

Analiza szeregów czasowych uwagi dodatkowe

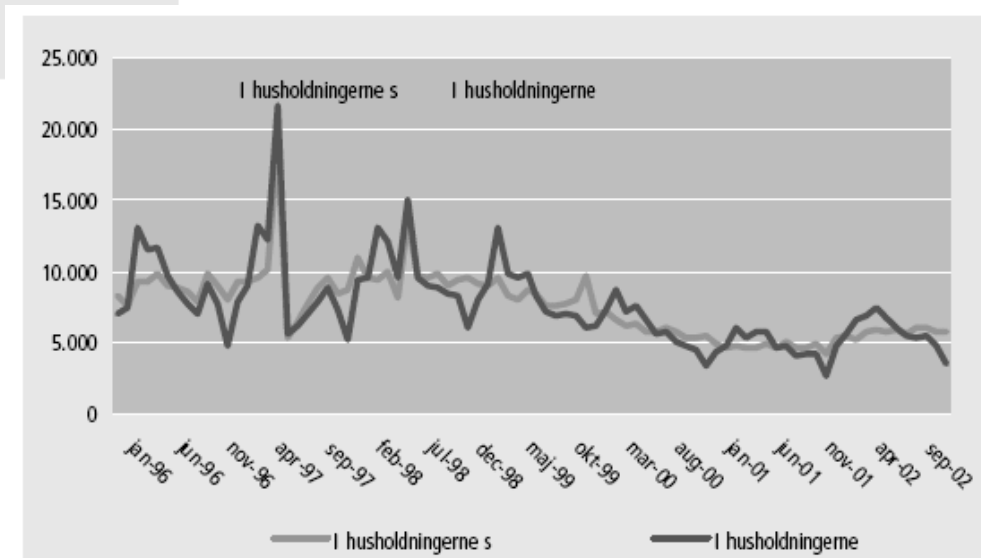
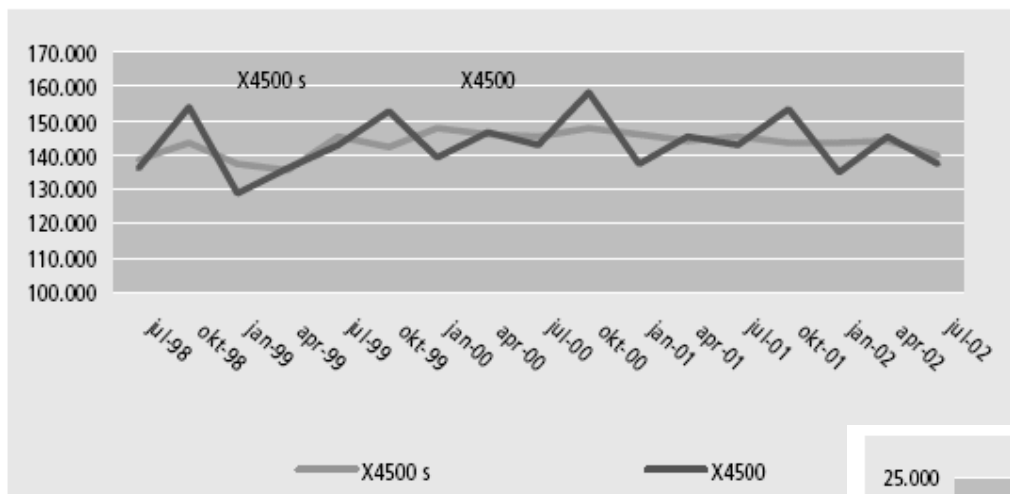
Jerzy Stefanowski

Politechnika Poznańska

Zaawansowana Eksploracja Danych

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

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

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
- 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:

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

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	...

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

Średnie ruchome

- Stosuj, gdy zaobserwowany w okresie badawczym poziom wartości zmiennej prognozowanej jest względnie stały, z pewnymi niewielkimi odchyleniami przypadkowymi.
- **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)

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)

Dostępność w Statistica

The image displays two overlapping dialog boxes from the Statistica software interface, illustrating the available transformation options for time series data.

