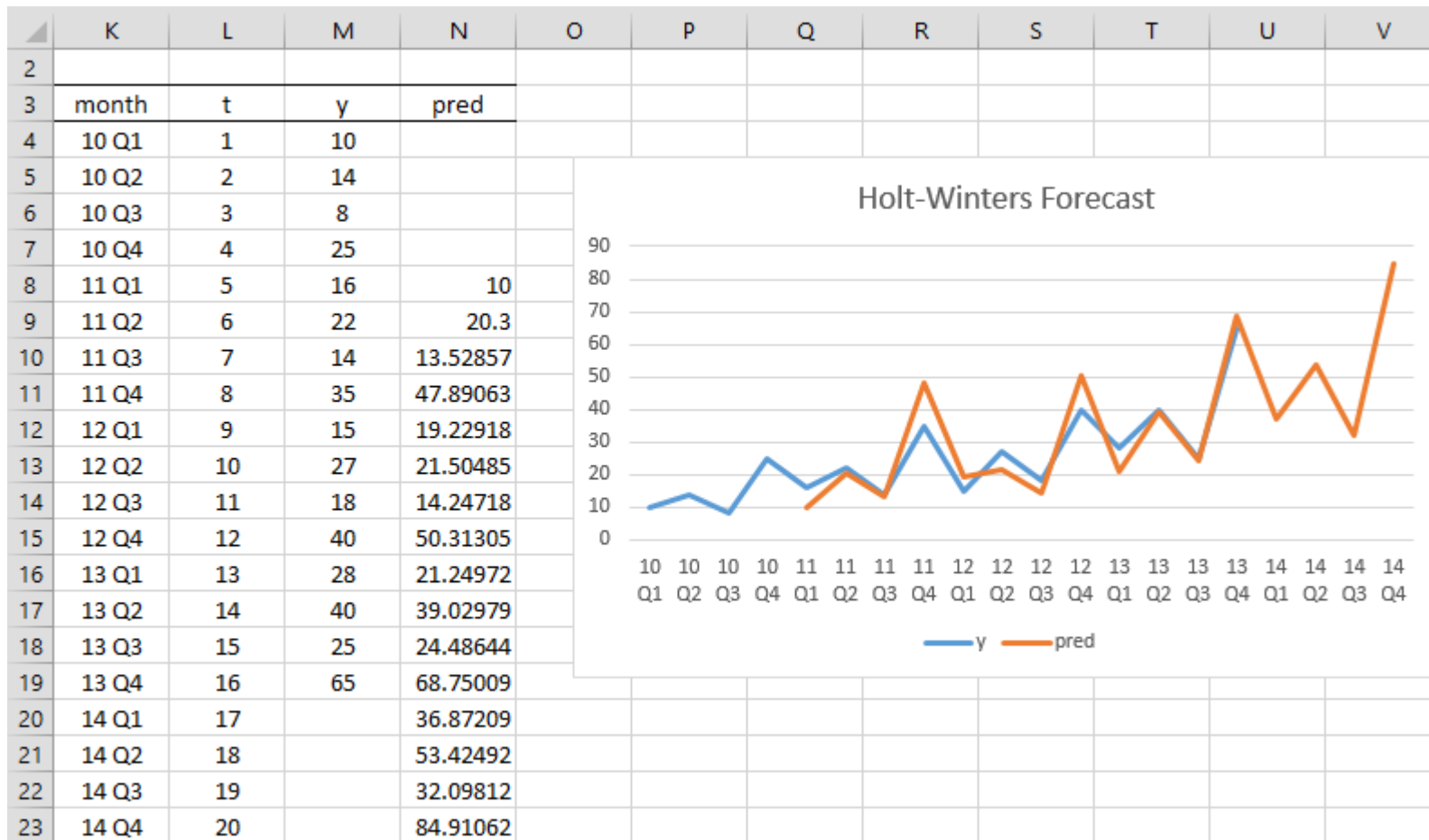
- Równania na moment $t > n$
- Wersja addytywna

$$y_t^* = F_n + T_n(t - n) + S_{t-r}$$

- Wersja multiplikatywna

$$y_t^* = (F_n + T_n(t - n)) \cdot S_{t-r}$$

Przykład użycia p. Wintersa



Estymacja trendu

- Obserwacja kształtu graficznego wykresu oraz próba doboru funkcji (lub składanego zbioru funkcji)
- Można wykorzystać modelowanie funkcji – regresja
- Przykład

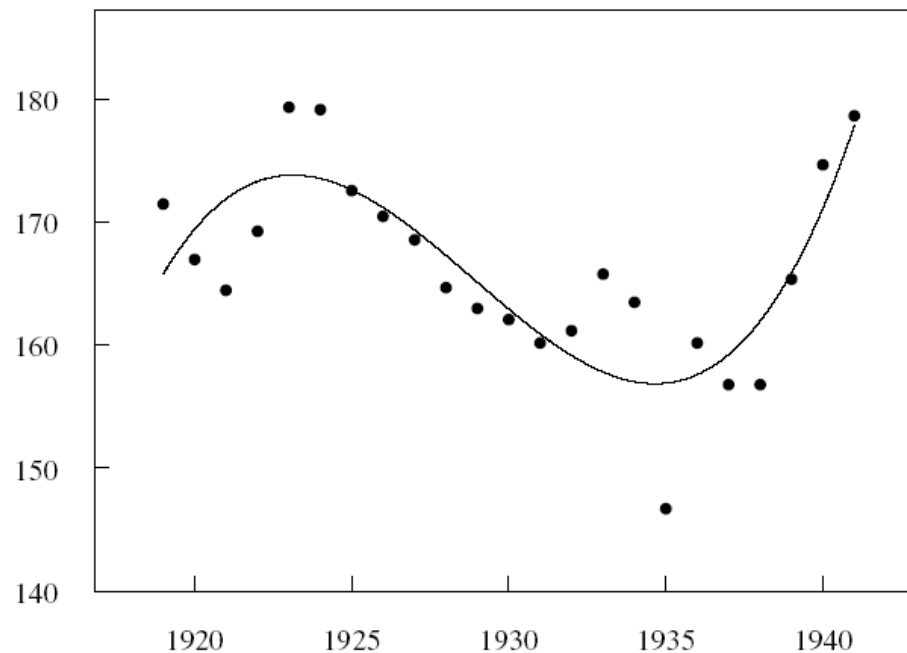


Figure 1. A cubic function fitted to data on meat consumption in the United States, 1919–1941.