Przekształcenia szeregów czasowych (Time Series Transformations) dialog box:

- Przekształć zmienną:** SZEREG_G: Miesięczna liczba pasażerów (w tysiącach)
- Przekształcenia: $x=f(x)$**
 - Dodaj stałą ($x=x+C$) C= -104
 - Potęga ($x=x^{**}C$) C= 2,00
 - Pierwiastek ($x=x^{**}1/C$) C= 2,00
 - Logarytm naturalny ($x=\ln(x)$)
 - Przekształcenie wykładnicze ($x=\exp(x)$)
- Odjęcie średniej ($x=x-M$)** M= 280,3
- Standaryzacja ($x=(x-M)/S$)** M= 280,3 S= 119,97
 - Szacuj średnią i odchylenie standardowe z danych
- Odjęcie trendu ($x=x-(a+b*t)$)** a= 87,653 b= 2,6572
- Autokor. ($x=x-(a+b*x(\text{opóźn}))$)** a= 0, b= 1, opóźn= -1
 - Szacuj a/b z danych
- Wygładzanie**
 - Średnia ruch. N-punkt. N= 2 Poprz. Ważona
 - Mediana ruch. N-punkt. N= 2 Poprzedzające
 - Proste wyrównywanie wykładnicze alfa= ,20

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.= 0
- Residualizacja ($x=x-(a+b*y(\text{opóźn}))$) a= 0, b= 1, opóźn= -1
 - Szacuj a/b z danych

Przesunięcie

- Przesuń szereg
- Przesuń szereg

Filtrowanie i sumowanie

- Filtr 4253H
- Różnicowa
- Sumowanie

Transformations of Variables: Series_G.sta dialog box:

- OK (Transform selected series)
- Table with columns: Lock, Variable, Long variable (series) name
- Row 1: L, SERIES_G, Monthly passenger totals (in 1000's)
- Options: Cancel, Options
- Number of backups per variable (series): 5
- Buttons: Save variables, Delete
- Navigation: x=f(x) | Smoothing | x=f(x,y) | Shift | Difference, integrate | Fourier | Review & plot | Autocorrs | Descriptives
- OK (Transform selected series)
- Transformation options:
 - N-pts mov. averg. N= 2 Prior Weighted Specify weights
 - N-pts mov. median N= 2 Prior
 - Simple exponential alpha= ,20
 - 4253H Filter