Analiza sprzedaży obuwia produkowanego przez pewne przedsiębiorstwo

Tablica 5.3
Sprzedaż obuwia dziecięcego przedsiębiorstwa A
w latach 1991–1998

Lata	Sprzedaż w mln zł
1991	3
1992	4
1993	6
1994	10
1995	11
1996	12
1997	14
1998	18

Źródło: Dane umiowne.

Zastosujemy model liniowy

$$y_t^* = b_0 + b_1 \cdot t$$

Obliczmy b_0 i b_1 na podstawie wzorów wynikających z metody MNK

$$y_t^* = 1,51 + 1,83 \cdot t$$

Przykład prognozy na 1999 rok

$$y_9 = 1,51 + 1,83 \cdot 9 = 17,98 \text{ mln zł.}$$

Dostępność w Statistica

Przekształcenia szeregów czasowych [?] [X]

Przekształć zmienną: **SZEREG_G: Miesięczna liczba pasażerów (w tysiącach)** **OK (Przekształć)**

Przekształcenia: $x=f(x)$

☒ Dodaj stałą ($x=x+C$) C= [↑] [↓]

☐ Potęga ($x=x^C$) C= [↑] [↓]

☐ Pierwiastek ($x=x^{1/C}$) C= [↑] [↓]

☐ Logarytm naturalny ($x=\ln(x)$)

☐ Przekształcenie wykładnicze ($x=\exp(x)$)

☐ Odjęcie średniej ($x=x-M$) M= [↑] [↓]

☐ Standaryzacja ($x=(x-M)/S$) M= [↑] [↓] S= [↑] [↓]

☒ Szacuj średnią i odchylenie standardowe z danych

☐ Odjęcie trendu ($x=x-(a+b*t)$) a= [↑] [↓] b= [↑] [↓]

☐ Autokor. ($x=x-(a+b*x(\text{opóźn}))$) a= [↑] [↓] b= [↑] [↓]

☒ Szacuj a/b z danych opóźn= [↑] [↓]

Wygładzanie

☐ Średnia ruch. N-punkt. N= [↑] [↓] ☐ Poprz. ☐ Ważona

☐ Mediana ruch. N-punkt. N= [↑] [↓] ☐ Poprzedzające

☐ Proste wyrównywanie wykładnicze alfa= [↑] [↓]

Przekształcenia do analizy widmowej

Wybierz przekształcenie dla zmiennej pierwotnej. Z panelu początkowego możesz wywołać wygładzanie wykładnicze.

Przekształcenia dwóch szeregów

☐ Różnica ($x=x-y(\text{opóź.})$) opóź.= [↑] [↓]

☐ Residualizacja ($x=x-(a+b*y(\text{opóźn}))$) a= [↑] [↓] b= [↑] [↓]

☒ Szacuj a/b z danych

Przesunięcie

☐ Przesuń szereg o opóźnienie

☐ Przesuń szereg o opóźnienie

Filtrowanie i Sumowanie

☐ Filtr 4253H

☐ Różnicowa

☐ Sumowanie

Transformations of Variables: Series_G.sta [?] [-] [X]

OK (Transform selected series)

Lock	Variable	Long variable (series) name
L	SERIES_G	Monthly passenger totals (in 1000's)

Cancel Options ▾

Number of backups per variable (series): [↑] [↓] Save variables Delete

☒ $x=f(x)$ ☐ Smoothing ☒ $x=f(x,y)$ ☐ Shift ☐ Difference, integrate ☐ Fourier ☐ Review & plot ☐ Autocorr ☐ Descriptives

OK (Transform selected series)

Transformation

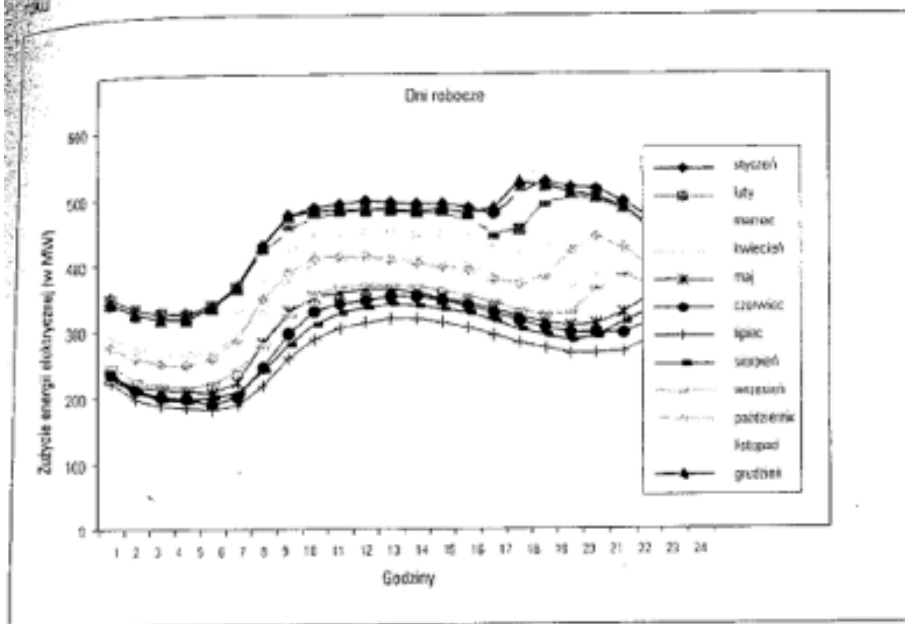
☒ N-pts mov. averg. N= [↑] [↓] ☐ Prior ☐ Weighted Specify weights

☐ N-pts mov. median N= [↑] [↓] ☐ Prior

☐ Simple exponential alfa= [↑] [↓]

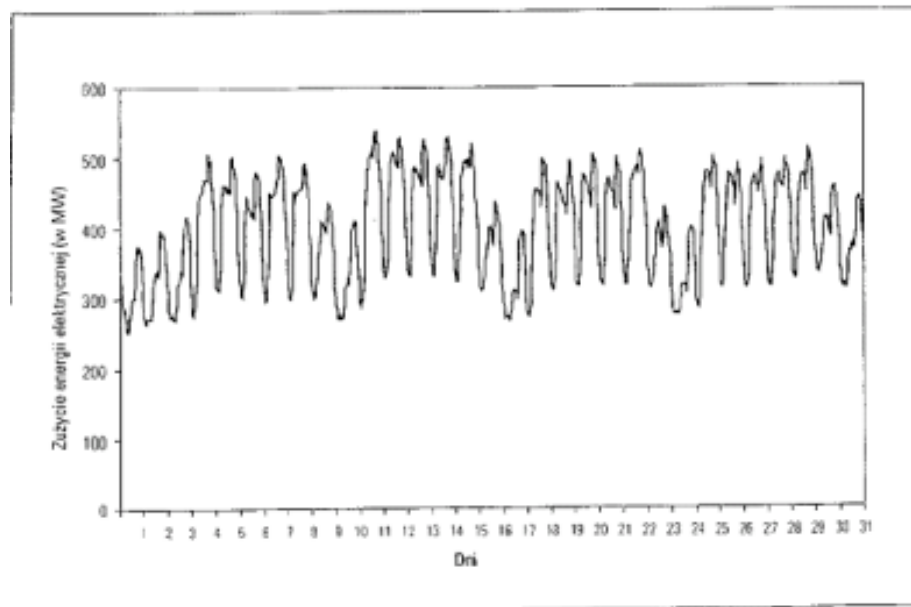
☐ 4253H Filter

Ilustracja 6.2. Zapotrzebowanie na energię elektryczną w typowym dniu roboczym w kolejnych miesiącach



Okresowe zmiany
-
Dekompozycja szeregu

Ilustracja 6.3. Miesięczne zapotrzebowanie na energię elektryczną w styczniu



Za Witkowska

Zmiany okresowe

Składowa okresowa może wystąpić w postaci wahań:

- **cyklicznych** - długookresowe, rytmiczne wahania (cykl koniunkturalny gospodarki, cykl rozwoju populacji nabywców danego produktu, itp.),
- **sezonowych** – krótkookresowe do 1 roku, odzwierciedlają wpływ zachowań wynikający z „kalendarza” (np. rytm pracy w skali tygodnia, dnia, pory roku, świąt, ...).

W strone wskaźników sezonowości (metoda Census)

W ogólności przyjmuje się addytywną lub multiplikatywną formę modelu.

- **Model addytywny:**

$$y_t = f(t) + g(t) + h(t) + \zeta_t$$

- **Model multiplikatywny:**

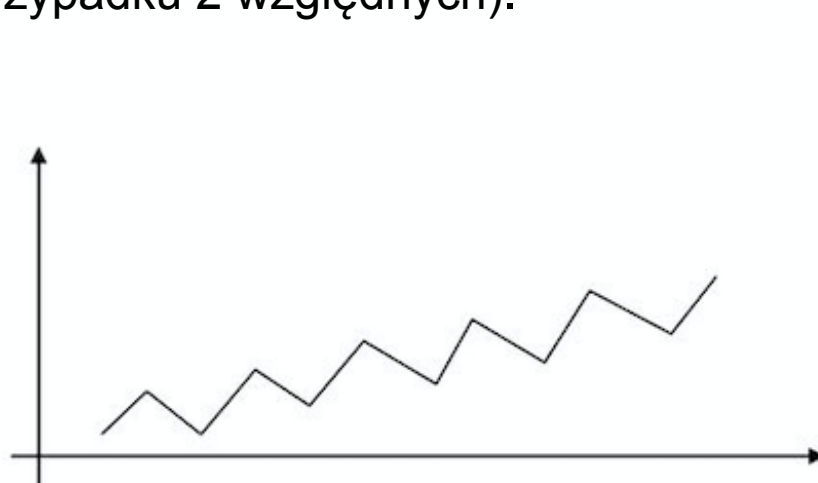
$$y_t = f(t) \cdot g(t) \cdot h(t) \cdot \zeta_t$$

- gdzie $f(t)$ – funkcja trendu,
 $g(t)$ – funkcja czasu charakteryzująca wahania sezonowe,
 $h(t)$ – funkcja czasu charakteryzująca wahania cykliczne,
 ζ_t - składnik losowy.

- Stosuje się także modele mieszane.

Idea wskaźników

- Polega na wyznaczeniu **wskaźników sezonowości** poszczególnych faz cyklu.
- **Amplitudy wahań** (wynikające z porównania wartości rzeczywistych zmiennej z wartościami teoretycznymi uzyskanymi z modelu) dzieli się na:
 - Bezwzględnie stałe (w analogicznych fazach cyklu).
 - Względnie stałe (wielkości amplitud zmieniają się w przybliżeniu w tym samym stosunku).
- Okresowość może mieć względem trendu charakter **addytywny** (w przypadku 1 bezwzględnych stałych wahań) lub **multiplikatywny** (w przypadku 2 względnych).