Table 1. The Japanese Yen / US Dollar Rate: 1983Q1 - 1998Q4

Period	Actual	1-Q RW	3-Q MA	3-Q MAF	XS($\alpha=0.8$)	SE_RW
Mar-83	239.3	Missing	Missing	Missing	Missing	
Jun-83	239.8	239.3	Missing	Missing	Missing	
Sep-83	236.1	239.8	238.40	Missing	239.70	
Dec-83	232	236.1	235.97	238.40	236.82	16.81
Mar-84	224.75	232	230.95	235.97	232.96	52.56
Jun-84	237.45	224.75	231.40	230.95	226.39	161.29
Sep-84	245.4	237.45	235.87	231.40	235.24	63.20
Dec-84	251.58	245.4	244.81	235.87	243.37	38.19
Mar-85	250.7	251.58	249.23	244.81	249.94	0.77
Jun-85	248.95	250.7	250.41	249.23	250.55	3.06
Sep-85	216	248.95	238.55	250.41	249.27	1085.70
Dec-85	200.6	216	221.85	238.55	222.65	237.16
Mar-86	179.65	200.6	198.75	221.85	205.01	438.90
Jun-86	163.95	179.65	181.40	198.75	184.72	246.49
Sep-86	153.83	163.95	165.74	181.40	168.10	106.50
Dec-86	160.1	153.83	159.23	165.74	156.52	41.86
Mar-87	145.65	160.1	153.13	159.23	159.38	208.80
Jun-87	146.75	145.65	150.83	153.13	148.40	1.21
Sep-87	146.35	146.75	146.25	150.83	147.08	0.16
Dec-87	122	146.35	138.37	146.25	146.50	592.92
Mar-88	124.5	122	130.95	138.37	126.90	6.25
Jun-88	132.2	124.5	126.23	130.95	124.98	59.29
Sep-88	134.3	132.2	130.33	126.23	130.76	4.41
Dec-88	125.9	134.3	130.80	130.33	133.59	70.56
Mar-89	132.55	125.9	130.92	130.80	127.44	44.22
Jun-89	143.95	132.55	134.13	130.92	131.53	129.96
Sep-89	139.35	143.95	138.62	134.13	141.47	21.16
Dec-89	143.4	139.35	142.23	138.62	139.77	16.40
Mar-90	157.65	143.4	146.80	142.23	142.67	203.06
Jun-90	152.85	157.65	151.30	146.80	154.65	23.04
Sep-90	137.95	152.85	149.48	151.30	153.21	222.01
Dec-90	135.4	137.95	142.07	149.48	141.00	6.50
Mar-91	140.55	135.4	137.97	142.07	136.52	26.52
Jun-91	138.15	140.55	138.03	137.97	139.74	5.76
Sep-91	132.95	138.15	137.22	138.03	138.47	27.04
Dec-91	125.25	132.95	132.12	137.22	134.05	59.29
Mar-92	133.05	125.25	130.42	132.12	127.01	60.84
Jun-92	125.55	133.05	127.95	130.42	131.84	56.25
Sep-92	119.25	125.55	125.95	127.95	126.81	39.69
Dec-92	124.65	119.25	123.15	125.95	120.78	29.16
Mar-93	115.35	124.65	119.75	123.15	123.87	86.49
Jun-93	106.51	115.35	115.50	119.75	117.05	78.15
Sep-93	105.1	106.51	108.99	115.50	108.62	1.99
Dec-93	111.89	105.1	107.83	108.99	105.80	46.10
Mar-94	102.8	111.89	106.60	107.83	110.67	82.63
Jun-94	98.95	102.8	104.55	106.60	104.37	14.82
Sep-94	98.59	98.95	100.11	104.55	100.03	0.13
Dec-94	99.83	98.59	99.12	100.11	98.88	1.54
Mar-95	88.38	99.83	95.60	99.12	99.64	131.10
Jun-95	84.77	88.38	90.99	95.60	90.63	13.03
Sep-95	98.18	84.77	90.44	90.99	85.94	179.83
Dec-95	102.91	98.18	95.29	90.44	95.73	22.37

Table 3-1 (continued)

Period	Actual	1-Q RW	3-Q MA	3-Q MAF	XS($\alpha=0.8$)	SE_RW
Mar-96	106.49	102.91	102.53	95.29	101.47	12.82
Jun-96	109.88	106.49	106.43	102.53	105.49	11.49
Sep-96	111.45	109.88	109.27	106.43	109.00	2.46
Dec-96	115.98	111.45	112.44	109.27	110.96	20.52
Mar-97	123.97	115.98	117.13	112.44	114.98	63.84
Jun-97	114.3	123.97	118.08	117.13	122.17	93.51
Sep-97	121.44	114.3	119.90	118.08	115.87	50.98
Dec-97	129.92	121.44	121.89	119.90	120.33	71.91
Mar-98	133.39	129.92	128.25	121.89	128.00	12.04
Jun-98	139.95	133.39	134.42	128.25	132.31	43.03
Sep-98	135.72	139.95	136.35	134.42	138.42	17.89
Dec-98	115.2	135.72	130.29	136.35	136.26	421.07
Mar-99		115.2	Missing	130.29	119.41	
MSE						218.94

Notes:

1-Q RW: 1-quarter random walk process

3-Q MA : 3-quarter moving average

3-Q MAF: 3-quarter moving average forecast

XS($\alpha=0.8$): 1-quarter exponential smoothing with $\alpha =0.8$

SE_RW: Squared errors by using random walk forecast

SE_MA: Squared errors by 3-quarter moving-average forecast

SE_XS: Squared errors by using exponential-smoothing forecast

MSE: Mean squared errors

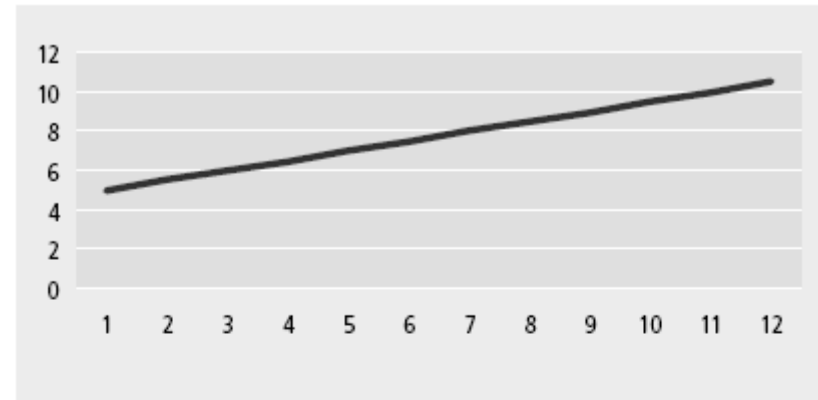
Wskaźniki sezonowości

- Klasyczny model addytywny T+C+S

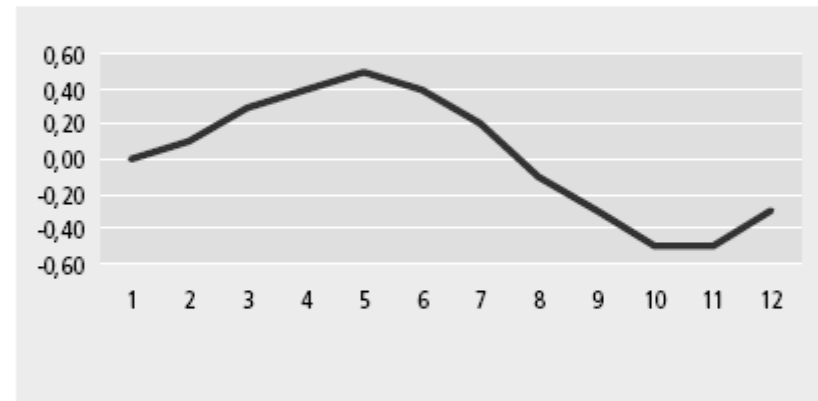
Component series in an additive model

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

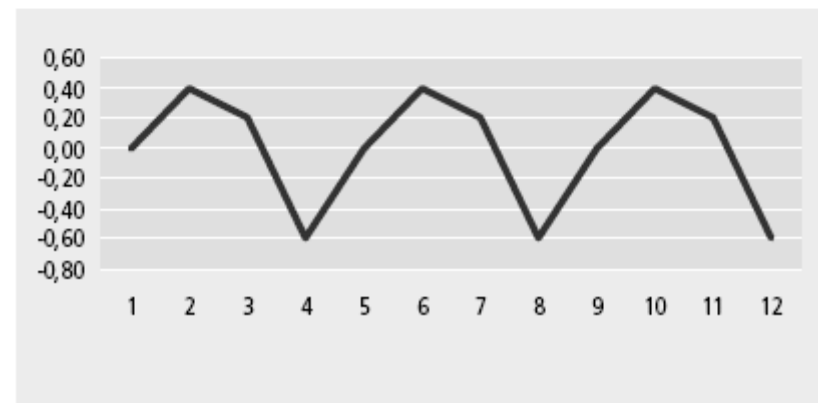
Trend



Cycle



Seasonal



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

Seasonally adjusted figures in the additive model

Year	Quarter	X	Seasonal factor S	Seasonally adjusted series
				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