Addytywny



Multiplikatywny

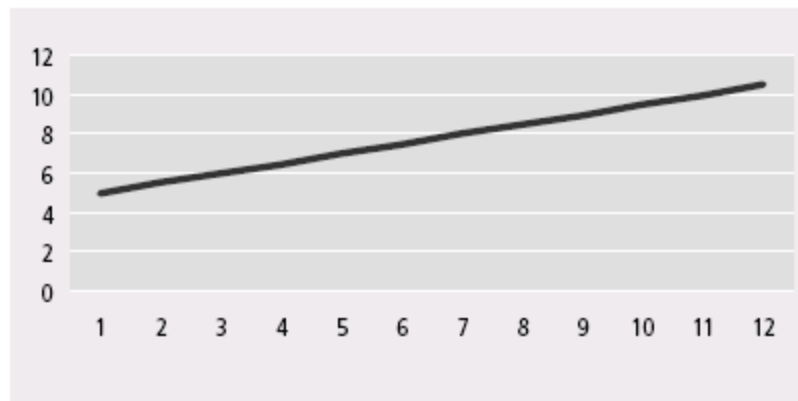
Wskaźniki sezonowości

Klasyczny model addytywny T+C+S

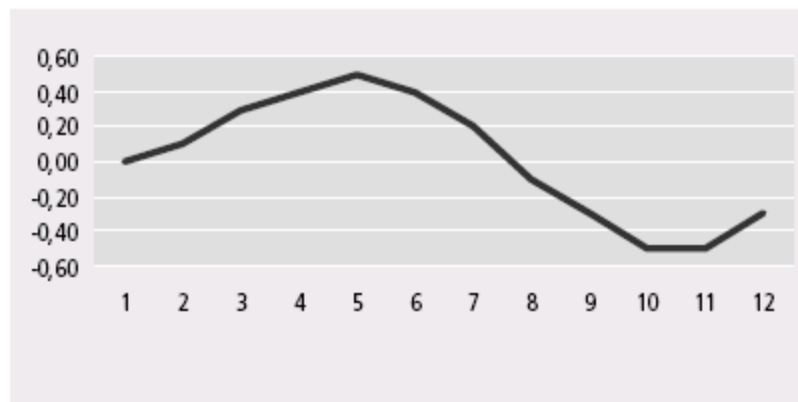
Component series in an additive model

Year	Quarter	Trend T	Cycle C	Season S
1	1	5,0	0,0	0,0
	2	5,5	0,1	0,4
	3	6,0	0,3	0,2
	4	6,5	0,4	-0,6
2	1	7,0	0,5	0,0
	2	7,5	0,4	0,4
	3	8,0	0,2	0,2
	4	8,5	-0,1	-0,6
3	1	9,0	-0,3	0,0
	2	9,5	-0,5	0,4
	3	10,0	-0,5	0,2
	4	10,5	-0,3	-0,6

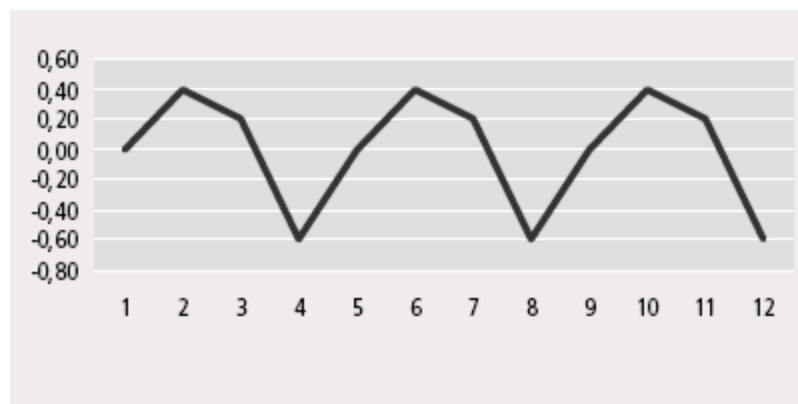
Trend



Cycle



Seasonal



Wskaźniki sezonowości

- Wahania okresowe tworzą cykl sezonowy – długość cyklu / okres wahań.
- Okres / sezon składa się z faz (kształtowanie się przebiegu, np. szybki wzrost, lekki wzrost, spadek, itd.)
- Liczba faz w cyklu, np. 12 faz dotyczących danych miesięcznych, 4 faz dotyczących danych kwartalnych -> długość okresu (sezonu).

Dekompozycja - sposób postępowania:

- wyodrębnienie trendu,
- eliminacja trendu z szeregu czasowego,
- eliminacja wahań przypadkowych,
- obliczenie wskaźników sezonowości

Model addytywny: $y_{ti} = \hat{y}_{ti} + s_i + \zeta_t$

Model multiplikatywny: $y_{ti} = \hat{y}_{ti} \cdot s_i \cdot \zeta_t$

gdzie: $t=1,2,\dots,n; i=1,2,\dots,r$,

y_{ti} — rzeczywista wartość zmiennej w okresie t w i -tej fazie cyklu,

$\hat{y}_{ti}$ — teoretyczna wartość zmiennej w okresie t w i -tej fazie cyklu,

s_i — wskaźnik sezonowości dla i -tej fazy cyklu,

ζ_t — składnik losowy,

r — liczba faz cyklu.

■

Eliminacja trendu: $z_{ti} = y_{ti} - \hat{y}_{ti}$ (addytywny) lub $z_{ti} = \frac{y_{ti}}{\hat{y}_{ti}}$

Eliminacja oddziaływania składnika losowego →
obliczenie surowych wskaźników sezonowości.

$$z_i = \frac{1}{k} \sum_{j=0}^{k-1} z_{i+j \cdot r, i}$$

gdzie k – liczba jednoimiennych faz w badanym szeregu
czasowym, i – wybrana faza, r – liczba faz w okresie.

Wskaźniki podlegają korekcji:

$$s_i = z_i - q \quad (\text{dla modelu addytywnego})$$

$$\text{lub } s_i = \frac{z_i}{q} \quad (\text{dla modelu multiplikatywnego})$$

$$\text{gdzie } q = \frac{1}{r} \sum_{i=1}^r z_i$$

Suma wskaźników sezonowości powinna być równa
zeru (model addytywny) lub liczbie faz tworzących
okres (model multiplikatywny).

Przykład wyznaczania wskaźników sezonowości

Dzienna sprzedaż mleka (w litrach)

Tabela 4.1

Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	63	47	43	45	30
Drugi	65	48	44	46	31
Tzeci	68	50	48	50	33

Źródło: Raporty sklepowe.

Analiza dziennej sprzedaży mleka (Dittmann, 1998)

Zauważ, wahania sezonowe - okres tygodnia

Zastosujemy model multiplikatywny i dwie metody

Metoda I - wygładzanie średnią ruchomą
(niecentrowana, 5-elementowa)

Tabela 4.2

Wygładzanie szeregu niecentrowaną średnią ruchomą

Tydzień	Dzień tygodnia	Obserwacja	Niecentrowana średnia ruchoma	Wygładzenie
Pierwszy	poniedziałek	63	-	-
	wtorek	47	-	-
	środa	43	45,6	0,943
	czwartek	45	46,0	0,978
	piątek	30	46,2	0,649
Drugi	poniedziałek	65	46,4	1,401
	wtorek	48	46,6	1,030
	środa	44	46,8	0,940
	czwartek	46	47,4	0,970
	piątek	31	47,8	0,649
Trzeci	poniedziałek	68	48,6	1,399
	wtorek	50	49,4	1,012
	środa	48	49,8	0,964
	czwartek	50	-	-
	piątek	33	-	-

Źródło: Obliczenia własne.

Tabela 4.3

Sezowe wskaźniki sezonowości

Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	-	-	0,943	0,978	0,649
Drugi	1,401	1,030	0,940	0,970	0,649
Tzeci	1,399	1,012	0,964	-	-
Średnie	1,400	1,021	0,949	0,974	0,649

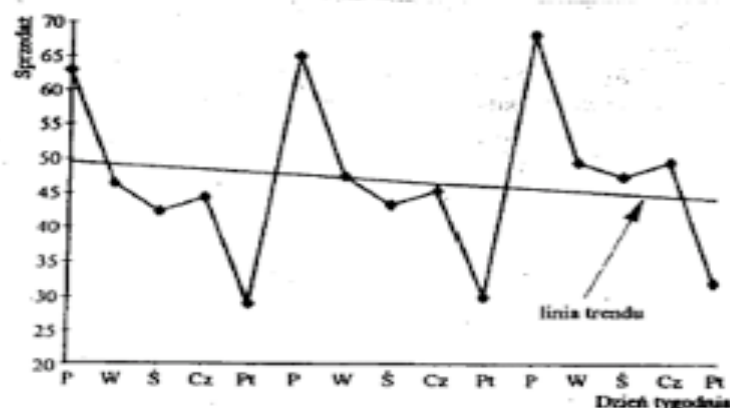
Źródło: Obliczenia własne.

$$q - \text{współczynnik korygujący} = \frac{1,4 + 1,021 + 0,949 + 0,974 + 0,649}{5} =$$

wskaźniki sezonowości:

$$= 0,9986$$

Dzień	Poniedziałek	Wtorek	Środa	Czwartek	Piątek
Wskaźnik	1,402	1,023	0,950	0,975	0,650



Rys. 4.4. Dzienna sprzedaż mleka (w litrach)

II Metoda analityczna

MNK-funkcja trendu
 $\hat{y}_t = 50,3 - 0,368 \cdot t$

Tabela 4.4

Wygładzanie szeregu funkcją trendu

Tydzień	Dzień tygodnia	Obserwacje	Wartości trendu	Wygładzenie
Pierwszy	poniedziałek	63	49,9	1,263
	wtorek	47	49,6	0,948
	środa	43	49,2	0,874
	czwartek	45	48,8	0,922
	piątek	30	48,5	0,619
Drugi	poniedziałek	65	48,1	1,351
	wtorek	48	47,7	1,006
	środa	44	47,4	0,928
	czwartek	46	47,0	0,979
	piątek	31	46,6	0,665
Trzeci	poniedziałek	68	46,3	1,469
	wtorek	50	45,9	1,089
	środa	48	45,5	1,055
	czwartek	50	45,1	1,109
	piątek	33	44,8	0,737

Źródło: Obliczenia własne.

Tabela 4.5

Serowe wskaźniki sezonowości

Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	1,263	0,948	0,874	0,922	0,619
Drugi	1,351	1,006	0,928	0,979	0,655
Trzeci	1,469	1,089	1,055	1,109	0,737
Srednie	1,361	1,014	0,952	1,003	0,674

Źródło: Obliczenia własne.

Wskaźniki sezonowości:

$q = 0,9992$

Dzień	Poniedziałek	Wtorek	Środa	Czwartek	Piątek
Wskaźnik	1,360	1,013	0,952	1,002	0,673

Wykorzystanie średniej w dekompozycji

Moving Averages in the additive model

Year	Quarter	Time Series X	Moving Average G
1	1	5,0	...
	2	6,0	...
	3	6,5	6,26
	4	6,3	6,86
2	1	7,5	7,39
	2	8,3	7,81
	3	8,4	8,15
	4	7,8	8,44
3	1	8,7	8,74
	2	9,4	9,13
	3	9,7	...
	4	9,6	...

Estimates of the seasonal component in the additive model

Year	Quarter	Difference D=X-G
1	1	...
	2	...
	3	0,24
	4	-0,56
2	1	0,11
	2	0,49
	3	0,25
	4	-0,64
3	1	-0,04
	2	0,27
	3	...
	4	...

First estimate of the seasonal factors

Quarter	D=X-G			Average
	Year 1	Year 2	Year 3	S'
1	...	0,11	-0,04	0,04
2	...	0,49	0,27	0,38
3	0,24	0,25	...	0,25
4	-0,56	-0,64	...	-0,60
Total				0,07

Final estimate of the seasonal factors

Quarter	D=X-G			Average
	Year 1	Year 2	Year 3	S
1	...	0,11	-0,04	0,02
2	...	0,49	0,27	0,36
3	0,24	0,25	...	0,23
4	-0,56	-0,64	...	-0,62
Total				-0,01

Seasonally adjusted figures in the additive model

		Observation	Seasonal factor	Seasonally adjusted series
Year	Quarter	X	S	X-S
1	1	5,0	0,0	5,0
	2	6,0	0,4	5,6
	3	6,5	0,2	6,3
	4	6,3	-0,6	6,9
2	1	7,5	0,0	7,5
	2	8,3	0,4	7,9
	3	8,4	0,2	8,2
	4	7,8	-0,6	8,4
3	1	8,7	0,0	8,7
	2	9,4	0,4	9,0
	3	9,7	0,2	9,5
	4	9,6	-0,6	10,2

D.Witkowska (str. 154) dane nt. kwartalnych nakładów inwestycyjnych w latach 1992-1997

Trend liniowy $y=75,28t+1880,95$

Sezonowość: kwartalna ($r = 4$); 6 kolejnych sezonów

Surowe wskaźniki:

$Z_1=0,661$; $Z_2=0,9$; $Z_3=0,939$;

$Z_4=1,468$;

Skorygowane wskaźniki sezonowości:

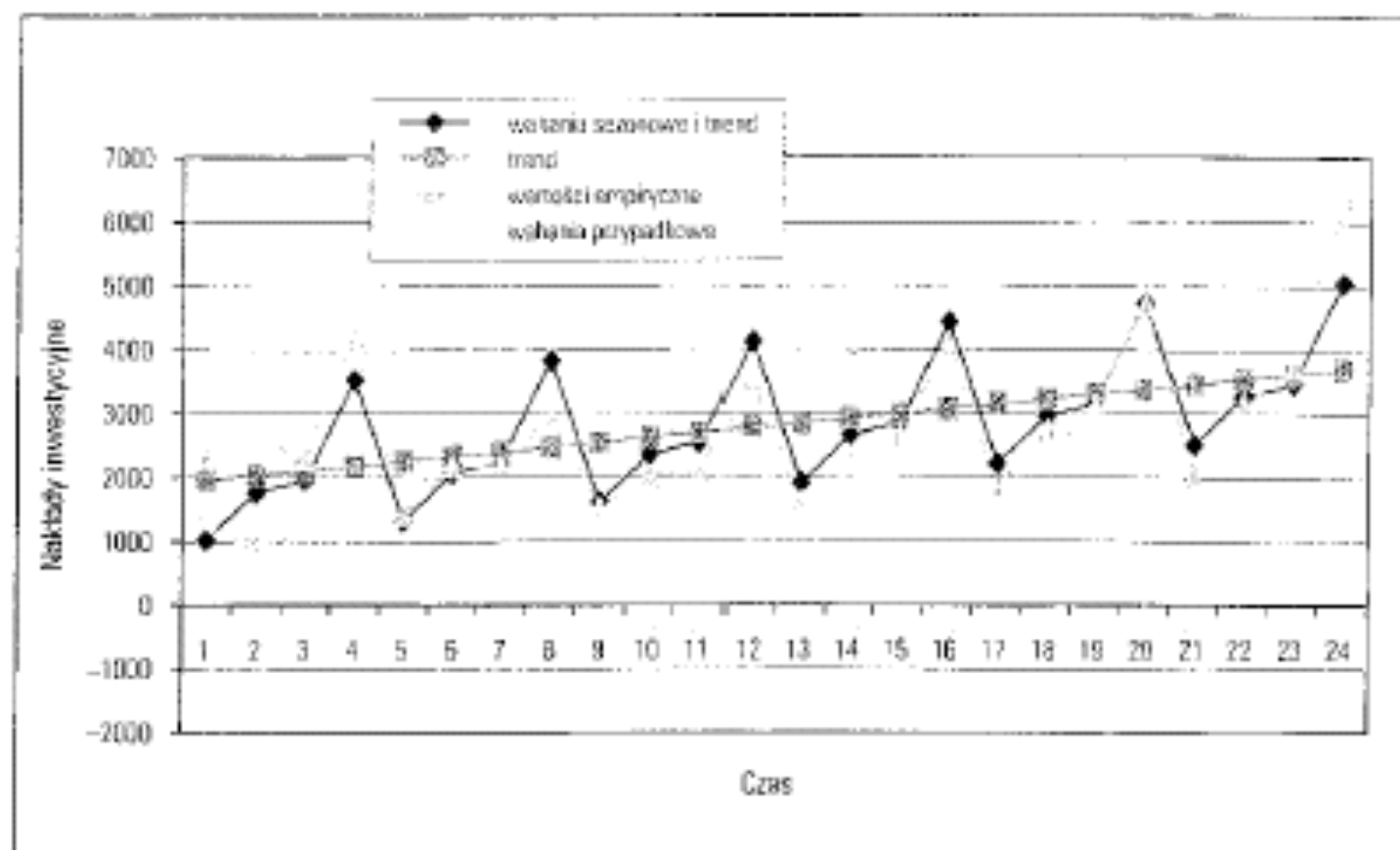
$S_1=0,667$; $S_2=0,907$; $S_3=0,946$;

$S_4=1,48$;

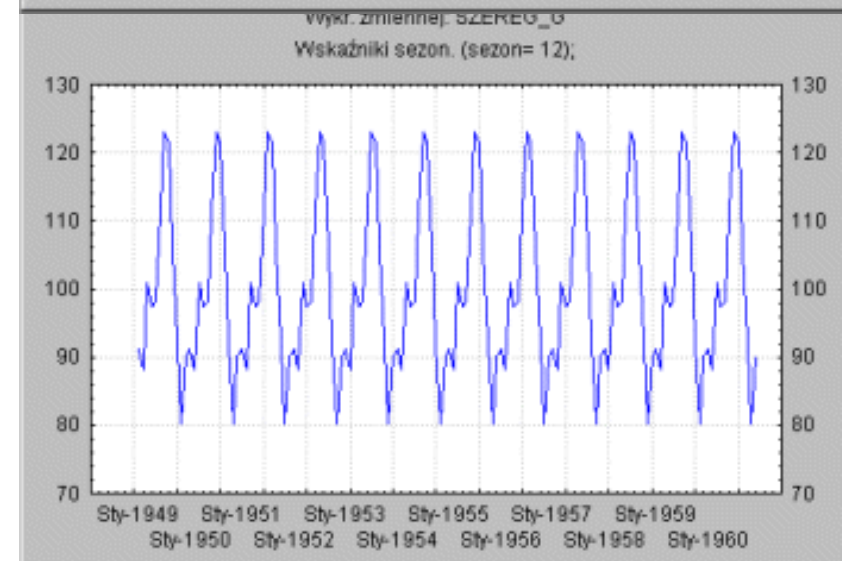
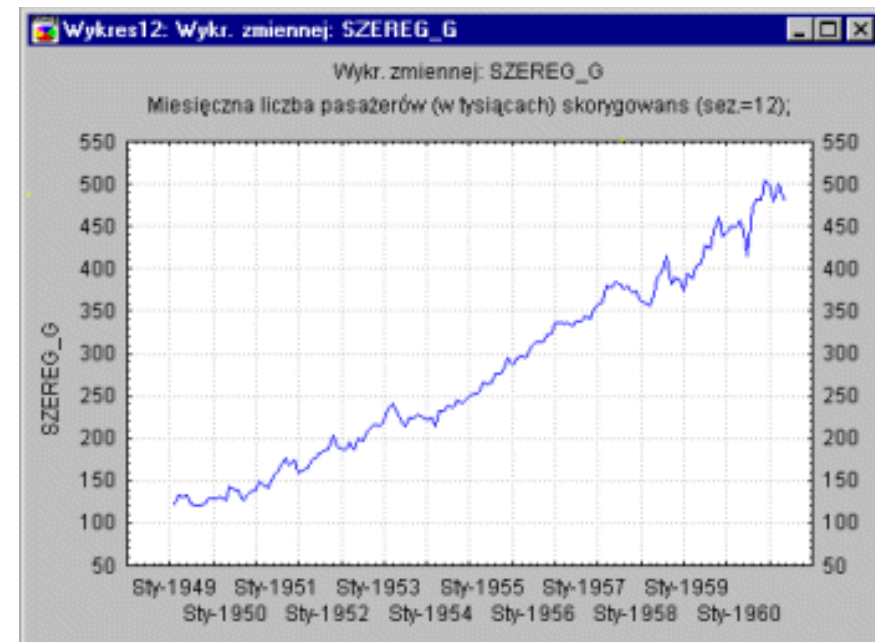
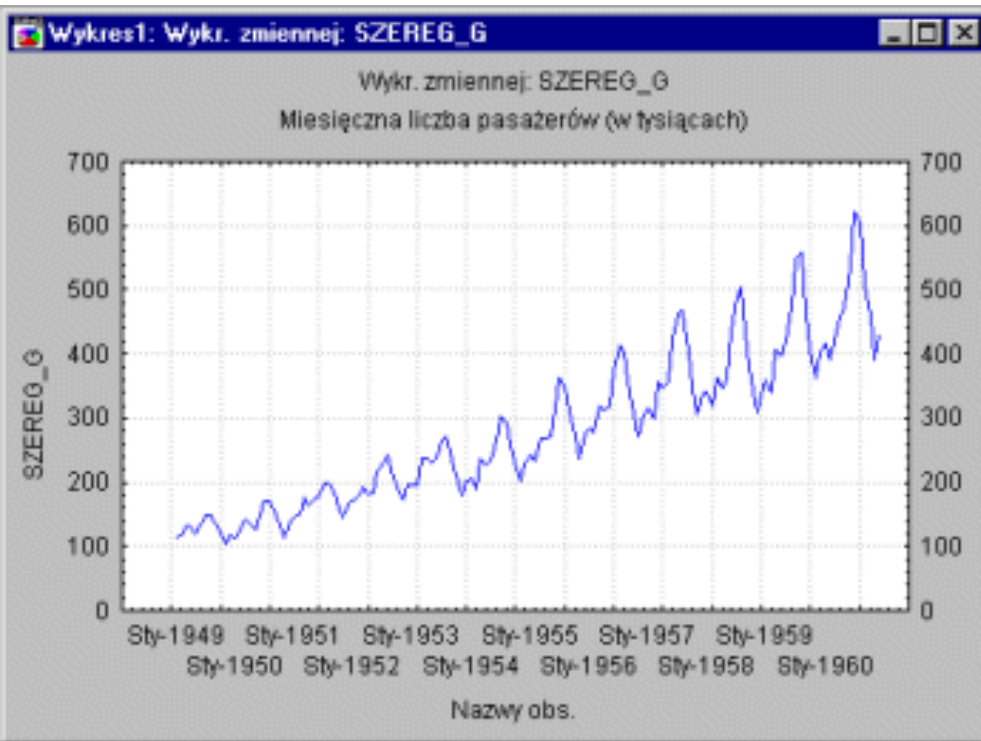
Tabela 6.2. Obliczenia do przykładu