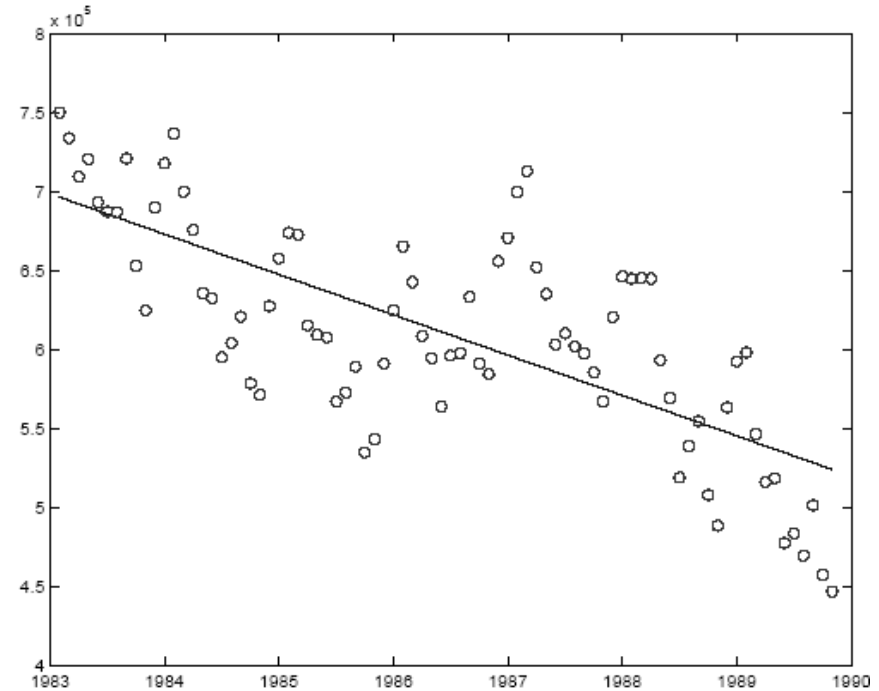
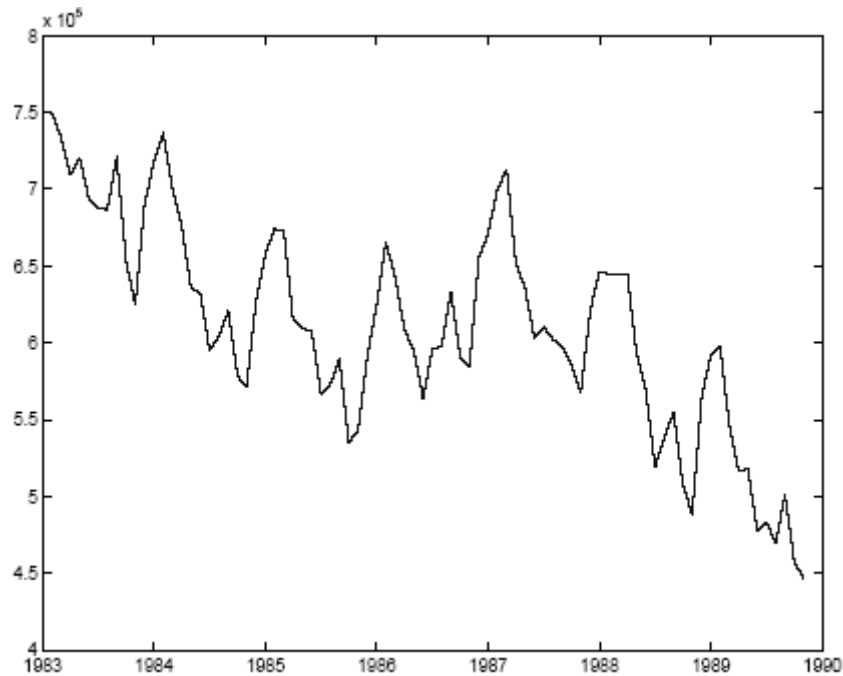
Final estimate of the seasonal factors