Okresy	t	y_t	$\hat{y}_t$	S_t
1	2	3	4	5
1992, I kwartał	1	2270,50	1956,23	0,67
1992, II kwartał	2	2656,83	2031,50	0,91
1992, III kwartał	3	2285,21	2106,78	0,95
1992, IV kwartał	4	4206,14	2182,05	1,48
1993, I kwartał	5	1424,96	2257,33	0,67
1993, II kwartał	6	2165,40	2332,60	0,91
1993, III kwartał	7	2193,31	2407,88	0,95
1993, IV kwartał	8	2976,59	2483,15	1,48
1994, I kwartał	9	1545,18	2558,43	0,67
1994, II kwartał	10	2032,53	2633,71	0,91
1994, III kwartał	11	2117,94	2708,98	0,95
1994, IV kwartał	12	3486,68	2784,26	1,48
1995, I kwartał	13	1618,58	2859,53	0,67
1995, II kwartał	14	2248,79	2934,81	0,91
1995, III kwartał	15	2672,11	3010,08	0,95
1995, IV kwartał	16	4126,51	3085,36	1,48
1996, I kwartał	17	1847,15	3160,63	0,67
1996, II kwartał	18	2754,43	3235,91	0,91
1996, III kwartał	19	3125,50	3311,19	0,95
1996, IV kwartał	20	4719,98	3386,46	1,48
1997, I kwartał	21	2040,60	3461,74	0,67
1997, II kwartał	22	3174,23	3537,01	0,91
1997, III kwartał	23	3710,47	3612,29	0,95
1997, IV kwartał	24	6323,84	3687,56	1,48
Suma	×	67725,46	×	×

Ilustracja 6.7. Dekompozycja szeregu czasowego za pomocą metody analitycznej



Przykład analizy przewozów linii lotniczej PANAM

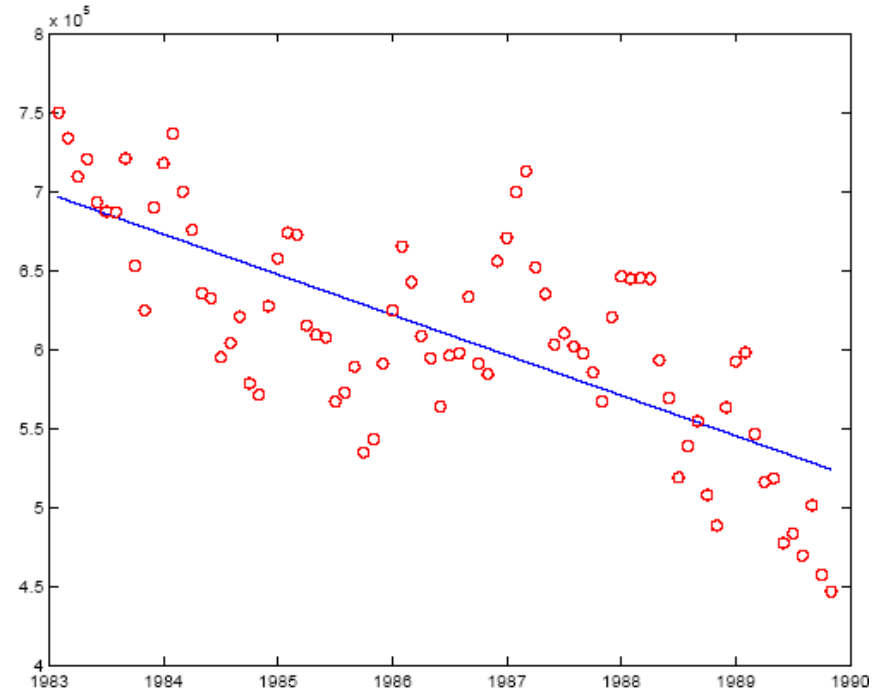
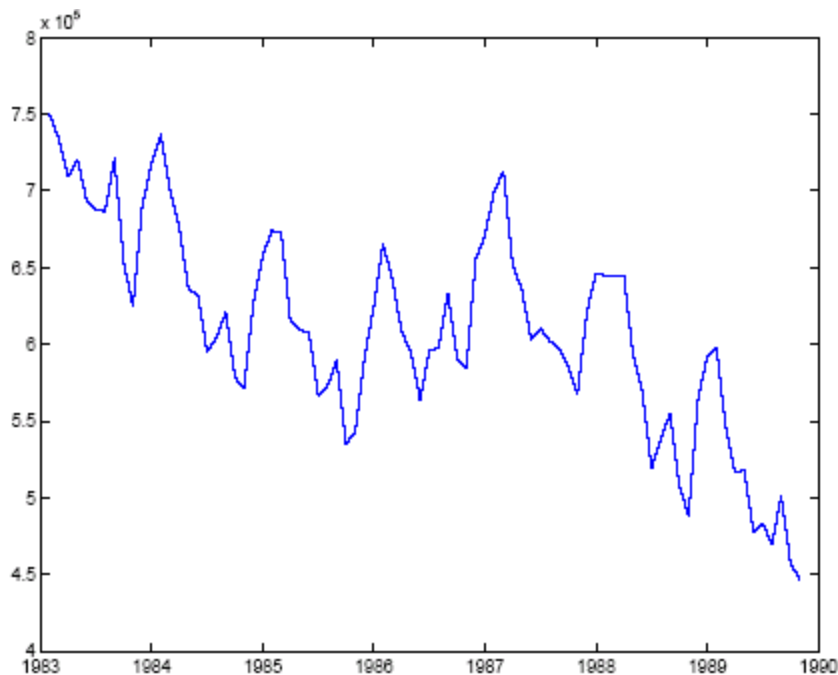


Spójrz szczegóły dodatkowe,
dawne slajdy

Dekompozycja: trend + wahania

dane o miesięcznym bezrobociu w Australii

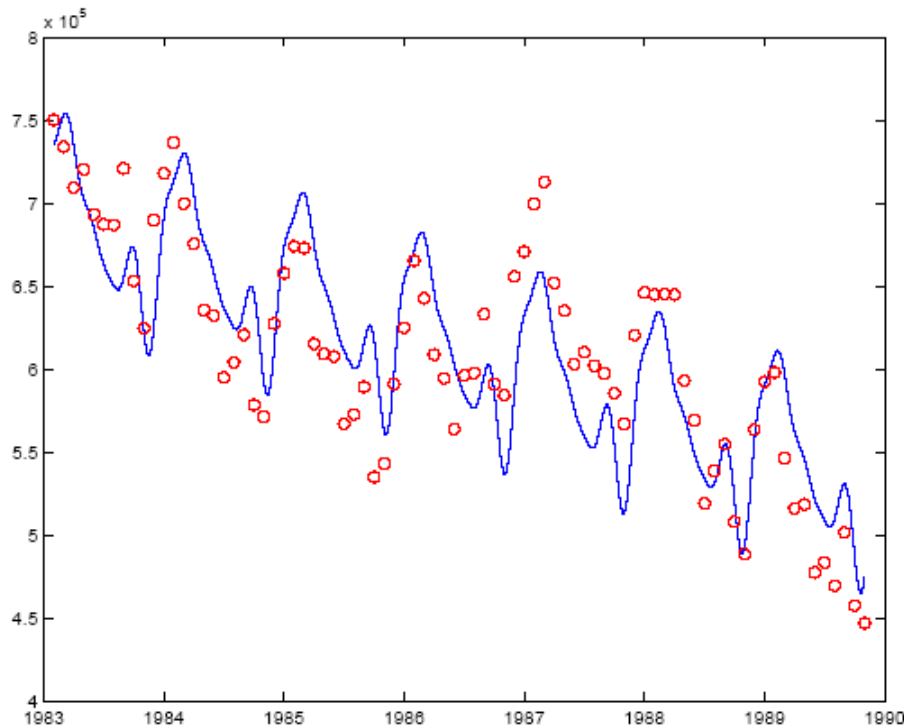
- Za: (Hipel and McLeod, 1994)



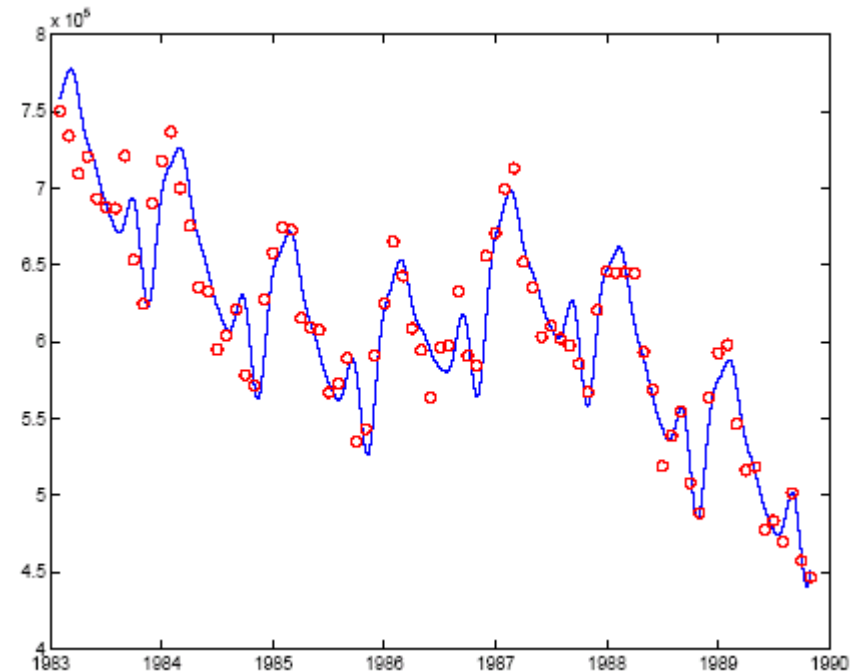
Dekompozycja: trend + wahania

dane o miesięcznym bezrobociu w Australii

- Trend+wahania



Predykcja wartości zmiennych



Uwagi końcowe

Klasyczne metody – dostępne we wszystkich pakietach statystycznych i wielu innych

Wiele innych metod

- Modelowanie z wykorzystaniem wielomianów
- Modele autoregresyjne (ARMA, ARIMA)
- Możliwość wykorzystania zaawansowanej regresji oraz sieci neuronowych

Ponadto

- Szeregi wielo-wymiarowe
- Obecność dodatkowych zmiennych objaśniających

Inne zadania eksploracji szeregów czasowych

Dziękuję za uwagę

Pytania lub komentarze?

Więcej informacji w publikacjach
oraz materiały dodatkowe!

Kontakt:
Jerzy.Stefanowski@cs.put.poznan.pl