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

Dekompozycja: trend + wahania

dane o miesięcznym bezrobociu w Australii

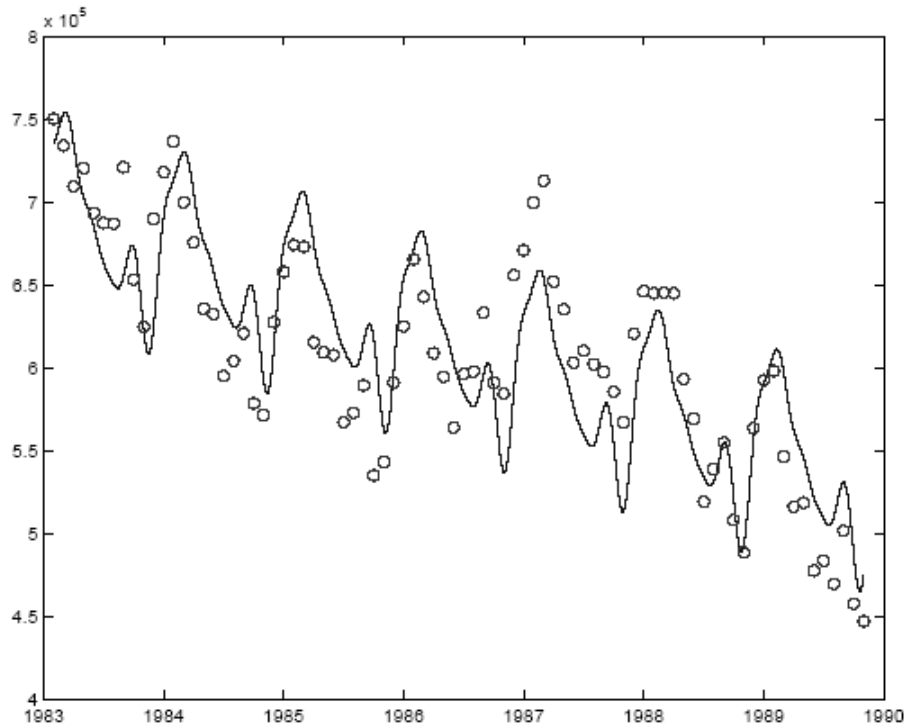
- Za: (Hipel and McLeod, 1994)



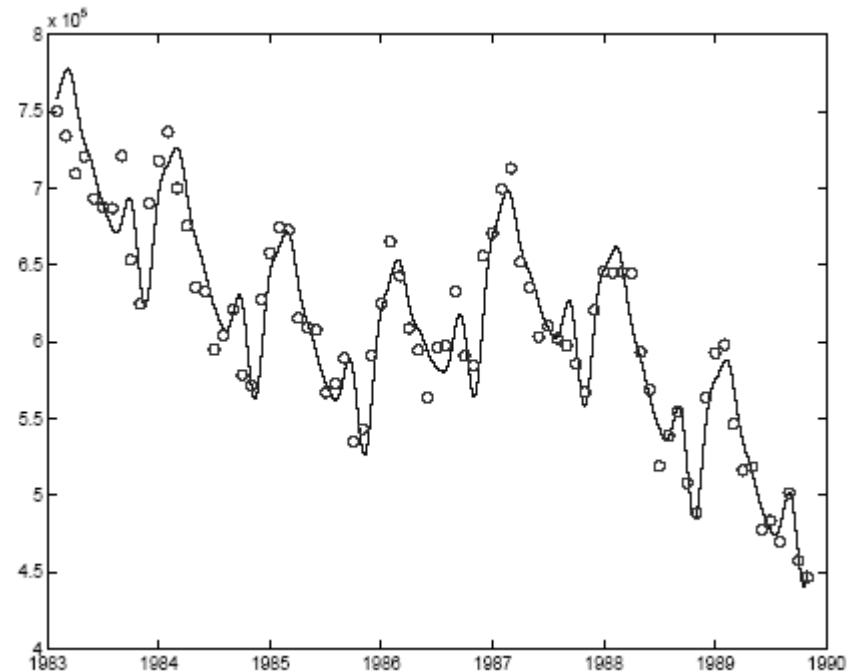
Dekompozycja: trend + wahania

dane o miesięcznym bezrobociu w Australii

- Trend+wahania



Predykcja wartości zmiennych



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

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 T1 nachyleniu w funkcji liniowej trendu, gdyby ją wyznaczać

Dostępność w Statistica

Sezonowe i nesezonowe wyrównywanie wykładnicze

Zabezpiecz Zmienna Długa nazwa zmiennej (szeregu) OK (wykonaj wyrównywanie wykładnicze)

L	SZEREG	G	Długa nazwa zmiennej (szeregu)
			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 zmiennę Usun

Przeglądaj i kreśl zmiennę

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

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

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